2015-12-01

Specialized Replication Operons Control Rhizobial Plasmid Copy Number in Developing Symbiotic Cells

Clarice Lorraine Perry
Brigham Young University

Specialized Replication Operons Control Rhizobial Plasmid Copy Number in Developing Symbiotic Cells

Clarice Lorraine Perry

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

Joel S. Griffitts, Chair
Bill McCleary
Scott Weber

Department of Microbiology and Molecular Biology

Brigham Young University

December 2015

ABSTRACT

Specialized Replication Operons Control Rhizobial Plasmid Copy Number in
in Developing Symbiotic Cells

Clarice Lorraine Perry
Department of Microbiology and Molecular Biology, BYU
Master of Science

The rhizobium – legume symbiosis is a complex process that involves genetic cooperation from both bacteria and plants. Previously, our lab described naturally occurring accessory plasmids in rhizobia that inhibit this cooperation. A transposon mutagenesis was performed on the plasmids to detect the genetic factor that blocked nitrogen fixation. Several of the plasmids were found to possess a replication operon that when disrupted by transposon insertion, restored symbiotic function. This study describes an in-depth investigation into one of those plasmids, pHRC377, and into its replication operon. The operon, which we have called *repA2C2*, comes from the *repABC* family of replication and partitioning systems commonly found in alphaproteobacteria. In this study we show that this operon is not necessary for pHRC377 replication in LB culture or free living cells, but is necessary for plasmid amplification in the plant, specifically during rhizobial differentiation into nitrogen fixing bacteroids. We also show how the other *repABC* type operons on pHRC377 function in relation to plasmid maintenance and copy number during endoreduplication and how they do not have the same phenotypic effect as *repA2C2*.

Keywords: *Sinorhizobium meliloti*, *Medicago truncatula*, *repABC*, symbiosis, nitrogen fixation

ACKNOWLEDGEMENTS

I would like to first and foremost thank my long time mentor and committee chair Joel Griffitts. I would also like to thank the other members of my committee Bill McCleary and Scott Weber for their excellent suggestions and support. It would be a major blunder not to mention the support of the Griffitts lab and especially Skip Price without whom I would have been lost. Finally I would like to thank my family and my husband for their support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Nitrogen Fixation

It has been estimated that 60-65% of the earth's atmospheric nitrogen is produced by biological processes (Zahran 1999, Newton 2000). These biological processes include the ability of some prokaryotic organisms to "fix" nitrogen or take biologically inert N_2 and convert it into biologically useful NH_4^+. For decades farmers have used chemically fixed nitrogen in agriculture because of its ease of use. Nitrogen is chemically fixed for nitrogen fertilizer through the Haber-Bosch process. Chemical fixation is environmentally taxing both in production and application, with up to 50% of nitrogen fertilizer being leached into our water supplies. Nitrogen leaching can lead to severe pollution problems. Biological nitrogen fixation is a more economically sound and environmentally friendly way to grow crops, but much has yet to be discovered to apply biological nitrogen fixation to large-scale agriculture.

Prokaryotes, spread across a wide range of archaea, bacteria and cyanobacteria, are known to fix nitrogen using an enzyme called nitrogenase. Some of the most agriculturally significant of these organisms are known as rhizobia - *Azorhizobium*, *Bradyrhizobium*, *Mesorhizobium*, *Rhizobium*, and *Sinorhizobium*. These alphaproteobacteria form symbiotic relationships with legumes (beans, peas, soybeans, alfalfa, clover, etc). The physical manifestation of this symbiotic relationship is nodules that form on the roots of the legumes (Figure 1). Rhizobia colonize their plant hosts intracellularly using *nod* (nodulation) genes to induce nodule development and *nif* (nitrogen fixation) genes to reduce N_2 (Masson-Boivin, Giraud et al. 2009). One of the best studied of the rhizobia – legume symbioses is the relationship between *Sinorhizobium meliloti* and *Medicago truncatula* (Gage 2004, Mergaert,

Uchiumi et al. 2006, Jones, Kobayashi et al. 2007, Kereszt, Mergaert et al. 2011, Oldroyd, Murray et al. 2011, Haag, Arnold et al. 2013).

Figure 1. Medicago truncatula and nodules. Images taken 28 days after inoculation. The pink nodules indicate symbiotic compatibility resulting in nitrogen fixation.

S. meliloti as a Model Organism

S. meliloti has been used as a model organism for rhizobia-legume symbiosis for many years. Its genome sequence was first published in *Science* in 2001. The strain that was sequenced, Rm1021, is one of the most common lab strains of *S. meliloti* in use today (Meade, Long et al. 1982). The three papers detailed *S. meliloti*'s three part genome: a single chromosome (3,654 kb), and two megaplasmids, pSymA (1,354kb), and pSymB (1,683kb) (Barnett, Fisher et al. 2001, Finan, Weidner et al. 2001, Galibert, Finan et al. 2001). The pSymA and pSymB megaplasmids are considered secondary chromosomes because they are large and stably maintained. These megaplasmids have been important for nitrogen fixation symbiosis research. Nod factor-based communication was discovered on pSymA and it also contains all the nitrogen

fixation genes used in symbiosis (Faucher, Maillet et al. 1988). pSymB also encodes genes essential for symbiosis such as gene clusters involved in exopolysaccharide synthesis (Cangelosi, Hung et al. 1987). Up until November of 2015 there have been 2861 papers linked to the word '*meliloti*' on PubMed. Despite all that is known from *S. meliloti* about symbiosis there is much yet to be discovered, especially concerning the later stages of symbiotic development.

M. truncatula as a Model Organism

Although *M. truncatula* has been studied for many years, especially in its relationship with *S. meliloti,* its genome was not fully sequenced until 2011 (Young, Debelle et al. 2011). It was found through this sequencing project that a whole-genome duplication had occurred approximately 58 million years ago. It was postulated that this event allowed for the genetic development that led to symbiotic relationships. The strain that was sequenced in 2011 was cultivar A17, which along with another common lab cultivar A20 come from the Jemalong series (Penmetsa and Cook 2000). *M. truncatula* is readily nodulated by *S. meliloti*. It is typically used in association with *S. meliloti* because of its small genome (500-600 Mpbs), fast reproductive cycle, and natural genetic diversity (Thoquet, Gherardi et al. 2002). Also, *M. truncatula* is self-fertilizing which makes it amenable to genetic analysis.

S. meliloti - M. truncatula Symbiosis

When an *S. meliloti* bacterial cell senses that an *M. truncatula* plant root is secreting flavonoids nearby, it turns on Nod factor production. Nod factors are lipochitooligosaccharides that contain different functional groups that can be recognized by different plants. When a plant recognizes compatible Nod factors it corrals the bacterial cell and its descendants into a complex network of microscopic conduits, termed infection threads, through which the bacteria infiltrate

the developing nodule tissue. Once this network is developed enough, thousands of bacterial cells are endocytosed into plant nodule cells (Figure 2).

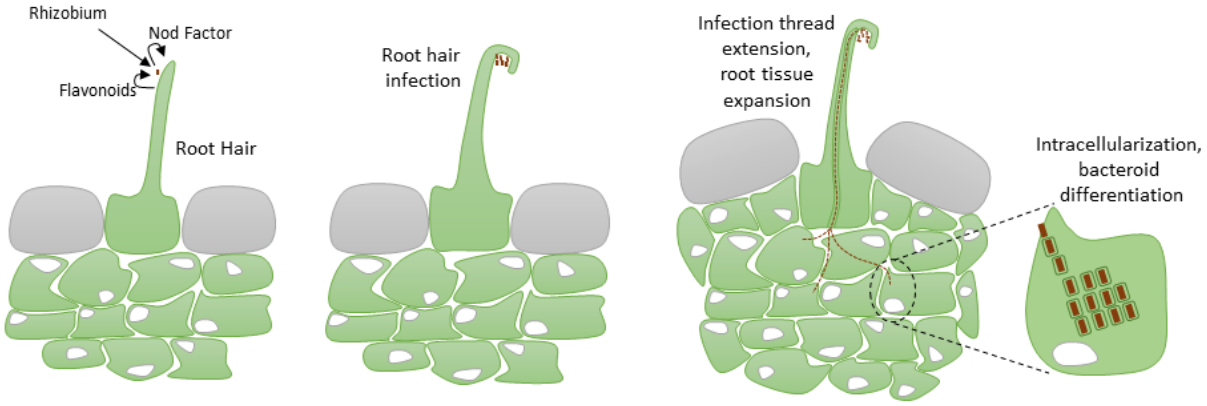

Figure 2. S. meliloti – M. truncatula symbiosis. Plant root hairs excrete flavonoids into the soil which are sensed by rhizobia. Rhizobia move towards the root hairs and in turn excrete Nod factors. When the root recognizes compatible Nod factors the root hair curls and the plant cells become permissible to bacterial infection. The rhizobia colonize an infection thread down the root hair and into the cortex cells. This prompts the cortex cells to divide forming the outgrowth that becomes a nodule. The rhizobia are endocytosed into plant cells where they undergo further development before fixing nitrogen.

Every endocytosed bacterial cell is surrounded by a plant membrane and the resulting structure is called a symbiosome. Plant root cells produce hundreds of different nodule-specific cysteine-rich (NCR) peptides in response to cell invasion by the bacteria (Mergaert, Nikovics et al. 2003, Maunoury, Redondo-Nieto et al. 2010, Nallu, Silverstein et al. 2014, Penterman, Abo et al. 2014). The type and amount of NCRs vary between plant species. NCR peptides are delivered to the symbiosome through the plant secretory system, where they cause developmental changes in the bacteria. This involves endoreduplication of the bacterial genome and enlargement of the bacterial cells without division (Van de Velde, Zehirov et al. 2010, Haag, Baloban et al. 2011, Penterman, Abo et al. 2014). Once the bacteria fully develop, they are called bacteroids (Figure 3). Bacteroids are terminally differentiated, meaning they can no longer proliferate (Mergaert, Uchiumi et al. 2006, Van de Velde, Zehirov et al. 2010). They fix nitrogen for the plant and in

return the plant feeds the bacteroids carbon from photosynthesis. Proliferating *S. meliloti* can still be isolated from the nodule because some remain undeveloped in infection threads. These free-living bacteria benefit the most from symbiosis because they have the ability to proliferate after nodule senescence.

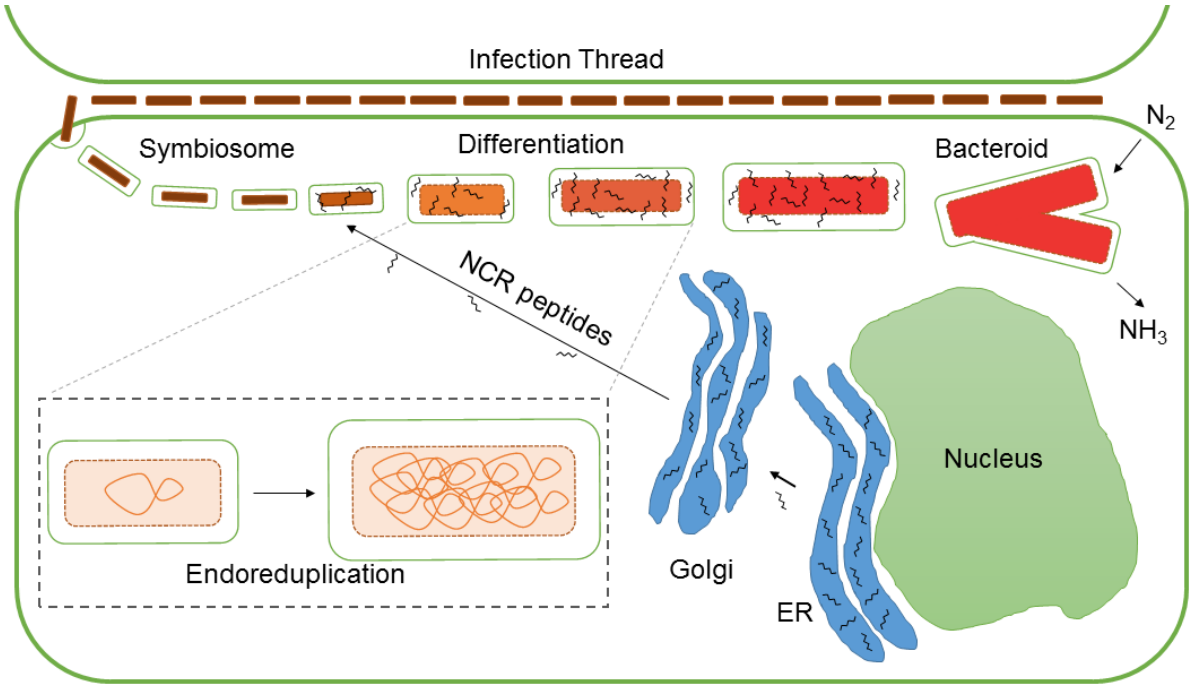

Figure 3. Bacteroid differentiation. Rhizobia are endocytosed into the plant cells from the infection thread to form membrane bound symbiosomes. Plant secretory pathways shuttle NCR peptides to the symbiosomes which cause the bacteria to undergo developmental changes including, but not limited to, many rounds of endoreduplication, cell enlargement without division, and activation of nitrogen fixation genes.

Symbiotic Incompatibility

There are many molecular checkpoints that both the bacteria and plant have to go through to form a symbiotic relationship. For example, the bacteria must produce the right Nod factor which will signal the plant to allow the bacteria inside the nodule (Wais, Keating et al. 2002). But Nod factor recognition is only part of the story. There have been studies where rhizobia are engineered to make a new Nod factor that would theoretically allow it to interact with new plant hosts. This type of engineering usually results in the formation of nodules on previously

incompatible plant hosts, but the resulting nodules are almost always unable to fix nitrogen (Roche, Maillet et al. 1996, Barran, Bromfield et al. 2002). There have been many observations of strains that induce Fix+ (nitrogen-fixing) nodules on some plant hosts, but Fix- (non-nitrogen fixing) nodules on others (Van Berkum, Elia et al. 2006, Kereszt, Mergaert et al. 2011, Crook, Lindsay et al. 2012). This is interesting because the bacteria have all the genes necessary to fix nitrogen for the plant, but because of some incompatibility at a late-stage in development they will not. Little is known about the molecular basis of this late-stage incompatibility.

Nitrogen Fixation-Blocking Plasmids

In 2012 the Griffitts lab published a paper describing wild isolates of *S. meliloti* that form ineffective symbiotic relationships with certain *Medicago* genotypes (Crook, Lindsay et al. 2012). Numerous *S. meliloti* isolates from the United States Department of Agriculture collection were grown on different plant hosts to test their host range. Host range is an indication of which plants an isolate of *S. meliloti* will fix nitrogen for (Fix^+, large pink nodules) and which plants it is incompatible with (Fix^-, small white nodules). All bacterial strains reported could fix nitrogen for *Medicago lupulina* and *Medicago italica,* but the results were mixed when the strains were paired with cultivars of *Medicago truncatula*, A17 and A20 (See Table 1). Most strains could not fix nitrogen for these two cultivars, but it was found that the symbiotic incompatibility with A20 and A17 could be alleviated spontaneously when pink nitrogen fixing nodules were discovered alongside the white non-fixing nodules. The spontaneous mutant strains inhabiting the pink nodules were isolated and called gain of compatibility (GOC) mutants. It was predicted that a genetic element must ultimately control the nitrogen fixation blocking phenotype because the trait was heritable and the bacteria did not revert back to a less permissive host range.

Table 1. Host range of certain S. meliloti strains.

Strain*	LU	IT	PR	A17	A20
B464	Fix^{+}	Fix^{+}	Fix^{+}	Fix^{+}	Fix^{+}
C017	Fix^{+}	Fix^{+}	Fix^{-*}	Fix^{-*}	Fix^{-*}
B469	Fix^{+}	Fix^{+}	Fix^{-*}	Fix^{-*}	Fix^{-*}
B800	Fix^{+}	Fix^{+}	Fix^{-}	Fix^{+}	Fix^{-*}
C377	Fix^{+}	Fix^{+}	Fix^{-*}	Fix^{-}	Fix^{-*}

*Strain-host compatibility as indicated by nitrogen fixing (Fix^{+}) and non-fixing (Fix^{-}) nodules 30 days after inoculation. Fix^{-*} indicates events of gain of compatibility (GOC) mutants. LU = *Medicago lupulina*; IT = *M. italica*; PR = *M. praecox*; A17 = *M. truncatula* cv. A17; and A20 = *M. truncatula* cv. Reprinted with permission (Crook, Lindsay et al. 2012).

To discover the genetic cause of the nitrogen fixation blocking phenotype, large scale sequencing was done on one of the bacterial strains C017 and its GOC derivative. When the sequence reads were mapped back to the genome of the lab strain Rm1021, it was found that large sections of DNA were missing in the GOC strain. This missing DNA was similar to DNA found in a known *S. meliloti* accessory plasmid. To test whether or not a similar accessory plasmid could be present in C017 a special electrophoretic gel technique (Eckhardt electrophoresis) was used. It was found that the C017 GOC derivative was indeed missing a large accessory plasmid about 300 kb in size. It was then shown in every single GOC derivative, for every wild-type strain tested, a large accessory plasmid was missing. Re-introduction of accessory plasmids from parent strains into GOC derivatives also showed a reversion to the original Fix^{-} phenotype. These plasmids are now referred to as pHR (plasmid affecting Host Range).

It was concluded that a genetic factor(s) on these plasmids prevented the strains from being able to fix nitrogen on certain hosts and that it must be connected to the termination of late stage development that had been observed by other researchers (Van Berkum, Elia et al. 2006, Kereszt, Mergaert et al. 2011). Late-stage incompatibility was shown because bacteria harboring

pHRs would still grow in infection threads forming small nodules and even endocytose into the plants cells, but then they fail to fix nitrogen.

While the overall genetic cause of the nitrogen fixation blocking phenotype was linked to the accessory plasmids, the specific factors on each of the plasmids was not discovered until recently. A new chapter to this story was published in 2015. This paper focused on the strain B800 and its accessory plasmid pHRB800. A transposon mutagenesis was performed on pHRB800 to establish the genetic basis of the nitrogen blocking factor. This process involved inserting a selectable marker onto the plasmid and an origin of transfer (*oriT*) for plasmid conjugation. The plasmid was moved to *Agrobacterium* where it was mutagenized with transposon insertions. The plasmid was then moved back into its respective GOC strain. The resulting mutagenized strains were tested on plants and any Fix+ nodules were collected. The transposon insertions were confirmed in Fix+ nodule bacteria by arbitrary PCR and the place of the insertions was sequenced (Price, Tanner et al. 2015).

Using the results from the transposon mutagenesis, the nitrogen fixation blocking factor of pHRB800 from the bacterial strain B800 was characterized. It was shown that an M16 peptidase encoded by pHRB800 prevents plant-stimulated differentiation of rhizobial cells. Specifically this peptidase appears to degrade the NCR peptides that the plant transports inside symbiosomes (Figure 3). If the NCR peptides are degraded, functional bacteroids never form (Van de Velde, Zehirov et al. 2010, Wang, Griffitts et al. 2010, Farkas, Maroti et al. 2014, Penterman, Abo et al. 2014, Horvath, Domonkos et al. 2015).

This peptidase, however, was not found on the other pHR plasmids. In similar mutagenesis screens done on pHRC377, pHRC017, and pHRB469 there was a different outcome that is not so easily explained. For these three plasmids, most of the transposon insertions preventing the fixation-blocking phenotype are found in a conserved *repABC*-type operon that we refer to as *repA2C2*. A smaller number of insertions have been identified in a pair of genes that may encode transcriptional regulators (Figure 4).

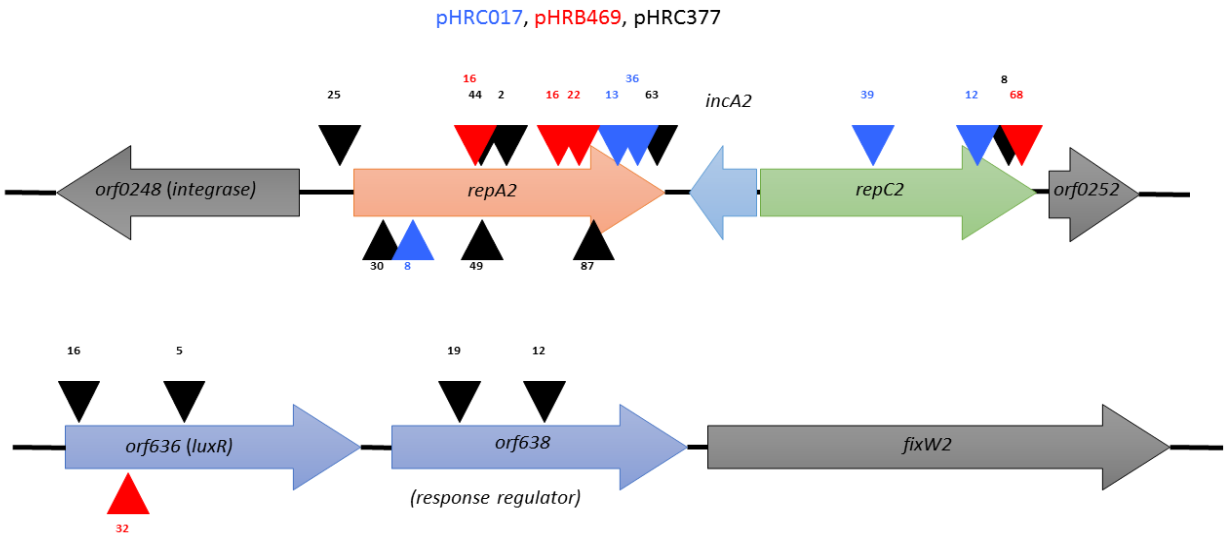

Figure 4. Himar 1 transposon insertions affecting the nitrogen fixation blocking phenotype. Pictured are the relative insertion sites for the mapped transposon insertions into pHRC017, pHRB469, and pHRC377. 29 insertions were detected in *repA2*, 7 in *repC2*, 2 in *luxR*, and 2 in a response regulator next to *luxR*. See Table 6 (Supplementary) for transposon insertion strains.

S. meliloti Genome and *repABC*-type Replication Operons

Earlier in the introduction it was noted that *S. meliloti* has a tripartite genome with a chromosome and two megaplasmids important for symbiosis: pSymA and pSymB. The *S. meliloti* chromosome is replicated by an *oriC*/DnaA mechanism; however, pSymA, pSymB, and most accessory plasmids found in wild isolates are *repABC* family plasmids (Galibert, Finan et al. 2001). The *repABC* plasmids are very common in alphaproteobacteria. Generally, plasmids whose replication is controlled by *repABC* operons are larger in size, very stable, and have a

copy number of one. Megaplasmids with *repABC* operons, such as pSymA and pSymB that have important symbiosis or housekeeping genes tend to have a similar GC content to the chromosome of their respective bacterial host (Cevallos, Cervantes-Rivera et al. 2008, Pinto, Pappas et al. 2012). It is thought that the longer a *repABC* family plasmid has been with its host evolutionarily, the more similar the GC content is.

Diversity and operation of *repABC* operons allows for many incompatibility groups among *repABC* plasmids (Petersen, Brinkmann et al. 2009, Mazur and Koper 2012, Zebracki, Koper et al. 2015). Incompatibility groups are formed when plasmids sharing the same host have different enough replication machinery to not interfere with each other during plasmid segregation into daughter cells or during DNA replication. For example *Rhizobium leguminosarum* biovar *viciae* strain 3841 has six plasmids that are all a part of the *repABC* family (Young, Crossman et al. 2006). The specific incompatibility factors are thought to be contained in binding differences of RepA and B and the counter transcribed RNA that is encoded between *repB* and *C*.

A typical *repABC* operon is transcribed as a polycistronic message from the promoter region upstream of *repA*. The main differences between *repABC* operons include transcriptional regulatory elements, the number and position of *par*-sites, and the presence of peptide encoding minigenes (Cevallos, Cervantes-Rivera et al. 2008, Pinto, Pappas et al. 2012) (Figure 5). For the purpose of this introduction, focus will be placed on the most basic genetic structure of a *repABC* operon: *repAB*, *parS* sites, *repC*, and the small countertranscribed RNA hereafter referred to as ctRNA.

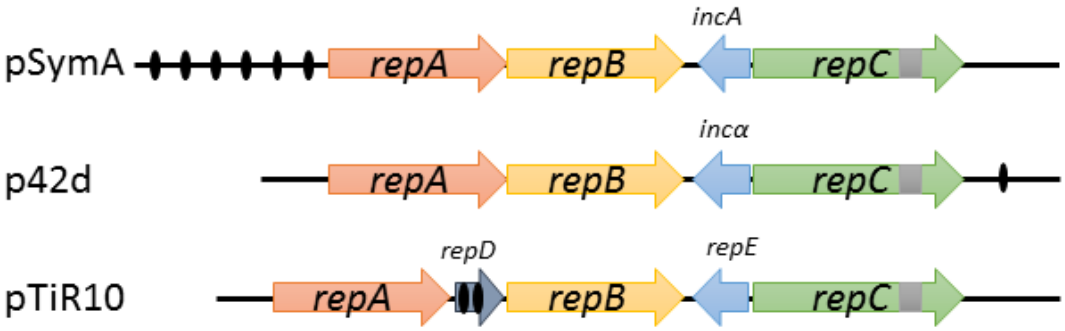

Figure 5. Different operons from the repABC family. pSymA of *S. meliloti*, p42d of *Rhizobium etli* str. CFN42, and pTiR10 of *Agrobacterium tumefaciens*. AT-rich regions that are believed to contain the plasmid origin of replication are shown in grey. Arrows immediately upstream of *repC* represent counter transcribed RNAs (ctRNA); in the case of pSymA this ctRNA is called *incA*. Black ovals represent *parS* sites.

The Function of RepA and RepB Proteins

The RepA and B proteins are part of a large family of partitioning proteins found in many bacteria. They are thought to relate most closely to the Type I systems of plasmid partitioning because RepA acts as a Walker-type ATPase (commonly known as ParA) and RepB (commonly known as ParB) acts as its partner in forming the partitioning complex.

RepA has dual functionality: it binds the operator of a *repABC* operon to repress *repA* transcription and it facilitates plasmid partitioning (Pappas and Winans 2003, Perez-Oseguera and Cevallos 2013, Zebracki, Koper et al. 2015) (Figure 6). The DNA binding motif used by RepA is predicted to have a helix-loop-helix much like a ParA protein (Dunham, Xu et al. 2009). It can bind specifically or nonspecifically to DNA sequences depending on whether or not ADP or ATP is present. It can also form dimers and filaments depending on its association with ATP or ADP. As a Walker-type ATPase, RepA, has been postulated to move plasmids by one of two mechanisms: cycles of polymerization and de-polymerization as it interacts with the plasmid DNA or by a concentration gradient of dimers that forms in the nucleoid (Kiekebusch and Thanbichler 2014).

The partitioning complex comes together when RepB dimers bind to a centromere-like 16-nucleotide palindromic sequences called *parS* (Figure 6). The *parS* site(s) can be located upstream, within, or downstream of the *repABC* operon. The first RepB dimer serves to nucleate the binding of more dimers. RepB stabilizes the ATPase action of RepA causing it to bind tighter to the *repA* operator. This interaction between RepB and RepA and the plasmid DNA eventually leads to proper plasmid separation into daughter cells (Cevallos, Cervantes-Rivera et al. 2008, Pinto, Pappas et al. 2012, Kiekebusch and Thanbichler 2014).

The Function of RepC and ctRNA

The RepC protein is the only protein that is absolutely necessary for a *repABC* operon to support plasmid replication (Tabata, Hooykaas et al. 1989, Ramirez-Romero, Tellez-Sosa et al. 2001). There have even been examples of some naturally occurring plasmids that replicate with only *repAC* or *repC* genes present (Bartosik, Bialkowska et al. 1997, Bartosik, Wlodarczyk et al. 1997, Izquierdo, Venkova-Canova et al. 2005, Young, Crossman et al. 2006, Mazur and Koper 2012, Perez-Segura, Perez-Oseguera et al. 2013). The RepC protein is the initiator of replication and acts on an origin of replication (*oriV*) found within the *repC* open reading frame. This origin is thought to be localized to AT-rich region found in all *repC* genes (Cervantes-Rivera, Pedraza-Lopez et al. 2011, Pinto, Flores-Mireles et al. 2011). Having the origin inside the *repC* gene also adds another level of regulation to the *repABC* operon because it cannot be transcribed if RepC is bound. It has been hypothesized and some evidence has shown that *repC* acts only in *cis* (Pinto, Flores-Mireles et al. 2011, Pinto, Pappas et al. 2012). For example, in Pinto, Flores-Mireles et al. 2011, they showed that over-expression of RepC in *Agrobacterium tumefaciens* caused an increase of copy number in *cis* but not in *trans*. There may be several reasons how this could happen, including low protein expression and poor diffusion of RepC within the cell. It also

might be linked to the fact that usually there are several plasmids of the *repABC* family in a single cell and sequestering RepC would help prevent replication incompatibility (Pinto, Flores-Mireles et al. 2011).

Whether by itself or contained in a *repABC* operon, *repC* is almost always accompanied by a counter transcribed RNA (ctRNA) gene which is encoded in the region between *repB* and *C.* This untranslated ctRNA is about 50 nucleotides long and acts as a regulator of *repC.* It is thought to act as a translation repressor (Venkova-Canova, Soberon et al. 2004, Chai and Winans 2005, Cervantes-Rivera, Romero-Lopez et al. 2010). Binding of ctRNA to the repABC transcript gives rise to a stem loop structure that sequesters the Shine-Dalgarno sequence of *repC*. When the ctRNA is not bound to the transcript the *repABC* transcript folds differently allowing *repC* to be translated. This ctRNA is part of the incompatibility system of *repABC* operons because it can work in *trans*. It has been shown that when two plasmids have similar ctRNA genes they cannot coexist in the same bacterium (Venkova-Canova, Soberon et al. 2004, MacLellan, Smallbone et al. 2005). Most recently it was discovered that the ctRNA must be expressed highly to exhibit incompatibility (Yip, Ding et al. 2015).

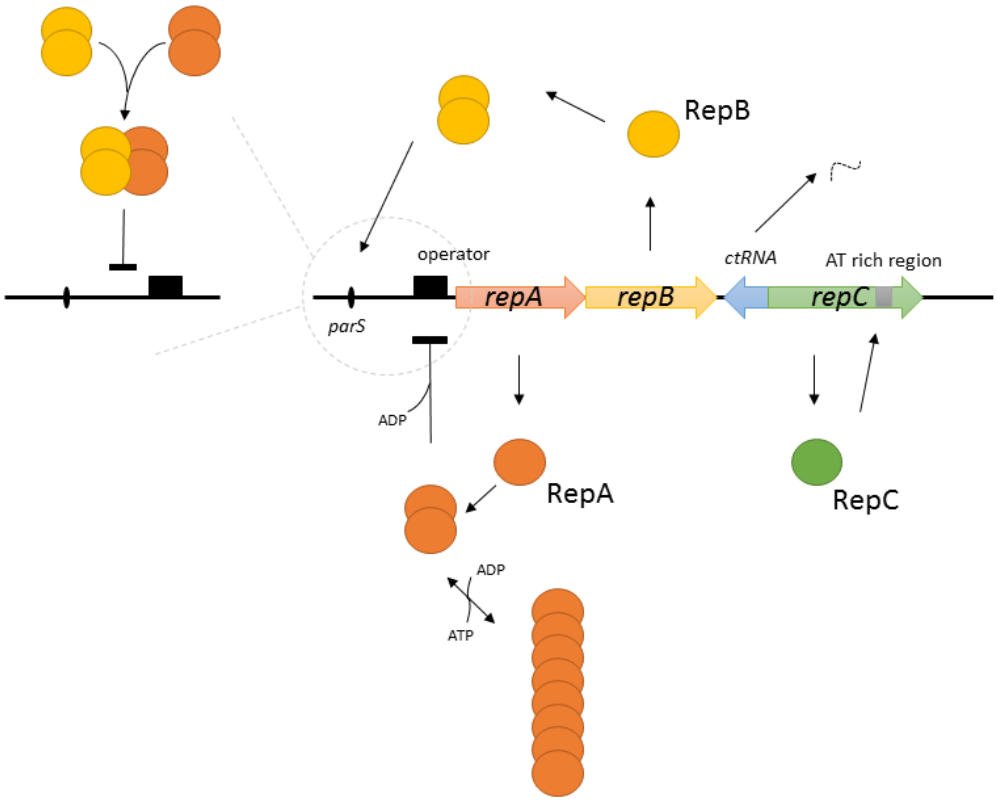

Figure 6. Transcription and regulation of the repABC operon. Each *repABC* operon is transcribed as a polycistronic message from the promoter region upstream of *repA*. Repression of *repABC* happens when RepA-ADP binds to the operator and when both RepA and B complex together and bind to the operator region. RepA has the ability to dimerize or oligomerize depending on whether it is bound to ADP or ATP. The ctRNA will bind to the mRNA transcript of *repABC* upstream of *repC* causing the Shine-Dalgarno of *repC* to be sequestered preventing translation.

RESULTS

Sequencing of pHRC377 was done as previously shown for pHR plasmids pHRC017 (Crook, Lindsay et al. 2012) and pHRB800 (Price, Tanner et al. 2015). This resulted in a circular molecule 188,525 bps long with ~300 predicted gene products. Around 61.5% of the gene products (partial or full length) matched to known proteins in the NCBI database while 31.5% matched to hypothetical proteins and 7% had no blast results. Of the 61.5% of gene products with predictable functions, about 19% were related to transposases and 11% were transcriptional regulators. Plasmid mobility genes *traG* and *traA* were found. Almost all gene products had homologs in *S. meliloti* or other rhizobial species. The pHRC377 plasmid also was found to have one complete *repABC* operon (*repA1B1C1*), and several incomplete operons, *repA2C2*, *repA3* (full length), *repC3* (partial), and *repB4C4* (*repB4* partial *C4* full length) (**Figure 7**).

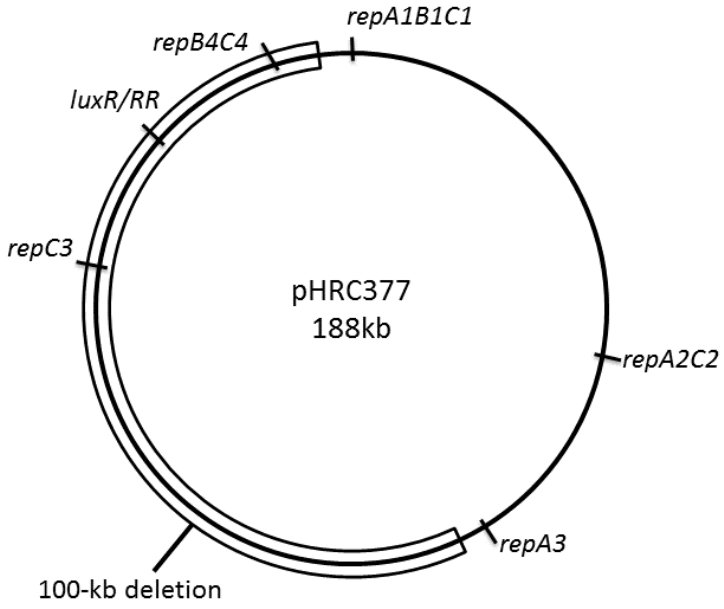

Figure 7. Genetic map of pHRC377. All full and partial replication operons are highlighted. The *repA1B1C1* operon, *repA2C2*, *repA3*, *repC3* (only the C terminal domain), and *repB4* (partial) *repC4*. Also highlighted are the *luxR* like gene and its hypothesized response regulator gene (*luxR/RR*) which were found to affect the nitrogen fixation blocking phenotype.

The pHRC377 accessory plasmid blocks nitrogen fixation on *M. truncatula* accessions A17 and A20 as previously shown (Table 1). A plasmid-specific transposon mutagenesis of pHRC377 was performed to identify plasmid mutants that no longer block nitrogen fixation. This mutagenesis resulted in several mutants with full or partial ability to fix nitrogen for plants (A17 and A20) that they were previously incompatible with. Transposon insertions appeared in *repC2* resulting in phenotypically normal nodules and healthy plants, and in *repA2* resulting in plants that were healthier than uninnoculated plants but not as healthy as an insertion in *repC2*. Additionally, transposon insertions in a pair of contiguous genes encoding putative transcriptional regulators also led to a Fix^+ phenotype, but these mutants were also not as strongly Fix^+ as the *repC2* mutants. This was again evident because plants inoculated with these strains were not as green as normal Fix^+ plants. One of these two transcriptional regulators is in the LuxR family, and the other is in the response regulator family (hereafter both are referred to as *luxR/RR*) (Figure 4). They are part of a three gene operon that includes a hydrogen peroxidase that was not found in the transposon mutagenesis. Also during the mutagenesis screen, a mutant of pHRC377 that had a ~100 kb deletion was found. This deletion encompassed the region of the plasmid with *luxR/RR* genes, but not *repA2C2*. This smaller plasmid gave a full symbiotic compatibility. This information led us to create the model that pHRC377 (along with other plasmids pHRC017 and pHRB469 shown in Figure 4) does not rely on *repA2C2* for the actual cause of the nitrogen fixation blocking phenotype. There must be another factor contained in the 100kb deletion that blocks nitrogen fixation. Rather, loss of the function of *repA2C2* and most importantly *repC2* must affect amplification of pHRC377 which in turn would affect whether or not the nitrogen fixation blocking factor is carried on to later generations of cells or perpetuated

during endoreduplication in the symbiosome. (See Table **2** for a list of the strains used in this paper organized by replication operon).

Table 2. Strains used in experiments sorted by replication operon.

Strains*	Strain type	Fixation	Resistance
Base strains			
C377	contains unaltered pHRC377	Fix $^-$	Sm^R
C378	C377 without pHRC377 (GOC)	Fix^+	Sm^R
C389	pHRC377 marked with Km^R	Fix $^-$	Sm^RKm^R
repA1B1C1			
CP37	C377 disruption loop-in in *repC1*	N/A	Sm^RKm^R
CP38	C377 non-disruption loop-in *repC1*	N/A	Sm^RKm^R
PP539	C378 minimal plasmid with *repA1C1B1*	Fix^+	Sm^RKm^R
repA2C2			
CP03	C377 transposon insertion in *repA2*	weak Fix^+	Sm^RKm^R
CP05	C377 transposon insertion in *repC2*	Fix^+	Sm^RKm^R
CP23	C377 clean deletion of *repA2C2*	Fix^+	Sm^RKm^R
PP596	C377 clean deletion of *repA2*	Fix $^-$	Sm^RKm^R
CP28	C377 disruption loop-in in *repC2*	Fix^+	Sm^RKm^R
CP29	C377 non-disruption loop-in in *repC2*	Fix $^-$	Sm^RKm^R
PP538	C378 minimal plasmid with *repA2C2*	Fix^+	Sm^RKm^R
repB4C4			
PP588	C377 disruption loop-in in *repC4*	N/A	Sm^RKm^R
PP589	C377 non-disruption loop-in in *repC4*	N/A	Sm^RKm^R
PP541	C378 minimal plasmid with *repB4C4*	Fix^+	Sm^RKm^R
pSymB repABC			
PP540	C378 minimal plasmid with *repABC*	Fix^+	Sm^RKm^R
luxR/RR			
CP07	C378 transposon insertion in *luxR*	weak Fix^+	Sm^RKm^R

*Strains used in experiments are sorted by which replication operon they effect. The ability to fix nitrogen (fix^+/fix^-) is also indicated if known and antibiotic resistance markers are noted. Km^R always marks modified pHRC377 or is contained on a minimal plasmid. Sm^R = streptomycin resistance. Km^R = kanamycin/ neomycin resistance. See Table 8 (Supplemental) for more information.

To start testing our model of the influence of *repA2C2* on pHRC377 and to confirm its importance for replication, a clean deletion of *repA2C2* was made. This pHRC377 mutant gave the same result as a transposon insertion in *repC2*, restoration of symbiotic compatibility (see Table 2). More interestingly a clean in-frame deletion of *repA2,* which allowed the *repA2C2*

promoter to read directly into *repC2,* did not alleviate the nitrogen fixation blocking phenotype (Figure **8**) indicating that only replication initiation controlled by *repC2* and not plasmid partitioning, which would be controlled by *repA2* is important in the action of *repA2C2* on blocking nitrogen fixation.

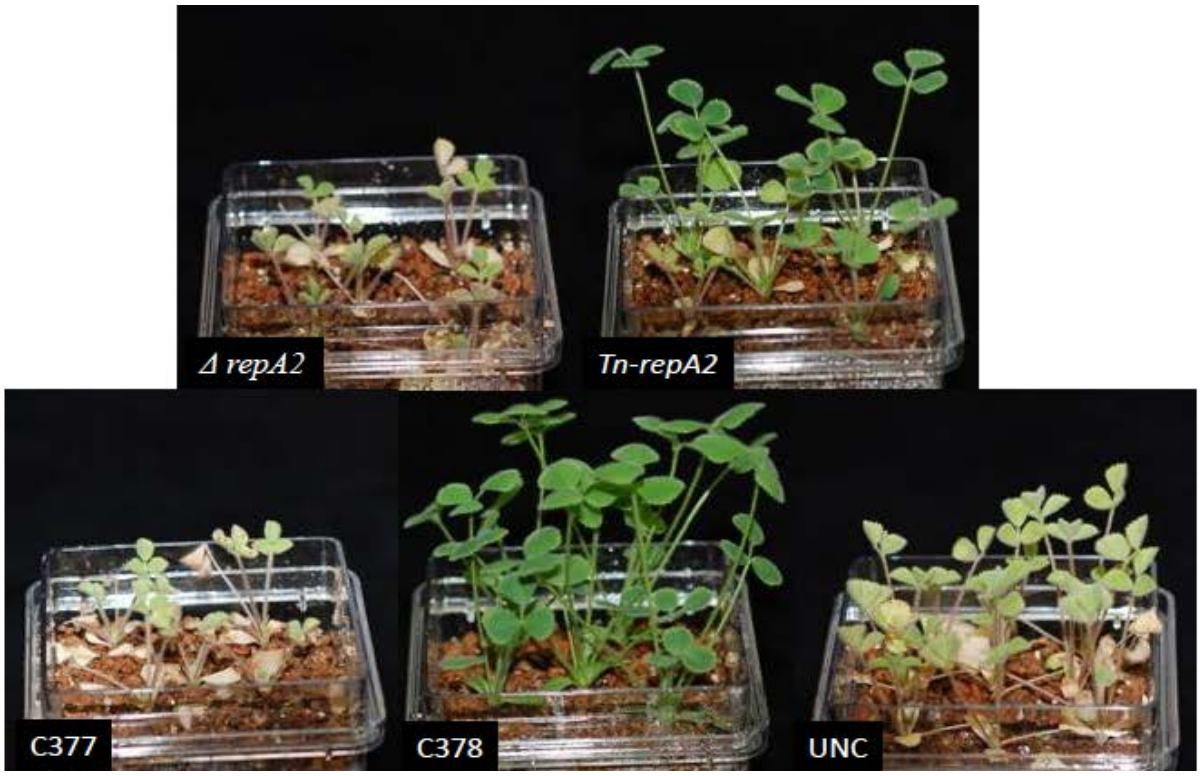

Figure 8. Clean in-frame deletion of repA2. Images are taken 28 days after inoculation of strains grown on A20 *M. truncatula.* A clean in-frame deletion of *repA2* does not alleviate the nitrogen fixation blocking phenotype resulting in unhealthy plants. A transposon insertion in *repA2* results in partially healthy plants while GOC strain C378 shows fully healthy plants resulting from complete symbiotic compatibility. *Δ repA2* = strain PP596 clean in-frame deletion of *repA2*. *Tn-repA2* = strain CP03 transposon insertion in *repA2*. C377 = wild type pHRC377. C378 = GOC of C377. UNC = uninoculated.

We next wanted to look at the expression of pHRC377 *in planta*. So we tagged pHRC377 with a constitutively active form of GUS (β-glucuronidase) that would either insert into *repC2* and disrupt it or insert nearby, not disrupting it. We inoculated plants and after 10 days of growth stained for glucuronidase activity. 10 days of growth would allow for symbiotic genes to be

expressed, but would not allow for the beginning of nodule senescence triggered by incompatible bacteria. What we found is that the disrupted version of *repC2* resulted in even staining throughout the nodule while the non-disruption version gave intense staining in the tip of nodule and little to no staining in the rest of the nodule (Figure 9). Our observations are consistent with the model that the disruption strain propagates throughout the nodule and therefore gives even staining throughout while in the non-disruption strain, plasmid levels spike where bacteroids would normally start to develop, and then in a zone of the nodule where differentiated bacteroids would normally reside, there is no staining. This is because bacteroids fail to differentiate in this zone because of the expression of the pHRC377 encoded nitrogen fixation blocking factor.

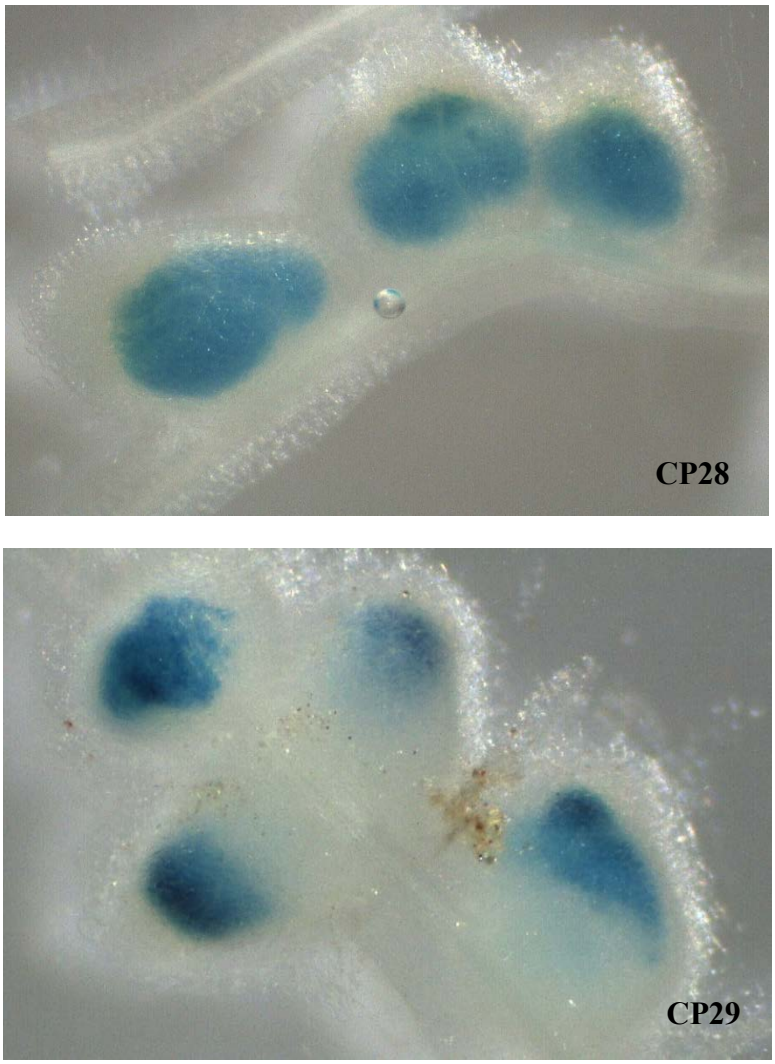

Figure 9. GUS staining for pHRC377 expression. 10 day old nodules harboring strains CP28 (*repC2* disruption) and CP29 (*repC2* non-disruption) were stained for glucuronidase activity. CP28 has even staining throughout the nodules indicating an even distribution of bacteria harboring pHRC377. CP29 has deep staining at the tips of the nodules with no stain below indicating expression of pHRC377 at the tip but nowhere else.

The *repA2C2* operon contains a full length *repA* gene, a ctRNA gene, and a full length *repC* gene (Figure 10). When a protein sequence alignment was done between *repA2C2* and *repABC* from pSymA and pSymB, the RepC proteins from *repA2C2* and *repABC* operon of pSymA were 98% identical, with 396 out of 402 amino acids being exactly the same. The intergenic region containing the ctRNA was also very similar. RepC2 from *repA2C2* and RepC

from pSymB were only 52% identical. RepA2 was 35% identical to RepA of pSymA and 37% identical to RepA of pSymB. This type of diversity reflects the natural diversity found among *repABC*-type operons in alphaproteobacteria. Alignments of other replication proteins on pHRC377 and in *S. meliloti* can be found in Table 7 (Supplementary).

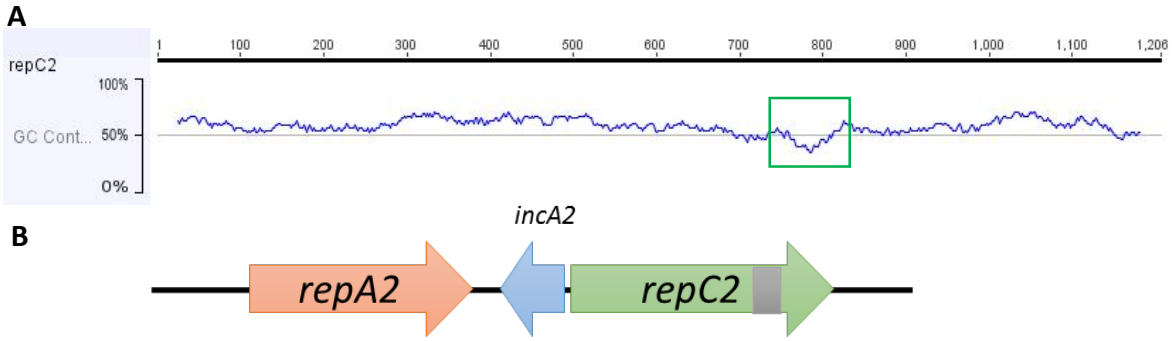

Figure 10. The repA2C2 operon. A, Graphic of GC content of *repC2*. The dip in the GC content near the end of *repC2* correlates with the predicted origin of replication. (Image created with Geneious version 7.1 created by Biomatters.) B, Graphic of the hypothesized *repA2C2* origin. The putative replication origin is the grey box

It was hypothesized that since a smaller version of pHRC377 containing undisrupted *repA2C2* allowed symbiotic compatibility, *repA2C2* could not be the ultimate determinant for the nitrogen fixation blocking phenotype. Instead, *repA2C2* might affect plasmid maintenance or affect the copy number of pHRC377 during endoreduplication in the symbiosome. This stimulated us to commence an investigation into the replication properties of *repA2C2* and of general pHRC377 maintenance. At the same time we decided to test the other possibly important replication operons of pHRC377 to get a more comprehensive picture of pHRC377 maintenance in *S. meliloti*. Other operons were considered important if they contained a complete *repC* gene. As previously explained in the introduction, *repC* is the initiator of replication, contains the origin of replication within its genetic space, and is the only gene of a *repABC*-type operon to be shown to be necessary for replication. Therefore both *repA1B1C1* and *repB4C4* were included in

our experiments, but were hypothesized to not affect the nitrogen fixation blocking phenotype (see Table 2).

To test the ability of *repA2C2* to be sufficient for plasmid replication, a minimal plasmid was created in *E. coli* that could then be conjugated into *S. meliloti* by tri-parental mating. We started by attempting to ligate the *repA2C2* region into a standard *E. coli* cloning vector that created a high copy number plasmid, but found it impossible to get these constructs to transform into *E. coli*. When the high copy number origin was switched for a low copy origin (pSC101), *E. coli* transformants were obtained. In the end the minimal plasmid contained an origin of transfer (*oriT*), an SC101 *E. coli* replication origin, a kanamycin/neomycin resistance gene (Km^R), and *repA2C2* with its promoter region (see Figure 11). We also created similar minimal plasmids for the other *repABC* operons containing a complete *repC* gene (*repA1B1C1* and *repB4C4*). Minimal plasmids containing *luxR/RR* and the *repABC* operon from pSymB were included as negative and positive controls, respectively, for this test. The minimal plasmids were created and transformed into *E. coli* and then mated into *S. meliloti* C378 (the GOC derivative of C377 containing no pHRC377). It was found that all *repABC* operons tested from pHRC377 were able to sustain replication of the minimal plasmids in C378. The minimal plasmid with *lux/RR*, as expected, was not able to sustain replication (See Table 3).

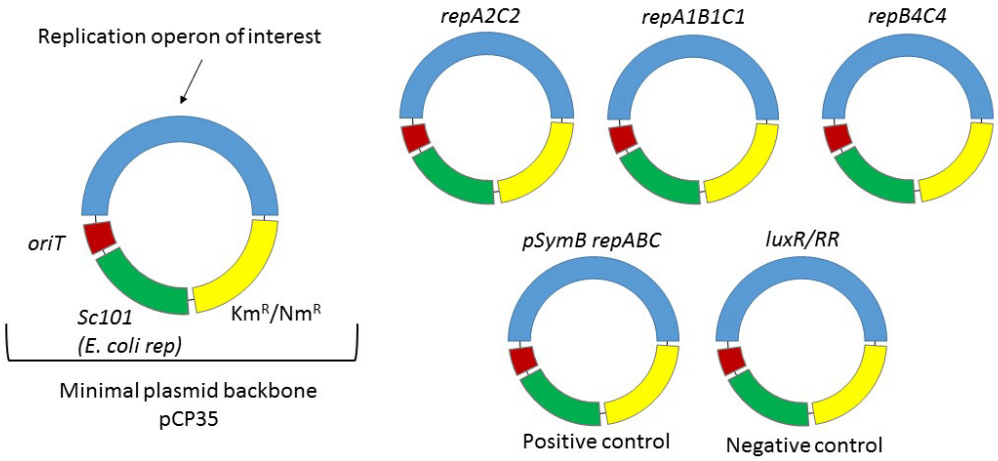

Figure 11. Outline of minimal plasmids. Each minimal plasmid backbone (pCP35) contained an *E. coli* replication origin *Sc101*, a mobilizable *oriT* gene, and a selectable marker (Km^R/Nm^R).

Table 3. Minimal plasmid replication.

Plasmid*	Able to maintain replication in *E. coli*?	Able to maintain replication in *S. meliloti*?
pCP35:*repA1B1C1*	Yes	Yes
pCP35:*repA2C2*	Yes	Yes
pCP35:*repB4C4*	Yes	Yes
pCP35:*luxR/RR*	Yes	No
pCP35:*pSymB repABC*	Yes	Yes

*The pCP35 plasmid backbone contains *Sc101:oriT*:Km. Each origin was inserted non-directionally with its promoter region.

The replication operons were further tested to see if they are sufficient for plasmid maintenance after a longer period growth. Each minimal plasmid strain was grown for five days with serial dilution in liquid culture with streptomycin (all strains are streptomycin resistant (Sm^R) regardless of plasmid presence). After five days each culture was diluted and plated onto LB with streptomycin selection. These colonies were then transferred to LB plates with additional neomycin selection. Any colonies that had lost the plasmid were no longer resistant to neomycin and were counted. Both complete operons *repA1B1C1* and pSymB *repABC*

experienced limited to no plasmid loss. The *repA2C2* minimal plasmid strain showed on average 22% of colonies maintained the plasmid while the *repB4C4* minimal plasmid strain had 1.33% of colonies maintain the plasmid (see Table 4). From these data it was hypothesized that *repA1B1C1* is the main replication operon of pHRC377 while *repA2C2* and *repB4C4* must play secondary or non-essential roles.

Table 4. Minimal plasmid maintenance in liquid culture.

Strain*	Percent of colonies that maintained the plasmid
PP539 (*repA1B1C1)*	92 %
PP538 (*repA2C2)*	22 %
PP540 (*repB4C4)*	1.33 %
PP541 (*pSymB repABC)*	100 %

*Each minimal plasmid strain was tested for ability to maintain plasmid presence (indicated by resistance to neomycin) in liquid culture after 5-day serial dilution growth. Numbers are averages of three replicates, n = 150.

Further evidence to test the hypothesis of *repA1B1C1* as the main replication origin of pHRC377 was needed. To do this we decided to test the growth abilities of strains with modified versions of pHRC377 in liquid culture. We created loop-in disruption and non-disruption strains of the 3 main *repC* genes on pHRC377. Each loop-in event marked pHRC377 with Km^R. Each of these strains was grown in liquid culture overnight to saturation with streptomycin/neomycin selection to ensure plasmid presence. They were then transferred to larger cultures with either streptomycin or streptomycin/neomycin selection and their growth was monitored over a 30 hour period (Figure 12). It was found that both *repC2* and *repC4* disruptions had no significant effect on growth. Conversely, the *repC1* disruption did have a significant effect. The strain containing pHRC377 with disrupted *repC1* (CP37) grew slower in both types of growth medium selection. Growth data therefore support the hypothesis that *repA1B1C1* is the main replication operon for pHRC377 and that *repA2C2* and *repB4C4* are non-essential for growth in liquid culture.

Interestingly growth was not completely stopped through the *repC1* disruption indicating the other replication genes and operons on pHRC377 still function, but less optimally.

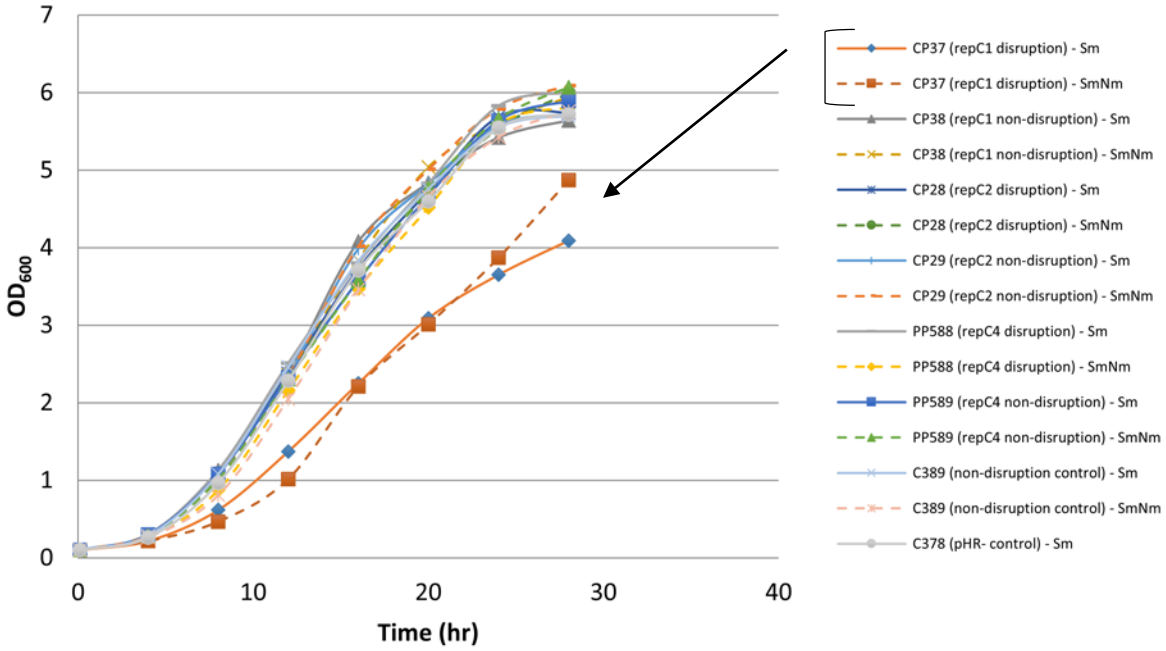

Figure 12. The importance of repC1 for pHRC377 maintenance in liquid culture. Growth curve data over a 30 hr period. Out of all the strains tested only CP37 (*repC1* disruption) showed a significant difference in growth pattern from the controls (CP389, C378).

If *repA2C2* is not essential for pHRC377 maintenance in liquid culture then it was hypothesized that it might be important for maintenance during bacterial division in the nodule or during endoreduplication in the symbiosome. Testing pHRC377 maintenance during bacterial division in nodules was the most approachable problem because *S. meliloti* cells are easily recoverable from crushed nodules. These recoverable bacteria are found in infection threads and have not undergone the developmental changes (like endoreduplication) that allow them to fix nitrogen and become terminally differentiated. To test plasmid maintenance, transposon mutants CP05 (*Tn-repC2*) and CP07 (*Tn-luxR*) were used to inoculate A20 *M. truncatula*. 28 days after

inoculation, recoverable bacteria were obtained from crushed nodules and plated onto LB with streptomycin. After three days colonies were then transferred to neomycin selection. Any colonies that had lost neomycin resistance and therefore lost pHRC377 were recorded (see Table 5). As a control, two-day saturated cultures of these strains were also tested in the same way. It was found that less than 1% of colonies had lost neomycin resistance regardless of the strain or growth condition used. It was concluded that disruption of *repC2* does not affect pHRC377 maintenance in recoverable bacteria from nodules. It follows that *repA2C2* function must be important for pHRC377 replication during endoreduplication in the symbiosome, beyond the point where colonies can be recovered from the cells.

Table 5. pHRC377 maintenance in recoverable bacteria from nodules.

Growth condition/Strain*	Percent of Colonies that Maintained Resistance
Liquid CP05 (*Tn-repC2)*	99.5 %
Liquid CP07 (*Tn-luxR)*	99.8 %
Nodule CP05 (*Tn-repC2)*	99.7 %
Nodule CP07 (*Tn-luxR)*	99.8 %

* Strains were grown for 2 days in liquid culture or 28 days on plants. Bacteria were diluted and plated from culture or recovered from nodules. Colonies were then transferred to neomycin selection and resistant colonies were counted. Percentages are averages of three replicates, n = ~600. Differences between strains and growth methods are not statistically significantly ($p < 0.001$).

A final assay was needed to test if the copy number of pHRC377 changed during endoreduplication in the symbiosome when *repA2C2* was disrupted. As was explained in the introduction most bacteria contained in nodules have been endocytosed into plant cells and undergone many physiological changes to be able to fix nitrogen. These changes include many rounds of endoreduplication which can result in 64 or 128 copies of the genome (Mergaert, Uchiumi et al. 2006). Bacterial cells are called bacteroids at this point. The *repA2C2* operon could be doing one of three things during endoreduplication. It could facilitate basal replication of pHRC377 inside the symbiosome, it could facilitate endoreduplication of pHRC377 along

with the rest of the genome, or it could cause hyper-endoreduplication. Since bacteroids are no longer recoverable we decided to extract bacterial DNA from nodules. Then the amount of DNA present for different targets could be tested using qPCR. With qPCR we would have the ability to compare the relative amount of DNA for multiple targets in a sample and the ease of comparing different samples to each other. Genomic DNA preparations were created from 2-day saturated liquid cultures and nodules 28 days after inoculation. The strains used were the same used for testing plasmid maintenance in recoverable bacteria from nodules, CP05 and CP07. Primers selecting for targets on the chromosome and on pHRC377 were used to assess relative copy number. The chromosomal target was used as an internal copy number reference. In liquid culture conditions, both plasmid derivatives exhibited copy number that was similar to each other and to the chromosome. However, DNA samples from nodules show that pHRC377 (*Tn-luxR*) increases slightly relative to the chromosome, while pHRC377 (*Tn-repC2*) copy number decreases sharply relative to the chromosome (Figure 13). This observation supports the hypothesis that *repA2C2* is important for copy number of pHRC377 during endoreduplication in the symbiosome but not important for plasmid maintenance in liquid medium.

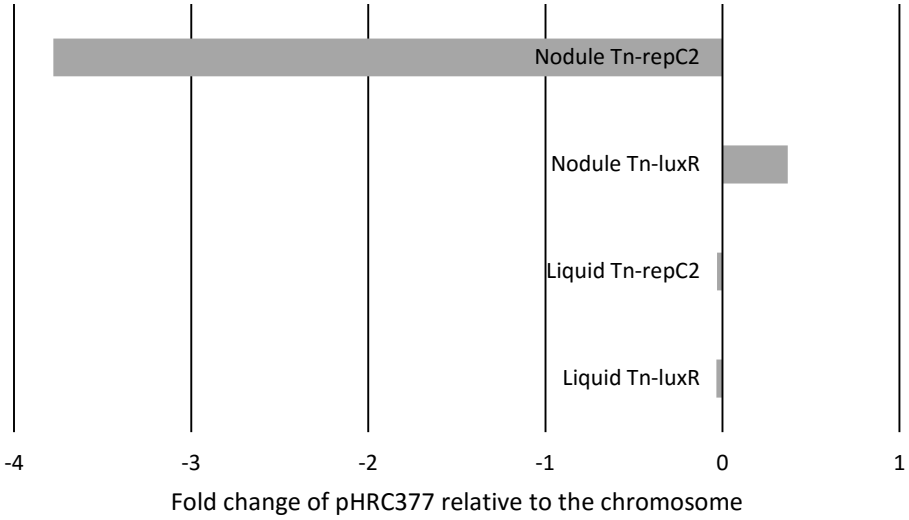

Figure 13. Fold change of pHRC377 in liquid culture and in nodules. The fold change of pHRC377 was determined using delta delta Ct values generated from Ct values from qPCR. Each sample represents the average of 3 biological replicates. *Tn-repC2* = CP05. *Tn-luxR* = CP07**.**

DISCUSSION

S. meliloti has been shown to carry a great diversity of accessory plasmids including two mega-plasmids that are essential for its symbiotic relationship with *Medicago* species. These types of plasmids typically have a *repABC*-type operon controlling their replication. Multiple *repABC* operons in a single organism is fairly common due to their natural diversity and ability to have many incompatibility groups, but descriptions of multiple *repABC* operons on the same plasmid is less common. To our knowledge an in-depth description of multiple *repABC* operons on a single plasmid one of which controls endoreduplication in late stage symbiotic development has never been shown. The data presented here describes pHRC377, an accessory plasmid found in *S. meliloti*. This plasmid contains one complete replication operon (*repA1B1C1*) that has properties that could classify it as the main replication operon of pHRC377, a secondary replication operon (*repA2C2*) shown to effect the ability of pHRC377 to replicate during endoreduplication in the symbiosome, and one other full length *repC* gene (*repC4*) which has weak replication properties. It is important to note that *repA2C2* was also found to control the nitrogen fixation blocking phenotype for two other accessory plasmids pHRC017 and pHRB469. The pHRC377 plasmid was chosen for this study because it was fully sequenced and was found to be the easiest to genetically manipulate.

Plasmid maintenance was the chosen focus of this study because it was known through a spontaneous large deletion mutant of pHRC377 that the ability of pHRC377 to block nitrogen fixation was lost when half of the plasmid not containing *repA2C2* was also lost. So *repA2C2* was not the sole cause of the nitrogen fixation blocking phenotype, but allowing it to happen. Another interesting property of *repA2C2* is the fact that its RepC2 protein is almost identical to

the RepC protein encoded by pSymA. This could indicate as other researchers have shown, a recent recombination event (Castillo-Ramirez, Vazquez-Castellanos et al. 2009).

Endoreduplication in *S. meliloti* is still under investigation. Most recently it has been shown that low concentrations of NCR peptides can cause endoreduplication in *S. meliloti* (Penterman, Abo et al. 2014). NCR peptides block cell division and disrupt Z-ring function. They cause a transcriptional response in the bacterial cell which changes the expression of cell-cycle regulators and cell division genes. Further research has shown that the cell cycle master regulator CtrA may have an important role in symbiotic cell development. Its depletion causes cells to elongate and their genomes to amplify (Pini, De Nisco et al. 2015). Perhaps with pHRC377 only *repA2C2* can be regulated by changes in cell cycle regulators like CtrA. Without the ability to respond to cell cycle regulators pHRC377 will not endoreduplicate during symbiotic development and so its nitrogen fixation blocking property will be lost.

Maybe the diverse collection of different replication operons on pHRC377 hints at an unknown deficiency in the *repA1B1C1* operon as does the importance of *repA2C2* in the symbiosome. Perhaps further genetic studies and comparisons of the other accessory plasmids with *repA2C2* would further elucidate this point. They also might implicate the reason why accessory plasmids would need to harbor so many replication genes. Whether it has become their job to spread genetic diversity of replication machinery or as scavengers collecting spare parts to keep themselves functioning.

MATERIALS AND METHODS

Bacterial Growth Conditions and Media

Specific strains used in this study are listed in Table 8 (Supplementary), *Escherichia coli, Sinorhizobium meliloti,* and *Agrobacterium tumefaciens* cultures were grown at 37ºC, 30ºC, and 30ºC, respectively, in Luria-Bertani (LB) medium or on LB agar plates. Media were supplemented with following antibiotics as needed: 2.5 μg/mL tetracycline, 3 μg/mL gentamicin for *E. coli*, 15 μg/mL gentamicin for *S. meliloti/A. tumefaciens*, 30 μg/mL kanamycin, 30 μg/mL chloramphenicol, 50 μg/mL rifampicin, 100 μg/mL neomycin, 100 μg/mL ampicillin, and 200 μg/mL streptomycin.

Growth Curve Analyses

Growth curve analyses of *S. meliloti* derivatives were initiated using strains that had grown to saturation overnight at 30ºC in LB medium containing streptomycin and neomycin to ensure plasmid maintenance. At time zero, the overnight cultures were used to inoculate 20 ml of fresh LB media to achieve an OD_{600} equivalent of 0.1 in 125-ml Erlenmeyer flasks. Culture media were again supplemented with streptomycin/neomycin to ensure plasmid maintenance during the course of the growth curve. Identical control cultures supplemented with only streptomycin were used to monitor growth under non-selective conditions. The cultures were incubated at 30ºC at 225 rpm and OD_{600} measurements were recorded every 4 hours for a total of 36 hours. To account for the degree to which plasmid loss contributed to the growth of the control cultures (supplemented only with streptomycin), the 36-hour cultures were plated on LB agar plates containing streptomycin and patched onto LB agar plates containing streptomycin/neomycin to determine the percent plasmid loss of the culture.

Plasmid Maintenance Assays

Strains were serially passaged in liquid LB-streptomycin for approximately 100 generation (5 days of serial passage every 24 hr) and plated at different time points on LB-streptomycin plates. After 3-4 days of growth at 30ºC, individual colonies were patched onto both LB-streptomycin/neomycin and LB-streptomycin plates and grown for 3-4 days at 30ºC. Percent plasmid loss was calculated based on the total number of colonies that lost neomycin resistance.

Transposon Mutagenesis

Plasmid-specific transposon mutagenesis of pHRB469, pHRC017, and pHRC377 was achieved using a mating-out procedure as described by Price, Tanner et. al. 2015. Transposon insertion sites for mutant strains that yielded a Fix^+ phenotype and maintained the respective pHR were mapped onto the pHRs using arbitrary-PCR. Table 6 (Supplementary) lists the specific transposon insertion sites that were mapped for all three pHRs. PCR analysis of *repA2C2* was used to determine the number/percent of insertions into this locus in pHRC377.

Plasmid and Strain Construction

Plasmids, strains, and oligonucleotides used in this study are listed in Tables 8, 9, and 10 (Supplementary). Plasmids were constructed using standard molecular techniques with enzymes purchased from New England Biolabs (Boston, MA). All custom oligonucleotides were purchased from Invitrogen.

Mobilization of plasmids between strains was mediated via tri-parental matings with helper *E. coli* B001 (DH5α harboring pRK600) followed by selection on the appropriate antibiotics. Clean deletion strains were created using the suicide vector pJQ200sk followed by selection on gentamicin and counterselection on sucrose.

Annotation and Sequence Analysis

Sequencing and Annotation of pHRC377 was done as previously shown (Price, Tanner et al. 2015). For pyrosequencing of pHRC377, the plasmid was marked with an *oriT/neo* cassette (pJG498) and conjugated into plasmid-free *A. tumefaciens* UBAPF2 to yield strain C396. C396 was grown overnight in 50 mL of LB, pelleted, and resuspended in 5 mL of Qiagen P1 buffer. Five milliliters of Qiagen P2 buffer was added to the cells and incubated at room temperature for 10 min before the addition of 7.5 mL of ice- cold Qiagen N3 buffer. Lysates were incubated for 30 min on ice and centrifuged twice at 10,000 × g for 30 min to remove cellular debris. DNA was precipitated with isopropanol, washed with 70% ethanol, dried, and resuspended in Tris-EDTA (TE). Samples were treated with 200 μg of proteinase K and incubated at 42°C for 2 h, followed by chloroform extraction, isopropanol/sodium acetate precipitation, and resuspension in TE. The 454 library preparation was performed according to the rapid library preparation protocol, followed by sequencing on the 454 Genome Analyzer FLX (Roche). Assembly into contigs was performed using Newbler (version 2.5.3), and contigs corresponding to *A. tumefaciens* were removed. Contig edge ambiguities were resolved by PCR using oPP166–oPP171. The sequence corresponding to pJG498 then was removed from the assembly to reconstitute the native pHRC377 sequence. Reads were remapped to this assembly using Geneious (version 5.3.4) to confirm the sequence. The final assembly was preliminarily annotated using DNA Master, using Glimmer (version 3.02) to predict ORFs and BLASTx to assign putative functions. This sequence will soon be accessible in GenBank. The annotated sequence was used to identify replication origins or partial replication origins.

Protein sequence alignments were done using Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/).

Plant Growth and Nodulation

M. truncatula A20 (Jemalong A20, Sharon Long Laboratory, Department of Biology, Stanford University, Stanford, CA) plants were grown in a 4:1 Turface:Vermiculite mixture (Turface Athletics, Buffalo Grove, IL, U.S.A.; Thermo-O-Rock West Inc., Chandler, AZ, U.S.A.). Seeds were scarified in 100% sulfuric acid, surface-sterilized in 6.5% bleach, vernalized at 4°C for 2 days, and germinated in petri plates for 2 days. Seedlings were then planted in sterile the Turface:Vermiculite mixture and allowed to grow for 4 days prior to inoculation. Plants were maintained under nitrogen-limiting conditions as indicated for individual experiments. Plants were watered with standard nodulation medium (SNM) as described in Crook, Lindsey et al. 2012 [1 mM KH2PO4, 0.5 mM MgSO4, 0.5 mM CaCl2·2H2O, 50 μM Na2-EDTA·2H2O, 50 μM FeSO4·7H2O, 30 μM H3BO3, 2.5 μM MnSO4·H2O, 0.35 μM ZnSO4·7H2O, 0.4 μM Na2MoO4·2H2O, 0.6 μM CuSO4, and 0.1 μM CoCl2)]

Genomic DNA preparations

For *in vitro* grown bacteria, saturated overnight cultures were pelleted and resuspended in TE buffer. For *in planta* experiments, 28-day-old whole nodules were harvested and crushed in TE buffer. The same standard genomic DNA preparation protocol was followed for both samples. Briefly, 10% SDS and proteinase K (20 mg/ml) were added to the resuspended cultures/nodules. The tubes were incubated at 50ºC for 20 min, and 100 μl 5M NaCl was added to the samples. The samples were centrifuged, and the supernatant was moved to a new tube. A chloroform extraction was performed on the samples followed by RNase A (final concentration 0.2 μg/μl) treatment for 20 minutes at 37ºC. A second chloroform extraction was performed, and the DNA was precipitated using isopropanol. The DNA was pelleted and resuspened in TE buffer. DNA

concentration was determined using a Nanodrop® spectrophotometer (ThermoScientific, Waltham, MA).

qPCR

IDT PrimerQuest Tool (https://www.idtdna.com/Primerquest/Home/Index) was used to design primers for targets on the chromosome, pSymA, and pHRC377 (see Table 11 (Supplementary)). qPCR analyses were performed in an Applied Biosystems StepOneTM Real-time PCR System (Foster City, CA) using the following program: 5 min at 94ºC; and 40 cycles of 15 sec at 95ºC and 60 sec at 60ºC. Fluorescence data was acquired during the extension step at 60ºC. A melt curve was also performed to check for product specificity. The master mix for the reactions contained: 0.5 µl forward primer (12.5 µM), 0.5 µl reverse primer (12.5 µM), 1 µl DNA from genomic prep (1 or 10 µg/ml), 3 µl H_2O, and 5 µl iTaq™ Universal SYBR® Green Supermix (Bio-Rad, Hercules, California). All reactions were performed in triplicate and included a negative TE-only control. C_t values for each triplicate were averaged, and differences in genome copy number between samples was calculated using standard methods ($2^{-\Delta\Delta Ct}$).

GUS staining

GUS reporter strains were generated by integrating reporter plasmids into the various replication origins such that *repC1*or *repC2*, remained intact or was disrupted by the integration. X-GLUC was used to measure GUS expression in 14-d-old *M. truncatula* accession A20 nodules (Price, Tanner et al. 2015). Briefly, nodules were fixed (2% paraformaldehyde, 0.1 M sodium citrate) on ice for 30 min, washed three times with 50 mM sodium citrate, stained (0.5 mg/mL X-GLUC, 50 mM sodium citrate, 1 mM potassium ferrocyanide, 1 mM potassium ferricyanide, 0.1% Triton- X, 10% methanol, vacuum infiltration) for 3-5 h at 37°C in the dark, washed with

ddH2O, bleached for 3 min, washed with ddH2O, and imaged under a Leica EZ4D dissecting microscope (Leica Micro- systems, Inc.).

REFERENCES

Barnett, M. J., R. F. Fisher, T. Jones, C. Komp, A. P. Abola, F. Barloy-Hubler, L. Bowser, D. Capela, F. Galibert, J. Gouzy, M. Gurjal, A. Hong, L. Huizar, R. W. Hyman, D. Kahn, M. L. Kahn, S. Kalman, D. H. Keating, C. Palm, M. C. Peck, R. Surzycki, D. H. Wells, K. C. Yeh, R. W. Davis, N. A. Federspiel and S. R. Long (2001). "Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti pSymA megaplasmid." Proceedings of the National Academy of Sciences of the United States of America **98**(17): 9883-9888.

Barran, L. R., E. S. P. Bromfield and D. C. W. Brown (2002). "Identification and cloning of the bacterial nodulation specificity gene in the Sinorhizobium meliloti - Medicago laciniata symbiosis." Canadian Journal of Microbiology **48**(9): 765-771.

Bartosik, D., A. Bialkowska, J. Baj and M. Wlodarczyk (1997). "Construction of mobilizable cloning vectors derived from pBGS18 and their application for analysis of replicator region of a pTAV202 mini-derivative of Paracoccus versutus pTAV1 plasmid." Acta Microbiol Pol **46**(4): 387-392.

Bartosik, D., M. Wlodarczyk and C. M. Thomas (1997). "Complete nucleotide sequence of the replicator region of Paracoccus (Thiobacillus) versutus pTAV1 plasmid and its correlation to several plasmids of Agrobacterium and Rhizobium species." Plasmid **38**(1): 53-59.

Cangelosi, G. A., L. Hung, V. Puvanesarajah, G. Stacey, D. A. Ozga, J. A. Leigh and E. W. Nester (1987). "Common loci for Agrobacterium tumefaciens and Rhizobium meliloti exopolysaccharide synthesis and their roles in plant interactions." J Bacteriol **169**(5): 2086-2091.

Castillo-Ramirez, S., J. F. Vazquez-Castellanos, V. Gonzalez and M. A. Cevallos (2009). "Horizontal gene transfer and diverse functional constrains within a common replication-partitioning system in Alphaproteobacteria: the repABC operon." BMC Genomics **10**: 536.

Cervantes-Rivera, R., F. Pedraza-Lopez, G. Perez-Segura and M. A. Cevallos (2011). "The replication origin of a repABC plasmid." Bmc Microbiology **11**.

Cervantes-Rivera, R., C. Romero-Lopez, A. Berzal-Herranz and M. A. Cevallos (2010). "Analysis of the Mechanism of Action of the Antisense RNA That Controls the Replication of the repABC Plasmid p42d." Journal of Bacteriology **192**(13): 3268-3278.

Cevallos, M. A., R. Cervantes-Rivera and R. M. Gutierrez-Rios (2008). "The repABC plasmid family." Plasmid **60**(1): 19-37.

Chai, Y. and S. C. Winans (2005). "A small antisense RNA downregulates expression of an essential replicase protein of an Agrobacterium tumefaciens Ti plasmid." Mol Microbiol **56**(6): 1574-1585.

Crook, M. B., D. P. Lindsay, M. B. Biggs, J. S. Bentley, J. C. Price, S. C. Clement, M. J. Clement, S. R. Long and J. S. Griffitts (2012). "Rhizobial Plasmids That Cause Impaired Symbiotic Nitrogen Fixation and Enhanced Host Invasion." Molecular Plant-Microbe Interactions **25**(8): 1026-1033.

Dunham, T. D., W. Xu, B. E. Funnell and M. A. Schumacher (2009). "Structural basis for ADP-mediated transcriptional regulation by P1 and P7 ParA." EMBO J **28**(12): 1792-1802.

Farkas, A., G. Maroti, H. Durgo, Z. Gyorgypal, R. M. Lima, K. F. Medzihradszky, A. Kereszt, P. Mergaert and E. Kondorosi (2014). "Medicago truncatula symbiotic peptide NCR247 contributes to bacteroid differentiation through multiple mechanisms." Proc Natl Acad Sci U S A **111**(14): 5183-5188.

Faucher, C., F. Maillet, J. Vasse, C. Rosenberg, A. A. van Brussel, G. Truchet and J. Denarie (1988). "Rhizobium meliloti host range nodH gene determines production of an alfalfa-specific extracellular signal." J Bacteriol **170**(12): 5489-5499.

Finan, T. M., S. Weidner, K. Wong, J. Buhrmester, P. Chain, F. J. Vorholter, I. Hernandez-Lucas, A. Becker, A. Cowie, J. Gouzy, B. Golding and A. Puhler (2001). "The complete sequence of the 1,683-kb pSymB megaplasmid from the N-2-fixing endosymbiont Sinorhizobium meliloti." Proceedings of the National Academy of Sciences of the United States of America **98**(17): 9889-9894.

Gage, D. J. (2004). "Infection and invasion of roots by symbiotic, nitrogen-fixing rhizobia during nodulation of temperate legumes." Microbiology and Molecular Biology Reviews **68**(2): 280-+.

Galibert, F., T. M. Finan, S. R. Long, A. Puhler, P. Abola, F. Ampe, F. Barloy-Hubler, M. J. Barnett, A. Becker, P. Boistard, G. Bothe, M. Boutry, L. Bowser, J. Buhrmester, E. Cadieu, D. Capela, P. Chain, A. Cowie, R. W. Davis, S. Dreano, N. A. Federspiel, R. F. Fisher, S. Gloux, T. Godrie, A. Goffeau, B. Golding, J. Gouzy, M. Gurjal, I. Hernandez-Lucas, A. Hong, L. Huizar, R. W. Hyman, T. Jones, D. Kahn, M. L. Kahn, S. Kalman, D. H. Keating, E. Kiss, C. Komp, V. Lalaure, D. Masuy, C. Palm, M. C. Peck, T. M. Pohl, D. Portetelle, B. Purnelle, U. Ramsperger, R. Surzycki, P. Thebault, M. Vandenbol, F. J. Vorholter, S. Weidner, D. H. Wells, K. Wong, K. C. Yeh and J. Batut (2001). "The composite genome of the legume symbiont Sinorhizobium meliloti." Science **293**(5530): 668-672.

Griffitts, J. S. and S. R. Long (2008). "A symbiotic mutant of Sinorhizobium meliloti reveals a novel genetic pathway involving succinoglycan biosynthetic functions." Mol Microbiol **67**(6): 1292-1306.

Haag, A. F., M. F. F. Arnold, K. K. Myka, B. Kerscher, S. Dall'Angelo, M. Zanda, P. Mergaert and G. P. Ferguson (2013). "Molecular insights into bacteroid development during Rhizobium-legume symbiosis." Fems Microbiology Reviews **37**(3): 364-383.

Haag, A. F., M. Baloban, M. Sani, B. Kerscher, O. Pierre, A. Farkas, R. Longhi, E. Boncompagni, D. Herouart, S. Dall'angelo, E. Kondorosi, M. Zanda, P. Mergaert and G. P. Ferguson (2011). "Protection of Sinorhizobium against host cysteine-rich antimicrobial peptides is critical for symbiosis." PLoS Biol **9**(10): e1001169.

Horvath, B., A. Domonkos, A. Kereszt, A. Szucs, E. Abraham, F. Ayaydin, K. Boka, Y. Chen, R. Chen, J. D. Murray, M. K. Udvardi, E. Kondorosi and P. Kalo (2015). "Loss of the nodule-specific cysteine rich peptide, NCR169, abolishes symbiotic nitrogen fixation in the Medicago truncatula dnf7 mutant." Proc Natl Acad Sci U S A.

Izquierdo, J., T. Venkova-Canova, M. A. Ramirez-Romero, J. Tellez-Sosa, I. Hernandez-Lucas, J. Sanjuan and M. A. Cevallos (2005). "An antisense RNA plays a central role in the replication control of a repC plasmid." Plasmid **54**(3): 259-277.

Jones, K. M., H. Kobayashi, B. W. Davies, M. E. Taga and G. C. Walker (2007). "How rhizobial symbionts invade plants: the Sinorhizobium-Medicago model." Nature Reviews Microbiology **5**(8): 619-633.

Kereszt, A., P. Mergaert and E. Kondorosi (2011). "Bacteroid Development in Legume Nodules: Evolution of Mutual Benefit or of Sacrificial Victims?" Molecular Plant-Microbe Interactions **24**(11): 1300-1309.

Kiekebusch, D. and M. Thanbichler (2014). "Plasmid segregation by a moving ATPase gradient." Proc Natl Acad Sci U S A **111**(13): 4741-4742.

MacLellan, S. R., L. A. Smallbone, C. D. Sibley and T. M. Finan (2005). "The expression of a novel antisense gene mediates incompatibility within the large repABC family of alpha-proteobacterial plasmids." Mol Microbiol **55**(2): 611-623.

Masson-Boivin, C., E. Giraud, X. Perret and J. Batut (2009). "Establishing nitrogen-fixing symbiosis with legumes: how many rhizobium recipes?" Trends Microbiol **17**(10): 458-466.

Maunoury, N., M. Redondo-Nieto, M. Bourcy, W. Van de Velde, B. Alunni, P. Laporte, P. Durand, N. Agier, L. Marisa, D. Vaubert, H. Delacroix, G. Duc, P. Ratet, L. Aggerbeck, E. Kondorosi and P. Mergaert (2010). "Differentiation of symbiotic cells and endosymbionts in Medicago truncatula nodulation are coupled to two transcriptome-switches." PLoS One **5**(3): e9519.

Mazur, A. and P. Koper (2012). "Rhizobial plasmids - replication, structure and biological role." Central European Journal of Biology **7**(4): 571-586.

Meade, H. M., S. R. Long, G. B. Ruvkun, S. E. Brown and F. M. Ausubel (1982). "Physical and genetic characterization of symbiotic and auxotrophic mutants of Rhizobium meliloti induced by transposon Tn5 mutagenesis." J Bacteriol **149**(1): 114-122.

Mergaert, P., K. Nikovics, Z. Kelemen, N. Maunoury, D. Vaubert, A. Kondorosi and E. Kondorosi (2003). "A novel family in Medicago truncatula consisting of more than 300 nodule-specific genes coding for small, secreted polypeptides with conserved cysteine motifs." Plant Physiol **132**(1): 161-173.

Mergaert, P., T. Uchiumi, B. Alunni, G. Evanno, A. Cheron, O. Catrice, A. E. Mausset, F. Barloy-Hubler, F. Galibert, A. Kondorosi and E. Kondorosi (2006). "Eukaryotic control on bacterial cell cycle and differentiation in the Rhizobium-legume symbiosis." Proceedings of the National Academy of Sciences of the United States of America **103**(13): 5230-5235.

Nallu, S., K. A. Silverstein, P. Zhou, N. D. Young and K. A. Vandenbosch (2014). "Patterns of divergence of a large family of nodule cysteine-rich peptides in accessions of Medicago truncatula." Plant J **78**(4): 697-705.

Newton, W. E. (2000). Nitrogen fixation in perspective. Nitrogen fixation: from molecules to crop productivity. Netherlands, Springer**:** 3-8.

Oldroyd, G. E. D., J. D. Murray, P. S. Poole and J. A. Downie (2011). "The Rules of Engagement in the Legume-Rhizobial Symbiosis." Annual Review Genetics, Vol 45 **45**: 119-144.

Pappas, K. M. and S. C. Winans (2003). "The RepA and RepB autorepressors and TraR play opposing roles in the regulation of a Ti plasmid repABC operon." Mol Microbiol **49**(2): 441-455.

Penmetsa, R. V. and D. R. Cook (2000). "Production and characterization of diverse developmental mutants of Medicago truncatula." Plant Physiol **123**(4): 1387-1398.

Penterman, J., R. P. Abo, N. J. De Nisco, M. F. Arnold, R. Longhi, M. Zanda and G. C. Walker (2014). "Host plant peptides elicit a transcriptional response to control the Sinorhizobium meliloti cell cycle during symbiosis." Proc Natl Acad Sci U S A **111**(9): 3561-3566.

Penterman, J., R. P. Abo, N. J. De Nisco, M. F. F. Arnold, R. Longhi, M. Zanda and G. C. Walker (2014). "Host plant peptides elicit a transcriptional response to control the Sinorhizobium meliloti cell cycle during symbiosis." Proceedings of the National Academy of Sciences of the United States of America **111**(9): 3561-3566.

Perez-Oseguera, A. and M. A. Cevallos (2013). "RepA and RepB exert plasmid incompatibility repressing the transcription of the repABC operon." Plasmid **70**(3): 362-376.

Perez-Segura, G., A. Perez-Oseguera and M. A. Cevallos (2013). "The repAC replication system of the Rhizobium leguminosarum pRL7 plasmid is functional: Implications regarding the origin and evolution of repABC plasmids." Plasmid **69**(1): 49-57.

Petersen, J., H. Brinkmann and S. Pradella (2009). "Diversity and evolution of repABC type plasmids in Rhodobacterales." Environ Microbiol **11**(10): 2627-2638.

Pini, F., N. J. De Nisco, L. Ferri, J. Penterman, A. Fioravanti, M. Brilli, A. Mengoni, M. Bazzicalupo, P. H. Viollier, G. C. Walker and E. G. Biondi (2015). "Cell Cycle Control by the Master Regulator CtrA in Sinorhizobium meliloti." PLoS Genet **11**(5): e1005232.

Pinto, U. M., A. L. Flores-Mireles, E. D. Costa and S. C. Winans (2011). "RepC protein of the octopine-type Ti plasmid binds to the probable origin of replication within repC and functions only in cis." Mol Microbiol **81**(6): 1593-1606.

Pinto, U. M., K. M. Pappas and S. C. Winans (2012). "The ABCs of plasmid replication and segregation." Nature Reviews Microbiology **10**(11): 755-765.

Price, P. A., J. Jin and W. E. Goldman (2012). "Pulmonary infection by Yersinia pestis rapidly establishes a permissive environment for microbial proliferation." Proc Natl Acad Sci U S A **109**(8): 3083-3088.

Price, P. A., H. R. Tanner, B. A. Dillon, M. Shabab, G. C. Walker and J. S. Griffitts (2015). "Rhizobial peptidase HrrP cleaves host-encoded signaling peptides and mediates symbiotic compatibility." Proc Natl Acad Sci U S A.

Quandt, J. and M. F. Hynes (1993). "Versatile suicide vectors which allow direct selection for gene replacement in gram-negative bacteria." Gene **127**(1): 15-21.

Ramirez-Romero, M. A., J. Tellez-Sosa, H. Barrios, A. Perez-Oseguera, V. Rosas and M. A. Cevallos (2001). "RepA negatively autoregulates the transcription of the repABC operon of the Rhizobium etli symbiotic plasmid basic replicon." Molecular Microbiology **42**(1): 195-204.

Roche, P., F. Maillet, C. Plazanet, F. Debelle, M. Ferro, G. Truchet, J. C. Prome and J. Denarie (1996). "The common nodABC genes of Rhizobium meliloti are host-range determinants." Proceedings of the National Academy of Sciences of the United States of America **93**(26): 15305-15310.

Tabata, S., P. J. Hooykaas and A. Oka (1989). "Sequence determination and characterization of the replicator region in the tumor-inducing plasmid pTiB6S3." J Bacteriol **171**(3): 1665-1672.

Thoquet, P., M. Gherardi, E. P. Journet, A. Kereszt, J. M. Ane, J. M. Prosperi and T. Huguet (2002). "The molecular genetic linkage map of the model legume Medicago truncatula: an essential tool for comparative legume genomics and the isolation of agronomically important genes." BMC Plant Biol **2**: 1.

Van Berkum, P., P. Elia and B. D. Eardy (2006). "Multilocus sequence typing as an approach for population analysis of Medicago-nodulating rhizobia." Journal of Bacteriology **188**(15): 5570-5577.

Van de Velde, W., G. Zehirov, A. Szatmari, M. Debreczeny, H. Ishihara, Z. Kevei, A. Farkas, K. Mikulass, A. Nagy, H. Tiricz, B. Satiat-Jeunemaitre, B. Alunni, M. Bourge, K. Kucho, M. Abe, A. Kereszt, G. Maroti, T. Uchiumi, E. Kondorosi and P. Mergaert (2010). "Plant Peptides Govern Terminal Differentiation of Bacteria in Symbiosis." Science **327**(5969): 1122-1126.

Venkova-Canova, T., N. E. Soberon, M. A. Ramirez-Romero and M. A. Cevallos (2004). "Two discrete elements are required for the replication of a repABC plasmid: an antisense RNA and a stem-loop structure." Molecular Microbiology **54**(5): 1431-1444.

Wais, R. J., D. H. Keating and S. R. Long (2002). "Structure-function analysis of nod factor-induced root hair calcium spiking in Rhizobium-legume symbiosis." Plant Physiol **129**(1): 211-224.

Wang, D., J. Griffitts, C. Starker, E. Fedorova, E. Limpens, S. Ivanov, T. Bisseling and S. Long (2010). "A nodule-specific protein secretory pathway required for nitrogen-fixing symbiosis." Science **327**(5969): 1126-1129.

Yip, C. B., H. Ding and M. F. Hynes (2015). "Counter-transcribed RNAs of Rhizobium leguminosarum repABC plasmids exert incompatibility effects only when highly expressed." Plasmid **78**: 37-47.

Young, J. P., L. C. Crossman, A. W. Johnston, N. R. Thomson, Z. F. Ghazoui, K. H. Hull, M. Wexler, A. R. Curson, J. D. Todd, P. S. Poole, T. H. Mauchline, A. K. East, M. A. Quail, C. Churcher, C. Arrowsmith, I. Cherevach, T. Chillingworth, K. Clarke, A. Cronin, P. Davis, A. Fraser, Z. Hance, H. Hauser, K. Jagels, S. Moule, K. Mungall, H. Norbertczak, E.

Rabbinowitsch, M. Sanders, M. Simmonds, S. Whitehead and J. Parkhill (2006). "The genome of Rhizobium leguminosarum has recognizable core and accessory components." Genome Biol **7**(4): R34.

Young, N. D., F. Debelle, G. E. Oldroyd, R. Geurts, S. B. Cannon, M. K. Udvardi, V. A. Benedito, K. F. Mayer, J. Gouzy, H. Schoof, Y. Van de Peer, S. Proost, D. R. Cook, B. C. Meyers, M. Spannagl, F. Cheung, S. De Mita, V. Krishnakumar, H. Gundlach, S. Zhou, J. Mudge, A. K. Bharti, J. D. Murray, M. A. Naoumkina, B. Rosen, K. A. Silverstein, H. Tang, S. Rombauts, P. X. Zhao, P. Zhou, V. Barbe, P. Bardou, M. Bechner, A. Bellec, A. Berger, H. Berges, S. Bidwell, T. Bisseling, N. Choisne, A. Couloux, R. Denny, S. Deshpande, X. Dai, J. J. Doyle, A. M. Dudez, A. D. Farmer, S. Fouteau, C. Franken, C. Gibelin, J. Gish, S. Goldstein, A. J. Gonzalez, P. J. Green, A. Hallab, M. Hartog, A. Hua, S. J. Humphray, D. H. Jeong, Y. Jing, A. Jocker, S. M. Kenton, D. J. Kim, K. Klee, H. Lai, C. Lang, S. Lin, S. L. Macmil, G. Magdelenat, L. Matthews, J. McCorrison, E. L. Monaghan, J. H. Mun, F. Z. Najar, C. Nicholson, C. Noirot, M. O'Bleness, C. R. Paule, J. Poulain, F. Prion, B. Qin, C. Qu, E. F. Retzel, C. Riddle, E. Sallet, S. Samain, N. Samson, I. Sanders, O. Saurat, C. Scarpelli, T. Schiex, B. Segurens, A. J. Severin, D. J. Sherrier, R. Shi, S. Sims, S. R. Singer, S. Sinharoy, L. Sterck, A. Viollet, B. B. Wang, K. Wang, M. Wang, X. Wang, J. Warfsmann, J. Weissenbach, D. D. White, J. D. White, G. B. Wiley, P. Wincker, Y. Xing, L. Yang, Z. Yao, F. Ying, J. Zhai, L. Zhou, A. Zuber, J. Denarie, R. A. Dixon, G. D. May, D. C. Schwartz, J. Rogers, F. Quetier, C. D. Town and B. A. Roe (2011). "The Medicago genome provides insight into the evolution of rhizobial symbioses." Nature **480**(7378): 520-524.

Zahran, H. H. (1999). "Rhizobium-legume symbiosis and nitrogen fixation under severe conditions and in an arid climate." Microbiol Mol Biol Rev **63**(4): 968-989, table of contents.

Zebracki, K., P. Koper, M. Marczak, A. Skorupska and A. Mazur (2015). "Plasmid-Encoded RepA Proteins Specifically Autorepress Individual repABC Operons in the Multipartite Rhizobium leguminosarum bv. trifolii Genome." PLoS One **10**(7): e0131907.

APPENDIX

Table 6. Supplementary Himar 1 transposon insertions in C377, C017, and B469.

Mutant Designation*	*Himar* 1 location (strand)	Gene	Phenotype
pHRC377			
R2P2	52784 (+)	*repA2*	weak Fix$^+$
R2P5	164193 (+)	*luxR*	weak Fix$^+$
R2P8	54726 (+)	*repC2*	Fix+
R2P12	164796 (+)	*RR*	weak Fix$^+$
R2P16	163477 (+)	*luxR*	weak Fix$^+$
R2P19	164554 (-)	*RR* *repA2*	weak Fix$^+$
R2P25	51979 (+)	*promoter*	weak Fix+
R2P30	52232 (-)	*repA2*	weak Fix+
R2P44	52614 (+)	*repA2*	weak Fix+
R2P49	52615 (-)	*repA2*	weak Fix+
R2P63	53511 (-)	*repA2*	weak Fix+
R2P87	53154 (-)	*repA2*	weak Fix+
pHRC017			
R2P8	90418 (-)	*repA2*	weak Fix+
R2P12	92615 (+)	*repC2*	Fix+
R2P13	91120 (-)	*repA2*	Fix+
R2P36	91185 (+)	*repA2*	weak Fix+
R2P39	92183 (-)	*repC2*	Fix+
pHRB469	(based off of pHRC017)		
R2P16	91099 (+)	*repA2*	weak Fix+
R2P22	91185 (+)	*repA2*	weak Fix+
R2P32	262500 (-)	*luxR*	weak Fix+

*Mutant transposon strains were created as described in materials and methods. Mutants were collected from fix$^+$ nodules and retested for retention of phenotype. Weak fix$^+$ phenotypes were determined by site as they were compared to GOC derivative positive controls. The pHRC377 plasmid was selected for further study because it was the easiest to manipulate genetically and had been sequenced.

Table 7. Supplementary Protein sequence alignments repABC.

	pSymA RepA	pSymB RepA	pHRC377 RepA1	pHRC377 RepA2	pHRC377 RepA3
pSymA RepA*	100	42.86	42.75	34.93	33.33
pSymB RepA	42.86	100	57.72	36.51	36.29
pHRC377 RepA1	42.75	57.72	100	38.34	35.55
pHRC377 RepA2	34.93	36.51	38.34	100	64.89
pHRC377 RepA3	33.33	36.29	35.55	64.89	100

	pSymA RepB	pSymB RepB	pHRC377 RepB1	pHRC377 RepB4
pSymA RepB	100	34.81	30.06	24.26
pSymB RepB	34.81	100	38.77	21.59
pHRC377 RepB1	30.06	38.77	100	23.43
pHRC377 RepB4	24.26	21.59	23.43	100

	pSymA RepC	pSymB RepC	pHRC377 RepC1	pHRC377 RepC2	pHRC377 RepC3	pHRC377 RepC4
pSymA RepC	100	51.62	58.85	98.23	12.86	25.68
pSymB RepC	51.62	100	49.26	51.26	17.33	25.89
pHRC377 RepC1	58.85	49.26	100	58.33	15.28	26.54
pHRC377 RepC2	98.23	51.26	58.85	100	12.86	25.75
pHRC377 RepC3	12.86	17.33	15.28	12.86	100	72
pHRC377 RepC4	25.68	25.89	26.54	25.75	72	100

*****Every protein alignment was done using Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/). Each number represents the percent of amino acid similarity between proteins. All RepA, B, and C proteins commonly found in *S.* meliloti were included in the analysis.

Table 8. Supplementary Bacterial strains used in this study.

Strain*	Genotype	Source/reference
B001	*E.coli* DH5α + pRK600 conjugation plasmid, Cm^R	(Griffitts and Long 2008)
B100	*S. meliloti* RM1021	(Griffitts and Long 2008)
C237	*A. tumefaciens* UBAPF2, Rf^R	(Crook, Lindsay et al. 2012)
C377	*S. meliloti* M256, Sm^R	(Crook, Lindsay et al. 2012)
C378	*S. meliloti* C377 pHR^- ($pHRC377^-$), Sm^R	(Crook, Lindsay et al. 2012)
C389	*S. meliloti* C377 + $pHRC377Km^R$ (pJG498), Sm^R	(Crook, Lindsay et al. 2012)
C396	*A. tumefaciens* UBAPF2 + $pHRC377Km^R$(pJG498), Sm^R	(Crook, Lindsay et al. 2012)
CP03	*S. meliloti* C377 + Himar 1 insertion R2P2 in repA2, Sm^R , Km^R	This study
CP05	*S. meliloti* C377 + *Himar*1 insertion R2P8 in *repC2*, Sm^R, Km^R	This study
CP07	*S. meliloti* C377 + *Himar*1 insertion R2P5 in *luxR*, Sm^R, Km^R	This study
CP23	*S. meliloti* C389 *repA2C2* clean deletion, Sm^R, Km^R	This study
CP28	*S. meliloti* C377 + pCP23 (loop-in disruption *repC2* with Ptrp:GUS), Sm^R, Km^R	This study
CP29	*S. meliloti* C377 + pCP24 (loop-in non-disruption *repC2* with Ptrp:GUS) Sm^R, Km^R	This study
CP37	*S. meliloti* C377 + pCP16 (disruption loop-in *repC1*), Sm^R, Km^R	This study
CP38	*S. meliloti* C377 + pCP26 (non-disruption loop-in *repC1*), Sm^R, Km^R	This study
PP538	*S. meliloti* C378 + pCP36, Sm^R, Km^R	This study
PP539	*S. meliloti* C378 + pCP39, Sm^R, Km^R	This study
PP540	*S. meliloti* C378 + pCP40, Sm^R, Km^R	This study
PP541	*S. meliloti* C378 + pCP42, Sm^R, Km^R	This study
PP596	*S. meliloti* C389 + *repA2* clean deletion, Sm^R, Km^R	This study

Table 9. Supplementary Plasmids used in this study.

Plasmid	Relevant Features	Source/reference
pCP16	pJG194(Griffitts and Long 2008) carrying *repC1* internal fragment, Km^R	This study
pCP18	pJQ200 (Quandt and Hynes 1993) derivative carrying *repA2C2* upstream and downstream regions for knockout, Gm^R	This study
pCP23	pPP244 carrying internal *repC2* fragment and Prp:GUS, Km^R	This study
pCP24	pPP244 carrying fragment of *repC2* to do a non-disruption loop-in and Ptrp:GUS, Km^R	This study
pCP26	pJG194(Griffitts and Long 2008) carrying fragment of *repC1* to do a non-disruption loop-in, Km^R	This study
pCP33	pQJ200 (Quandt and Hynes 1993) derivative carrying *repA2* upstream and downstream regions for an in frame knockout, Gm^R	This study
pCP35	Plasmid carrying *sc101* replication origin from pJG385 with *oriT* from pJG194 (Griffitts and Long 2008), Km^R	This study
pCP36	pCP35 carrying *repA2C2*, Km^R	This study
pCP38	pCP35 carrying *luxR*/response regulator, Km^R	This study
pCP39	pCP35 carrying *repA1B1C1*, Km^R	This study
pCP40	pCP35 carrying *repABC* from pSymB, Km^R	This study
pCP42	pCP35 carrying *repB4C4*, Km^R	This study
pJG498	pJG194(Griffitts and Long 2008) carrying pHRC377 intergenic region, Km^R	(Crook, Lindsay et al. 2012)
pPP244	pJG194 (Griffitts and Long 2008) carrying *Ptrp* promoter from pRF771 (Price, Jin et al. 2012)–GUS fusion, Km^R	This study

Table 10. Supplementary Oligonucleotides used in this study.

Oligonucleotide	Sequence	Source/reference
oCP71	GCGGGATCCGGGCTGACGCGTACAGGAAACA	This study
oCP72	TTATGATGGATCGAAGTTATCCCGGGCTCTTTC	This study
oCP73	CCGGGATAACTTCGATCCATCATAACGGTTCCGG	This study
oCP74	CTTATACACATCTAGATGTGAAC	This study
oCP92	CGCGGATCCGCGAGATCATTTCGGAATCAGG	This study
oCP93	GCGGAATTCGGTGCTCGTAGAAGTGCCGG	This study
oCP96	AAGGCCCTCCTCGCGATAGA	This study
oCP97	CACCGGCTCGAAACAGCTTG	This study
oCP99	GCGGGATCCGACGAAACTGGCGGAATCGAAGAG	This study
oCP100	CGCTCTAGACTCCATTTCCTCGAGGATTTGCTGT	This study
oCP101	GCACTGGCGCAAAGACGTAG	This study
oCP102	GCGCTCGAGCATGAGGTCGCGCCAACTTC	This study
oCP103	GCGCTCGAGATTGGCAGACGCCCAGCCATG	This study
oCP105	CACGCTCGAGCCGTTGATGATACCGCTGCC	This study
oCP106	CAGCCTCGAGGTCGACGGTATCGATAAG	This study
oCP107	GCCGGTACCTGGGCTGCCCTTCCT	This study
oCP108	CGGATTATCCCGTGACAGGTC	This study
oCP109	GCTCTTGTATCTATCAGTGAAGC	This study
oCP116	TCGATCACGTTGAAGACGCGGAA	This study
oCP117	CGAAGGAAGAAACTTCACACCAGCAGA	This study
oCP118	CGCTCTAGAGCACTGGCGCAAAGACGTAG	This study
oCP119	CGGTACAGCCCTGATGCTCC	This study

oCP120	CAGCTCTAGATGCTCGCGCAGGATCGAAATG	This study
oCP121	CATGTCTAGACCTTGCAGAAACATCGCCTCTC	This study
oCP122	GCGTCTAGAGCGAGTTGGAGCTAGTCTCTGGG	This study
oCP123	CGCTCTAGACCCAAAGATGAGCTCGCTTGC	This study
oCP124	GCGTCTAGACCTGGAAAGCTGGCTCAAGGTG	This study
oCP126	CGCTCTAGAGTGCGCCCGAGATTCCTGTC	This study
oCP127	GCGTCTAGACTGAAAAGCGTTCTGCGGGAC	This study
oCP128	CGCTCTAGACCGGGAAGATTTGGAACGGC	This study
oCP129	CAAGGATTCCGCGTTCTCGC	This study
oCP130	CCGTCAATGCGGAGGTGATGG	This study
oCP131	GCTGCCAAACGCCGTCTTGG	This study
oCP132	GCGTCTAGAGTAGCCGAGGGTCATTCCGC	This study
oCP133	GCGTCTAGACGGCCATGCGCAAATTGGTC	This study
oCP134	CGCGGATCCCTTGCCGTTCGGACTGTCCT	This study
oCP135	TGAGGGAATGTATGCCACAAGGAATATAAACGCAAAAGAA	This study
oCP136	TTCTTTTGCGTTTATATTCCTTGTGGCATACATTCCCTCA	This study
oCP137	CCGTCGTGAGGGAATGTATGCCACAAGGAATATAAACGCAAAAGAAAAAGGCCCC	This study
oCP138	GGGGCCTTTTTCTTTTGCGTTTATATTCCTTGTGGCATACATTCCCTCACGACGG	This study
oCP139	CGAAACCGGCCTACTTGC	This study
oCP140	CGCATCGATGAACCAGTTCC	This study
oCP141	CGAGCTAAAGGCGTCGATCG	This study
oCP142	CTCTTCGATTCCGCCAGTTTCGTC	This study
oCP143	CCACCGACAGAGAACGACGT	This study
oCP144	ACGTCGTTCTCTGTCGGTGG	This study
oCP145	CGGTCAATGCGGTGGAGAAGAA	This study
oCP146	GCGCAGGTATTGGATAGCATCG	This study
oCP147	CGATGCTATCCAATACCTGCGC	This study
oCP148	CCCAACGACGGACCGCAG	This study
oCP149	CTGCGGTCCGTCGTTGGG	This study
oCP150	GGAAGTGTGACGACGCCCT	This study
oCP151	AGGGCGTCGTCACACTTCC	This study
oCP152	CGCGGTACCGCTTCTAACGGTGAACAGTTGTTC	This study
oPP040	GCTCTAGAGGCCATGCGCAAATTGGTCATG	This study
oPP052	ACGGCATTATCACATACATTCCCTCACGACGGTT	This study
oPP053	AGGGAATGTATGTGATAATGCCGTCGAAGCTTGA	This study
oPP166	CCGCTCACCTACAGCTTTGAAAAGAC	This study
oPP167	TGAAGAACGCGAGTCACCATGCCG	This study
oPP168	GTTAGCCGCTTCACTATACATCCGAG	This study
oPP169	CCTTTTCCGTTGTAAGAGGTGCGG	This study
oPP170	CGGCTACTATCTCCTCCTACACCAAG	This study
oPP171	CAGATCTCATCGCGAAATCTCACCTG	This study
oPP405	GATCGAATTCCATTTTGGAGTGGCTTGTGGAGC	This study
oPP406	GATCCTCGAGCTGCCGGTCGTATTTCTCAAGGGTC	This study
oPP407	GATCCTCGAGCGATGACAAGGGCGGCGAAAAG	This study

Table 11.Supplementary qPCR Oligonucleotides.

Oligonucleotide	Target	Sequence	Source/reference
oCP67	pHRC377	GCTCAGCAAGGCTGTAGTATT	This study
oCP68	pHRC377	GATCAACCCACCGATGATACTG	This study
oPP178	*S. meliloti* chromosome	CTGCTGCTCACCTTCTTCTT	This study
oPP179	*S. meliloti* chromosome	CTTGAGATACTGGACGGACTTG	This study