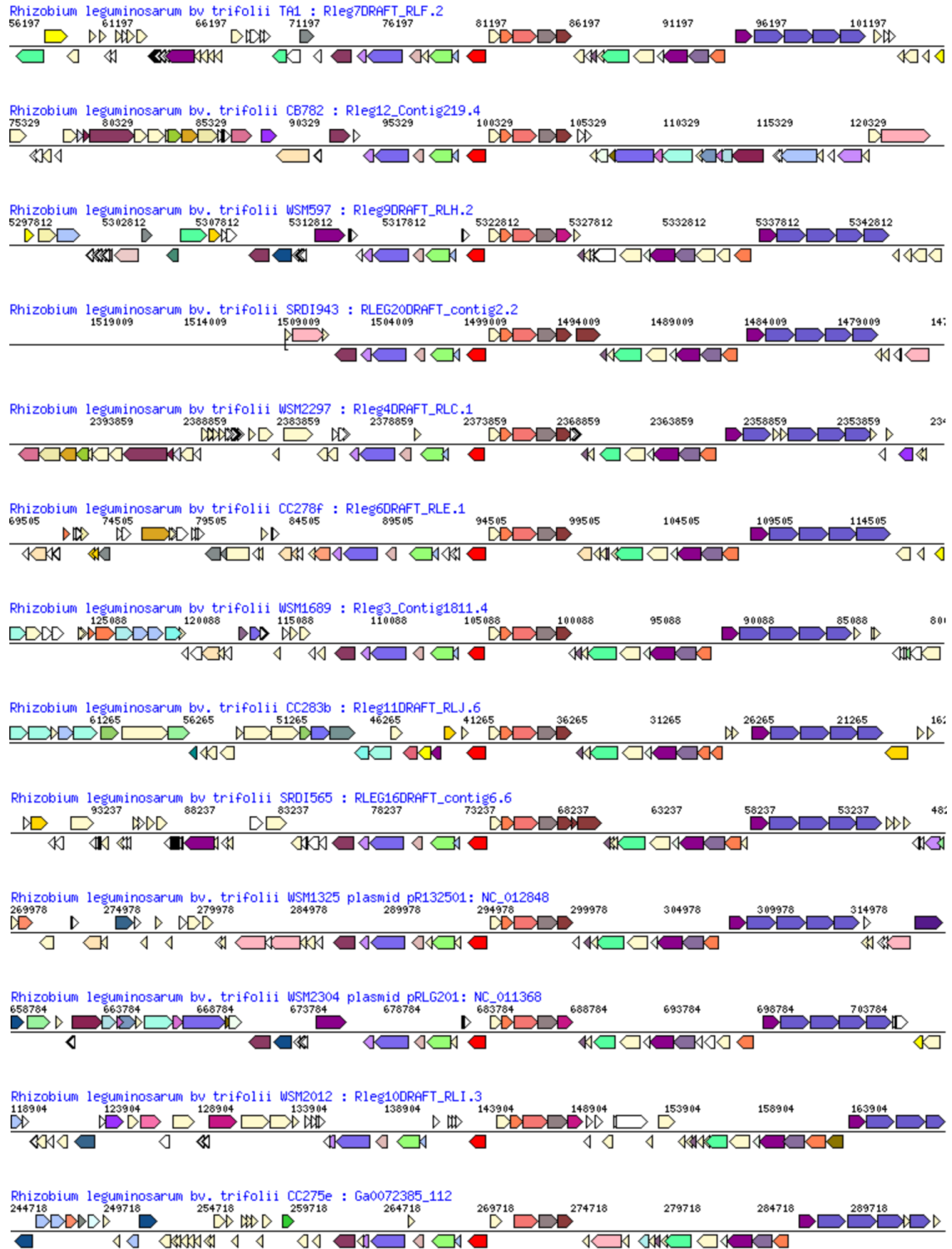


SUPPLEMENTAL MATERIAL

A)



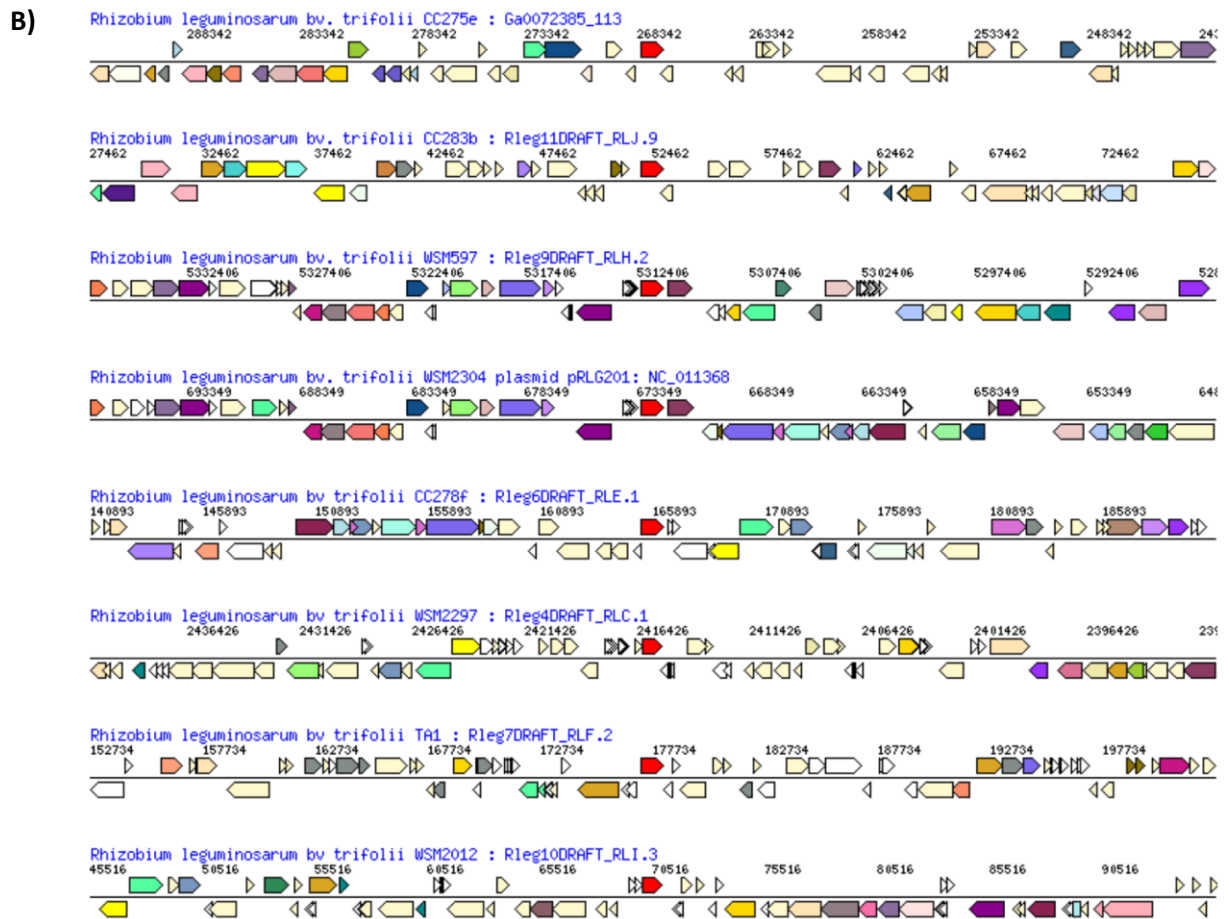


Fig. S1 Conserved gene neighborhoods surrounding *nodD1* and *nodD2* in 13 strains of *Rhizobium leguminosarum* bv. *trifolii* determined using the IMG database (1, 2). A) Genetic context of 13 *nodD1* genes; B) Genetic context of eight *nodD2* genes. The *nodD* genes are coloured red.

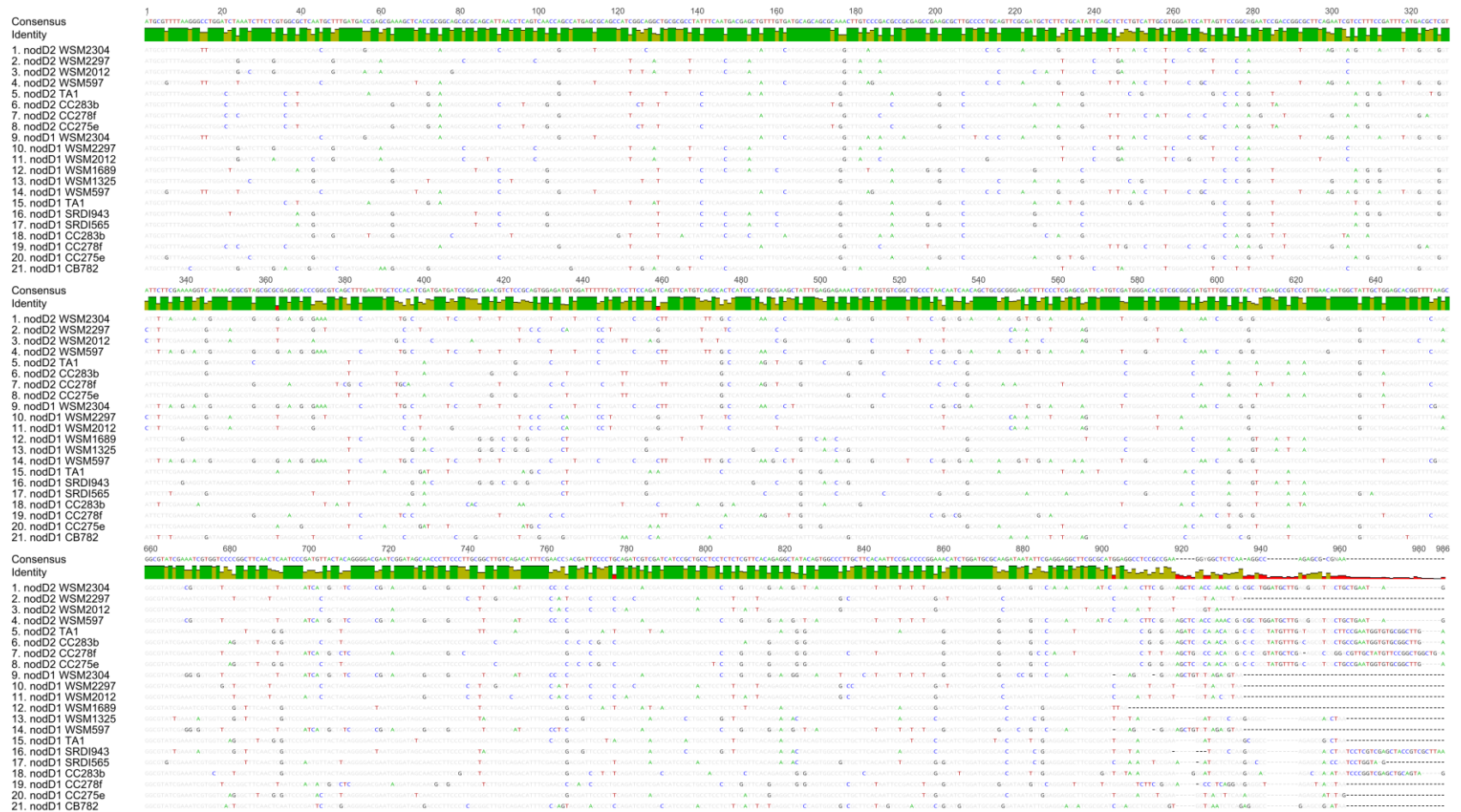


Fig. S2 Alignment of 21 *nodD1* and *nodD2* nucleotide sequences from 13 strains of *Rhizobium leguminosarum* bv. *trifolii*. The *nodD* and strain of origin is indicated on the left. The top line indicates the consensus identity and the bar below illustrates the level of conservation. Nucleotides that vary from the consensus are highlighted.

c)

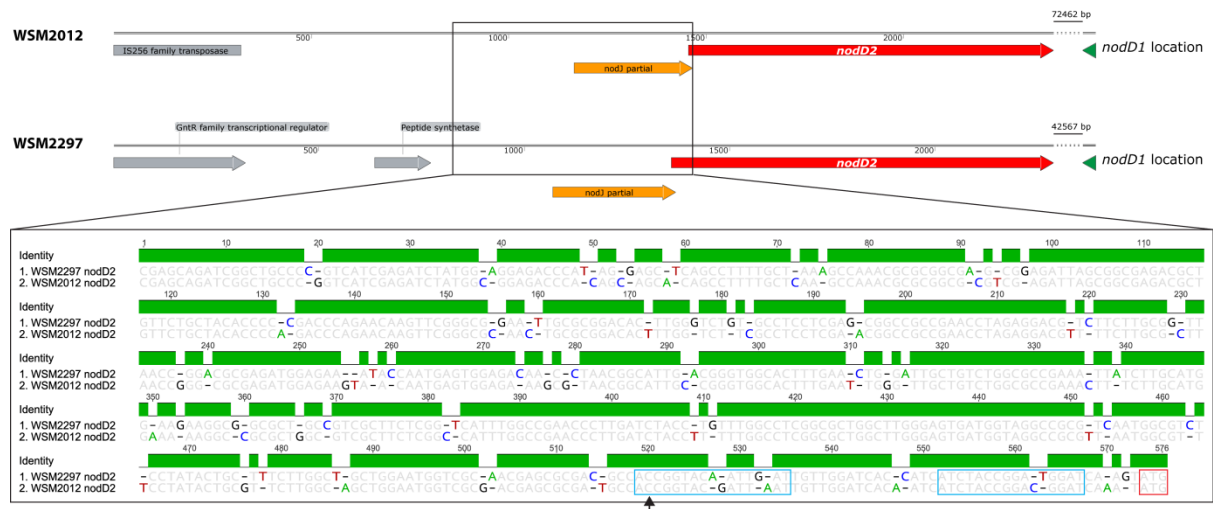


Fig. S3. The genetic context and sequence alignments of the *nodD2* upstream promoter regions in the eight *Rhizobium leguminosarum* bv. *trifolii* strains found to encode *nodD2*. There appeared to be three distinct promoter region subgroups, therefore the alignments were separated into three groups across the eight strains to enable interpretation of the conservation between similar regions. The genetic context images were produced by importing sequence from IMG into SnapGene Viewer. The alignments were performed using Geneious and the green bar above sequence indicates conservation. A) *nodD2* upstream regions of strains TA1, CC283b, CC275e, and CC278f; B) *nodD2* upstream regions of strains WSM2304 and WSM597; C) *nodD2* upstream regions of strains WSM2012 and 2297, The *nod*-box motifs are boxed in blue and the base substitution disrupting the LysR motif is indicated by a vertical black arrow. The start codon of *nodD2* is boxed in red.

Table S1. Distance matrix of nucleotide similarity for the 21 *nodD* genes in the 13 strains of *Rhizobium leguminosarum* bv. *trifolii*. The highest and lowest percentage identities are indicated in green and orange, respectively.

nodD1_WSM2297																					
nodD2_WSM2297	100																				
nodD1_WSM2012	97	98																			
nodD2_WSM2012	94	95	94																		
nodD1_CB782	90	91	90	90																	
nodD1_CC275e	85	86	85	86	84																
nodD1_TA1	85	86	85	85	84	95															
nodD1_CC283b	85	85	85	85	83	89	89														
nodD1_SRDI943	84	84	84	84	83	90	90	88													
nodD1_WSM1325	84	84	83	84	82	90	91	89	97												
nodD1_WSM1689	85	85	85	84	84	92	90	90	97	94											
nodD1_SRDI565	86	86	86	85	84	91	90	89	94	92	94										
nodD2_CC275e	84	85	84	85	83	90	89	86	87	89	90	89									
nodD2_CC283b	85	85	84	86	84	90	90	87	88	90	91	90	99								
nodD2_TA1	84	84	84	84	82	88	89	86	87	89	91	89	92	92							
nodD1_CC278f	84	84	84	83	82	83	83	83	83	83	84	84	85	85	86						
nodD2_CC278f	84	84	84	83	82	83	82	81	80	82	84	82	83	83	84	93					
nodD2_WSM2304	78	79	79	79	78	78	78	75	76	78	79	77	77	77	79	87	84				
nodD2_WSM597	79	79	79	79	78	78	78	76	76	77	79	77	77	77	79	87	84	99			
nodD1_WSM597	78	78	78	78	78	78	78	76	77	77	79	78	79	79	80	87	84	96	97		
nodD1_WSM2304	79	79	79	79	78	79	78	78	78	78	79	78	79	79	81	88	86	95	95	97	

* Note values are rounded to the nearest whole number.

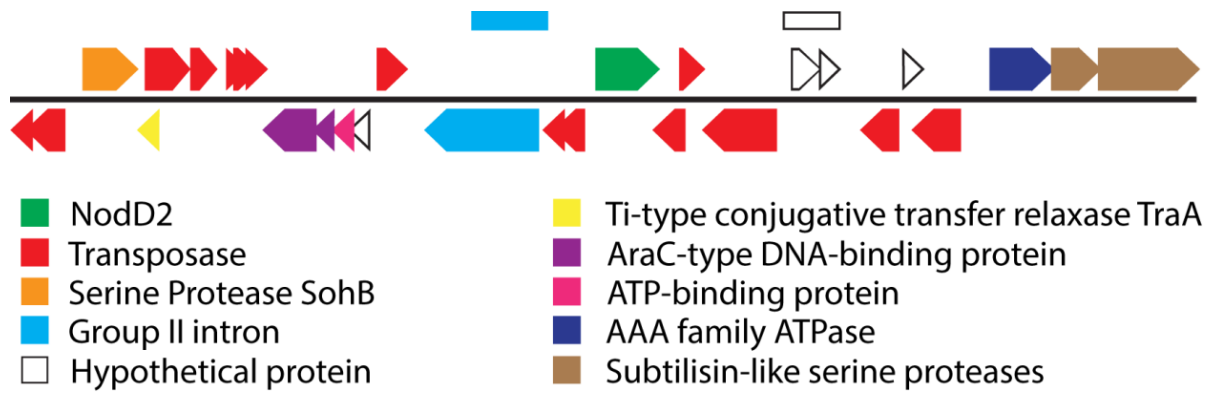


Fig. S4 The genetic context of *nodD2* in *Rhizobium leguminosarum* bv. *trifolii* strain TA1. Open reading frames surrounding *nodD2*. The boxes above indicate the location of the flanking arms that were designed to facilitate construction of a *nodD2* mutant. The left arm and right arm are indicated by a light blue and white box, respectively. The putative functions of genes surrounding *nodD2* were determined by BLAST search analysis.

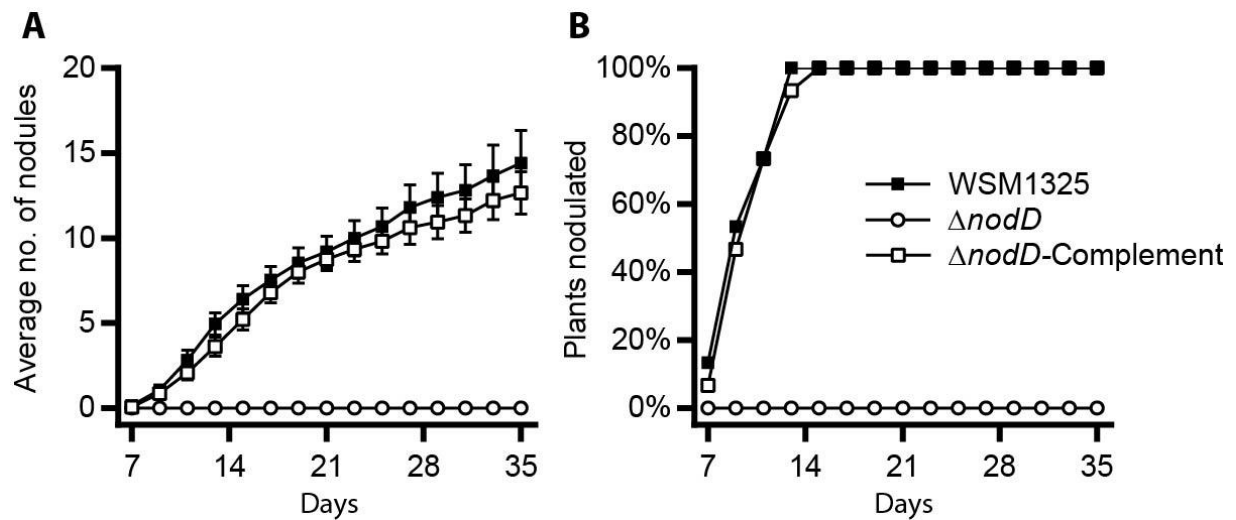


Fig. S5 Nodulation characteristics of WSM1325, WSM1325 $\Delta nodD$, and WSM1325 $\Delta nodD$ complemented with *nodD* on white clover. A) Nodulation kinetics showing the average number of nodules per plant. B) The percentage of plants nodulated. The nodulation kinetics was recorded from seven to 35 dpi. Error bars represent the SEM.

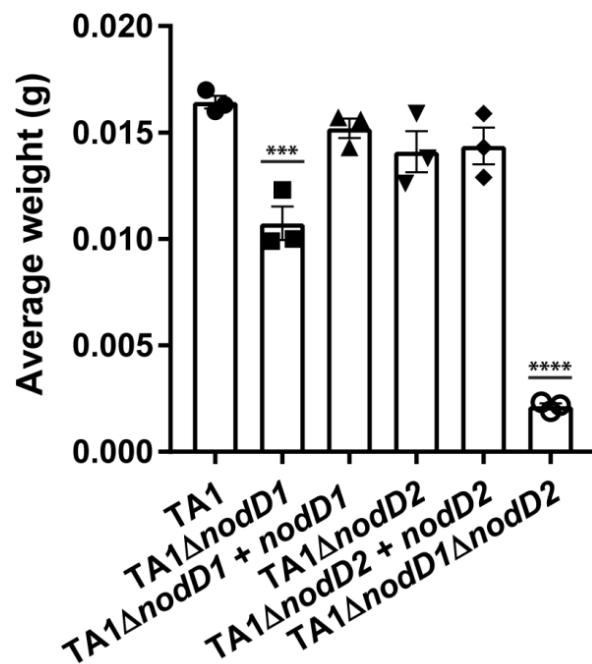


Fig. S6 The dry shoot weights of white clover inoculated with TA1 *nodD* mutant derivatives at 42 dpi. For every *Rhizobium leguminosarum* bv. *trifolii* strain three biological replicates were each inoculated onto 10 plants. The average shoot weight of the 10 plants per replicate is represented by a symbol. The error bars represent \pm the SEM. The asterisks indicate significant difference from the wild-type (one-way ANOVA with Dunnett's multiple comparisons post hoc test, *** $P < 0.001$, **** $P < 0.0001$).

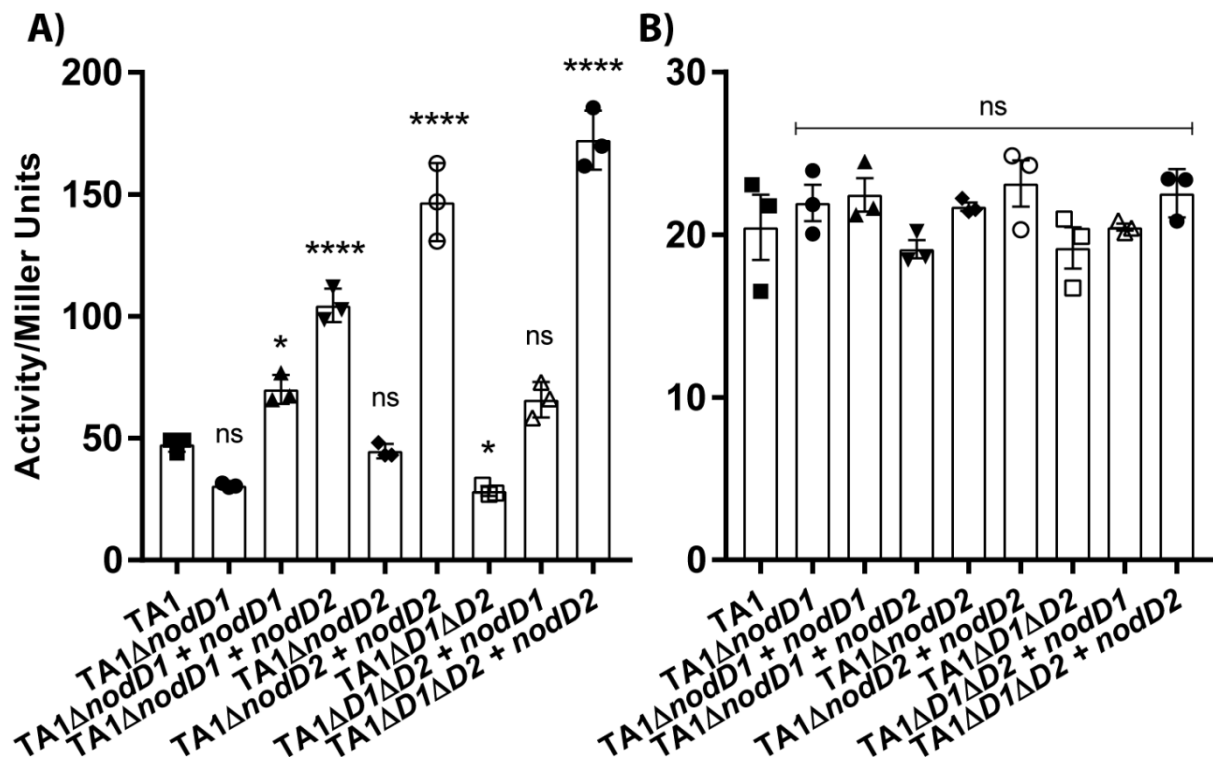


Fig. S7 β -galactosidase assays showing uninduced expression of the *nodA* and *nodF* promoters in TA1 *nodD* mutant derivatives containing promoter reporter plasmids. A) *nodA* promoter induction. B) *nodF* promoter induction. Each symbol represents a biological replicate. Error bars represent \pm SEM. The asterisks indicate significant difference from the wild-type (One-way ANOVA with Dunnett's multiple comparisons post hoc test, * $P < 0.05$, **** $P < 0.0001$).

Table S2. Distance matrix of amino acid similarity for the 21 NodD proteins in the 13 strains of *Rhizobium leguminosarum* bv. *trifolii*. The highest and lowest percentage identities are indicated in green and orange, respectively.

nodD1_WSM2297																						
nodD2_WSM2297	100																					
nodD1_WSM2012	98	98																				
nodD1_CB782	92	92	91																			
nodD2_WSM2012	95	95	95	92																		
nodD1_SRDI943	88	88	88	83	88																	
nodD1_WSM1689	88	88	88	85	88	98																
nodD1_SRDI565	87	87	87	83	87	93	97															
nodD1_WSM1325	88	88	88	84	88	97	98	96														
nodD1_CC283b	87	87	87	83	88	88	93	89	91													
nodD1_CC275e	87	87	87	84	87	89	92	89	91	88												
nodD1_TA1	89	89	88	86	89	89	92	90	92	87	97											
nodD2_CC275e	85	85	84	82	86	87	91	87	90	85	88	88										
nodD2_CC283b	86	86	85	83	86	88	92	88	90	86	89	89	99									
nodD2_TA1	85	85	85	81	86	86	92	87	89	86	87	86	94	94								
nodD1_CC278f	85	85	85	84	85	84	86	85	85	84	85	84	86	86	88							
nodD2_CC278f	84	84	83	81	83	81	85	83	83	82	83	83	86	86	88	93						
nodD1_WSM597	80	80	79	79	80	78	81	79	78	78	79	80	81	81	82	90	86					
nodD1_WSM2304	80	80	80	79	81	80	82	80	80	80	80	81	82	82	84	91	88	97				
nodD2_WSM597	80	80	79	77	80	76	81	77	78	76	79	80	80	80	81	88	85	95	94			
nodD2_WSM2304	80	80	79	77	80	76	81	77	78	76	79	80	80	80	81	88	84	95	95	99		

* Note values are rounded to the nearest whole number.

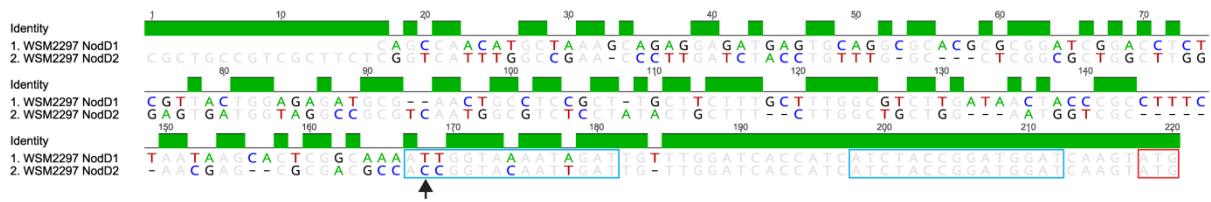
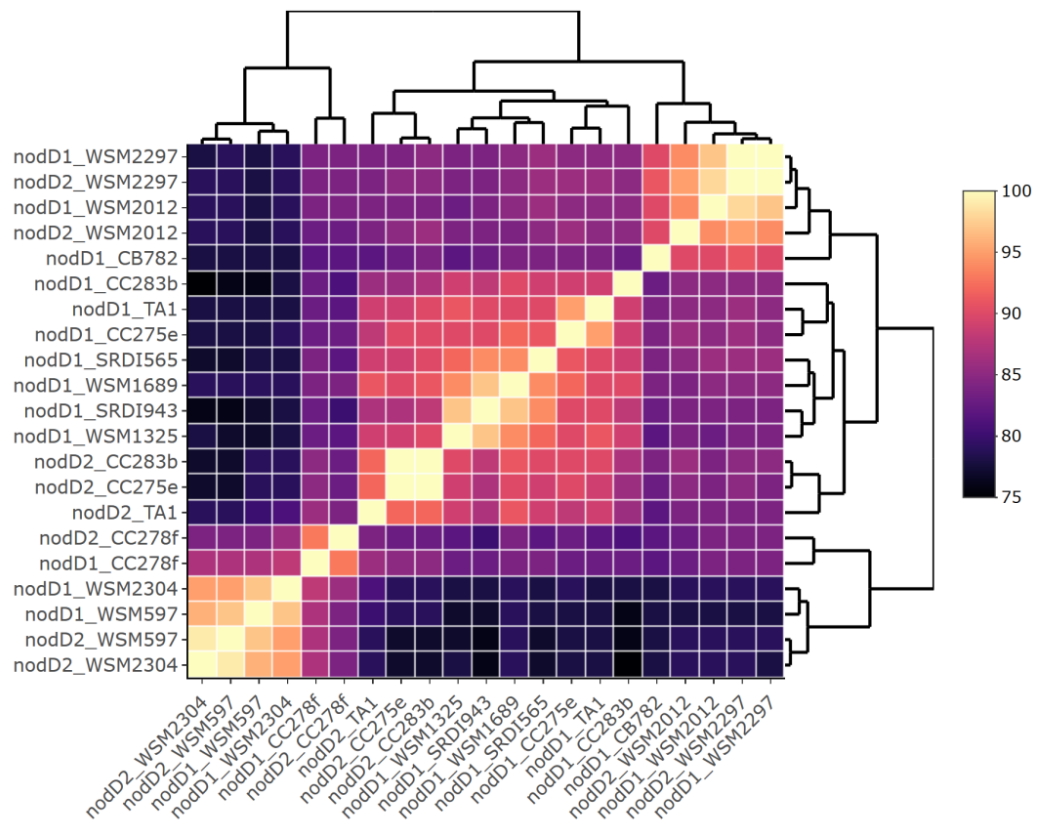
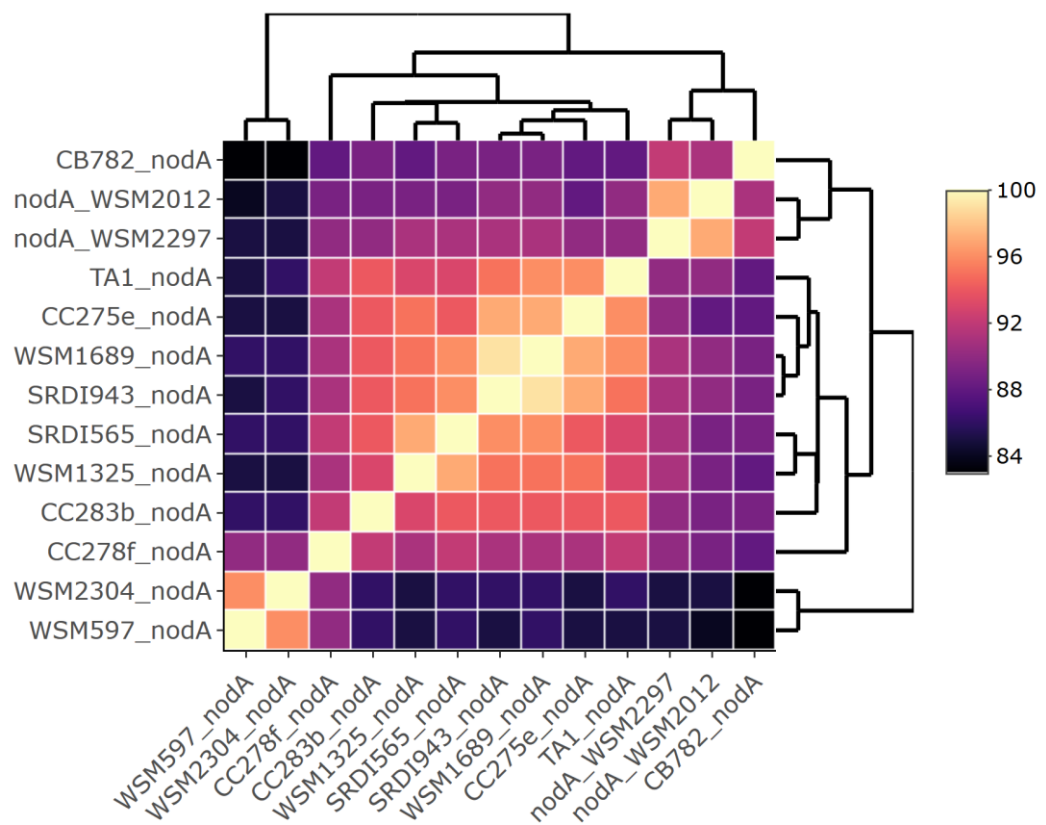


Fig. S8 Nucleotide sequence alignment of the start codon and preceding 197 bp upstream of *Rhizobium leguminosarum* bv. *trifolii* strain WSM2297 *nodD1* and *nodD2* genes. The *nod*-box motifs are boxed in blue and the base substitution disrupting the LysR motif is indicated by a vertical black arrow. The start codon of *nodD1* and *nodD2* is boxed in red. The P- and D-half *nod*-box motifs are located farthest and closest to the *nodD* start codons, respectively.

A)



B)



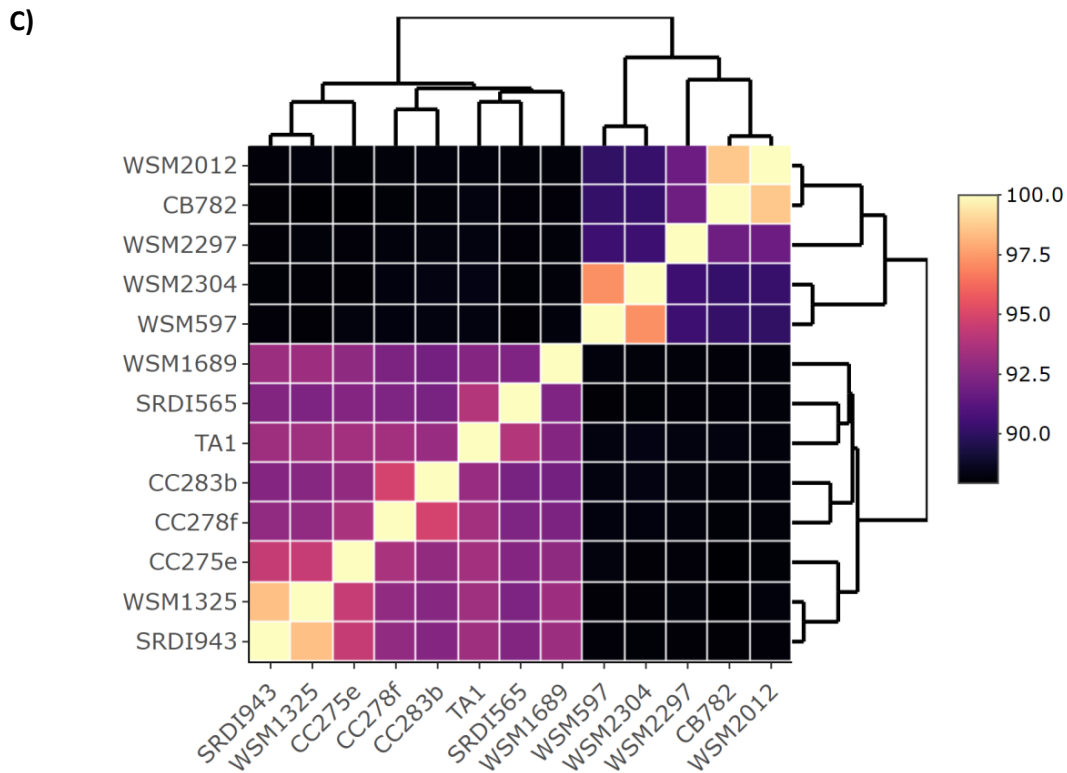


Fig. S9 Heat maps showing pairwise conservation between whole genome, *nodD*, and *nodA* gene sequences across 13 *Rhizobium leguminosarum* bv. *trifolii* strains. A) Heat map of *nodD1* and *nodD2* nucleotide sequences generated from a distance identity matrix constructed using ClustalX2. Distance values range from 75 – 100%. B) Heat map of *nodA* nucleotide sequences generated from a distance identity matrix constructed using ClustalX2. Distance values range from 83 – 100%. C) Heat map of pairwise whole genome ANI determined using OrthoANI. Distance values range from 87.5 – 100%. The heat maps were plotted using heatmaply (3) in RStudio (4, 5). Dendrograms on the top and right indicate the relatedness based on sequence conservation.

Table S3. The IMG database genome ID of the *Rhizobium leguminosarum* strains investigated throughout this study.

<i>Rhizobium leguminosarum</i> strains	IMG Genome ID
<i>Rhizobium leguminosarum</i> bv. <i>trifolii</i>	
strains	
WSM1325	644736401
SRDI565	2517287029
SRDI943	2517093000
WSM1689	2510065019
CB782	2510065076
TA1	2510461076
CC275e	2630968330
CC283b	2507525018
WSM2297	2508501100
WSM2012	2509276033
WSM2304	643348569
WSM597	2509276021
CC278f	2509276052

Rlv strains

UPM791	2859833134
Vc2	2515154113
248	2515075009
TOM	2516653047
WSM1481	2516653077
UPM1131	2513237084
BIHB 1217	2844454524
VF39	2513237103
Vh3	2515154114
BIHB 1148	2844462476
128C53	2515154115
GB30	2513237162
UPM1137	2513237085
Ps8	2515154116
3841	639633055
WSM1455	2509276044

REFERENCES

1. Markowitz VM, Chen I-MA, Palaniappan K, Chu K, Szeto E, Grechkin Y, Ratner A, Jacob B, Huang J, Williams P. 2011. IMG: the integrated microbial genomes database and comparative analysis system. *Nucleic Acids Res* 40:D115-D122.
2. Mavromatis K, Chu K, Ivanova N, Hooper SD, Markowitz VM, Kyrpides NC. 2009. Gene context analysis in the Integrated Microbial Genomes (IMG) data management system. *PLoS One* 4:e7979-e7979.
3. Galili T, O'callaghan A, Sidi J, Sievert C. 2017. heatmaply: an R package for creating interactive cluster heatmaps for online publishing. *Bioinformatics* 34:1600-1602.
4. R Core Team. 2013. R: A language and environment for statistical computing.
5. RStudio Team. 2015. RStudio: integrated development for R. RStudio, Inc, Boston, MA