Title

Evolution of a multi-step phosphorelay signal transduction system in Ensifer: recruitment of the sigma factor RpoN and a novel enhancer-binding protein triggers acid-activated gene expression

Journal

Molecular Microbiology, 103(5)

ISSN

0950-382X

Authors

Tian, Rui

Heiden, Stephan

Osman, Wan AM

et al.

Publication Date

2017-03-01

DOI

10.1111/mmi.13592

Peer reviewed

Molecular Microbiology (2017) 103(5), 829–844
doi:10.1111/mmi.13592
First published online 9 January 2017

Evolution of a multi-step phosphorelay signal transduction system in *Ensifer*: recruitment of the sigma factor RpoN and a novel enhancer-binding protein triggers acid-activated gene expression

Rui Tian,[1] Stephan Heiden,[1] Wan A. M. Osman,[1] Julie K. Ardley,[1] Euan K. James,[2] Margaret M. Gollagher,[3] Ravi Tiwari,[1] Rekha Seshadri,[4] Nikos C. Kyrpides[4] and Wayne G. Reeve[1]*

[1]*School of Veterinary and Life Sciences, Murdoch University, Murdoch, WA 6150, Australia.*
[2]*The James Hutton Institute, Invergowrie, Dundee DD2 5DA, UK.*
[3]*Curtin University Sustainability Policy Institute, Curtin University, Bentley, WA, Australia.*
[4]*DOE Joint Genome Institute, Walnut Creek, CA, USA.*

Summary

Most *Ensifer* strains are comparatively acid sensitive, compromising their persistence in low pH soils. In the acid-tolerant strain *Ensifer medicae* WSM419, the acid-activated expression of *lpiA* is essential for enhancing survival in lethal acidic conditions. Here we characterise a multi-step phosphorelay signal transduction pathway consisting of TcsA, TcrA, FsrR, RpoN and its cognate enhancer-binding protein EbpA, which is required for the induction of *lpiA* and the downstream *acvB* gene. The *fsrR, tcrA*, *tcsA* and *rpoN* genes were constitutively expressed, whereas *lpiA* and *acvB* were strongly acid-induced. RACE mapping revealed that *lpiA*/*acvB* were co-transcribed as an operon from an RpoN promoter. In most *Ensifer* species, *lpiA/acvB* is located on the chromosome and the sequence upstream of *lpiA* lacks an RpoN-binding site. Nearly all *Ensifer meliloti* strains completely lack *ebpA, tcrA*, *tcsA* and *fsrR* regulatory loci. In contrast, *E. medicae* strains have *lpiA/acvB* and *ebpA/tcrA/tcsA/fsrR* co-located on the pSymA megaplasmid, with *lpiA/acvB* expression coupled to an RpoN promoter. Here we provide a model for the expression of *lpiA/acvB* in *E. medicae*. This unique acid-activated regulatory system provides insights into an evolutionary process which may assist the adaptation of *E. medicae* to acidic environmental niches.

Accepted 28 November, 2016. *For correspondence. E-mail W.Reeve@murdoch.edu.au; Tel. +61 8 9360 2631; Fax +61 8 9360 6486.

Introduction

Medicago spp. are important pasture legumes that are nodulated by strains of rhizobia belonging to either *Ensifer* (*Sinorhizobium*) *meliloti* or the closely related *Ensifer medicae* (Rome *et al.*, 1996; Béna *et al.*, 2005). *Medicago* productivity in nitrogen deficient soils is directly impacted by acidity (Munns, 1968), particularly since the *Ensifer* microsymbiont is especially acid-sensitive (Howieson *et al.*, 1988). In acidic soils, the decreased survival of *Ensifer* inoculant strains has prevented their persistence from one season to the next, with detrimental impacts on the re-emergence of a productive Medic stand in the absence of further inoculation (O'Hara *et al.*, 1989). To identify superior microsymbionts better suited to acidic soils, Howieson *et al.* (1988) sourced strains from nodules of *Medicago* plants growing in acidic soil in the Eastern Mediterranean region. One of the first of the acid-tolerant strains to be characterized and used as an inoculant was *E. medicae* WSM419 (Howieson and Ewing, 1986).

Gene inactivation and expression studies have shown that at least three regulatory systems govern acid response in WSM419: *phrR* (*pH* regulated *r*egulator) (Reeve *et al.*, 1998), *actR/S* (*ac*id *t*olerance *r*egulator/*s*ensor) (Tiwari *et al.*, 1996) and *fsrR* (*f*used *s*ensor *r*egulator *R*) (Reeve *et al.*, 2006), which appear to be regulated independently of one another. The *actR/S* genes are essential for acid-tolerance and constitute a signal transduction pathway (Tiwari *et al.*, 1996). The *phrR* gene is induced by general stress and in acid conditions

is upregulated fivefold by an unidentified regulation system (Reeve *et al.*, 1998). The *fsrR* gene encodes a regulatory protein essential for the acid-activation of the *lpiA* (*l*ow *p*H-*i*nducible) gene (Reeve *et al.*, 2006). The *lpiA* gene is one of the most acid-activated (~20-fold) at the sub-lethal pH value of 5.7 relative to pH 7.0 (Reeve *et al.*, 1999). Its expression is specifically induced after cell exposure to acid, thus identifying for the first time a pH-specific regulatory circuit in *E. medicae*. The induction of *lpiA* dropped to threefold in the *fsrR* knockout background but was not completely eliminated, suggesting the involvement of additional regulatory genes. The *lpiA* gene neighbourhood contains several candidate regulatory genes, including *tcsA* and *tcrA* (*t*wo *c*omponent *s*ensor and *t*wo *c*omponent *r*egulator), which are both located directly upstream of *fsrR*. Previous studies have also postulated a role for the alternative sigma factor RpoN (sigma-54) in the regulation of *lpiA* expression (Reeve *et al.*, 2006), as the regulatory region upstream of *lpiA* has an RpoN consensus binding motif similar to the published consensus 5′-TG**G**$_{-24}$CACG-N$_4$-TTG**C**$_{-12}$W-3′ (Barrios *et al.*, 1999; Dombrecht *et al.*, 2002). Hence, RpoN could be important for the expression of both *lpiA* and the downstream *acvB* (*ac*id *v*irulence gene) gene, which appear to be in an operon (Reeve *et al.*, 2006). As RpoN requires a cognate enhance-binding protein (EBP) to initiate transcription (Popham *et al.*, 1989; Morett and Segovia, 1993), we hypothesized that the *ebpA* gene located in the *lpiA* gene neighbourhood upstream of *tcsA* could encode the canonical EBP required for RpoN-dependent acid-activated *lpiA* expression.

In this paper we have investigated the roles of EbpA, TcrA, TcsA and RpoN in the expression of *lpiA* and present a regulatory model which could explain how the *lpiA/acvB* operon is acid activated in *E. medicae*. As the genome sequences of a number of *Ensifer* spp., including WSM419, have now been published (Reeve *et al.*, 2010; Reeve *et al.*, 2015), we also compared the *lpiA/acvB* gene neighbourhood across *Ensifer* species.

Results

Identification of regulatory genes for the acid-activation of lpiA *in* E. medicae

Bioinformatics analysis of the *E. medicae* WSM419 genome located *lpiA/acvB* on the symbiotic megaplasmid pSMED02 (Fig. 1A and B). Within the predicted LpiA protein we identified a putative membrane domain (COG0392) and a lysylphosphatidylglycerol synthetase C-terminal domain (COG2898) (Table S1). The AcvB protein contained a Type IV secretory pathway VirJ component associated with intracellular trafficking, secretion and vesicular transport (COG3946). Proteins encoded by other genes in the *lpiA* neighbourhood contained COG domains associated with signal transduction mechanisms (*fsrR*, *tcrA*, *tcsA* and *ebpA*) and carbohydrate transport and metabolism (Smed_5955). This suggested that *fsrR*, *tcrA*, *tcsA* and *ebpA* formed part of a signal transduction regulatory cascade required for transcriptional activation of *lpiA*.

By analyzing protein family (PF) domains within these encoded regulatory proteins we identified further evidence of signal transduction (Fig. 1C and Table S1): a response regulatory (RR) receiver (REC) domain (PF00072) in FsrR, TcrA and EbpA; the histidine kinase (HK) HisKA domains (PF00512 and PF07568) in TcsA and FsrR, respectively; and a CHASE3 HK sensory domain (PF05227) in TcsA. An additional HATPase_C domain (PF02518) was identified in FsrR and TcsA. The presence of both RR and HK domains in FsrR indicated this protein was a fused two-component signal transduction protein. A DNA binding domain was found in EbpA (PF02954) but not in FsrR or TcrA. In addition, EbpA contained a sigma-54 interaction domain (PF00158).

Gene knockouts and phenotyping

Using homologous recombination, we constructed knockout mutations in *ebpA* (MUR2347 and MUR2348), *tcsA* (MUR2121), *tcrA* (MUR2090), *acvB* (MUR2124) and *rpoN* (MUR2088) in *E. medicae* WSM419 (Table S2) and verified all mutations by PCR using extragenic primers (Table S3). In TY-buffered broth, the mean generation times of the wild-type and mutants were unaltered at pH 7.0 (2.5 h) and at pH 5.7 (8 h). Previous studies have shown that lysinylation of the membrane lipid phosphatidylglycerol (PG) confers resistance to cationic growth inhibitors, such as polymyxin B, under acidic growth conditions (Sohlenkamp *et al.*, 2007 and Arendt *et al.*, 2013). The *Pseudomonas aeruginosa acvB* ortholog PA0919 has an alpha/beta hydrolase fold (IPR029058) and has been shown to hydrolyse aminoacyl-PG (resulting in the release of alanine, glycine or lysine from PG). The inactivation of PA0919 confers sensitivity to cationic growth inhibitors, including polymyxin B (Arendt *et al.*, 2013). As *E. medicae* WSM419 AcvB also contains this alpha beta hydrolase fold domain (Table S1), we therefore characterized the phenotypes of the wild-type and various mutants in response to polymyxin B. At pH 7.0, no growth occurred for any strain incubated in the presence of 3 μg ml^{-1} of the antibiotic. In contrast, at pH 5.7, cells were able to grow in the presence of 3 μg ml^{-1} polymyxin B (Table 1), albeit with slight differences in the number of

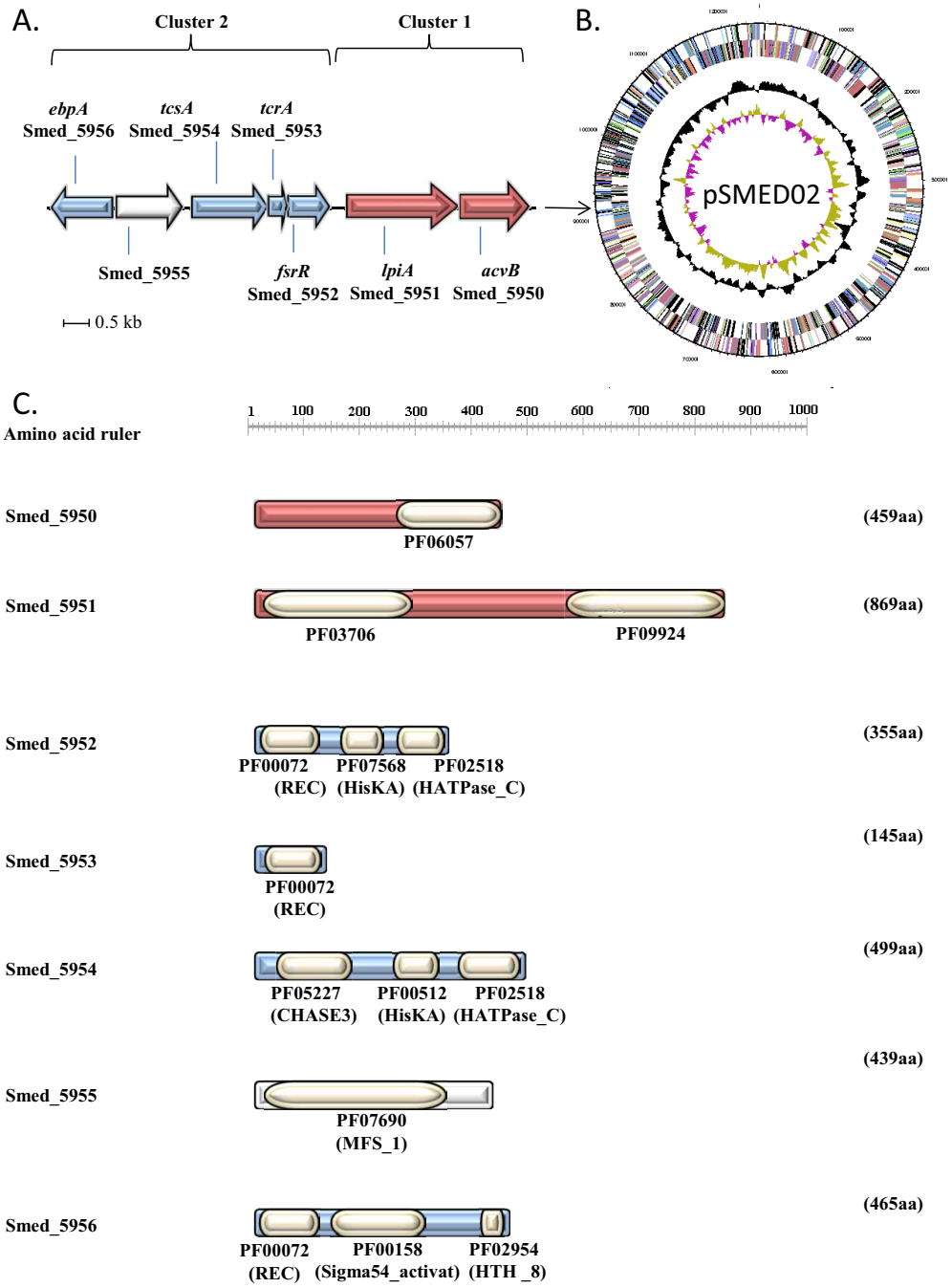


Fig. 1. A. The arrangement of genes within the *lpiA* gene neighbourhood of *Ensifer medicae* WSM419.
B. Map of the WSM419 symbiotic megaplasmid revealing the location of the *lpiA* gene region.
C. Protein domain structures within each of the proteins encoded by genes in the *lpiA* neighbourhood. The N-terminus HisKA dimerization/phosphoacceptor domain contains the conserved histidine residue (the TcsA H box = H_{259}) necessary for autophosphorylation-induced dimer formation based on alignments with previously characterized HisKA proteins.

generations possible. In these conditions, growth of the *E. medicae acvB* mutant was slightly reduced, indicating increased sensitivity to polymyxin B compared to the wild-type. Similar to the *acvB* mutant, growth of the *fsrR* and *tcrA* mutants was slightly reduced compared to the wild-type, whereas growth of the *tcsA* and *ebpA* mutants was more inhibited in these conditions (Table 1).

We assessed the symbiotic phenotypes of the wild-type and mutants on *Medicago murex* Willd., *Medicago sativa* L., *Medicago polymorpha* L. and *Medicago*

Table 1. Number of generations of acid-adapted wild-type WSM419 and derived mutants in TY broth buffered to pH 5.7 and supplemented with polymyxin B.

	Polymyxin B (μg ml^{-1})	
Strains	0	3.0
WSM419 (wild-type)	7.5 ± 0.1	7.7 ± 0.0
MUR2121 (*tcsA*)	7.3 ± 0.0	5.5 ± 0.8*
MUR2090 (*tcrA*)	7.4 ± 0.0	6.8 ± 0.0**
MUR1973 (*fsrR*)	7.4 ± 0.0	6.9 ± 0.0**
MUR2124 (*acvB*)	7.4 ± 0.1	7.1 ± 0.6***
MUR2347 (*ebpA*)	7.4 ± 0.0	5.8 ± 0.6*

Data are the mean of two biological replicates ±SD. Data were analyzed using the unequal variance *t*-test (Ruxton, 2006).
*$P \leq 0.002$.
**$P \leq 0.001$.
***$P \leq 0.05$.

truncatula Gaertn. harvested after 6 weeks (post-seedling inoculation). Plants showed no obvious differences in size, color, nodule number or top dry weight except when MUR2088 (*rpoN* mutant) was used as the inoculant. In this case, the plants were stunted, displayed necrotic symptoms, a higher nodule number and produced less top dry weight, indicating that the *rpoN* mutant failed to fix nitrogen. In co-inoculation experiments with the wild type at low pH, we found no difference in nodulation competitiveness for the *lpiA* and *acvB* mutants. We compared micrographs of nodule sections taken from 6-week-old *M. murex* plants inoculated with the wild-type strain (Fig. 2A–C) with those taken from plants inoculated with MUR1169 (*lpiA* mutant) (Fig. 2D–F) and MUR2124 (*acvB* mutant) (Fig. 2G–I). In general, there were no obvious differences in the structure of nodules formed by the two mutant strains compared to those formed by the wild-type strain, except that the *acvB* mutant bacteroids were abnormally shaped (e.g., often elongated and occasionally pleomorphic) compared to the wild-type and *lpiA* mutant bacteroids (Fig. 2I), but this had no obvious impact on plant biomass. Using a *lpiA–gusA* fusion, we determined that expression of *lpiA* was predominantly

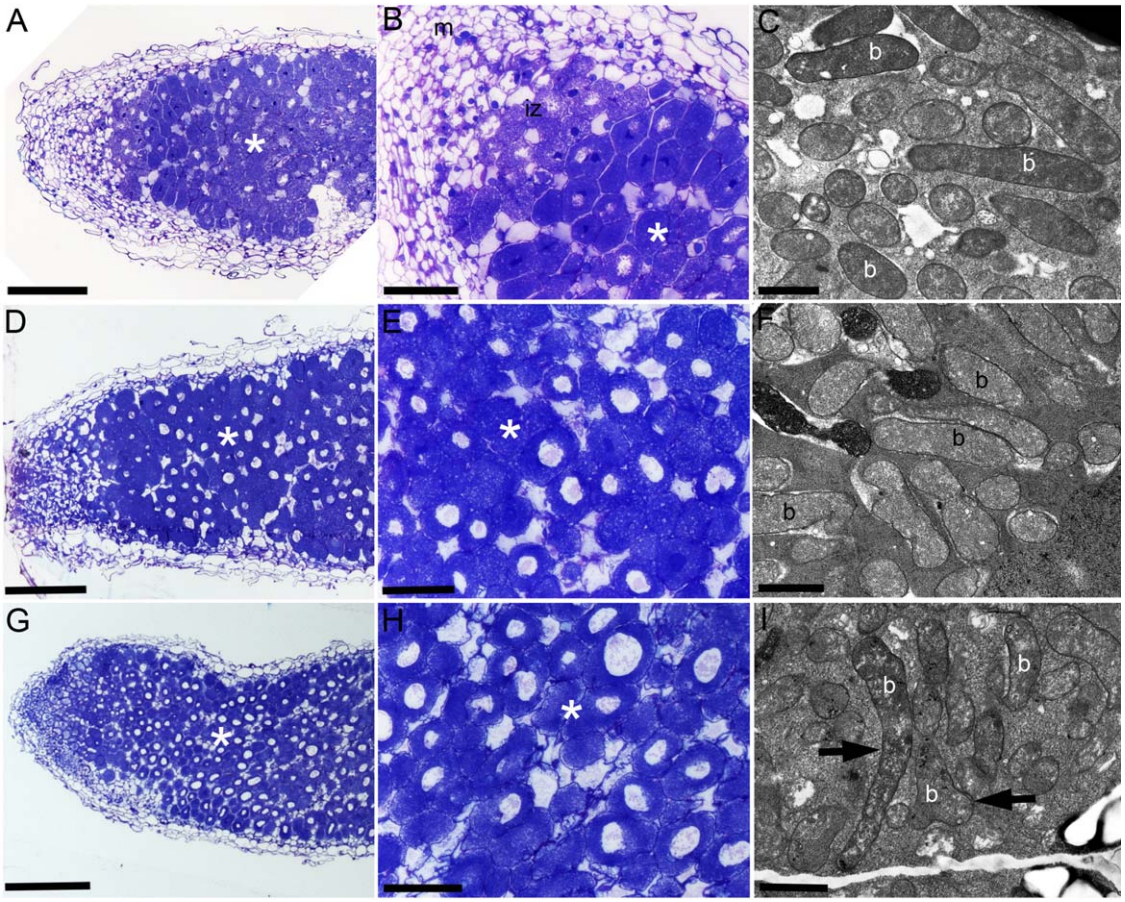


Fig. 2. A. Micrographs of sections of *Medicago murex* nodules containing *Ensifer medicae* WSM419 (A–C), MUR1169 (*lpiA* mutant, D–F) and MUR2124 (*acvB* mutant, G–I) using light (A, B, D, E, G and H) and transmission electron (C, F and I) microscopy. The tip of the nodules is to the left in panels (A, D and G), and the N-fixing zone is marked with asterisk in (A), (B), (D), (E), (G) and (H). Bacteroids are marked with b in (C), (F) and (I) and elongated and pleomorphic bacteroids are arrowed in (I). Bars = 200 μm (A, D and G), 50 μm (B, E and H) and 1 μm (C, F and I).

localized to the apical meristem in nodules of *M. murex* inoculated with MUR1169, after overnight staining with X-Glc (results not shown).

Gene expression in wild-type and isogenic mutant backgrounds

We mobilized the broad host range plasmid pCRS536 (containing an acid-inducible *lpiA-gusA* fusion) into wild-type and mutant backgrounds and compared the fold induction of GUS activity at pH 5.7, relative to pH 7.0 (Fig. 3A). As expected, GUS was induced (26.9-fold) at pH 5.7 in the wild-type background MUR1589. A mutation in *fsrR* (MUR1793) or *tcrA* (MUR2094) reduced the acid-induction of the *lpiA-gusA* fusion to 3.0- and 5.9-fold respectively. In contrast, GUS was not induced in *ebpA*, *rpoN* or *tcsA* mutant backgrounds. These results indicated that FsrR, TcsA, TcrA, RpoN and EbpA are all required for the complete acid-induction of *lpiA*.

In addition to verifying the acid-activated expression of *lpiA*, the expression of *fsrR*, *tcsA*, *tcrA*, *acvB* and *rpoN* was investigated using qRT-PCR. The expression ratio of each target gene at pH 5.7 and pH 7.0 was calculated using the constitutively expressed *actA* gene as a reference, revealing that while *fsrR*, *tcsA*, *tcrA* and *rpoN* were constitutively expressed, there was $>$18-fold acid-induction of both *lpiA* and *acvB* (Fig. 3B).

The transcription start site for lpiA *and* acvB *is located downstream of a RpoN promoter*

Transcription of *acvB* is acid-activated to a similar extent to that observed for *lpiA*, and therefore, we hypothesized that these two genes may constitute an RpoN-activated operon. Sequence analysis of RACE generated PCR amplicons showed that *lpiA* and *acvB* were both transcribed from the same transcription start site (TSS), located 14 bases downstream of an RpoN binding site and 206 bases upstream of the *lpiA* start codon (Fig. 3C).

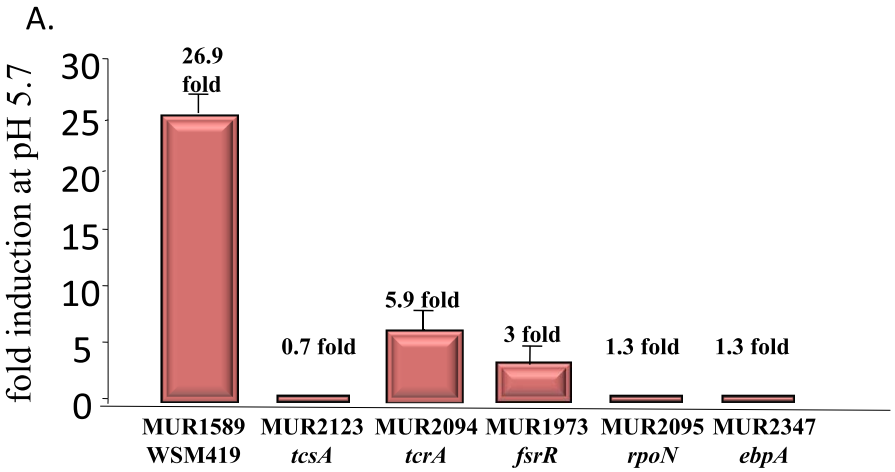
There are five lpiA *gene neighbourhood architectures in* Ensifer *spp*

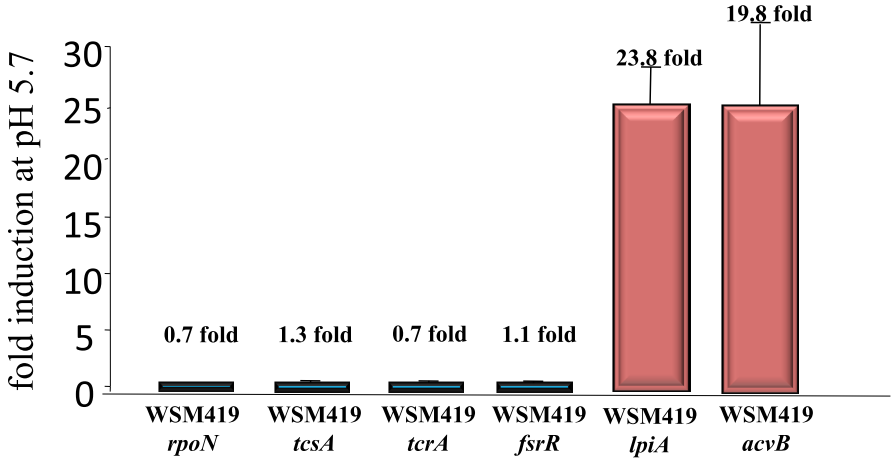
To determine whether the *lpiA* gene region was conserved in other *Ensifer* spp., we analyzed the *lpiA* gene neighbourhood across sequenced *Ensifer* genomes, using the Integrated Microbial Genome (IMG) analysis portal and BLASTP. In the finished WSM419 genome, the *lpiA* gene region is located on the symbiotic megaplasmid pSMED02. In the remaining draft *E. medicae* genomes, three strains (WSM244, WSM1115 and WSM4191) had *lpiA*-containing scaffolds that also harboured symbiotic loci. Although the *lpiA*-containing scaffold in the remaining strain (WSM1369) was too small to determine the presence of symbiotic loci, it is reasonable to assume that the *lpiA* gene region is borne on the symbiotic megaplasmid in *E. medicae* strains. In contrast, in the finished genomes of *Ensifer fredii* strains (HH103, NGR234 and USDA 257) and *E. meliloti* strains (1021, AK83, BL225C, GR4 and SM11), *lpiA* was located on the chromosome.

BLASTP analysis of the *lpiA* gene region identified five different *lpiA* gene neighborhood architectures (Class I–V) in 43 *Ensifer* strains (Table S4). By mapping these gene neighbourhood architectures to the *Ensifer lpiA* phylogenetic tree, we identified five different architectures that were associated with distinct *Ensifer* clades within the tree (Fig. 4). Class I–IV architectures all contained *lpiA*/*acvB* (Cluster 1) as well as the regulatory loci *fsrR*, *tcrA*, *tcsA* and *ebpA* (Cluster 2). Class I architecture found in four *E. medicae* strains (WSM244, WSM419, WSM1115 and WSM1369) contained *acvB*, *lpiA*, *fsrR*, *tcrA*, *tcsA*, Smed_5955 and *ebpA*. Class II architecture of *E. medicae* WSM4191 contained two additional open reading frames (ORFs), encoding a truncated cation/multi-drug transporter MdtC and a serine protease (labelled *degP/htrA* in Fig. 4), between the *fsrR* and *lpiA* genes. Class III architecture contained ORFs encoding a full length MdtC, a RND family efflux transporter MdtA and a serine protease (*degP*/*htrA*) between the *fsrR* and *lpiA* genes. This architecture was found in the broad host range strain *E. fredii* USDA 257 and in *E. meliloti* Mlalz-1, a microsymbiont of *Medicago laciniata* (L.) Mill. *Ensifer* sp. strains TW10 and PC2 contained Class IV architecture, in which the Cluster 1 *lpiA*/*acvB* genes are in a separate chromosomal location to Cluster 2 regulatory genes. Class V architecture, which contains Cluster 1 but completely lacks Cluster 2 regulatory genes, was found in all *E. meliloti* strains, *E. arboris* LMG 14919, other *E. fredii* strains, *Ensifer* sp. WSM1721 and the non-symbiont *E. adhaerens* Casida A. This gene neighborhood appears to be highly conserved. *Ensifer meliloti* strain Mlalz-1 was unusual in that it contained two *lpiA* paralogues: one (locus tag A3CADRAFT_05694) within Class III architecture and the other (locus tag A3CADRAFT_01190) within Class V architecture. The latter encoded a protein with highest identity (95.6–100%) with LpiA of *E. meliloti* strains, whereas the LpiA encoded by A3CADRAFT_05694 had highest protein identity with LpiA of *E. medicae* strains (96.8–97.0%).

The RpoN promoter resides downstream of fsrR *in strains containing regulatory cluster 2*

An RpoN promoter motif (5′-TGG$_{-26}$CAYA-CYGT-STGC$_{-14}$W-3′) with homology to an established





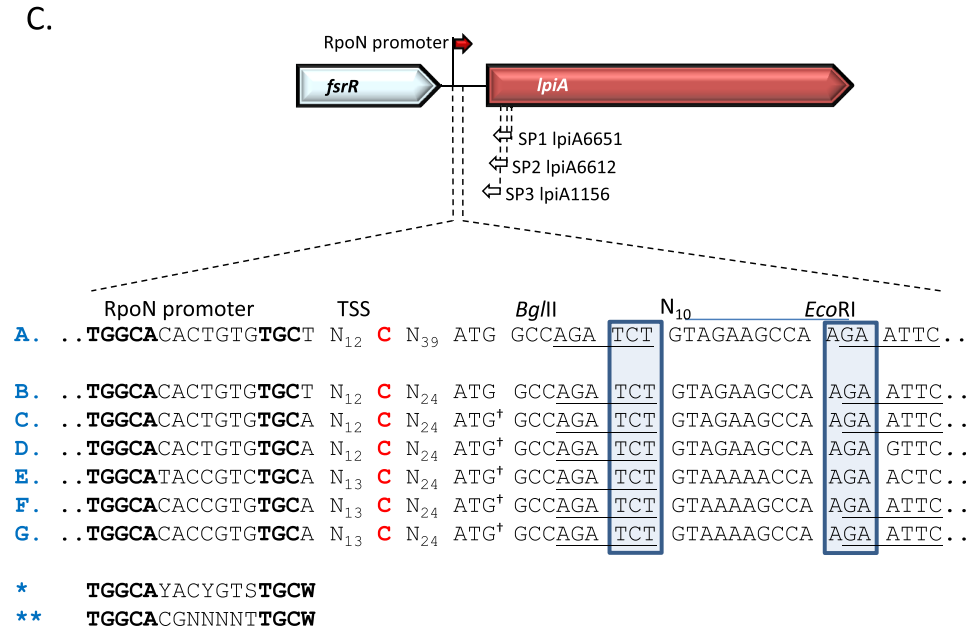


Fig. 3. Expression of genes in *Ensifer medicae* and *lpiA* transcript start site localization in *Ensifer* spp.
A. Fold induction of GUS in wild-type and mutant backgrounds containing a pCRS536 *lpiA-gusA* fusion. Data are the mean of five biological replicates ±SD. Data were analyzed using the unequal variance *t*-test (Ruxton, 2006). All values generated from mutants were significantly different from the wild-type ($P \leq 0.001$).
B. Fold induction of wild-type genes at pH 5.7 relative to pH 7.0 as determined by qRT-PCR. Data are the mean of three biological replicates ±SD. Data were analyzed using the unequal variance *t*-test (Ruxton, 2006); ***$P \leq 0.001$.
C. Promoter and transcript localization for the *lpiA*/*acvB* operon in *Ensifer medicae* WSM419 (A) and WSM1369 (B) and related sequences of *Ensifer* strains WSM4191 (C), Mlalz-1 (D), USDA257 (E), PC2 (F) and TW10 (G). The experimentally determined transcription start site (TSS) for *lpiA*/*acvB* in WSM419 is indicated in red. The bolded sequence represents the RpoN binding motif. The SP primers used to determine the TSS in WSM419 are shown. †ATG start codon for the *degP/htrA* gene. The asterisk indicates the derived RpoN promoter consensus for this region. The double asterisks indicate the bolded bases in the RpoN binding site representing the conserved bases in the Rhizobiales (Dombrecht *et al.*, 2002) consensus RpoN-binding sequence.

consensus 5′-TG**G**$_{-24}$CACG-N$_4$-TTG**C**$_{-12}$W-3′ was identified in the intergenic region downstream of *fsrR* in *Ensifer* strains that have Class I, II, III and IV gene architectures (Fig. 4). Strains devoid of regulatory Cluster 2 (i.e., those with Class V architecture) did not have the RpoN promoter consensus downstream of *fsrR*. Sequences that contained the RpoN consensus also featured a palindromic motif (TCT N_{10} AGA) 48 bases downstream of the TSS identified in WSM419.

Aligning the WSM419 intergenic sequence to sequences of *Ensifer* strains with Class I to IV architecture produced two alignments for each strain that overlapped at a 5 bp box (CTGAG; purple boxed text Fig. 5 between *fsrR* and *lpiA*). Class I strains had a single box in this region, while Class II strains had two CTGAG boxes that flanked *degP*/*htrA*-truncated *mdtC*. Class III D strains also had two boxes that flanked *degP*/*htrA*-*mdtA*-full length *mdtC*. Class III E and Class IV sequences contained a single CTGAG box upstream of *degP*/*htrA*. Another 5 bp box (TCATG; blue boxed text Fig. 5) was found in Class II, III and IV architectures but was absent from the region in Class I architecture). Notably, Class II architecture contained a *mdtA*-*mdtC* deletion between the TCATG boxes located upstream of *mdtA* and within *mdtC* in Class III D architecture. The observed rearrangements have positioned the RpoN promoter proximal to either *degP*/*htrA* or *lpiA*.

Discussion

Model of lpiA/acvB *regulation in* E. medicae

The *E. medicae* WSM419 *lpiA* gene is co-located with *acvB* in an operon. In free-living cells, *lpiA* is acid-induced and is required to enhance survival after lethal acid shock (Reeve *et al.*, 2006). Expression of *lpiA* occurs in the tips of indeterminate *Medicago* nodules prior to rhizobial differentiation into bacteroids. In other rhizobia, *lpiA* is upregulated in free-living cells of *E. meliloti* 1021 and *Rhizobium tropici* CIAT899 exposed to acid conditions (Vinuesa *et al.*, 2003; Hellweg *et al.*, 2009; de Lucena *et al.*, 2010) and is downregulated in bacteroids of *E. fredii* NGR234 and *E. meliloti* 2011 (Li *et al.*, 2013; Roux *et al.*, 2014). As *E. medicae* WSM419 *lpiA* and *acvB* knockout mutants are capable of forming nitrogen fixing symbioses with *Medicago* hosts, and *acvB* and *lpiA* knockout mutants are competitive for nodulation, it would appear that these genes are not essential for symbiotic nitrogen fixation in *Medicago* symbioses. However, in CIAT899, *lpiA* and *acvB* are required for nodulation competitiveness on the phaseoloid host *Phaseolus vulgaris*, in addition to being required for acid tolerance (Vinuesa *et al.*, 2003). It is interesting to note that the observed altered bacteroid morphology of the WSM419 *acvB* mutant resembles that of *R. tropici* CIAT899 *acvB* (*atvA*) mutants (Vinuesa *et al.*, 2003).

At low pH, a *lpiA*-deficient mutant of CIAT899 was more sensitive than the wild-type to cationic peptides (Sohlenkamp *et al.*, 2007), as has also been shown for a *mprF* (*lpiA* orthologue) mutant of the Gram-positive pathogen *Staphylococcus aureus* (Peschel *et al.*, 2001). Similarly, exposure of *E. meliloti* 1021 to cationic nodule-specific cysteine-rich peptides induces the expression of both *lpiA* (locus tag SMc00611) and *acvB* (annotated as locus tags SMc00612 and SMc00613) (Penterman *et al.*, 2014; Tiricz *et al.*, 2013). The characterized *R. tropici* LpiA and *S. aureus* MprF are required for the synthesis of lysyl-phosphatidylglycerol (LPG), which results in a more positive membrane surface charge, postulated to cause the repulsion of host defensins, such as cationic antimicrobial peptides (Peschel *et al.*, 2001; Sohlenkamp *et al.*, 2007). The *acvB* gene, which is cotranscribed with *lpiA*, encodes a protein containing the VirJ secretion domain (PF06057) postulated to be a host interaction determinant (Seshadri *et al.*, 2015). This protein contains four alpha/beta-hydrolase folds (IPRO29058) and hence could be an aminoacyl-phosphatidylglycerol hydrolase. In *Pseudomonas aeruginosa* (Arendt *et al.*, 2013) and *Enterococcus faecium* (Smith *et al.*, 2013), this enzyme fine tunes the concentration of aminoacylated phosphatidylglycerol. Similar to the phenotype obtained for the *E. medicae acvB* mutant, a mutation in the *P. aeruginosa acvB* ortholog PA0919 resulted in increased sensitivity to polymixin B under acid growth conditions (Arendt *et al.*, 2013). These findings suggest a role for *lpiA/acvB* in membrane modification in response to environmental stresses and symbiotic and pathogenic interactions.

By sequential knockouts of genes clustered in the WSM419 *lpiA* gene neighbourhood, coupled with expression studies, we have shown that the observed 20-fold acid-induction of *lpiA* requires a multi-phospho-relay signal transduction pathway (Fig. 6). In WSM419, the proteins in the pathway consist of a histidine kinase (HK) sensor component TcsA, the response regulatory (RR) components TcrA and FsrR and an enhancer-binding protein EbpA. TcsA is a membrane spanning protein, containing an N-terminus sensing domain (CHASE3) flanked by two predicted transmembrane helices (Krogh *et al.*, 2001), a phospho-acceptor domain (HisKA) and a ATP hydrolyzing domain (HTPase_C). These three domains are characteristic features of a HK protein (Mascher *et al.*, 2006). The CHASE3 domain is found in the 'periplasmic-sensing' HK proteins and is required for the detection of an extracellular signal (Mascher *et al.*, 2006). The N-terminus HisKA dimerization/phosphoacceptor domain contains the conserved

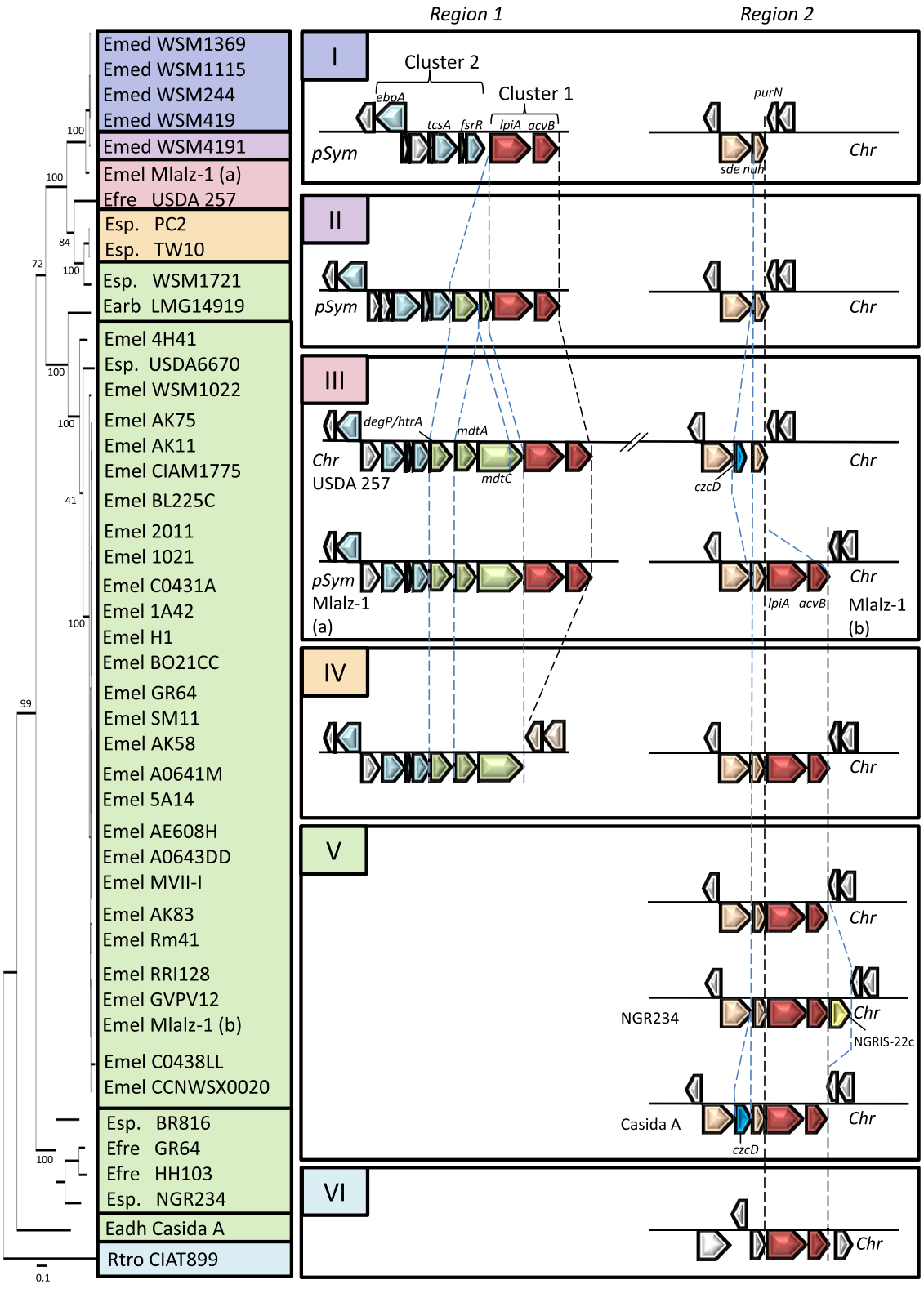

Region 1
Region 2
Emed WSM1369
Emed WSM1115
Emed WSM244
Emed WSM419
Emed WSM4191
Emel Mlalz-1 (a)
Efre USDA 257
Esp. PC2
Esp. TW10
Esp. WSM1721
Earb LMG14919
Emel 4H41
Esp. USDA6670
Emel WSM1022
Emel AK75
Emel AK11
Emel CIAM1775
Emel BL225C
Emel 2011
Emel 1021
Emel C0431A
Emel 1A42
Emel H1
Emel BO21CC
Emel GR64
Emel SM11
Emel AK58
Emel A0641M
Emel 5A14
Emel AE608H
Emel A0643DD
Emel MVII-I
Emel AK83
Emel Rm41
Emel RRI128
Emel GVPV12
Emel Mlalz-1 (b)
Emel C0438LL
Emel CCNWSX0020
Esp. BR816
Efre GR64
Efre HH103
Esp. NGR234
Eadh Casida A
Rtro CIAT899
100
100
84
72
100
100
100
41
100
99
100
0.1
I
Cluster 2
Cluster 1
ebpA
tcsA
fsrR
lpiA
acvB
pSym
purN
sde
nuh
Chr
II
pSym
Chr
III
degP/htrA
mdtA
mdtC
Chr
USDA 257
czcD
Chr
pSym
Mlalz-1
(a)
lpiA
acvB
Chr
Mlalz-1
(b)
IV
Chr
V
Chr
NGR234
Chr
NGRIS-22c
Casida A
Chr
czcD
VI
Chr

Fig. 4. A *lpiA* phylogenetic tree for 43 *Ensifer* strains and *Rhizobium tropici* CIAT899 (out group) with associated *lpiA* gene neighbourhoods. Classes I to VI are coloured using the colour scheme represented in the phylogenetic tree. Locations of Cluster 1 and Cluster 2 are indicated where known. Note that *E. meliloti* Mlalz-1 has two Cluster 1 paralogues designated as "(a)" or "(b)" in the phylogenetic tree and the Class III gene neighbourhood. In PC2 and TW10 (Class IV neighbourhood), the replicons containing Cluster 2 could not be identified but are located on scaffolds separate to those containing Cluster 1. Abbreviations are as follows: Chr, chromosome; Eadh, *Ensifer adhaerens*; Earb, *Ensifer arboris*; Efre, *Ensifer fredii*; Emed, *Ensifer medicae*; Emel, *Ensifer meliloti*; Esp., *Ensifer* species; pSym, symbiotic plasmid; Rtro, *Rhizobium tropici*. Gene symbols include *ebpA*, enhancer-binding protein gene; *tcsA*, two component sensor gene; *fsrR*, fuse sensor regulator gene; *lpiA*, low pH inducible gene; *acvB*, acid virulence induced gene; *sde*, selenocysteine-containing dehydrogenase gene; *nuh*, nudix hydrolase gene; *purN*, phosphoribosylglycinamide formyltransferase gene; *degP*/*htrA*, serine protease precursor gene; *mdtA*, multi-drug transporter gene; *mdtC*, multi-drug transporter gene; *czcD*, cobalt zinc cadmium resistance gene.

histidine residue (the TcsA H box = H_{259}) necessary for autophosphorylation-induced dimer formation (Laub and Goulian, 2007). TcsA autophosphorylation would enable phosphotransfer to occur with an appropriate regulator containing a conserved aspartate in a REC protein domain. Cross-talk of the regulatory proteins with another histidine kinase sensor (Laub and Goulian, 2007) would not be expected, as a mutation in *tcsA* completely abolishes acid induction of the *lpiA*/*acvB* operon. Since an HK and its preferred RR partner are generally proximally located (Laub *et al*., 2007), we suggest that phosphoryl group transfer would preferentially occur from TcsA to the REC domain-containing regulator TcrA, although a phosphoryl group could also be transferred from TcsA to the REC domains of the regulatory proteins FsrR and EbpA. Kinetic preference for phosphoryl group transfer would need to be determined using phosphotransfer profiling (Laub and Goulian, 2007). The possibility that TcrA and FsrR are both phosphorylated by TcsA is supported by the finding that the absence of either TcrA or FsrR caused a significant, but not total, loss of *lpiA* acid induction. The lack of a DNA binding domain in both of these RR proteins indicates that these proteins relay the phosphoryl group by protein–protein interaction. As FsrR contains a HK HisKA domain PF00512 (also present in TcsA), this suggests that FsrR could act as a histidine phosphotransfer (HPt) protein. In addition, this protein contains a HATPase domain, suggesting that FsrR could autophosphorylate in response to a cytoplasmic signal. Such intracellular monitoring, of pH for example, has been postulated to be essential for adaptation to an acidic environment (Dilworth *et al*., 2001). The transfer of a phosphoryl group from a HK histidine to a RR aspartate is conserved in other two-component regulatory systems (Stock *et al*., 2000). The predicted multi-phospho relay (His–Asp–His–Asp) described in this paper would, therefore, represent a branched 'One-to-Many' signal transduction pathway (Laub and Goulian, 2007) with conserved phosphoryl group transfer between His and Asp. Similarly, Falord *et al*. (2012) have identified a five-component signal transduction network regulating the expression of the *lpiA* ortholog *mprF* in the pathogen *S. aureus*.

In WSM419, the final part of the pathway would involve interaction of the cognate enhancer-binding protein (EbpA) with the alternative sigma factor RpoN (Dixon and Kahn, 2004) to enable transcription from the TSS upstream of *lpiA*. We have shown that a knockout mutation in *ebpA* completely abolishes *lpiA* induction in WSM419. EbpA contains the required RR REC domain, a sigma-54-activated domain and a HTH_8 DNA binding domain required for signal transfer, RpoN interaction and DNA binding, respectively. These domains are characteristic of Group 1 EBPs, which become activated after transfer of a phosphoryl group from a cognate HisKA-containing protein (Bush and Dixon, 2012). Hence, EbpA could become phosphorylated either via FsrR, or be activated by direct phosphorylation by TcsA, albeit weakly. Regardless, the activation of EbpA by phosphorylation would be necessary for the assembly of the polymeric active form of EBP required to initiate RpoN-dependent transcription (Popham *et al*., 1989; Morett and Segovia, 1993). EBPs are known to bind to palindromic sequences on the DNA, usually located upstream of an RpoN promoter (Jyot *et al*., 2002). In *Ensifer* strains that possess Cluster 2 regulatory loci (*fsrR*, *tcrA*, *tcsA* and *ebpA*), we have identified a palindromic sequence (TCT N_{10} AGA) downstream to the RpoN promoter located between *Bgl*II and *Eco*RI sites upstream of *lpiA*. The region between these two sites has previously been shown to be critical for the acid-activation of *lpiA* (Reeve *et al*., 2006).

Our analysis of sequenced *Ensifer* genomes has now provided some insights into how genome rearrangement has led to the evolution of this novel system of *lpiA/acvB* regulation. In *E. medicae*, the gene neighbourhood containing *lpiA/acv*B (Cluster 1) and the adjacent *fsrR*, *tcrA*, *tcsA* and *ebpA* (Cluster 2) regulatory loci is highly conserved and located on a symbiotic plasmid in all strains examined (Class I and II architecture types). *E. meliloti* Mlalz-1 (Class III), a microsymbiont of the highly specific host *M. laciniata* (van Berkum *et al*., 2007), seems to have acquired paralogues of *E. medicae* Cluster 1 and Cluster 2 loci via horizontal gene transfer. All other *Ensifer* spp. appear to harbour *lpiA/acvB* on the chromosome. Based on the sequence alignments, we hypothesize that the evolution of Class I architecture found in *E. medicae*

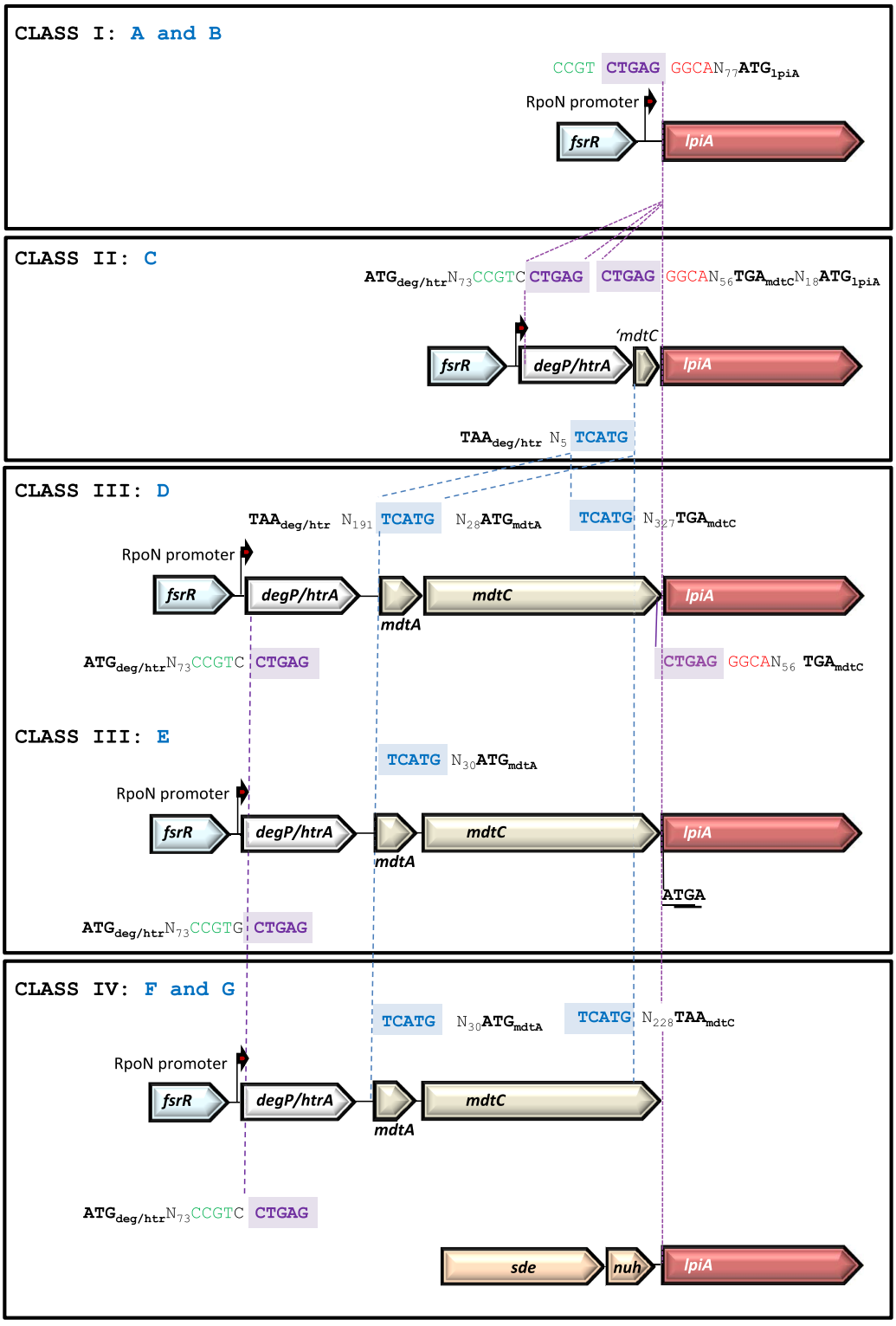


Fig. 5. Rearrangements of genes within the different *lpiA* gene neighborhood architecture types in *Ensifer* species. Based on sequence alignments, five base sequences that flank gene rearrangements were identified and are shown in purple or blue. The location of each RpoN binding sequence is indicated with a red arrow. *Ensifer medicae* WSM419 (A) and WSM1369 (B), *Ensifer medicae* WSM4191 (C), *Ensifer meliloti* Mlalz-1 (D), *Ensifer fredii* USDA257 (E), *Ensifer* sp. PC2 (F) and *Ensifer* sp. TW10 (G).

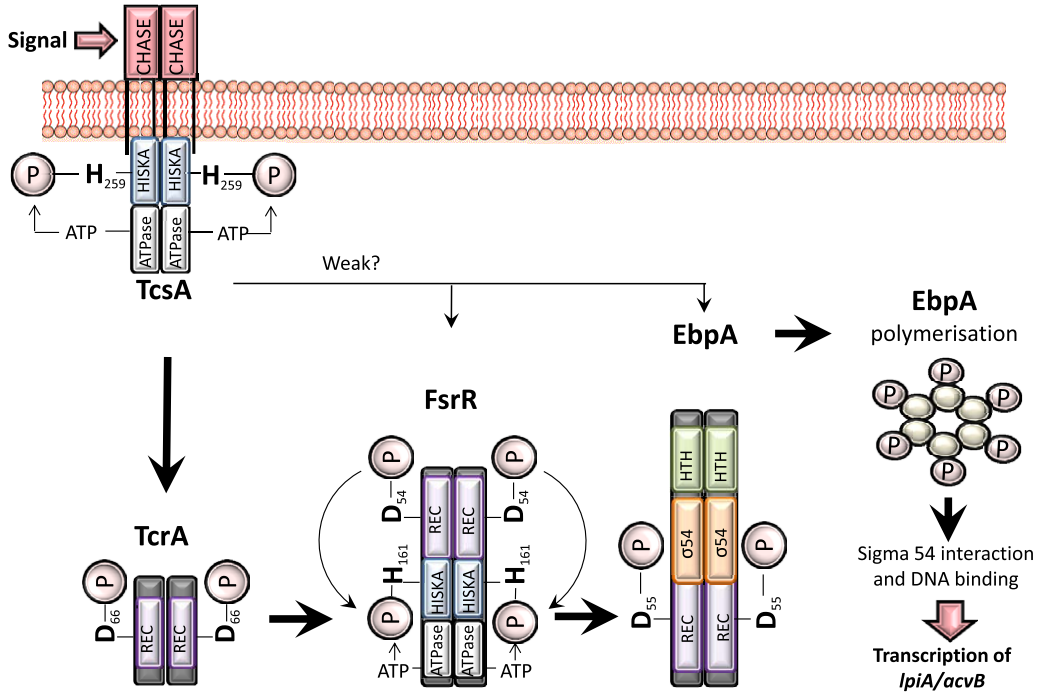


Fig. 6. Model of the regulation of *lpiA/acvB* gene expression in *Ensifer medicae* WSM419 via a multi phospho signal transduction cascade. Conserved aspartate and histidine amino acid involved in phospho transfer were identified from NCBI annotation data and protein sequence alignments.

would have occurred from a deletion between the CTGAG repeats found in Class III D or Class II variants. Notably, Class II architecture would have arisen from the deletion between TCATG box repeats of Class III D architecture. A number of insertion sequences, such as member of the Tn*3* family, are known to cause the duplication of five base pairs at the site of insertion (Siguier *et al*., 2014). These mobile genetic elements are important agents of evolution (Dobrindt *et al*., 2004; Frost *et al*., 2005; Abby and Daubin, 2007) and are speculated to assist in microbial adaptation to environmental niches. In *Ensifer* spp., the gene rearrangements that we have described here may be important in the evolution of enhanced acid tolerance of *E. medicae* strains, which predominantly interact with *Medicago* spp. adapted to acid soils (Garau *et al*., 2005).

Experimental Procedures

Bacterial strains, plasmids and media

Bacterial strains and plasmids used in this work are listed in Table S2. *Escherichia coli* strains were grown at 37°C in Lysogeny Broth (LB) (Miller, 1972) or Antibiotic Medium No. 3 (AM3) (Oxoid; Thermo Fisher, Adelaide, Australia) when using gentamicin. *Ensifer* strains were grown at 28°C in TY media (Beringer, 1974) supplemented with 6 mM $CaCl_2$, or minimal MJMM medium (Reeve *et al*., 2002) buffered with 20 mM HEPES (pH 7.0) or MES (pH 5.7), adding agar at 1.5% (w/v) where required. We supplemented media with antibiotics at the following concentrations (μg ml^{-1}): ampicillin (100), chloramphenicol (20), gentamicin (40; 10 for *E. coli*), kanamycin (50) and tetracycline (20; 12.5 for *E. coli*). Starter cultures of WSM419, MUR1169 and MUR2088 were grown for 3 days at 28°C in 5 ml TY with antibiotics. The cell pellets were collected, washed and re-suspended in saline (0.89% w/v NaCl). They were subsequently sub-cultured into 5 ml of TY at pH 7.0 and 5.7. The $OD_{600\ nm}$ of the cultures in TY was standardized to 0.01 for pH 7.0 and of 0.05 for pH 5.7. Both pH 7.0 and 5.7 TY cultures were then incubated for 24 h at 28°C. A sample of each TY culture was centrifuged and re-suspended in 30 ml of TY at pH 7.0 or 5.7 to provide an $OD_{600\ nm}$ of 0.03. Both 30 ml cultures of pH 7.0 and 5.7 for each strain were grown at 28°C with shaking at 200 rpm, and a 1 ml of aliquot was removed from each culture every 4 h for $OD_{600\ nm}$ measurement.

Mean generation time (MGT) at neutral and acidic pH and phenotypic characterization at low pH in the presence of polymyxin B

Starter cultures of WSM419 and its mutants were grown for 2 days at 28°C in 5 ml TY with antibiotics. The cell

pellets were collected, washed and re-suspended in saline (0.89%, w/v, NaCl). They were subsequently sub-cultured into 5 ml of buffered TY at pH 7.0 and pH 6.0 to an $OD_{600\ nm}$ of 0.01 and 0.02 for pH 7.0 and pH 6.0 grown cultures respectively. The cultures were then incubated overnight at 28°C, following which a sample of each culture was centrifuged and re-suspended in 50 ml of TY at pH 7.0 or pH 5.7 to provide an $OD_{600\ nm}$ of 0.025 and 0.05, respectively and then grown at 28°C with shaking at 200 rpm. Duplicate one ml aliquots were removed from each of two biological replicate cultures every 3 (pH 7.0) and 5 h (pH 5.7) for $OD_{600\ nm}$ measurements to determine the mean generation time for each strain.

For phenotypic characterization at low pH in the presence of polymyxin B, cultures were prepared as above but the $OD_{600\ nm}$ of the final resuspended cultures was standardized to 0.01 in TY broth supplemented with 3 $\mu g\ ml^{-1}$ polymyxin B (Sigma-Aldrich, Castle Hill, NSW, Australia) at pH 7.0 or pH 5.7. The cultures were grown at 28°C with shaking at 200 rpm for 48 h. The $OD_{600\ nm}$ of each culture was measured, and the number of generations (n) was calculated based on the formula: $n = \log^{(\text{Final OD reading/starting OD reading})}/\log 2$.

DNA manipulation and analysis

Plasmid or genomic DNA isolation, DNA manipulation, transformation and DNA sequencing and analysis were as reported by Reeve *et al.* (2006). We identified potential proteins using the BLASTP algorithm in IMG-ER (Markowitz *et al.*, 2009), NCBI (http://www.ncbi.nlm.nih.gov/BLAST/) or *E. meliloti* BLAST server (http://sequence.toulouse.inra.fr/rhime/public/Access/RhimeFormRA.html). We identified clusters of orthologous groups (COGs) (Tatusov *et al.*, 2000) and protein family (PF) (Finn *et al.*, 2014) domains from the Conserved Domain Database (Marchler-Bauer *et al.*, 2015). Transmembrane domains were identified using TMHMM2.0 (Krogh *et al.*, 2001).

Phylogenetic analysis of lpiA *homologs*

DNA sequences of 45 *lpiA* homologs were retrieved from the JGI IMG database (http://jgi.doe.gov/) from 44 strains. The database was searched using the BLASTN program (with default parameters) with the *lpiA* from *E. medicae* WSM419 being used as a query sequence. The phylogenetic analysis was performed using MEGA 6.0 software (Tamura *et al.*, 2013). The translated DNA sequences were aligned with MUSCLE and the phylogenetic tree was built using the Maximum Likelihood method based on the JTT (+G + I + F) model (Jones *et al.*, 1992) that produced the lowest Bayesian Information Criterion value in the Best-Fit Substitution Model. Boot strap analysis (Felsenstein, 1985) with 500 replicates was performed to assess the support of the clusters.

PCR amplification

We amplified DNA using *Taq* polymerase (Invitrogen; ThermoFisher, Adelaide, Australia) or *Pfu* DNA polymerase (Promega, Sydney, Australia) using a BIORAD™ iCycler and the following reactions and thermal cycling conditions. Reactions contained 50 ng of plasmid or 100 ng of genomic DNA and 0.5 μl each of a forward and a reverse primer (50 μM) (Tables S5 and S6). For standard PCR amplifications, we used 1.5 U of *Taq* DNA polymerase in a reaction that contained 5 μl 5× PCR polymerization buffer (Fisher-Biotech, Perth, Australia), 2.5 μl of 10× $MgCl_2$ stock and PCR grade H_2O (Fisher-Biotech) to make a final volume of 25 μl. We optimized the $MgCl_2$ concentration for each reaction (Tables S5 and S6) and used the following thermal cycling conditions: 1 cycle at 94°C for 4 min, 30 cycles of denaturation at 94°C for 15 s, annealing for 45 s (temperature range between 50 and 60°C) and extension at 72°C for 45 s (or 65°C for 6 min for a product size of 3 kb or more), with a final cycle at 72°C for 5 min. For proofreading amplifications, we used 1.5 U of *Pfu* polymerase, 5 μl of 10× *Pfu* polymerization buffer (Promega) and PCR grade H_2O (Fisher-Biotech) to make a final volume of 50 μl. Thermal cycling conditions were as follows: 1 cycle at 96°C for 1 min, 35 cycles of denaturation at 96°C for 45 s, annealing at 58°C for 45 s and extension at 68°C for 2 min with a final cycle at 68°C for 5 min.

Construction of single crossover mutations in tcsA, tcrA, acvB *and* rpoN

Intragenic fragment sizes of 300–600 bp within *tcsA* *tcrA*, *rpoN* or *acvB* were PCR amplified (using primers in Table S3) and cloned into the pGEM-T plasmid (Invitrogen). A *Spe*I digested CAS-1116 antibiotic cassette containing a promoterless *gusA* reporter gene and associated antibiotic resistance genes (*nptII* and *aacCI*) was ligated to the *Spe*I site of the multiple cloning site of each recombinant plasmid. The pGEM-T sequence was removed from recombinant plasmids containing the intragenic fragments of *tcsA* and *acvB* by *Xho*I restriction and religation. We transformed plasmids into BW20767 and then mobilized them into WSM419 to create single-crossover mutants (Reeve *et al.*, 1999). Transconjugants from each mating were selected on JMM plates containing chloramphenicol and kanamycin, patched for gentamicin sensitivity and screened by PCR to confirm single crossover mutations. Amplification reactions contained WIL3 primer (binding to the *gusA*

gene on CAS-1116) and a primer that was upstream to each intragenic sequence (Table S3).

Deletion of CAS-1116 from single crossover mutants

The plasmid pCM157 containing the *cre* gene was conjugated into MUR2088, MUR2090, MUR2121 and MUR2124 single crossover mutants using the helper plasmid pRK2013 (Reeve *et al.*, 2002) to excise the *loxP* flanked CAS-1116. We replica patched chloramphenicol and tetracycline resistant transconjugants for kanamycin and gentamicin sensitivity. PCR amplification was used to verify the excision of CAS-1116 in each mutant background. Each primer pair (Table S3) was designed such that one primer bound upstream and one downstream of the intragenic PCR fragment used to mutate each gene.

Construction of double crossover mutations in ebpA

Using the primers F_19253/R_21236 (Table S5), we amplified a 2,018 bp fragment containing the *ebpA* gene and cloned this into pGEM-T Easy. Recombinant plasmid was restriction digested with *Bgl*II (site within *ebpA*) and ligated with a 2,237 bp *Bam*HI fragment containing the Ω-Km interposon from pHP45 ΩKm (Prentki and Krisch, 1984). The *Not*I *ebpA* fragment containing the Ω-Km interposon was cloned into the *Not*I site of pJQ200KS (Quandt and Hynes, 1993) to create pSH200 ΩKm1 and pSH200 ΩKm2 and these constructs were each transformed into BW20767. We mobilized plasmids from this strain into WSM419 using a triparental conjugation (Reeve *et al.*, 2002) and forced double-crossover recombination by selecting for chloramphenicol, naladixic acid and kanamycin resistance on TY. Transconjugants were replica patched for sucrose tolerance and gentamicin sensitivity. Double crossover mutants were confirmed by PCR using the primers KO1fw/KO1rev (Table S5).

Mobilization of pCRS536 in E. medicae *mutants*

The broad host range plasmid pCRS536, which contains an acid-activated *lpiA-gusA* fusion was mobilized into MUR2093, MUR2092, MUR2122, MUR2125, MUR2347 and MUR2348 using a biparental mating (Reeve *et al.*, 2002) and transconjugants selected for chloramphenicol, kanamycin and tetracycline resistance transconjugants.

Nodulation and nitrogen fixation studies

Seeds of *Medicago* spp. were surface sterilized, germinated and sown as previously described (Reeve *et al.*, 1999). Immediately after planting, seedlings were inoculated with wild-type *E. medicae* WSM419 or *E. medicae* mutant strains. Plants were watered with sterile nutrient solution (Howieson *et al.*, 1995) and harvested 6 weeks after sowing. The nodule count and top dry weight for each isolate was recorded. Nodules were stained for GUS activity as previously described (Reeve *et al.*, 2006). Competition studies were performed as previously described (Reeve *et al.*, 2006) using a 1:1 river sand:yellow sand mix that had been washed with 1% (v/v) H_2SO_4 and rinsed six times with water buffered with 25 mM MES (pH 5.5) or 25 mM HEPES (pH 7.0). Plants were watered with buffered solutions.

Nodule microscopy

M. murex nodules infected with the wild-type *E. medicae* WSM419 or *E. medicae* mutant strains were fixed, embedded and sectioned according to Beck *et al.* (2008), except that the sections were viewed using a JEOL JEM1400 transmission electron microscope.

Gene expression using reporter fusions

We examined *lpiA–gusA* activity within nodules by using X-GlcA to stain root systems of 6-week old *M. murex* plants inoculated with MUR1169 (*lpiA* mutant).

For quantitative assays, we first inoculated cultures into 5 ml of MJMM broth (pH 7.0) and grew them at 28°C to an $OD_{600\ nm}$ of approximately 0.8 in the presence of appropriate antibiotics. After overnight incubation at 28°C, suspensions were centrifuged and resuspended in MJMM broth (at pH 7.0 or 5.7) to an $OD_{600\ nm}$ of 0.25–0.5. Specific GUS activity was determined as described previously (Reeve *et al.*, 1999; Reeve *et al.*, 2006). Five biological replicate assays per strain were used in statistical analysis. β-Glucuronidase-specific activity was expressed as nmol *p*-nitrophenol (pNP) produced min^{-1} $(OD_{595\ nm})^{-1}$ at 28°C.

RNA extraction

E. medicae WSM419 was cultured in MJMM at pH 7.0 and the cells collected by centrifugation, washed and resuspended in saline (0.89%, w/v, NaCl). Cells were then subcultured into 50 ml of MJMM medium at pH 7.0 and 5.7 at an initial $OD_{600\ nm}$ of 0.05 and 0.25, respectively, and subsequently incubated for 16 h at 28°C. The cells (10 ml) were centrifuged at 8,000 *g* for 5 min at 4°C and the pellets washed with DEPC-treated saline (0.89%, w/v, NaCl). The cells were then resuspended in 500 μl of EDTA (5 mM), to which we

added 0.1 ml lysozyme (2 mg ml^{-1}), 0.1 ml 10% SDS and 0.1 ml proteinase K (1.2 mg ml^{-1}). The suspensions were incubated at 65°C for 10 min after which 0.4 ml of chilled 5 M NaCl was added, and the mixture centrifuged (1,600 *g*) for 20 min at 4°C. Nucleic acid was isolated from the supernatant by adding one volume of isopropanol, inverting to mix and then centrifuging (1,600 *g*) for 15 min at 4°C. The resulting pellet was resuspended in 132 μl of RNase-free PCR grade water (Fisher-Biotech). Following the addition of DNase buffer (15 μl) and DNaseI (3 μl of 2 U μl^{-1}) (Fisher-Biotech), the reaction was incubated at 37°C for 30 min. RNA was purified using standard phenol/chloroform/isoamylalcohol (25:24:1) extraction and ethanol precipitation procedures. RNA pellets were washed with 1 ml of 70% cold ethanol, air-dried and dissolved in 100 μl of DEPC-treated water.

Quantitative gene expression using qRT-PCR

For each treatment, we synthesized cDNA using Superscript III first strand synthesis supermix (Invitrogen) according to the manufacturer's protocol. cDNA was used as the template for qRT-PCR using the following reaction conditions in a Rotor-Gene RG300: 50°C for 2 min, 95°C for 2 min, followed by 40 cycles of 95°C for 15 s and 60°C for 30 s. The sample spectrum was acquired using a 470 nm excitation filter and a 510 nm detection with a gain of +5. Rotor Gene 6 (Corbett Research, Australia) software was used to analyse the data. All reactions were performed in triplicate independently.

The Pfaffl mathematical model was used to determine the expression of a target gene relative to a reference gene (Pfaffl, 2001). We calculated the relative expression ratio of each target gene at low pH (pH 5.7) and neutral pH (pH 7.0) from the PCR efficiencies and the crossing point deviation of the low pH sample (pH 5.7) versus the neutral pH sample (pH 7.0). The mathematical model of relative expression ratios was calculated using the following equation: $\text{ratio} = (E_{\text{target}})^{\Delta \text{CPtarget (pH 7.0–5.7)}} / (E_{\text{reference}})^{\Delta \text{CPreference (pH 7.0–5.7)}}$.

Transcription start site identification using 5′-RACE

The TSS for *lpiA* and *acvB* was identified using a 5′/3′ RACE kit (Roche, Australia) as detailed by the manufacturer. The specific primers (SP1, SP2 and SP3) used to target the *lpiA* or *acvB* TSS are presented in Table S6. Total RNA extracted from *E. medicae* WSM419 cultured at pH 5.7 was used as a template for cDNA synthesis. Amplification of RACE-generated cDNA using SP1 (either lpiA 6651 or acvB 4552R) was confirmed using nested PCR primer sets (Table S6). We did not obtain PCR amplicons from a reaction containing RACE-generated cDNA template and the primers tcsA3685F/tcrA4421R, confirming the absence of contaminating DNA. The resulting RACE cDNA PCR generated product was cloned into pGEM-T to generate pGEM-lpiA and pGEM-acvB, which were then sequenced using M13 universal reverse primer (Table S6). The *lpiA* transcription start site was determined by aligning the sequencing reads with the *E. medicae* WSM419 genome sequence.

Acknowledgements

The authors would like to thank Regina Carr (The Centre for Rhizobium Studies, Murdoch University) for help with the glasshouse component to establish symbiotic phenotypes and Professor Janet Sprent for critically reading the manuscript.

Author contributions

Tian, Osman, Heiden, Tiwari and Reeve designed and conducted the experiments. James performed the root nodule sectioning, staining and microscopy. Osman, Ardley and Reeve performed the bioinformatics analysis. Gollagher, Kyrpides, Seshadri and Reeve were involved in the sequencing and analysis of the rhizobial genomes. Ardley, Osman, Reeve and Tian wrote the article. All authors read and edited the article and approved the final version for publication.

Conflict of Interest

There are no competing interests.

Funding

We gratefully acknowledge the funding received from Curtin University Sustainability Policy Institute and from the Murdoch University Small Research Grants Scheme in 2016.

Availability of data and materials

The genome of *Ensifer medicae* WSM419 has been deposited in the GenBank database with the accession numbers CP000738.1 (chromosome), CP000739.1 (pSMED01), CP000740.1 (pSMED02) and CP000741.1 (pSMED03). The original *lpiA* gene neighbourhood sequence was deposited in the GenBank database with the accession number AF199025.3.

References

Abby, S., and Daubin, V. (2007) Comparative genomics and the evolution of prokaryotes. *Trends Microbiol* **15**: 135–141.

Arendt, W., Groenewold, M.K., Hebecker, S., Dickschat, J.S., and Moser, J. (2013) Identification and characterization of a periplasmic aminoacyl-phosphatidylglycerol hydrolase responsible for *Pseudomonas aeruginosa* lipid homeostasis. *J Biol Chem* **288**: 24717–24730.

Barrios, H., Valderrama, B., and Morett, E. (1999) Compilation and analysis of sigma(54)-dependent promoter sequences. *Nucleic Acids Res* **27**: 4305–4313.

Beck, S.L.M.V., Woodall, K., Doerrler, W.T., James, E.K., and Ferguson, G.P. (2008) The *Sinorhizobium meliloti* MsbA2 protein is essential for the legume symbiosis. *Microbiology* **154**: 1258–1270.

Béna, G., Lyet, A., Huguet, T., and Olivieri, I. (2005) *Medicago-Sinorhizobium* symbiotic specificity evolution and the geographic expansion of *Medicago*. *J Evol Biol* **18**: 1547–1558.

Beringer, J.E. (1974) R factor transfer in *Rhizobium leguminosarum*. *J Gen Microbiol* **84**: 188–198.

Bush, M., and Dixon, R. (2012) The role of bacterial enhancer binding proteins as specialized activators of σ^{54}-dependent transcription. *Microbiol Mol Biol R* **76**: 497–529.

de Lucena, D.K.C., Pühler, A., and Weidner, S. (2010) The role of sigma factor RpoH1 in the pH stress response of *Sinorhizobium meliloti*. *BMC Microbiol* **10**: 265.

Dilworth, M.J., Howieson, J.G., Reeve, W.G., Tiwari, R.P., and Glenn, A.R. (2001) Acid tolerance in legume root nodule bacteria and selecting for it. *Aust J Exp Agric* **41**: 435–446.

Dixon, R., and Kahn, D. (2004) Genetic regulation of biological nitrogen fixation. *Nat Rev Microbiol* **2**: 621–631.

Dobrindt, U., Hochhut, B., Hentschel, U., and Hacker, J. (2004) Genomic islands in pathogenic and environmental microorganisms. *Nat Rev Microbiol* **2**: 414–424.

Dombrecht, B., Marchal, K., Vanderleyden, J., and Michiels, J. (2002) Prediction and overview of the RpoN-regulon in closely related species of the Rhizobiales. *Genome Biol* **3**: research0076–research0071.

Falord, M., Karimova, G., Hiron, A., and Msadek, T. (2012) GraXSR proteins interact with the VraFG ABC transporter to form a five-component system required for cationic antimicrobial peptide sensing and resistance in *Staphylococcus aureus*. *Antimicrob. Agents Chemother* **56**: 1047–1058.

Felsenstein, J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**: 783–791.

Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., *et al*. (2014) Pfam: the protein families database. *Nucleic Acids Res* **42**: D222–D230.

Frost, L.S., Leplae, R., Summers, A.O., and Toussaint, A. (2005) Mobile genetic elements: the agents of open source evolution. *Nat Rev Microbiol* **3**: 722–732.

Garau, G., Reeve, W.G., Bräu, L., Yates, R.J., James, D., Tiwari, R., *et al*. (2005) The symbiotic requirements of different *Medicago* spp. suggest the evolution of *Sinorhizobium meliloti* and *S. medicae* with hosts differentially adapted to soil pH. *Plant Soil* **276**: 263–277.

Hellweg, C., Pühler, A., and Weidner, S. (2009) The time course of the transcriptomic response of *Sinorhizobium meliloti* 1021 following a shift to acidic pH. *BMC Microbiol* **9**: 37.

Howieson, J.G., and Ewing, M.A. (1986) Acid tolerance in the *Rhizobium meliloti-Medicago* symbiosis. *Aust J Agric Res* **37**: 55–64.

Howieson, J.G., Ewing, M.A., and D'antuono, M.F. (1988) Selection for acid tolerance in *Rhizobium meliloti*. *Plant Soil* **105**: 179–188.

Howieson, J.G., Loi, A., and Carr, S.J. (1995) *Biserrula pelecinus* L. – a legume pasture species with potential for acid, duplex soils which is nodulated by unique root-nodule bacteria. *Aust J Agric Res* **46**: 997–1009.

Jones, D.T., Taylor, W.R., and Thornton, J.M. (1992) The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* **8**: 275–282.

Jyot, J., Dasgupta, N., and Ramphal, R. (2002) FleQ, the major flagellar gene regulator in *Pseudomonas aeruginosa*, binds to enhancer sites located either upstream or atypically downstream of the RpoN binding site. *J Bacteriol* **184**: 5251–5260.

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305**: 567–580.

Laub, M.T., Biondi, E.G., and Skerker, J.M. (2007) Phosphotransfer profiling: systematic mapping of two-component signal transduction pathways and phosphorelays. *Methods Enzymol* **423**: 531–548.

Laub, M.T., and Goulian, M. (2007) Specificity in two-component signal transduction pathways. *Annu Rev Genet* **41**: 121–145.

Li, Y., Tian, C.F., Chen, W.F., Wang, L., Sui, X.H., and Chen, W.X. (2013) High-resolution transcriptomic analyses of *Sinorhizobium* sp. NGR234 bacteroids in determinate nodules of *Vigna unguiculata* and indeterminate nodules of *Leucaena leucocephala*. *PLOS One* **8**: e70531.

Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer, L.Y., *et al*. (2015) CDD: NCBI's conserved domain database. *Nucleic Acids Res* **43**: D222–D226.

Markowitz, V.M., Mavromatis, K., Ivanova, N.N., Chen, I.M., Chu, K., and Kyrpides, N.C. (2009) IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics* **25**: 2271–2278.

Mascher, T., Helmann, J.D., and Unden, G. (2006) Stimulus perception in bacterial signal-transducing histidine kinases. *Microbiol Mol Biol R* **70**: 910–938.

Miller, J.H. (1972) *Experiments in Molecular Genetics Cold Spring Harbor Laboratory*. New York: Cold Spring Harbor.

Morett, E., and Segovia, L. (1993) The sigma 54 bacterial enhancer-binding protein family: mechanism of action and phylogenetic relationship of their functional domains. *J Bacteriol* **175**: 6067–6074.

Munns, D.N. (1968) Nodulation of *Medicago sativa* in solution culture. 1. Acid-sensitive steps. *Plant Soil* **28**: 129–146.

O'Hara, G.W., Goss, T.J., Dilworth, M.J., and Glenn, A.R. (1989) Maintenance of intracellular pH and acid tolerance in *Rhizobium meliloti*. *Appl Environ Microbiol* **55**: 1870–1876.

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

Peschel, A., Jack, R.W., Otto, M., Collins, L.V., Staubitz, P., Nicholson, G., *et al*. (2001) *Staphylococcus aureus* resistance to human defensins and evasion of neutrophil killing via the novel virulence factor MprF is based on modification of membrane lipids with l-lysine. *J Exp Med* **193**: 1067–1076.

Pfaffl, M.W. (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* **29**: e45.

Popham, D.L., Szeto, D., Keener, J., and Kustu, S. (1989) Function of a bacterial activator protein that binds to transcriptional enhancers. *Science* **243**: 629–635.

Prentki, P., and Krisch, H.M. (1984) In vitro insertional mutagenesis with a selectable DNA fragment. *Gene* **29**: 303–313.

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

Reeve, W., Ardley, J., Tian, R., Eshragi, L., Yoon, J.W., Ngamwisetkun, P., *et al*. (2015) A genomic encyclopedia of the root nodule bacteria: assessing genetic diversity through a systematic biogeographic survey. *Stand Genom Sci* **10**: 14.

Reeve, W., Chain, P., O'hara, G., Ardley, J., Nandesena, K., Bräu, L., *et al*. (2010) Complete genome sequence of the *Medicago* microsymbiont *Ensifer* (*Sinorhizobium*) *medicae* strain WSM419. *Stand Genom Sci* **2**: 77–86.

Reeve, W.G., Bräu, L., Castelli, J., Garau, G., Sohlenkamp, C., Geiger, O., *et al*. (2006) The *Sinorhizobium medicae* WSM419 *lpiA* gene is transcriptionally activated by FsrR and required to enhance survival in lethal acid conditions. *Microbiology* **152**: 3049–3059.

Reeve, W.G., Tiwari, R.P., Kale, N.B., Dilworth, M.J., and Glenn, A.R. (2002) ActP controls copper homeostasis in *Rhizobium leguminosarum* bv. viciae and *Sinorhizobium meliloti* preventing low pH-induced copper toxicity. *Mol Microbiol* **43**: 981–991.

Reeve, W.G., Tiwari, R.P., Wong, C.M., Dilworth, M.J., and Glenn, A.R. (1998) The transcriptional regulator gene *phrR* in *Sinorhizobium meliloti* WSM419 is regulated by low pH and other stresses. *Microbiology* **144**: 3335–3342.

Reeve, W.G., Tiwari, R.P., Worsley, P.S., Dilworth, M.J., Glenn, A.R., and Howieson, J.G. (1999) Constructs for insertional mutagenesis, transcriptional signal localization and gene regulation studies in root nodule and other bacteria. *Microbiology* **145**: 1307–1316.

Rome, S., Fernandez, M.P., Brunel, B., Normand, P., and Cleyet-Marel, J.C. (1996) *Sinorhizobium medicae* sp. nov., isolated from annual *Medicago* spp. *Int J Syst Bacteriol* **46**: 972–980.

Roux, B., Rodde, N., Jardinaud, M.F., Timmers, T., Sauviac, L., Cottret, L., *et al*. (2014) An integrated analysis of plant and bacterial gene expression in symbiotic root nodules using laser-capture microdissection coupled to RNA sequencing. *The Plant J* **77**: 817–837.

Ruxton, G.D. (2006) The unequal variance *t*-test is an underused alternative to Student's *t*-test and the Mann–Whitney *U* test. *Behav Ecol* **17**: 688–690.

Seshadri, R., Reeve, W.G., Ardley, J.K., Tennessen, K., Woyke, T., Kyrpides, N.C., and Ivanova, N.N. (2015) Discovery of novel plant interaction determinants from the genomes of 163 Root Nodule Bacteria. *Sci Rep* **5**: 16825.

Siguier, P., Gourbeyre, E., and Chandler, M. (2014) Bacterial insertion sequences: their genomic impact and diversity. *FEMS Microbiol Rev* **38**: 865–891.

Smith, A.M., Harrison, J.S., Sprague, K.M., and Roy, H. (2013) A conserved hydrolase responsible for the cleavage of aminoacylphosphatidylglycerol in the membrane of *Enterococcus faecium*. *J Biol Chem* **288**: 22768–22776.

Sohlenkamp, C., Galindo-Lagunas, K.A., Guan, Z.Q., Vinuesa, P., Robinson, S., Thomas-Oates, J., *et al*. (2007) The lipid lysyl-phosphatidylglycerol is present in membranes of *Rhizobium tropici* CIAT899 and confers increased resistance to polymyxin B under acidic growth conditions. *Mol Plant Microbe Interact* **20**: 1421–1430.

Stock, A.M., Robinson, V.L., and Goudreau, P.N. (2000) Two-component signal transduction. *Annu Rev Biochem* **69**: 183–215.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013) MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* **30**: 2725–2729.

Tatusov, R.L., Galperin, M.Y., Natale, D.A., and Koonin, E.V. (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res* **28**: 33–36.

Tiricz, H., Szűcs, A., Farkas, A., Pap, B., Lima, R.M., Maróti, G., *et al*. (2013) Antimicrobial nodule-specific cysteine-rich peptides induce membrane depolarization-associated changes in the transcriptome of *Sinorhizobium meliloti*. *Appl Environ Microbiol* **79**: 6737–6746.

Tiwari, R.P., Reeve, W.G., Dilworth, M.J., and Glenn, A.R. (1996) Acid tolerance in *Rhizobium meliloti* strain WSM419 involves a two-component sensor-regulator system. *Microbiology* **142**: 1693–1704.

van Berkum, P., Badri, Y., Elia, P., Aouani, M.E., and Eardly, B.D. (2007) Chromosomal and symbiotic relationships of rhizobia nodulating *Medicago truncatula* and *M. laciniata*. *Appl Environ Microbiol* **73**: 7597–7604.

Vinuesa, P., Neumann-Silkow, F., Pacios-Bras, C., Spaink, H.P., Martínez-Romero, E., and Werner, D. (2003) Genetic analysis of a pH-regulated operon from *Rhizobium tropici* CIAT899 involved in acid tolerance and nodulation competitiveness. *Mol Plant Microbe Interact* **16**: 159–168.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web-site.