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

Permalink
https://escholarship.org/uc/item/3h89t27c

Journal
Molecular Microbiology, 103(5)

ISSN
0950-382X

Authors
Tian, R
Heiden, S
Osman, WAM
et al

Publication Date
2017-03-01

DOI
10.1111/mmi.13592

Peer reviewed
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, Stephan Heiden, Wan A. M. Osman, Julie K. Ardley, Euan K. James, Margaret M. Gollagher, Ravi Tiwari, Rekha Seshadri, Nikos C. Kyrpides and Wayne G. Reeve

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 melloti* 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.

Introduction

*Medicago* spp. are important pasture legumes that are nodulated by strains of rhizobia belonging to either *Ensifer (Sinorhizobium) melloti* 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 regulator) (Reeve et al., 1998), *actR/S* (acid tolerance regulator/sensor) (Tiwari et al., 1996) and *fsr* (fused sensor regulator) (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 (low pH-inducible) 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 (two component sensor and two component regulator), 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’TGG−24CACG−N7TTGC−12W3 (Barrios et al., 1999; Dombrecht et al., 2002). Hence, RpoN could be important for the expression of both lpiA and the downstream acvB (acid virulence 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-buffed 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
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 pSMED02*.
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

| 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. |
|---|---|---|
| Strains | Polymyxin B (µg ml⁻¹) |
| | 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 ≤ 0.002.
**P ≤ 0.001.
***P ≤ 0.05.

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).

© 2016 John Wiley & Sons Ltd, Molecular Microbiology, 103, 829–844
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 fold respectively. In contrast, GUS was not induced in acid-induction of the *lpiA-gusA* fusion (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, Sm_b_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, *Ensiler* sp. WSM1721 and the non-symbiont *E. adhaerens* Casida A. This gene neighborhood appears to be highly conserved. *Ensiler melliloti* 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 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.
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 ≤ 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 ≤ 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), MilaZ-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.

A. . . TGGCA CACTGT GTGC T
   B. . . TGGCA CACTGT GTGC T
   C. . . TGGCA CACTGT GTGC T
   D. . . TGGCA CACTGT GTGC T
   E. . . TGGCA CACTGT GTGC T
   F. . . TGGCA CACTGT GTGC T
   G. . . TGGCA CACTGT GTGC T

* TGGCA YACYGT GTGC W
** TGGCA CGNNTT GTGC W

© 2016 John Wiley & Sons Ltd, Molecular Microbiology, 103, 829–844
consensus 5’-TGG<sub>24</sub>CACG-N<sub>14</sub>-TTGC<sub>12</sub>W-3’ was identified in the intergenic region downstream of fsrR in <i>Ensifer</i> 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<sub>10</sub> AGA) 48 bases downstream of the TSS identified in WSM419.

Aligning the WSM419 intergenic sequence to sequences of <i>Ensifer</i> 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 <i>E. medicae</i> 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 <i>Medicago</i> nodules prior to rhizobial differentiation into bacteroids. In other rhizobia, lpiA is upregulated in free-living cells of <i>E. medicae</i> 1021 and <i>Rhizobium tropici</i> CIAT899 exposed to acid conditions (Vinuesa et al., 2003; Hellweg et al., 2009; de Lucena et al., 2010) and is downregulated in bacteroids of <i>E. fredii</i> NGR234 and <i>E. meliloti</i> 2011 (Li et al., 2013; Roux et al., 2014). As <i>E. medicae</i> WSM419 lpiA and acvB knockout mutants are capable of forming nitrogen fixing symbioses with <i>Medicago</i> 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 <i>Medicago</i> symbioses. However, in CIAT899, lpiA and acvB are required for nodulation competitiveness on the phaseoloid host <i>Phaseolus vulgaris</i>, 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 <i>R. tropici</i> 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 <i>Staphylococcus aureus</i> (Peschel et al., 2001). Similarly, exposure of <i>E. meliloti</i> 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 <i>R. tropici</i> LpiA and <i>S. aureus</i> MprF are required for the synthesis of lysyl-phosphatidylglycerol (PG), 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 <i>Pseudomonas aeruginosa</i> (Arendt et al., 2013) and <i>Enterococcus faecalis</i> (Smith et al., 2013), this enzyme fine tunes the concentration of aminoacylated phosphatidylglycerol. Similar to the phenotype obtained for the <i>E. medicae</i> acvB mutant, a mutation in the <i>P. aeruginosa</i> 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-phosphorelay 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
<table>
<thead>
<tr>
<th>Region 1</th>
<th>Region 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cluster 2</strong></td>
<td><strong>Cluster 1</strong></td>
</tr>
<tr>
<td>pSym</td>
<td>purN</td>
</tr>
<tr>
<td>Chr</td>
<td>Chr</td>
</tr>
</tbody>
</table>

### Region 2

- **Cluster 1**
  - pSym
  - Chr

### Region 3

- **Cluster 2**
  - degh/hnuA
  - mbA
  - Chr

### Region 4

- **Cluster 1**
  - pSym
  - Chr

### Region 5

- **Cluster 1**
  - NGR234
  - Chr

- **Cluster 2**
  - Casida A
  - Chr

### Region 6

- Eadh Casida A
- Rtro CIAT899

---

© 2016 John Wiley & Sons Ltd, *Molecular Microbiology*, 103, 829–844
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 Mialz-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 melloti; 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; nuf, nudix hydrolase gene; pusN, phosphoribosylglycinamide formyltransferase gene; degP/hrA, serine protease precursor gene; mdtA, multi-drug transporter gene; mdtC, multi-drug transporter gene; cccD, cobalt zinc cadmium resistance gene.

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 EBPs 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 N10 AGA) downstream to the RpoN promoter located between BglII and EcoRI 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/acvB (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. melloti Mialz-1 (Class III), a microsymbiont of the highly specific host M. lacticola (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).
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 Tn3 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 CaCl₂, 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 subcultured into 5 ml of TY at pH 7.0 and 5.7. The OD₆₀₀ 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 OD₆₀₀ 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 OD₆₀₀ nm measurement.

![Fig. 6. Model of the regulation of lpiA/acvB gene expression in Ensiler 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.](image-url)
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\textsubscript{600} 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\textsubscript{600} of 0.025 and 0.05, respectively and then grown at 28°C with shaking at 200 rpm. Duplicate 1 ml aliquots were removed from each of two biological replicate cultures every 3 (pH 7.0) and 5 h (pH 5.7) for OD\textsubscript{600} 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\textsubscript{600} of the final resuspended cultures was standardized to 0.01 in TY broth supplemented with 3 μg ml\textsuperscript{-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\textsubscript{600} of each culture was measured, and the number of generations (n) was calculated based on the formula: 
\[ n = \frac{\log(\text{Final OD reading}) - \log(\text{starting OD reading})}{\log(2)} \]

**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. melloti* 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 BIOTRAD™ 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\textsubscript{2} stock and PCR grade H\textsubscript{2}O (Fisher-Biotech) to make a final volume of 25 μl. We optimized the MgCl\textsubscript{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\textsubscript{2}O (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* digested CAS-1116 antibiotic cassette containing a promoterless *gusA* reporter gene and associated antibiotic resistance genes (*nptII* and *aacCI*) was ligated to the *SpeI* 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 *XhoI* 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 BglII (site within ebpA) and ligated with a 2,237 bp BamHI fragment containing the Ó-Km interposon from pH45 ÓKm (Prentki and Krisch, 1984). The Nofl ebpA fragment containing the Ó-Km interposon was cloned into the Nofl site of pJQ200KS (Quandt and Hynes, 1993) to create pSH200 ÓKm1 and pSH200 ÓKm2 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. Transconjugants were replica patched for sucrose tolerance and 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 OD600 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 (OD₅₉₅ nm)⁻¹ at 28°C.

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 total 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) H₂SO₄ 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 OD₆₀₀ 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 OD₆₀₀ 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 (OD₅₉₅ nm)⁻¹ at 28°C.

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.

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 OD₆₀₀ 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 DNasel (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: $\text{ratio} = \left(\frac{E_{\text{target}}^{\Delta C P_{\text{target}}} (pH 7.0-6.7)}{E_{\text{reference}}^{\Delta C P_{\text{reference}}} (pH 7.0-6.7)}\right)^{-1}$.

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 Ensifer 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


Supporting information

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