Cryptic plasmid pSKU146 from the wall-less plant pathogen *Spiroplasma kunkelii* encodes an adhesin and components of a type IV translocation-related conjugation system

Robert E. Davis*a,*, Ellen L. Dally*a, Rasa Jomantiene*a,b, Yan Zhao*a, Bruce Roe*c, ShaoPing Lin*c, Jonathan Shao*a

*a Molecular Plant Pathology Laboratory, USDA-Agricultural Research Service, Beltsville, MD 20705, USA
b Microbial Molecular Biology Group, Fitosvirus Laboratory, Institute of Botany, Vilnius 2021, Lithuania
c Advanced Center for Genome Technology, Department of Chemistry and Biochemistry, University of Oklahoma, Norman, OK 73019, USA

Received 17 March 2004, revised 21 June 2004

Communicated by K. Smalla

Abstract

A cryptic plasmid of the wall-less plant pathogenic mollicute, *Spiroplasma kunkelii* CR2-3X, was cloned and its sequence analyzed. The 14,615 bp plasmid, designated pSKU146, has a nucleotide content of 28 mol% G + C, and contains 18 potential protein-coding regions (open reading frames, ORFs), of which six encode proteins that exhibit similarity to virulence-associated proteins involved in cell-to-cell adhesion or conjugal DNA transfer. One ORF encodes a 96 kDa protein, SkARP1, that is highly similar to SARPI adhesin involved in attachment of *Spiroplasma citri* to insect vector gut membrane. Five ORFs encode proteins similar to TraE and Mob in walled bacteria, and to ORFs found in the integrative, conjugative element (ICEF) of *Mycoplasma fermentans*, respectively. Presence of domains similar to proteins of the Type IV secretion system in pathogenic bacteria suggests that spiroplasma possesses a related translocation system. Plasmid pSKU146 also contains two identical *oriT* regions each containing a nick sequence characteristic of the IncP conjugative plasmid family, as well as a 58 bp palindromic sequence, palSK1. Features in pSKU146 suggest that the plasmid functions as a mobile genetic element in conjugative transmission of spiroplasma pathogenicity-related genes.

Published by Elsevier Inc.

Keywords: Virulence; Mobilization; Conjugation; Pathogenicity; Vir domain

* Corresponding author.
E-mail address: davisr@ba.ars.usda.gov (R.E. Davis).

0147-619X/S - see front matter. Published by Elsevier Inc.
1. Introduction

Spiroplasmas are helical, motile, cell wall-less prokaryotes that are classified, along with members of genus *Mycoplasma* and other wall-less bacteria, in class Mollicutes. In their descent from the Gram-positive, low G+C *Bacillus–Clostridium* group of walled bacteria, spiroplasmas underwent massive evolutionary genome reduction, while acquiring capabilities for arthropod parasitism, and in some cases plant pathogenicity. First discovered in association with plant disease (Davis and Worley, 1973; Davis et al., 1972), spiroplasmas comprise broadly diverse species, most of which parasitize a variety of insects or ticks (Gasparich, 2002). As in other bacteria, plasmids presumably have played key roles in spiroplasma evolution, and contributed to the diversity of *Spiroplasma* species, through horizontal exchange of genes encoding virulence factors, niche adaptation factors, and factors conferring competitive advantages on the spiroplasma cell. In the present work, we characterized a plasmid from a plant pathogen, *Spiroplasma kunkelii*.

*Spiroplasma kunkelii* is the causative agent of corn stunt disease, a major factor limiting maize production in the Americas (Davis and Worley, 1973; Davis et al., 1972; Whitcomb and Williamson, 1975). This spiroplasma is transmitted between maize plants exclusively by phloem-feeding leafhopper insects, in which it multiplies. In maize plants, the spiroplasma resides specifically in sieve cells of the plant’s phloem tissue, and induces symptoms including general stunting, broad chlorotic stripes, poor filling of ears, sterility of tassels, and plant death. Although plasmids have been reported in *S. kunkelii* (Gasparich et al., 1993), none has previously been sequenced. Here we present the structural analysis of a previously unreported *S. kunkelii* plasmid, designated pSKU146, harboring regions potentially encoding proteins exhibiting significant homologies, respectively, with a spiroplasma adhesion-related protein and proteins involved in conjugal mating pair formation and DNA transfer.

2. Materials and methods

2.1. Cloning and sequencing of DNA

*Spiroplasma kunkelii* strain CR2-3X was grown in broth medium LD8A3 as described (Lee and Davis, 1989). Spiroplasma cells were embedded in agarose and high molecular weight DNA was prepared according to published procedures (Birren et al., 1999; Peterson et al., 2000). The high molecular weight DNA was partially digested with *Hind*III (New England BioLabs, Beverly, MA) and cloned in vector pBeloBac11 (Kim et al., 1996) according to Peterson et al. (2000). The cloned plasmid DNA was partially sequenced using automated sequencing and the data were combined with whole genome shotgun data from the Spiroplasma Genome Sequencing Project web site at www.genome.ou.edu/spiro.html. The sequence data from these two sources were assembled in silico using the SeqMan program of the sequence analysis software suite Lasergene (DNA-STAR, Madison, WI).

2.2. Nucleotide and amino acid sequence analyses

BLAST searches (Altschul et al., 1990), open reading frame (ORF) analysis, and COGnitor searches (Tatusov et al., 2001) were carried out at the National Center for Biotechnology Information (NCBI) web site (http://www.ncbi.nlm.nih.gov). In silico restriction mapping and multiple sequence alignments were performed using the MapDraw and Megalign programs, respectively, of the Lasergene sequence analysis software suite, and conserved regions were defined as common segments in the alignment. A signal sequence search was done using SignalP (Nielsen et al., 1997) (http://www.cbs.dtu.dk/services/SignalP), and the Simple Modular Architecture Research Tool SMART (Letunic et al., 2002; Schultz et al., 1998) (http://smart.embl-heidelberg.de) was used to search for signal sequences, transmembrane regions, and regions of low complexity. Protein mass and isoelectric points were estimated using the GeneQuest option of the Lasergene suite. A search for gene orthologs in completely sequenced gen-
omes of *Mycoplasma* species was carried out by using the ortholog table construction option at the Whitehead Institute Center for Genome Research (http://www-genome.wi.mit.edu/annotation/microbes/methanosarcina/keggmap.html).

### 3. Results and discussion

The complete nucleotide sequence of plasmid pSKU146 has been deposited in the GenBank database under GenBank Accession No. AY528560. Plasmid pSKU146 is a circular molecule of 14,615 bp, within the size range of naturally occurring extrachromosomal DNAs thus far reported in *Spiroplasma* spp. and similar to the size of extrachromosomal DNAs reported in *S. kunkelii* (Gasparich and Hackett, 1994; Ranhand et al., 1980; Razin et al., 1987; Salvado et al., 1989). Plasmid pSKU146 has a G + C content of 28 mol%. A search for open reading frames (ORFs) larger than 30 codons resulted in 18 potential protein-coding regions that were compared with current databases (Table 1). The ORFs were located in all three reading frames on the same DNA strand and had ATG as the translational start codon. The plasmid sequence was numbered arbitrarily starting at the first base of the inverted repeat in a palindromic sequence designated palSK1 (Fig. 1). The deduced amino acid sequences encoded by 13 ORFs displayed some similarities to known sequences (Table 1). The overall organization of the plasmid includes genes encoding a putative plasmid partitioning protein (ParA) and a putative adhesin (SkARP1), a region encoding putative Type IV secretion system-like components of conjugation machinery (ICEF-IA ORF15-like, ICEF-IA ORF16-like, and TraE protein with VirB4 domain similar to ICEF-II ORF17 of *Mycoplasma fermentans*), and a duplicated origin of transfer (oriT) region. These features will be discussed in detail below.

#### 3.1. Putative plasmid partitioning protein

ORF6 (bases 1602–2369) encodes a putative protein with ATPase activity related to Soj family proteins (COG1192) that are involved in chromosome and plasmid partitioning in other bacteria. A Pfam:ParA domain is predicted to occur at residues 82–193. ParA family ATPases are involved in bacterial plasmid partitioning, and an analysis of the similarity of ORF6 with described proteins in GenBank revealed similarity with the ATPase involved in chromosome partitioning in *Corynebacterium glutamicum* (NP_600639; 29% identity over 236 amino acids), and the ParA protein of *Leptospira interrogans* (NP_714527; 33% identity over 209 amino acids).

#### 3.2. Spiroplasma adhesin gene pathogenicity factor

ORF9 (bases 3740–6337) encodes a putative protein homolog of the SARP1 adhesion protein (P89) of *Spiroplasma citri*. The ORF9 putative protein, designated SkARP1, is a new member of the sarpin family of proteins described by Berg et al. (2001) and may function similarly as an adhesin of *S. kunkelii*. A putative signal peptide of 23 amino acids is present at the N-terminus of the preprotein. The predicted mature protein (SkARP1), after cleavage of the signal peptide, comprised 842 amino acids with a calculated molecular mass of 96 kDa. This contrasted with the mature SARP1 (here termed ScARP1) protein (GenBank Accession No. AJ297706) of *S. citri*, with a mass of 86 kDa (Berg et al., 2001). The SkARP1 signal peptide was the same in size and similar in amino acid sequence to that found in the ScARP1 preprotein. Low sequence similarity of the N-termini accounts for much of the size difference between the two mature proteins. The mature proteins also differed in numbers and identities of charged amino acids and have different isoelectric points (6.3 for SkARP1 and 5.5 for ScARP1).

It is known that ScARP1 is involved in the attachment of *S. citri* cells to gut cells of the insect vector, *Circulifer tenellus*, during early stages of infection in this host (Berg et al., 2001), and therefore it is quite likely that the *S. kunkelii* SkARP1 ortholog similarly is involved in attachment of *S. kunkelii* cells to gut tissue during infection of its insect vectors, which include *Dalbulus* spp. We postulate that amino acid sequence variations and differences in isoelectric points may reflect dif-
<table>
<thead>
<tr>
<th>ORF</th>
<th>Endpoints (nt)</th>
<th>G + C %</th>
<th>Ribosome-binding site/ start codon*</th>
<th>Product (aa/kDa)</th>
<th>Best BLAST hit b % identity (over aa)</th>
<th>Predicted function/ similar protein</th>
<th>Related COG c and gene name</th>
<th>Functional d category</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORF1</td>
<td>140–733</td>
<td>31.8</td>
<td>GAGTGTTGAGTTGAGTTGATATT/ATG</td>
<td>197/23.6</td>
<td>NP_814286 28 (164)</td>
<td>Hypothetical protein</td>
<td>No related COG</td>
<td></td>
</tr>
<tr>
<td>ORF2</td>
<td>789–893</td>
<td>12.4</td>
<td>AATTGAAACAAATACGGAGTAAATAAT/ATG</td>
<td>34/4.0</td>
<td>No significant hits</td>
<td>Hypothetical protein</td>
<td>No related COG</td>
<td></td>
</tr>
<tr>
<td>ORF3</td>
<td>895–1071</td>
<td>17.5</td>
<td>AAAATAGCATAAAAAAAGGGATTAAAT/ATG</td>
<td>58/7.2</td>
<td>NP_703879 33 (56)</td>
<td>Hypothetical protein</td>
<td>No related COG</td>
<td></td>
</tr>
<tr>
<td>ORF4</td>
<td>1034–1516</td>
<td>28.4</td>
<td>N.A./ATG</td>
<td>160/19.5</td>
<td>CAD58576 26 (125)</td>
<td>Hypothetical protein (TriL-like, MobC-like)</td>
<td>No related COG</td>
<td>N</td>
</tr>
<tr>
<td>ORF5</td>
<td>1095–1190</td>
<td>28.1</td>
<td>N.A./ATG</td>
<td>31/3.8</td>
<td>No significant hits</td>
<td>Hypothetical protein</td>
<td>No related COG</td>
<td></td>
</tr>
<tr>
<td>ORF6</td>
<td>1602–2369</td>
<td>23.8</td>
<td>AATATATAAGAAGGGAAACACAAA/ATG</td>
<td>255/29.7</td>
<td>NP_600639 29 (236)</td>
<td>Soj protein, PFAM ParA</td>
<td>Soj D</td>
<td></td>
</tr>
<tr>
<td>ORF7</td>
<td>2372–2626</td>
<td>26.7</td>
<td>TAGAACAAAAGGAAGAGTTGATATAAC/ATG</td>
<td>84/9.8</td>
<td>NP_228535 30 (82)</td>
<td>Hypothetical protein, RepA-like region</td>
<td>No related COG</td>
<td></td>
</tr>
<tr>
<td>ORF8</td>
<td>2774–3088</td>
<td>23.5</td>
<td>TTAAAATAAAATAGAGAAAAGGAGTAAATAAT/ATG</td>
<td>104/12.4</td>
<td>CAD20867 23 (89)</td>
<td>Hypothetical protein</td>
<td>No related COG</td>
<td></td>
</tr>
<tr>
<td>ORF9</td>
<td>3740–6337</td>
<td>30.1</td>
<td>TTATTGTGAAGAGTTGTTTTTTTTAAGTTGTT/ATG</td>
<td>865/98.4</td>
<td>CAC10363 71 (536)</td>
<td>SKARPl adhesion protein</td>
<td>No related COG SkarPl</td>
<td></td>
</tr>
<tr>
<td>ORF10</td>
<td>6337–6606</td>
<td>37.4</td>
<td>AAAACAAAAGGAGGGCGGTTGATGAATA/ATG</td>
<td>89/9.4</td>
<td>AAN85225 31 (69)</td>
<td>Hypothetical protein (ICEF-1A ORF15-like)</td>
<td>No related COG</td>
<td></td>
</tr>
<tr>
<td>ORF11</td>
<td>6618–8180</td>
<td>32.1</td>
<td>CAATCACAAGTTGAGCGGTTGTAAT/ATG</td>
<td>520/57.6</td>
<td>CAC10364 87 (82)</td>
<td>Hypothetical protein (ICEF-1A ORF16-like)</td>
<td>No related COG</td>
<td></td>
</tr>
<tr>
<td>ORF12</td>
<td>8170–10,692</td>
<td>28.2</td>
<td>AAAAAAAAGTTAAGGGAGAATGAAACAAAATG</td>
<td>840/96.9</td>
<td>AA074893 94 (840)</td>
<td>TraE protein with VirB4 domain</td>
<td>COG3451 TraE N</td>
<td></td>
</tr>
<tr>
<td>ORF13</td>
<td>10,696–11,034</td>
<td>28.6</td>
<td>AAAAAAGATTTTTTTGGACAAAACAAAC/ATG</td>
<td>112/13.3</td>
<td>NP_040346 33 (95)</td>
<td>Hypothetical protein</td>
<td>No related COG</td>
<td></td>
</tr>
<tr>
<td>ORF14</td>
<td>11,136–12,650</td>
<td>27.5</td>
<td>TCTTTAAAGGAAGGTGTTTAAAT/ATG</td>
<td>504/58.3</td>
<td>NP_114056 23 (425)</td>
<td>Mob protein with VirD4 domain</td>
<td>COG3505, 0433 N</td>
<td>Mob</td>
</tr>
<tr>
<td>ORF15</td>
<td>12,833–12,955</td>
<td>21.1</td>
<td>N.A./ATG</td>
<td>40/5.0</td>
<td>No significant hits</td>
<td>Hypothetical protein, membrane associated</td>
<td>No related COG</td>
<td></td>
</tr>
<tr>
<td>ORF16</td>
<td>13,002–13,133</td>
<td>25.8</td>
<td>N.A./ATG</td>
<td>43/5.0</td>
<td>No significant hits</td>
<td>Hypothetical protein, membrane associated</td>
<td>No related COG</td>
<td></td>
</tr>
<tr>
<td>ORF17</td>
<td>13,491–14,135</td>
<td>30.1</td>
<td>ACTATTATTGAGAGGAGAACATAATTT/ATG</td>
<td>214/25.5</td>
<td>NP_701124 26 (143)</td>
<td>Hypothetical protein</td>
<td>No related COG</td>
<td></td>
</tr>
<tr>
<td>ORF18</td>
<td>14,213–14,338</td>
<td>23.8</td>
<td>N.A./ATG</td>
<td>41/5.1</td>
<td>No significant hits</td>
<td>Hypothetical protein, membrane associated</td>
<td>No related COG</td>
<td></td>
</tr>
</tbody>
</table>

a Bold underlined nucleotides denote purine-rich region within the putative ribosomal-binding site (RBS).

b GenBank accession number or protein ID of the best BLAST hit, followed by the percent identity between the query and the best hit.

c COG stands for Cluster of Orthologous Groups.

d Functional classification based on the result of COG search.

e N.A., not applicable; putative RBS not identified.

f Similar to ICEF-II ORF17 of *Mycoplasma fermentans*. 
ferences in the ligand attachment sites of ScARP1 and SkARP1 in their respective insect hosts.

Based on work with *Enterococcus faecalis* (Hirt et al., 2000; Olmsted et al., 1991), it is possible that putative virulence factors SkARP1 and ScARP1 could be involved in formation of spiroplasma mating aggregates for conjugal DNA transfer. Studies of the aggregation substance (AS) surface protein encoded by *E. faecalis* virulence plasmid pCF10 indicate that AS is a virulence factor involved in the adhesion of bacterial cells to host tissues (Hirt et al., 2000), as well as a conjugation protein involved in the formation of a mating pair, enhancing conjugal transfer of the plasmid (Olmsted et al., 1991). A homologous protein is encoded by *E. faecalis* plasmid pAD1 (Francia et al., 2001). It would be interesting to know whether the SkARP1 and ScARP1 putative virulence factors likewise are involved in the formation of spiroplasma mating aggregates leading to conjugal DNA transfer.

3.3. Components of conjugation machinery, Type IV secretion system-like proteins

Plasmid pSKU146 contains ORFs that encode proteins having similarity to conjugative elements and components of the bacterial Type IV secretion system. The deduced amino acid sequence encoded by ORF10 possesses a signal peptide (residues 1–34) and a single transmembrane region (residues 62–80). ORF10 displayed similarity with ICEF-IA ORF15 of *M. fermentans* (GenBank Accession No. AAN85225; 31% identity over 69 amino acids). The ICEF of *M. fermentans* is a non-self replicating, chromosomally integrating DNA that encodes proteins similar to known conjugation proteins and that is capable of being transferred during a conjugation-like event (Calcult et al., 2002).

ORF11 (bases 6618–8180) encodes a hypothetical protein having similarity with a hypothetical protein of *S. citri* (GenBank Accession No. CAC10364; 87% identity over 82 amino acids) and similarity with several other proteins, including the integrative conjugal element ICEF-IA ORF16 of *M. fermentans* (GenBank Accession No. AAN85226; 27% identity over 147 amino acids). The deduced amino acid sequence encoded by ORF11 contains eight transmembrane regions (residues 36–58, 99–121, 153–175, 185–207, 260–282, 287–306, 327–349, and 353–372), and two regions of low complexity at residues 461–471 and 503–520, respectively. This predicted protein also shares topological similarities with membrane-trafficking proteins, in particular the bacterial Family 1 export proteins that export signal peptide-lacking proteins through the membrane.

ORF12 (bases 8170–10,692) encodes a TraE-like protein, a putative membrane-bound ATPase with a VirB4 domain related to Type IV secretory
pathway COG3451. This putative conjugal gene transfer protein is among the first to be found in class Mollicutes.

The protein encoded by ORF12 contains a signal peptide (residues 1–49), two transmembrane regions (50–72 and 79–93), and a low complexity region (244–253) predicted by SMART. The presence of a signal peptide and transmembrane regions is uncharacteristic of VirB4 proteins and may indicate that the ORF12 protein carries out additional functions. This protein has similarity with a putative membrane-bound ATPase of *S. kunkelii* (GenBank Accession No. AAO74893; 94% identity over 840 amino acids), as well as to other GenBank listed proteins that include a TRSE-like protein of *Mycoplasma pulmonis* (GenBank Accession No. NP_326214; 27% identity over 795 amino acids); protein ICEF-II ORF17 of the integrative, conjugative element of *M. fermentans* (ICEF) (GenBank Accession No. NP_047296; 22% identity over 601 amino acids); a putative ATPase TraE of *Staphylococcus epidermis* plasmid pMRC01 (GenBank Accession No. NP_765045; 21% identity over 522 amino acids); a putative pilus assembly protein of *Pseudomonas resinovorans* (GenBank Accession No. NP_758682; 19% identity over 660 amino acids); and a sex pilus assembly protein of *Vibrio cholerae* (GenBank Accession No. AAL59681; 19% identity over 409 amino acids).

Thus, the gene arrangement of pSKU146 ORFs 10 through 12 parallels the gene arrangement in ORFs 15 through 17 in the ICEF of *M. fermentans*, consistent with the concept that pSKU146 is also involved in DNA transfer. The results also raise the possibility that pSKU146 and the ICEF shared a common ancestor or donor that contributed pSKU146 ORF10-, ORF11-, and ORF12-like sequences to both elements.

ORF14 (bases 11,136–12,650) encodes another putative protein exhibiting significant similarity with bacterial conjugation proteins. Domains predicted in the deduced amino acid sequence were a signal peptide (residues 1–41), a transmembrane domain (residues 64–86), and a Pfam:TRAG domain (residues 103–491). The putative ORF14-encoded protein exhibits the same domain architecture as conjugal transfer protein VirD4 (GenBank Accession No. CAC15172) of *Agrobacterium tumefaciens* and has similarity with the Mob protein (GenBank Accession No. NP_114056) of *Streptococcus mutans* (23% identity over 425 amino acids) and COGs related to VirD4, Type IV secretion pathway COG3505, and predicted ATPase COG0433. A putative Walker A nucleotide-binding site, 139GT TGSGKT146, is present in the deduced amino acid sequence of the pSKU146-encoded Mob.

Since bacterial Mob proteins are involved in plasmid mobilization and transfer to a recipient cell, and the Tra-G-TraD family proteins and VirD4 are bacterial coupling proteins involved in passage of conjugally transferred DNA through the mating channel and in protein translocation via the Type IV secretion system (Christie, 2001), the presence of ORFs encoding domains similar to Vir and Tra proteins in pSKU146 suggests that *Spiroplasma* spp. utilize components related to the Type IV system for macromolecule transfer. A recent report of a putative pilus-like structure in electron micrographs of spiroplasma cells (Ozbek et al., 2003) is consistent with this concept. We therefore hypothesize that pSKU146 can be conjugally transferred to recipient cells through a Type IV-related secretory machinery involving proteins encoded by the plasmid. The possibility that additional proteins related to components of a Type IV translocation system are encoded by the *S. kunkelii* genome is currently under investigation.

Compared to the Gram-negative *A. tumefaciens*, we anticipate that the *S. kunkelii* genome (www.genome.ou.edu/spiro.html) encodes fewer Type IV secretion system-related putative conjugation proteins, but presumably encodes all that it requires for a functional translocation system. Notably, only four homologues of Type IV secretory pathway components have been previously found to be encoded by conjugative elements from unicellular Gram-positive bacteria (Grohmann et al., 2003). *Spiroplasma* is a member of the Firmicutes, Gram-positive low G + C bacteria. As noted previously (Waters, 1999), Gram-positive bacteria, lacking an outer membrane, may not require as many components as Gram-negative bacteria for formation of a conjugative bridge. *Spiroplasma*, lacking a wall and
bounded only by a single membrane, may require fewer components than are required for DNA transfer among bacteria having an outer membrane.

3.4. Type IV secretion system-related genes in Mollicutes

As discussed above, the Type IV secretion system is an intercellular macromolecule transfer machinery that is evolutionarily related to conjugation systems of Gram-negative bacteria and also is implicated in bacterial pathogenesis for its capacity to deliver virulence effectors to eukaryotic host cells (Christie, 2001). In addition to Spiroplasma (www.genome.ou.edu/spiro.html and this communication), new evidence has emerged indicating that Type IV secretion-related systems might exist among other Gram-positive, cell wall-less bacteria, since genes encoding components of such a system are found in the genomes of Mycoplasma species, and include the M. pulmonis TRSE-like protein that has similarity with the pSKU146 ORF12 putative protein as noted above. The TraE/TrsE family NTPase protein gene (Calcutt et al., 2002) and a putative TraG family NTPase gene in the integrative conjugal element (ICEF) of M. fermentans (GenBank Accession No. AY168953) encode proteins similar to the deduced amino acid sequence encoded by ORF12 of pSKU146 that contains a region similar to the Type IV secretion system-related VirB4 domain. A search for VirD4-related sequences among the completely sequenced genomes of Mollicutes reveals a sequence, related to the C-terminal portion of the plasmid pSKU146 putative Mob protein, in Mycoplasma pneumoniae that encodes a putative VirD4-like amino acid sequence (GenBank Accession No. MPN513). These findings seem to indicate that Vir-related proteins, Type IV translocation system-related elements, may be more common among wall-less bacteria than previously realized.

3.5. Ribosomal-binding sites and coordinate regulation

Several of the ORFs in pSKU146 are in juxtaposition or overlap. The start codon of ORF10, encoding a hypothetical protein exhibiting similarity to the M. fermentans mobile genetic element ICEF-IA ORF15, overlaps the termination codon of ORF9, which encodes a putative adhesin (SKARP1). The termination codon of ORF10 and the start of ORF11, encoding a protein with similarity to ICEF-IA ORF16, are separated by only 12 bases. The start of ORF12, encoding a TraE-like protein, overlaps the 3′-end of ORF11 by 10 bases. ORF13 begins 4 bases after the termination codon of ORF12, and ORF14, encoding a putative Mob protein, begins only 2 bases after the termination codon of ORF13. A putative ribosome-binding site (RBS) could be located upstream of each of the above ORFs, and in some cases is within the 3′-end of the respective ORF immediately upstream. Departure of some RBS from strict homology with the 3′-end of 16S rRNA may reflect translational regulation through modulating the formation of the translation initiation complex.

Based on the relative positions of the juxtaposed and overlapping ORFs, it is likely that at least some of these genes are coordinately expressed. It is possible that the tra (transfer)-related genes (ORFs 10, 11, 12, and 14) are co-transcribed with skarp1 as a single operon. Plasmid R1 contains a polycistronic tra-operon (Koraimann and Högenauer, 1989), and co-transcription has been demonstrated for tra genes of plasmid pIP501 from Streptococcus agalactiae (Kurenbach et al., 2002). The tra genes in pSKU146 may be expressed under the influence of the skarp1 adhesin gene (ORF9) promoter. This hypothesis suggests a possible relationship between the skarp1 gene and the genes encoding components of the transfer machinery. In addition, the possible coordinate expression could imply that the adhesin protein functions not only in spiroplasma cell–host cell adhesion, but also in spiroplasma cell-to-cell contact for conjugation, as reported earlier in the case of plasmid-encoded adhesin of E. faecalis involved in contact between mating cells (Hirt et al., 2000; Olmsted et al., 1991; Wirth, 1994).

3.6. Origin of transfer (oriT) and palindromic sequences

Plasmid pSKU146 contains two identical 175 bp direct repeat units, designated Repeat
Alpha (bases 377–551) and Repeat Beta (bases 1160–1334), in each of which is embedded the sequence ATCCTG, a sequence that is conserved at the oriT nick region among conjugative plasmids (Furuya and Komano, 2000; Herrera-Cervera et al., 1998). The ATCCTG in pSKU146 is part of an imperfect inverted repeat, that can potentially form a hairpin structure important for relaxosome formation and endonuclease protein binding that results in a single-stranded nick required for plasmid mobilization during conjugation. The ATCCTG putative nick region sequence in pSKU146 is identical to nick regions found in the IncP conjugative plasmid family (Waters, 1999) and is preceded 5 bases upstream by a 10-base AT-rich sequence. The putative nick site (nic) within this region likely lies between bases 494 and 495 and between bases 1277 and 1278 in oriTA and oriTB, respectively (Fig. 2). Repeat Alpha and Repeat Beta thus represent two putative oriT regions (oriTA and oriTB, respectively) that probably resulted from sequence duplication. Two oriT sequence regions have been reported in other plasmids (Avila et al., 1996; Becker and Meyer, 2003; Herrera-Cervera et al., 1998), including the conjugative virulence plasmid pAD1 of E. faecalis (Francia et al., 2001).

In pSKU146, oriTA is embedded in putative protein-coding region ORF1 (bases 140–733). oriTB is embedded in a putative protein-coding region (ORF4, bases 1034–1516) and spans most of ORF5 (bases 1095–1190). The region spanned by the two oriT regions and the sequence upstream of oriTA is characterized by a cluster of direct and inverted repeat sequences. The two oriT regions each contain at least seven inverted repeat sequences (IR), including the putative nick region (nick) at IR5 (Fig. 2). Four of these inverted repeats are located upstream and two are located downstream of the putative nick region, resulting in a structural order as follows: oriTA IR1β-IR2β-IR3β-IR4β-nick-IR6β-ORF2-ORF3-oriTB (IR1β-IR2β-IR3β-IR4β-nick-IR6β-IR7β).

Interestingly, the A + T content of pSKU146 oriTA and oriTB (69%) is lower than the A + T content of the intervening sequence between the two oriT regions (76%) and of sequences flanking the putative double oriT region, as well as being lower than the A + T content of the plasmid (72%) and that of the spiroplasma genome (74%). A similar observation has been made for the oriT region of conjugative element pRS01 from L. lactis subsp. lactis ML3 (Mills et al., 1998). In pSKU146, the A + T content of sequences flanking the oriTA-oriTB region is 86% for bases 14,245–14,483 (located upstream of oriTA), 71% for bases 1–378 (from beginning of the second inverted repeat in palSK1 to the beginning of oriTA), 76% from the end of oriTB to the start of the soj gene, and 76% in the soj gene.

A prominent feature of pSKU146 is a palindromic sequence, palSK1, of 58 bp (bases 1–29 abutting 14,587–14,615) located upstream of ORF1 and consisting of two abutted, inverted repeats having the sequence GGTAGGTTGTTATTGGTTGTAATTGCAATAAACATC.

Fig. 2. Nucleotide sequence and structure of putative oriTA region in plasmid pSKU146. Inverted repeat (IR) sequences are represented by opposing arrows. Paired bases are underlined. Direct repeats (DR) are indicated by non-opposing arrows. The putative nick region within the oriT region is in bold and double underlined. The putative nick site is indicated by a vertical arrow. An AT-rich sequence is indicated upstream of the putative nick region. oriTA and oriTB are separated by a 608 bp sequence and are identical in sequence (not shown).
A second prominent structural feature, designated IRSK, is located at base positions 12,810–12,855 and consists of two 20 bp inverted repeat units that are separated by 6 bp and have the sequence ACTAAAAAATAAAAACACCTtttatgAGGTGTTTTTATTTTTAGT. The IRSK sequence potentially forms a hairpin structure 160 bases downstream of the Mob gene (ORF14). The start codon for ORF15 is located within the six bases separating the inverted repeats of IRSK; this ORF may not be functional, since a putative RBS could not be identified. Whether palSK1 or IRSK is involved in plasmid mobilization, as noted for some palindromes (Priebe and Lacks, 1989), remains to be determined.

No ORFs with significant sequence similarity to any rep genes previously associated with plasmid origins of replication in Gram-positive bacteria were identified. However, a low scoring PFAM HTH_4 domain (amino acid positions 46–83) was identified in the putative protein product of ORF7, through a domain search using SMART. A subsequent BLASTP search, using the region spanning amino acids 46–83 as a query against the SMART protein database, revealed similarity between the 84-amino acid ORF7 protein and the 49-amino acid replication protein (RepA) (NP_862029) from L. lactis plasmid pSH72 (39% identity in 38 a.a. overlap; 68% similarity in 38 a.a. overlap) as well as the 52-amino acid RepA protein (CAA52852) from Lactobacillus fermentum (28% identity in 38 a.a. overlap; 71% similarity in 38 a.a. overlap). In addition, a sequence similar to the plasmid pGT5/pC194 rolling circle plasmid family double-stranded origin (dsO) sequence TTATCTTGATA occurs twice with a single base mismatch (bases 11,312–11,322, within ORF14 on the forward strand, and bases 13,324–13,334 on the reverse strand), and the core sequence CTTGATA is exactly matched four times (three times on the forward strand at bases 739–745, at 12,242–12,248 within TTATcaaCTTGATA in ORF14, and at 13,737–13,743 in ORF16, and once on the reverse strand at bases 13,803–13,809), in pSKU146. Finally, the putative Salmonella NTP1 plasmid origin of replication sequence AGGCCGTT occurs once in pSKU146 at bases 4035–4041, within ORF9.

### 3.7. Hypothetical proteins and spiroplasma virus protein homolog

The putative protein product of ORF1 exhibited similarity (28% identity over 164 residues) to a hypothetical protein from E. faecalis V583 (GenBank Accession No. NP_814286). Both ORF1 and ORF4 protein products exhibited similarities (ORF1 28% over 164 residues and ORF4 29% over 151 residues) to an unknown protein from the E. faecalis conjugative virulence plasmid pAD1 (GenBank Accession No. CAD58576) encoded by a cryptic plasmid that encodes a conjugative transfer system in Yersinia enterocolitica (Goelz et al., 2003; Strauch et al., 2003). This degree of similarity raises the possibility that the proteins encoded by these ORFs are the result of horizontal gene transfer and may be involved in horizontal transfer of plasmid pSKU146.

The ORF13 (bases 10,696–11,034) putative protein has similarities (36% identity over 95 residues, see Table 1) to a hypothetical protein of spiroplasma virus 1 (GenBank Accession No. NP_040346). However, of the remaining ORFs (ORF2, ORF3, ORF8, and ORF17), the deduced amino acid sequences displayed little or no significant similarity to previously deposited GenBank sequences using BLASTP searches. Although they most likely are expressed since they are preceded by putative ribosome-binding sites (RBS) (Table 1), any function(s) associated with these putative proteins thus remain cryptic. The putative products of the only remaining ORFs (ORF5, ORF15, ORF16, and ORF18) exhibited no similarity to known sequences, and since clear ribosomal-binding sites were not observed, these ORFs may not be translated into functional proteins. However, there is mounting evidence that several organisms, including M. pneumoniae express genes that lack clear ribosomal-binding sites (Weiner III et al., 2000). Thus, since the putative protein products of ORF15, ORF16, and ORF18 each contain a transmembrane segment (ORF15...
product residues 5–27, ORF16 product residues 4–23, and ORF18 product residues 7–29) predicted by SMART, it is likely that these ORFs result in expressed proteins that are membrane associated. It would be interesting to know whether they function, or have functioned, in macromolecule transfer in the spiroplasma or a progenitor.

In summary, plasmid pSKU146 harbors genes encoding putative proteins similar to known bacterial virulence factors and conjugative elements including Mob family proteins, origins of plasmid transfer, Vir domains, ICEF ORF-like proteins, and the ScARP1 adhesin of \textit{S. citri}, implicating a role of pSKU146-encoded genes in conjugal cell-to-cell communication and pathogenesis. Conjugative and mobilizable plasmids have not been previously described in genus \textit{Spiroplasma}, making the putative conjugative element pSKU146 the first of its type to be described in Spiroplasmataceae.

The finding of putative conjugative plasmid sequences in \textit{S. kunkelii} raises some important questions. For example, it will be of interest to learn whether other \textit{Spiroplasma} species carry the array of putative pathogenicity factors observed in pSKU146 and similar plasmids may influence the biology of pathogenic spiroplasmas and contribute to an evolving spiroplasma diversity. One may also ask whether conjugative plasmids in spiroplasmas may mediate genetic exchange with other genera of organisms, including phytoplasmas and \textit{Lieberobacter} spp.; phytoplasmas and \textit{Lieberobacter} spp. are wall-less and walled plant pathogens, respectively, that inhabit sieve cells of plant phloem tissue just as do spiroplasmas, but cannot be isolated in artificial culture. Some phytoplasmas even share both insect vector and plant host species with a spiroplasma. For example, maize bushy stunt (MBS) phytoplasma shares plant and insect hosts with \textit{S. kunkelii}, and periwinkle virescence phytoplasma (beet leafhopper transmitted virescence agent, VR) shares plant and insect hosts with \textit{S. citri} (Barros et al., 2001; Nault, 1980; Oldfield, 1984). Moreover, since \textit{S. kunkelii} is a member of the spiroplasma-\textit{Mycoplasma mycoides} clade (Weisburg et al., 1989), which contains \textit{M. mycoides} and \textit{Mycoplasma capricolum}, one may ask whether similar plasmids play a role in the biology of mammalian pathogenic \textit{Mycoplasma} spp., including the development and spread of antibiotic resistance.

Acknowledgments

\textit{Spiroplasma kunkelii} gene sequence data were provided by the \textit{Spiroplasma kunkelii} Genome Sequencing Project funded by U.S. Department of Agriculture, Agricultural Research Service Project Number: 1275-22000-144-02. The data were made available by Bruce Roe, ShaoPing Lin, HongGui Jia, HongMin Wu, and Doris Kupfer (University of Oklahoma, Department of Chemistry and Biochemistry, Norman, Oklahoma 73019) and Robert E. Davis (U.S. Department of Agriculture-Agricultural Research Service, Molecular Plant Pathology Laboratory, Beltsville, MD 20705). We thank J. Fletcher (Oklahoma State University, Stillwater) for kindly supplying \textit{Spiroplasma kunkelii} strain CR2-3X.

References


Calcutt, M.J., Lewis, M.S., Wise, K.S., 2002. Molecular genetic analysis of ICEF, an integrative conjugal element that is present as a repetitive sequence in the chromosome of


