Complete Nucleotide Sequence of Ubiquitous Plasmid pEA29 from *Erwinia amylovora* Strain Ea88: Gene Organization and Intraspecies Variation

GAYLE C. McGHEE AND ALAN L. JONES*

Department of Botany and Plant Pathology, Michigan State University, East Lansing, Michigan 48824-1312

Received 12 May 2000/Accepted 31 August 2000

The complete sequence of plasmid pEA29 from *Erwinia amylovora* strain Ea88 consists of 28,185 bp with a 50.2% G+C content. As deletions and insertions were detected in other derivatives of pEA29, its size actually varied from 27.6 to 34.9 kb. Thirteen open reading frames that encoded predicted proteins with similarities to known proteins from other bacteria were identified along with two open reading frames related to hypothetical proteins found in GenBank and six open reading frames with no similarities to existing GenBank entries. Predicted products of open reading frames with similarity to the thiamine biosynthetic genes thiO, thiG, and thiF; a betR gene coding for choline transport; an msrA gene for the enzyme methionine sulfoxide reductase; a putative methyl-accepting chemotaxis gene; an aldehyde dehydrogenase gene; an hns DNA binding gene; a LysR-type transcriptional regulator; and parA and parB partitioning genes were identified. A putative iteron-containing theta-type origin of replication with a AT-rich region and a gene for a RepA protein was identified. *PsI* and *KpnI* restriction patterns for pEA29 isolated from tree fruit strains of *E. amylovora* were homogeneous and different from those for pEA29 isolated from *Rubus* (raspberry) strains. All *Rubus* derivatives of pEA29 contained a point mutation that eliminated a *PsI* site and a 1,264-bp region that replaced 1,890 bp of sequence found in pEA29 from strain Ea88. This change eliminated a second *PsI* site and increased the length of a *KpnI* fragment. An insertion sequence, IS*EamI*, was detected in one *Rubus* strain, and transposon Tn5393 was detected in three apple strains in two separate locations on the plasmid. Plasmid-cured strains exhibited reduced virulence and modified colony morphology on minimal medium without thiamine, indicating that some of the genes in pEA29 play a role in the physiology or metabolism of *E. amylovora*.

The plasmid biology of *Erwinia amylovora*, a phytopathogen that causes fire blight, is gradually emerging. The pathogen infects plants in the family Rosaceae, including apple and pear in the subfamily Pomoideae and a few other rosaceous plants such as *Rubus* spp. All naturally occurring strains of *E. amylovora* contain a nontransmissible plasmid called pEA29 or pEA28 (15, 24), and up to 25% of strains also contain a 56-kb plasmid (48). Some streptomycin-resistant strains of *E. amylovora* carry the tandem strA and strB genes on plasmids. Many of these strains contain the conjugative plasmid pEa34 (28), which carries strA-strB on Tn5393. Other strains carry strA and strB genes on a small mobilizable relative of the IncQ broad-host-range plasmid RSF1010 named pEa8.7 or in transposon Tn5393 on pEA29 (28, 37). The geographic distribution of pEA29 is the same as the distribution of *E. amylovora*, while pEa34 and pEa8.7 have been detected only in strains of *E. amylovora* from Michigan and California, respectively. The distribution of the 56-kb plasmid is not correlated with the geographic distribution of strains of *E. amylovora*. None of the plasmids found in *E. amylovora* have been completely sequenced, but the sequence of pEa8.7 is probably nearly identical to that of plasmid RSF1010, which has been sequenced (45).

Plasmid pEA29 has received considerable attention because its high stability in *E. amylovora* suggests that it may confer some traits advantageous to the bacterium. Also, PCR-based detection of pEA29 is the most commonly used method for the detection and identification of *E. amylovora* and is based on amplification using primers within the 1.1-kb *PsI* fragment (5, 25, 29). The lengths of the amplification products vary between strains due to differences in the numbers of 8-bp repeats found in the amplified fragments (21, 44), a characteristic that proved to be unstable in the host strain (44). Restriction maps of pEA29 from two isolates have been published by two groups, with some minor differences between the maps (15, 24). The plasmid is not easily eliminated by classical chemical and physical treatments; eviction by incompatibility is required for curing bacteria of pEA29 (15, 24). Plasmid-cured strains are still pathogenic but exhibit reduced virulence (16, 24) and altered growth on minimal medium without thiamine (MM − T) (16, 24). Except for the probable location of the origin of replication on the 4.4-kb *PsI* fragment and the sequence of a 1.1-kb *PsI* fragment, little is known about the genetic makeup of pEA29. No specific plasmid sequences have been associated with bacterial virulence or physiology. Increased knowledge of the genetic makeup of pEA29 may be of value in explaining why this plasmid is maintained by *E. amylovora*.

Sequence analysis of pEA29 was conducted to determine the genetic makeup of this indigenous plasmid. A genetic map was generated based on sequence similarities to GenBank entries for other bacteria. Physiological experiments were conducted with plasmid-cured strains to evaluate their oxidase sensitivities, chemotactic responses, and thiamine auxotrophy. Finally, based on the sequence analysis of pEA29 from strain Ea88, we were able to characterize restriction length polymorphisms found in plasmid pEA29 from *Rubus* strains and in a few streptomycin-resistant strains of *E. amylovora* from apple.

* Corresponding author. Mailing address: Michigan State University, 103 Center for Integrated Plant Systems, East Lansing, MI 48824. Phone: (517) 355-4573. Fax: (517) 353-5598. E-mail: jonesa@pilot.msu.edu.
The remaining sequence was filled in by sequencing directly from pEA29. The ori-

fientations of sequences from each fragment were verified by sequencing across

\( Pst \) sites. Sequences from other strains were obtained by sequencing directly from

pEA29. All sequencing was performed at the Michigan State University

Department of Energy Plant Research Laboratory sequencing facility using ABI dye-terminator chemistry. Sequenced fragments were analyzed and assembled using the Lasergene software of DNASTAR Inc. (Madison, Wis.). The majority of the assembled sequence was derived from single sequencing runs, the

sequences from which were proofread by eye; where positions were ambiguous, they were reported as "N." Sequence analysis was conducted using BLAST version 2.0 (1) at the National Center for Biotechnology Information website (http://www.ncbi.nlm.nih.gov/BLAST) and the Pfam database of protein families search engine (http://pfam.wustl.edu/) (3). Open reading frames of 375 nucleotides or more were identified, and the sequences of their predicted products were compared to sequences in the GenBank protein database using the BLAST P algorithm and to sequences in the Pfam database. The algorithms BLASTN and BLAST X were also used to compare pEA29 DNA sequences to those in the GenBank database.

PCR. *Erwinia amylovora* strain Ea88 was used as a positive control, and 10 \( \mu l \) of water was used as a negative control for all reactions. Primers AJ75 (5'-CTGTA TCACCGGTTCTGGCAGAT) and AJ76 (5'-ACCCCGCAGTATAGTGCGCA TA) were used to detect pEA29 as previously described (29). Primers AJ484 (5'-CCTGCTGGCCGCGCCGATCC) and AJ485 (5'-AAAAACCGCAGCAAC CGCATAGTGGT) were used to detect a variable region of pEA29 initially found in *Rubus* strain MR1. Primers AJ486 (5'-AACCGCATTAAAAACGCGATTGC AC) and AJ487 (5'-CTGATCGAACTGTCCGGACAC) were used to detect the *srcA* gene of pEA29. Amplification of the *srcA* gene from the chromo-

somes of *E. coli* and *Erwinia chrysanthemi* was conducted using primers AJ631 (5'-GATCTTCCGATCTGGTGTCCTTCT) and AJ632 (5'-GTTGGTCGTCCTT CGGGTAGTAA) with chromosomal DNA from *E. chrysanthemi* strain 4150 as a positive control. PCRs were carried out in a PTC-150 programmable thermal controller (MJ Research, Inc., Watertown, Mass.). The PCR mixtures (50 \( \mu l \)) consisted of 1× PCR buffer (10 mM Tris-HCl, pH 8.3), 1.5 mM MgCl\textsubscript{2}, 0.1% Tween 20, 200 \( \mu M \) concentrations of each deoxynucleoside triphosphate, 20 \( \mu M \) each primer, 2.5 \( U \) of Taq DNA polymerase (Gibco BRL), and either 0.1 to 100 ng of template DNA or 1 \( \mu l \) of bacterial culture diluted to an optical density at 600 nm between 0.1 and 0.5. Reaction mixtures were covered with 50 \( \mu l \) of light mineral oil prior to placement in the thermal controller. Cycling parameters were

94°C for 4 min followed by 40 cycles of 94°C for 1 min, 52°C (variable region),
5°C (plasmid mseqA), or 48°C (chromosomal mseqA) for 1 min, and 72°C for 2 min, and a final incubation at 72°C for 5 min. Products were analyzed by electrophoresis in 1% (wt/vol) agarose gels in TBE, ethidium bromide for 1 h at 10 V/cm, and visualized by ethidium bromide staining.

**Southern analysis and chromosomal DNA isolation.** Genomic DNAs from *E. amylovora* strains MR1, Ea88, and Ea110 were isolated using a genomic DNA miniprep protocol (56). DNA (20 μg) from each strain was cut with EcoRI, loaded onto a 0.8% (wt/vol) agarose gel in TBE, electrophoresed for 1 h at 10 V/cm, and visualized by ethidium bromide staining. The DNAs were transferred to nylon membranes as previously described (40). A 425-bp probe for the mseqA gene was generated by PCR with primers A636 and A632 as described above, except 1.5 μM of each primer, 0.25 mM dATP, dCTP, dGTP, and dTTP, 135 μM of TTP, and 65 μM digoxigenin-11-dUTP (Roche Molecular Biochemicals) were used. The probe was gel purified using a Quantum Prep Freeze and Squeeze gel extraction kit (Bio-Rad Laboratories). Southern hybridizations were conducted according to the instructions of the manufacturer (Roche Molecular Biochemicals) with standard hybridization buffer at 58°C and detection with CDP-Star.

**Plasmid curing of pEA29.** Plasmid curing of pEA29 was accomplished by electroporating plasmid pC9 (Table 1) into electroporation-competent *E. amylovora* strains and subculturing the resulting ampicillin-resistant transformants several times on LB agar plus ampicillin. Individual colonies were screened for the loss of pEA29 using the PCR primers AJ75 and AJ76 (28) found at nucleotide position 25800 to 25830 and 25989 to 26064, respectively.

**Pathogenicity testing.** Immature Bartlett pears (Pyrus communis) were washed for 20 min in tap water and air dried. The fruits were inoculated with a sterile needle dipped into a single colony of a 48-h-old *E. amylovora* culture grown on LB agar. The pears were incubated in a humidified chamber for 7 days at room temperature (10, 14).

**Oxidative stress disk inhibition assay.** *E. amylovora* strains Ea88, Ea88, MR1, MR1, MR1, 2, 2, 9, 2, 9, RBA4, and RBA4 were grown to an optical density of 0.5 at 600 nm (OD600) in modified Cerevisiae broth cultures at 27°C. Aliquots (250 μl) of culture were added to 2.5 ml of 50°C MM plus Fe salt and poured over the surfaces of MM plus Fe plates. After 2 h, sterile filter disks impregnated with 100 μl of a 400 mM hydrogen peroxide solution were placed in the center of each plate. Plates were incubated at room temperature for 48 h before zones of inhibition were measured.

**Chemotaxis experiments.** Both wild-type and plasmid-cured *E. amylovora* strains Ea88, Ea110, and MR1 were stab inoculated into the surfaces of MM plus 5% plant juice supplemented with 0.1% sucrose, sucinic acid, L-malic acid, fumaric acid, choline, or D-serine at a concentration of 0.1 M except for malic acid, whose concentration was 0.001 M. Media were standardized to a pH of 7.2. Plates were poured to uniform thickness. Zones of migration were measured after 9 days at room temperature.

**Nucleotide sequence accession numbers.** The following nucleotide sequences were entered into GenBank: the complete sequence for plasmid pEA29 from *E. amylovora* strain Ea88 (accession no. AF264948), a substitution sequence found in pEA29 from raspberry strains MR1 (accession no. AF264949) and 2-95 (accession no. AF264950), and insertion element IS12158. Twenty putative open reading frames (ORFs) that were at least 375 nt in length were identified in pEA29 (Table 2, Fig. 1). A single putative 76-amino-acid ORF (ORF 12) was also included because of its association with ORF 13. Amino acid translations for all ORFs were compared to protein entries in GenBank using the BLAST X algorithm (1), and significant matches were found for 15 of these ORFs. Two of the 15 ORFs (15 and 17) were related to hypothetical proteins found in GenBank. The predicted products of ORFs 2, 7, 11, 18, 19, and 20 had no significant matches in the databases (Table 2).

**Results**

**Nucleotide sequence and gene organization.** The complete nucleotide sequence of plasmid pEA29 from strain Ea88 consisted of 28,185 bp and had a G+C content of 50.2%. The first base of the unique BamHI site in pEA29 was designated the first nucleotide position in the sequence. The locations of the BamHI, PstI, SalI, and PvuII restriction sites were in agreement with their locations in the published restriction map of pEA29 from strain Ea7/74 (15); however, an additional PvuII site which generated a 192-bp PvuII fragment was located at nt positions 6033 to 6225. This fragment may have been undetected in the previous restriction analysis due to its small size. Five of the 11 predicted Clai restriction sites (nt positions 7839, 10417, 12158, 13457, and 18064) were refractory to digestion by Clai and overlapped a dam methylase recognition sequence. If methylated, the dam recognition sequence blocks the action of Clai. This indicates that *E. amylovora* possesses a methylase similar to that encoded by the dam gene in *E. coli* (27).

The analysis of the whole nucleotide sequence by the BLAST N algorithm (1) revealed little nucleotide similarity to sequences in GenBank. Two 36-bp direct repeats with associated flanking sequences at nt positions 11374 to 11467 and 15692 to 15870 were the only significant hits across the entire length of pEA29.

The repeat regions were 85% identical to regions in IS911; however, none of the internal structure of IS911 was present in the pEA29 sequence. The repeats were also similar to repeats in p0157, a large plasmid found in *E. coli* 0157 strains (43). As expected, homology was also observed between pEA29 from strain Ea88 to a region of pEA29 previously sequenced from *E. amylovora* strain CA11 (accession no. U19254) and used for detection of *E. amylovora* (29); the sequences from Ea88 and CA11 were 97% identical. Located within this region were five copies of the previously reported 8-bp tandem repeat (nt positions 25765 to 26699) (44).

**Twenty putative open reading frames (ORFs)** that were at least 375 nt in length were identified in pEA29 (Table 2, Fig. 1). A single putative 76-amino-acid ORF (ORF 12) was also included because of its association with ORF 13. Amino acid translations for all ORFs were compared to protein entries in GenBank using the BLAST X algorithm (1), and significant matches were found for 15 of these ORFs. Two of the 15 ORFs (15 and 17) were related to hypothetical proteins found in GenBank. The predicted products of ORFs 2, 7, 11, 18, 19, and 20 had no significant matches in the databases (Table 2). The predicted product of ORF 1 showed 30% identity to a replication protein of a *Pseudomonas aeruginosa* plasmid (54); therefore, ORF 1 was designated repA. The repA gene was preceded by a tandem series of four 22-bp repeats, two 11-bp repeats, and another 22-bp repeat located 15 bp downstream of the direct repeat region (Fig. 2A). Two additional 11-bp repeats (numbers 6 and 7) were found upstream of a 94-bp AT-rich region (72.8% A+T) (Fig. 2). The predicted products of ORFs 12 and 13 resembled the partitioning proteins ParA and ParB, respectively. ParA from pEA29 was grouped in the ParA family by Pfam analysis. No ParB family members were listed in the Pfam database. Positioned in the same orientation on the plasmid, parA and parB were separated by a short segment of sequence containing eight copies of an 8-bp direct repeat (nt 16552 to 16585). The repeat consensus sequence was ATGAGTT(A/T)T.

The predicted product of ORF 3 was most closely related to a histone-like nucleoid structuring protein (H-NS) found in *Proteus mirabilis* and associated with DNA binding and transcriptional regulation. Comparison of the protein to known Pfam family groups also placed the protein in the H-NS histone family. The sequence targeted in pEA29 by the H-NS protein is unknown. The predicted protein product of ORF 4 was a LysR transcriptional regulator and was most closely related to a hypothetical transcriptional regulator found in *E. coli*. Using Pfam analysis, a helix-turn-helix region, characteristic of LysR transcriptional regulators, was found in the N terminus of the predicted protein. The predicted ORF 5 gene product, AldD, is related to aldehyde dehydrogenases by Pfam analysis. The closest GenBank match was a hypothetical aldehyde dehydrogenase found in *E. coli* (7). The product of ORF 6 had significant relatedness to a methyl-accepting chemotaxis protein (MCP) (Tsr) from *E. coli* (8). Tsr mediates chemotactic behavior by sensing environmental concentrations of serine and signaling changes to the flagellar apparatus. The strongest match within the 415-amino-acid alignment was within the linker regions, methylation segments K1 and R1, and signaling domains of Tsr proteins. Low amino acid similarity to existing GenBank entries in the sensing domain of this putative chemotaxis protein suggests that it may be a previously uncharacterized chemotaxis protein within the MCP family. Grouping into the MCP family was confirmed by the alignment of the signaling domain of the protein from pEA29 with other MCP signaling domains found in the Pfam database.

The predicted product of ORF 8 was grouped by Pfam within the Thf-MoeB-HesA family. Thf is required for thiamine
the plasmid next to ORFs 9 and 10, the thiG
have named this gene thiF
also showed similarity. Despite its similarity with moeB
although thiF
in the cyanobacterium
thase sulfurylase, and HesA is involved in heterocyst formation
biosynthesis, MoeB is thought to encode molybdopterin syn-
Fourth gene of the thiGOF gene cluster appears to be a single operon on pEA29.
No ThiG or ThiO family members were listed in the Pfam
gene cluster appears to be a single operon on pEA29.

<table>
<thead>
<tr>
<th>ORF</th>
<th>nt positions (start-stop)*</th>
<th>No. of amino acids</th>
<th>Predicted protein size (kDa)</th>
<th>Proposed gene name</th>
<th>Predicted function(s) of the product/Pfam protein family grouping; homologyb</th>
<th>GenBank accession no. of closest relative</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>13–990</td>
<td>325</td>
<td>37.5</td>
<td>repA</td>
<td>Replication protein; 30% identical (69 of 228 aa*) and 55% similar (125 of 228 aa) to a 277-aa protein in P. aeruginosa</td>
<td>JC2565</td>
</tr>
<tr>
<td>2</td>
<td>1530–1910</td>
<td>126</td>
<td>14.4</td>
<td>ORF 2</td>
<td>Unknown</td>
<td>AAF61240.1</td>
</tr>
<tr>
<td>3</td>
<td>2991–2587</td>
<td>134</td>
<td>15.2</td>
<td>hns</td>
<td>Histone-like nucleoid structuring protein (H-Ns), DNA binding/H-Ns histone family; 42% identical (57 of 133 aa) and 51% similar (70 of 133 aa) to 134-aa H-Ns protein in P. mirabilis</td>
<td>P77309</td>
</tr>
<tr>
<td>4</td>
<td>5489–4632</td>
<td>285</td>
<td>31.3</td>
<td>trlA</td>
<td>LysR-type transcriptional regulator/helix-turn-helix family I, LysR type; 54% (153 of 282 aa) identical and 70% similar (199 of 282 aa) to a 293-aa hypothetical transcriptional regulator in E. coli</td>
<td>P76149</td>
</tr>
<tr>
<td>5</td>
<td>5708–7093</td>
<td>461</td>
<td>49.7</td>
<td>aldD</td>
<td>Aldehyde dehydrogenase/alddehyde dehydrogenase family; 57% identical (262 of 452 aa) and 73% similar (325 of 452 aa) to a 462-aa E. coli aldehyde dehydrogenase gene</td>
<td>CAA23676</td>
</tr>
<tr>
<td>6</td>
<td>8889–7376</td>
<td>507</td>
<td>55.5</td>
<td>ctpE</td>
<td>MCF, attractant unknown/MCF signaling domain family; 43% identical (179 of 415 aa) and 62% similar (259 of 415 aa) to a 536-aa MCP in E. coli</td>
<td>CAA45974</td>
</tr>
<tr>
<td>7</td>
<td>10072–11354</td>
<td>427</td>
<td>47.7</td>
<td>ORF 7</td>
<td>Unknown</td>
<td>AAC72412.1</td>
</tr>
<tr>
<td>8</td>
<td>12513–11548</td>
<td>321</td>
<td>34.2</td>
<td>thiF</td>
<td>Thiamine biosynthesis/ThiF family; 32% identical (78 of 237 aa) and 39% similar (96 of 237 aa) to the 460-aa MoeB protein in H. sapiens; 32% identical (66 of 206 aa) and 48% similar (100 of 206 aa) to the 267-aa ThiF protein in Archaeoglobus fulgidus</td>
<td>AACB0689.1</td>
</tr>
<tr>
<td>9</td>
<td>13287–12529</td>
<td>252</td>
<td>27.2</td>
<td>thiG</td>
<td>Thiamine biosynthesis; 62% identical (157 of 250 aa) and 74% similar (187 of 250 aa) to a 323-aa protein in R. etli</td>
<td>AAC45974</td>
</tr>
<tr>
<td>10</td>
<td>14534–13485</td>
<td>349</td>
<td>38.5</td>
<td>thiO</td>
<td>Amino acid oxidase flavoprotein; 52% identical (158 of 302 aa) and 65% similar (197 of 302 aa) to a 327-aa protein in R. etli</td>
<td>AAC45974</td>
</tr>
<tr>
<td>11</td>
<td>15154–15576</td>
<td>140</td>
<td>15.8</td>
<td>ORF 11</td>
<td>Unknown</td>
<td>AAC45974</td>
</tr>
<tr>
<td>12</td>
<td>16484–16254</td>
<td>76</td>
<td>8.7</td>
<td>parB</td>
<td>Partitioning protein; 42% identical (14 of 33 aa) and 63% similar (21 of 33 aa) to a 73-aa ParB protein in Pseudomonas alcaligenes</td>
<td>AACD40335</td>
</tr>
<tr>
<td>13</td>
<td>17204–16584</td>
<td>206</td>
<td>22.1</td>
<td>parA</td>
<td>Partitioning protein/ParA family; 35% identical (69 of 199 aa) and 57% similar (114 of 199 aa) to a 214-aa protein in Mycobacterium celatum</td>
<td>AAD42964</td>
</tr>
<tr>
<td>14</td>
<td>18201–18778</td>
<td>194</td>
<td>21.4</td>
<td>tnpR</td>
<td>Resolvase/recombinase family; 83% identical (160 of 194 aa) and 92% similar (179 of 194 aa) to a 195-aa protein in E. coli transposon Tn2501</td>
<td>POS523</td>
</tr>
<tr>
<td>15</td>
<td>19259–19774</td>
<td>171</td>
<td>18.9</td>
<td>ORF 15</td>
<td>Unknown; 53% identical (76 of 142 aa) and 71% similar (102 of 143 aa) to a 143-aa hypothetical protein in E. coli</td>
<td>P77460</td>
</tr>
<tr>
<td>16</td>
<td>22094–20064</td>
<td>676</td>
<td>76.0</td>
<td>betT</td>
<td>High-affinity choline transport gene/BCCT transporter family; 78% identical (527 of 672 aa) and 89% similar (596 of 672 aa) to a 676-aa protein in E. coli</td>
<td>P71447</td>
</tr>
<tr>
<td>17</td>
<td>22506–23504</td>
<td>332</td>
<td>37.6</td>
<td>ORF 17</td>
<td>Unknown; 71% identical (220 of 311 aa) and 84% similar (260 of 311 aa) to a 313-aa hypothetical protein in S. enterica serovar Typhimurium</td>
<td>P37415</td>
</tr>
<tr>
<td>18</td>
<td>23620–23168</td>
<td>150</td>
<td>15.9</td>
<td>ORF 18</td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>23888–23490</td>
<td>132</td>
<td>14.9</td>
<td>ORF 19</td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>24343–23960</td>
<td>127</td>
<td>13.8</td>
<td>ORF 20</td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>25065–25601</td>
<td>178</td>
<td>19.6</td>
<td>msrA</td>
<td>Peptide methionine sulfoxide reductase/peptide methionine sulfoxide reductase family; 52% identical (90 of 170 aa) and 66% similar (115 of 170 aa) to a 206-aa protein in D. radiodurans; 47% identical (67 of 142 aa) and 61% similar (86 of 142 aa) to a 212-aa protein in E. coli; 48% identical (68 of 142 aa) and 59% similar (83 of 142 aa) to a 213-aa protein in E. chrysanthemi</td>
<td>AAF11403, AACC77176, CAA10143</td>
</tr>
</tbody>
</table>

* Ascending and descending nucleotide positions indicate that the gene is located on the positive and negative strands, respectively.

b Percent identity and similarities are based upon alignments to regions of genes, not total gene length.

c Although smaller than 375 nt, parB is included because it is often associated with parA.

d aa, amino acids.
isolation. The msrA broth or MM were observed when plasmid-cured strains were grown on LB creamy, mucoid, fluid colonies. No morphological changes is found in Table 2.

...msrA genes revealed several areas of high homology. Primers AJ631 and AJ632 were designed to amplify a 425-bp region from E. chrysanthemi for use as a DNA probe for detecting an msrA gene. Sequence analysis of the 425-bp fragment amplified from E. chrysanthemi confirmed the expected msrA sequence. The msrA chromosomal genes from E. coli and E. chrysanthemi and the msrA gene from pEA29 were detected by Southern analysis using the E. chrysanthemi msrA probe, but no msrA sequences were detected with this probe in E. amylovora chromosomal DNA. An msrA product was amplified from pEA29 with the plasmid-derived primers AJ486 and AJ487 from E. amylovora strains Ea88 and MR1, and msrA genes were amplified from the chromosome-derived primers AJ631 and AJ632. No msrA PCR product was amplified with either primer pair from plasmid-cured strains of E. amylovora. Therefore, the Southern and PCR analyses both indicate that msrA is present in E. amylovora and is not duplicated on the chromosome.

Growth and virulence properties of the plasmid-cured strains. Several experiments confirmed that the plasmid-cured strains Ea88 , Ea110 , and 2-95 exhibited the expected altered growth phenotype on MM - T (16). All plasmid-cured strains exhibited translucent, raised, viscous colonies on MM - T, while the parental wild-type strains produced creamy, mucoid, fluid colonies. No morphological changes were observed when plasmid-cured strains were grown on LB broth or MM + T. Plasmid-cured E. amylovora strain Ea110 exhibited delayed virulence and a reduction in the severity of disease symptoms compared to those characteristics of the parent strain Ea110 on immature Bartlett pears, as previously reported for other cured strains of E. amylovora (16, 24). In addition, immature pears inoculated with strain Ea110 exhibited decreased amounts of viscous ooze compared to levels in pears inoculated with strain Ea110 (Fig. 4). Rubus strains are host specific; therefore, they were not tested on pear. In the oxidase stress assay, comparison of inhibition zone sizes between plasmid-cured and wild-type strains from apple, pear, and raspberry were not statistically different. Both plasmid-cured and wild-type strains of Ea88, Ea110, and MR1 were attracted to L-aspartic acid, succinic acid, L-malic acid, fumaric acid, and l-serine.

Variation in pEA29. Restriction analysis with enzymes PstI and KpnI was used to analyze pEA29 from strains isolated from tree fruit, ornamental, and Rubus hosts, including three strains with the streptomycin resistance transposon Tn5393 (Table 1). Restriction digests of pEA29 from pear strain Ea88 and from strains of E. amylovora from Indian hawthorne and apple were nearly identical. However, digests of pEA29 from Rubus strain MR1 were distinct from those for Ea88 and the other tree fruit strains (Fig. 5A and 6). PstI digests of pEA29 from all raspberry strains yielded a 5-kb fragment, rather than the expected 4.11- and 0.82-kb PstI fragments, and a 13.8-kb PstI fragment, rather than the 10.86- and 3.59-kb fragments found in Ea88 (Fig. 5B). Sequence analysis of the 5-kb fragment from strain MR1 revealed a point mutation at nt position 2645 that eliminated the PstI restriction site. KpnI restriction analysis of pEA29 from each raspberry strain of E. amylovora revealed a fragment with 629 fewer nucleotides than were in the corresponding 11.73-kb fragment found in Ea88 (Fig. 5B and 6, lanes 2 to 4). Sequence analysis revealed that a 1,890-bp region in pEA29 from strain Ea88 (nt positions 9635 to 11525), which contained a PstI site and ORF 7, was missing in pEA29 from strains MR1 and 2-95 (Fig. 5C). A different 1,264-bp sequence was present in the two raspberry strains. This sequence variation accounted for the 13.8-kb PstI fragment and for the KpnI restriction fragment length polymorphism detected in raspberry strain MR1. It was also observed that the first 51 nucleotides in the 1,264-bp region in MR1 were identical to 51 nucleotides on the reverse strand of pEA29 in Ea88. No ORFs
greater than 375 nt were detected in the new sequence, and no homology to existing GenBank entries was detected for several smaller ORFs of 150 to 225 nt (data not shown).

Although pEA29 was approximately 28.2 kb in most strains, it was found to range in size from 27.6 to 34.9 kb due to the presence of deletions and insertions. Because of the deletion described above, pEA29 in strain MR1 was 2.6 or 0.6 kb smaller than pEA29 in strain EA88. Plasmid pEA29 in raspberry strain 2-95 was 28.9 kb. Restriction analysis of this plasmid revealed a 5.4-kb *Kpn*I fragment rather than a 4.1-kb *Kpn*I fragment (Fig. 6). Sequence analysis of the fragment revealed a 1,293-bp insertion sequence, designated IS*Eam1*, at nt position 15692 (Fig. 7). A 100-bp direct repeat was found at the 3' end of IS*Eam1*, and an identical 100-bp direct repeat was found upstream from the insertion site (Fig. 7C). The nucleotide sequence for IS*Eam1* coded for a transposase gene and for an ORF of 94 amino acids (Fig. 7B) with resemblance to an *E. coli* hypothetical protein (accession no. I41306) and to a hypothetical protein found in IS*600* (accession no. P16939). The nucleotide sequence from this ORF exhibited homology to *Yersinia pestis* plasmid pCD1 (accession no. AF053946; 88% identity over 79 bp) and *Y. enterocolitica* plasmid pYVA26 (accession no. U59895; 83% identity over 119 bp). Plasmid pEA29 in the streptomycin-resistant strains DM22, W2, and BCN20 was 34.9 kb. Sequence analysis of strains W2 and DM22 revealed that Tn*5393* had inserted into pEA29 at nt position 17527. The insertion increased the size of the 4.08-kb *Kpn*I fragment to 10.8 kb (Fig. 6, lanes 6 to 7). In strain BCN20, Tn*5393* had inserted in pEA29 at nt position 12360; it disrupted ORF 8, and the 11.7-kb *Kpn*I fragment was expanded to 18.5 kb (Fig. 6, lane 8).

**DISCUSSION**

Plasmid-cured strains of *E. amylovora* have consistently exhibited reduced virulence (references 16 and 24 and this study). Knowledge of the nucleotide sequence of pEA29 has provided insight into the genetic organization of the plasmid. It is possible that several of the genes found in pEA29 interact such that plasmid-free strains are less adaptive for survival in plant tissues. Another possibility is that one or more of the genes identified in pEA29 are virulence determinants that may contribute to the ability of the bacterium to cause fire blight. The peptide methionine sulfoxide reductase (*msrA*) is a viru-
lence determinant in the phytobacterium *E. chrysanthemi* (13). Like *msrA* mutants of *E. chrysanthemi*, *msrA*-minus plasmid-free strains of *E. amylovora* exhibit reduced virulence on pear slices and reduced systemic invasion of apple and pear seedlings (16, 24). However, *msrA*-minus strains of *E. amylovora* did not exhibit increased sensitivity to H$_2$O$_2$ as reported for *msrA* mutants of *E. chrysanthemi* (13) and *E. coli* (33). The putative MCP may enable *E. amylovora* to detect an attractant in the cell’s environment, providing a chemotaxis stimulus as reported for *E. coli* (8), or the choline transport protein (*BetT*) may help protect the pathogen against osmotic stress (22). The *betT* gene is typically associated with the glycine betaine pathway. While choline alone provides no osmoprotective properties (49), glycine betaine, the final product of the pathway, has been shown to protect *E. coli* from osmotic stress (49), and the genes coding for the rest of the glycine betaine pathway may be located on the bacterial chromosome. If present, the pathway could explain the ability of *E. amylovora* to survive the osmotically stressful environment present in the blossom (38), the typical route of colonization in new fire blight infections. Finally, altered exopolysaccharide synthesis, as observed on MM – T, may explain the delay in symptom expression observed with plasmid-free strains (references 24 and 16 and this study). Although we have identified a few genes that may be involved in virulence, a functional analysis of each gene is needed to establish which are virulence factors in *E. amylovora*. Strains of *E. amylovora* cured of pEA29 have been reported to exhibit thiamine auxotrophy (24) and altered exopolysaccharide production on MM – T (16). In the present study we report the detection of a cluster of three putative genes (*thiO, thiG, and thiF*) involved in thiamine biosynthesis. In *E. coli*, *thiFGHS1* and *dss* are necessary for thiazole biosynthesis and *thiE* is necessary for the biosynthesis of thiamine phosphate from thiazole and pyrimidine (4). It was suggested that *thiO* has amino acid oxidase activity important in the biosynthesis of thiazole from cysteine (32), *thiG* is involved in the synthesis of 4-methyl-5-(beta-hydroxyethyl) thiazole, and *thiF* catalyzes both the adenylation of ThiS and the transfer of sulfur from C5-cysteine to an intermediate in thiazole synthesis. In the soil bacterium *R. etli* four thiamine biosynthetic genes (*thiC, thiO, thiG, and thiE*) were clustered in plasmid pb (32), and in *E. coli* a five-gene operon (*thiC, thiE, thiF, thiG, and thiH*) was detected in the chromosome (52). In both *E. coli* and *Salmonella enterica* serovar Typhimurium the genes involved in thiamine biosynthesis, salvage, and transport were scattered throughout the chromosome in three separate operons and four single-gene loci (6, 11, 20, 52, 53). Two thiamine biosynthesis operons and a single gene locus in *B. subtilis* have been characterized (57). The thiazole biosynthesis cluster in *B. subtilis* does not contain *thiH*; instead, *B. subtilis* contains a gene with similarity to the *thiO* gene from *R. etli* (4). To date, 12 genes involved in thiamine biosynthesis in prokaryotes have been identified. Other genes involved in thiamine biosynthesis, including *thiC, thiE*, and *thiS*, may be located on the *E. amylovora* chromosome. It has been suggested that genes for thiamine biosynthesis are located in a 4.4-kb *PstI* fragment of pEA28 in strain CFBP1430 (24). When we examined two CFBP1430 derivatives by PCR, *thiO, thiG, and thiF* were detected in the same locations on the plasmid as in strain Ea88 (data not shown). Since Laurent et al. (24) were able to restore thiamine prototrophy to a Thi$^{-}$ strain of *E. coli* with a clone containing the 4.4-kb *PstI* fragment of pEA28, there is a possibility that pEA29 in strain CFBP1430 contains additional genes for thiamine biosynthesis. We confirmed earlier reports that the replication region for plasmid pEA29 was located on the 4.4-kb *PstI* fragment and that introducing a plasmid carrying this fragment into *E. amylovora* resulted in the loss of pEA29 from *E. amylovora* by incompatibility (references 16 and 24 and this study). Sequence analysis of the replication region of pEA29 revealed a cluster of direct repeats termed iters (12); iters are binding sites for RepA proteins and play a role in the control of plasmid copy number (18, 34). RepA was identified based on similarity to replication proteins from other plasmids and its location.

**FIG. 3.** Restriction and physical maps and sequence data for a region of pEA29 from *E. amylovora* Ea88 with homology to regions in Tn2501 and Tn2502. Triple lines (——) indicate sequenced regions of Tn2501. (A) Physical and genetic map of Tn2501 compared to those of the partial res site and resolvase (tnpR) gene found on pEA29 (nt 18084 to 18785). Positions for the transposase gene (tnpA), res site, and resolvase gene (tnpR) of Tn2501 are indicated by the hatched box, white box, and latticework box, respectively. Gene orientation is indicated by arrows. Restriction sites are as follows: *Bam* III (B), *Bgl* II (Bg), *Hin* III (H), *Sal* I (S), *Sma* I (Sm), and *Sst* I (S). Inverted repeats for Tn2502 are indicated by IR-L (left) and IR-R (right). (B) Sequence alignment of the res regions of Tn2501, Tn2502, and pEA29 (nt 18084 to 18188). Dyad symmetry regions consistent with Tn2501 and Tn2502 res sites I, II, and III are boxed, as is the corresponding pEA29 sequence. Areas where the sequence is in agreement with Tn2501 are indicated by asterisks.
adjacent to the iterons. The configuration of the direct and partial repeats and the AT-rich region is consistent with a theta-type plasmid replication mechanism (12). However, the \( \text{parA} \) and \( \text{parB} \) loci were located some distance from the \( \text{repA} \) locus. Plasmid pEA29 is extremely stable in \( \text{E. amylovora} \), and the \( \text{parA} \) and \( \text{parB} \) genes likely contribute to the ubiquitous occurrence of this plasmid by providing active DNA partitioning functions during cell division. As expected, because pEA29 is nontransmissible, the nucleotide sequence was devoid of genes involved in conjugal transfer and plasmid mobilization. Other putative genes found on pEA29 may involve regulation of gene transcription or substrate utilization. The H-NS DNA binding protein coded for by the \( \text{hns} \) gene has been implicated in several functions. H-NS can act as either a transcription activator or a repressor and may be important for DNA compaction (46, 47) or alteration of DNA topology by generating negative supercoiling (36, 51). The protein preferentially binds to curved strands of DNA and has no known sequence specificity. Among the genes regulated by H-NS proteins are the \( \text{proV} \), \( \text{proW} \), and \( \text{proX} \) osmoregulation genes found in the \( \text{proU} \) operon of \( \text{E. coli} \) (26) and the \( \text{virF} \) virulence gene in \( \text{Shigella} \) spp. and \( \text{E. coli} \) (14). A characteristic shared by genes dependent on regulation by H-NS is their induction by extreme environmental stresses such as osmotic or cold shock. It is unknown at this time which genes are regulated by the

---

**FIG. 4.** Surface and cross section of immature pears stab inoculated with wild-type and plasmid-cured \( \text{E. amylovora} \) strain Ea110 and incubated for 7 days at room temperature under high humidity. (A and C) Pears inoculated with strain Ea110; (B and D) pears inoculated with strain Ea110. Pears inoculated with plasmid-cured strains exhibited less water soaking and ooze production, and symptom expression was delayed and remained localized around the site of inoculation. Pears inoculated with wild-type strains exhibited extensive water soaking and a proliferation of ooze droplets (arrow), and necrosis developed rapidly and extended beyond the inoculation site.

---

**FIG. 5.** (A) Comparison of pEA29 \( \text{PstI} \) and \( \text{KpnI} \) restriction enzyme maps for \( \text{E. amylovora} \) strains Ea88 and MR1 from pear and raspberry, respectively. The maps were normalized to the unique \( \text{BamHI} \) restriction site. (B) Sizes in kilobases of the restriction fragments shown in the restriction maps above. (C) Map of a variable region of pEA29 found in two raspberry strains of \( \text{E. amylovora} \). The locations of the 1,264-bp sequence of pEA29 found in raspberry strains MR1 and 2-95 and the corresponding 1,890-bp sequence found in pear strain Ea88 are shown. Shaded bars indicate that the sequences flanking the variable region are identical in the three strains. \( \text{KpnI} \) sites (K) (nt positions 2011 and 13642) used to confirm the fragment size variation are shown. Dotted arrows indicate the locations of primers AJ484 and AJ485, used to detect length variation in this region. Plasmid pEA29 of strains MR1 and 2-95 is 629 bp smaller than the 28,185-bp plasmid found in Ea88. The substituted sequence starts at bp 9635 and returns to \( \text{Ea88} \) (pEA29) homology at nt 11525. A hypothetical ORF (ORF 7; solid arrow) and the restriction sites for \( \text{PstI} \) (P), \( \text{SalI} \) (S), and methylated \( \text{ClaI} \) (\( \text{Cm} \)) are missing in pEA29 of strains MR1 and 2-95.
H-NS protein in *E. amylovora* or whether the protein acts as an activator or repressor for that gene. A second transcriptional regulator found on pEA29 is the LysR-type transcriptional regulator coded for by *trlA*. LysR-type transcriptional regulators are one of the most common types of autoregulated positive transcriptional regulators and are typically divergently transcribed from a promoter that is very close to and often overlaps the promoter of their target gene (42). The closest divergently transcribed gene relative to *trlA* on pEA29 is the *aldD* gene, which codes for a putative aldehyde dehydrogenase.

In microbes, aldehyde dehydrogenases have narrow inducer and substrate specificities and enhance the flexibility of a microbe’s degradative, fermentative, and biosynthetic pathways. In some instances, aldehyde dehydrogenases may provide a key oxidative link between primary and secondary metabolites. In the glycine betaine osmoregulation pathway, the soluble NAD-dependent betaine aldehyde dehydrogenase found in the pathway converts glycinebetaine aldehyde to glycine betaine in the final step (23). Although not the strongest match to the *aldD* gene in *E. amylovora*, the 490-amino-acid betaine aldehyde dehydrogenase protein found in *E. coli* (accession no. M77739) did share 31% identity (140 of 448 amino acids) and 48% similarity (220 of 448 amino acids) with *aldD*. Further examination of the *aldD* gene may reveal the target substrate of the predicted product. The predicted product of ORF 7 has no matches in GenBank. However, because of the large size of the predicted product, some comment is warranted. ORF 7 was found in all apple, pear, and Indian hawthorn strains examined.
in this study, but it was not detected in any of the Rubus strains. It is possible that the predicted product of ORF 7 is necessary for infection of tree fruit and ornamental hosts. This might explain why Rubus strains are not able to infect pears in pathogenicity testing. Alternatively, it is possible that the product of ORF 7 is not conserved in Rubus strains because it has a minor function. Further study is needed to elucidate the function of the hypothetical protein.

A link was observed between the organization of pEA29 and the strain’s original host plant. Derivatives of pEA29 found in strains isolated from apple, pear, and Indian hawthorn were homogenous; however, significant sequence variation was found to exist between strains from plants in the subfamily Pomoideae and those from Rubus plants. Derivatives of pEA29 found in strains isolated from Rubus exhibited restriction fragment length polymorphism when they were digested with KpnI and PsI. All Rubus derivatives lacked two PsI sites and were missing 1,890 bp of sequence found in tree fruit strains. Rubus strains contained an entirely different 1,264-bp region. Since the Rubus strains that we examined came from different geographic regions, Michigan in the United States and Nova Scotia in Canada, this variation in sequence may be a common feature of pEA29 found in strains of E. amylovora from Rubus.

The detection of transposon Tn5393 on pEA29 is interesting with respect to circulation of antibiotic resistance genes in the environment. Tn5393 is found on the conjugal plasmid pEA34 in E. amylovora and Pantoaea agglomerans and carries the strA and strB genes for streptomycin resistance (10, 19). The detection of Tn5393 on pEA29 was a recent event (19, 28) and represents further evolution of streptomycin resistance in E. amylovora in Michigan apple orchards. The insertion of Tn5393 into different sites on pEA29 in field strains indicates that resistance in these strains involved separate insertion events. Since pEA29 is nonpermeable, a strain containing pEA29:Tn5393 would not be an important source for further dissemination of the strA-strB gene throughout microbial communities.

There were remnants of several insertion sequences detected on pEA29 during sequencing that resembled insertion elements identified previously in other unrelated bacteria. A remnant of Tn2501 and direct repeats found in IS911 were detected in all derivatives of pEA29. Residual traces of several genes, including a remnant from a parA gene located downstream from the repeats located in the 1.1-kb PsI fragment, were found. Kim and Geider (21) reported finding a 108-amino-acid ORF with similarity to ParA from Agrobacterium tumefaciens at this location on the plasmid. We located a 65-amino-acid ORF at bp 26,241 to 26,438 exhibiting the highest similarity to the Mycobacterium celatum parA gene (accession no. AAD42964.1) and had no BLAST hit to the 222-amino-acid ParA protein from Agrobacterium tumefaciens (accession no. S07280). The presence of a partial parA gene may explain the occurrence of the 5-bp repeats found consistently in this region. Remnants of a single gene which once coded for an outer membrane protein belonging to the peptidase family S18 (ompC family) were identified in gaps between some of the large ORFs on pEA29 (data not shown). The closest GenBank match by BLAST X homology was the plasmigen activator gene in Yersinia pestis (accession no. P17811) coding for a coagulase or fibrinolysin precursor. These remnants may provide important clues to the evolution of pEA29 and E. amylovora. 15Eam1 was found adjacent to an IS911 direct repeat in pEA29. This region of the plasmid may be highly variable because it is an attractive site for insertion. Very few insertion sequence-like elements have been detected in E. amylovora previously, and they may be important for tracing local outbreaks of disease.

ACKNOWLEDGMENTS

This research was supported in part by the Michigan Agricultural Experiment Station and USDA/CSREES grant number 97-34367-3967.

We thank E. L. Schnabel for helpful discussions and for reviewing the manuscript and R. Roberts, G. Braun, and G. E. Halcomb for providing strains of E. amylovora.

REFERENCES


