Erwinia amylovora CRISPR Elements Provide New Tools for Evaluating Strain Diversity and for Microbial Source Tracking

Gayle C. McGhee, George W. Sundin*

Department of Plant Pathology and Centers for Microbial Ecology and Pathogenesis, Michigan State University, East Lansing, Michigan, United States of America

Abstract

Clustered regularly interspaced short palindromic repeats (CRISPRs) comprise a family of short DNA repeat sequences that are separated by non repetitive spacer sequences and, in combination with a suite of Cas proteins, are thought to function as an adaptive immune system against invading DNA. The number of CRISPR arrays in a bacterial chromosome is variable, and the content of each array can differ in both repeat number and in the presence or absence of specific spacers. We utilized a comparative sequence analysis of CRISPR arrays of the plant pathogen Erwinia amylovora to uncover previously unknown genetic diversity in this species. A total of 85 E. amylovora strains varying in geographic isolation (North America, Europe, New Zealand, and the Middle East), host range, plasmid content, and streptomycin sensitivity/resistance were evaluated for CRISPR array number and spacer variability. From these strains, 588 unique spacers were identified in the three CRISPR arrays present in E. amylovora, and these arrays could be categorized into 20, 17, and 2 patterns types, respectively. Analysis of the relatedness of spacer content differentiated most apple and pear strains isolated in the eastern U.S. from other continents. E. amylovora strains from Rubus and Indian hawthorn contained mostly unique spacers compared to apple and pear strains, while strains from loquat shared 79% of spacers with apple and pear strains. Approximately 23% of the spacers matched known sequences, with 16% targeting plasmids and 5% targeting bacteriophage. The plasmid pEU30, isolated in E. amylovora strains from the western U.S., was targeted by 55 spacers. Lastly, we used spacer patterns and content to determine that streptomycin-resistant strains of E. amylovora from Michigan were low in diversity and matched corresponding streptomycin-sensitive strains from the background population.

Introduction

The enterobacterial pathogen Erwinia amylovora is the causal agent of fire blight, a devastating disease of apple, pear, and other plants in the Rosaceae family. The bacterium initiates infection of flowers following colonization of intercellular spaces on stigmas and growth to relatively large population sizes (ca. 10^7 to 10^8 cfu/flower). Following flower infection, E. amylovora invades the plant and migrates internally producing biofilms within the plant vascular system, and can also emerge as bacterial ooze (cells embedded in exopoly saccharide) that can be transmitted to and infect other nearby plants [1,2]. Management of fire blight is difficult, and is exacerbated by the development of streptomycin resistance in E. amylovora populations [3].

The E. amylovora species exhibits very low levels of genetic diversity. For example, comparison of the complete genome sequences of two E. amylovora strains isolated from apple and pear on different continents revealed 99.99% sequence identity [4,5]. In addition, other typing methods, based on pulsed field gel electrophoresis, REP-PCR fingerprinting, ribotyping, and variable number of tandem repeat analyses revealed the difficulties in accurately and effectively differentiating strains [6,7,8,9,10,11]. Only E. amylovora strains isolated from Rubus spp. are sufficiently diverged from the apple and pear genotypes to enable easy molecular differentiation [12,13]. E. amylovora is also closely related to other tree fruit pathogens E. pyrifoliae and Erwinia sp. isolated from pear in Asia and to the nonpathogenic species E. billingiae and E. tasmaniensis [14,15,16].

Information on the presence and characterization of laterally-acquired DNA sequences and indigenous plasmids in E. amylovora has been reported in a few studies. Almost all E. amylovora strains contain a plasmid of approximately 29 kb termed the ubiquitous plasmid pEA29 [17,18]. This nonconjugative plasmid plays a role in virulence linked to the carriage of thiamin-biosynthetic genes [19]. In addition to pEA29, variation in plasmid profiles or lack of pEA29 has been used to differentiate strains, although the contribution of other plasmids detected in E. amylovora to virulence has not been demonstrated [20,21,22,23], except for the recently described pEI70 which affects aggressiveness in an immature pear infection model [24]. However, the presence of different plasmid patterns in various strains indicates that the E. amylovora species has been subject to plasmid invasion and colonization during its life.
| Strain | Geographic location        | Host       | Year | Known plasmids | Streptomycin Sensitivity/ location of strAB genes | Ribotype | groEL seq type |
|-------|---------------------------|------------|------|----------------|--------------------------------------------------|----------|---------------|
| BH    | Hart, MI                  | Apple      | 2008 | pEA29          | Sensitive                                        |          | 1             |
| DP11  | Michigan                  | Pear       | 1993 | pEA29          | Sensitive                                        | 1        | 1             |
| DR5   | Michigan                  | Apple      | 1993 | pEA29          | Sensitive                                        | 1        | 1             |
| Ea(T1)2 | Michigan                | Apple      | 1997 | pEA29          | Sensitive                                        |          | 1             |
| Ea(T3)2 | Michigan                | Apple      | 1997 | pEA29          | Sensitive                                        |          | 1             |
| Ea110 | Ingham County, MI        | Apple      | 1975 | pEA29          | Sensitive                                        |          | 1             |
| EL01  | East Lansing, MI         | Apple      | 1993 | pEA29          | Sensitive                                        | 1        | 1             |
| GH9   | Michigan                  | Apple      | 1993 | pEA29          | Sensitive                                        | 1        | 1             |
| K2    | Michigan                  | Apple      | 1993 | pEA29          | Sensitive                                        | 1        | 1             |
| L14   | Michigan                  | Apple      | 1993 | pEA29          | Sensitive                                        | 1        | 1             |
| MK1   | Michigan                  | Apple      | 1993 | pEA29          | Sensitive                                        | 1        | 1             |
| NW17-4 | Northwest MI         | Apple      | 2011 | pEA29          | Sensitive                                        |          | 1             |
| Pn    | Hart, MI                  | Apple      | 2008 | pEA29          | Sensitive                                        |          | 1             |
| RB02  | Michigan                  | Apple      | 1993 | pEA29          | Sensitive                                        | 3        | 1             |
| RB07  | Michigan                  | Apple      | 1993 | pEA29          | Sensitive                                        | 1        | 1             |
| RL3   | Michigan                  | Apple      | 1993 | pEA29          | Sensitive                                        | 1        | 1             |
| RRP12 | Michigan                  | Apple      | 1993 | pEA29          | Sensitive                                        | 1        | 1             |
| BBA-8 | Southwest MI              | Apple      | 2007 | pEA29          | MR, pEA29 bp 17527                                 |          | 1             |
| BCN20 | Southwest MI              | Apple      | 1995 | pEA29          | MR, pEA29 bp 12360                                 |          | 1             |
| CA11  | Southwest MI              | Crab apple | 1993 | pEA29, pEA34   | MR, pEA34                                        |          | 1             |
| DM1   | Southwest MI              | Apple      | 1994 | pEA29          | MR, pEA29 bp 17527                                 |          | 1             |
| EaRoo29 | Southwest MI         | Apple      | 1997 | pEA29          | MR, pEA29 bp 17527 and pEA34                       |          | 1             |
| GR5B1 | Grand Rapids, MI         | Apple      | 2004 | pEA29, pEA34   | MR, pEA29 bp 1515 and pEA34                       |          | 1             |
| HS10  | Southwest MI              | Apple      | 1993 | pEA29, pEA34   | MR, pEA34                                        |          | 1             |
| KL    | Ionia County, MI          | Apple      | 2008 | pEA29          | MR, pEA29 bp 1515                                 |          | 1             |
| KR    | Ionia County, MI          | Apple      | 2007 | pEA29          | MR, pEA29 bp 1515                                 |          | 1             |
| MA-1  | Southwest MI              | Apple      | 2007 | pEA29          | MR, pEA29 bp 1515                                 |          | 1             |
| MC-5  | Southwest MI              | Apple      | 2007 | pEA29          | MR, pEA29 bp 1515                                 |          | 1             |
| M15-1 | Southwest MI              | Apple      | 2002 | pEA29          | MR, pEA29 bp 1515                                 |          | 1             |
| NW H26 | Northwest MI            | Apple      | 2011 | pEA29          | MR, pEA29 bp 1515                                 |          | 1             |
| NW1-1 | Northwest MI              | Apple      | 2011 | pEA29          | MR, pEA29 bp 17527                                 |          | 1             |
| NW18-6 | Northwest MI           | Apple      | 2011 | pEA29          | MR, pEA29 bp 1515                                 |          | 1             |
| NW2-11 | Northwest MI           | Pear       | 2011 | pEA29          | MR, pEA29 bp 1515                                 |          | 1             |
| NW21-4 | Northwest MI           | Pear       | 2011 | pEA29          | MR, pEA29 bp 1515                                 |          | 1             |
| NW2A  | Northwest MI              | Pear       | 2011 | pEA29          | MR, pEA29 bp 1515                                 |          | 1             |
| NW3-1 | Northwest MI              | Apple      | 2011 | pEA29          | MR, pEA29 bp 17527                                 |          | 1             |
| RA    | Grand Rapids, MI          | Apple      | 2005 | pEA29          | MR, pEA29 bp 1515                                 |          | 1             |
| RM5   | Southwest MI              | Apple      | 2007 | pEA29          | MR, pEA29 bp 17527                                 |          | 1             |
| S5    | Southwest MI              | Apple      | 1994 | pEA29          | HR, mutation rpsL                                   |          | 1             |
| SB1-9 | Southwest MI              | Apple      | 2003 | pEA29          | MR, pEA29 bp 17527                                 |          | 1             |
| W4    | Southwest MI              | Apple      | 2007 | pEA29          | MR, pEA29 bp 17527                                 |          | 1             |
| Ea273 | New York                  | Apple      | 1971 | pEA29          | Sensitive                                         |          | 1             |
| 6–97  | Canada                    | Apple      | 2000 | pEA29          | Sensitive                                         |          | 1             |

**Western USA**

| 87–70 | Washington State          | Apple      | pEA29 | Sensitive   | 1        | 1             |
| 87–73 | Washington State          | Apple      | pEA29 | Sensitive   | 1        | 1             |
Table 1. Cont.

| Strain | Geographic location | Host | Year | Known plasmids | Streptomycin Sensitivity/ location of strAB genes | Ribotype | groEL seq type
|--------|---------------------|------|------|----------------|-----------------------------------------------|----------|-----------------
| Ca1R   | California          | Apple| 1995 | pEA29         | Sensitive                                    | 3        | 2               |
| Ca3R   | California          | Apple| 1995 | pEA29         | MR                                           | 1        | 1               |
| Ea88   | Washington State    | Pear | 1995 | pEA29         | HR, mutation in rpsL                         | 3        | 2               |
| FB93-9 | Idaho               | Apple| 1998 | pEA29         | Sensitive                                    | 1        | 1               |
| JL1189 | Washington          | Pear | 1988 | pEA29, pEU30  | HR, mutation in rpsL                         | 3        | 2               |
| LA004  | Washington          | Pear | 1995 | pEA29, pEU30  | HR, mutation in rpsL                         | 3        | 2               |
| La092  | Washington          | Pear | 1988 | pEA29, pEU30  | HR, mutation in rpsL                         | 3        | 2               |
| LP101  | Washington          | Apple| 1995 | pEA29         | Sensitive                                    | 3        | 2               |
| OR1    | Oregon              | Pear | 1995 | pEA29         | HR, mutation in rpsL                         | 3        | 2               |
| OR6    | Oregon              | Pear | 1995 | pEA29         | HR, mutation in rpsL                         | 3        | 2               |
| UT5P4  | Utah                | Apple| 2000 | pEA29, pEU30  | Sensitive                                    | 1        |                 |
| UTFer3 | Utah                | Pear | 2000 | pEA29         | HR, mutation in rpsL                         | 1        |                 |
| UTRU2  | Utah                | Pear | 2000 | pEA29, pEU30  | HR, mutation in rpsL                         | 1        |                 |
| WSDA 16| Washington          | Apple| 1995 | pEA29         | Sensitive                                    | 1        |                 |

**Europe, Middle East and New Zealand**

| Strain | Geographic location | Host | Year | Known plasmids | Streptomycin Sensitivity/ location of rpsL | Ribotype | groEL seq type
|--------|---------------------|------|------|----------------|-------------------------------------------|----------|-----------------
| 1596   | Spain               | unknown| 1995 | pEA70 only     | Sensitive                                  | 1        |                 |
| B3     | Serb/Montenegro     | unknown| 1995 | No plasmids    | Sensitive                                  | 1        |                 |
| CFBP1430| France             | Hawthorne| 1995 | pEA29         | Sensitive                                  | 1        |                 |
| Ea1189 | Germany             | unknown| 1995 | pEA29         | Sensitive                                  | 1        |                 |
| Ea322  | France              | Pear  | 1995 | pEA29, pCP60   | Sensitive                                  | 1        |                 |
| Leb A-1| Lebanon             | Pear  | 1995 | pEA29, pEL60   | HR, mutation in rpsL                       | 1        |                 |
| Leb A-16| Lebanon            | Pear  | 1995 | pEA29, pEL60   | HR, mutation in rpsL                       | 1        |                 |
| Leb A-19| Lebanon            | Pear  | 1995 | pEA29         | Sensitive                                  | 1        |                 |
| Leb A-3| Lebanon             | Quince| 1995 | pEA29, pEL60   | Sensitive                                  | 1        |                 |
| Leb B-66| Lebanon            | Apple | 1995 | pEA29, pEL60   | HR, mutation in rpsL                       | 1        |                 |
| NZR3   | New Zealand         | unknown| 1995 | pEA29         | HR, mutation in rpsL                       | 1        |                 |
| NZR5   | New Zealand         | unknown| 1995 | pEA29         | HR, mutation in rpsL                       | 1        |                 |
| NZS24  | New Zealand         | unknown| 1995 | pEA29         | Sensitive                                  | 1        |                 |
| OT-1   | England             | unknown| 1995 | pEA29         | Sensitive                                  | 1        |                 |

**E. amylovora from Rubus USA**

| Strain | Geographic location | Host | Year | Known plasmids | Streptomycin Sensitivity/ location of rpsL | Ribotype | groEL seq type
|--------|---------------------|------|------|----------------|-------------------------------------------|----------|-----------------
| IL5    | Illinois            | Raspberry| 1977 | pEA29, small plasmids | Sensitive | 4 | 1 |
| IL6    | Illinois            | Raspberry| 1977 | pEA29, small plasmids | Sensitive | 4 | 1 |
| MR1    | Alcona Co., MI      | Raspberry| 1995 | pEA29         | Sensitive                                  | 4        | 3 |
| OKR1   | Oklahoma            | Raspberry| 2002 | pEA29         | Sensitive                                  | 1        |       |
| RBA4   | Alpena Co., MI      | Apple  | 1995 | pEA29         | Sensitive                                  | 4        | 3 |
| Rkk3   | Michigan            | Raspberry| pEA29 | Sensitive | 4 | 3 |

**E. amylovora from Indian Hawthorn**

| Strain | Geographic location | Host   | Year | Known plasmids | Streptomycin Sensitivity/ location of rpsL | Ribotype | groEL seq type
|--------|---------------------|--------|------|----------------|-------------------------------------------|----------|-----------------
| IH2-3  | South Carolina      | Indian Hawthorn| 1998 | pEA29, small plasmids | Sensitive | 1 |
| IH3-1  | South Carolina      | Indian Hawthorn| 1998 | pEA29, small plasmids | Sensitive | 1 |

**E. amylovora from Loquat**

| Strain | Geographic location | Host | Year | Known plasmids | Streptomycin Sensitivity/ location of rpsL | Ribotype | groEL seq type
|--------|---------------------|------|------|----------------|-------------------------------------------|----------|-----------------
| TxLo3  | Texas               | Loquat| 2011 | pEA29         | Sensitive                                  | 2        |       |
| TxLo4  | Texas               | Loquat| 2011 | pEA29         | Sensitive                                  | 2        |       |
| TxLo6  | Texas               | Loquat| 2011 | pEA29         | Sensitive                                  | 2        |       |
| TxLo7  | Texas               | Loquat| 2011 | pEA29         | Sensitive                                  | 2        |       |

**Other Erwinias**

| Strain | Geographic location | Host   | Year | Known plasmids | Streptomycin Sensitivity/ location of rpsL | Ribotype | groEL seq type
|--------|---------------------|--------|------|----------------|-------------------------------------------|----------|-----------------
| E. pyrifolia Ep1/96| Korea | Pyrus pyrifolia| 1996 | pEP36, small plasmids | Sensitive | 5 |
| E. pyrifolia Ep4/97| Korea | Pyrus pyrifolia| 1997 | pEP36, small plasmids | Sensitive | 5 |
| E. pyrifolia Ep16/96| Korea | Pyrus pyrifolia| 1996 | pEP36, small plasmids | Sensitive | 5 |
| Erwinia spp. Ejp617| Japan  | Nashi pear | 1996 | pEJ30, small plasmids | Sensitive | 4 |
history. Streptomycin-resistant (SmR) strains of *E. amylovora* isolated in Michigan harbor the transposon Tn5393 that encodes the streptomycin-resistance genes *strA-strB* [25,26]. This transposon was thought to be obtained by *E. amylovora* from the orchard epiphyte *Pantoea agglomerans* on the plasmid pEA34 [25]. A subsequent genetic analysis recently demonstrated that only very few strain types are responsible for the dissemination of streptomycin resistance in Michigan, and that Tn5393 had moved to pEA29 in these strains [27]. This observation suggests that gene transfer events resulting in the acquisition of resistance genes may be relatively rare in the *E. amylovora* population. Finally, there have been a few reports including some sequence availability of bacteriophage that can infect *E. amylovora* [28,29], and experiments by Schnabel and Jones [30] established differential sensitivity of *E. amylovora* isolates from Michigan to a panel of five bacteriophage.

*E. amylovora* is thought to have originated in North America, and was first observed on apple in New York in the 1700s following the introduction of apple to the continent by European settlers [31]. Fire blight has since spread to apple and pear in over 46 countries. Epidemiology and strain tracking provides critical to topics such as determining outbreak centers for disease Fire blight has since spread to apple and pear in over 46 countries.

### Table 1. Cont.

| Strain               | Geographic location | Host          | Year | Known plasmids            | Streptomycin Sensitivity/ location of *strAB* genes | Ribotype† | groEL seq type‡ |
|---------------------|---------------------|---------------|------|---------------------------|---------------------------------------------------|----------|-----------------|
| *Erwinia* spp. Ejp556 | Japan               | Nashi pear, 1994 | pE30, small plasmids
| *Erwinia* spp. Ejp557 | Japan               | Nashi pear, 1994 | pE30, small plasmids

*S* = sensitive, minimum inhibitory concentration (MIC) < 100 μg ml⁻¹, *MR* = medium resistance MIC between 100 μg ml⁻¹ and 1000 μg ml⁻¹, and *HR* = MIC > 2000 μg ml⁻¹. Streptomycin phenotypes and MICs for some strains have been reported previously.

Transposon Tn5393 harboring *strA* and *strB* resistance genes present on plasmid pEA29 at bp 1515 (29–1), bp 17527 (29–2) or bp 12360 (29–3), or on conjugal plasmid pEA34

Ribotyping previously reported in [8].

Four hundred and nine bp of the groEL gene was amplified and sequenced using primers groEL-A and groEL-B. Pattern 1 is the predominant pattern and is found in sequenced strains ATCC 49942 and CFBP1430. Pattern 2 has 2 nucleotide changes from a C to a T at bp 329 and 335; Pattern 3 has a single bp change of T to C at bp 299; Pattern 4 has 18 bp changes at positions 59, 53, 134, 140, 146, 152, 155, 215, 228, 281, 299, 302, 314, 341, 374, 413, 422, and 437. Pattern 5 is similar to pattern 4 except that there is an additional bp change at position 203.

*TS* = This study.

doi:10.1371/journal.pone.0041706.t001

---

Materials and Methods

Bacterial Strains

The bacterial strains characterized for CRISPR array number and content for this study are listed in Table 1. Many of these strains had been characterized previously by ribotyping [8], and the plasmid content of each strain is known. In addition, the SmR phenotype, genetic mechanism of streptomycin resistance, and location of Tn5393, if present, is summarized in Table 1. All strains were stored in 15% glycerol at −80°C prior to use. *Erwinia amylovora* and other *Erwinia* spp. strains were maintained on LB agar and cultured at 28°C. Where necessary, streptomycin (100 or 2,000 μg ml⁻¹) was added to LB for resistance screening.

DNA Extraction, Amplification, and Sequencing

Genomic DNA preps were prepared for each strain by suspending single colonies in 100 μl of a lysis buffer (0.5 M KCl, 0.01 M Tris-HCL [pH 8.5], 1% Tween 20) and boiling for 10 min. The lysate was used as a template in PCR reactions. All PCR reactions (50 μl) contained 1× PCR buffer (10 mM Tris-
HCl at pH 8.3), 1.5 mM MgCl₂, 0.2 mM of each deoxynucleotide triphosphate, 20 pM of each primer, 2.5 U of Taq DNA polymerase (Invitrogen; Carlsbad, CA), and 3 μl of bacterial lysate.

Oligonucleotide primers used to amplify the CRISPR 1, 2, and 3 array sequences were designed using the E. amylovora Ea273 (ATCC 49946) genome sequence (GenBank accession number NC_013971). Primer sequences are listed in Table S1. The cycling parameters were 94°C for 5 min followed by 40 cycles of 94°C for 30 seconds, annealing temperatures of 58°C (CRISPR 1 and 2) or 55°C (CRISPR 3) for 30 seconds, followed by 72°C for 4 min (CRISPR 1 and 2) or 45 seconds (CRISPR 3) with a final extension time for 7 min at 72°C. Amplification parameters for E. pyrifoliae CRISPRs 1–4 were identical to those listed above for CRISPRs 1 and 2 from isolates of E. amylovora.

Amplification and sequencing of a partial region of groEL, a housekeeping gene coding for Hsp60 commonly used to compare bacterial strains, was also reported in this study. The groEL gene...
has been previously used to compare *E. amylovora* both intra-
species and inter-species to closely related *Erwinia* species. 
Amplification of the partial *groEL* gene from the *E. amylovora*
chromosome was performed as previously described using primers 
groEL-A and groEL-B [12,22,27].

Amplified PCR products were purified using the QIAquick PCR Purification Kit (QIAGEN, Valencia, CA) and sequenced. All
sequencing was performed at the Michigan State University 
Research Technology Support Facility using ABI dye-terminator
chemistry and ABI 3730 genetic analyzer (Applied Biosystems, 
Foster City, CA). Where necessary, primer walking was used to 
complete sequencing.

**CRISPR Array Analysis and Alignment**

Sequences were assembled using the DNASTAR Lasergene 
Software Suite for Sequence Analysis Version 7.2.0 (DNASTAR, 
Inc.; Madison, WI). CRISPR array spacer and repeat patterns 
were generated using the CRISPR recognition tool (CRT) Version 
1.0 (http://www.room220.com/crt) [37]. The repeat sequences 
within the CRISPRs 1, 2, and 3 were assessed for existing 
homology to known sequences using BLAST searches of the 
GenBank database (http://blast.ncbi.nlm.nih.gov/Blast.cgi). 
Alignments of repeat sequences to those present in other species 
were done using WebLogo (http://weblogo.berkeley.edu/logo. 
cgi). Each spacer in the array was also assessed for homology to 
known sequences using BLASTn. All spacers were compared 
among strains to identify unique spacers and to identify conserved 
patterns of spacers among strains.

Cluster analysis was accomplished by the generation of an 
unweighted-pair group method (UPGMA)-based tree including 
all of the *E. amylovora* strains analyzed in this study. CRISPR spacers 
from the CRISPR 1, 2, and 3 arrays were concatenated and 
converted to a binary matrix based on presence or absence of a 
particular spacer sequence. The distance matrix was calculated 
using the Jaccard coefficient (http://genomes.urv.es/UPGMA/) 
with 1,000 bootstrap replications.

**Use of CRISPR Spacer Pattern in Strain Tracking**

CRISPR spacer patterns were utilized to determine the identity 
of genotypes of *E. amylovora* in Michigan that were sensitive to 
streptomycin or that were identical to the three genotypes that 
had acquired the streptomycin resistance determinant Tn5393 that 
carries the streptomycin resistance genes *strA* and *strB*. Similarly, 
we also examined the CRISPR genotype of *E. amylovora* S5, a
natural spontaneous chromosomal Sn<sup>R</sup> strain also isolated in 
Michigan. A total of 17 Sn<sup>R</sup> *E. amylovora* strains (isolated between 
1975 and 2011) and 24 Sm<sup>R</sup> (Tn5393) *E. amylovora* strains (isolated 
between 1993 and 2011) were selected for this analysis.

**GenBank Accession Numbers**

CRISPR array sequences were submitted to the NCBI database 
and assigned an accession number. A list of accession numbers by 
strain and array number (CR1, CR2, or CR3) is available in Table 
S2.

**Results**

**Global Genetic Organization of CRISPR Loci**

Eighty-five strains of *Erwinia amylovora* varying in geographic 
isoation, plant host of isolation, plasmid content, and streptomycin 
sensitivity were evaluated for CRISPR array number and 
spacer variability (Table 1). Representative strains from Michigan, 
the eastern and western United States, Canada, Europe, the 
Middle East, and New Zealand were included in the study. 
Comparison of a 409-bp fragment of the *groEL* genes among 
strains revealed little polymorphism except for between strains 
from *Rubus* and the rest of the strains (Table 1).

We identified three arrays of spacer sequences associated with 
CRISPRs (CRISPR 1-3) present in *E. amylovora*. After our studies 
were initiated, genome sequences of two *E. amylovora* strains were 
completed, confirming the presence of three CRISPR spacer 
arrays in both sequenced strains [4,5]. Arrays 1-3 are clustered 
together between nt 854678 to 879523 on the *E. amylovora* 
chromosome (nt position is relative to GenBank accession number
NC_013971). The eight *cse* and *cas* genes are located between 
CRISPR spacer arrays 1 and 2 (Fig. 1) in an orientation that is 
termed the *E. coli*-type and is conserved in *Escherichia coli* [37,38]. 
The 2.56-kb region between the 3′ end of the CRISPR array 1 and 
the *cas3* gene and the 9.46-kb region between CRISPR arrays 
2 and 3 encoded *E. amylovora* housekeeping genes (Fig. 1).

The genetic organization of the *cse* and *cas* genes found in the 
necrogenic *Erwinia* spp., i.e. *E. amylovora*, *E. pyrifoliae*, and *Erwinia* 
sp. from Japan, is identical and the translated Cse and Cas 
proteins share high levels of amino acid identity (95% and 92% 
identities of Cas1 from Epj617 and Epj1/96, respectively; 89% 
identities of Cas3 from Epj617 and Epj1/96; 92% identities of 
Cse1 from Epj617 and Epj1/96). In contrast, the gene organization in *E. tasmaniensis*, a related non-pathogenic bacte-

**Table 2. Number of spacers located in CRISPR arrays 1 and 2 from *E. amylovora* strains isolated from apple and pear, *Rubus*, Indian Hawthorn, and loquat.**

|                  | Total spacers* | Minimum/Maximum spacer no. | Avg. no. of spacers ± SD |
|------------------|----------------|---------------------------|--------------------------|
|                  | C1  | C2  | C1  | C2  | C1  | C2  |
| *E. amylovora* (85) |     |     |     |     |     |     |
| Midwest, Eastern US and Canada (43) | 36  | 34  | 27–36 | 23–34 | 34.7±3 | 26.3±4 |
| Western US (16) | 130 | 62 | 12–96 | 32–49 | 58.5±39 | 42.5±8 |
| Europe and Middle East (11) | 36  | 34  | 35–36 | 34  | 35.7±1 | 34±0 |
| New Zealand (3) | 35  | 34  | 35  | 34  | 35±0  | 34±0  |
| Ep Rubus (6) | 195 | 138 | 33–59 | 32–42 | 51.1±9 | 36.3±0 |
| Ep Indian hawthorn (2) | 29  | 42  | 29  | 42  | 29±0  | 42±0  |
| Ep Loquat (4) | 84  | 30  | 84  | 30  | 84±0  | 30±0  |

*The total count includes repeated spacers, if any.
doi:10.1371/journal.pone.0041706.t002
CRISPR Elements of *Erwinia amylovora*

- **Pattern 1**
- **Pattern 2**
- **Pattern 3**
- **Pattern 4**
- **Pattern 5**
- **Pattern 6**
- **Pattern 7**
- **Pattern 8**
- **Pattern 9**
- **Pattern 10**
- **Pattern 11**
- **Pattern 12**
- **Pattern 13**
- **Pattern 14**
- **Pattern 15**
- **Pattern 16**
- **Pattern 17**
- **Pattern 18**
- **Pattern 19**
- **Pattern 20**

Apple/pear/quince:

- **ATCC 49946**
- **CFBP1430**

Loquat:

- **Pattern 11**

Hazelnut:

- **Pattern 16**
- **Pattern 17**

Indian Hawthorn:

- **Pattern 10**
**Figure 3.** Graphic representation of spacers grouped into patterns from CRISPR array CR1 of the 85 *E. amylovora* strains examined in this study. Hosts of *E. amylovora* strains harboring the respective spacer patterns shown are listed on the left. Individual spacer sequences are represented by boxes; spacers were considered unique if they contained >5 nucleotide differences compared to other spacer sequences. Each unique spacer (588 identified in this study) was assigned a number. Spacers of similar color shown in the same columns under the same number are identical. Empty areas indicate the corresponding spacer in other similar patterns is not present. Open boxes refer to spacers that were only detected once in the collection. The **“** symbol refers to a 31-bp insertion sequence identified in 11 western U.S. strains from apple and pear and 4 loquat strains. The **“–”** symbol refers to the location of a 10-bp insert (5’–gttgtgttgtgtg–3’) observed in *E. amylovora* strains isolated from *Rubus*. Spacers shaded grey and boxed in red are spacers found in both CR1 and CR2 arrays, while spacers shared between *Rubus* strains are shaded in yellow or in dark blue. The following spacer patterns were predominantly isolated from *E. amylovora* strains from apple and pear isolated in the Midwest, eastern U.S., and Canada (CR1: patterns 1–4 and 9) and in the western U.S. (CR1: patterns 5–8 and 12–15).

doi:10.1371/journal.pone.0041706.g003
**Figure 4. Graphic representation of spacers grouped into patterns from CRISPR arrays CR2 and CR3 of the 85 *E. amylovora* strains examined in this study.** Hosts of *E. amylovora* strains harboring the respective spacer patterns shown are listed on the left. Individual spacer sequences are represented by boxes; spacers were considered unique if they contained >5 nucleotide differences compared to other spacer sequences. Each unique spacer (588 identified in this study) was assigned a number. Spacers of similar color shown in the same columns under the same number are identical. Empty areas indicate the corresponding spacer in other similar patterns is not present. Open boxes refer to spacers that were only detected once in the collection. Spacers shaded grey and boxed in red are spacers found in both CR1 and CR2 arrays, while spacers shared between Rubus strains are shaded in yellow or in dark blue. The following spacer patterns were predominantly isolated from *E. amylovora* strains from apple and pear isolated in the Midwest, eastern U.S., and Canada (CR2: patterns 21–28) and in the western U.S. (CR2: patterns 29, 32, and 34). doi:10.1371/journal.pone.0041706.g004
Searches in both the CRISPR spacer database and to GenBank by BLASTn revealed no significant matches greater than 18 nt. This insert was not observed in CR1 pattern 13 (Fig. 3) which may have undergone a homologous recombination event that deleted this insert and the flanking spacers. A second novel insert was detected in CR1 arrays 16 and 18 from 3 *E. amylovora* Rubus strains. This 10-bp insert consisted of 5 repeating GT pairs and was located between spacer 462 and 463 downstream of the 462 spacer.

Five *E. amylovora* strains (Ea110, Ea(T1)2, Ea(T3)2, EaRoo29, and *Rubus* isolate RKK3) with known differing profiles of bacteriophage sensitivity to ΦEa1, ΦEa7, ΦEa100, ΦEa125, and ΦEa116C were included in this study [29 and unpublished data]. Three CRISPR genotypes were identified in the 5 strains. Although their phage sensitivity differed, CRISPR genotype 4-23-38 was shared between Ea110 and EaRoo29, and genotype 4-22-38 was shared between isolates Ea(T1)2 and Ea(T3)2.

The grouping of strains based on combined patterns for CRISPR spacer arrays 1, 2, and 3 resulted in the identification of 28 distinct genotypes with clear differentiation by host of isolation and geographic location (Table 3). Similarity of genotypes among *E. amylovora* strains isolated from apple and pear in the eastern U.S. was the most readily-apparent feature identified. However, we observed an almost complete lack of overlap (only one exception) of CRISPR genotype among apple and pear strains isolated from the eastern or western U.S. (Table 3). Of two CRISPR genotypes identified among apple and pear strains isolated in Europe, the Middle East, or New Zealand, one each was identified in a strain isolated in the eastern or western U.S. (Table 3). The genotypes of strains isolated from *Rubus*, Indian
Figure 5. Cluster analysis of concatenated spacer patterns from CRISPR arrays CR1, CR2, and CR3 of 85 *E. amylovora* strains. Strains isolated from apple and pear are distinguished by geographic origin and shown in red (midwestern and eastern U.S. and Canada), blue (Western U.S.), or green (Europe, Middle East, and New Zealand). Strains from other hosts are shown in orange (Indian hawthorn), light green (loquat), and magenta (*Rubus*). Triangles to the right of strain names delineate streptomycin-resistant strains. Bootstrap values >50% are shown.

doi:10.1371/journal.pone.0041706.g005
Figure 6. CRISPR spacer distribution in plasmids and bacteriophages determined by sequence identity. Percent distribution of CRISPR spacers from (A) all 85 *E. amylovora* strains examined in this study, (B) 73 *E. amylovora* strains isolated from apple or pear, and (C) 6 *E. amylovora* strains isolated from *Rubus* sharing sequence identity with plasmids, bacteriophage, and chromosomal sequences as well as percentage of spacers with no known homology to sequences in the GenBank database. Spacers from all *E. amylovora* strains shown in (A) with plasmid homology are further subdivided into percentages with homology to plasmids from *E. amylovora*, plasmids from *E. pyrifoliae* and *Erwinia* sp. pathogens, plasmids from non-pathogenic *Erwinia* spp., plasmids from other Enterobacteriaceae, and other plasmids. Spacers from *E. amylovora* with bacteriophage homology are further subdivided into percentages with homology to *E. tasmaniensis* phage φEt88 and to other phage sequences. A total of 14 spacers with identity to plasmid sequences matched >1 plasmid. These s were counted a single time when placed into group homologies (i.e., plasmid, phage, and chromosomal) and multiple times as necessary for plasmid characterizations.

doi:10.1371/journal.pone.0041706.g006
Figure 7. Target locations of CRISPR spacer sequences from *E. amylovora* strains in plasmids pEU30 and pEA72. Spacers are numbered as in Fig. 3 and shown in blue, maroon, and orange based on their presence in CRISPR arrays from *E. amylovora* strains isolated from apple and pear, *Rubus*, and loquat, respectively. The annotated gene map showing the open reading frames of pEU30 is included, from [20]. A partial annotation of pEA72 is shown as relevant to target locations in this plasmid.

doi:10.1371/journal.pone.0041706.g007
hawthorn, or loquat were all unique when compared to the apple and pear strains due to the presence of additional unique spacers (Table 3). A complete listing of the combined CRISPR spacer array pattern for each of the 85 strains analyzed in this study is presented in Table S4.

Cluster analysis of the concatenated CRISPR sequence arrays led to the differentiation of genotypic groups that further highlighted the similarities and differences among strains related to geographic source and host of isolation (Fig. 5). Of particular interest was the separation of most of the apple and pear strains by geographical location of isolation, and the overall similarity in arrays of loquat and western U.S. apple and pear strains (Fig. 5). It should be noted that branch lengths in the tree are biased by the effect of size of the CRISPR array regions. Thus, the clustering and branching identified is not indicative of \textit{E. amylovora} phylogeny but is likely caused by rapid divergence or replacement of CRISPR arrays within individual strains.

**CRISPR Spacer Repertoire from \textit{E. pyrifoliae} and \textit{Erwinia} sp**

Although \textit{E. amylovora} and \textit{E. pyrifoliae} shared 100% sequence identity for CR1 and CR2 repeats, they did not share spacer homology. Examination of \textit{E. pyrifoliae} CRISPR genotypes revealed conservation of a single genotype among strains Ep16/96, Ep1/96 and Ep4/97 with spacer organization and sequence conserved (data not shown) suggesting that the limited number of isolates available to us for use in this study were closely related to one another. Conversely, examination of the CRISPR profiles from 3 strains of \textit{Erwinia} sp. isolated from Japan thought to be closely related to both \textit{E. amylovora} and \textit{E. pyrifoliae}, revealed conserved repeat sequences among strains but no single conserved CRISPR genotype and a diverse spacer repertoire. At the time experiments were conducted, there were no genome sequences available for Japanese \textit{Erwinia}. Despite repeated attempts, we were unable to amplify the CR1 array from Ejp556 or Ejp557 using primers EPyF-1 and EPyR-1 or any of the \textit{E. amylovora} CR1 primers listed (Table S1); however, amplification of the CR2 and CR3 arrays from Japanese \textit{Erwinia} sp. strain Ejp617 (GenBank accession CP002124) now available [15], the sequence of the CRISPR arrays from Ejp617 are known. Spacers from Japanese \textit{Erwinia} CR1 (Ejp617 only) 2 and 3 arrays shared no homology to spacers from \textit{E. amylovora}. A

**Table 4. CRISPR array genotype for CRISPR arrays CR1, CR2, and CR3 in streptomycin sensitive and SmR \textit{E. amylovora} strains isolated from apple in Michigan. Patterns listed in bold for streptomycin-sensitive strains were also observed in SmR strains.**

| CR1-CR2-CR3 Pattern | No. of SmR strains isolated in Michigan | No of SmR strains isolated in Michigan |
|----------------------|----------------------------------------|----------------------------------------|
| 2-22-38              | 3                                      | 0                                      |
| 2-23-38              | 0                                      | 1                                      |
| 3-24-38              | 3                                      | 0                                      |
| 4-22-38              | 2                                      | 0                                      |
| 4-23-38              | 1                                      | 20                                     |
| 4-24-38              | 1                                      | 0                                      |
| 4-25-38              | 1                                      | 0                                      |
| 4-27-38              | 3                                      | 1                                      |
| 4-28-38              | 1                                      | 0                                      |
| 5-27-38              | 1                                      | 0                                      |
| 6-24-38              | 1                                      | 0                                      |
| 9-23-38              | 0                                      | 2                                      |

![Figure 8. Target locations of CRISPR spacer sequences from \textit{E. amylovora} strains in bacteriophage \textit{ΦEt88} from \textit{E. tasmaniensis}. The annotated gene map, showing the open reading frames of \textit{ΦEt88}, was constructed using the sequence available in GenBank (Accession number FQ482085).](http://www.plosone.org/doi:10.1371/journal.pone.0041706.g008)

**doi:**10.1371/journal.pone.0041706.t004
single spacer, designated Ep3, was shared between E. pyrifoliae CR1 and Ep617 CR2 but was not found Ep556 or Ep557 CR2 and CR3 arrays. A diagrammatic representation of the spacer arrays from E. pyrifoliae and Erwinia sp. strains is shown in Fig. S1.

Relationship of CRISPR Spacer Sequences with known Plasmid and Bacteriophage Sequences

The sequence identity of the 588 individual CRISPR spacers identified in this study with other known sequences was determined using BLASTn analysis, and a total of 77% of the spacers had no match to the GenBank databases. The remaining 23% of spacers examined shared nucleotide identity with plasmids (16%), bacteriophage (5%) or bacterial chromosomal DNA sequences (2%) (Fig. 6). We defined positive relationships as the sharing of sequence identity among at least 25 of 32 nucleotides (>78%) in a given spacer with an existing sequence(s) in the GenBank database.

Almost all of the 95 spacers matching known plasmid sequences targeted plasmids previously identified in E. amylovora, other Erwinia spp., or other Enterobacteriaceae (Fig. 6), with a disproportionate bias toward pEU30 (targeted by 55 spacers), a conjugative plasmid only present in E. amylovora strains isolated from the western U.S. [20]. E. amylovora strains that were isolated from apple, loquat, and Rubus hosts possessed spacers with homology to pEU30. Forty-four of the 51 spacers shared 100% sequence identity with the pEU30 sequence from E. amylovora UTR2 with an apparent unbiased distribution of targets within the plasmid (Fig. 7A). The majority of the spacers with homology to pEU30 were present in the CR1 array; only 6 of 55 pEU30-spacers were located in CR2 arrays. A total of 5 of 16 of the E. amylovora strains from the western United States analyzed in this study harbored pEU30. Two of these five strains (UTR2 and UT5P4) were isolated in Utah and had no spacers matching pEU30. However, the remaining three strains (JJ1189, LA092, and LA004), isolated from pear, harbored pEU30, but also contained numerous spacers that matched the pEU30 sequence. Any Rubus or loquat strains of E. amylovora with spacers matching pEU30 did not harbor the plasmid. In addition, none of the spacers with homology to pEU30 from Rubus or loquat strains were shared with the three pear strains from the western U.S.

A single spacer (no. 214 in this study; Table S3) from Rubus strain Rkk3 shared 100% identity (32/32 bp) with the repE gene of pEA29, the ubiquitous plasmid found in E. amylovora [18]. Other significant nucleotide matches of spacers to plasmids in the necrotizing Erwinia group included E. amylovora pEA72, a 71.4 kb plasmid found in Ea273/ATCC 49946 (NC_013971) (Fig. 7B), pEL60 from E. amylovora strain Leb B60 [20], pEP36 from E. pyrifoliae [22], pET35 and pET46 from E. tassmaniensis [40], pEB170 from Erwinia billingiae [40], pEF30 and pEL02 from Erwinia sp. from Japan [12,15], and pEAR4.3 from E. amylovora Rubus strain II-5 [13]. Unlike with pEU30, most spacers targeting E. amylovora pEA72 targeted sequences located close to the replication origin of the plasmid (Fig. 7B).

E. amylovora strain Ea273 (= ATCC 49946) was the only strain examined from the midwest/eastern U.S./Canada group with a CRISPR array lacking the spacer with homology to E. amylovora pEA72 (no. 70; Table S3). Ea273 is also the only strain examined in the study possessing E. amylovora pEA72 (data not shown). Significant spacer similarity was also observed with plasmids isolated from bacterial species in the Enterobacteriaceae family, including plasmids from Salmonella enterica subsp. enterica, Yersinia pestis, Yersinia pseudotuberculosis, and Pectobacterium carotovorum subsp. carotovorum (data not shown).

Of the 26 spacers identified with identity to bacteriophage sequences, 22 were located within E. tassmaniensis bacteriophage ΦEa100 (Fig. 8). Four of the 22 spacers were found in strains from apple and pear, 1 of the spacers was found in strains from apple and pear and loquat, and the remaining 17 were identified in E. amylovora strains recovered from Rubus. The only other notable match to bacteriophage was found in spacer no. 449 (Table S3) isolated from western strains of E. amylovora. This spacer had similarity (25/29 bp) to Burkholderia bacteriophage ΦL2-2 to the coding region for a bacteriophage capsid scaffolding protein (data not shown). Sequences for E. amylovora bacteriophage ΦEa100 and ΦEa1 are available [30]. Comparison of E. amylovora isolates reported to be resistant to one or both of these bacteriophage had no spacer homology to either phage.

Typically, CRISPR spacers with homology to chromosomal DNA targeted sequences that were similar to bacteriophage integrases, a topoisomerase gene, and a pilV gene (data not shown).

Use of CRISPR Spacer Patterns for Tracking the Dispersal of SmR E. amylovora in Michigan

CRISPR genotypes of SmR and SmS E. amylovora strains isolated in Michigan were compared. The SmR strains, isolated early as 1975, exhibited diverse CRISPR spacer pattern profiles among CRISPR arrays 1, and 2, with 10 genotypes observed among 17 strains (Table 4). In contrast, 20 of 24 SmR strains from Michigan (first isolated in 1993) exhibited an identical CRISPR genotype (genome 4-23-38, Table 5). The correlation of CRISPR genotype with streptomycin sensitivity or resistance status resulted in the identification of potential SmR ancestral strains for 21 of the 24 Tn5393-containing SmR strains studied (Table 4). In addition, the CRISPR genotype of the spontaneously SmR strain S5, containing a mutation in the spiL gene [25,27] was also observed in three Sm+ Michigan genotypes (Table 4).

Discussion

Analysis of CRISPR spacer sequences and patterns revealed considerably more genetic diversity in E. amylovora than had been known previously. CRISPR genotyping enabled the differentiation of strains that were shown in previous studies to be contained within the same ribotype, pulsed-field gel electrophoresis (PFGE) group, and groEL sequence group [6,8,27]. E. amylovora has been considered a fairly homogeneous species with a low level of genetic diversity although there are obvious differences between genomes of strains isolated from apple and strains isolated from Rubus sp. [13]. This is due to the hypothesis of a recent evolutionary bottleneck associated with the colonization of apple hosts in North America in the 1700s. The previous host(s) of the progenitor strain(s) first infecting apple and pear is unknown, and, to our knowledge, a comprehensive phylogenetic analysis of E. amylovora including a large number of strains isolated in North America from wild Rosaceae hosts has not been done. While the CRISPR locus is not useful for phylogenetic analyses [41], we and others [42] have shown that CRISPR spacer array genotyping is a potential tool that could be used to identify progenitor strain(s) of E. amylovora that are currently infecting apple and pear today. In this study, the E. amylovora strains isolated from loquat were closest in genetic relatedness of CRISPR spacer content with apple and pear strains.

CRISPR spacer sequences are thought to provide a historical context of mobile sequences an organism encounters because individual spacers are inserted at the same position, adding on to the existing spacer assembly [36]. For most of its life cycle, E. amylovora is believed to be present within the interior of plants with
significant epiphytic growth only occurring on the stigma surfaces of flowers [2]. However, the recent identification of *E. amylovora* pathogenicity island sequences suggestive of functioning in association with insect hosts has broadened the habitats in which this bacterium may dwell [43]. Thus, the diversity of *E. amylovora* habitats (plant and insect) could potentially increase the ecological context of CRISPR spacer evolution with exposure to distinct microbes.

Fire blight disease and *E. amylovora* were known to spread from North America to New Zealand in the 1910s, to Europe in the 1950s, and subsequently to the Middle East [44,45]. Our results of strain genotyping based on CRISPR spacer content lead us to hypothesize that an *E. amylovora* strain(s) from the eastern U.S. is the likely source of fire blight disease spread into New Zealand and Europe. This is due to the observation that the most prevalent eastern U.S. genotype only differs from the genotypes observed in European and New Zealand strains by a small set of deleted spacers (Table 3 and Fig. 3). Spacer deletion is thought to be a common route to CRISPR genotypic differentiation [46]. Based on previous work using PFGE analyses, Jock and Geider [45] concluded that *E. amylovora* was originally spread from North America to England, and from there likely throughout Europe. PFGE results also suggested that *E. amylovora* was not repeatedly introduced from North America, rather only a few strains, in effect a bottleneck were associated with spread of *E. amylovora* to Europe [6]. Our CRISPR spacer data, and those of a previous study [42], corroborate these results, as relatively little diversity is observed among European *E. amylovora* strains while North American strains contain a higher level of diversity.

Our current results also suggest that the sources of *E. amylovora* strains initially infecting apple and pear in the eastern and western U.S. could be distinct. *E. amylovora* WSDA 16, 87–70, and 87–73 are the only western U.S. strains with similar CRISPR genotype to eastern U.S. strains (Fig. 5). It seems likely that these strains were transported from the eastern U.S. to the western U.S. through human activity such as via movement of contaminated nursery stock.

We observed a clear delineation in spacer content and diversity in the CRISPR 1 and 2 arrays in the *E. amylovora* strains studied, based on geographical location of isolation and plant host. This differentiation of spacer content lends credence to the hypothesis that CRISPR spacers reflect a geographic component of host interactions and an environmental niche component [32,36,39]. Geographic differentiation within CRISPR loci has been used previously to reconstruct the routes of transmission of *Yersinia pestis* strains from natural plague foci in China [47] and to make inferences about viral biogeography, host-virus interactions, and genome dynamics in *Sulfurobacter islandicus* [48]. For the most part, *E. amylovora* apple strains from the western U.S. harbored a completely distinct set of CRISPR spacers compared to corresponding strains from the eastern U.S. and other continents, and many of the spacers carried by the western strains targeted pEU30, a plasmid that is exclusively found in a subset of western U.S. apple strains [20].

Differentiation of CRISPR spacer content based on host of isolation adds to the possibility of an environmental niche component affecting the evolution of CRISPR loci. The *E. amylovora* strains isolated from *Rubus*, are readily differentiable from apple strains by various typing methods, and this has been confirmed at the genome sequence level [13]. These strains are also differentiated by host specificity in that *Rubus* strains are not pathogenic on apple or pear [49,50]. We found that these strains are also clearly distinct in terms of CRISPR genotype. Our results, along with the clear phylogenetic distinction of *E. amylovora* strains isolated from apple and from *Rubus* [8,12], suggests that these strains have been evolving in isolation from each other for an evolutionarily long period of time. Our results are also similar to those observed in *E. coli* when phylogenetic distance is small, a high relatedness of spacer repertoire is observed [41]. As phylogenetic distance increases, spacer relatedness decreases. However, a second aspect of the *E. coli* analyses indicated that spacer repertoire relatedness among strains followed either of two paths: the spacer content was either closely identical or completely different [41]. This radical replacement or replenishment of spacers with unique spacers was interpreted to indicate that turnover of spacers is not gradual [41]. Our results with *E. amylovora* corroborate these previous observations with *E. coli*.

The utility of CRISPR sequences for strain tracking on a local level was demonstrated in this study as we detected similar CRISPR genotypes in Michigan populations of Sm<sup>S</sup> *E. amylovora* and in corresponding Sm<sup>R</sup> strains that had either acquired Tn5393 or were spontaneous Sm<sup>R</sup> mutants. Thus, CRISPR analysis was more sensitive than comparative *groEL* sequencing or ribotyping which were used previously in an attempt to differentiate these strains [3,27]. These results are important in that they suggest that Sm<sup>R</sup> *E. amylovora* populations in Michigan evolved from indigenous populations which also suggests that the resistance has arisen in locally-adapted genotypic backgrounds.

The diversity and distribution of plasmid sequences inhabiting *E. amylovora* has received increased attention in recent years as researchers attempt to define the pan-genome of this species [21,24,51,52]. Identification of 95 spacers targeting plasmids found in *Erwinia* spp. in this study provides evidence of prior interactions and attempts to avoid invasions of specific plasmids during the life history of these strains. Of particular interest are spacers targeting plasmids reported from the epiphytic organisms *E. billingiae* and *E. tasmaniensis* and other related pathogenic species *E. pyrifoliae* and *Erwinia* sp. isolated in Japan (Fig. 6). Our observations either suggest interactions of *E. amylovora* with these other species or mobility of targeted plasmids into *E. amylovora* at points during the life history of this pathogen. Another question originating from our analyses is why is pEU30 targeted by so many spacers? Analysis of the complete sequence of pEU30 [20] suggested that the plasmid is relatively innocuous; aside from a *virB*-type system encoding conjugation machinery, the plasmid does not encode any known genes of ecological or pathogenic importance. The lack of traits encoding a positive fitness benefit might be the very reason that pEU30 is frequently targeted for elimination. In addition, we found that three strains that harbored pEU30 also contained CRISPR spacers targeting the plasmid. Since it is known that 100% nucleotide identity is required for sequence elimination by the CRISPR system [34], this could be an example of a plasmid-bacterial host “arms race” in which the plasmid has evolved through mutation to escape CRISPR surveillance. An alternate hypothesis is that self-targeting CRISPRs are involved in gene regulation; however, a recent comprehensive analysis suggested that self targeting is more a consequence of autoimmunity [53].

In a recent study with *E. coli* in which 926 unique spacer sequences were identified, none of these were found to match any known sequenced enterophages [39]. This discontinuity between CRISPR sequences and bacteriophage sequences could be due to the low availability of phage sequences compared to plasmid environmental diversity. We identified 22 spacers targeting known phage sequences, and most of these targeted phage *ΦE688*, which was previously identified in *E. tasmaniensis* [14]. Our results could also be due to the low availability of phage sequences or to the lack of encounters between the *E. amylovora* examined in this study and
these characterized phages. The potential of phage deployment for fire blight disease management has been assessed by several groups [29,30,34,35]. Since the sensitivity or resistance to infection by specific phage can be affected by genes in addition to the CRISPR loci, much more information would be necessary to predict the sensitivity of E. amylovora strains to phage under development for fire blight control.

In summary, we characterized CRISPR spacer diversity among 85 E. amylovora strains and found that this locus is robust for differentiating genotypes. We find that CRISPR analysis could be particularly useful for strain tracking on a local and possibly on a regional level. Also, the almost completely distinct composition of CRISPR arrays between E. amylovora strains isolated in the eastern and western U.S. indicates the potential that there were multiple introductions of this pathogen from native Rosaceae hosts to apple and pear hosts brought to and transported across North America by European settlers.

Supporting Information

Figure S1 Graphic representation of spacers grouped into patterns from CRISPR arrays CR1, CR2, and CR3 of E. pyrifoliae and Erwinia sp. strains. Individual spacer sequences are represented by boxes; spacers were considered unique if they contained >5 nucleotide differences compared to other spacer sequences. Each unique spacer was assigned a number with an Ep (E. pyrifoliae) or EjP (Erwinia sp.) prefix. Only spacer Ep3 was shared among E. pyrifoliae and Erwinia sp. strain 617. Empty areas indicate the corresponding spacer in other similar patterns is not present. ND indicates that the sequence of CR1 was not determined for Erwinia sp. 556 and 557.

References

1. Koczan JM, McGrath MJ, Zhao Y, Sundin GW (2009) Contribution of Erwinia amylovora exopolysaccharides amylovoran and levans to biofilm formation: Implications in pathogenicity. Phytopathology 99: 1237–1244.
2. Norelli JL, Jones AL, Aldwinckle HS (2003) Fire blight management in the twenty-first century: Using new technologies that enhance host resistance in apple. Plant Dis 87: 756–765.
3. McManus PS, Stockwell VO, Sundin GW, Jones AL (2002) Antibiotic use in plant agriculture. Annu Rev Phytopathol 40: 443–465.
4. Sebaihia M, Bocsanczy AM, Biehl BS, Quail MA, Perna NT, et al. (2010) Complete genome sequence of the plant pathogen Erwinia amylovora strain ATCC 192: 2020–2021.
5. Smits THM, Rezzonico F, Kamber T, Blom J, Goesmann A, et al. (2010) Complete genome sequence of the fire blight pathogen Erwinia pyrifoliae DSM 12163(T) and comparative genomic insights into plant pathogenicity. BMC Genom 11: 2.
6. Torrent J, Barny M-A, Katoujansky A, Dufiche P, Vanneste JL (1989) Characterization of a ubiquitous plasmid in Erwinia amylovora. Mol Plant-Microbe Interact 2: 160–165.
7. Steinberger EM, Beer SV (1990) Characterization of a 56-kb plasmid of Erwinia amylovora lacking the ubiquitous plasmid pEA29. Acta Hortic 279: 271–277.
8. Llop P, Cabrera J, Smits THM, Dreo T, Barbe S, et al. (2011) Erwinia amylovora novel plasmid pE70: complete sequence, biogeography, and role in aggressiveness in the fire blight phytopathogen. PLoS ONE 6: e20951.
9. Chiou C-S, Jones AL (1993) Nucleotide sequence analysis of a transposon (Tn5395) carrying streptomycin resistance genes in Erwinia amylovora and other gram-negative bacteria. J Bacteriol 175: 732–740.
10. McManus PS, Jones AL (1994) Epidemiology and genetic analysis of streptomycin-resistant Erwinia amylovora from Michigan and evaluation of oxytetracycline for control. Phytopathology 84: 627–633.
11. Park DH, Thapa SP, Choi BS, Kim WS, Hur JH, et al. (2011) Complete genome sequence of Japanese Erwinia strain EgiP617, a bacterial shoot blight pathogen of pear. J Bacteriol 193: 586–587.
12. Smits THM, Jaenicke S, Rezzonico F, Kamber T, Goesmann A, et al. (2010) Complete genome sequence of the fire blight pathogen Erwinia pyrifoliae DSM 12163(T) and comparative genomic insights into plant pathogenicity. BMC Genom 11: 2.
13. Foster GC, McGhee GC, Jones AL (2004) Nucleotide sequences, genetic organization, and distribution of pEU30 and pEL60 from Erwinia amylovora. Appl Environ Microbiol 70: 7539–7544.
14. McGhee GC, Jones AL (2003) Genetic analysis of chromosomal and plasmid DNAs of Erwinia pyrifoliae and Erwinia amylovora. Genom 11: 2.
15. Koczan JM, McGrath MJ, Zhao Y, Sundin GW, Jones AL, et al. (2009) Contribution of Erwinia amylovora exopolysaccharides amylovoran and levans to biofilm formation: Implications in pathogenicity. Phytopathology 99: 1237–1244.
16. Norelli JL, Aldwinckle HS, Jones AL (2003) Fire blight management in the twenty-first century: Using new technologies that enhance host resistance in apple. Plant Dis 87: 756–765.
17. Smits THM, Blom J, Goesmann A, et al. (2010) Complete genome sequence of the fire blight pathogen Erwinia pyrifoliae DSM 12163(T) and comparative genomic insights into plant pathogenicity. BMC Genom 11: 2.
18. Torrent J, Barny M-A, Katoujansky A, Dufiche P, Vanneste JL (1989) Characterization of a ubiquitous plasmid in Erwinia amylovora. Mol Plant-Microbe Interact 2: 160–165.
19. Steinberger EM, Beer SV (1990) Characterization of a 56-kb plasmid of Erwinia amylovora lacking the ubiquitous plasmid pEA29. Acta Hortic 279: 271–277.
20. Foster GC, McGhee GC, Jones AL, Sundin GW (2003) Genetic analysis of chromosomal and plasmid DNAs of Erwinia pyrifoliae and Erwinia amylovora. Appl Environ Microbiol 70: 7539–7544.
21. Llop P, Cabrera J, Smits THM, Dreo T, Barbe S, et al. (2011) Erwinia amylovora novel plasmid pE70: complete sequence, biogeography, and role in aggressiveness in the fire blight phytopathogen. PLoS ONE 6: e20951.
22. Chiou C-S, Jones AL (1993) Nucleotide sequence analysis of a transposon (Tn5395) carrying streptomycin resistance genes in Erwinia amylovora and other gram-negative bacteria. J Bacteriol 175: 732–740.
23. McManus PS, Jones AL (1994) Epidemiology and genetic analysis of streptomycin-resistant Erwinia amylovora from Michigan and evaluation of oxytetracycline for control. Phytopathology 84: 627–633.
24. Llop P, Cabrera J, Smits THM, Dreo T, Barbe S, et al. (2011) Erwinia amylovora novel plasmid pE70: complete sequence, biogeography, and role in aggressiveness in the fire blight phytopathogen. PLoS ONE 6: e20951.
20. Lehman SM, Kropinski AM, Castle AJ, Svircev AM (2009) Complete genome of the broad-host-range Escherichia coli phage Phi E21-4 and its relationship to Salmonella phage Felix O1. Appl Environ Microbiol 75: 2139–2147.

21. Muller I, Kube M, Reinhardt R, Jelkmann W, Geider K (2011) Complete genome sequences of three Escherichia coli phages isolated in North America and a bacteriophage induced from an Erwinia tasmaniensis strain. J Bacteriol 193: 795–796.

22. Schnabel EL, Jones AL (2001) Isolation and characterization of five CRISPR elements of E. coli. J Bacteriol 183: 7319–7328.

23. Lehman SM, Kropinski AM, Castle AJ, Svircev AM (2009) Complete genome of the broad-host-range Escherichia coli phage Phi E21-4 and its relationship to Salmonella phage Felix O1. Appl Environ Microbiol 75: 2139–2147.

24. Muller I, Kube M, Reinhardt R, Jelkmann W, Geider K (2011) Complete genome sequences of three Escherichia coli phages isolated in North America and a bacteriophage induced from an Erwinia tasmaniensis strain. J Bacteriol 193: 795–796.

25. Schnabel EL, Jones AL (2001) Isolation and characterization of five CRISPR elements of E. coli. J Bacteriol 183: 7319–7328.

26. Lehman SM, Kropinski AM, Castle AJ, Svircev AM (2009) Complete genome of the broad-host-range Escherichia coli phage Phi E21-4 and its relationship to Salmonella phage Felix O1. Appl Environ Microbiol 75: 2139–2147.

27. Muller I, Kube M, Reinhardt R, Jelkmann W, Geider K (2011) Complete genome sequences of three Escherichia coli phages isolated in North America and a bacteriophage induced from an Erwinia tasmaniensis strain. J Bacteriol 193: 795–796.

28. Lehman SM, Kropinski AM, Castle AJ, Svircev AM (2009) Complete genome of the broad-host-range Escherichia coli phage Phi E21-4 and its relationship to Salmonella phage Felix O1. Appl Environ Microbiol 75: 2139–2147.

29. Muller I, Kube M, Reinhardt R, Jelkmann W, Geider K (2011) Complete genome sequences of three Escherichia coli phages isolated in North America and a bacteriophage induced from an Erwinia tasmaniensis strain. J Bacteriol 193: 795–796.

30. Schnabel EL, Jones AL (2001) Isolation and characterization of five CRISPR elements of E. coli. J Bacteriol 183: 7319–7328.