Genomics 112 (2020) 3762–3772

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Comparative genomic analysis of *Erwinia amylovora* reveals novel insights in phylogenetic arrangement, plasmid diversity, and streptomycin resistance

Michael Parcey[a,b], Steven Gayder[a,b], Vivian Morley-Senkler[c], Guus Bakkeren[c], José Ramón Úrbez-Torres[c], Shawkat Ali[d], Alan J. Castle[e], Antonet M. Svircev[b,*]

[a] Centre for Biotechnology, Brock University, St. Catharines, ON L2S 3A1, Canada
[b] Agriculture and Agri-Food Canada, Vineland Station, ON L0R 2E0, Canada
[c] Agriculture and Agri-Food Canada, Summerland, BC V0H 1Z0, Canada
[d] Agriculture and Agri-Food Canada, Kentville, NS B0L 1C0, Canada
[e] Department of Biological Sciences, Brock University, St. Catharines, ON L2S 3A1, Canada

ARTICLE INFO

Keywords:
Erwinia amylovora
Fire blight
Phytopathogen
Comparative genomics
Streptomycin resistance

ABSTRACT

Erwinia amylovora is a destructive pathogen of Rosaceous plants and an economic concern worldwide. Herein, we report 93 new *E. amylovora* genomes from North America, Europe, the Mediterranean, and New Zealand. This new genomic information demonstrates the existence of three primary clades of *Amygdaloideae* (apple and pear) infecting *E. amylovora* and suggests all three independently originate from North America. The comprehensive sequencing also identified and confirmed the presence of 7 novel plasmids ranging in size from 2.9 to 34.7 kbp. While the function of the novel plasmids is unknown, the plasmids pEAR27, pEAR28, and pEAR35 encoded for type IV secretion systems. The *strA-strB* gene pair and the K43R point mutation at codon 43 of the *rpsL* gene have been previously documented to confer streptomycin resistance. Of the sequenced isolates, *rpsL*-based streptomycin resistance was more common and was found with the highest frequency in the Western North American clade.

1. Introduction

Erwinia amylovora, the causal agent of fire blight, is an important agricultural plant pathogen that infects the *Amygdaloideae* subfamily (previously *Spiroideae*) which contains apples and pears as well as the *Rubus* genus which contains raspberries [1]. The symptoms of fire blight were first documented in the Hudson Valley region of New York State, USA in 1793 [2] and later in 1887 in California [3]. The movement of infected plant material by humans has rapidly spread the pathogen around the world resulting in a minimal number of regions globally where the pathogen is not present [4].

The *E. amylovora* disease cycle is initiated in the spring, when bacterial ooze is produced in cankers of infected trees. The bacterial cells are spread by wind, rain and pollinating insects to the various hosts where they colonize the stigma of a blossom [5]. Moisture accumulates on the blossom and washes the *E. amylovora* into the hypanthium where the bacterium enters the vascular tissue of the plant [6]. The infected blossoms act as a source of secondary inoculum and, under optimal weather conditions, lead to shoot infections and death of the entire tree. At the end of the growing season, *E. amylovora* overwinters in the infected tree and will form new cankers in the spring [5]. In this way, the population of *E. amylovora* is constantly segregating itself from other environmental bacteria.

Effective fire blight management relies chiefly on the prevention of blossom infections. In Canada and the United States, antibiotics such as streptomycin are applied during open bloom [7]. This seasonal application of streptomycin has led to the development of antibiotic resistant *E. amylovora* populations. There are two primary mechanisms for streptomycin resistance in *E. amylovora*. The *strA-strB* gene pair encode for streptomycin-inactivating enzymes which are common to other *Enterobacterales* [8,9] and are found in *E. amylovora* on transposon *Tn5393* [10] and on plasmid pEA8.7 [11]. *E. amylovora* can also escape streptomycin activity through the modification of its target, the S12 protein of the 30S small ribosomal subunit encoded by the *rpsL* gene. A point mutation of the *rpsL* gene, at codon 43 or 88, has been shown to minimize streptomycin sensitivity in *E. amylovora*, even at exceedingly high concentrations [12,13].

The advance of next-generation sequencing platforms allowed the sequencing and characterization of *E. amylovora* genomes. The first two isolates sequenced were CFBP1430 [14] and ATCC 49946 [15] from

* Corresponding author.
E-mail address: Antonet.Svircev@canada.ca (A.M. Svircev).

https://doi.org/10.1016/j.ygeno.2020.04.001
Received 7 February 2020; Received in revised form 16 March 2020; Accepted 1 April 2020
Available online 04 April 2020

France and New York State, respectively. Comparisons to later European sequenced genomes showed a remarkable genetic similarity of 99.99% [16]. Recently, larger sequencing projects have been undertaken to examine the genomic diversity of *Amygdaloideae*-infecting isolates found in the eastern United States [17]. Two primary clades of *E. amylovora* were found in the United States; the Widely-Prevalent clade and the Eastern U.S. clade. Four isolates from the western United States did not fit into either of these clades which suggested two additional clades called the Western U.S. 1 and Western U.S. 2 clades [17]. The genetic diversity of a species or clade is a measure of the variability within that population. This is often represented by the Watterson theta (θ_w) and the pairwise difference (θ_π), where the former is the expected level of variability per site and the latter is the observed variability [70]. The difference between these two values (θ_π - θ_w), known as Tajima's D, can be used to determine if a specific population has rapidly expanded [18].

In this study, we investigated the how the isolates of *E. amylovora* from North America vary from those found in other geographic regions. A total of 127 *E. amylovora* genomes were examined. Our analysis suggested that the sampled North American *E. amylovora* collection which can infect *Amygdaloideae* plants, could be separated into 8 clades. Our study also confirmed that the Widely-Prevalent clade is the only known clade which was found outside of North America. The genomic data was examined to determine the occurrence of plasmids within the *E. amylovora* isolates. We used these results to produce a phylogeny based on the pEA29 plasmid and to identity 7 novel plasmids. Lastly, the genomes were screened for known genes responsible for streptomycin resistance to determine if there was any correlation between antibiotic resistance, clade development, and plasmid occurrence.

2. Results and discussion

2.1. Genomic diversity

In 2017 and 2018, whole-genome sequencing was performed on 93 *E. amylovora* isolates from Canada and around the world to generate draft genomes. These sequences were combined with 34 previously sequenced genomes [16,17] for a total of 127 genomes. The chromosomal sequences of this large data set were aligned to CFBP 1430 and the gaps generated from the draft genomes were removed. This whole-genome alignment which was 3,449,098 bp in length and contained 77,361 variable sites. As a species, *E. amylovora* showed an average similarity of 99.62% (pairwise difference, 3.76E-3 per site) with a Watterson theta of 4.14E-3 per site. This is a very low level of expected diversity in a species. *Pseudomonas syringae*, a more extensively studied plant pathogen, consists of 41 different pathovars some of which are capable of infection of apples and pears. The pathovars *P. syringae pv. tomato* and *P. syringae pv. kiwi* have similar Watterson thetas of 6.92E-3 and 8E-3 per site respectively, where the Watterson theta of the species approaches 5.89E-2 per site [19–21]. From this perspective, *E. amylovora* as a species that has a high level of genomic conservation.

Whole-genome alignment was used to produce a phylogeny based on variable sites (Fig. 1). The major genetic division within *E. amylovora* is determined by its host preference where isolates can be clustered roughly in the *Amygdaloideae*-infecting (A.I.; Fig. 1A) and the *Rubus*-infecting (R.I.; Fig. 1D, purple) superclades. The A.I. superclade is compose of the three primary *E. amylovora* clades; the Widely-Prevalent clade (Fig. 1B, red), the Eastern N.A. clade (Fig. 1C, green) and the Western N.A. clade (Fig. 1C, blue). The clades of the A.I. superclade are approximately equidistance from one another with average pairwise differences ranging from 6.09E-4 to 6.85E-4 per site (Table 1). Located in between the A.I. and R.I. superclades was the “B-Group” superclade (Fig. 1A; orange): a collection of distinct strains which showed little sequence similarity to each other or any of the other clades.

The Widely-Prevalent clade contains the vast majority of sequenced *E. amylovora* isolates as well as every isolate collected outside of North America. The Widely-Prevalent clade has a remarkable low intracladal pairwise difference of 7.37E-5 per site (Watterson theta, 1.87E-4 per site). This low level of diversity, combined with its commonality, led to the original conclusion that genomes from all *E. amylovora* were 99.99% similar [14]. The comparison of the pairwise difference and the Watterson theta generates a negative Tajima's D which can suggest population expansion. This would coincide with the pan-global expansion of *E. amylovora* starting in the early 1900's [4]. While the Widely-Prevalent clade has a vast geographic distribution, its genome has diverged very slowly and has only 3034 variable sites. A large portion of these variable sites, 1169 specifically, are singletons representing variants found in only one of the isolates within the clade. This shows that while the Widely-Prevalent clade is changing genetically, it is doing so at an extremely slow rate as some isolates within this clade were collected up to 57 years apart (1959 to 2016). This may be due to both the size of the genome and the number of proteins which it encodes. The genome of *E. amylovora* is 3.8 Mbp in length while other plant pathogens such as *P. syringae pv. syringae* B728a (NC_007005.1) and *Pectobacterium carotovorum* subsp. *carotovorum PC1* (NC_012917.1) have much larger genomes of 6.1 and 4.9 Mbp, respectively. With these larger genomes also comes a larger number of protein coding sequences (CDS). Using Glimmer3 [22], we predicted *P. syringae pv. syringae* B728a of having 3218 genes and *P. carotovorum* subsp. *carotovorum PC1* of having 2931 genes: *E. amylovora* isolate CFBP1430 is predicted to have only 2012 genes. The smaller genome and fewer protein coding genes would suggest an increased number of essential genes required by *E. amylovora* due to the lack of redundancy [23–25]. Additionally, a niche pathogenic life cycle is likely to put a high degree of pressure on *E. amylovora* to conserve its gene functions [26]. As such, these two factors are likely acting to conserve the genome of *E. amylovora*.

The recent effort to isolate and sequence *E. amylovora* in the eastern United States led to an observed decrease in similarity within the A.I. superclade to 99.90% [17]. This decrease in similarity was due to the first reports of the Eastern N.A. clade, Western N.A clade and isolate CA3R, the first member of the B-Group. This level of similarity is much closer to what was observed amongst the clades in this study at 99.93% (pairwise difference, 7.07E-4 per site). When compared to the previous study, the intraclade similarity and Watterson theta of the Eastern N.A. clade remained consistent at 99.996% (pairwise difference, 4.18E-05 per site) and 5.30E-05 per site respectively. While the intraclade similarity of Western N.A. clade was consistent at 99.997% (pairwise difference, 3.41E-05 per site), the Watterson theta increased to 6.49E-05 per site. This corresponds to a negative Tajima's D similar to the Widely-Prevalent clade, and suggests population of the Western N.A. clade is expanding.

The *E. amylovora* CA3R and ATCC BAA-2158 were the first members of the B-Group superclade to be sequenced though their connection was unknown at the time. CA3R was isolated from apple [27] while BAA-2158 was isolated from blackberry [28]. Despite this, BAA-2158 was later shown to be infectious on pear fruit [29]. While *Rubus*-infecting isolates are unable to infect *Amygdaloideae* hosts, *Amygdaloideae*-infecting isolates are able to infect *Rubus* hosts [30]. This, and the phylogenic placement of BAA-2158 in the B-Group, would suggest that BAA-2158 may be an *Amygdaloideae*-infecting strain which was collected from a *Rubus* host.

There were an additional eight isolates which fell into the B-Group superclade after sequencing. All of these isolates were collected in North America. The intraclade similarity of B-Group is 99.76% (pairwise difference, 2.4E-03 per site) which is approximately the same as their relation to primary A.I. clades (99.74% similarity). This would indicate that the clades which make up the B-Group superclade are not grouped together based on their similarity to one another: they are clustered together based on their dissimilarity to the A.I. and R.I. superclades. Likewise, the R.I. superclade is also highly diverse with no more than 3 isolates per clade, a pairwise identity of only 99.89%, and 13,185 variable sites between the sequences (Watterson theta, 1.15E-3

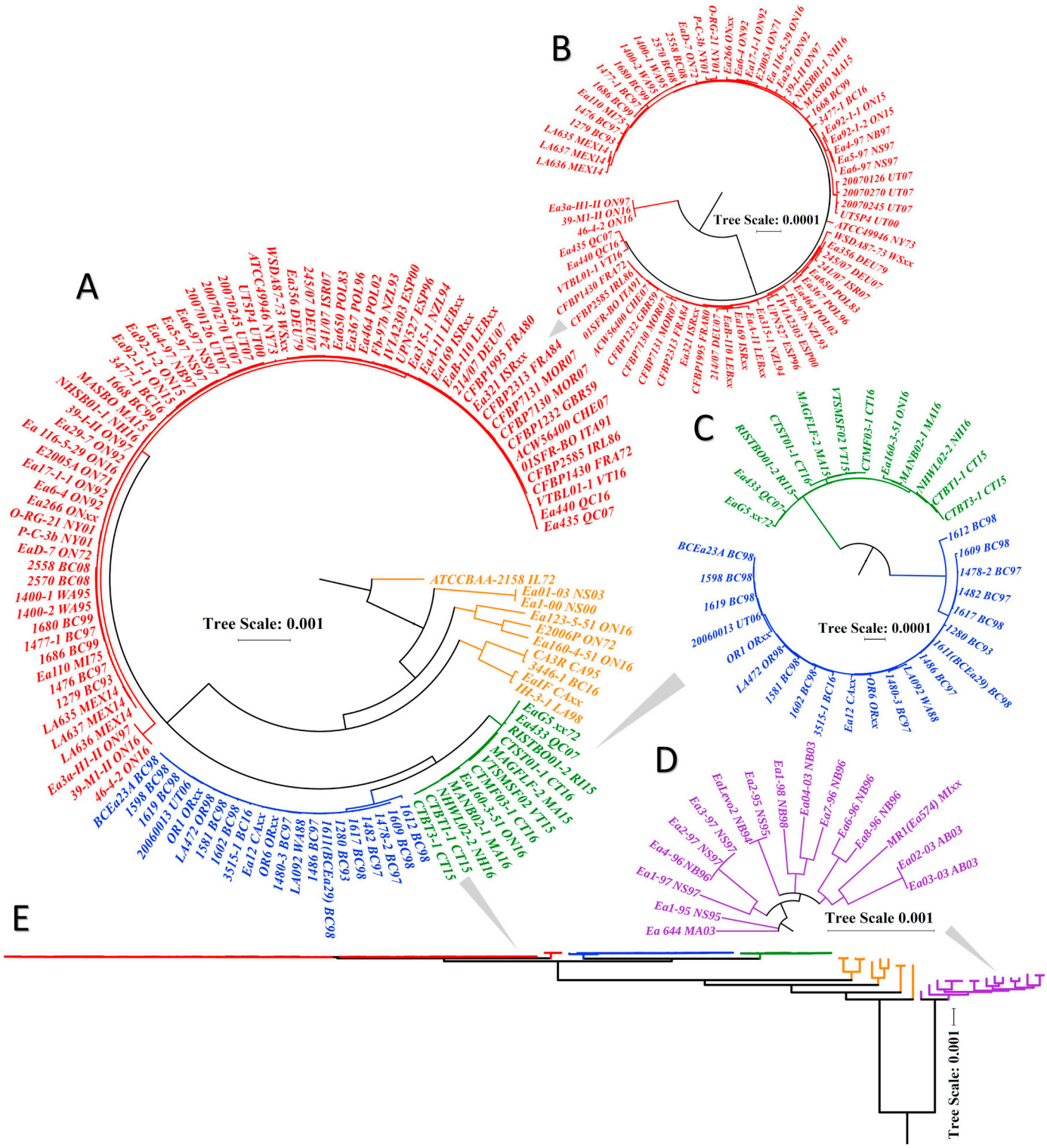

Fig. 1. Phylogeny of *Erwinia amylovora* based on the whole-genome alignment. (A) Phylogeny of *Amygdaloideae*-infecting group of *E. amylovora*. (B) Phylogeny of the high prevalence clade of *E. amylovora*. (C) Phylogeny of the low prevalence clade of *E. amylovora*. (D) Phylogeny of *Rubus*-infecting group of *E. amylovora*. (E) Phylogeny of all sequenced *E. amylovora* chromosomes. The phylogenetic tree was generated in RAxML with a bootstrap value of 1000 based on the merged chromosomal alignment generated by REALPHY. Only branches with bootstrap frequency 70% or greater are shown. The coding after each isolate's name is state/province (two letters), country (three letters), and year of isolation. Colours indicate the clade which the isolate belongs to (red, Widely-Prevalent; blue, Western N.A.; green, Eastern N.A.; orange, B-Group; purple, *Rubus*-infecting). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Pairwise differences of the whole-genome alignment of *E. amylovora*.

		Widely-Prevalent	North America		B-Group	Rubus
			Eastern	Western		
Widely-Prevalent		7.37E-05				
North America	Eastern	6.85E-04	4.18E-05			
	Western	6.70E-04	6.09E-04	3.41E-05		
B-Group		2.62E-03	2.62E-03	2.64E-03	2.40E-03	
Rubus		1.43E-02	1.43E-02	1.43E-02	1.42E-02	1.09E-03

* Grey boxes show intraclade pairwise differences

Grey boxes show intraclade pairwise differences.

per site). If it can be assumed that the genetic conservation of the A.I. superclade is due to the structure of the pathogenicity cycle and host-specificity, then the diversity found in the R.I. and B-Group clades would suggest that their reservoirs are likely not the hosts which they were isolated from.

2.2. Plasmids in E. amylovora

2.2.1. Plasmid pEA29

A distinguishing genomic characteristic of *E. amylovora* is its ubiquitous plasmid pEA29 found in almost all isolates. This plasmid is non-transmissible, increases fitness when colonizing young tissue, and has been used extensively for identification of *E. amylovora* [31,32]. While the pEA29 plasmid has been reported as large as 35,000 bp due to the insertion of transposons such as *Tn*5393 [31], the size of all of the sequenced pEA29 plasmids within this study fall between 27,319 and 28,876 bp when a circular alignment is taken into account. Due to the non-transmissible nature of pEA29 and the conservation of the core genome, a phylogeny based on the sequence of pEA29 was produced to determine if similar clade conservation was observed.

The phylogeny produced using the sequences of pEA29 was nearly identical to that of the chromosomal phylogeny (Fig. 2). Each previously observed clade is represented within pEA29 phylogeny with the major genetic division still between the A.I. (Fig. 2A) and R.I. (Fig. 2B) superclades. The intra-clade pairwise difference remained vastly unchanged between the chromosomal sequences and pEA29 (Table 2) with the exception of the R.I. superclade and Western N.A. clade which saw decreases to 8.56E-04 and 7.68E-06 per site, respectively. Additionally, while the three major clades of A.I. *E. amylovora* are equidistance from each other when considering only their chromosomes, the pEA29 plasmids of the Eastern and Western N.A. clades show a higher similarity to each other and a greater dissimilarity to the pEA29 plasmids of the Widely-Prevalent clade. This similarity between the Eastern N.A. and Western N.A. clades would suggest a common ancestor distinct from the Widely-Prevalent clade. Overall, the conserved sequence similarity between the chromosome and pEA29 would suggest that they are influenced by the same selective pressures. In this way, the pEA29 can be considered an extension of the chromosome. This may prove useful as a phylogenetic tool. Long-read sequencing platforms, such as nanopore sequencing, continue to increase the feasibility of rapidly and cheaply sequencing whole plasmids [33]. In *E. amylovora*, this could be exploited to identify the clade of isolates not destined for whole genome sequencing.

2.2.2. Large accessory plasmids

There are more than 13 previously documented plasmids found in lower frequency in *E. amylovora*, most of which are cryptic, that is, no phenotypic traits have been ascribed to them [34]. The two most common plasmids other than pEA29 found within this study are pEA72 and pEU30. The plasmid pEA72 (71,487 bp) appears in only 8 sequenced isolates in our study: all of which are from eastern North America and six of these isolates belong to the Eastern N.A. clade. The plasmid pEU30 (30,314 bp) was also found in 8 sequenced *E. amylovora* isolates but, in contrast, these isolates were primarily from western North America and the Western N.A. clade. The plasmid pEA78 (78,740 bp) was found within LA637 from Mexico and CA3R isolated in California, USA. The presence of pEA72, pEU30, and pEA78 has not been linked to a specific function however all three plasmids were predicted to harbor a type IV secretion system (T4SS). The other two large plasmids were found in isolates from the Widely-Prevalent clade and which were isolated outside of North America. Plasmid pEL60 was originally isolated from Israel and Lebanon while pEI70 was found in isolates from France, Morocco, and Switzerland. The function of these plasmids is unknown [35].

Aside from the previously observed large plasmids, we report the presence of three novel plasmids within the R.I. clades. These plasmids were identified *in silico* and confirmed through restriction digestion (data not shown). Based on their size and clade of isolation, these plasmids were designated as pEAR27, pEAR28, and pEAR35. pEAR28 (28,558 bp) is a novel plasmid found in Ea1–98 and pEAR35 plasmid (34,732 bp) is found in EaLevo2 and Ea2–95. Both plasmids encode for unrelated P-type T4SS proteins. The pEAR27 (27,479 bp) plasmid, found in Ea03–03, also encoded for T4SS proteins. While the homology of these proteins is ~40 to 50% to known T4SS proteins, the gene order doesn't reflect a standard F-type, P-type or I-type T4SS [36] and so its primary function is cannot be inferred.

2.2.3. Small accessory plasmids

There were eight small plasmids found within this study and these small plasmids are unique to the B-Group and R.I. superclades. Within the B-Group superclade itself, only isolate Ea123–5-51 does not have at least one small plasmid. Four of these small plasmids have been previously described; pEA1.7 [37], pEAR4.3 [38], pEAR5.2 [38], and pEA8.7 [11]. Of these four plasmids only pEA8.7 has a known function: it harbours the *strA/strB* genes which confer streptomycin resistance [11].

In addition to the previously described plasmids, there were 4 novel small plasmids identified in this study. pEA2.9 (2847 bp) was the smallest of the novel plasmids sequenced in this study. Found in isolates EaIF and 3446–1, pEA2.9 is predicted to have only two genes: a replicase and a hypothetical protein with low homology to an Abi gene. If this hypothetical protein also has a similar function, the role of this plasmid would likely be to grant bacteriocin immunity [39]. The novel plasmid pEA4.0 (4068 bp) was found in isolates Ea01–03 and Ea1–00 and plasmid pEA5.8 (5800 bp) was found in isolates EaIH-3-1, 3446–1 and EaIF: all members of the B-Group superclade. All predicted proteins of these plasmids except for the relaxase and replicase homologs were hypothetical and therefore the function of these plasmids is unknown.

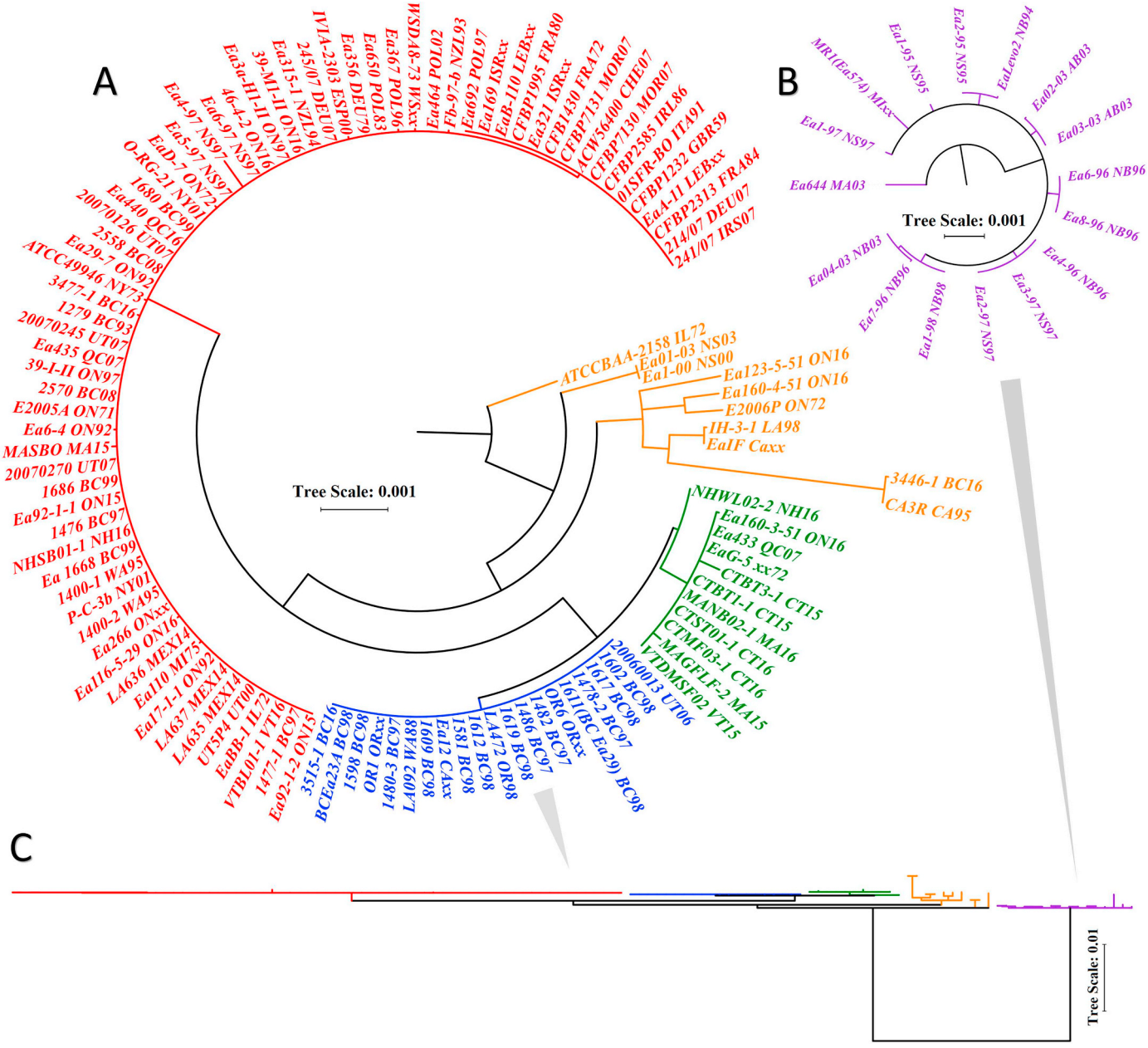

Fig. 2. Phylogeny of *Erwinia amylovora* inferred from the alignment of pEA29. (A) Phylogeny of *Amygdaloideae*-infecting group of *E. amylovora*. (B) Phylogeny of *Rubus*-infecting group of *E. amylovora*. (C) Phylogeny of all *E. amylovora* pEA29 sequences. The phylogenetic tree was generated in RAxML with a bootstrap value of 1000 based on the merged chromosomal alignment generated by REALPHY. Only branches with bootstrap frequency 70% or greater are shown. The coding after each isolate's name is state/province (two letters), country (three letters), and year of isolation. Colours indicate the clade which the isolate belongs to (red, Widely-Prevalent; blue, Western N.A.; green, Eastern N.A.; orange, B-Group; purple, *Rubus*-infecting). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The largest of the novel, small plasmids was pEA6.0 (5944 bp), found in isolates Ea1–95 and Ea7–96 of the R.I. superclade. In addition to its predicted relaxase homolog, an endonuclease and a cytosine methyltransferase were also identified. Presumably, these genes work together as a protective restriction system within pEA6.0 to grant bacteriophage resistance [40]. While this is supported by the inability of some *Podoviridae* phage to cause infection in Ea7–96 [41], further study of this plasmid would be required for confirmation. These novel plasmids were identified *in silico* and confirmed through restriction digestion (data not shown).

2.2.4. *E. amylovora* plasmid sequence similarity

One of the major similarities observed across all large plasmids of *E. amylovora* is the presence of genes coding for the T4SS. This can be seen in the amino acid similarity between pEA72, pEA78, and pEL60 as well as the relation between pEAR35 and pEAR28 (Fig. 3). The shortest of the large plasmids, such as pEU30, pEAR27, pEAR28, and pEAR35 only encode the T4SS proteins while the only large plasmids which are not predicted to code for a T4SS are pEI70 and the non-conjugative pEA29. As such, the main role of these plasmids may be providing the T4SS itself. While T4SS are mostly commonly used for horizontal DNA transfer [42], we would have expected greater genetic diversity and more plasmids in these samples if this were the case. Alternatively, some 13% of T4SS are used for transfer of proteins [43]. This is done to eliminate bacteria competitors [44] or, in the case with plant pathogen *Agrobacterium tumefaciens*, to transfer effector proteins into the plant

Table 2
Pairwise differences of the pEA29 plasmids of *E. amylovora*

		Widely-Prevalent	North America		B-Group	Rubus
			Eastern	Western		
Widely-Prevalent		7.49E-05				
North America	Eastern	2.07E-03	6.85E-05			
	Western	2.07E-03	2.82E-04	7.68E-06		
B-Group		3.18E-03	2.64E-03	2.73E-03	2.71E-03	
Rubus		3.63E-02	3.63E-02	3.64E-02	3.67E-02	8.56E-04

* Grey boxes show intraclade pairwise differences

Grey boxes show intraclade pairwise differences.

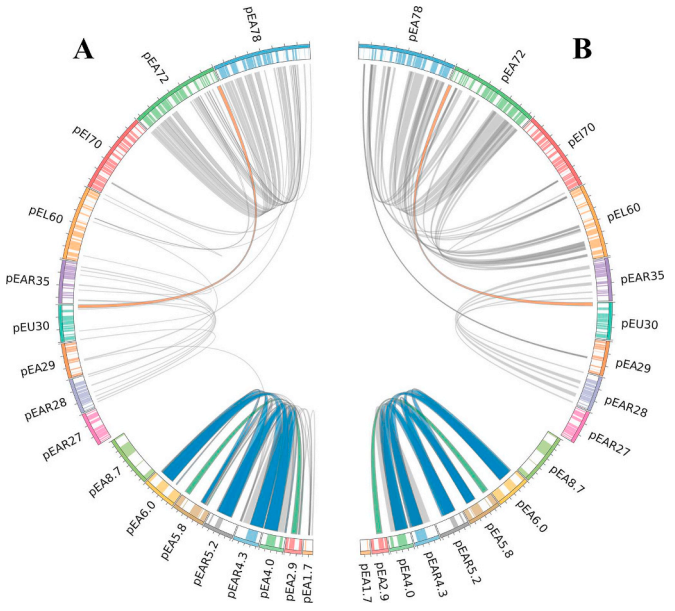

Fig. 3. The molecular relation of *E. amylovora* plasmids found within this study. (A) Relation of plasmids based on nucleotide identity. Ribbons indicate regions with an e-value greater than 1E-03. (B) Relation of plasmids based on amino acid identity. Ribbons indicate regions with an e-value greater than 1E-70. The colored bands represent the predicted protein coding sequences of the plasmids. The coloured ribbons indicate regions of high similarity: orange, ~3000 bp region of 100% similarity; blue, the relaxase gene of the small plasmids; green, the replicase gene of the small plasmids. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

cell [45]. In *E. amylovora* bacterial competition, virulence, and exopolysaccharide production have been linked to the endogenous Type-III and Type-VI secretion systems [46–48]. It is possible that T4SS systems may serve a similar function or enhance the function of existing systems however further investigation is required.

The conjugative gene of the small plasmids was another commonality amongst plasmids. A similar *mobA*-like relaxase homolog was found within the predicted genes of pEA4.0, pEAR4.3, pEAR5.2, pEA5.8, and pEA6.0 (Fig. 3: blue ribbon), which could be used in their conjugative transfer. Additionally, pEA5.8 and pEA2.9 shared a similar DNA replicase gene (Fig. 3: green ribbon) which would likely be used for mobilization during conjugation [49]. Lastly, there was no significant genetic similarity between large plasmids and the small plasmids found in *E. amylovora*. This is likely to minimize incompatibility between the plasmids [50].

Table 3
Isolates of *E. amylovora* with *rpsL* mutations that confere streptomycin resistance.

Clade	Isolate	Location	rpsL Codon 43 Sequence	
			DNA	Amino Acid
Wild type	–		AAA	K43
High Prevalence	20,070,126	USA, UT	AGA	K43R
High Prevalence	92–1-1	Canada, ON	AGA	K43R
High Prevalence	92–1-2	Canada, ON	AGA	K43R
High Prevalence	Ea116–5-29	Canada, ON	ACA	K43T
High Prevalence	Ea241	Israel	AGA	K43R
High Prevalence	Ea315–1	New Zealand	AGA	K43R
High Prevalence	LA635	Mexico	AGA	K43R
High Prevalence	LA637	Mexico	AGA	K43R
Eastern N.A.	Ea160–3-51	Canada, ON	ATA	K43I
Western N.A.	1280	Canada, BC	AGA	K43R
Western N.A.	1478–2	Canada, BC	AGA	K43R
Western N.A.	1480–3	Canada, BC	AGA	K43R
Western N.A.	1482	Canada, BC	AGA	K43R
Western N.A.	1486	Canada, BC	AGA	K43R
Western N.A.	1609	Canada, BC	AGA	K43R
Western N.A.	1611 (BCEa29)	Canada, BC	AGA	K43R
Western N.A.	1612	Canada, BC	AGA	K43R
Western N.A.	Ea12	USA, CA	AGA	K43R
Western N.A.	LA092	USA, WA	AGA	K43R
Western N.A.	OR1	USA, OR	AAT	K43N
Western N.A.	OR6	USA, OR	AGA	K43R
B-Group	Ea160–4-51	Canada, ON	AAC	K43N
B-Group	Ea123–5-51	Canada, ON	AGA	K43R

2.3. Streptomycin resistance

The genetic capacity for streptomycin resistance (SmR) was identified in 24 *E. amylovora* isolates examined in this study; 23 which contained a mutation at codon 43 of *rpsL* (Table 3). The SmR isolates which were sequenced in this study were then confirmed experimentally to have streptomycin resistance at 1000 ppm if this information was not previously published [51]. This *rpsL* mutations was found most commonly within the Western N.A. clade in which 12 out of the 21 had point mutations; 11 were K43R mutations and a single isolate displayed a K43N mutation. In addition, there were 8 isolates from the Widely-Prevalent clade which again all showed the K43R mutation except Ea116–5-29 which had a K43T mutation. This increased prevalence of SmR in the Western N.A. clade may be the cause of the expansion which was seen in the chromosomal analysis. The Eastern N.A. clade only had one isolate, Ea160–3-51, which showed SmR with a SNP generating an K43I variant. Likewise, only two isolates of the B-Group showed SNPs within the *rpsL* gene resulting in isolates Ea160–4-51 and Ea123–5-51 having K43N and K43R mutations, respectively. While no isolates from the *Rubus*-Infecting superclade had known streptomycin resistance, there were still mutations within the *rpsL* gene. When compared to the

A.I. clades, the R.I. isolates had 2 synonymous SNPs within the 72nd and 78th codon. In addition, isolate MR1 showed a nucleotide deletion within the poly(A) repeat of codon 43 and 44. As this type of indel is a common sequencing error, this deletion in MR1 needs to be confirmed. If proven correct, a deletion of this nature would cause the truncation of the RpsL protein reducing its size from 124 amino acids to 56 amino acids. There were no isolates within this study which showed a point mutation within the 88th codon of *rpsL* which has been related to SmR in *Pectobacterium carotovorum* (*Erwinia carotovora*) [52].

While *rpsL*-based SmR was relatively common amongst the sequenced isolates, only a single isolate, CA3R, contained the *strA/strB* gene set. CA3R is known to have had these genes located on its pEA8.7 plasmid [11]. This plasmid was previously sequenced [17] but never described. It can be found as the 18th contig of the CA3R assembly (NQKC01000018.1) and is the first sequence for pEA8.7 from *E. amylovora*. As previously presumed, it is an exact match to the RSF1010 plasmid from *E. coli* [37]. The *strA/strB* genes can also be transferred *via* the *Tn*5393 transposon, which can insert itself into the pEA34 and pEA29 plasmids [53,54]. While the *Tn*5393 transposon has been sequenced, there are currently no published sequences for pEA34, pEA29:*Tn*5393, or a complete *E. amylovora* genome which contains either plasmid.

3. Conclusions

E. amylovora was first observed in Hudson Valley, New York in 1793 [2] and this region has been suggested as the point of origin of all *E. amylovora* [4,17]. This hypothesis holds for Widely-Prevalent clade which has spread globally but not the other clades of *E. amylovora*. No geographical location, whether it be France, New Zealand, Israel, or Poland, produced enough selective pressure to cause significant genetic variation of the genome of the Widely-Prevalent clade. The natural rate of mutation is also insignificant, otherwise mutations which are not influenced by selection, such as synonymous mutations and mutations in the intergenic regions, would have increased the observed number of variable sites. For example, *E. amylovora* was first observed in New Zealand in 1919 [55] and yet isolates from that region, such as EA315–1, have remained nearly indistinguishable from all the other isolates of the Widely-Prevalent clade. These observations, along with the separation of the R.I. and A.I. superclades, would suggest that the driving force for genetic diversity within *E. amylovora* is host selection. The continuous cultivation of apples and pears in afflicted regions is therefore likely the cause of the genomic conservation of *E. amylovora*. It is therefore reasonable to assume that once a strain of *E. amylovora* enters the pathogenicity cycle in apples and pears it remain vastly unchanged. This means that each clade of *E. amylovora* is likely to have their own original reservoir host and point of origin within North America. As such, only the Widely-Prevalent clade would have originated from the Hudson Valley.

Isolates from the Western N.A. clade appear to be expanding as observed through the prevalence of rare variants and their negative Tajami's D value. We hypothesize that this is due to the clade's predisposition to develop *rspL*-based SmR. The Western N.A. clade had the greatest number of streptomycin resistant isolates and only one of the Widely-Prevalent isolates which had SmR was found in western North America. The use of streptomycin, especially in British Columbia, Canada, seems to have selected for SmR isolates from the Western N.A. clade, at least between 1993 and 1998 when the survey was completed [51]. The increased prevalence of Western N.A. isolates in British Columbia may also suggest that this clade originated from the British Columbia/Washington State area. It is also interesting that while the plasmid pEU30 does not have SmR genes, all 6 isolates from western North America which have this plasmid also have *rpsL*-based SmR. This could suggest a potential correlation between pEU30 and the development of *rpsL*-based SmR however further investigation is required.

In summary, we have shown that the Widely-Prevalent, Western N.A., and Eastern N.A. are the three primary clades which cause fire blight in North America and that the Widely-Prevalent clade is still the only clade which has a global distribution. While there are many clades which compose the B-Group, they are infrequently isolated and genetically diverse. Likewise, R.I. isolates show significantly greater genetic variation than A.I. isolates. Seven novel plasmids were described within this study, all of which were found in the highly diverse B-Group and R.I superclades. The function of the large plasmids may be related to the T4SS they code, whereas the smaller plasmids have an assortment of functions, likely related to competition with other microbes and providing streptomycin resistance. Yet, the overall impact of these plasmids on pathogenicity is still unknown.

4. Materials and methods

4.1. Sequencing and assembly

Total DNA of *E. amylovora* was isolated using the DNeasy UltraClean Microbial Kit (#12224–50, Qiagen). Library preparation and sequencing were performed by the Sequencing and Bioinformatics Consortium (University of British Columbia, Vancouver, Canada). DNA was quantified using Qubit and diluted for library preparation using Nextera XT, according to manufacturer's instructions (Illumina). QC'd libraries were pooled and sequenced on a NextSeq Mid Output lane to generate paired-end 150 bp reads. Raw base call data (bcl) was converted into FastQ format using the bcl2fastq conversion software from Illumina. *De novo* assemblies were constructed from trimmed, paired-end reads using the CLC Genomics Workbench v9 (Qiagen). These draft genomes were combined with 34 genomic sequences of *E. amylovora* available from GenBank. Isolates used in this study and their accession numbers from GenBank can be found in Table 4.

4.2. Genomic diversity and phylogeny

The contigs of each genome were concatenated using the MeDuSa scaffolder [56]. These scaffolds were then aligned using RealPhy [57]. Due to the limitations of the program, only CFBP1430 could be used as a reference. Pairwise difference (p-distance), the number of variable sites, and Watterson theta were calculated using MEGA X [58]. Phylogeny was produced in RaxML 8 [59] using the maximum-likelihood model and a bootstrap value of 1000. Phylogenies were visualized using iTOL [60] with a bootstrap cut off of 70%. The names within the phylogenies are the isolate name is followed by the location and year they were isolated. Countries are designated by their IBAN alpha-3 codes. The provinces and states of Canada and the United States respectively are designated by their two letter postal abbreviations.

4.3. Plasmid identification and phylogeny

Newly sequenced isolates in this study were not screened experimentally for the presence of plasmids, therefore *in silico* screening was performed. A method similar to PlasmidSeeker [61] was developed in Biopython [62] after the PlasmidSeeker was unable to recognize pEA29. The sequence coverage of the contigs which did not concatenate to the scaffold (if available) for each isolate was compared to the median coverage of the assembly. The sequences which had a coverage greater than the median minus 1 standard deviation were compared to known *Erwinia* plasmids as well as the GenBank [63] for plasmid similarities.

Novel plasmids were extracted and confirmed through restriction digestion. Briefly, isolates were grown in nutrient broth (NB; BD Diagnostics) at 27 °C (165 rpm) until an OD_{600} of 0.6 was reached. The plasmids were then extracted using the Monarch® Plasmid Miniprep Kit (#T1010S, NEB) and restricted using *Sna*BI (#R0130S, NEB), and/or *Nhe*I (#R0131S, NEB) as per the manufacturer's protocol. The banding pattern of the restricted plasmids was then observed through in gel

Table 4
The genomes of *E. amylovora* which were analysed in this study.

Clade	Strain ID	Location	Year	Isolate Host	Plasmids	Accession	Reference
Widely-Prevalent	1279	Canada, BC	1993	*Malus domestica*	pEA29	JAAEXY000000000	This study; [51]
Widely-Prevalent	1476	Canada, BC	1997	*Malus domestica*	pEA29	JAAEXW000000000	This study; [51]
Widely-Prevalent	1477–1	Canada, BC	1997	*Malus domestica*	pEA29	JAAEXB000000000	This study; [51]
Widely-Prevalent	1668	Canada, BC	1999	*Malus domestica*	pEA29	JAAEXM000000000	This study
Widely-Prevalent	1680	Canada, BC	1999	*Malus domestica*	pEA29	JAAEXL000000000	This study; [51]
Widely-Prevalent	1686	Canada, BC	1999	*Malus domestica*	pEA29	JAAEXK000000000	This study; [51]
Widely-Prevalent	2558	Canada, BC	2008	*Pyrus communis*	pEA29	JAAEXJ000000000	This study
Widely-Prevalent	2570	Canada, BC	2008	*Malus domestica*	pEA29	JAAEXI000000000	This study
Widely-Prevalent	3477–1	Canada, BC	2016	*Malus domestica*	pEA29	JAAEWT000000000	This study
Widely-Prevalent	Ea4–97	Canada, NB	1997	*Malus domestica*	pEA29	JAAEUQ000000000	This study
Widely-Prevalent	Ea5–97	Canada, NS	1997	*Malus domestica*	pEA29	JAAEVE000000000	This study
Widely-Prevalent	Ea6–97	Canada, NS	1997	*Malus domestica*	pEA29	JAAEVA000000000	This study
Widely-Prevalent	39-I-II	Canada, ON	1997	*Malus domestica*	pEA29, pEA72	JAAEWR000000000	This study
Widely-Prevalent	39-M1-II	Canada, ON	2016	*Crataegus mollis*	pEA29	JAAEWQ000000000	This study
Widely-Prevalent	46–4-2	Canada, ON	2016	*Malus sp.*	pEA29	JAAEWP000000000	This study
Widely-Prevalent	E2005A	Canada, ON	1972	*Malus domestica*	pEA29	JAAEWJ000000000	This study
Widely-Prevalent	Ea116–5-29	Canada, ON	2016	*Malus domestica*	pEA29	JAAEWC000000000	This study
Widely-Prevalent	Ea17–1-1	Canada, ON	1992	*Malus domestica*	pEA29	JAAEVW000000000	This study
Widely-Prevalent	Ea266	Canada, ON	N/A	*Malus sp.*	pEA29	NZ_CAOY00000000	[16]
Widely-Prevalent	Ea29–7	Canada, ON	1992	*Malus domestica*	pEA29	JAAEVM000000000	This study
Widely-Prevalent	Ea3a-H1-II	Canada, ON	1997	Unknown	pEA29	JAAEVL000000000	This study
Widely-Prevalent	Ea6–4	Canada, ON	1992	*Malus domestica*	pEA29	JAAEVD000000000	This study
Widely-Prevalent	Ea92–1-1	Canada, ON	2015	*Malus domestica*	pEA29	JAAEUX000000000	This study
Widely-Prevalent	Ea92–1-2	Canada, ON	2015	*Malus domestica*	pEA29	JAAEUW000000000	This study
Widely-Prevalent	EaD-7	Canada, ON	1972	*Malus domestica*	pEA29	JAAEUT000000000	This study
Widely-Prevalent	Ea435	Canada, QC	2007	*Malus domestica*	pEA29	JAAEVJ000000000	This study
Widely-Prevalent	Ea440	Canada, QC	2016	*Pyrus communis*	pEA29	JAAEVI000000000	This study
Widely-Prevalent	CFBP 1232	England	1959	*Pyrus communis*	pEA29	CAPB00000000	[16]
Widely-Prevalent	CFBP 1430	France	1972	*Crataegus sp.*	pEA29	GCA_000091565.1	[14]
Widely-Prevalent	CFBP 1995	France	1980	*Malus domestica*	pEA29, pEI70	JAAEWN000000000	This study; Paulin, J. P.
Widely-Prevalent	CFBP 2313	France	1984	*Pyrus communis*	pEA29, pEI70	JAAEWM000000000	This study; Paulin, J. P.
Widely-Prevalent	214/07	Germany	2007	*Malus domestica*	pEA29	JAAEWX000000000	This study; Moltmann, E.
Widely-Prevalent	245/07	Germany	2007	*Malus domestica*	pEA29	JAAEWV000000000	This study; Moltmann, E.
Widely-Prevalent	Ea356	Germany	1979	*Cotoneaster sp.*	pEA29	CAOX00000000	[16]
Widely-Prevalent	CFBP 2585	Ireland	1986	*Sorbus sp.*	pEA29	CAOZ00000000	[16]
Widely-Prevalent	Ea169	Israel	N/A	*Pyrus communis*	pEA29	JAAEWW000000000	This study; Manulis, S.
Widely-Prevalent	Ea241	Israel	2007	*Pyrus communis*	pEA29	JAAEVX000000000	This study; Manulis, S.
Widely-Prevalent	Ea321	Israel	N/A	*Pyrus communis*	pEA29	JAAEVO000000000	This study; Manulis, S.
Widely-Prevalent	01SFR-BO	Italy	1991	*Sorbus sp.*	pEA29	CAOZ00000000	[16]
Widely-Prevalent	EaA-11	Lebanon	N/A	*Malus domestica*	pEA29, pEL60	JAAEUV000000000	This study; Sundin, G.
Widely-Prevalent	EaB-110	Lebanon	N/A	*Malus domestica*	pEA29	JAAEUU000000000	This study; Sundin, G.
Widely-Prevalent	LA635	Mexico	2014	*Malus domestica*	pEA29	CBVS000000000	[67]
Widely-Prevalent	LA636	Mexico	2014	*Malus domestica*	pEA29	CBVT000000000	[67]
Widely-Prevalent	LA637	Mexico	2014	*Unknown*	pEA29, pEA78	CBVU000000000	[67]
Widely-Prevalent	CFBP 7130	Morocco	2007	*Pyrus communis*	pEA29, pEI70	JAAEWL000000000	This study; Paulin, J. P.
Widely-Prevalent	CFBP 7131	Morocco	2007	*Pyrus communis*	pEA29	JAAEWK000000000	This study; Paulin, J. P.
Widely-Prevalent	Ea315–1	New Zealand	1994	*Malus domestica*	pEA29	JAAEVP000000000	This study; Vanneste, J.
Widely-Prevalent	Fb-97b	New Zealand	1993	*Malus domestica*	pEA29	JAAEUP000000000	This study; Vanneste, J.
Widely-Prevalent	Ea367	Poland	1996	*Pyracantha sp.*	pEA29	JAAEVN000000000	This study; Sobiczewski, P.
Widely-Prevalent	Ea464	Poland	2002	*Pyrus communis*	pEA29	JAAEVH000000000	This study; Sobiczewski, P.
Widely-Prevalent	Ea650	Poland	1983	*Crataegus monogyna*	pEA29	JAAEVC000000000	This study; Sobiczewski, P.
Widely-Prevalent	IVIA2303	Spain	2000	*Pyrus communis*	pEA29	JAAEUN000000000	This study; Lopez, M.
Widely-Prevalent	UPN527	Spain	1996	*Malus domestica*		CAPC00000000.1	[16]
Widely-Prevalent	ACW 56400	Switzerland	2007	*Pyrus communis*	pEA29, pEI70	AFHN000000000	[16]
Widely-Prevalent	MASHBO	USA, MA	2015	*Pyrus communis*	pEA29	NQJK00000000	[17]
Widely-Prevalent	Ea110	USA, MI	1975	*Malus domestica*	pEA29	NQJZ00000000	[17,68]
Widely-Prevalent	NHSB01–1	USA, NH	2016	*Malus domestica*	pEA29	NQJW00000000	[17]
Widely-Prevalent	O-RG-21	USA, NY	2001	*Malus domestica*	pEA29	JAAEUL000000000	This study; Aldwinckle, H.
Widely-Prevalent	P-C-3b	USA, NY	2001	*Malus domestica*	pEA29	JAAEUK000000000	This study; Aldwinckle, H.
Widely-Prevalent	ATCC 49946	USA, NY	1973	*Malus domestica*	pEA29, pEA72	GCA_000027205	[15]
Widely-Prevalent	20,070,126	USA, UT	2007	*Malus domestica*	pEA29, pEU30	JAAEXF000000000	This study; [69]
Widely-Prevalent	20,070,245	USA, UT	2007	*Malus domestica*	pEA29	JAAEXE000000000	This study; [69]
Widely-Prevalent	20,070,270	USA, UT	2007	*Pyrus communis*	pEA29	JAAEXG000000000	This study; [69]
Widely-Prevalent	UT5P4	USA, UT	2000	*Malus domestica*	pEA29	NQJS00000000	[17,54]
Widely-Prevalent	VTBL01–1	USA, VT	2016	*Malus domestica*	pEA29	NQJR00000000	[17]
Widely-Prevalent	1400–1	USA, WA	1995	*Malus domestica*	pEA29	JAAEXD000000000	This study
Widely-Prevalent	1400–2	USA, WA	2016	*Malus domestica*	pEA29	JAAEXC000000000	This study

(*continued on next page*)

Table 4 (*continued*)

Clade	Strain ID	Location	Year	Isolate Host	Plasmids	Accession	Reference
Widely-Prevalent	WSDA87–73	USA, WA	N/A	*Malus domestica*	pEA29	NQJQ00000000	[17,27]
Eastern N.A.	Ea160–3-51	Canada, ON	2016	*Pyrus communis*	pEA29	JAAEVZ000000000	This study
Eastern N.A.	Ea433	Canada, QC	2007	*Malus domestica*	pEA29	JAAEVK000000000	This study
Eastern N.A.	EaG5	Unknown (Presumed: Canada, ON)	1972	*Pyrus communis*	pEA29, pEA72	JAAEUS000000000	This study
Eastern N.A.	CTBT1–1	USA, CT	2015	*Pyrus communis*	pEA29, pEA72	NQJP00000000	[17]
Eastern N.A.	CTBT3–1	USA, CT	2015	*Pyrus communis*	pEA29, pEA72	NQJO00000000	[17]
Eastern N.A.	CTMF03–1	USA, CT	2016	*Pyrus communis*	pEA29	NQKB00000000	[17]
Eastern N.A.	CTST01–1	USA, CT	2016	*Malus domestica*	pEA29, pEA72	NQKA00000000	[17]
Eastern N.A.	MAGFLF-2	USA, MA	2015	*Malus domestica*	pEA29	NQJN00000000	[17]
Eastern N.A.	MANB02–1	USA, MA	2016	*Malus domestica*	pEA29	NQJX00000000	[17]
Eastern N.A.	NHWL02–2	USA, NH	2015	*Malus domestica*	pEA29	NQJV00000000	[17]
Eastern N.A.	RISTBO01–2	USA, RI	2015	*Malus domestica*	pEA72	NQJM00000000	[17]
Eastern N.A.	VTDMSF02	USA, VT	2015	*Malus domestica*	pEA29, pEA72	NQJL00000000	[17]
Western N.A.	1280	Canada, BC	1993	*Malus domestica*	pEU30	JAAEXX000000000	This study; [51]
Western N.A.	1478–2	Canada, BC	1997	*Malus domestica*	pEA29, pEU30	JAAEXA000000000	This study; [51]
Western N.A.	1480–3	Canada, BC	1997	*Malus domestica*	pEA29, pEU30	JAAEWZ000000000	This study; [51]
Western N.A.	1482	Canada, BC	1997	*Pyrus communis*	pEA29	JAAEXV000000000	This study; [51]
Western N.A.	1486	Canada, BC	1997	*Malus domestica*	pEA29	JAAEXU000000000	This study; [51]
Western N.A.	1581	Canada, BC	1998	*Malus domestica*	pEA29	JAAEXT000000000	This study; [51]
Western N.A.	1598	Canada, BC	1998	*Pyrus communis*	pEA29	JAAEXS000000000	This study; [51]
Western N.A.	1602	Canada, BC	1998	*Malus domestica*	pEA29	JAAEXR000000000	This study; [51]
Western N.A.	1609	Canada, BC	1998	*Malus domestica*	pEA29	JAAEXQ000000000	This study; [51]
Western N.A.	1611 (BCEa29)	Canada, BC	1998	*Pyrus communis*	pEA29	JAAEWY000000000	This study; [51]
Western N.A.	1612	Canada, BC	1998	*Malus domestica*	pEA29	JAAEXP000000000	This study; [51]
Western N.A.	1617	Canada, BC	1998	*Malus domestica*	pEA29	JAAEXO000000000	This study; [51]
Western N.A.	1619	Canada, BC	1998	*Malus domestica*	pEA29	JAAEXN000000000	This study; [51]
Western N.A.	3515–1	Canada, BC	2016	*Malus domestica*	pEA29	JAAEWS000000000	This study
Western N.A.	BCEa23A	Canada, BC	1998	*Pyrus betulifolia*	pEA29	JAAEWO000000000	This study
Western N.A.	Ea12	USA, CA	N/A	*Pyrus communis*	pEA29, pEU30	JAAEWB000000000	This study; Lindow, S.
Western N.A.	LA472	USA, OR	1998	*Pyrus communis*	pEA29	JAAEUM000000000	This study; Stockwell, V.
Western N.A.	OR1	USA, OR	N/A	*Pyrus communis*	pEA29	NQJU00000000	[17,27]
Western N.A.	OR6	USA, OR	N/A	*Pyrus communis*	pEA29, pEU30	NQJT00000000	[17,27]
Western N.A.	20,060,013	USA, UT	2006	*Pyrus communis*	pEA29	JAAEXH000000000	This study; [69]
Western N.A.	LA092	USA, WA	1988	*Pyrus communis*	pEA29	NQJY00000000	[17,27]
B-Group	3446–1	Canada, BC	2016	*Pyrus communis*	pEA2.9, pEA5.8, pEA29	JAAEWU000000000	This study
B-Group	Ea01–03	Canada, NS	2003	*Pyrus communis*	pEA4.0, pEA29, pEU30	JAAEWH000000000	This study
B-Group	Ea1–00	Canada, NS	2000	*Pyrus communis*	pEA4.0, pEA29, pEU30	JAAEWD000000000	This study
B-Group	E2006P	Canada, ON	1997	*Pyrus communis*	pEAR4.3, pEA29	JAAEWI000000000	This study
B-Group	Ea123–5-51	Canada, ON	2016	*Malus domestica*	pEA29	JAAEWA000000000	This study
B-Group	Ea160–4-51	Canada, ON	2016	*Pyrus communis*	pEAR4.3, pEA29	JAAEVY000000000	This study
B-Group	CA3R	USA, CA	1995	*Malus domestica*	pEA8.7, pEA29, pEA78	NQKC00000000	[17,27]
B-Group	EaIF	USA, CA	N/A	*Malus domestica*	pEA2.9, pEA5.8, pEA29	JAAEUR000000000	This study; Lindow, S.
B-Group	ATCC BAA-2158	USA, Illinois	1972	*Rubus sp.*	pEAR4.3, pEAR5.2, pEA29	PRJEA52823	[37]
B-Group	IH-3-1	USA, LA	1998	*Rhaphiolepis indica*	pEA1.7, pEA5.8, pEA29	JAAEUO000000000	This study; Sundin, G.
Rubus-Infecting	Ea02–03	Canada, AB	2003	*Rubus idaeus*	pEA29	JAAEWG000000000	This study
Rubus-Infecting	Ea03–03	Canada, AB	2003	*Rubus idaeus*	pEAR27, pEA29	JAAEWF000000000	This study
Rubus-Infecting	Ea04–03	Canada, NB	2003	*Rubus idaeus*	pEA29	JAAEWE000000000	This study
Rubus-Infecting	Ea1–98	Canada, NB	1998	*Rubus idaeus*	pEAR28, pEA29	JAAEVT000000000	This study
Rubus-Infecting	Ea4–96	Canada, NB	1996	*Rubus idaeus*	pEA29	JAAEVG000000000	This study
Rubus-Infecting	Ea6–96	Canada, NB	1996	*Rubus idaeus*	pEA29	JAAEUQ000000000	This study
Rubus-Infecting	Ea7–96	Canada, NB	1996	*Rubus idaeus*	pEA6.0, pEA29	JAAEUQ000000000	This study
Rubus-Infecting	Ea8–96	Canada, NB	1996	*Rubus idaeus*	pEA29	JAAEUQ000000000	This study
Rubus-Infecting	EaLevo2	Canada, NB	1994	*Rubus idaeus*	pEA29, pEAR35	JAAEUQ000000000	This study
Rubus-Infecting	Ea1–95	Canada, NS	2016	*Rubus idaeus*	pEA6.0, pEA29	JAAEVV000000000	This study
Rubus-Infecting	Ea1–97	Canada, NS	1997	*Rubus idaeus*	pEA29	JAAEVU000000000	This study
Rubus-Infecting	Ea2–95	Canada, NS	1995	*Rubus idaeus*	pEA29, pEAR35	JAAEVS000000000	This study
Rubus-Infecting	Ea2–97	Canada, NS	1997	*Rubus idaeus*	pEA29	JAAEVR000000000	This study
Rubus-Infecting	Ea3–97	Canada, NS	1997	*Rubus idaeus*	pEA29	JAAEVQ000000000	This study
Rubus-Infecting	Ea644	USA, MA	2003	*Rubus idaeus*	pEA29	CAPD00000000	[16]
Rubus-Infecting	MR1 (Ea574)	USA, MI	N/A	*Rubus sp.*	pEA29	CAPE00000000	[16]

N/A – Information is unknown.

electrophoresis to confirm the *in silico* predictions.

All pEA29 sequences were used to produce a phylogeny. Sequences were first oriented to the same start position using CSA [64], aligned using MAFFT [65], and then a phylogeny was produced using RaxML 8 as above. The genes of the plasmids were predicted using Glimmer3 [22]. The circular diagram showing the relatedness of the plasmids was created in Circos [66]. Regions of nucleotide identity and amino acid identity were determined using BLASTn and BLASTp (NCBI), respectively. The e-value cut-off for nucleotide identity was 1E-3 and the e-value cut of for amino acid identity 1E-70. All read frames were considered when determining amino acid identity and overlapping regions were combine into a single band within the figure.

4.4. Streptomycin resistance

Isolates were screened for the presence of the strA/strB genes and mutations in the rpsL gene using BLASTn. Any isolates which were sequenced within this study, and showed the genomic capacity for

streptomycin resistance, were then confirmed experimentally. Each isolate was grown from frozen stock in NB at 27 °C (165 rpm) for 5 h before being streaked on nutrient agar (BD Diagnostics) amended with 0 ppm, 100 ppm, and 1000 ppm streptomycin (#S9137, Sigma-Aldrich).

Acknowledgements

We would like to acknowledge and thank the following individuals: Julie Boulé, Rob Linning, and Sherry Sun for the collection and preparation of *E. amylovora* isolates from British Columbia, Canada; Darlene Nesbitt, Kinga Bodnár, Emily Bordeleau, and Curtis Burton for the collection and preparation of *E. amylovora isolates* from Ontario, Canada; Gordon Braun and Eric Bevis for the collection and preparation of *E. amylovora* strains in Atlantic Canada. Lastly, we would like to thank all of the researchers who have generously shared their *E. amylovora* isolates with us. This research was funded by Agriculture and Agri-Food Canada Growing Forward II and Canadian Agricultural Partnership grants as well as grant RGPIN-2016-05590 from the Natural Sciences and Engineering Research Council of Canada.

Author statement

All the authors have agreed on the contents of the manuscript and its submission to Genomics, and confirm that neither the manuscript nor any of its parts are currently under consideration or published in another journal. Genomic sequences generated in this study have been submitted to GenBank and will be released upon acceptance of this manuscript. The authors also declare that there is no conflict of interest to disclose.

References

[1] T. van der Zwet, N. Orolaza-Halbrendt, W. Zeller, Symptomatology of Fire Blight and Host Range of *Erwinia amylovora*, in: Fire Blight: History, Biology, and Management, Phytopathology (2016) 45–64.
[2] W. Denning, On the decay of apple trees, NY Soc. Promotion. Agric. Arts Manuf. Trans. 2 (1794) 219–222.
[3] M.W. Gardner, P.A. Ark, Early references to pear blight in California, Plant. Dis. Rptr. 48 (1924) 871–872.
[4] T. van der Zwet, N. Orolaza-Halbrendt, W. Zeller, Spread and Current Distribution of Fire Blight, in: Fire Blight: History, Biology, and Management, Phytopathology (2016) 15–36.
[5] T. van der Zwet, N. Orolaza-Halbrendt, W. Zeller, The Disease Cycle of Fire Blight, in: Fire Blight: History, Biology, and Management, Phytopathology (2016) 155–182.
[6] P.L. Pusey, The role of water in epiphytic colonization and infection of pomaceous flowers by *Erwinia amylovora*, Phytopathology 90 (2000) 1352–1357.
[7] V.O. Stockwell, B. Duffy, Use of antibiotics in plant agriculture, Rev. Sci. Tech. 31 (2012) 199–210.
[8] C. Pezzella, A. Ricci, E. DiGiannatale, I. Luzzi, A. Carattoli, Tetracycline and streptomycin resistance genes, transposons, and plasmids in *Salmonella enterica* isolates from animals in Italy, Antimicrob. Agents Chemother. 48 (2004) 903–908.
[9] K.J. Shaw, P.N. Rather, R.S. Hare, G.H. Miller, Molecular genetics of aminoglycoside resistance genes and familial relationships of the aminoglycoside-modifying enzymes, Microbiol. Rev. 57 (1993) 138–163.
[10] G.W. Sundin, C.L. Bender, Dissemination of the *strA-strB* streptomycin-resistance genes among commensal and pathogenic bacteria from humans, animals, and plants, Mol. Ecol. 5 (1996) 133–143.
[11] E.L. Palmer, B.L. Teviotdale, A.L. Jones, A realtive of the broad-host-range plasmid RSF1010 detected in *Erwinia amylovora*, Appl. Environ. Microbiol. 63 (1997) 4604–4607.
[12] H. Forster, G.C. McGhee, G.W. Sundin, J.E. Adaskaveg, Characterization of streptomycin resistance in isolates of *Erwinia amylovora* in California, Phytopathology 105 (2015) 1302–1310.
[13] Y. Zhang, X. Yang, F.Y. Zhou, A.F. Zhang, X.F. Zhu, Y. Chen, M.G. Zhou, T.C. Gao, Detection of a mutation at codon 43 of the *rpsL* gene in *Xanthomonas oryzae* pv. Oryzicola and *X. oryzae* pv. Oryzae by PCR-RFLP, Genet. Mol. Res. 14 (2015) 18587–18595.
[14] T.H. Smits, F. Rezzonico, T. Kamber, J. Blom, A. Goesmann, J.E. Frey, B. Duffy, Complete genome sequence of the fire blight pathogen *Erwinia amylovora* CFBP 1430 and comparison to other *Erwinia spp.* Mol. Plant-Microbe Interact. 23 (2010) 384–393.
[15] M. Sebaihia, A.M. Bocsanczy, B.S. Biehl, M.A. Quail, N.T. Perna, J.D. Glasner, G.A. DeClerck, S. Cartinhour, D.J. Schneider, S.D. Bentley, J. Parkhill, S.V. Beer, Complete genome sequence of the plant pathogen *Erwinia amylovora* strain ATCC 49946, J. Bacteriol. 192 (2010) 2020–2021.
[16] R.A. Mann, T.H. Smits, A. Buhlmann, J. Blom, A. Goesmann, J.E. Frey, K.M. Plummer, S.V. Beer, J. Luck, B. Duffy, B. Rodoni, Comparative genomics of 12 strains of *Erwinia amylovora* identifies a pan-genome with a large conserved core, PLoS One 8 (2013) e55644.
[17] Q. Zeng, Z. Cui, J. Wang, K.L. Childs, G.W. Sundin, D.R. Cooley, C.H. Yang, E. Garofalo, A. Eaton, R.B. Huntley, X. Yuan, N.P. Schultes, Comparative genomics of Spiraeoideae-infecting *Erwinia amylovora* strains provides novel insight to genetic diversity and identifies the genetic basis of a low-virulence strain, Mol. Plant Pathol. 19 (2018) 1652–1666.
[18] S. Biswas, J.M. Akey, Genomic insights into positive selection, Trends Genet. 22 (2006) 437–446.
[19] R. Cai, S. Yan, H. Liu, S. Leman, B.A. Vinatzer, Reconstructing host range evolution of bacterial plant pathogens using *Pseudomonas syringae* pv. Tomato and its close relatives as a model, Infect. Genet. Evol. 11 (2011) 1738–1751.
[20] H.C. McCann, E.H.A. Rikkerink, F. Bertels, M. Fiers, A. Lu, J. Rees-George, M.T. Andersen, A.P. Gleave, B. Haubold, M.W. Wohlers, D.S. Guttman, P.W. Wang, C. Straub, J.L. Vanneste, P.B. Rainey, M.D. Templeton, Correction: Genomic Analysis of the Kiwifruit Pathogen *Pseudomonas syringae* pv. actinidiae Provides Insight into the Origins of an Emergent Plant Disease, PLoS Pathogens 9 (2013).
[21] S.F. Sarkar, D.S. Guttman, Evolution of the core genome of *Pseudomonas syringae*, a highly clonal, endemic plant pathogen, Appl. Environ. Microbiol. 70 (2004) 1999–2012.
[22] A.L. Delcher, K.A. Bratke, E.C. Powers, S.L. Salzberg, Identifying bacterial genes and endosymbiont DNA with glimmer, Bioinformatics 23 (2007) 673–679.
[23] I. Matic, T. Bergmiller, M. Ackermann, O.K. Silander, Patterns of evolutionary conservation of essential genes correlate with their compensability, PLoS Genetics (2012) 8.
[24] J.I. Glass, N. Assad-Garcia, N. Alperovich, S. Yooseph, M.R. Lewis, M. Maruf, C.A. Hutchison 3rd, H.O. Smith, J.C. Venter, Essential genes of a minimal bacterium, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 425–430.
[25] G. Fang, E. Rocha, A. Danchin, How essential are nonessential genes? Mol. Biol. Evol. 22 (2005) 2147–2156.
[26] M.P. Francino, The ecology of bacterial genes and the survival of the new, Int. J. Evol. Biol. 2012 (2012) 394026.
[27] P.S. McManus, A.L. Jones, Genetic fingerprinting of *Erwinia amylovora* strains Isolted from tree-fruit crops and *Rubus spp.* Phytopathology 85 (1995) 1547–1553.
[28] S.M. Ries, Occurrence of fire blight on thornless blackberry in Illinois, Plant Dis. Rep. 61 (1977) 232–235.
[29] E.M. Steinberger, G.-Y. Cheng, S.V. Beer, Characterization of a 56-kb plasmid of *Erwinia amylovora* Ea322: its noninvolvement in pathogenicity, Plasmid 24 (1990) 12–24.
[30] P.G. Braun, P.D. Hildebrand, Infection, carbohydrate utilization, and protein profiles of apple, pear, and raspberry isolates of *Erwinia amylovora*, Can. J. Plant Pathol. 27 (2005) 338–346.
[31] G.C. McGhee, A.L. Jones, Complete nucleotide sequence of ubiquitous plasmid pEA29 from *Erwinia amylovora* strain Ea88: gene organization and intraspecies variation, Appl. Environ. Microbiol. 66 (2000) 4897–4907.
[32] H. Falkenstein, P. Bellemann, S. Walter, W. Zeller, K. Geider, Identification of *Erwinia amylovora*, the fire blight pathogen, by colony hybridization with DNA from plasmid pEA29, Appl. Environ. Microbiol. 54 (1988) 2798–2802.
[33] J.K. Lemon, P.P. Khil, K.M. Frank, J.P. Dekker, Rapid Nanopore sequencing of plasmids and resistance gene detection in clinical isolates, J. Clin. Microbiol. 55 (2017) 3530–3543.
[34] P. Llop, S. Barbe, M.M. Lopez, Functions and origin of plasmids in *Erwinia* species that are pathogenic to or epiphytically associated with pome fruit trees, Trees (Berl West) 26 (2012) 31–46.
[35] P. Llop, J. Cabrefiga, T.H. Smits, T. Dreo, S. Barbe, J. Pulawska, A. Bultreys, J. Blom, B. Duffy, E. Montesinos, M.M. Lopez, *Erwinia amylovora* novel plasmid pEI70: complete sequence, biogeography, and role in aggressiveness in the fire blight phytopathogen, PLoS One 6 (2011) e28651.
[36] T.D. Lawley, W.A. Klimke, M.J. Gubbins, L.S. Frost, F factor conjugation is a true type IV secretion system, FEMS Microbiol. Lett. 224 (2003) 1–15.
[37] G.C. McGhee, E.L. Schnabel, K. Maxson-Stein, B. Jones, V.K. Stromberg, G.H. Lacy, A.L. Jones, Relatedness of chromosomal and plasmid DNAs of *Erwinia pyrifoliae* and *Erwinia amylovora*, Appl. Environ. Microbiol. 68 (2002) 6182–6192.
[38] R. Powney, T.H. Smits, T. Sawbridge, B. Frey, J. Blom, J.E. Frey, K.M. Plummer, S.V. Beer, J. Luck, B. Duffy, B. Rodoni, Genome sequence of an *Erwinia amylovora* strain with pathogenicity restricted to *Rubus* plants, J. Bacteriol. 193 (2011) 785–786.
[39] M. Kjos, L. Snipen, Z. Salehian, I.F. Nes, D.B. Diep, The Abi proteins and their involvement in bacteriocin self-immunity, J. Bacteriol. 192 (2010) 2068–2076.
[40] T.H. Bestor, DNA methylation: evolution of a bacterial immune function into a regulator of gene expression and genome structure in higher eukaryotes, Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. 326 (1990) 179–187.
[41] S. Gayder, M. Parcey, A.J. Castle, A.M. Svircev, Host range of bacteriophages against a world-wide collection of *Erwinia Amylovora* determined using a quantitative PCR assay, Viruses (2019) 11.
[42] E. Cabezon, J. Ripoll-Rozada, A. Pena, F. de la Cruz, I. Arechaga, Towards an integrated model of bacterial conjugation, FEMS Microbiol. Rev. 39 (2015) 81–95.
[43] Z. Esna Ashari, N. Dasgupta, K.A. Brayton, S.L. Broschat, An optimal set of features for predicting type IV secretion system effector proteins for a subset of species based on a multi-level feature selection approach, PLoS One 13 (2018) e0197041.
[44] D.P. Souza, M.O. Andrade, C.E. Alvarez-Martinez, G.M. Arantes, C.S. Farah, R.K. Salinas, A component of the *Xanthomonadaceae* type IV secretion system

combines a VirB7 motif with a N0 domain found in outer membrane transport proteins, PLoS Pathog. 7 (2011) e1002031.
[45] Y.G. Li, P.J. Christie, The *Agrobacterium* VirB/VirD4 T4SS: Mechanism and architecture defined through in vivo mutagenesis and chimeric systems, in: S.B. Gelvin (Ed.), Agrobacterium Biology: From Basic Science to Biotechnology, Springer International Publishing, Cham, 2018, pp. 233–260.
[46] Y. Tian, Y. Zhao, L. Shi, Z. Cui, B. Hu, Y. Zhao, Type VI secretion systems of *Erwinia amylovora* contribute to bacterial competition, virulence, and exopolysaccharide production, Phytopathology 107 (2017) 654–661.
[47] D. Khokhani, C. Zhang, Y. Li, Q. Wang, Q. Zeng, A. Yamazaki, W. Hutchins, S.S. Zhou, X. Chen, C.H. Yang, Discovery of plant phenolic compounds that act as type III secretion system inhibitors or inducers of the fire blight pathogen, *Erwinia amylovora*, Appl. Environ. Microbiol. 79 (2013) 5424–5436.
[48] A. Degrave, M. Fagard, C. Perino, M.N. Brisset, S. Gaubert, S. Laroche, O. Patrit, M.A. Barny, *Erwinia amylovora* type three-secreted proteins trigger cell death and defense responses in *Arabidopsis thaliana*, Mol. Plant-Microbe Interact. 21 (2008) 1076–1086.
[49] F. Lorenzo-Diaz, C. Fernandez-Lopez, R. Lurz, A. Bravo, M. Espinosa, Crosstalk between vertical and horizontal gene transfer: plasmid replication control by a conjugative relaxase, Nucleic Acids Res. 45 (2017) 774–785.
[50] M. Shintani, Z.K. Sanchez, K. Kimbara, Genomics of microbial plasmids: classification and identification based on replication and transfer systems and host taxonomy, Front. Microbiol. 6 (2015) 242.
[51] P.L. Sholberg, K.E. Bedford, P. Haag, P. Randall, Survey of *Erwinia amylovora* isolates from British Columbia for resistance to bactericides and virulence on apple, Can. J. Plant Pathol. 23 (2001) 60–67.
[52] A.M. Barnard, N.J. Simpson, K.S. Lilley, G.P. Salmond, Mutations in *rpsL* that confer streptomycin resistance show pleiotropic effects on virulence and the production of a carbapenem antibiotic in *Erwinia carotovora*, Microbiology 156 (2010) 1030–1039.
[53] K.A. Tancos, S. Villani, S. Kuehne, E. Borejsza-Wysocka, D. Breth, J. Carol, H.S. Aldwinckle, K.D. Cox, Prevalence of streptomycin-resistant *Erwinia amylovora* in New York apple orchards, Plant Dis. 100 (2016) 802–809.
[54] G.C. McGhee, G.W. Sundin, *Erwinia amylovora* CRISPR elements provide new tools for evaluating strain diversity and for microbial source tracking, PLoS One 7 (2012) e41706.
[55] A.H. Cockayne, Fireblight and its control; The hawthorn question, N.Z. J. Agric. 23 (1921) 30–36.
[56] E. Bosi, B. Donati, M. Galardini, S. Brunetti, M.F. Sagot, P. Lio, P. Crescenzi, R. Fani, M. Fondi, MeDuSa: a multi-draft based scaffolder, Bioinformatics 31 (2015) 2443–2451.
[57] F. Bertels, O.K. Silander, M. Pachkov, P.B. Rainey, E. van Nimwegen, Automated reconstruction of whole-genome phylogenies from short-sequence reads, Mol. Biol. Evol. 31 (2014) 1077–1088.
[58] S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, MEGA X: molecular evolutionary genetics analysis across computing platforms, Mol. Biol. Evol. 35 (2018) 1547–1549.
[59] A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies, Bioinformatics 30 (2014) 1312–1313.
[60] I. Letunic, P. Bork, Interactive tree of life (iTOL) v4: recent updates and new developments, Nucleic Acids Res. 47 (W1) (2019) W256–W259.
[61] M. Roosaare, M. Puustusmaa, M. Mols, M. Vaher, M. Remm, PlasmidSeeker: identification of known plasmids from bacterial whole genome sequencing reads, PeerJ 6 (2018) e4588.
[62] P.J. Cock, T. Antao, J.T. Chang, B.A. Chapman, C.J. Cox, A. Dalke, I. Friedberg, T. Hamelryck, F. Kauff, B. Wilczynski, M.J. de Hoon, Biopython: freely available Python tools for computational molecular biology and bioinformatics, Bioinformatics 25 (2009) 1422–1423.
[63] K. Clark, I. Karsch-Mizrachi, D.J. Lipman, J. Ostell, E.W. Sayers, GenBank, Nucleic Acids Res. 44 (2016) D67–D72.
[64] F. Fernandes, L. Pereira, A.T. Freitas, CSA: an efficient algorithm to improve circular DNA multiple alignment, BMC Bioinformatics 10 (2009) 230.
[65] T. Nakamura, K.D. Yamada, K. Tomii, K. Katoh, Parallelization of MAFFT for large-scale multiple sequence alignments, Bioinformatics 34 (2018) 2490–2492.
[66] M. Krzywinski, J. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S.J. Jones, M.A. Marra, Circos: an information aesthetic for comparative genomics, Genome Res. 19 (2009) 1639–1645.
[67] T.H. Smits, V.M. Guerrero-Prieto, G. Hernandez-Escarcega, J. Blom, A. Goesmann, F. Rezzonico, B. Duffy, V.O. Stockwell, Whole-genome sequencing of *Erwinia Amylovora* strains from Mexico detects single nucleotide polymorphisms in *rpsL* conferring streptomycin resistance and in the *avrRpt2* effector altering host interactions, Genome Announc (2014) 2.
[68] D.F. Richie, E.J. Klos, Isolation of *Erwinia amylovora* bacteriophage from aerial parts of apple trees, Phytopathology 67 (1977) 101–104.
[69] C.K. Evans, Survey results of *Erwinia Amylovora* in Utah for resistance to streptomycin and investigations comparing kasugamycin (Kasumin) to streptomycin and oxytetracycline for control of fire blight, Acta Hortic. (2008) 433–437.
[70] G.A. Watterson, On the number of segregating sites in genetical models without recombination, Theor. Popul. Biol. 7 (2) (1975) 256–276, https://doi.org/10.1016/0040-5809(75)90020-9.