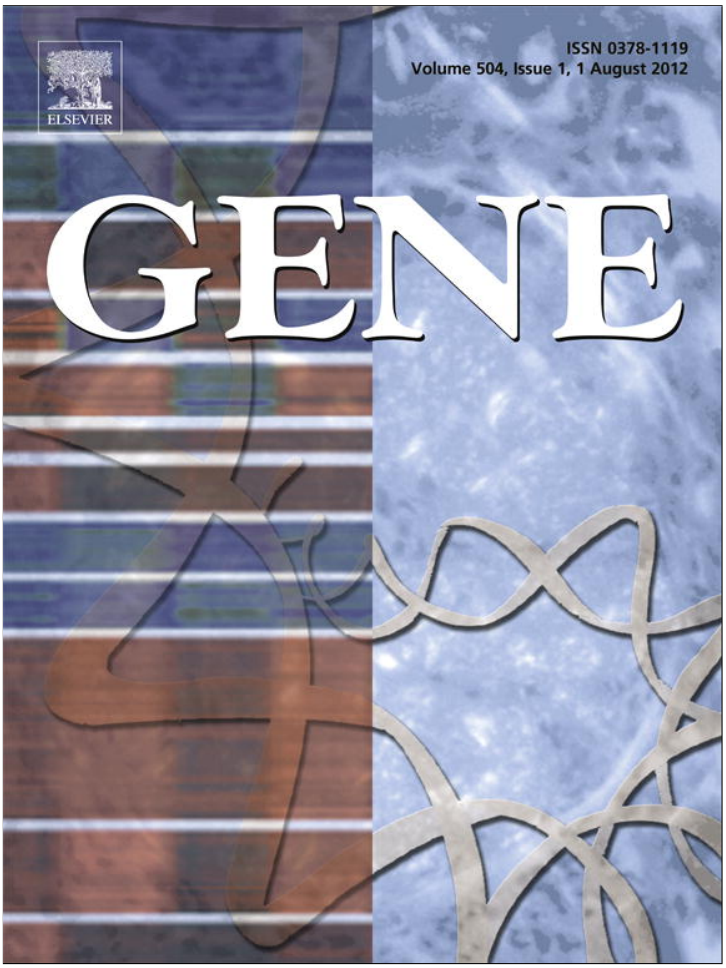

Gene 504 (2012) 6–12

Contents lists available at SciVerse ScienceDirect

Gene

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

Comparative analysis of the Hrp pathogenicity island of *Rubus*- and Spiraeoideae-infecting *Erwinia amylovora* strains identifies the IT region as a remnant of an integrative conjugative element

Rachel A. Mann [a,b,c,d], Jochen Blom [e], Andreas Bühlmann [f], Kim M. Plummer [a,c], Steven V. Beer [b], Joanne E. Luck [a,c,d], Alexander Goesmann [e], Jürg E. Frey [f], Brendan C. Rodoni [a,c,d], Brion Duffy [f], Theo H.M. Smits [f,*]

[a] *Cooperative Research Centre for National Plant Biosecurity, 621 Burwood Hwy, Knoxfield, VIC 3180, Australia*
[b] *Department of Plant Pathology and Plant-Microbe Biology, Cornell University, Ithaca, NY 14853, USA*
[c] *Department of Botany, La Trobe University, Bundoora, VIC 3086, Australia*
[d] *Department of Primary Industries, 621 Burwood Hwy, Knoxfield, VIC 3180, Australia*
[e] *CeBiTec, University of Bielefeld, Bielefeld, Germany*
[f] *Agroscope Changins-Wädenswil ACW, Swiss National Competence Center for Fire Blight, Plant Protection Division, CH-8820 Wädenswil, Switzerland*

ARTICLE INFO

Article history:
Accepted 2 May 2012
Available online 11 May 2012

Keywords:
Fire blight
Hrp type III secretion system
hrpPAI
Genomics

ABSTRACT

The Hrp pathogenicity island (hrpPAI) of *Erwinia amylovora* not only encodes a type III secretion system (T3SS) and other genes required for pathogenesis on host plants, but also includes the so-called island transfer (IT) region, a region that originates from an integrative conjugative element (ICE). Comparative genomic analysis of the IT regions of two Spiraeoideae- and three *Rubus*-infecting strains revealed that the regions in Spiraeoideae-infecting strains were syntenic and highly conserved in length and genetic information, but that the IT regions of the *Rubus*-infecting strains varied in gene content and length, showing a mosaic structure. None of the ICEs in *E. amylovora* strains were complete, as conserved ICE genes and the left border were missing, probably due to reductive genome evolution. Comparison of the hrpPAI region of *E. amylovora* strains to syntenic regions from other *Erwinia* spp. indicates that the hrpPAI and the IT regions are the result of several insertion and deletion events that have occurred within the ICE. It also suggests that the T3SS was present in a common ancestor of the pathoadapted *Erwinia* spp. and that insertion and deletion events in the IT region occurred during speciation.

1. Introduction

The host range of *Erwinia amylovora* includes more than 180 species from 39 genera (Bradbury, 1986; van der Zwet and Keil, 1979). Within *E. amylovora* there are no formal species-specific pathovars; however strains can be divided into two distinct groups on the basis of host specificity (Braun and Hildebrand, 2005). The first group, here referred to as 'Spiraeoideae-infecting' strains, infect a wide range of hosts in the subfamily Spiraeoideae including *Malus* spp. (apple), *Pyrus* spp. (pear), *Crataegus* spp. (hawthorn) and *Cotoneaster* spp. (cotoneaster). The second group, referred to as '*Rubus*-infecting' strains, have a host range limited to plants in the genus *Rubus*, including raspberry and blackberry (Braun and Hildebrand, 2005; Maes et al., 2001; McManus and Jones, 1995). Pathogenicity assays have shown that Spiraeoideae-infecting strains of *E. amylovora* can be weakly pathogenic to raspberry but *Rubus*-infecting strains are not pathogenic to apple (McIntosh and Gala varieties) (Asselin et al., 2011; Braun and Hildebrand, 2005).

Genetic variation has been reported between Spiraeoideae-infecting and *Rubus*-infecting strains with studies consistently separating the two groups (Laby and Beer, 1992; McManus and Jones, 1995; Momol et al., 1997; Rezzonico et al., 2012). The basis of this variation in pathogenicity is still elusive; however, protein profiles secreted in an *hrp* (hypersensitive response and pathogenicity)-inducing medium were distinctly different for Spiraeoideae-infecting and *Rubus*-infecting strains (Braun and Hildebrand, 2005). DNA–DNA hybridization has shown Spiraeoideae-infecting strain Ea110 to share 88% (±2.6%) similarity to *Rubus*-infecting strain IL-5 and 70% (±15.6%) similarity to *Rubus*-infecting strain MR-1 (McGhee et al., 2002). While the *Rubus*-infecting strains were not directly compared, this indicates not only variation between Spiraeoideae-infecting and *Rubus*-infecting stains but also genetic variation among *Rubus*-infecting strains. Spiraeoideae-infecting strains of *E. amylovora* are

Abbreviations: ICE, integrative conjugative element; PAIs, pathogenicity islands; T3SS, type III secretion system; T4SS, type IV secretion system; hrpPAI, Hrp pathogenicity island; HAE, Hrp-associated enzymes; HEE, *hrp* effectors and elicitors; IT region, island transfer region; CDS, coding sequence.

* Corresponding author at: Agroscope Changins-Wädenswil ACW, Swiss National Competence Center for Fire Blight, Plant Protection Division, Schloss 1, CH-8820 Wädenswil, Switzerland. Tel.: +41 44 783 6189; fax: +41 44 783 6305.
E-mail address: theo.smits@acw.admin.ch (T.H.M. Smits).

0378-1119/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2012.05.002

considered to be relatively homogeneous based on subtractive hybridization (Triplett et al., 2006); however, genetic variation has been observed using pulse field gel electrophoresis analysis (Jock et al., 2002; Zhang and Birch, 1997) and clustered regularly interspaced short palindromic repeats (Rezzonico et al., 2011), and differential virulence to apple cultivars has also been observed (Norelli et al., 1984).

In many plant pathogenic bacteria, including *E. amylovora*, the type III secretion system (T3SS) plays an integral role in the transport of effectors into the host cell (Collmer et al., 2009). It comprises a specialized protein-export apparatus that can inject effectors into host cells (Cornelis and Van Gijsegem, 2000) which is relatively conserved across plant pathogenic bacteria (Cornelis, 2006). The genes that encode the T3SS apparatus are generally clustered in genomic islands of pathogenicity genes known as pathogenicity islands (PAIs) (Arnold et al., 2003; Winstanley and Hart, 2001). The genome of the Spiraeoideae-infecting *E. amylovora* strain CFBP 1430 contains three T3SS PAIs (Smits et al., 2010b); however, only one, the hrpPAI, is required for pathogenicity on host plants (Zhao et al., 2009). This and the published hrpPAIs of *E. amylovora* strains ATCC 49946 and Ea321 (Oh and Beer, 2005; Sebaihia et al., 2010) all span 62 kb of genomic DNA that are virtually identical in sequence (Smits et al., 2010b). The genes encoded within are divided into 4 regions: the Hrp-associated enzymes (HAE) region; the *hrp*/*hrc* (*hrp* genes and hypersensitive response conserved genes) region, which encodes the T3SS apparatus and hypersensitive response genes; the *hrp* effectors and elicitors (HEE) region; and the island transfer (IT) region (Oh and Beer, 2005). Comparative genomics and in-depth analysis of the gene content led to the recent hypothesis, that the IT region is a remnant of an insertion event by an integrating conjugative element (ICE), and that it may have entered the genome at a different time point as the Hrp region (Smits et al., 2011).

ICE are a family of mobile genetic elements that vary in size from 10 to 500 kb, encode for mobility proteins, and commonly exhibit mosaic G + C content and codon usage. Typical ICEs are flanked by direct repeats and carry phage-like integrase genes that allow for site specific integration, often into tRNA genes, as well as plasmid-like replication, recombination functions and conjugative machinery that allow them to contribute to horizontal gene transfer (Binnewies et al., 2006; Juhas et al., 2009; Mavrodi et al., 2009; Seth-Smith and Croucher, 2009; Wozniak and Waldor, 2010). Within an ICE there are often genes referred to as 'cargo' genes that are not required by the ICE for self-transmission, but can increase the fitness of the host bacteria (e.g., virulence genes).

The aim of this work was to compare the sequences of the hrpPAI of five *E. amylovora* strains (two Spiraeoideae-infecting strains and three *Rubus*-infecting strains) with each other and with those of two closely related species (*Erwinia pyrifoliae* and *Erwinia piriflorinigrans*). We describe the genetic variation within the IT regions and draw conclusions towards the evolution of the species examined. The ability of different *E. amylovora* strains to infect different hosts is likely to be reflected in their genetic makeup, and thus variability, especially in horizontally acquired regions, may have relevance for pathogenicity and host specificity.

2. Material and methods

2.1. Strain selection

Five complete or draft genome sequences of *E. amylovora* strains were used (Table 1): two from Spiraeoideae-infecting strains (CFBP 1430 and ATCC 49946) and three from *Rubus*-infecting strains (ATCC BAA-2158, Ea644 and MR-1). The draft genome sequences of *E. amylovora* Ea644 and MR-1 were determined in the frame of a larger study, for which details will be published at a later time point (Mann et al., unpublished). Additional published genome sequences from other *Erwinia* spp. were used for comparison (Table 1).

2.2. Genome annotation

Genes were predicted using a combined strategy (McHardy et al., 2004) based on the gene prediction programs Glimmer (Salzberg et al., 1998) and Critica (Badger and Olsen, 1999). Subsequently, the potential function of each predicted gene was automatically assigned using the GenDB annotation pipeline (Meyer et al., 2003). The resulting genome annotation was curated manually, and metabolic pathways were identified using the KEGG pathways (Kanehisa et al., 2002) tool in GenDB.

2.3. Software

Routine sequence manipulations were done using several subroutines of the LASERGENE package (DNASTAR, Madison, WI). Comparative genome analysis of *Erwinia* spp. genome sequences was done using EDGAR (Blom et al., 2009) with the settings as described previously (Smits et al., 2010b). DNA-level sequence comparisons were done using the progressive alignment option of the Mauve comparison software (version 2.3.1 (Darling et al., 2004)).

3. Results and discussion

3.1. Comparison of the hrpPAI of *E. amylovora* strains

Comparison of the genomes of two Spiraeoideae-infecting strains (CFBP 1430, ATCC 49946) showed that the hrpPAI is conserved in both strains (Smits et al., 2010b). The recent sequencing of the genome of the *Rubus*-infecting strain ATCC BAA-2158 (Powney et al., 2011) showed that the IT region of this strain was much larger than that of Spiraeoideae-infecting *E. amylovora* strains. Therefore, a more detailed analysis was done, for which the hrpPAIs of two further *Rubus*-infecting isolates (MR-1 and Ea644) were sequenced.

The hrpPAIs of the three *Rubus*-infecting strains ATCC BAA-2158, Ea644 and MR-1 were all larger than the Spiraeoideae-infecting

Table 1
Genome sequences used in this study.

Species	Strain	Origin	Host	Acc. number	Reference
Erwinia amylovora	CFBP 1430	France, 1972	*Crataegus* sp.	FN434113	Smits et al. (2010b)
	ATCC 49946	NY, USA, 1973	*Malus domestica*	FN666575	Sebaihia et al. (2010)
	ATCC BAA-2158	IL, USA, 1972	*Rubus idaeus*	FR719186[a]	Powney et al. (2011)
	Ea644	MA, USA, 2003	*R. idaeus* cv. Polana	Unpublished	Mann et al., unpublished
	MR-1 (Ea574)	MI, USA	*R. idaeus*	Unpublished	Mann et al., unpublished
Erwinia pyrifoliae	DSM 12163^T	South Korea, 1996	*Pyrus pyrifoliae*	FN392235	Smits et al., (2010a)
	Ep1/96	South Korea, 1996	*P. pyrifoliae*	FP236842	Kube et al. (2010)
Erwinia sp.	Ejp617	Japan	*P. pyrifoliae*	CP002124	Park et al. (2011)
Erwinia piriflorinigrans	CFBP 5888^T	Spain, 2000	*Pyrus communis*	In progress	Smits et al. (submitted for publication)
Erwinia tasmaniensis	Et1/99	Tasmania, Australia, 1999	*Malus domestica*	CU468135	Kube et al. (2008)
Erwinia billingiae	Eb661	United Kingdom	*P. communis*	FP236843	Kube et al. (2010)

[a] Relevant contig only.

strains, and each hrpPAI varied in length, coding sequence (CDS) number and content. The hrpPAI of ATCC BAA-2158 was 93.5 kb in length and contained 95 predicted CDS; the hrpPAI of Ea644 was 95 kb in length and contained 113 predicted CDS, and the hrpPAI of MR-1 was 95 kb in length and contained 108 predicted CDS. The hrpPAIs of Ea644 and MR-1 were the most similar among the *Rubus*-infecting strains sharing 92.5% nucleotide identity.

All strains of *E. amylovora* compared in this study share a high degree of synteny and sequence similarity across the hrpPAI (see supplemental data) except for the IT region which ended with a CDS of an integrase and a tRNA-Phe. The G + C content of the hrpPAI of each strain varied between 52% and 54%, which is consistent with the average G + C content for the *E. amylovora* genome (approx 53.5% G + C) (Smits et al., 2010b). However, the G + C content of the IT regions is highly variable (Oh and Beer, 2005).

In the *Rubus*-infecting strains, the predicted integrase next to the tRNA-Phe at the flank of the hrpPAI encodes a 276 amino acid product sharing 63% sequence identity to a *Yersinia pseudotuberculosis* 32777 complete phage integrase gene. In Spiraeoideae-infecting isolates, this integrase [EAMY_0574-0575] is non-functional due to two frameshift mutations that have produced an early stop codon (Oh and Beer, 2005). None of the *Rubus*-infecting strains possess this frameshift mutation in this integrase.

3.2. Variation in the IT region of E. amylovora strains

The IT regions of Spiraeoideae-infecting strains of *E. amylovora* are highly conserved across the 18 kb region (>99% nucleotide identity and identical synteny), and the major difference observed between the available sequences appears to be the annotation rather than the actual sequence information. This is largely due to differences in the annotation platforms used by different researchers. For this analysis, the sequence of the Spiraeoideae-infecting *E. amylovora* strain CFBP 1430 was used as the type sequence that was annotated on the same platform (GenDB) as the *Rubus*-infecting isolates. For each of sequences of the individual Spiraeoideae-infecting strains, the G + C content of CDS in this region varies between 32% and 65%.

The IT regions of the *Rubus*-infecting strains all vary in sequence identity and length (Fig. 1). The IT region of ATCC BAA-2158 spans 50 kb and encodes 52 CDS whereas the IT regions of Ea644 and MR-1 both span approximately 53 kb and encode 69 and 62 predicted CDS, respectively. The G + C content within the regions, but not between strains, varies as well between 32% and 66%. Annotation of the variable IT regions of all strains contain predicted CDS that are indicative of an ICE, as it contains (most of) the components that qualify this region as such (de la Cruz et al., 2010; Guglielmini et al., 2011).

All strains share sequence identity (97–99% nucleotide identity) for the first 5.5 kb of the IT region containing genes for the transcriptional regulator of levansucrase *rlsA* [EAMY_0559], the chaperone protein encoding *clpB* [EAMY_0560], and several hypothetical genes. As the *rlsA* region is conserved in all *E. amylovora* strains and in other pathoadapted *Erwinia* spp., irrespective of their host, it can be questioned whether this part of the IT region belongs to the ICE, and should rather be included in the HEE region.

Following this 5.5 kb region, there is variation between Spiraeoideae-infecting and *Rubus*-infecting strains but also among *Rubus*-infecting strains (Fig. 1). These variable regions extend for 11.5 kb in the Spiraeoideae-infecting strains, 42.5 kb in ATCC BAA-2158 and 45.5 kb in MR-1 and Ea664. The IT regions regain sequence similarity across all strains in the 2 kb at the 3′ end of the hrpPAI (96–100% nucleotide identity) with the relaxase *rel* and the integrase gene *xerD*, truncated in the Spiraeoideae-infecting strains, and a tRNA-Phe.

No direct repeats were found flanking the variable ICE in the IT region of any *E. amylovora* strain used in this study and therefore no recombination related attachment sites were identified. Searches for repeats (direct, overlapping and palindromic) also yielded no significant matches for the identification of the origin of transfer. Therefore, the IT region of all the *E. amylovora* strains contains coding regions/domains indicative of an integrative conjugative element (ICE), but based on the absence of a recognizable AttL, it can be speculated that none of the ICE are complete.

3.3. Typical components of the integrative conjugative element (ICE) in the E. amylovora IT region

To qualify as an active, self-transmissible ICE the element specifically needs three components i) a relaxase, ii) a type 4 secretion system (T4SS) and iii) a coupling factor (de la Cruz et al., 2010; Guglielmini et al., 2011). *Rubus*-infecting strain ATCC BAA-2158 lacks the T4SS, but has the other two required ICE-related components. The *Rubus*-infecting *E. amylovora* strains Ea644 and MR-1 encode the relaxase (frameshifted in Ea644), the TraD/G coupling factor (frameshifted in Ea644) and the T4SS *pil* gene cluster. The *pil* gene cluster of strain MR-1 is missing the genes encoding the pilin protein PilS, prepilin peptidase PilU, the pilus adhesion protein PilV and the pilus retraction protein PilT. These T4SS genes are, however, found in the ICE of strain Ea644, which encodes 10 *pil* genes, *pilLNOPQRSTUV*, a combination also found in functional ICE including the *Salmonella enterica* subsp. *enterica* ICE CTn*scr94* region (Hochhut et al., 1997), but Ea644 encodes the frameshifted relaxase and coupling factor. The left border and accompanying genes of the ICE could not be identified in the three *Rubus*-infecting *E. amylovora* strains analyzed in this study, and it is therefore doubtful whether the ICE can still be excised from these strains, even when using the required proteins *in trans*. The G + C content of the *pil* cluster in Ea644 (59% G + C) and MR-1 (60% G + C) is distinctly higher than the average G + C content of the *E. amylovora* chromosome (53.5% G + C).

3.4. The variable gene content of the ICE in E. amylovora strains

There are 20 predicted CDS that are common between the IT regions of all *Rubus*-infecting strains but are absent from the Spiraeoideae-infecting strains (some of these CDS are frameshifted and thus probably inactive in individual strains). Most of these CDS encode ICE-specific hypothetical proteins and includes the *ssb* gene encoding a single-stranded DNA-binding protein [EAIL5_0582, MR1_draft2_898 & EA644_draft_833].

A 7 kb region in the ICE of ATCC BAA-2158 has no sequence similarity to genes in any of the other sequenced strains (Fig. 1). Within this region, there are only four CDS including genes that encode a putative type VI secretion system effector protein Hcp [EAIL5_0574] and a remnant of a replication protein [EAIL5_0572] indicating that this region may have originated from a plasmid. Additional singleton ICE genes in the ATCC BAA-2158 genome include the *topB* gene encoding a putative topoisomerase IA [EAIL5_0578] and a putative lipoprotein [EAIL5_0593]. The *Rubus*-infecting strains ATCC BAA-2158 and Ea 644 both encode a predicted ATP-dependent endonuclease [EAIL5_0599 and EA644_draft_873, respectively] that is not present in the third *Rubus*-infecting strain MR-1 or the Spiraeoideae-infecting strains.

3.5. ICE in the hrpPAI of other pathoadapted Erwinia spp.

The genomes of pathoadapted *Erwinia* species *E. pyrifoliae* DSM 12163, *Erwinia* sp. Ejp617, *E. piriflorinigrans* CFBP 5888 and *Erwinia tasmaniensis* Et1/99 all contain a Hrp T3SS (Kamber et al., 2012; Smits et al., 2010b). With the exception of *E. tasmaniensis* Et1/99 (see supplemental data), each species also contains CDS that are characteristic of an ICE (Fig. 1) (de la Cruz et al., 2010; Guglielmini et al., 2011). The ICE in *E. piriflorinigrans* CFBP 5888 contains both the AttL site and the common mobility genes, and is nearly complete in its complement of ICE genes (Mavrodi et al., 2009; Smits et al.,

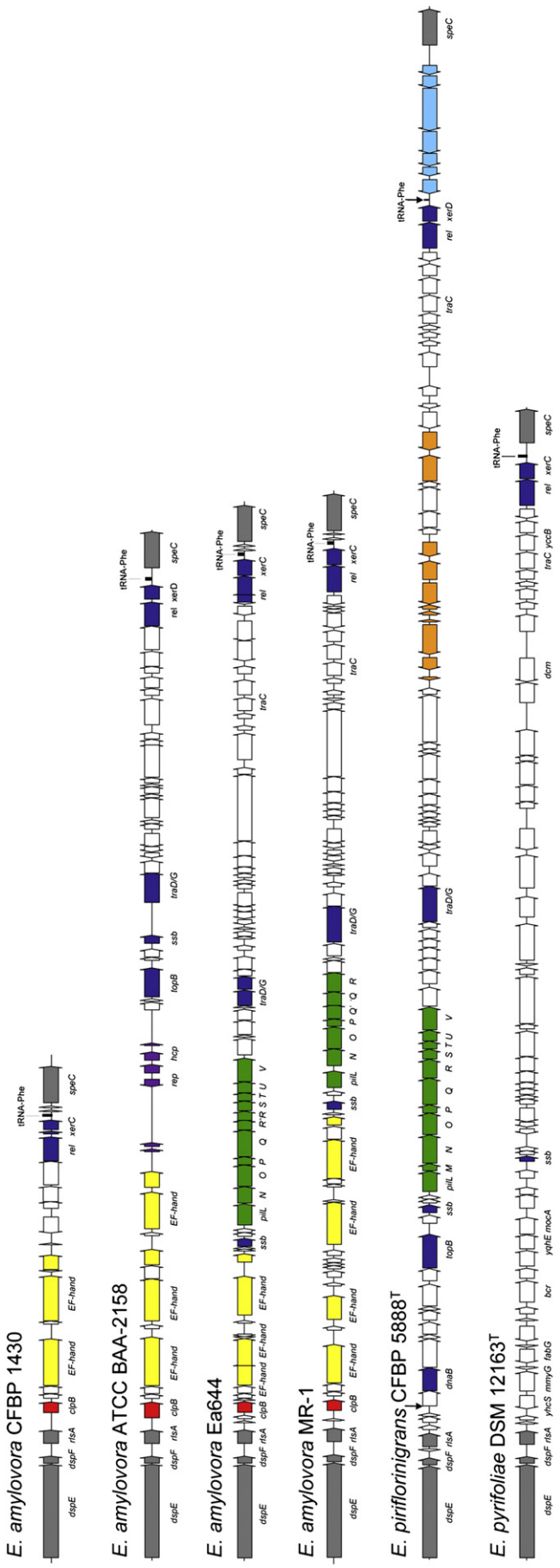

Fig. 1. Graphical representation of the IT regions from different *E. amylovora* strains (CFBP 1430, ATCC BAA-2158, Ea644, and MR-1): *E. piriflorinigrans* CFBP 5888^T, *E. pyrifoliae* DSM 12163^T and *Erwinia* sp. Ejp617. To include some omnipresent reference points, the regions were chosen between *dspE/A* and *speC*. Colors are used to highlight specific features: gray, conserved regions between all genomes; dark blue, characteristic components of the ICE; green, type 4 secretion system *pil*-genes; red, the *clpB* gene; yellow, EF-hand domain protein-coding genes; purple, singleton genes in the ICE of ATCC BAA-2158; orange, singleton genes in the ICE of *E. piriflorinigrans* CFBP 5888^T; light blue, fimbrial gene cluster outside the ICE of *E. piriflorinigrans* CFBP 5888^T. Genes in white can be ICE-related genes, but also singletons. The arrows above the map of *E. piriflorinigrans* CFBP 5888^T indicate the position of the 50-base repeats representing AttL and AttR.

submitted for publication). In contrast, the ICE in *E. pyrifoliae* DSM 12163 lacks several typical features, such as the T4SS pilus genes and the coupling factor (Smits et al., 2010a; Smits et al., 2011). Additionally, this strain has several larger insertions that may represent cargo DNA that could have been inserted afterwards. The reduction of ICE genes is even more pronounced in *Erwinia* sp. Ejp617 (Park et al., 2011), which has lost the majority of the typical ICE genes to a final size of only 12 kb, and contains a frameshifted integrase gene, comparable to that of the Spiraeoideae-infecting strains of *E. amylovora*.

When comparing the near complete ICE of *E. piriflorinigrans* CFBP 5888 with that of all *E. amylovora* strains, the most obvious difference is in the region containing the *clpB* gene and multiple copies of the genes encoding putative EF hand domain proteins (pfam00036) [EAMY_0560-EAMY_0565 and related genes in *Rubus*-infecting strains] (Fig. 1). Based on our comparative genomics approach, it may be hypothesized that this region was acquired after the entry of the ICE (Fig. 2), and that this event may have removed the AttL site and some of the conserved ICE genes, including the *dnaB* gene. This would have rendered the ICE immobile. Subsequently, the removal of inactive remnants has occurred, which has resulted in the loss of smaller parts of the ICE in the *Rubus*-infecting isolates, and large parts in the Spiraeoideae-infecting isolates (Fig. 3). The extent of the individual deletions corresponds to the phylogenetic difference between these three *Rubus*-infecting isolates (Rezzonico et al., 2012) and needs to be examined further using more strains.

3.6. Evolution of the IT region in Erwinia spp.

In contrast to the variable region flanking the tRNA-Phe in the *Rubus*-infecting strains, the Spiraeoideae-infecting *E. amylovora* strains are all monomorphic across this region, containing, as sparse remnants of the ICE, three CDS with ICE-like domains of unknown function, together with a frameshifted integrase (Oh et al., 2005; Smits et al., 2011). The deletions in this region (Fig. 3) would support the hypothesis of reductive genome evolution in *E. amylovora* (Kamber et al., 2012). Whereas in the *Rubus*-infecting strains this process is not as advanced when compared to the Spiraeoideae-infecting strains, the process in the latter seems to have come to a state of complete inactivation that does not pose a metabolic burden anymore to the Spiraeoideae-infecting *E. amylovora* strains.

Comparative genomics with *Erwinia* genomes supports the fact that the insertion of the Hrp cluster is more ancient than the insertion of the IT region (Smits et al., 2011). The genome of the epiphyte *Erwinia billingiae* Eb661 (Kube et al., 2010) lacks all T3SSs present in *E. amylovora* (Smits et al., 2010b); hence it may be concluded that the Hrp T3SS was acquired by the common ancestor of the pathoadapted species *E. amylovora, E. pyrifoliae, E. piriflorinigrans* and *E. tasmaniensis*. The overall genome synteny in *E. billingiae* Eb661 is quite different to that of the pathoadapted *Erwinia* spp., and resembles more that of *Pantoea* spp. (Kamber et al., 2012). It is still difficult to identify the original insertion site of the Hrp genes. The first consistent region ends at EAMY_0498 in *E. amylovora* CFBP 1430, far upstream of the Hrp-region (EAMY_0525-EAMY_0551). The G + C content and GC-skew of the Hrp region resemble that of the complete chromosome of *E. amylovora* CFBP 1430 (Smits et al., 2010b), indicating that adaptive mechanisms may already have been at work, whereas for the IT region, both parameters clearly deviate from the backbone values. This again points to an ancient introduction of the Hrp region in the genome of the pathoadapted *Erwinia* spp.

4. Conclusion

The major differences between the hrpPAIs of the *Rubus*-infecting and Spiraeoideae-infecting strains of *E. amylovora* are within the IT region and seem to be due to the presence of an ICE (or remnants thereof), which is also present in other pathoadapted *Erwinia* spp. Comparative analysis of the *E. amylovora* hrpPAI reveals that the region between *rlsA* and the tRNA-Phe has undergone extensive recombination in *E. amylovora* strains, rather than a single insertion event. Compared to the Spiraeoideae-infecting strains, the three *Rubus*-infecting strains are all variable across this region; longer; mosaic in composition and G + C content; and encode CDS indicative of an ICE. The extensive deletion of regions within the ICE of the Spiraeoideae-infecting strains is indicative of genome reduction, which has been observed in the pathoadapted species of *Erwinia* (Kamber et al., 2012). It may thus be concluded that the hrpPAI as defined earlier (Oh and Beer, 2005) is the result of several insertion and deletion events (Fig. 3), before and during the specialization of the pathoadapted *Erwinia* species (Fig. 2) (Kamber et al., 2012;

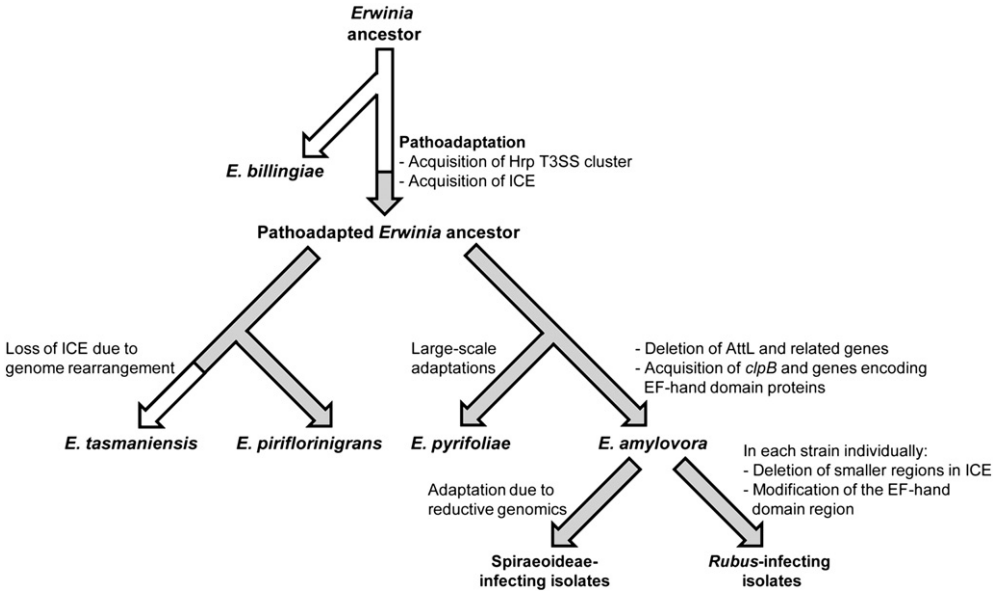

Fig. 2. Genealogy of *Erwinia* spp., combined with the hypothesis for the development of the different ICE regions therein. Gray regions in arrows represent the presence of the ICE in the species.

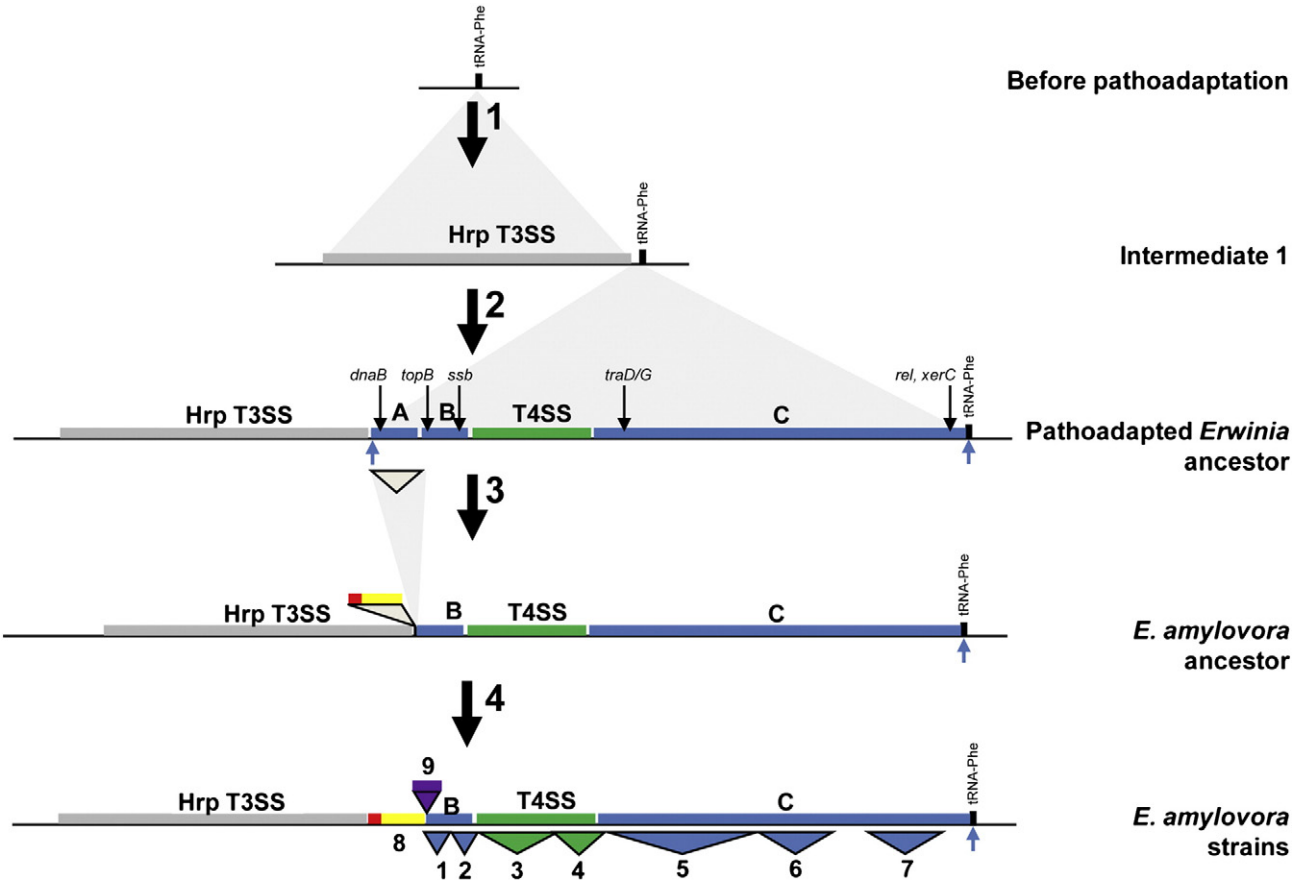

Fig. 3. Hypothesis for the development of the hrpPAI in *Erwinia amylovora* isolates. Step 1: insertion of the Hrp T3SS (gray bar) during pathoadaptation. Step 2: insertion of the ICE (blue and green bars) during pathoadaptation. Genes mentioned in the text are indicated with arrows. Arrows below indicate the AttL and AttR regions. Step 3: deletion of the ICE domain A containing *dnaB* and AttL and replacement by *clpB* (red bar) and the genes encoding EF hand domain proteins (yellow bar) after species separation. Step 4: Reduction and adaptation of the region in individual *E. amylovora* strains. All Spiraeoideae-infecting isolates have deleted regions 1–7 to yield the current situation (see Fig. 1). *Rubus*-infecting isolate ATCC BAA-2158 has deletions in regions 3, 4 and 6, the insertion of region 9 and modifications in region 8. *Rubus*-infecting isolate Ea644 has deletions in region 1, and smaller deletions in the regions 5–7, while *Rubus*-infecting isolate MR-1 has deletions in regions 1, 4 and 7, and modifications in region 8.

Smits et al., 2011), and that the name IT (island transfer) region thus has to be revised.

Acknowledgments

The authors would like to acknowledge the support of the Australian Government's Cooperative Research Centre for National Plant Biosecurity, Horticulture Australia Limited, a special grant provided by the USDA CSREES for research on fire blight in New York, the European Union ESF-FP7-KBBE Project Q-Detect (no. 245047), the Swiss Secretariat for Education and Research (SER-Nr. C09.0029), and the Swiss Federal Office of Agriculture (BLW Nr. 08.02). It was conducted in part within the European Science Foundation funded research network COST Action 864.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gene.2012.05.002.

References

Arnold, D.L., Pitman, A., Jackson, R.W., 2003. Pathogenicity and other genomic islands in plant pathogenic bacteria. Mol. Plant Pathol. 4, 407–420.
Asselin, J.A., Bonasera, J., Kim, J., Oh, C.-S., Beer, S.V., 2011. Eop1 from a *Rubus* strain of *Erwinia amylovora* functions as a host-range limiting factor. Phytopathology 101, 935–944.
Badger, J.H., Olsen, G.J., 1999. CRITICA: coding region identification tool invoking comparative analysis. Mol. Biol. Evol. 16, 512–524.
Binnewies, T.T., et al., 2006. Ten years of bacterial genome sequencing: comparative-genomics-based discoveries. Funct. Integr. Genomics 6.
Blom, J., et al., 2009. EDGAR: a software framework for the comparative analysis of prokaryotic genomes. BMC Bioinforma. 10, 154.
Bradbury, J.F., 1986. Guide to Plant Pathogenic Bacteria. CAB International, Wallingford (GB).
Braun, P.G., Hildebrand, P.D., 2005. Infection, carbohydrate utilization, and protein profiles of apple, pear, and raspberry isolates of *Erwinia amylovora*. Can. J. Plant Pathol. 27, 338–346.
Collmer, A., Schneider, D.J., Lindeberg, M., 2009. Lifestyles of the effector rich: genome-enabled characterization of bacterial plant pathogens. Plant Physiol. 150, 1623–1630.
Cornelis, G.R., 2006. The type III secretion injectisome. Nat. Rev. Microbiol. 4, 811–825.
Cornelis, G.R., Van Gijsegem, F., 2000. Assembly and function of type III secretory systems. Annu. Rev. Microbiol. 54, 735–774.
Darling, A.C.E., Mau, B., Blattner, F.R., Perna, N.T., 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 14, 1394–1403.
de la Cruz, F., Frost, L.S., Meyer, R.J., Zechner, E.L., 2010. Conjugative DNA metabolism in Gram-negative bacteria. FEMS Microbiol. Rev. 34, 18–40.
Guglielmini, J., Quintais, L., Garcillán-Barcia, M.P., de la Cruz, F., Rocha, E.P.C., 2011. The repertoire of ICE in prokaryotes underscores the unity, diversity and ubiquity of conjugation. PLoS Genet. 7, e1002222.
Hochhut, B., Jahreis, K., Lengeler, J.W., Schmid, K., 1997. CTn*scr94*, a conjugative transposon found in Enterobacteria. J. Bacteriol. 179, 2097–2102.
Jock, S., Donat, V., López, M.M., Bazzi, C., Geider, K., 2002. Following spread of fire blight in Western, Central and Southern Europe by molecular differentiation of *Erwinia amylovora* strains with PFGE analysis. Environ. Microbiol. 4, 106–114.
Juhas, M., Van der Meer, J.R., Gaillard, M., Harding, R.M., Hood, D.W., Crook, D.W., 2009. Genomic islands: tools of bacterial horizontal gene transfer and evolution. FEMS Microbiol. Rev. 33, 376–393.
Kamber, T., Smits, T.H.M., Rezzonico, F., Duffy, B., 2012. Genomics and current genetic understanding of *Erwinia amylovora* and the fire blight antagonist *Pantoea vagans*. Trees Struct. Funct. 26, 227–238.

Kanehisa, M., Goto, S., Kawashima, S., Nakaya, A., 2002. The KEGG databases at GenomeNet. Nucleic Acids Res. 30, 42–46.
Kube, M., et al., 2008. The genome of *Erwinia tasmaniensis* strain Et1/99, a non-pathogenic bacterium in the genus *Erwinia*. Environ. Microbiol. 10, 2211–2222.
Kube, M., et al., 2010. Genome comparison of the epiphytic bacteria *Erwinia billingiae* and *E. tasmaniensis* with the pear pathogen *E. pyrifoliae*. BMC Genomics 11, 393.
Laby, R.J., Beer, S.V., 1992. Hybridization and functional complementation of the hrp gene cluster from *Erwinia amylovora* strain Ea321 with DNA from other bacteria. Mol. Plant Microbe Interact. 5, 412–419.
Maes, M., et al., 2001. Influence of amylovoran production on virulence of *Erwinia amylovora* and a different amylovoran structure in *E. amylovora* isolates from *Rubus*. Eur. J. Plant Pathol. 107, 839–844.
Mavrodi, D.V., Loper, J.E., Paulsen, I.T., Thomashow, L.S., 2009. Mobile genetic elements in the genome of the beneficial rhizobacterium *Pseudomonas fluorescens* Pf-5. BMC Microbiol. 9, 8.
McGhee, G.C., et al., 2002. Relatedness of chromosomal and plasmid DNAs of *Erwinia pyrifoliae* and *Erwinia amylovora*. Appl. Environ. Microbiol. 68, 6182–6192.
McHardy, A.C., Goesmann, A., Pühler, A., Meyer, F., 2004. Development of joint application strategies for two microbial gene finders. Bioinformatics 20, 1622–1631.
McManus, P.S., Jones, A.L., 1995. Genetic fingerprinting of *Erwinia amylovora* strains isolated from tree-fruit crops and *Rubus* spp. Phytopathology 85, 1547–1553.
Meyer, F., et al., 2003. GenDB—an open source genome annotation system for prokaryote genomes. Nucleic Acids Res. 31, 2187–2195.
Momol, M.T., Momol, E.A., Lamboy, W.F., Norelli, J.L., Beer, S.V., Aldwinckle, H.S., 1997. Characterization of *Erwinia amylovora* strains using random amplified polymorphic DNA fragments (RAPDs). J. Appl. Microbiol. 82, 389–398.
Norelli, J.L., Aldwinckle, H.S., Beer, S.V., 1984. Differential host × pathogen interactions among cultivars of apple and strains of *Erwinia amylovora*. Phytopathology 74, 136–139.
Oh, C.-S., Beer, S.V., 2005. Molecular genetics of *Erwinia amylovora* involved in the development of fire blight. FEMS Microbiol. Lett. 253, 185–192.
Oh, C.-S., Kim, J.F., Beer, S.V., 2005. The Hrp pathogenicity island of *Erwinia amylovora* and identification of three novel genes required for systemic infection. Mol. Plant Pathol. 6, 125–138.
Park, D.H., et al., 2011. Complete genome sequence of Japanese *Erwinia* strain Ejp617, a bacterial shoot blight pathogen of pear. J. Bacteriol. 193, 586–587.
Powney, R., et al., 2011. Genome sequence of an *Erwinia amylovora* strain with restricted pathogenicity to *Rubus* plants. J. Bacteriol. 193, 785–786.
Rezzonico, F., Smits, T.H.M., Duffy, B., 2011. Diversity, evolution and functionality of clustered regularly interspaced short palindromic repeat (CRISPR) regions in fire blight pathogen *Erwinia amylovora*. Appl. Environ. Microbiol. 77, 3819–3829.
Rezzonico, F., Braun-Kiewnick, A., Mann, R.A., Goesmann, A., Rodoni, B., Duffy, B. and Smits, T.H.M., 2012. Lipopolysaccharide biosynthesis genes discriminate between *Rubus*- and Spiraeoideae-infective genotypes of *Erwinia amylovora*. Mol. Plant. Pathol. doi:10.1111/J.1364-3703.2012.00807.x.
Salzberg, S.L., Delcher, A.L., Kasif, S., White, O., 1998. Microbial gene identification using interpolated Markov models. Nucleic Acids Res. 26, 544–548.
Sebaihia, M., et al., 2010. Complete genome sequence of the plant pathogen *Erwinia amylovora* strain ATCC 49946. J. Bacteriol. 192, 2020–2021.
Seth-Smith, H., Croucher, N.J., 2009. Breaking the ICE. Nat. Rev. Microbiol. 7, 328–329.
Smits, T.H.M., et al., 2010a. Complete genome sequence of the fire blight pathogen *Erwinia pyrifoliae* DSM 12163^T and comparative genomic insights into plant pathogenicity. BMC Genomics 11, 2.
Smits, T.H.M., et al., 2010b. Complete genome sequence of the fire blight pathogen *Erwinia amylovora* CFBP 1430 and comparison to other *Erwinia* spp. Mol. Plant Microbe Interact. 23, 384–393.
Smits, T.H.M., Rezzonico, F., Duffy, B., 2011. Evolutionary insights from *Erwinia amylovora* genomics. J. Biotechnol. 155, 34–39.
Smits, T.H.M., Rezzonico, F., López, M.M., Blom, J., Goesmann, A., Frey, J.E. and Duffy, B., submitted for publication. Comparative analysis of the genome of *Erwinia piriflorinigrans* CFBP 5888^T reveals its pathogenic potential and its phylogenetic position within the *Erwinia* species pathogenic to *Spiraeoideae*. BMC Genomics.
Triplett, L.R., Zhao, Y., Sundin, G.W., 2006. Genetic differences between blight-causing *Erwinia* species with differing host specificities, identified by suppression subtractive hybridization. Appl. Environ. Microbiol. 72, 7359–7364.
van der Zwet, T., Keil, H.L., 1979. Fire Blight: A Bacterial Disease of Rosaceous Plants. U.S. Govt. Print, Washington (US).
Winstanley, C., Hart, C.A., 2001. Type III secretion systems and pathogenicity islands. J. Med. Microbiol. 50, 116–126.
Wozniak, R.A.F., Waldor, M.K., 2010. Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. Nat. Rev. Microbiol. 8, 552–563.
Zhang, L., Birch, R.G., 1997. The gene for albicidin detoxification from *Pantoea dispersa* encodes an esterase and attenuates pathogenicity of *Xanthomonas albilineans* to sugarcane. Proc. Natl. Acad. Sci. U. S. A. 94, 9984–9989.
Zhao, Y., Sundin, G.W., Wang, D., 2009. Construction and analysis of pathogenicity island deletion mutants of *Erwinia amylovora*. Can. J. Microbiol. 55, 457–464.