Evolutionary insights from *Erwinia amylovora* genomics

Theo H.M. Smits*, Fabio Rezzonico, Brion Duffy

Agroscope Changins-Wädenswil ACW, Division of Plant Protection, Swiss National Competence Center for Fire Blight, CH-8820 Wädenswil, Switzerland

ABSTRACT

Evolutionary genomics is coming into focus with the recent availability of complete sequences for many bacterial species. A hypothesis on the evolution of virulence factors in the plant pathogen *Erwinia amylovora*, the causative agent of fire blight, was generated using comparative genomics with the genomes *E. amylovora*, *Erwinia pyrifoliae* and *Erwinia tasmaniensis*. Putative virulence factors were mapped to the proposed genealogy of the genus *Erwinia* that is based on phylogenetic and genomic data. Ancestral origin of several virulence factors was identified, including levan biosynthesis, sorbitol metabolism, three T3SS and two T6SS. Other factors appeared to have been acquired after divergence of pathogenic species, including a second flagellar gene and two glycosyltransferases involved in amylovoran biosynthesis. *E. amylovora* singletons include 3 unique T3SS effectors that may explain differential virulence/host ranges. *E. amylovora* also has a unique T1SS export system, and a unique third T6SS gene cluster. Genetic analysis revealed signatures of foreign DNA suggesting that horizontal gene transfer is responsible for some of these differential features between the three species.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Erwinia amylovora causes fire blight, the most important threat to pome fruit production (i.e., apple, pear, and quince) globally, and it affects a wide-variety of rosaceous plants (Duffy et al., 2005; Norelli et al., 2003). *E. amylovora* is native to the North Eastern USA, and it was the first phytopathogenic bacterium described ever (Bonn and van der Zwet, 2000). *E. amylovora* has quarantine status outside North America, and it is a controversial issue in fruit trade with fire blight-free countries like Australia, Japan and all areas of the Southern Hemisphere aside from New Zealand (Calvin and Krissoff, 1998; Roberts et al., 1998).

Fire blight symptoms include wilting of flowers and shoots (shepherd's crook), necrosis, release of ooze, and cankers. The pathogen enters through nectaries in flowers, hydathodes and wounds resulting in blossom, shoot or rootstock blight symptoms (Thomson, 2000). *E. amylovora* growth is highly dependent on weather conditions, which has been used to develop disease forecasting models, and it is passively vectored by flower-foraging insects and rain (Johnson and Stockwell, 1998). Epidemics can develop rapidly and result in death of individual plants or entire orchards within a single season, leading to severe economic losses (Bonn and van der Zwet, 2000).

Although important insights have been acquired regarding this important phytopathogenic bacterium, much remains uncertain about the genetics of *E. amylovora* (Oh and Beer, 2005). The recent publication of the genomes of three species from the genus *Erwinia* [*E. amylovora* CFBP 1430 (Smits et al., 2010b), *E. pyrifoliae* DSM 12163 (Smits et al., 2010a) and *E. tasmaniensis* Et1/99 (Kube et al., 2008a)] provides basic scientific information to infer the evolution and relatedness of the species within this genus.

2. Evolution of species within the genus *Erwinia*: a genome-based hypothesis

We compared complete genomes of three *Erwinia* species (Smits et al., 2010b) in order (1) to resolve the *Erwinia* core genome, and (2) to identify differences that clarify the basis of *E. amylovora* pathogen success in contrast to *E. pyrifoliae* which has a considerably more limited host range on pome fruits (Kim et al., 1999), and *E. tasmaniensis* Et1/99 which is a pome fruit epiphyte and is not reported to be a phytopathogen of other species (Geider et al., 2006).

The pan-genome of *Erwinia* contains 5650 CDS, of which 2474 (43%) are core CDS (Smits et al., 2010b). According to the definition (Medini et al., 2005), a core genome includes all genes responsible for the basic aspects of the biology of the genus and its major phenotypic traits. In the dispensable genome, genes that confer a selective advantage are included (Medini et al., 2005). Notably, a large number of genes in the dispensable genomes of the *Erwinia* spp. encode small hypothetical proteins that have no counterparts in any other genome, and may in fact not encode functional proteins (Smits et al., 2010b). It should be noted that, in comparison to other Enterobacterial genomes, the sequenced genomes of the three *Erwinia* spp. are smaller, possibly due to genome erosion.

* Corresponding author. Tel.: +41 44 783 6189; fax: +41 44 783 6305.
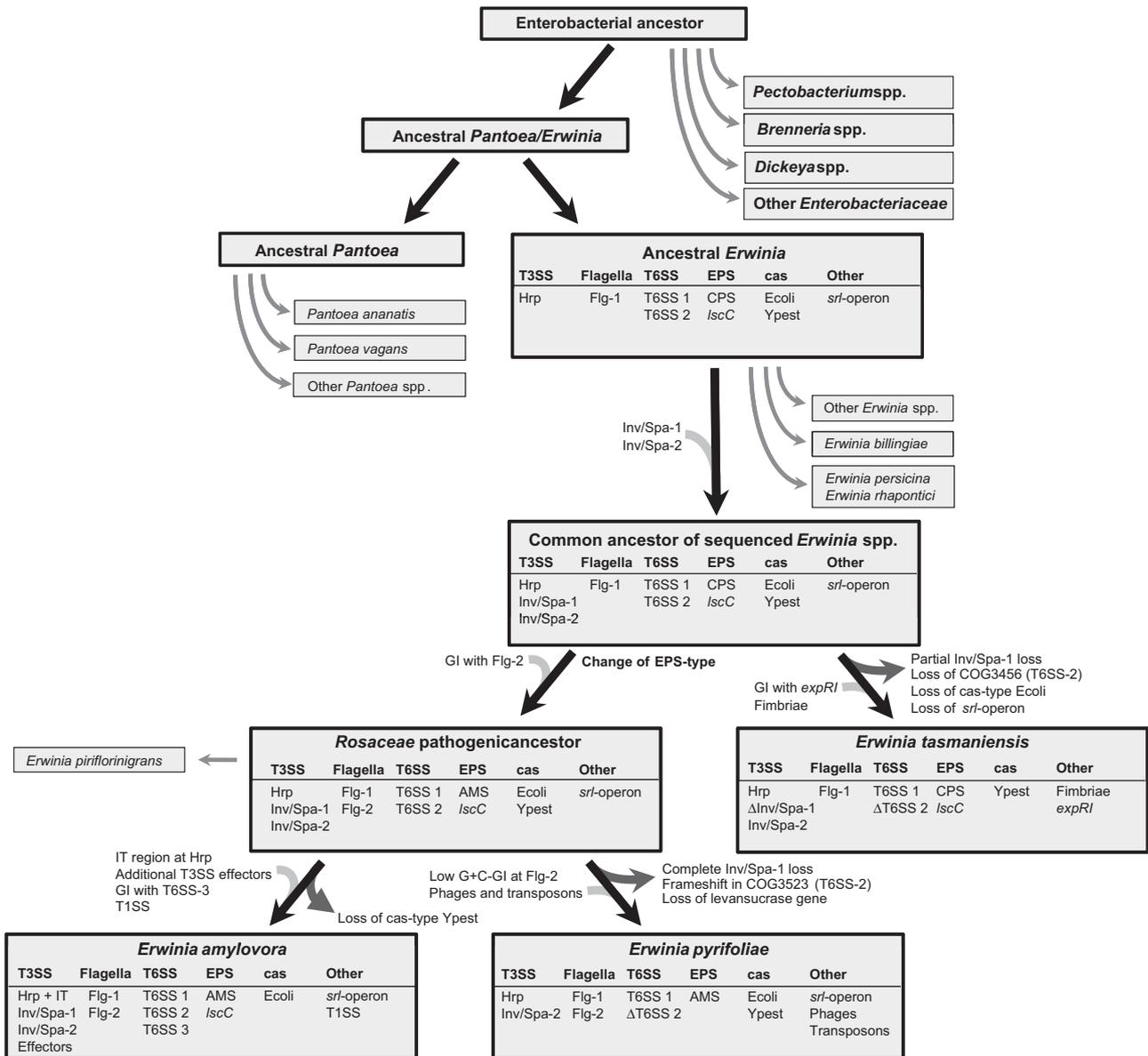*E-mail address:* theo.smits@acw.admin.ch (T.H.M. Smits).

Fig. 1. Proposed genealogy of *Erwinia* spp. and putative virulence factors therein based on genome sequence information. The delineation of the different tree structure follows below.

Enterobacterial ancestor → **Ancestral *Pantoea/Erwinia*** (grey arrows to: *Pectobacterium* spp., *Brenneria* spp., *Dickeya* spp., Other *Enterobacteriaceae*)

Ancestral *Pantoea* (grey arrows to: *Pantoea ananatis*, *Pantoea vagans*, Other *Pantoea* spp.)

Ancestral *Erwinia*

T3SS	Flagella	T6SS	EPS	cas	Other
Hrp	Flg-1	T6SS 1	CPS	Ecoli	*srl*-operon
		T6SS 2	*lscC*	Ypest	

Inv/Spa-1, Inv/Spa-2 (grey arrows to: Other *Erwinia* spp., *Erwinia billingiae*, *Erwinia persicina* / *Erwinia rhapontici*)

Common ancestor of sequenced *Erwinia* spp.

T3SS	Flagella	T6SS	EPS	cas	Other
Hrp	Flg-1	T6SS 1	CPS	Ecoli	*srl*-operon
Inv/Spa-1		T6SS 2	*lscC*	Ypest	
Inv/Spa-2					

GI with Flg-2, **Change of EPS-type**

GI with *expRI*, Fimbriae; Partial Inv/Spa-1 loss, Loss of COG3456 (T6SS-2), Loss of cas-type Ecoli, Loss of *srl*-operon

Rosaceae pathogenic ancestor → *Erwinia piriflorinigrans*

T3SS	Flagella	T6SS	EPS	cas	Other
Hrp	Flg-1	T6SS 1	AMS	Ecoli	*srl*-operon
Inv/Spa-1	Flg-2	T6SS 2	*lscC*	Ypest	
Inv/Spa-2					

Erwinia tasmaniensis

T3SS	Flagella	T6SS	EPS	cas	Other
Hrp	Flg-1	T6SS 1	CPS	Ypest	Fimbriae
ΔInv/Spa-1		ΔT6SS 2	*lscC*		*expRI*
Inv/Spa-2					

IT region at Hrp, Additional T3SS effectors, GI with T6SS-3, T1SS; Loss of cas-type Ypest

Low G+C-GI at Flg-2, Phages and transposons; Complete Inv/Spa-1 loss, Frameshift in COG3523 (T6SS-2), Loss of levansucrase gene

Erwinia amylovora

T3SS	Flagella	T6SS	EPS	cas	Other
Hrp + IT	Flg-1	T6SS 1	AMS	Ecoli	*srl*-operon
Inv/Spa-1	Flg-2	T6SS 2	*lscC*		T1SS
Inv/Spa-2		T6SS 3			
Effectors					

Erwinia pyrifoliae

T3SS	Flagella	T6SS	EPS	cas	Other
Hrp	Flg-1	T6SS 1	AMS	Ecoli	*srl*-operon
Inv/Spa-2	Flg-2	ΔT6SS 2		Ypest	Phages
					Transposons

Fig. 1. Proposed genealogy of *Erwinia* spp. and putative virulence factors therein based on genome sequence information. The delineation of mentioned genera and/or species is indicated by grey arrows. The "soft rot *Erwinia*" genera *Pectobacterium*, *Dickeya* and *Brenneria* (Hauben and Swings, 2005) and other *Enterobacteriaceae* separate earlier in evolution from the *Pantoea/Erwinia* lineage (Rezzonico et al., 2009). Insertions and deletions of putative virulence factors between the levels within the *Erwinia* ancestry are indicated. *Abbreviations:* T3SS: type III secretion system; Inv/Spa: T3SS related to that of *Salmonella enterica*; Flg: flagellum; T6SS: type VI secretion system; EPS: exopolysaccharide; CPS: stewartan-like exopolysaccharide; AMS: amylovoran; *lscC*: levansucrase gene; cas: subtype of *cas*-genes according to Haft et al. (2005); *srl*-operon: sorbitol metabolism operon; *expRI*: genes encoding autoinducer-2 biosynthesis; GI: genomic island.

A genealogy, based on the genes present in the core or dispensable genomes, was generated (Fig. 1) for the origin and evolution of the three genome-sequenced species from the genus *Erwinia*. At the separation of the Enterobacterial ancestor in many species, a common ancestor for the genera *Erwinia* and *Pantoea* was formed (Rezzonico et al., 2009). This ancestor split then in the two genera, from which the different species originate (Grimont and Grimont, 2005; Hauben and Swings, 2005). The former "soft rot *Erwinia* species", nowadays represented by the genera *Brenneria*, *Pectobacterium* and *Dickeya* (Hauben et al., 1998; Hauben and Swings, 2005; Samson et al., 2005), the latter two for which genome sequences are available, must have separated before. Whereas the genomes of *Erwinia* spp. (Kube et al., 2008a; Smits et al., 2010a,b) and *Pantoea* spp. (De Maayer et al., 2010; Smits et al., in press, 2010c) are quite comparable in their gene content and percentage identity between core genes (Hauben et al.,

1998; Rezzonico et al., 2009), the genomes of *Pectobacterium* spp. (Bell et al., 2004; Glasner et al., 2008; Toth et al., 2006) and *Dickeya* spp. (genome sequences available at GenBank, but unpublished) are more distantly related based on the identity of core genes.

The genealogy within the genus *Erwinia* agrees largely with the phylogeny (Hauben and Swings, 2005) (Fig. 1), but introduces additional levels of differentiation in the tree. The ancestral *Erwinia* may have had several steps in the evolution until it reached the common ancestor for the three genome-sequenced *Erwinia*, thereby separating from all other *Erwinia* spp. From this common ancestor, the non-pathogenic *E. tasmaniensis* separated from the pathogenic *Erwinia* ancestor, that later differentiated in the species *E. amylovora* and *E. pyrifoliae*. The hypothesis is supported by the levels of gene identity between the individual species (Smits et al., 2010b), and will be substantiated below.

A question that cannot be answered until genome sequences are available is the exact position of *Rubus* isolates of *E. amylovora* and the so-called Japanese *Erwinia* strains. *Rubus* isolates are commonly included in *E. amylovora*, although genomic differences to pome fruit isolates were found (Triplett et al., 2006). The Japanese *Erwinia* spp. were recently assigned to *E. pyrifoliae* based on a polyphasic approach (Geider et al., 2009). Additionally, the position of all other *Erwinia* spp. (Hauben and Swings, 2005) is still unresolved (Fig. 1). The genealogy leaves space for them to separate between the level of the ancestral *Erwinia* and the common ancestor of the sequenced *Erwinia* spp. For the novel species *E. piriflorinigrans* that causes necrosis on pear blossoms (López et al., in press; Roselló et al., 2006), a separation close to the level of the pathogenic *Erwinia* ancestor can be anticipated, as it is more closely related to *E. amylovora* and *E. pyrifoliae*.

3. Type III secretion systems

In the genome sequence of *E. amylovora* CFBP 1430, three complete Type III Secretion Systems (T3SS) were found: the "classical" Hrp T3SS involved in phytopathogenicity (Oh et al., 2005), and two Inv/Spa T3SSs on genomic islands PAI-2 and PAI-3 with high homology with the insect–pathogen *S. glossinidius* str. *morsidans* and to the mammalian pathogens *Salmonella* and *Yersinia* (Smits et al., 2010b; Zhao et al., 2009). The latter two systems appear non-functional in virulence of *E. amylovora* on both apple seedlings and in immature pear tests (Zhao et al., 2009). Both PAI-2 and PAI-3 have a significantly lower G+C content as the host, and may be acquired before the common ancestor of the three genome-sequenced *Erwinia* spp. After the separation of *E. tasmaniensis*, the insertion of a second mobile element has deleted a part of PAI-2 in *E. tasmaniensis* Et1/99 (Kube et al., 2008a), whereas PAI-2 has been excised and lost in *E. pyrifoliae* as seen in the genome of DSM 12163[T] (Smits et al., 2010a).

In contrast to PAI-2 and PAI-3, the G+C content and GC skew of the Hrp T3SS system resembles that of the host, contradicting the current assumption that it has also been acquired (Oh and Beer, 2005; Oh et al., 2005). This assumption was originally based on the finding of a region with variable G+C-content directly flanking the Hrp T3SS (Oh et al., 2005), which in fact may constitute a remnant of an Integrating Conjugative Element (ICE) (Fig. 2) (Gaillard et al., 2006; Mohd-Zain et al., 2004).

In *E. amylovora* CFBP 1430, some T3SS effectors (*eop2*, *hopPtoC* and *avrRpt2*) were identified, that all were singletons (Smits et al., 2010b). These genes must have been acquired after the separation of *E. amylovora* and *E. pyrifoliae*, and may play a role in the stronger virulence of *E. amylovora* on apple trees (Zhao et al., 2006).

4. Flagella

Two sets of genes encoding flagellar biosynthesis and chemotaxis related proteins were found in the genome of *E. amylovora* CFBP 1430 (Smits et al., 2010b): a complete gene set spread over four gene clusters (Flg-1), and a second set contained in a single gene cluster (Flg-2). The operons encoding Flg-1 closely resemble the operons for biosynthesis of flagella found in other Enterobacteria and is spread over the genome in all *Erwinia* spp. while Flg-2 matches an analogous region found in *E. pyrifoliae* DSM 12163[T] (Smits et al., 2010a). Flg-1 can be considered ancestral, whereas Flg-2 is acquired at the level of the pathogenic ancestor for *E. amylovora* and *E. pyrifoliae*.

With respect to *E. amylovora* CFBP 1430, the region upstream of Flg-2 differs in *E. pyrifoliae* DSM 12163[T] where the insertion of a low-G+C genomic island was obvious (Smits et al., 2010a,b). We propose that the low-G+C genomic island in *E. pyrifoliae* was only inserted at the establishment of *E. pyrifoliae* as separate species.

Flg-2 is absent in *E. tasmaniensis* Et1/99, which contains amongst others the quorum-sensing (QS) genes *expRI* at the equivalent position in its genome (Kube et al., 2008a). Both *E. amylovora* and *E. pyrifoliae* lack a QS signal-generating enzyme (Rezzonico and Duffy, 2008; Smits et al., 2010a,b).

5. Type VI secretion systems

Three Type VI secretion system (T6SS) gene clusters were identified in the genome of *E. amylovora* CFBP 1430 [see supplemental Figure 2 in Smits et al., 2010b]. T6SS cluster 1 is shared with *E. pyrifoliae* DSM 12163[T] and *E. tasmaniensis* Et1/99, with the exception of some genes encoding hypothetical proteins that do not belong to the core genes of T6SSs (Bingle et al., 2008) and some differences in the genes encoding VgrG effector proteins. Compared to the small T6SS cluster 2 in *E. amylovora* CFBP 1430, a frameshift is found in one of the genes in *E. pyrifoliae* DSM 12163[T] and a gene is lost in the corresponding cluster in *E. tasmaniensis* Et1/99 (Smits et al., 2010a,b). Both clusters can be considered ancestral, and the modifications must have been introduced at the level of species differentiation. T6SS cluster 3 is solely identified in *E. amylovora* CFBP 1430 (Smits et al., 2010b), and might be acquired at the level of species differentiation. For T6SS cluster 3, some genes encoding core proteins appear missing, that could be complemented by the genes present in T6SS cluster 2.

6. Amylovoran biosynthesis

The exopolysaccharide (EPS) amylovoran, a pathogenicity factor contributing to biofilm formation of *E. amylovora* (Koczan et al., 2009), is produced by both *E. amylovora* and *E. pyrifoliae* (Bernhard et al., 1996; Kube et al., 2008a), but not by *E. tasmaniensis* (Kim et al., 2002). This is reflected in the genetic makeup of the gene cluster responsible for the biosynthesis, in which two glycosyltransferases (*amsD* and *amsE*) in the *ams* cluster of *E. amylovora* CFBP 1430 were found, while two different glycosyltransferases are present in the *cps* cluster in *E. tasmaniensis* Et1/99 (Smits et al., 2010a). The latter two genes rather resemble the glycosyltransferases of *Pantoea stewartii* pv. *stewartii* DC283 (Coplin et al., 1996).

It is difficult to establish the time point of exopolysaccharide differentiation within the evolution of *Erwinia* spp. with the current data set. It can be suggested that the *Erwinia* ancestor originally produced an EPS resembling stewartan of *P. stewartii* (Coplin et al., 1996; Kube et al., 2008b), but that the differentiation of EPS production has taken place at or after the separation of the pathogenic *Erwinia* from *E. tasmaniensis*. This would indicate that the genes involved in amylovoran production (Bernhard et al., 1996) are rather acquired.

7. CRISPR-associated genes

Clustered regularly interspaced short palindromic repeats (CRISPRs) and CRISPR-associated sequence (Cas) proteins constitute a putative prokaryotic RNA-interference-based immune system (Makarova et al., 2006; Sorek et al., 2008) protecting against bacteriophages or plasmids (Mojica et al., 2005). The annotation of the genome of *E. amylovora* CFBP 1430 revealed eight genes with homology to *cas*-genes (Smits et al., 2010b) of the Ecoli subtype (Haft et al., 2005; Rezzonico et al., submitted for publication). Orthologues thereof were detected in the genome of *E. pyrifoliae* DSM 12163[T], but not in *E. tasmaniensis* Et1/99. The latter strain contains *cas*-genes of the Ypest subtype (Haft et al., 2005; Rezzonico et al., submitted for publication), and orthologues of these were found as well in *E. pyrifoliae* DSM12163[T].
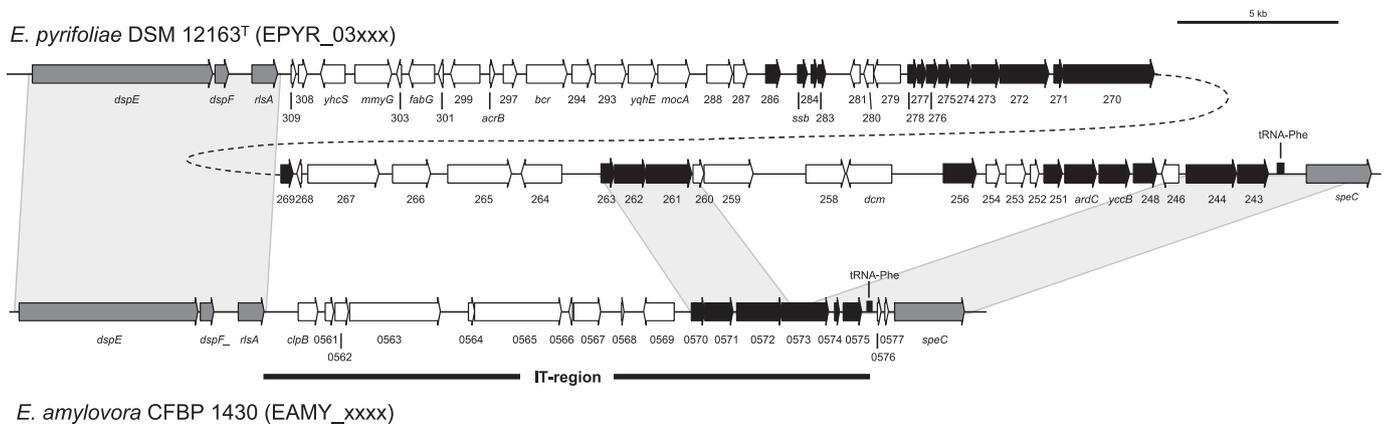
Fig. 2. Comparison of the island transfer (IT) region in *E. amylovora* CFBP 1430 with the integrative conjugative element (ICE) in *E. pyrifoliae* DSM 12163[T], flanking the Hrp type III secretion system. Numbers below the genes indicate locus tag numbers (locus tag prefix at the organism name). Genes indicated in grey are border regions of the insertions, while conserved ICE genes are indicated black. Orthologous genes between *E. amylovora* CFBP 1430 and *E. pyrifoliae* DSM 12163[T] are shaded.

It is very well possible that the presence of multiple CRISPR/Cas subtypes in *E. pyrifoliae* DSM 12163[T] represents the ancestral situation of the true *Erwinia* spp., and that both *E. amylovora* CFBP 1430 and *E. tasmaniensis* Et1/99 have lost one or the other set at the species differentiation time point. This is supported by the fact that a remnant of a CRISPR repeat region having the Ypest subtype repeats is still present in *E. amylovora* CFBP 1430 as CRISPR3 region (Rezzonico et al., submitted for publication; Smits et al., 2010b). In the genome sequence of *E. tasmaniensis* Et1/99 no such scar was observed for the Ecoli subtype.

8. Other putative virulence systems

E. amylovora CFBP 1430 has a Type I secretion system, that exports a protease (Zhang et al., 1999). As the genomes of the two other species do not contain this gene cluster, it must have been acquired at the species differentiation level.

A difference between *E. tasmaniensis* Et1/99 and the pathogenic species is the presence of a sorbitol metabolism gene cluster in the latter two species (Kube et al., 2008a; Smits et al., 2010b). This feature may be ancestral, as also other *Erwinia* spp. are able to grow with sorbitol (Hauben and Swings, 2005). Therefore, it is proposed that the *srl*-operon is lost from the genome of *E. tasmaniensis* Et1/99.

The exopolysaccharide levan is produced by *E. amylovora* strains only, but both *E. amylovora* CFBP 1430 and *E. tasmaniensis* Et1/99 contain the levansucrase gene *lscC* (Gross et al., 1992; Kube et al., 2008a). As other *Erwinia* spp. also produce levan (Hauben and Swings, 2005), it seems likely that the gene is ancestral, and got lost in *E. pyrifoliae* at the species differentiation level (Smits et al., 2010a).

9. Role of horizontal gene transfer in evolution of *Erwinia*

The genomes of the three *Erwinia* spp. are largely collinear, showing a relatively low number of rearrangements (Smits et al., 2010b). This indicates that evolution must have taken a different strategy for differentiation of the individual species. Apart from the already discussed variations in the virulence systems, a large role in the differentiation of the *Erwinia* spp. can be appointed to horizontal gene transfer. Genomic islands containing transposon-related genes were numerous in *E. pyrifoliae* DSM 12163[T] (Smits et al., 2010a), while less prominent in *E. amylovora* CFBP 1430 and *E. tasmaniensis* Et1/99 (Smits et al., 2010b). The genomes of *E. pyrifoliae* DSM 12163T and *E. tasmaniensis* Et1/99 have several regions that contain phage-related genes, while only few remnants

were observed in *E. amylovora* CFBP 1430 (Smits et al., 2010a). A prominent feature of the genome of *E. tasmaniensis* Et1/99 is the presence of multiple clusters encoding fimbriae, largely absent in the genomes of the other two species (Smits et al., 2010a). For many, mainly smaller regions that are differential between the *Erwinia* species, the insertion mode cannot be reconstructed with the current data set.

10. The *Erwinia amylovora* species pan-genome

E. amylovora originated in North America, from where it has relatively recently spread first to New Zealand in the 1910s, to the U.K. and Northern Europe in the late 1950s, and more widely in Europe and the Middle East in the 1960s (Bonn and van der Zwet, 2000). It continues to spread across Europe and the Middle East, with advance eastward threatening the native origin of apple germplasm resources in Central Asia (Jock et al., 2002).

A comparison of the recently published genomes of *E. amylovora* strains CFBP 1430, isolated from *Crataegus* (Paulin and Samson, 1973; Smits et al., 2010b) and the apple isolate Ea273 (Sebaihia et al., 2010) showed that they are more than 99.99% identical on sequence level. Other whole-genome-targeting molecular methods like AFLP (Rico et al., 2004) illustrated the high similarity between individual strains as well, while PFGE revealed only minimal differences (Jock et al., 2002; Jock and Geider, 2004; Zhang and Geider, 1997). Only one replacement of a small region of 700 bp was observed that has an effect on the pan-genome of the species in that it added a single hypothetical protein in *E. amylovora* Ea273 (Smits et al., 2010b).

The factor currently known with the most influence on the *E. amylovora* pan-genome is the presence of diverse plasmids in some strains of *E. amylovora*. The plasmid pEA29 is nearly ubiquitous in *E. amylovora* strains worldwide (Llop et al., 2006; McGhee and Jones, 2000). Two RSF1010-based plasmids conferring streptomycin resistance (Palmer et al., 1997) and plasmid pEU30 (Foster et al., 2004) were only detected in USA west coast isolates, whereas a 72 kb plasmid was detected in *E. amylovora* Ea273 and further strains from the USA (Sebaihia et al., 2010; Steinberger et al., 1990). The plasmid pEL60 was only detected in isolates from the Middle-East area (Israel and Lebanon) (Foster et al., 2004; Rezzonico et al., submitted for publication), and the plasmid pEI70 in European isolates (Llop et al., 2006) (P. Llop, J. Cabrefiga, T.H.M. Smits, T. Dreo, S. Barbé, J. Pulawska, A. Bultreys, B. Duffy, E. Montesinos and M.M. López, 2010, manuscript in preparation). Apart from pEA29, all plasmids are found at a

regional scale, and are not omnipresent in all isolates from the region.

Resequencing of 4 additional European isolates of *E. amylovora* (T.H.M. Smits and B. Duffy, unpublished) revealed a near full coverage of the chromosome, while no new insertions were observed. This would mean that the pan-genome of *E. amylovora* affecting *Spiraeoidae* can be considered closed, apart from the plasmid content. From the 5 European *E. amylovora* genome sequences, only a single base change in around 36–150 kb was calculated (T.H.M. Smits and B. Duffy, unpublished). This may be explained by the relatively rapid dispersal of the organism over Europe, starting from around 1957 in the UK. Considering the average mutation rate in bacteria, the number of polymorphisms observed cannot be high unless a strong selective advantage is present. The current breeding strategy in pome fruit production is the vegetative proliferation of commercialized high-value varieties, that are in many cases highly sensitive to fire blight. This strategy counteracts natural resistance of the host against the pathogen, and allows *E. amylovora* to not having to adapt its genome to changes in host resistance.

11. Perspectives

With the available genome data of three species (Kube et al., 2008a; Smits et al., 2010a,b), preliminary conclusions were drawn for the evolution of *Erwinia* spp. However, we were not able to allocate some of the modifications at present, as genome information from the less related *Erwinia* spp. is still missing. The GOLD database (www.genomesonline.org) indicates sequencing projects for *E. billingiae* Eb661 (Mergaert et al., 1999) and *E. aphidicola* DSM 19347 (Harada et al., 1997), but draft sequences are not available yet. Whenever the genome sequences for these species and eventually others are made publicly available, it will be possible to have a closer look at some of the features therein and conclude on the features with currently insecure allocations.

Contributions

T.H.M.S. performed the comparative genomics and wrote the manuscript. F.R. participated in data analysis and in writing the manuscript. B.D. conceived of and supervised the project and participated in writing the manuscript. All authors read and approved the final manuscript.

Acknowledgements

References

Bell, K.S., Sebaihia, M., Pritchard, L., Holden, M.T.G., Hyman, L.J., Holeva, M.C., Thomson, N.R., Bentley, S.D., Churcher, L.J.C., Mungall, K., Atkin, R., Bason, N., Brooks, K., Chillingworth, T., Clark, K., Doggett, J., Fraser, A., Hance, Z., Hauser, H., Jagels, K., Moule, S., Norbertczak, H., Ormond, D., Price, C., Quail, M.A., Sanders, M., Walker, D., Whitehead, S., Salmond, G.P.C., Birch, P.J.R., Parkhill, J., Toth, I.K., 2004. Genome sequence of the enterobacterial phytopathogen *Erwinia carotovora* subsp. *atroseptica* and characterization of virulence factors. Proc. Natl. Acad. Sci. U. S. A. 101, 11105–11110.

Bernhard, F., Schullerus, D., Bellemann, P., Geider, K., Nimtz, M., Majerczak, D.R., Coplin, D.L., 1996. Genetics and complementation of DNA regions involved in amylovoran synthesis of *Erwinia amylovora* and stewartan synthesis of *Erwinia stewartii*. Acta Hortic. 411, 269–274.

Bingle, L.E.H., Bailey, C.M., Pallen, M.J., 2008. Type VI secretion: a beginner's guide. Curr. Opin. Microbiol. 11, 3–8.

Bonn, W.G., van der Zwet, T., 2000. Distribution and economic importance of fire blight. In: Vanneste, J.L. (Ed.), Fire Blight: The Disease and Its Causative Agent, *Erwinia amylovora*. CAB International, Wallingford, UK, pp. 37–53.

Calvin, L., Krissoff, B., 1998. Technical barriers to trade: a case study of phytosanitary and U.S.-Japanese apple trade. J. Agric. Res. Econom. 23, 351–366.

Coplin, D.L., Majerczak, D.R., Bugert, P., Geider, K., 1996. Nucleotide sequence analysis of the *Erwinia stewartii cps* gene cluster for synthesis of stewartan and comparison to the *Erwinia amylovora ams* cluster for synthesis of amylovoran. Acta Hortic. 411, 251–257.

De Maayer, P., Chan, W.Y., Venter, S.N., Toth, I.K., Birch, P.R.J., Joubert, F., Coutinho, T.A., 2010. The genome sequence of *Pantoea ananatis* LMG20103, the causative agent of *Eucalyptus* blight and dieback. J. Bacteriol. 192, 2936–2937.

Duffy, B., Schärer, H.-J., Bünter, M., Klay, A., Holliger, E., 2005. Regulatory measures against *Erwinia amylovora* in Switzerland. EPPO Bull. 35, 239–244.

Foster, G.C., McGhee, G.C., Jones, A.L., Sundin, G.W., 2004. Nucleotide sequences, genetic organization, and distribution of pEU30 and pEL60 from *Erwinia amylovora*. Appl. Environ. Microbiol. 70, 7539–7544.

Gaillard, M., Vallaeys, T., Vorhölter, F.-J., Minoia, M., Werlen, C., Sentchilo, V., Pühler, A., van der Meer, J.R., 2006. The *clc* element of *Pseudomonas* sp. strain B13, a genomic island with various catabolic properties. J. Bacteriol. 188, 1999–2013.

Geider, K., Auling, G., Du, Z., Jakovljevic, V., Jock, S., Völksch, B., 2006. *Erwinia tasmaniensis* sp. nov., a non-phytopathogenic bacterium from apple and pear trees. Int. J. Syst. Evol. Microbiol. 56, 2937–2943.

Geider, K., Auling, G., Jakovljevic, V., Völksch, B., 2009. A polyphasic approach assigns the pathogenic *Erwinia* strains from diseased pear trees in Japan to *Erwinia pyrifoliae*. Lett. Appl. Microbiol. 48, 324–330.

Glasner, J.D., Marquez-Villavicencio, M., Kim, H.-S., Jahn, C.E., Ma, B., Biehl, B.S., Rissman, A.I., Mole, B., Yi, X., Yang, C.-H., Dangl, J.L., Grant, S.R., Perna, N.T., Charkowski, A.O., 2008. Niche-specificity and the variable fraction of the *Pectobacterium* pan-genome. Mol. Plant–Microbe Interact. 21, 1549–1560.

Grimont, P.A.D., Grimont, F., 2005. Genus XXIII. Pantoea. In: Brenner, D.J., Krieg, N.R., Staley, J.T., Garrity, G.M. (Eds.), Bergey's Manual of Systematic Bacteriology. 2nd edition. Vol. 2: the Proteobacteria, part B: The Gammaproteobacteria. Springer, New York, pp. 713–720.

Gross, M., Geier, G., Rudolph, K., Geider, K., 1992. Levan and levansucrase synthesized by the fireblight pathogen *Erwinia amylovora*. Physiol. Mol. Plant Pathol. 40, 371–381.

Haft, D.H., Selengut, J., Mongodin, E.F., Nelson, K.E., 2005. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. PLoS Comp. Biol. 1, e60.

Harada, H., Oyaizu, H., Kosako, Y., Ishikawa, H., 1997. *Erwinia aphidicola*, a new species isolated from pea aphid, *Acyrthosiphon pisum*. J. Gen. Appl. Microbiol. 43, 349–354.

Hauben, L., Moore, E.R.B., Vauterin, L., Steenackers, M., Mergaert, J., Verdonck, L., Swings, J., 1998. Phylogenetic position of phytopathogens within the *Enterobacteriaceae*. Syst. Appl. Micribiol. 21, 384–397.

Hauben, L., Swings, J., 2005. Genus XIII. *Erwinia*. In: Brenner, D.J., Krieg, N.R., Staley, J.T., Garrity, G.M. (Eds.), Bergey's Manual of Systematic Bacteriology. 2nd edition. vol. 2: The Proteobacteria, part B: The Gammaproteobacteria. Springer, New York, pp. 670–679.

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.

Jock, S., Geider, K., 2004. Molecular differentiation of *Erwinia amylovora* strains from North America and of two Asian pear pathogens by analyses of PFGE patterns and *hrpN* genes. Environ. Microbiol. 6, 480–490.

Johnson, K.B., Stockwell, V.O., 1998. Management of fire blight: a case study in microbial ecology. Annu. Rev. Phytopathol. 36, 227–248.

Kim, W.-S., Gardan, L., Rhim, S.-L., Geider, K., 1999. *Erwinia pyrifoliae* sp. nov., a novel pathogen that affects Asian pear trees (*Pyrus pyrifolia* Nakai). Int. J. Syst. Bacteriol. 49, 899–906.

Kim, W.-S., Schollmeyer, M., Nimtz, M., Wray, V., Geider, K., 2002. Genetics of biosynthesis and structure of the capsular exopolysaccharide from the Asian pear pathogen *Erwinia pyrifoliae*. Microbiology 148, 4015–4024.

Koczan, J.M., McGrath, M.J., Zhao, Y., Sundin, G.W., 2009. Contribution of *Erwinia amylovora* exopolysaccharides amylovoran and levan to biofilm formation: implications to pathogenicity. Phytopathology 99, 1237–1244.

Kube, M., Migdoll, A.M., Müller, I., Kuhl, H., Beck, A., Reinhardt, R., Geider, K., 2008a. The genome of *Erwinia tasmaniensis* strain Et1/99, a non-pathogenic bacterium in the genus *Erwinia*. Environ. Microbiol. 10, 2211–2222.

Kube, M., Reinhardt, R., Jakovljevic, V., Jock, S., Geider, K., 2008b. The genomic sequence of the fire blight antagonist *Erwinia tasmaniensis* compared with virulence regions of *E. amylovora*. Acta Hortic. 793, 141–144.

Llop, P., Donat, V., Rodríguez, M., Cabrefiga, J., Ruz, L., Palomo, J.L., Montesinos, E., López, M.M., 2006. An indigenous virulent strain of *Erwinia amylovora* lacking the ubiquitous plasmid pEA29. Phytopathology 96, 900–907.

López, M.M., Roselló, M., Llop, P., Ferrer, S., Christen, R., Gardan, L., in press. *Erwinia piriflorinigrans* sp. nov., a novel pathogen that causes necrosis of pear blossoms. Int. J. Syst. Evol. Microbiol., doi:10.1099/ijs.0.020479-0.

Makarova, K.S., Grishin, N.V., Shabalina, S., Wolf, Y.I., Koonin, E.V., 2006. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. Biol. Direct. 1, 7.

McGhee, G.C., Jones, A.L., 2000. Complete nucleotide sequence of ubiquitous plasmid pEA29 from *Erwinia amylovora* strain Ea88: gene organization and intraspecies variation. Appl. Environ. Microbiol. 66, 4897–4907.

Medini, D., Donati, C., Tettelin, H., Masignani, V., Rappuoli, R., 2005. The microbial pan-genome. Curr. Opin. Genet. Dev. 15, 589–594.

Mergaert, J., Hauben, L., Cnockaert, M.C., Swings, J., 1999. Reclassification of non-pigmented *Erwinia herbicola* strains from trees as *Erwinia billingiae* sp. nov. Int. J. Syst. Bacteriol. 49, 377–383.

Mohd-Zain, Z., Turner, S.L., Cerdeño-Tárraga, A.M., Lilley, A.K., Inzana, T.J., Duncan, A.J., Harding, R.M., Hood, D.W., Peto, T.E., Crook, D.W., 2004. Transferable antibiotic resistance elements in *Haemophilus influenzae* share a common evolutionary origin with a diverse family of syntenic genomic islands. J. Bacteriol. 186, 8114–8122.

Mojica, F.J.M., Díez-Villaseñor, C., García-Martínez, J., Soria, E., 2005. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J. Mol. Evol. 60, 174–182.

Norelli, J.L., Jones, A.L., Aldwinckle, H.S., 2003. Fire blight management in the twenty-first century: using new technologies that enhance host resistance in apple. Plant Dis. 87, 756–765.

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.

Palmer, E.L., Teviotdale, B.L., Jones, A.L., 1997. A relative of the broad-host-range plasmid RSF1010 detected in *Erwinia amylovora*. Appl. Environ. Microbiol. 63, 4604–4607.

Paulin, J.-P., Samson, R., 1973. Le feu bactérien en France. II.—Caractères des souches d'Erwinia amylovora (Burrill) Winslow et al., 1920, isolées du foyer franco-belge. Ann. Phytopathol. 5, 389–397.

Rezzonico, F., Duffy, B., 2008. Lack of genomic evidence of AI-2 receptors suggests a non-quorum sensing role for *luxS* in most bacteria. BMC Microbiol. 8, 154.

Rezzonico, F., Smits, T.H.M., Duffy, B., 2010. Diversity and functionality of CRISPR regions in fire blight pathogen *Erwinia amylovora*. Appl. Environ. Microbiol.

Rezzonico, F., Smits, T.H.M., Montesinos, E., Frey, J.E., Duffy, B., 2009. Genotypic comparison of *Pantoea agglomerans* plant and clinical strains. BMC Microbiol. 9, 204.

Rico, A., Ortiz-Barredo, A., Ritter, E., Murillo, J., 2004. Genetic characterization of *Erwinia amylovora* strains by amplified fragment length polymorphism. J. Appl. Microbiol. 96, 302–310.

Roberts, R.G., Hale, C.N., van der Zwet, T., Miller, C.E., Redlin, S.C., 1998. The potential for spread of *Erwinia amylovora* and fire blight via commercial apple fruit; a critical review and risk assessment. Crop Prot. 17, 19–28.

Roselló, M., Peñalver, J., Llop, P., Gorris, M.T., Chartier, R., García, F., Montón, C., Cambra, M., López, M.M., 2006. Identification of an *Erwinia* sp. different from *Erwinia amylovora* and responsible for necrosis on pear blossoms. Can. J. Plant Pathol. 28, 30–41.

Samson, R., Legendre, J.B., Christen, R., Fischer-Le Saux, M., Achouak, W., Gardan, L., 2005. Transfer of *Pectobacterium chrysanthemi* (Burkholder et al. 1953) Brenner et al. 1973 and *Brenneria paradisiaca* to the genus *Dickeya* gen. nov. as *Dickeya chrysanthemi* comb. nov. and *Dickeya paradisiaca* comb. nov. and delineation of four novel species, *Dickeya dadantii* sp. nov., *Dickeya dianthicola* sp. nov., *Dickeya dieffenbachiae* sp. nov. and *Dickeya zeae* sp. nov. Int. J. Syst. Evol. Microbiol. 55, 1415–1427.

Sebaihia, M., Bocsanczy, A.M., Biehl, B.S., Quail, M.A., Perna, N.T., Glasner, J.D., DeClerck, G.A., Cartinhour, S., Schneider, D.J., Bentley, S.D., Parkhill, J., Beer, S.V., 2010. Complete genome sequence of the plant pathogen *Erwinia amylovora* strain ATCC 49946. J. Bacteriol. 192, 2020–2021.

Smits, T.H.M., Jaenicke, S., Rezzonico, F., Kamber, T., Goesmann, A., Frey, J.E., Duffy, B., 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., Rezzonico, F., Kamber, T., Blom, J., Goesmann, A., Frey, J.E., Duffy, B., 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., Kamber, T., Goesmann, A., Ishimaru, C.A., Stockwell, V.O., Frey, J.E., Duffy, B., in press. The genome sequence of the biocontrol agent *Pantoea vagans* strain C9-1. J. Bacteriol., doi:10.1128/JB.01122-01110.

Smits, T.H.M., Rezzonico, F., Pelludat, C., Goesmann, A., Frey, J.E., Duffy, B., 2010c. Genomic and phenotypic characterization of a non-pigmented variant of *Pantoea vagans* biocontrol strain C9-1 lacking the 530 kb megaplasmid pPag3. FEMS Microbiol. Lett. 308, 48–54.

Sorek, R., Kunin, V., Hugenholtz, P., 2008. CRISPR—a widespread system that provides acquired resistance against phages in bacteria and archaea. Nat. Rev. Microbiol. 6, 181–186.

Steinberger, E.M., Cheng, G.-Y., Beer, S.V., 1990. Characterization of a 56-kb plasmid of *Erwinia amylovora* Ea322: its noninvolvement in pathogencity. Plasmid 24, 12–24.

Thomson, S.V., 2000. Epidemiology of fire blight. In: Vanneste, J.L. (Ed.), Fire Blight: The Disease and Its Causative Agent, *Erwinia amylovora*. CAB International, Wallingford, UK, pp. 9–36.

Toth, I.K., Pritchard, L., Birch, P.J.R., 2006. Comparative genomics reveals what makes an enterobacterial plant pathogen. Annu. Rev. Phytopathol. 44, 305–336.

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.

Zhang, Y., Bak, D.D., Heid, H., Geider, K., 1999. Molecular characterization of a protease secreted by *Erwinia amylovora*. J. Mol. Biol. 289, 1239–1251.

Zhang, Y., Geider, K., 1997. Differentiation of *Erwinia amylovora* strains by pulsed-field gel electrophoresis. Appl. Environ. Microbiol. 63, 4421–4426.

Zhao, Y., He, S.Y., Sundin, G.W., 2006. The *Erwinia amylovora* avrRpt2$_{EA}$ gene contributes to virulence on pear and AvrRpt2$_{EA}$ is recognized by *Arabidopsis* RPS2 when expressed in *Pseudomonas syringae*. Mol. Plant–Microbe Interact. 19, 644–654.

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.