environmental microbiology reports

Environmental Microbiology Reports (2022) **14**(4), 584–590 doi:10.1111/1758-2229.13074

Brief Report

Genome sequencing provides new insights on the distribution of *Erwinia amylovora* lineages in northern Italy

Davide Albanese,[1*] Christian Cainelli,[2] Valeria Gualandri,[2,3] Simone Larger,[1] Massimo Pindo[1] and Claudio Donati[1]

[1]*Unit of Computational Biology, Research and Innovation Centre, Fondazione Edmund Mach, Via E. Mach 1, San Michele all'Adige, 38010, Italy.*

[2]*Center for Technology Transfer, Fondazione Edmund Mach, Via E. Mach 1, San Michele all'Adige, 38010, Italy.*

[3]*Center of Agriculture, Food and Environment (C3A), University of Trento, Trento, Italy.*

Summary

***Erwinia amylovora* is a Gram-negative bacterium that colonizes a wide variety of plant species causing recurrent local outbreaks of fire blight in crops of the Rosaceae family. Recent genomic surveys have documented the limited genomic diversity of this species, possibly related to a recent evolutionary bottleneck and a strong correlation between geography and phylogenetic structure of the species. Despite its economic importance, little is known about the genetic variability of co-circulating strains during local outbreaks. Here, we report the genome sequences of 82 isolates of *E. amylovora*, collected from different host plants in a period of 16 years in Trentino, a small region in the Northeastern Italian Alps that has been characterized by recurrent outbreaks of fire blight in apple orchards. While the genome isolated before 2018 are closely related to other strains already present in Europe, we found a novel subclade composed only by isolates that were sampled starting from 2018 and demonstrate that the endemic population of this pathogen can be composed by mixture of strains.**

Received 22 December, 2021; accepted 13 April, 2022. *For correspondence. E-mail davide.albanese@fmach.it.

Introduction

Erwinia amylovora is the causative agent of fireblight, a destructive disease of apple and pear trees and other rosaceous plants that can severely damage the pear and apple production in a region or country (Mansfield *et al*., 2012). The typical symptoms of infected plants are brown to black colour of twigs, flowers and leaves, production of exudates and typical shepherd's crook of infected shoots (Vanneste, 2000). Since its first identification at the end of the 18th century in the Hudson Valley in New York State, *E. amylovora* has progressively expanded its area first to the West Coast of the United States of America and subsequently to Europe, where it is one of the major threats to several important crops. Large-scale genome sequencing projects have shown that the major subdivision of *E. amylovora* is associated with host specificity, defining a *Rubus*-infecting clade and a *Spiraeoideae*-infecting clade (Zeng *et al*., 2018; Parcey *et al*., 2020). Despite a general low level of variability at the genomic level, apple and pear tree infecting strains can be partitioned according to their geographical distribution into three clades, namely, the Widely prevalent clade, the Eastern North American clade and the Western North American clade, with a heterogeneous group of strains ('B-group'). Almost all the genomes isolated outside North America (in Europe, Middle East, New Zealand, Eastern Asia and North Africa) belong to the Widely prevalent clade (Parcey *et al*., 2020).

The genome of *E. amylovora* is approximately 3.8 Mbp. Whole-genome comparative analysis has shown a high degree of genomic conservation between isolates, with an estimated average nucleotide identity (ANI) of 99.90% (Zeng *et al*., 2018). This remarkably low level of diversity has been attributed to an evolutionary bottleneck due to the recent colonization of fruit trees from an unknown natural reservoir (Malnoy *et al*., 2012) that has coincided with the beginning of large-scale cultivation of apples in North America. At the level of genes, the core genome was estimated to include an average of 89% of the genes of each genome based on the comparative analysis of 12 genomic sequences (Mann *et al*., 2013). Besides the chromosome,

© 2022 The Authors. *Environmental Microbiology Reports* published by Society for Applied Microbiology and John Wiley & Sons Ltd. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

E. amylovora strains also contain a variable combination of plasmids, including the non-transmissible pEA29 plasmid, which is consistently found in almost all sequenced isolates. Variable combinations of other, less conserved plasmids are commonly found (Parcey *et al*., 2020).

Pathogenicity of *E. amylovora* has been associated to several chromosomally encoded features, including a Type III secretion system, the amylovoran synthesis gene cluster and the multidrug efflux pump AcrAB (Maes *et al*., 2001; Lee *et al*., 2010; Vrancken *et al*., 2013; Piqué *et al*., 2015). Production of the exopolysaccharide amylovoran has been shown to be essential for the formation of biofilm (Koczan *et al*., 2009) and to be one of the major determinants of the degree of virulence both in apple (Lee *et al*., 2010) and in Rubus-infecting strains. The latter produces amylovoran with different structural properties than other strains (Maes *et al*., 2001) The Hrp pathogenicity island, a 62 kb region flanked by genes that suggest that it might have been acquired by horizontal gene transfer, encodes a Type III secretion system that has been shown to deliver several proteins into host cells, including the HrpN and HrpW hairpin proteins, the dspA/E protein and EopB (Oh and Beer, 2005). Type VI secretion systems have been identified in several Erwinia species (De Maayer *et al*., 2011). The presence and functionality of Type VI secretion systems in *E. amylovora* has been shown to be associated to bacterial competition and virulence and to alter exopolysaccharide production and expression of motility and chemotaxis-associated genes (Tian *et al*., 2017).

Besides chromosomally encoded proteins, plasmids commonly found in *E. amylovora* have been shown to have a role in virulence. Strains cured from the pEA29 plasmid have been shown to be less virulent (Falkenstein *et al*., 1989), although strains lacking pEA29 have been found in symptomatic hosts (Llop *et al*., 2006). Similarly, strains cured from pEI70 plasmid show decreased aggressiveness compared to the wild type (Llop *et al*., 2011). Type IV secretion systems are predicted to be encoded by the accessory plasmids pEA72, pEU30 and pEA78 (Parcey *et al*., 2020).

Due to the excellent climatic conditions that enhance the quality of the fruit, apples are one of the most important crops in Trentino, an Italian region located in the North East of Italy. Apple trees are grown on approx. 12 000 ha, and their production accounts for about 65% of apples harvested in Italy. The main production areas are Val di Non, a valley of about 630 km^2 located west of the Adige river close to the Adamello mountain group, and Valsugana, a valley of about 980 km^2 east of the Adige river. Fire blight appeared for the first time in Trentino in 2003 in two locations on two pear trees. From 2003 to 2019 the contagion situation was stable and sporadic and over the years the stabilization of the disease was attributed to the effectiveness of the control measures adopted for the containment of the disease. This relatively stable situation changed in 2020, when a serious epidemic has affected particularly the orchards located in the Valsugana valley.

To understand if the major 2020 outbreak is to be attributed to the introduction of a more virulent variant of *E. amylovora*, possibly originated outside of the Trentino region, or to strains already present in the area from previous years, we assembled the genome sequences of 82 isolates of *E. amylovora* collected in Trentino both from Val di Non and Valsugana, exhaustively covering the strains circulating in the area since the first appearance of the pathogen in the area.

Results and discussion

We selected 82 isolates of *Erwinia amylovora* collected in Trentino over a 16-year timespan (2005–2020). In the vast majority of the cases the isolates were recovered from *Malus domestica*, one of the major crops in the sampled area, but the sampled hosts also include *Cydonia oblonga*, *Eriobotrya japonica*, *Cotoneaster* sp. and *Crataegus* sp. (Supplementary Table 1). The draft genomes assembled (see Methods) have a length between 3 810 012 and 3 948 503 bp and are composed of a number of scaffolds ranging between 124 and 388, with an N50 between 359 899 and 599 290. All the sequenced genomes were estimated 100% complete with a level of contamination below 1% evaluating the presence of clade-specific single-copy genes (Albanese and Donati, 2021). Summary statistics about the quality of the assembled genomes are reported in Supplementary Table 2.

In addition to the newly sequenced genomes, 134 complete and draft genome sequences were downloaded from GenBank (for a complete list of the sequences included in the present study see Supplementary Table 1). In order to guarantee homogeneity of the data, both newly sequenced genomes and genomes downloaded from GenBank were *de novo* annotated using the same bioinformatic pipeline (see Experimental procedures). The number of predicted genes in each genome of the complete dataset varied between 3338 and 3666. Pangenome analysis (see Experimental procedures) identified a core of 2935 genes that were present in more than 99% of the strains, complemented by 584 genes present in between 99% and 15% of the genomes and 3005 cloud genes present in less than 15% of the genomes. In total, the pangenome of *E. amylovora* included 6524 genes. Core genes produced a concatenated alignment of 2 111 642 bp. The high degree of sequence conservation of *E. amylovora* is confirmed by the Watterson estimator theta (see Supplementary Material).

© 2022 The Authors. *Environmental Microbiology Reports* published by Society for Applied Microbiology and John Wiley & Sons Ltd.,

The concatenated gene sequence alignment of the complete dataset was used to reconstruct a Maximum Likelihood phylogenetic tree (Fig. 1A and B). Since phylogenetic tree inference was performed using core gene alignments, non-ubiquitous genetic elements like some plasmids were not considered. The phylogenetic analysis confirms that *E. amylovora* isolates can be partitioned into different clades (bootstrap support >90%), in accordance with Parcey *et al*. (2020). In particular, we could identify a highly divergent *Rubus*-infecting clade (not shown, due to their high level of divergence from the rest of the species), a B-group clade, a North American specific clade subdivided into Eastern and Western subclades, and a Widely Prevalent clade that included, besides a group of isolates from North America, all the isolates from the rest of the World, and, in particular, all the newly sequenced isolates from Trentino (Fig. 1A). Within the Widely Prevalent clade, North American isolates formed a subclade that included also five recent isolates from South Korea (Song *et al*., 2021) and that was well separated from a second subclade that is predominantly European but also includes isolates from New Zealand, Middle East and North Africa. The European subclade can be further subdivided into two major branches, both of which include isolates from Trentino. One of these branches also included eight isolates from Europe (France, Switzerland, England and Italy), two from North Africa (Morocco) and one from Israel. A t-distributed stochastic neighbour embedding (t-SNE) analysis using the ANI as a measure of similarity (Fig. 1C) confirms the population structure of *E. amylovora*, including the heterogeneous nature of the B-Group. The subclades of the Widely Prevalent clade and their correlation with geographic origin of the isolates were clearly visible in the two-dimensional projection.

Using the Jaccard distances between the gene presence–absence profiles from the pangenome analysis, we could quantify the variability in the gene repertoire both within and between the clades. We found (Fig. 1C and D) that the B-group included a heterogeneous group

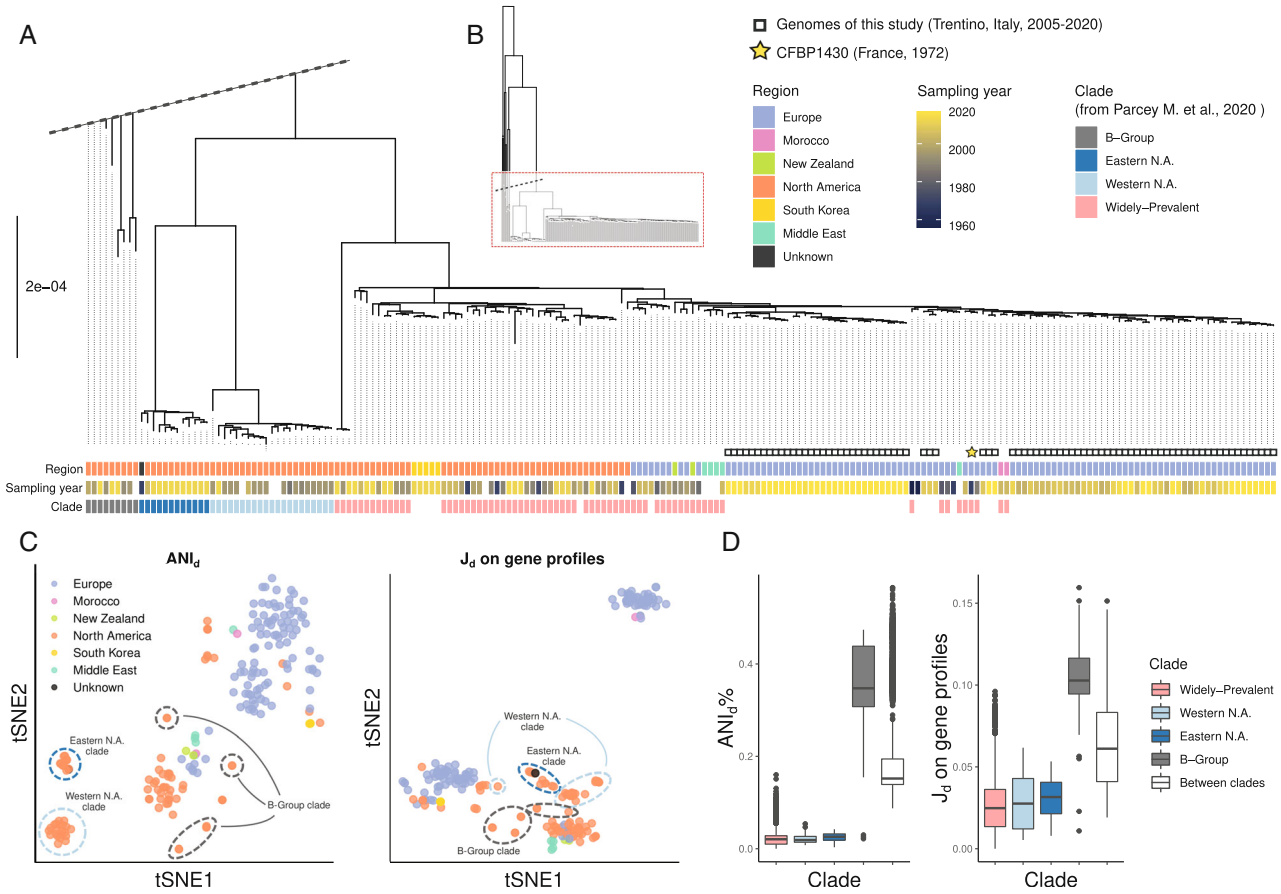

Fig. 1. A. Maximum Likelihood phylogenetic tree based on core gene alignment rooted at the 1994 *Rubus*-infecting EaLevo2 strain. B. Full tree (C) t-SNE based on the ANI (left) and on the Jaccard distance between profiles of orthologous gene presence–absence. Points are coloured according to the geographical origin of the isolate. Eastern North American clade, Western North American clade and B-Group are highlighted; all other points belong to the Widely Prevalent clade. D. Intra-clade distances based on ANI and Jaccard distances between profiles of orthologous gene presence–absence. Rubus-infecting strains are not shown.

© 2022 The Authors. *Environmental Microbiology Reports* published by Society for Applied Microbiology and John Wiley & Sons Ltd.,

of isolates, while the Eastern North American, Western North American and Widely Prevalent genomes have a comparable intra-clade variability, much smaller than the variability within the B-group. However, applying the t-SNE analysis on genes profiles, we could further partition the Widely Prevalent clade, identifying three distinct subgroups. One of these subgroups was composed almost exclusively by the North American and South Korean isolates, while the European isolates were distributed in the other two groups. The latter distinction, surprisingly well defined given the low ANI_d distance between the members of the two European subgroups (Fig. 1D), is due to the presence in one of the two of a large (70 kb) plasmid, the PEI70 (Llop *et al*., 2011), that so far had been identified only in a small group of isolates from France, Morocco and Switzerland (Parcey *et al*., 2020).

Exploiting the low level of sequence divergence between isolates from the Widely Prevalent clade and the relatively large time range spanned by the sampling (1959–2020), we could estimate the dates of the most recent common ancestor (MRCA) of the clade and the dates of the diversification of the major branches within the clade using an approximate maximum likelihood method (Sagulenko *et al*., 2018). We found that a strict molecular clock model could be used to describe the evolution of this set of sequences (Supplementary Fig. S1), with a scaled mutation rate inferred from the regression of the root to tip distance versus time of 1.159E-7 $\pm$ 1.66E-8 mutations per site and year. Using this approach (Fig. 2), we could estimate that the last common ancestor of the Widely Prevalent clade can be traced back in the first quarter of the 18th century (1721, CI 1657–1787, WP node in Fig. 2), and the MRCA of the European subclade is estimated in the mid-19th century (1849, CI 1817–1888 EU node).

Following the European branch of this tree we could further infer the dates of introduction of *E. amylovora* in the different parts of the continent. In particular, the MRCA of all the isolates from central and Eastern Europe (Germany, Poland and Belarus) could be dated in 1935 (CI 1909–1946, CE-EU node). In the same time range is the MRCA of the isolates from England, Ireland, France, Switzerland and Italy (CW-EU node, 1935, CI 1909–1941).

The molecular clock analysis allowed us to date also the time when the Italian isolates from Trentino diversified from the other European strains. One of the two Italian subclades (that also includes two isolates from Morocco, and isolates from France and Switzerland) had the MRCA that could be dated in 1958 (CI 1945–1960 node C1), while the second Italian subclade had a last common ancestor in 1965 (CI 1957–1982, node C2). Interestingly, we found that the isolates from 2020 could not be traced back to a single recently evolved group of closely related isolates, but rather belong to both the Italian lineages.

We explored the correlation between the phylogenetic analysis of the newly sequenced genomes and the year of sampling, the area within Trentino where they were sampled (Supplementary Fig. S2a,b) and the host species. Mapping the species of the infected hosts on the phylogenetic tree, we found no evident correlation between host and position on the phylogenetic tree (Supplementary Fig. S2c). We also found that a large fraction of the 2020 isolates from Alta and Bassa Valsugana are closely related, and well distinct from the isolates from previous years and from other areas of Trentino.

For the newly sequenced isolates, we assembled and characterized the plasmids that could be recovered from the whole genome sequencing. We found that all isolates contain the pEA29 plasmid (Parcey *et al*., 2020) (see Supplementary Fig. S2). Although strains lacking pEA29 have also been described (Llop *et al*., 2006) the pEA29 plasmid is commonly found in *E. amylovora* isolates and is known to modulate the development of symptoms during infection (Falkenstein *et al*., 1989). In addition, we could assemble two additional plasmids that could be classified as pEI70 and pEA30. While the pEA30 was only present in isolates sampled before 2020 (the most recent isolate harbouring pEA30 was from 2017), the pEI70 plasmid has been recovered from isolates that span the entire sampling time frame. Supplementary Fig. S2 indicates that there is only partial correlation between the tree structure and the presence of pEA30 and pEI70, suggesting that these plasmids could have been exchanged between co-infecting strains by lateral transfer. The pEA30 plasmid, that had been previously identified in a single isolate of *E. amylovora*, and that encodes a putative type IV secretion system for conjugative transfer (Mann *et al*., 2013; Christie, 2016) is distantly related to the IncU plasmids, which often include antibiotic resistance cassettes (Llop, 2015). However, no known antibiotic resistance gene has been identified in pEA30 (Mann *et al*., 2013). The pEI70 plasmid is a 65.8 kb genetic element that has been shown to be associated with increased aggressiveness in pear trees (Llop *et al*., 2011; Llop, 2015). The heterogeneity of the plasmid distribution that we have found among closely related isolates is in stark contrast with the high level of sequence conservation of the chromosome of *E. amylovora* (Parcey *et al*., 2020), and supports the hypothesis that plasmids play a role in the evolution and adaptation of this pathogen (Sundin, 2007).

Our results greatly expanded the repertoire of genetic information available for *E. amylovora* and allowed us to better define the population structure of strains circulating in Europe, demonstrating the geographical structuring of this species. These results also show that the endemic population of this pathogen in a relatively small

© 2022 The Authors. *Environmental Microbiology Reports* published by Society for Applied Microbiology and John Wiley & Sons Ltd.,

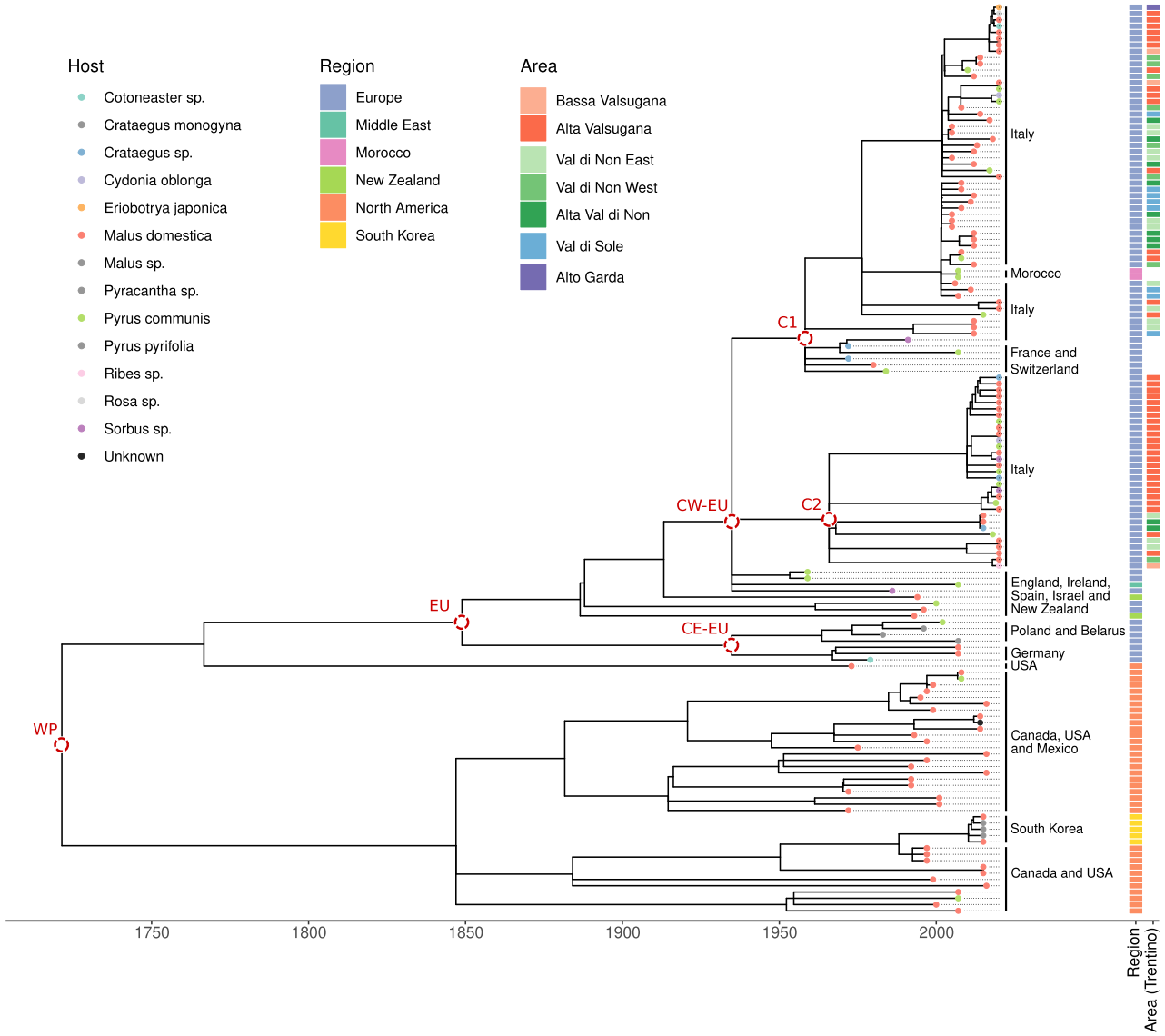

Fig. 2. Maximum-likelihood time-scaled phylogeny of the Widely Prevalent clade.

geographic area and homogenous from the agricultural management point of view such as Trentino is composed by a complex mixture of strains that have been independently introduced in different times, suggesting that epidemics outbreaks of the disease might be related to specific environmental factors, rather than to the introduction of new strains.

Experimental procedures

Selection of isolates, library construction and sequencing

In the period 2005–2020, bacterial cultures associated with natural infections of different plants were isolated in Trento Province. Isolates were selected on the basis of the typical levani form morphology on NSA medium followed by the lack of production of the fluorescent pigment on King's B medium. The identity of the isolates was then confirmed with Real-Time PCR (Gottsberger, 2010). Pure cultures were stored in glycerol at −80°C. Among these, 82 isolates were chosen to sequence the genome on the basis of the collection year, host plant species and, in the case of apple, cultivar. Thus, the selected isolates were transferred from the glycerol stock to NSA medium. *Erwinia amylovora* DNA was extracted according to the Nucleospin® Plant II protocol (Macherey-Nagel GmbH KG, Germany) starting from a 250 µl aqueous solution of pure culture. The DNA was eluted with 70 µl H_2O. The extracted DNA of 82 isolates have been used to construct the next-generation sequencing libraries, starting from 100 ng each and using the Kapa Hyperplus library kit (Roche) combined with the Kapa

© 2022 The Authors. *Environmental Microbiology Reports* published by Society for Applied Microbiology and John Wiley & Sons Ltd.,

Unique Dual Index adapters (Roche) following the manufacturer's instruction. The pooled library was quantified by qPCR and the fragment insertion size of 500 bp was evaluated on the Agilent 2200 TapeStation. The final library was sequenced at IGA Technology on NovaSeq 6000 in paired-end 150 mode.

Bioinformatic analysis

Bioinformatic methods, including raw sequences processing, genomes assembly and downstream analyses are described in the Supplementary Material.

Acknowledgements

The authors gratefully acknowledge technicians of the Technology Transfer Centre at Fondazione Edmund Mach, in particular: Delaiti Lodovico, Mattia Zaffoni and Nicola Andreatti for providing samples and their support, to the Phytosanitary Services of the Autonomous Province of Trento and to Dr. Zasso Rosaly for the 2005–2009 isolate collection.

Data Availability

Raw sequence data and assemblies are available at the National Center for Biotechnology Information (NCBI) under the BioProject accession PRJNA791304.

References

Albanese, D., and Donati, C. (2021) Large-scale quality assessment of prokaryotic genomes with metashot/prok-quality. *F1000Res* **10**: 822.

Christie, P. J. (2016) The mosaic type IV secretion systems. *EcoSal Plus* **7**(1). https://doi.org/10.1128/ecosalplus.ESP-0020-2015

De Maayer, P., Venter, S.N., Kamber, T., Duffy, B., Coutinho, T.A., and Smits, T.H.M. (2011) Comparative genomics of the type VI secretion systems of Pantoea and Erwinia species reveals the presence of putative effector islands that may be translocated by the VgrG and hcp proteins. *BMC Genomics* **12**: 1–15.

Falkenstein, H., Zeller, W., and Geider, K. (1989) The 29 kb plasmid, common in strains of Erwinia amylovora, modulates development of Fireblight symptoms. *Microbiology* **135**: 2643–2650.

Gottsberger, R.A. (2010) Development and evaluation of a real-time PCR assay targeting chromosomal DNA of *Erwinia amylovora*. *Lett Appl Microbiol* **51**: 285–292.

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

Lee, S.A., Ngugi, H.K., Halbrendt, N.O., O'Keefe, G., Lehman, B., Travis, J.W., *et al*. (2010) Virulence characteristics accounting for fire blight disease severity in apple trees and seedlings. *Phytopathology* **100**: 539–550.

Llop, P. (2015) Genetic islands in pome fruit pathogenic and non-pathogenic Erwinia species and related plasmids. *Front Microbiol* **6**: 874.

Llop, P., Cabrefiga, J., Smits, T.H.M., Dreo, T., Barbé, S., Pulawska, J., *et al*. (2011) Erwinia amylovora novel plasmid pEI70: complete sequence, biogeography, and role in aggressiveness in the fire blight phytopathogen. *PLoS One* **6**: e28651.

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

Maes, M., Orye, K., Bobev, S., Devreese, B., Van Beeumen, J., De Bruyn, A., *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.

Malnoy, M., Martens, S., Norelli, J.L., Barny, M.-A., Sundin, G.W., Smits, T.H.M., and Duffy, B. (2012) Fire blight: applied genomic insights of the pathogen and host. *Annu Rev Phytopathol* **50**: 475–494.

Mann, R.A., Smits, T.H.M., Bühlmann, A., Blom, J., Goesmann, A., Frey, J.E., *et al*. (2013) Comparative genomics of 12 strains of *Erwinia amylovora* identifies a pan-genome with a large conserved core. *PLoS One* **8**: e55644.

Mansfield, J., Genin, S., Magori, S., Citovsky, V., Sriariyanum, M., Ronald, P., *et al*. (2012) Top 10 plant pathogenic bacteria in molecular plant pathology. *Mol Plant Pathol* **13**: 614–629.

Oh, C.-S., and Beer, S.V. (2005) Molecular genetics of *Erwinia amylovora* involved in the development of fire blight. *FEMS Microbiol Lett* **253**: 185–192.

Parcey, M., Gayder, S., Morley-Senkler, V., Bakkeren, G., Úrbez-Torres, J.R., Ali, S., *et al*. (2020) Comparative genomic analysis of *Erwinia amylovora* reveals novel insights in phylogenetic arrangement, plasmid diversity, and streptomycin resistance. *Genomics* **112**: 3762–3772.

Piqué, N., Miñana-Galbis, D., Merino, S., and Tomás, J.M. (2015) Virulence factors of *Erwinia amylovora*: a review. *Int J Mol Sci* **16**: 12836–12854.

Sagulenko, P., Puller, V., and Neher, R.A. (2018) TreeTime: maximum-likelihood phylodynamic analysis. *Virus Evol* **4**: vex042.

Song, J.Y., Yun, Y.H., Kim, G.-D., Kim, S.H., Lee, S.-J., and Kim, J.F. (2021) Genome analysis of *Erwinia amylovora* strains responsible for a fire blight outbreak in Korea. *Plant Dis* **105**: 1143–1152.

Sundin, G.W. (2007) Genomic insights into the contribution of phytopathogenic bacterial plasmids to the evolutionary history of their hosts. *Annu Rev Phytopathol* **45**: 129–151.

Tian, Y., Zhao, Y., Shi, L., Cui, Z., Hu, B., and Zhao, Y. (2017) Type VI secretion systems of *Erwinia amylovora* contribute to bacterial competition, virulence, and exopolysaccharide production. *Phytopathology* **107**: 654–661.

Vanneste, J. (ed). (2000) *Fire Blight: The Disease and its Causative Agent, Erwinia amylovora*. Wallingford, England: CABI Publishing.

Vrancken, K., Holtappels, M., Schoofs, H., Deckers, T., and Valcke, R. (2013) Pathogenicity and infection strategies of the fire blight pathogen *Erwinia amylovora* in Rosaceae: state of the art. *Microbiology* **159**: 823–832.

© 2022 The Authors. *Environmental Microbiology Reports* published by Society for Applied Microbiology and John Wiley & Sons Ltd.,

Zeng, Q., Cui, Z., Wang, J., Childs, K.L., Sundin, G.W., Cooley, D.R., *et al*. (2018) 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**: 1652–1666.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Appendix S1. Supporting Information.
Appendix S2. Supporting Information.

© 2022 The Authors. *Environmental Microbiology Reports* published by Society for Applied Microbiology and John Wiley & Sons Ltd.,