



# The *ams* proteins and the amylovoran biosynthetic pathway: an extensive bioinformatic study

Lavinia Carlini<sup>1</sup> · Alfonso Esposito<sup>2</sup> · Luca Ambrosino<sup>3</sup> · Sharda Bharti<sup>1</sup> · Luca Mauro Invernizzi<sup>1</sup> · Silvano Piazza<sup>4</sup> · Stefano Benini<sup>1</sup>

Received: 18 April 2023 / Accepted: 26 September 2023  
© The Author(s) 2023

## Abstract

The genus *Erwinia* includes plant-pathogenic, non-pathogenic and even opportunistic human pathogens. Among them, *Erwinia amylovora* is one of the most studied species, as the etiological agent of the relevant necrotic plant disease fire blight. Its most important pathogenicity factor is amylovoran, an exopolysaccharide essential in the formation of a protective biofilm used by the bacterium to escape the plant protection mechanisms. Amylovoran biosynthesis, regulation and secretion require the concerted action of twelve genes clustered in a region spanning 16 kb, the *ams* operon. Most of the encoded proteins build and polymerize the repeating oligosaccharide units of galactose, glucose, glucuronic acid and pyruvate, others have signalling, regulation and secretion functions. We compared the *ams* operon organisation, predicted 3D structure of the encoded proteins of *Erwinia amylovora* in eighteen *Erwinia* species. We detected the presence of a paralog (for the *amsG* gene) and some rearrangements in the EPS biosynthetic operon in specific taxa that could be connected to the species-specific pathogenicity profile. The present study adds fundamental information about the genomic determinants of exopolysaccharide secretion in the *Erwinia* genus, and clarifies the evolutionary relationships among the genes in the *ams* operon.

**Keywords** *Erwinia* · Amylovoran · Ams protein · *ams* operon · Exopolysaccharide

## Introduction

*Erwinia amylovora* is the first phytopathogenic bacterium ever described (Vanneste 2000) and still one of the top ten most dangerous plant pathogens (Mansfield et al. 2012). It is one of the main species of the genus *Erwinia*, and the causal agent of fire blight, a devastating disease affecting a wide range of host species within Rosaceae (apple, pear, hawthorn, cotoneaster, rubus, etc.), with a potentially

disastrous economic impact on pome fruit commercial production (Griffith et al. 2003; Heitefuss 2012). The systemic infection of *E. amylovora* usually occurs in spring, and it starts when the bacterial cells enter the plants through the nectarthodes of flowers or other natural openings and slowly spreads to the vascular system, mainly in the xylem. Via the rapid movement within the plant, *E. amylovora* causes extensive wilting symptoms and sometimes complete die-back (Thomson et al. 1986). The mechanisms by which

✉ Stefano Benini  
stefano.benini@unibz.it

Lavinia Carlini  
lavinia.carlini@student.unibz.it

Alfonso Esposito  
alfonso.esposito@unibz.it

Luca Ambrosino  
luca.ambrosino@szn.it

Sharda Bharti  
sharda.bharti@student.unibz.it

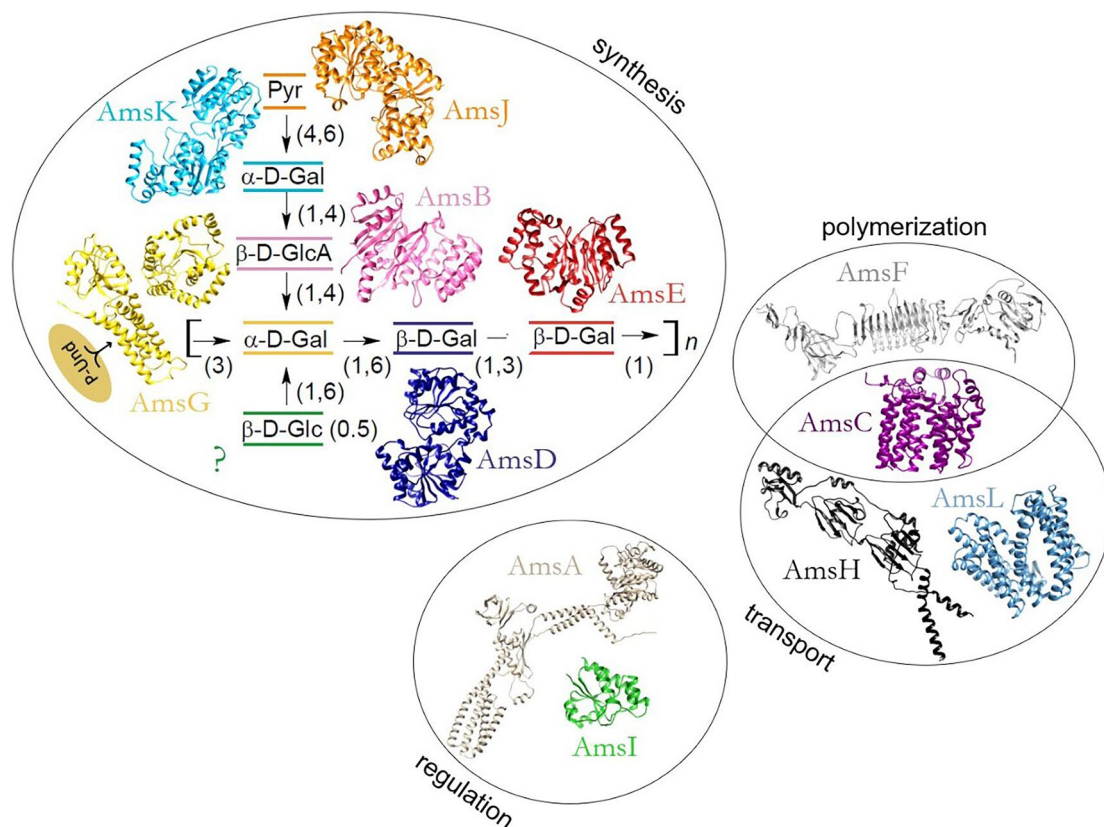
Luca Mauro Invernizzi  
luca Mauro Invernizzi@student.unibz.it

Silvano Piazza  
silvano.piazza@icgeb.org

- <sup>1</sup> Laboratory of Bioorganic Chemistry and Biocrystallography, Free University of Bolzano-Bozen, Piazza Università 1, 39100 Bolzano, Italy
- <sup>2</sup> Faculty of Medicine and Surgery, University of Enna “Kore”, Cittadella Universitaria, 94100 Enna, Italy
- <sup>3</sup> Research Infrastructure for Marine Biological Resources Department, Stazione Zoologica Anton Dohrn, 80121 Naples, Italy
- <sup>4</sup> International Centre for Genetic Engineering and Biotechnology Padriciano, 99, 34149 Trieste, Italy

*E. amylovora* triggers this relevant necrotic plant disease involve the migration in the apoplast, and amylovoran, an exopolysaccharide (EPS) essential in the formation of a protective biofilm, is associated with the motility phenotype (Yuan et al. 2022). Along with the type III secretion system (T3SS), the amylovoran biosynthetic pathway is one of the main virulence factors required for successful infection in *E. amylovora* as well as in other *Erwinia* species (Malnoy et al. 2012; Piqué et al. 2015). Mutants defective in the exopolysaccharide production are unable to form the required protective biofilms, which are fundamental to prevent recognition by the host plant defence system, obstruct the vascular system and facilitate movement through the plant tissue. In addition, amylovoran has also been shown to promote bacterial survival outside the host plant by preventing drying (Bellemann et al. 1994; Kharadi et al. 2021; Koczan et al. 2009; Vanneste 2000).

Besides the proteins involved in the galactose metabolism (Benini et al. 2017; Metzger et al. 1994), the *ams* operon determines the biosynthesis, chain elongation, transport and secretion of the amylovoran repeating unit, which is composed of four galactose, one glucuronic acid residue and final pyruvate. This operon was firstly identified in 1995 as a single transcript of 16 kb with 12 collinear open reading frames, represented by genes from *amsA* to *amsL* (Bugert and Geider 1995). Six of the *ams*-encoded proteins are essential to the synthesis of an amylovoran subunit (Fig. 1, upper part): the carbohydrate-active enzymes AmsG, AmsB, AmsD, AmsE, AmsK, which have a distinct role in glycosyl transfer, and the pyruvyl transferase AmsJ, play a key part in annealing the different sugar moieties to form a new polysaccharide unit. The consequent polymerization is instead achieved by AmsC and AmsF. While the *amsH* and *amsL*



**Fig. 1** Schematic representation of the amylovoran biosynthetic pathway. All protein structures are models obtained from AlphaFold analysis, except for AmsI, solved and deposited as PDB entry 4D74 (Salomone-Stagni et al. 2016). Proteins involved in the steps of synthesis, polymerisation, transport and regulation have been grouped into different sets. The synthesis steps governed by each protein are colour coded (and underlined in the same colour). At the beginning of the synthesis, a molecule of a membrane-bound lipidic carrier P-Und (undecaprenylphosphate) binds a molecule of UDP-galactose thanks to the glycosyltransferase AmsG (yellow). AmsD in blue and AmsE in red, add a second and a third molecule of UDP-Gal. To this trisaccharide, a molecule of UDP-Glucuronic acid is added by

AmsB (pink). The light blue protein AmsK adds a further UDP-Gal. AmsJ (orange) adds the pyruvate to end the growth of the side chain. Another yet unidentified glycosyltransferase is responsible for adding the glucose moiety in 50% of the repeating units. Proteins AmsB, AmsH, and AmsL (represented respectively in violet, black and turquoise) were suggested to be involved in oligosaccharide transport and assembly, as they contain transmembrane domains. While AmsA (beige) is a tyrosine kinase and activates amylovoran, AmsI (light green) appears to have a distinct function in the recycling of the lipid carrier. Finally, the enzyme AmsF (grey) processes the newly synthesized repeating units and is involved in their polymerization, adding them to another amylovoran chain

gene products are involved in oligosaccharide transport and assembly, the last two Ams proteins possess a tyrosine kinase (AmsA) and an acid phosphatase (AmsI) activity and regulate amylovoran synthesis (Langlotz et al. 2011; Wang et al. 2012a).

Extensive research has been dedicated to investigating the regulation of the biosynthetic pathway of amylovoran in *E. amylovora* (e.g., Benini 2021; Smits et al. 2017; Wang et al. 2011). Over the past years, there has been thorough examination of this complex process: the roles of the *rcaA* gene product and the RcsB protein in transcriptional activation and regulation of the *ams* operon has been extensively studied (Bugert and Geider 1995; Kelm et al. 1997). Furthermore, the orphan gene *amyR* has been identified as a novel negative regulator of amylovoran production (Wang et al. 2012b). The second messenger cyclic-di-GMP (c-di-GMP) has been recently reported to positively regulate the transcription of *amsG*, which encodes the first protein present in the amylovoran biosynthetic pathway (Kharadi and Sundin 2022). Moreover, current researches have shed light on the AraC-type transcription regulator YqhC, which has also been found to influence *amsG* expression and influence *E. amylovora* pathogenicity (Sahebi et al. 2022). However, the research on the exopolysaccharide biosynthesis in *Erwinia* spp. has been relatively limited. There are a few comparative studies investigating the responsible operon across a small range of *Erwinia* and *Pantoea* genomes (e.g., Langlotz et al. 2011; Kamber et al. 2012) but none of them exploits the wealth of genomic data available nowadays.

The significance of exopolysaccharide secretion in the pathogenic phenotype of *Erwinia* species is well recognized. Building upon earlier research, the present study aims to contribute by outlining the differences in the chromosomal organization of the EPS biosynthetic operon (*ams*, *cps*, *wce*, etc.) within the *Erwinia* genus. We used a comparative genomic approach on a dataset constituted of eighteen genomes belonging to the *Erwinia* spp. A genome-wide analysis was conducted including phylogenetic inference, orthology network analysis using BLAST and the Bidirectional-Best-Hit approach. To finalize the examination of the *ams* gene, analysis on the predicted 3D structure of the proteins involved in the amylovoran biosynthetic pathway were conducted.

## Materials and methods

### Download of the genomes and phylogenetic tree

In order to have a dataset covering most of the *Erwinia* species, we selected forty-six genomes including pathogens and non-pathogens. We selected the genomes with

the assembly level complete as of 29 May 2023. A phylogenetic tree, including all selected *Erwinia* species, *Duffryella gerundensis* and *Pantoea stewartii* DC283 as an outgroup, was built using PhyloPhlAn3 (Asnicar et al. 2020). Bootstrap support was calculated after 999 random resampling of the characters. The tree was drawn and annotated (i.e. presence or absence of genes was reported on each leaf of the tree) using IToL server 6.7.1 (Letunic and Bork 2021). A comprehensive list of all the genomes downloaded and the complete generated tree are available as Supplementary material (Fig. S1 and Table S1). Average Nucleotide Identity was computed using the software Pyani (Pritchard et al. 2016), and the output was plotted using DiMHepy (Esposito et al. 2019).

### Identification of *ams* genes

The search for homologous genes was done using BLAST (Altschul et al. 1990). Since the focus was on the among-species variation, and the identity among the different *E. amylovora* genomes was over 97%, only one genome was chosen per species (Table 1). The coding and protein sequences of the *ams* operon in each genome were identified using as a query the genes and proteins, respectively, of *E. amylovora* CFBP 1430 (Smits et al. 2010a, b), and as a database for each of the other seventeen *Erwinia* (Table 2). We considered as valid orthologs inference the hits having over 80% of coverage and at least 80% of nucleotide identity or 60% of protein identity. To account for the increasing phylogenetic distance between the query (i.e. the *E. amylovora* genes) and the other species, we repeated the search using as query the orthologs of *E. oleae*. This species was the most early diverging genome in the tree (therefore having the higher divergence from the *E. amylovora*) in which all orthologs were found. To further minimize the false negative detection due to low sequence identity, we also used the agnostic tool Prodigal (Hyatt et al. 2010), to detect additional coding sequences in the region of interest. To create the gene trees, we relied on the MEGAX program package (Tamura et al. 2021) using the neighbour-joining method (Fig. S2). The structure of the operon was depicted using the R-package GenoPlotR (Guy et al. 2010), using the coordinates found by the BLAST analysis, making sure that the operon was found on the same contig and including the genes at the 5' and 3' ends (gene *yegH* and *galG*, respectively). For simplicity, the orthologous genes detected in other species were named as "Ams" or "*ams*", but it is essential to recognize that individual names are designated for each protein or gene involved in exopolysaccharide synthesis across the different species.

**Table 1** List of the eighteen genomes selected for the comparative genomic study. The accession numbers under which each genome sequences was deposited (if sequence is fragmented, the assembly number was preferred), the associated size, habitat, related pathology and references to each genome sequence are given

<i>Erwinia</i> spp.	GenBank accession number/assembly number	Habitat/host	Plant pathogenicity	Reference
<i>Erwinia amylovora</i> CFBP 1430	FN434113.1	<i>Malus</i> sp. (apple tree)	Pathogen of Rosaceae Fire blight in apple and pear trees	(Smits et al. 2010b)
<i>Erwinia aphidicola</i> X001	JAMKCQ000000000.1	Gut aerobic bacteria from <i>Pea aphid</i>	Pathogen of Fabaceae Insect-associated bacteria	(Mouncey et al. unpublished)
<i>Erwinia billingiae</i> Eb661 <sup>T</sup>	FP236843.1	<i>Pyrus communis</i> (pear tree)	Non-pathogen	(Kube et al. 2010)
<i>Erwinia endophytica</i> A41C3	RQVV00000000.1	Potato stem	Non-pathogen	(Proença et al. 2019)
<i>Duffiella gerundensis</i> Eg EM595 <sup>T</sup>	LN907827.1	<i>Pyrus communis</i> (pear tree)	Non-pathogen	(Blom unpublished)
<i>Erwinia iniecta</i> B120 <sup>T</sup>	GCA_001267535.1	<i>Diuraphis noxia</i> (wheat aphid)	Non-pathogen Insect-associated bacteria	(Campillo et al. 2015)
<i>Erwinia mallotivora</i> BT-MARDI	GCA_000590885.1	<i>Carica papaya</i> and <i>Mallotus japonicus</i>	Pathogen of <i>Carica</i> sp. and Euphorbiaceae Papaya dieback disease	(Redzuan et al. 2014)
<i>Erwinia oleae</i> DAPP-PG531 <sup>T</sup>	GCA_000770305.1	<i>Olea europaea</i> (olive tree)	Associated to the olive knot	(Moretti et al. 2014)
<i>Erwinia persicina</i> NBRC 102418 <sup>T</sup>	GCA_001571305.1	<i>Piezodorus guildinii</i> (guts of redbanded stink bug) and <i>Leguminosae</i> (legume plants)	Pathogen of Leguminosae	(Hosoyama et al. unpublished)
<i>Erwinia phyllosphaerae</i> CMYE1 <sup>T</sup>	GCA_019132875.1	<i>Citrus maxima</i> (phyllosphere of pomelo)	Non-pathogen	(Pan et al. 2022)
<i>Erwinia piriflorinigrans</i> CFBP 5888 <sup>T</sup>	GCA_001050515.1	<i>Pyrus communis</i> (pear tree)	Pathogen of <i>Rosaceae</i>	(Smits et al. 2013)
<i>Erwinia psidii</i> IBSBF 435 <sup>T</sup>	GCA_003846135.1	<i>Psidium guava</i> (guava plants) and <i>Eucalyptus</i> spp. (eucalypt plantations)	Pathogen of Myrtaceae Bacterial Blight of Guava and Dieback of Eucalypt	(Alves et al. 2023)
<i>Erwinia pyrifoliae</i> DSM 12163 <sup>T</sup>	GCA_000026985.1	<i>Pyrus pyrofolia</i> (Asian pear tree/nashi)	Pathogen of Rosaceae Bacterial shoot blight of pear	(Smits et al. 2010a)
<i>Erwinia rhapontici</i> BY21311	CP085627.1	<i>Lens culinaris</i> Medik. (lentil) and <i>Cicer arietinum</i> L. (chickpea)	Pathogen of Fabaceae	(Wang unpublished)
<i>Erwinia tasmaniensis</i> Et1/99 <sup>T</sup>	GCA_000026185.1	<i>Malus</i> sp. (apple tree)	Non-pathogenic epiphytic bacteria, Considered pathoadapted	(Kube et al. 2008)
<i>Erwinia toletana</i> WS4403	GCA_017875465.1	<i>Olea</i> sp. (olive tree)	Associated to the pathogen of Oleaceae	(D'Agostino et al. unpublished)
<i>Erwinia tracheiphila</i> BHKY	GCA_021365465.1	<i>Cucurbita pepo</i> ssp. <i>Texana</i> (squash plant)	Pathogen of Cucurbitaceae	(LaSarre et al. 2022)
<i>Erwinia typographi</i> M043b	GCA_000773975.1	IT gut ( <i>Ips typographus</i> , bark beetles)	Non-pathogen Insect-associated bacteria	(Chan et al. unpublished)
<i>Pantoea stewartii</i> DC283	CP017581.1	<i>Zea mays</i> , USA		(Duong et al. unpublished)



**Table 2** For each protein of *Erwinia amylovora* CFBP 1430 considered as reference, locus tags, predicted function and gene name were listed

Protein	Function	gene	alternative gene name*	locus_tag	complement
AmsG	UDP-galactose-lipid carrier transferase	<i>amsG</i>	<i>wceG</i>	EAMY_2253	2294760..2296193
AmsH	Putative amylovoran export outer membrane protein	<i>amsH</i>	<i>wza</i>	EAMY_2252	2293505..2294647
AmsI	Low molecular weight protein-tyrosine-phosphatase	<i>amsI</i>	<i>wzb</i>	EAMY_2251	2293055..2293489
AmsA	Amylovoran biosynthesis membrane-associated protein	<i>amsA</i>	<i>wzc</i>	EAMY_2250	2290862..2293042
AmsB	Amylovoran biosynthesis glycosyltransferase	<i>amsB</i>	<i>wceB</i>	EAMY_2249	2289868..2290773
AmsC	Involved in oligosaccharide transport and polymerization	<i>amsC</i>	<i>wceL</i>	EAMY_2248	2288710..2289837
AmsD	Amylovoran biosynthesis glycosyltransferase	<i>amsD</i>	<i>wceN</i>	EAMY_2247	2287489..2288544
AmsE	Amylovoran biosynthesis glycosyltransferase	<i>amsE</i>	<i>wceM</i>	EAMY_2246	2286676..2287476
AmsF	Involved in polymerization	<i>amsF</i>	<i>wceF</i>	EAMY_2245	2284358..2286589
AmsJ	Amylovoran biosynthesis pyruvyl transferase	<i>amsJ</i>	<i>wcaK</i>	EAMY_2244	2282962..2284209
AmsK	Amylovoran biosynthesis glycosyltransferase	<i>amsK</i>	<i>wcaL</i>	EAMY_2243	2281734..2282957
AmsL	Involved in oligosaccharide assembly	<i>amsL</i>	<i>wzx</i>	EAMY_2242	2280397..2281737
YqhC	AraC-type DNA-binding domain-containing proteins	<i>yqhC</i>	-	EAMY_0464	543248..544165

\*The official name of each gene considered has been entered (Reeves et al. 1996)

## Protein similarity networks

Networks of orthologous proteins were predicted by using an in-house python software developed according to the pipeline described in two previous works (Ambrosino and Chiusano 2017; Ambrosino et al. 2018). In details orthologs, inferred by all-versus-all BLAST similarity searches (Camacho et al. 2009) together with a Bidirectional Best Hit (BBH) approach (Overbeek et al. 1999) were grouped into networks by using the NetworkX package (Hagberg et al. 2008).

The identification of Ams networks was obtained by using each Ams protein as a probe to scan the results and select the networks of interest. The graphical visualisation of these relationships was obtained via Cytoscape software (Shannon et al. 2003), generating networks in which nodes are proteins and a connection between each node (edges) were established upon inference of are homology.

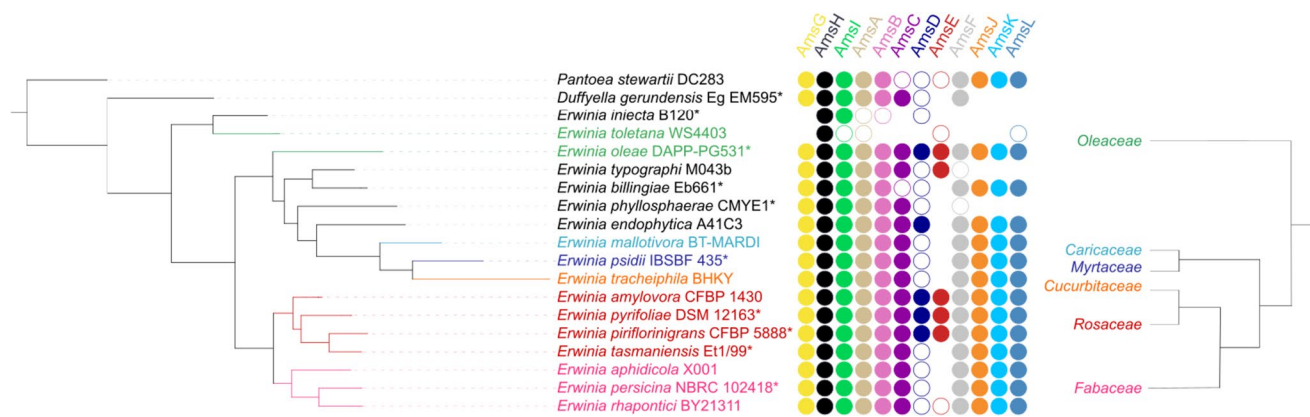
## Homology modelling and structural superposition

Three-dimensional (3D) homology models of the Ams proteins were built exploiting the AlphaFold server (Jumper et al. 2021). The top-score models generated were visualized using PyMOL v.2.4 (Schrödinger and DeLano 2020) and UCSF Chimera software (Pettersen et al. 2004). The MatchMaker tool of the suite UCSF Chimera was required for structural comparison. The search for conserved motifs (a recurring and ungapped sequence pattern that is approximately repeated within a set of related sequences) in some of the Ams proteins and YqhC was done using the MEME suite 5.5.1 (Bailey et al. 2015).

## Results

### Phylogenetic tree

To determine the phylogenetic relationships among the different *Erwinia* species, we built a phylogenetic tree using PhyloPhlAn3, a software that extracts the protein sequence of over 4000 markers and uses them as an input for the multiple sequence alignment and phylogenetic inference. The phylogenetic tree was built from a concatenated alignment of 5866 marker genes (total informative sites 49264), using RAXML (Fig. S1). The likelihood tree was created using the GTR substitution model with gamma-distributed rates. The final tree log-likelihood is -938871.63. All the conspecific genomes formed well-distinct monophyletic clusters, however within the same species, the genome of the isolates infecting different plants (i.e. *E. amylovora* RM1 and Ea644, that infect *Rubus* spp.) and the ones from isolates taken from geographically distant locations (*Erwinia* sp. Ejp617, reclassified as *E. pyrifoliae*), cluster as an outgroup to the other conspecific genomes. The earliest diverging species are *Duffryella gerundensis* (Soutar and Stavrinides 2022), *E. iniecta*, and *E. toletana* (Fig. 2). The species known to be pathogenic are not monophyletic, nevertheless the topology of the *Erwinia* spp. mirrors the one of the host plants (Cole and Hilger 2013). More specifically, Rosaceae and Fabaceae infecting species are monophyletic, whereas the ones infecting Myrtaceae, Cucurbitaceae and Oleaceae belong to distinct clades. The ANI ranged from 0.75 to 0.90, coherently with the values found among congeneric bacterial species, and the clustering pattern determined by the ANI is consistent with the one of the phylogenetic tree (Fig. S3).



**Fig. 2** Phylogenetic tree of the eighteen genomes considered for the study. On the right side of the picture, there is a paraphyletic representation of the angiosperm families including only the hosts of the *Erwinia* strains. In the middle, the genes are represented as dots, full

dots represent the presence of a complete CDS and empty dots represent the proteins found with an identity minor of 60%. Asterisks at the end of species name indicates the type strains.

### The *ams* operon displays a variable degree of conservation

The search for orthologs using the sequences of the *ams* genes from *E. amylovora* revealed that in *E. piriflorinigrans*, *E. pyrifoliae* and *E. oleae* orthologs for the whole operon were found with a complete sequence respecting all set parameters, whereas in the other species, the number of the proteins of interest ranged from 1 to 11 (Fig. 2). The most prevalent protein is AmsH (putative amylovoran export outer membrane protein), and the low molecular weight protein-tyrosine-phosphatase AmsI was found in all strains except *E. toletana*. The amylovoran biosynthesis membrane-associated protein AmsA, and the two proteins with roles in glycosyl transfer AmsB and AmsG were found in 16 genomes out of 18 studied ones, followed by the two proteins with transport and polymerization functions AmsC and AmsF in 15 and 14 genomes, respectively. AmsJ (with a Pyruvyl transferase function), AmsK (another galactose transferase) and AmsL (a transmembrane protein) in 13. The least present are the glycosyltransferases AmsD and AmsE, both found only in *E. amylovora*, *E. oleae*, *E. piriflorinigrans* and *E. pyrifoliae*. In all but one strains it was found a paralog of *amsG* located approximately 200 Kbp upstream of the operon. The only genome in which *amsG* was found as a single copy gene was in *E. tracheiphila*.

The regions from all genomes were extracted and compared in Fig. 3 where the main rearrangements of the *ams* genes and their homologs (*cps/wce* cluster) are visualised. Genes associated with the synthesis of exopolysaccharides in each species were always collinear. As pointed out in Kamber et al. 2012, there is a reciprocal translocation of the genes *amsC* and *amsB* between the genomes of *Pantoea* spp. and the *Erwinia* spp. In *E. billingiae* the homologous

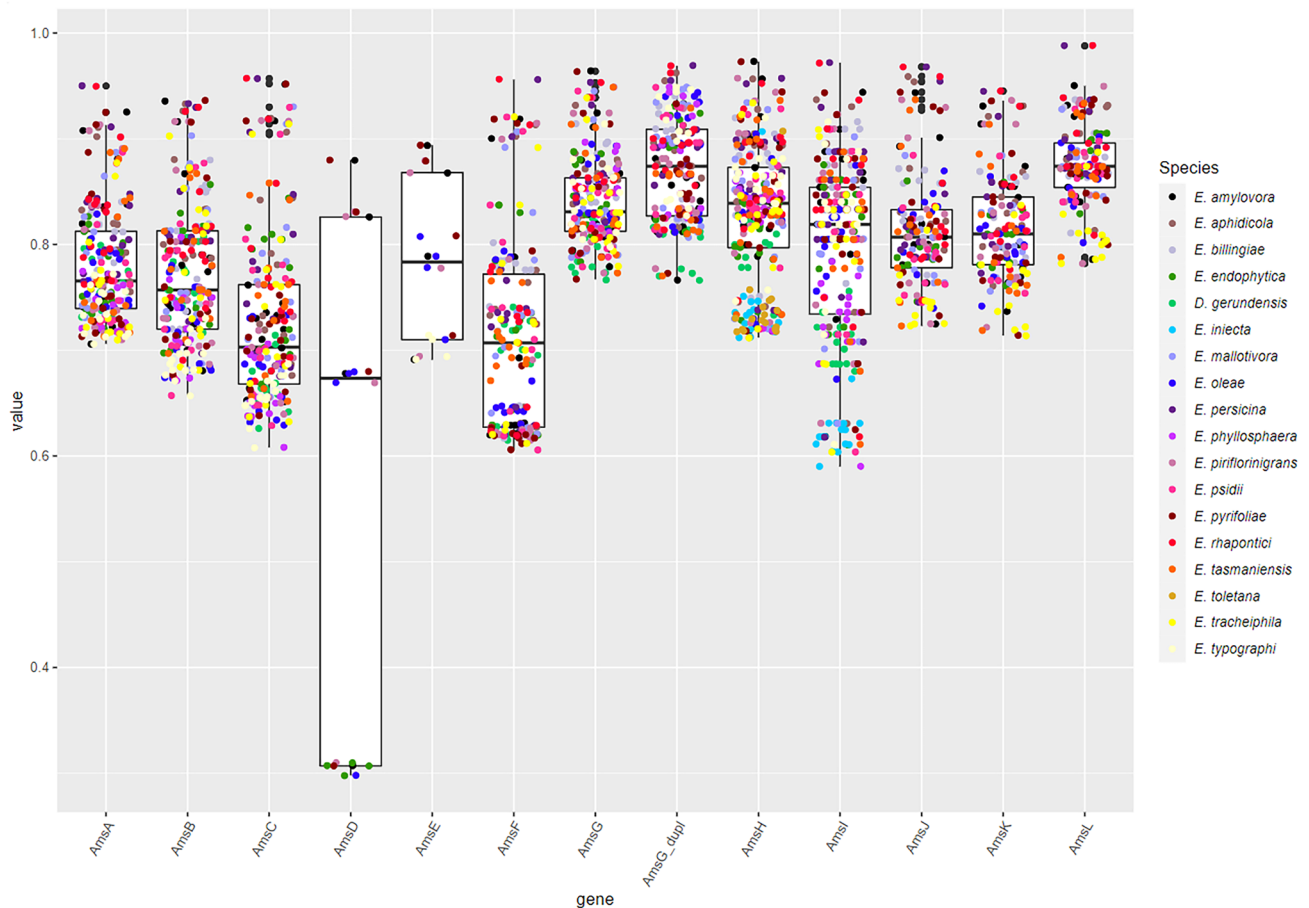
region to *amsC* contains a CDS that did not pass the threshold to be defined an orthologous protein. In *E. iniecta* and *E. toletana* a deletion of *amsG* occurred. In *D. gerundensis*, *E. phyllosphaerae* and *E. typographi*, where an homolog of *amsJ*, *amsK* and *amsL* is consistently missing, the homologous region presented a high sequence divergence. The pairwise identity matrix revealed different degrees of molecular divergence, with *amsG* and *amsL* being more conserved (with an id always over 75%), and *amsF* the most variable gene (Fig. 4).

### Prediction of orthologs and protein similarity networks

The BBH homology inference (Figs. 5 and S5) confirmed the one-to-one relationships for the proteins AmsA, AmsI and AmsH, which formed fully connected cliques (i.e. it was found in one copy for each genome). AmsB was also fully connected except for AmsB in *E. toletana*, which had a BBH only with *E. iniecta*. The same was found for AmsL, where AmsL in *E. toletana* had an homology inference with the one in *D. gerundensis*. The proteins AmsC, AmsF and AmsJ were located in almost all genomes with a one-to-one relationship. More complex patterns were discovered for the other Ams proteins. More specifically, the proteins AmsD and AmsK belonged to a homology network that was connected by a protein of *E. iniecta* annotated as a GTB-type-glycosyltransferase (WP\_052897582.1). The gene *amsG* was found with two paralogs in many species, although it was absent in *E. iniecta*. AmsE displays a one-to-one relationship with only seven other proteins and shows an orthology relationship with a protein in *E. persicina*, a glycosyltransferase family 2 protein (WP\_062742809.1), through a homology relationship inferred with *D. gerundensis*.

organisation, the species *E. pyrifoliae*, *E. piriflorinigrans* and *E. oleae* have been merged by taking *E. amylovora* as representative. The same was done for the species *E. aphidicola*, *E. endophytica*, *E. mallotivora*, *E. psidii*, *E. rhapontici*, *E. tasmaniensis* and *E. tracheiphila*., taking *E. persicina* as representative

According to Mann et al. 2013, *Erwinia amylovora* can be divided into host-specific genetically homogeneous strains colonizing different *Rosaceae* subfamilies (Mann et al. 2013). The analysis of the ANI among species suggests that this host-specific genomic homogeneity can also be detected at a higher taxonomic level. Indeed, the cluster of species infecting *Rosaceae* (*E. amylovora*, *E. pyrifoliae*, *E. piri-folinigrans*, *E. tasmaniensis*), is characterized by a higher similarity among them. The cluster of genomes infecting *Fabaceae* (*E. persicina*, *E. aphidicola*, *E. rhapontici*) also displays high similarity but to a minor extent in comparison with the *Rosaceae* infecting species. Furthermore, the high-resolution *Erwinia* phylogenetic tree topology was congruent with the phylogenetic tree of the *Angiosperm* families, suggesting that long-term interaction between host and pathogen has reciprocally influenced the evolution at the



**Fig. 4** Boxplots displaying the pairwise identity (y-axis) among all the combinations of each Ams protein (x-axis). Each boxplot represents the range of similarity score obtained by comparing the Ams

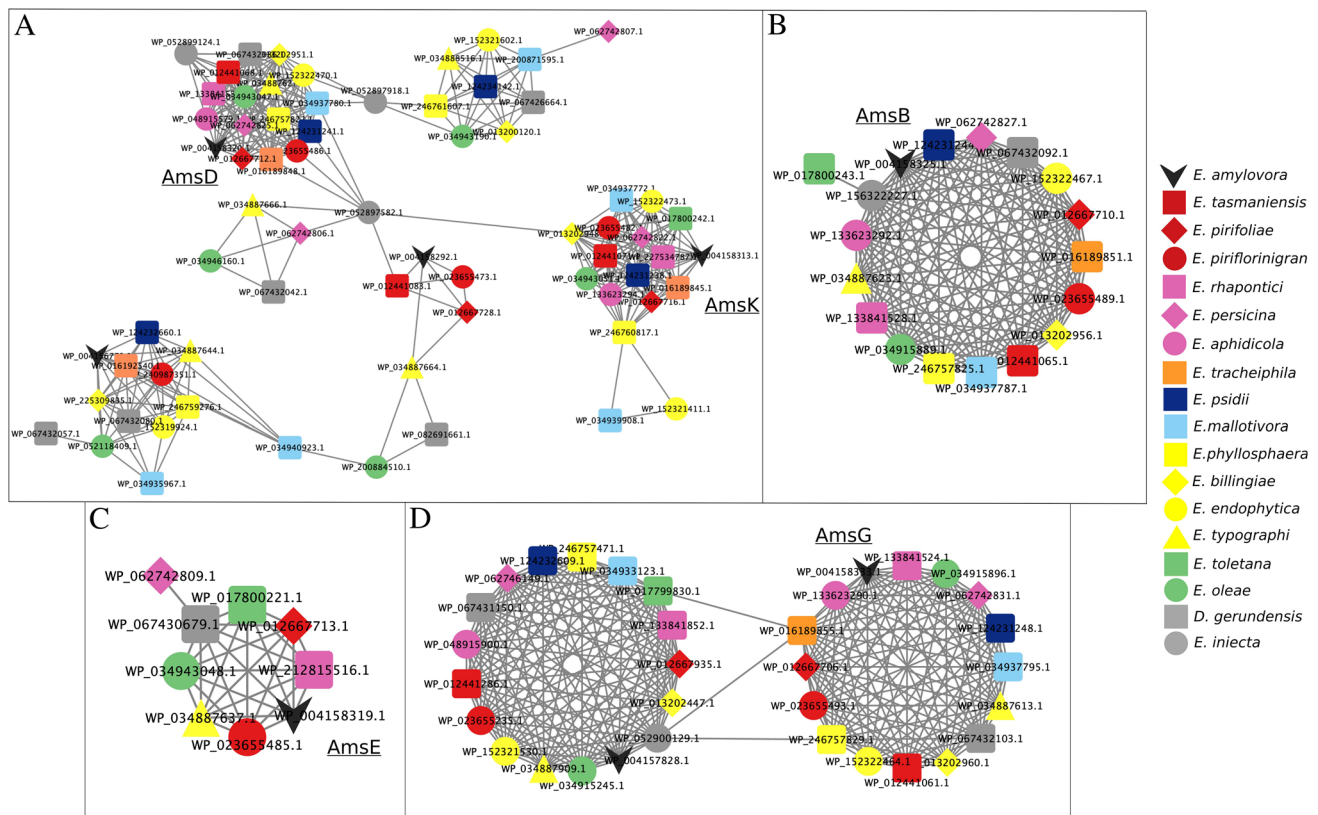
protein sequence from any given species of each of the selected 18 *Erwinia* species. Each dot represents a comparison and it is coloured according to the species used as a query

genomic level (Fig. 2). This phenomenon leads us back to the well-known “red queen hypothesis” (Papkou et al. 2019). Another evidence of this is the fact that *Rubus* infecting strains form a distinct group within the monophyletic clade of *E. amylovora* (Fig. S1). The species *Duffyella gerundensis* (formerly known as *Erwinia gerundensis*, Rezzonico et al. 2016) was also isolated from pome fruit, however it is placed very distant from the *Rosaceae*-infecting strains. This should be not interpreted as a contradiction because *D. gerundensis* changed the genus name, therefore it should be seen as an outgroup rather than an early diverging *Erwinia* lineage; and because this species is not pathogenic, it is rather commensal, therefore there is no strict co-evolution as in other host–pathogen relationships. Despite *E. tasmaniensis* is regarded as “pathoadapted”, housing the majority of pathogenicity factors for *Rosaceae*, no pathogenicity has been conclusively demonstrated thus far. While we have not specifically investigated this aspect in our research, it is plausible that regulatory mechanisms contribute to the pathogenicity observed. The same occurs with *E. toletana* and

*E. oleae*, which are grouped in the same cluster that infects the *Oleaceae* because although not strictly pathogenic, *E. toletana* co-occurs in infections caused by *Pseudomonas savastanoi* and modulates the disease severity (Buonaurio et al. 2015). Hence, the production of an EPS by *E. toletana* remains uncertain, yet there might exist other pathogenesis factors that exacerbate the condition caused by *P. savastanoi*.

Since the pathogenicity of *Erwinia amylovora* is connected to amylovoran secretion, we focused our attention on the biosynthetic operon *ams*. The rearrangements of genes in the operon could indicate that the gene products and/or the EPS structures are different (Moreno-Hagelsieb 2015). A previous study compared the *E. amylovora* *ams* operon to the *cps* region in *Pantoea stewartii*, which encodes for the biosynthetic pathway of stewartan, a related polysaccharide involved in the pathogenesis of this species (Langlotz et al. 2011). It was found that the pair of genes *amsB-C* and *amsD-E* underwent a translocation event between the two genomes. In the present study, we found that the order of the genes is conserved within the whole *Erwinia* genus. The



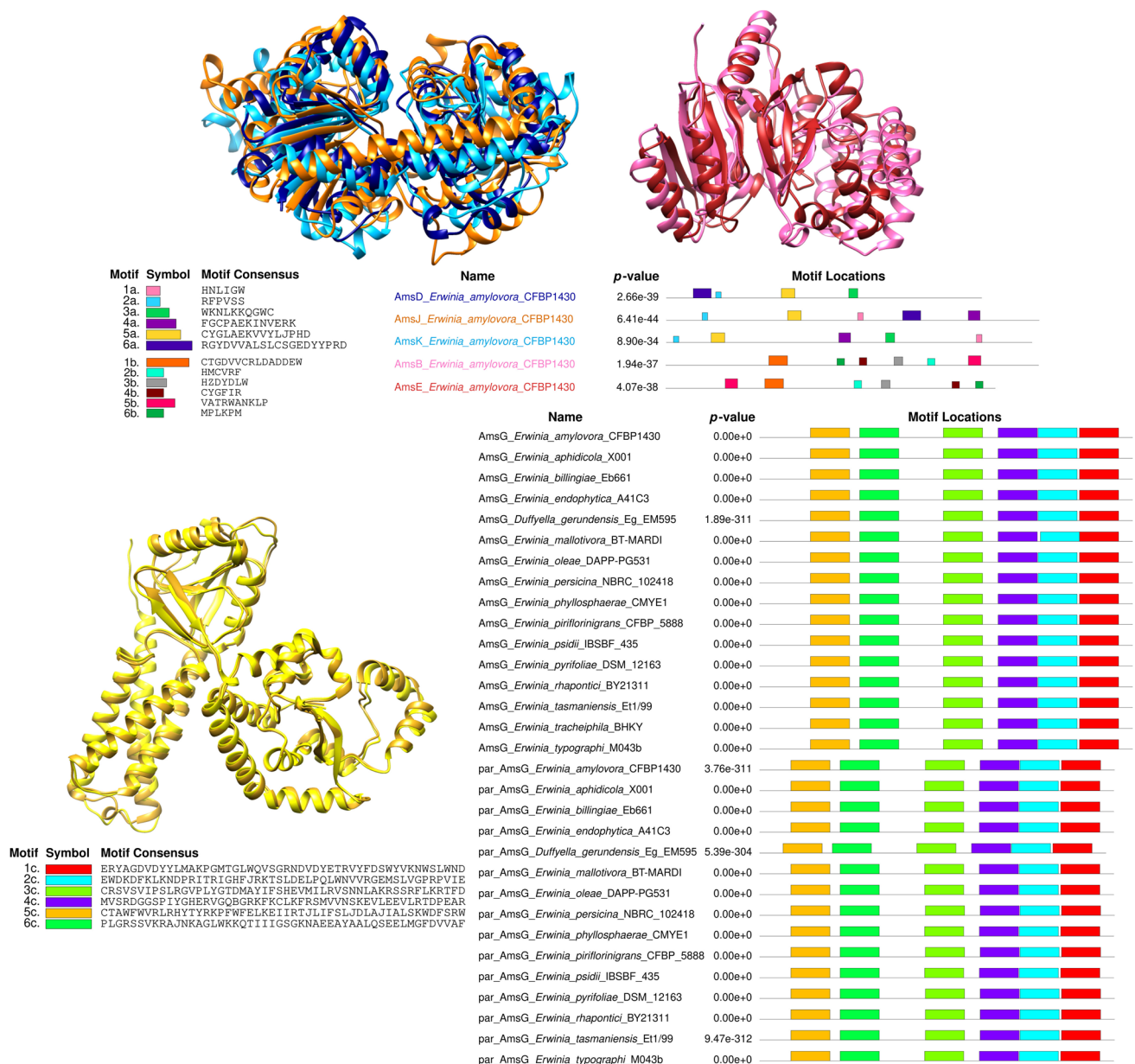


**Fig. 5** Panel of the most representative networks containing Ams orthologs in the considered *Erwinia* species. Circles, squares, triangles and diamonds represent proteins per species, according to the legend. Each species was represented with a symbol of the same colour code used throughout the manuscript and each protein was labelled with its ID. Grey lines represent orthology relationships

genes in the first part of the operon (*amsG*, *amsH* and *amsI*) are well conserved in their organization among all species in which they were detected (Fig. 3), additionally, they display a relatively lower level of divergence compared to the other genes (Fig. 4). Similarly, also *amsA* and *amsB* are well conserved, but their sequence similarity is not higher compared to the one of the rest of the genes. Other genes, such as *amsD* and *amsE* show re-arrangement such as translocations (*amsD* in *E. endophytica*) or were not found in many species. Indeed, the *amsE* gene is relatively rare among various *Erwinia* spp.. Yet, as demonstrated by previous research conducted by Kamber et al. 2012, there exists a notable similarity between other proteins located at the same position (such as WbdN), which, in turn, share only a 25% identity with AmsE. Furthermore, the same earlier research validates the observation that in the genomic regions corresponding to the location of *amsCDE*, the operon of *E. billingiae* has two other distinct genes, identified as an acyltransferase and a SGNH/GDSL hydrolase family protein, respectively. Those genes (particularly *amsD*) display a lower similarity among all species. However the network approach could

detect *amsD* in different species through a BBH approach, the wide range of sequence divergence among the *amsD* genes imply that the molecular clock runs faster for those genes than it does for the other genes in the operon. Probably due to a weaker or disruptive selective pressure. A similar situation was found also for the genes *amsJ-K* and *amsL* in the three non-pathogens *D. gerundensis*, *E. phyllosphaerae*, and *E. typographi*, suggesting that the biosynthetic pathway results in a polysaccharide with different secondary chains.

As demonstrated for the prokaryotic isomerases (Álvarez-Lugo and Becerra 2023), the duplication of genes is often associated with an increased expression (i.e. “gene dosage effect”). A paralog of the *amsG* gene was consistently found in all except one *Erwinia* species (that is *E. tracheiphila*). Being the first protein, AmsG catalyses a critical step in the biosynthetic pathways: it triggers the amylovoran unit synthesis by adding the first molecule of UDP-galactose to the membrane-bound lipid transporter. The presence of a paralog, which for example in *E. amylovora* is catalogued as another undecaprenyl-phosphate galactose phosphotransferase (*wbaP* gene, EAMY\_1987), highlights its crucial role



**Fig. 6** Superposition of the various Ams structures obtained with AlphaFold and MEME analysis in which are reported the conserved motifs (recurring, fixed-length patterns) between the considered proteins. On the top left is the overlay of AmsD (blue), AmsJ (orange)

and AmsK (light blue) structures. Beside, in red is the AmsE structure and in pink AmsB. Below we find in yellow AmsG and its “paralog” (par\_AmsG) with the corresponding MEME analysis

over the adjacent genes in the *ams* operon. Indeed, the down-regulation of *amsG* itself significantly hampered amylovoran secretion and motility (Yuan et al. 2022). The gene tree based on the sequences of both *amsG* paralogs in all species showed a clustering pattern that grouped together the two paralogs in different species rather than both the gene copies in each species, suggesting that the duplication occurred before the differentiation of the *Erwinia* spp. (Fig. S2). In *E. toletana* and *E. iniecta* the first gene of the operon *amsG* was deleted, implying that in those species the exopolysac-

charides biosynthetic pathway starts with the catalysis of an alternative protein or potentially not start at all, consequently leading to a significant change in their virulence potential.

Other exopolysaccharides have been characterized in the *Erwinia* species, one example is the pyrifolian, secreted by *E. pyrifoliae* that, as it was observed by Kim et al. 2002, displays a remarkable similarity with amylovoran in its biochemical structure. Pyrifolian is encoded by the *cps* operon, which, as we also observed, is orthologous to the *ams* operon in *E. amylovora*. Moreover, the cluster discovered in both *E.*

*tasmaniensis* and *E. billingiae* indicates that the structure of EPS of those species may bear a closer resemblance to Stewartan, the EPS secreted by *Pantoea stewartii* (Schollmeyer et al. 2012). A particular case encountered is that of *E. tracheiphila*, which, unlike other species, uses insect vectors for disease transmission (Mitchell and Hanks 2009). *amsD* and *amsE* were found at a low identity or not at all, and the gene *amsG* was found as a single copy. Evidence suggests that *E. tracheiphila* exhibits decreased EPS secretion (Mason 2013), implying that the different mechanisms of exopolysaccharide production and secretion could be connected to environmental adaptation. We screened this species also for the *yqhC* gene, and the Arac-like transcriptional regulator was not found, whereas it was found in all other *Erwinia* spp. (See supplementary material).

The use of the orthology network confirmed for most of the proteins a one-to-one relationship (Fig. S5). For AmsB and AmsL in *E. toletana*, it was possible to define an orthology relationship only with *E. iniecta* or *D. gerundensis*, respectively (Fig. 5B). This could be due to the higher sequence divergence of the *ams* genes in those species. The glycosyltransferases AmsD and AmsK belonged to the same network, however, they display a fourth-degree connection through an orthologous protein of *E. iniecta* that was the only node connecting the network of AmsD with the one of AmsK (Fig. 5A). These two proteins, as well as AmsE, belong to the large family of glycosyltransferase, which explains the presence of multiple proteins detected as paralogs to the ones of interest.

The relationships between protein sequence and 3D structure are complex, given the involvement of numerous factors in its determination (Krissinel et al. 2007). It could be expected that proteins with a more similar primary sequence display a high degree of overlap in their 3D structure. Indeed, the 3D structures of AmsD and AmsK overlapped, confirming their relatedness inferred also through the network analysis (Figs. 5 and 6). Conversely, proteins such as AmsB and AmsE had a similar 3D structure despite a higher divergence at the primary sequence level. This suggests a strict selection acting on the 3D structure yet tolerating a higher amino acid substitution rate. Indeed, it is also likely that the various glycosyltransferases resemble each other in structure because they catalyse the same reaction on slightly different substrates. Further structural analysis eventually revealed that AmsG and its duplicate, have similar sequence and the same secondary structure, emphasising how the function of the pathway's first glycosyltransferase may be crucial in the production of amylovoran. However, the structural analysis cannot be considered as predictive of a similar function. Future studies, aimed at understanding the substrate range of the two paralogues would be more reliable.

## Conclusion

The availability of genome sequences for many bacterial species provides a valuable resource for comparative genomics. The study of the *ams* operon organisation in different *Erwinia* strains allows a deeper understanding of the evolution of the exopolysaccharide biosynthesis gene clusters in those species. Generally, the operon gene order and the protein sequence exhibit resemblance to one another among the species, suggesting the preservation of their functions. Minor differences have been detected in the genomes from *Erwinia* species distantly related to *E. amylovora*.

The main findings regard the presence of a paralog of the glycosyltransferase *amsG* in most of the species, and the progressive sequence divergence of the genes at the 5'-end of the operon. This confirms the key role of the AmsG protein in the production of EPS and suggests the possibility of modifications to the chemical structure of the exopolysaccharide produced by each species, especially in the secondary chain. Moreover, the deletion of AmsG in *E. iniecta*, and *E. toletana* and the *yqhC* gene in *E. tracheiphila* suggests that in those species the exopolysaccharides production follows different biosynthetic pathways. Protein similarity networks revealed at least a one-to-one relationship between most of the Ams proteins, although in some species there were more complex orthology patterns among the *ams* genes. The 3D structure analysis allowed us to put forward hypotheses about the selective pressures acting at different levels on some of the proteins. Future research is needed in determining the effect of the genomic rearrangements found in the different species on the virulence and the chemical composition of the EPS, and in the last instance on the virulence of the species.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s42161-023-01532-9>.

**Author contributions** Conceptualization, L.C.; methodology, L.C., A.E.; software, L.C., L.A. and A.E.; validation, L.C., L.A. and A.E.; formal analysis, L.C., L.A., Sh.B., L.M.I. and A.E.; investigation, L.C., Sh.B. and A.E.; resources, S.B. and S.P.; data curation, L.C., L.M.I. and A.E.; writing-original draft preparation, L.C. and A.E.; writing-review and editing, L.C., Sh.B. and A.E.; visualization, L.C.; supervision, S.B. and S.P.; project administration, S.B.; funding acquisition, S.B.. All authors have read and agreed to the published version of the manuscript.

**Funding** Open access funding provided by Libera Università di Bolzano within the CRUI-CARE Agreement.

**Data availability** We used publicly available data see Table 1 and Table 2. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the



source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215(3):403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Álvarez-Lugo A, Becerra A (2023) The fate of duplicated enzymes in prokaryotes: The case of isomerases. *J Mol Evol* 91(1):76–92. <https://doi.org/10.1007/s00239-022-10085-x>
- Alves FHNS, Pereira IC, Santos SA, Monteiro ALR, Vidigal PMP, Alfenas AC et al (2023) Comprehensive prediction of plant cytoplasmic and apoplastic effectors underlying *Erwinia psidii* pathogenicity. *Plant Pathol* 72:130–143. <https://doi.org/10.1111/ppa.13636>
- Ambrosino L, Chiusano ML (2017) Transcriptologs: A transcriptome-based approach to predict orthology relationships. *Bioinform Biol Insights* 11. <https://doi.org/10.1177/1177932217690136>
- Ambrosino L, Ruggieri V, Bostan H, Miralto M, Vitulo N, Zouine M et al (2018) Multilevel comparative bioinformatics to investigate evolutionary relationships and specificities in gene annotations: An example for tomato and grapevine. *BMC Bioinform* 19(Suppl 15). <https://doi.org/10.1186/s12859-018-2420-y>
- Asnicar F, Thomas AM, Beghini F, Mengoni C, Manara S, Manghi P et al (2020) Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. *Nature Communications* 11(1):1–10. <https://doi.org/10.1038/s41467-020-16366-7>
- Bailey TL, Johnson J, Grant CE, Noble WS (2015) The MEME suite. *Nucleic Acids Res* 43(W1):W39–W49. <https://doi.org/10.1093/nar/gkv416>
- Bellemann P, Bereswill S, Berger S, Geider K (1994) Visualization of capsule formation by *Erwinia amylovora* and assays to determine amylovoran synthesis. *Int J Biol Macromol* 16(6):290–296. [https://doi.org/10.1016/0141-8130\(94\)90058-2](https://doi.org/10.1016/0141-8130(94)90058-2)
- Benini S (2021) Structural and functional characterization of proteins from the fire blight pathogen *Erwinia amylovora*. A review on the state of the art. *J Plant Pathol* 103(Suppl 1):51–63. <https://doi.org/10.1007/s42161-020-00682-4>
- Benini S, Toccafondi M, Rejzek M, Musiani F, Wagstaff BA, Wuerges J, Cianci M, Field RA (2017) Glucose-1-phosphate uridylyltransferase from *Erwinia amylovora*: Activity, structure and substrate specificity. *BBA - Proteins Proteomics* 1865:1348–1357. <https://doi.org/10.1016/j.bbapap.2017.08.015>
- Bugert P, Geider K (1995) Molecular analysis of the *ams* operon required for exopolysaccharide synthesis of *Erwinia amylovora*. *Mol Microbiol* 15(5):917–933. <https://doi.org/10.1111/j.1365-2958.1995.tb02361.x>
- Buonaurio R, Moretti C, Da Silva DP, Cortese C, Ramos C, Venturi V (2015) The olive knot disease as a model to study the role of interspecies bacterial communities in plant disease. *Front Plant Sci* 6(June):1–12. <https://doi.org/10.3389/fpls.2015.00434>
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: Architecture and applications. *BMC Bioinform* 10:1–9. <https://doi.org/10.1186/1471-2105-10-421>
- Campillo T, Luna E, Portier P, Fischer-Le Saux M, Lapitan N, Tisserat NA, Leach JE (2015) *Erwinia iniecta* sp. nov., isolated from Russian wheat aphid (*Diuraphis noxia*). *Int J Syst Evol Microbiol* 65(10):3625–3633. <https://doi.org/10.1099/ijsem.0.000466>
- Cole TCH, Hilger H (2013) *Angiosperm Phylogeny Poster* (2014):63166. <https://doi.org/10.7287/peerj.preprints.2320>
- Esposito A, Tamburini S, Triboli L, Ambrosino L, Chiusano ML, Jousson O (2019) Insights into the genome structure of four acetogenic bacteria with specific reference to the Wood-Ljungdahl pathway. *MicrobiologyOpen* 8(12):1–15. <https://doi.org/10.1002/mbo3.938>
- Griffith CS, Sutton TB, Peterson PD (2003) Fire blight: The foundation of phyto bacteriology. APS Press, St. Paul, MN, U.S.A
- Guy L, Kultima JR, Andersson SG (2010) genoPlotR: comparative gene and genome visualization in R. *Bioinformatics* 26(18):2334–2335. <https://doi.org/10.1093/bioinformatics/btq413>
- Hagberg A, Swart PJ, Schult DA (2008) Exploring network structure, dynamics, and function using NetworkX. *Proceedings of the 7th Python in Science Conference (SciPy2008)*, pp. 11–15, Pasadena, CA USA. [http://conference.scipy.org/proceedings/SciPy2008/paper\\_2/](http://conference.scipy.org/proceedings/SciPy2008/paper_2/)
- Heitefuss R (2012) Fire blight, history, biology, and management. *J Phytopathol* 160(7–8):440–440. <https://doi.org/10.1111/j.1439-0434.2012.01922.x>
- Hyatt D, Chen GL, LoCascio PF et al (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform* 11:119. <https://doi.org/10.1186/1471-2105-11-119>
- Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O et al (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* 596(7873):583–589. <https://doi.org/10.1038/s41586-021-03819-2>
- Kamber T, Smits THM, Rezzonico F (2012) Genomics and current genetic understanding of *Erwinia amylovora* and the fire blight antagonist *Pantoea vagans*. *Trees* 26:227–238. <https://doi.org/10.1007/s00468-011-0619-x>
- Kelm O, Kiecker C, Geider K, Bernhard F (1997) Interaction of the regulator proteins RcsA and RcsB with the promoter of the operon for amylovoran biosynthesis in *Erwinia amylovora*. *Mol Gen Genet* 256(1):72–83. <https://doi.org/10.1007/s004380050547>
- Kharadi RR, Schachterle JK, Yuan X, Castiblanco LF, Peng J, Slack SM (2021) Genetic dissection of the *Erwinia amylovora* disease cycle. *Annu Rev Phytopathol* 59(1):191–212. <https://doi.org/10.1146/annurev-phyto-020620-095540>
- Kharadi RR, Sundin GW (2022) CsrD regulates amylovoran biosynthesis and virulence in *Erwinia amylovora* in a novel cyclic-di-GMP dependent manner. *Mol Plant Pathol* 23(8):1154–1169. <https://doi.org/10.1111/mpp.13217>
- 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. <https://doi.org/10.1099/00221287-148-12-4015>
- Koczan JM, McGrath MJ, Zhao Y, Sundin GW (2009) Contribution of *Erwinia amylovora* exopolysaccharides amylovoran and levan to biofilm formation: Implications in pathogenicity. *Phytopathology* 99(11):1237–1244. <https://doi.org/10.1094/PHYTO-99-11-1237>
- Krissinel E (2007) On the relationship between sequence and structure similarities in proteomics. *Bioinformatics* 23(6):717–723. <https://doi.org/10.1093/bioinformatics/btm006>
- Kube M, Migdoll AM, Gehring I (2010) Genome comparison of the epiphytic bacteria *Erwinia billingiae* and *E. tasmaniensis* with the pear pathogen *E. pyrifoliae*. *BMC Genomics* 11:393. <https://doi.org/10.1186/1471-2164-11-393>
- Kube M, Migdoll AM, Müller I, Kuhl H, Beck A, Reinhardt R, Geider K (2008) The genome of *Erwinia tasmaniensis* strain Et1/99, a non-pathogenic bacterium in the genus *Erwinia*. *Environ Microbiol* 10:2211–2222. <https://doi.org/10.1111/j.1462-2920.2008.01639.x>



- Langlotz C, Schollmeyer M, Coplin DL, Nimtz M, Geider K (2011) Biosynthesis of the repeating units of the exopolysaccharides amylovoran from *Erwinia amylovora* and stewartan from *Pantoea stewartii*. *Physiol Mol Plant Pathol* 75(4):163–169. <https://doi.org/10.1016/j.pmp.2011.04.001>
- LaSarre B, Olawole OI, Paulsen AA, Halverson LJ, Gleason ML, and Beattie, G. A. (2022) Complete genome sequences of four strains of *Erwinia tracheiphila*: A resource for studying a bacterial plant pathogen with a highly complex genome. *Mol Plant Microbe Interact* 35(6):500–504. <https://doi.org/10.1094/MPMI-01-22-0008-A>
- Letunic I, Bork P (2021) Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* 49(W1):W293–W296. <https://doi.org/10.1093/nar/gkab301>
- Malnoy M, Martens S, Norelli JL, Barny MA, Sundin GW, Smits THM, Duffy B (2012) Fire blight: Applied genomic insights of the pathogen and host. *Annu Rev Phytopathol*. <https://doi.org/10.1146/annurev-phyto-081211-172931>
- Mann RA, Smits THM, Bühlmann A, Blom J, Goesmann A, Frey JE et al (2013) Comparative genomics of 12 strains of *Erwinia amylovora* identifies a pan-genome with a large conserved core. *PLoS ONE* 8(2). <https://doi.org/10.1371/journal.pone.0055644>
- 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*. <https://doi.org/10.1111/j.1364-3703.2012.00804.x>
- Mason SN (2012) Development of molecular tools for genetic analysis of *Erwinia tracheiphila* pathogenesis. Diss. Tennessee State University
- Metzger M, Bellemann P, Bugert P, Geider K (1994) Genetics of galactose metabolism of *Erwinia amylovora* and its influence on polysaccharide synthesis and virulence of the fire blight pathogen. *J Bacteriol* 176:450–459. <https://doi.org/10.1128/jb.176.2.450-459.1994>
- Mitchell RF, Hanks LM (2009) Insect frass as a pathway for transmission of bacterial wilt of cucurbits. *Environ Entomol* 38(2):395–403. <https://doi.org/10.1603/022.038.0212>
- Moreno-Hagelsieb G (2015) The power of operon rearrangements for predicting functional associations. *Comput Struct Biotechnol J* 13:402–406. <https://doi.org/10.1016/j.csbj.2015.06.002>
- Moretti C, Cortese C, Passos da Silva D, Venturi V, Firrao G, Buonauro R (2014) Draft Genome Sequence of *Erwinia oleae*, a Bacterium Associated with Olive Knots Caused by *Pseudomonas savastanoi* pv. *savastanoi*. *Genome Announc* 2(6):e01308–14. <https://doi.org/10.1128/genomeA.01308-14>
- Overbeek R, Fonstein M, D'Souza M, Push GD, Maltsev N (1999) The use of gene clusters to infer functional coupling. *Proc Natl Acad Sci USA* 96(6):2896–2901. <https://doi.org/10.1073/pnas.96.6.2896>
- Pan MK, Feng GD, Yao Q, Li J, Liu C, Zhu H (2022) *Erwinia phyllosphaerae* sp. nov., a novel bacterium isolated from phyllosphere of pomelo (*Citrus maxima*). *Int J Syst Evol Microbiol* 72(4). <https://doi.org/10.1099/ijsem.0.005316>
- Papkou A, Guzella T, Yang W, Koepper S, Pees B, Schalkowski R (2019) The genomic basis of red queen dynamics during rapid reciprocal host–pathogen coevolution. *Proc Natl Acad Sci USA* 116(3):923–928. <https://doi.org/10.1073/pnas.1810402116>
- Park DH, Thapa SP, Choi BS, Kim WS, Hur JH, Cho JM, Lim JS, Choi IY, Lim CK (2011) Complete genome sequence of Japanese *Erwinia* strain ejp617, a bacterial shoot blight pathogen of pear. *J Bacteriol* 193(2):586–587. <https://doi.org/10.1128/JB.01246-10>
- Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera - A visualization system for exploratory research and analysis. *J Comput Chem* 25(13):1605–1612. <https://doi.org/10.1002/jcc.20084>
- Piqué N, Miñana-Galbis D, Merino S, Tomás JM (2015) Virulence factors of *Erwinia amylovora*: A review. *OPEN ACCESS Int J Mol Sci* 16:16. <https://doi.org/10.3390/ijms160612836>
- Polsinelli I, Borruso L, Caliandro R, Triboli L, Esposito A, Benini S (2019) A genome-wide analysis of desferrioxamine mediated iron uptake in *Erwinia* spp. reveals genes exclusive of the *Rosaceae* infecting strains. *SciRep* 9(1). <https://doi.org/10.1038/s41598-019-39787-x>
- Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK (2016) Genomics and taxonomy in diagnostics for food security: Soft-rotting enterobacterial plant pathogens. *Anal Methods* 8(1):12–24. <https://doi.org/10.1039/c5ay02550h>
- Proença DN, Heine T, Senges CHR, Bandow JE, Morais PV, Tischler D (2019) Bacterial metabolites produced under iron limitation kill pinewood nematode and attract *Caenorhabditis elegans*. *Front Microbiol* 19(10):2166. <https://doi.org/10.3389/fmicb.2019.02166>
- Redzuan RA, Abu Bakar N, Rozano L, Badrun R, Mat Amin N, Mohd Raih MF (2014) Draft genome sequence of *Erwinia mallotivora* BT-MARDI, causative agent of papaya dieback disease. *Genome Announc* 2(3):e00375–e414. <https://doi.org/10.1128/genomeA.00375-14>
- Reeves PR, Hobbs M, Valvano MA, Skurnik M, Whitfield C, Coplin D, Kido N, Klena J, Maskell D, Raetz CR, Rick PD (1996) Bacterial polysaccharide synthesis and gene nomenclature. *Trends Microbiol* 4(12):495–503. [https://doi.org/10.1016/s0966-842x\(97\)82912-5](https://doi.org/10.1016/s0966-842x(97)82912-5)
- Rezzonico F, Smits THM, Born Y, Blom J, Frey JE, Goesmann A, Cleenwerck I, de Vos P, Bonaterra A, Duffy B, Montesinos E (2016) *Erwinia gerundensis* sp. nov., a cosmopolitan epiphyte originally isolated from pome fruit trees. *Int J Syst Evol Microbiol* 66(3):1583–1592. <https://doi.org/10.1099/ijsem.0.000920>
- Sahebi M, Tarighi S, Taheri P (2022) The Arac-like transcriptional regulator YqhC is involved in pathogenicity of *Erwinia amylovora*. *J Appl Microbiol* 132(2):1319–1329. <https://doi.org/10.1111/jam.15286>
- Salomone-Stagni M, Musiani F, Benini S (2016) Characterization and 1.57 Å resolution structure of the key fire blight phosphatase AmsI from *Erwinia amylovora*. *Acta Crystallogr Sect F Struct Biol Commun* 72(12):903–910. <https://doi.org/10.1107/S2053230X16018781>
- Sebahia M, Bocsanczy AM, Biehl BS, Quail MA, Perna NT, Glasner JD, DeClerck GA, Cartinhour S, Schneider DJ, Bentley SD, Parkhill J, Beer SV (2010) Complete genome sequence of the plant pathogen *Erwinia amylovora* strain ATCC 49946. *J Bacteriol* 192(7):2020–2021. <https://doi.org/10.1128/JB.00022-10>
- Schollmeyer M, Langlotz C, Huber A, Coplin DL, Geider K (2012) Variations in the molecular masses of the capsular exopolysaccharides amylovoran, pyrifolan and stewartan. *Int J Biol Macromol* 50:518–522. <https://doi.org/10.1016/j.ijbiomac.2012.01.003>
- Schrödinger L, DeLano W (2020) PyMOL
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13(11):2498–504. <https://doi.org/10.1101/gr.1239303>
- Smits HMT, Duffy B, Sundin GW, Zhao YF, Rezzonico F (2017) *Erwinia amylovora* in the genomics era: From genomes to pathogen virulence, regulation, and disease control strategies. *J Plant Pathol* 99(Special issue):7–23. <https://doi.org/10.4454/jpp.v99i0.391>
- Smits HMT, Rezzonico F, Kamber T, Blom J, Goesmann A, Frey JE, 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(4):384–393. <https://doi.org/10.1094/MPMI-23-4-0384>
- Smits TH, Jaenicke S, Rezzonico F (2010a) Complete genome sequence of the fire blight pathogen *Erwinia pyrifoliae* DSM

- 12163T and comparative genomic insights into plant pathogenicity. *BMC Genomics* 11:2. <https://doi.org/10.1186/1471-2164-11-2>
- Smits TH, Rezzonico F, López MM, Blom J, Goesmann A, Frey JE, Duffy B (2013) Phylogenetic position and virulence apparatus of the pear flower necrosis pathogen *Erwinia piriflorinigrans* CFBP 5888T as assessed by comparative genomics. *Syst Appl Microbiol* 36(7):449–456. <https://doi.org/10.1016/j.syapm.2013.04.003>
- Soutar CD, Stavrínides J (2022) Phylogenomic analysis of the *Erwiniaceae* supports reclassification of *Kalamielliella piersonii* to *Pantoea piersonii* comb. nov. and *Erwinia gerundensis* to the new genus *Duffiyella* gen. nov. as *Duffiyella gerundensis* comb. nov. *Mol Genet Genom* 297(1):213–225. <https://doi.org/10.1007/s00438-021-01829-3>
- Tamura K, Stecher G, Kumar S (2021) MEGA11: molecular evolutionary genetics analysis version 11. *Mol Biol Evol* 38(7):3022–3027. <https://doi.org/10.1093/molbev/msab120>
- Thomson SV (1986) The role of the stigma in fire blight infections. *Phytopathology* 76:476–482
- Vanneste JL (2000) What is fire blight? Who is *Erwinia amylovora*? How to control it? Fire blight: the disease and its causative agent, *Erwinia amylovora* 1–6. <https://doi.org/10.1079/9780851992945.0001>
- Wang D, Korban SS, Pusey PL, Zhao Y (2012b) AmyR is a novel negative regulator of amylovoran production in *Erwinia amylovora*. *PLoS ONE* 7(9):45038. <https://doi.org/10.1371/journal.pone.0045038>
- Wang D, Qi M, Calla B, Korban SS, Clough SJ, Cock PJA et al (2012a) Genome-wide identification of genes regulated by the *rcs* phosphorelay system in *Erwinia amylovora*. *Mol Plant Microbe Interact* 25(1):6–17. <https://doi.org/10.1094/MPMI-08-11-0207>
- Wang DP, Clough SJ, Zhao YF, Sundin GW, Korban SS, Toth IK (2011) Regulatory genes and environmental regulation of amylovoran biosynthesis in *Erwinia amylovora*. *Acta Hort* 896:195–202. <https://doi.org/10.17660/ActaHortic.2011.896.25>
- Yu M, Singh J, Khan A, Sundin GW, Zhao Y (2020) Complete genome sequence of the fire blight pathogen strain *Erwinia amylovora* Ea1189. *Mol Plant Microbe Interact* 33(11):1277–1279. <https://doi.org/10.1094/MPMI-06-20-0158-A>
- Yuan X, Eldred LI, Sundin GW (2022) Exopolysaccharides amylovoran and levan contribute to sliding motility in the fire blight pathogen *Erwinia amylovora*. *Environ Microbiol* 24(10):4738–4754. <https://doi.org/10.1111/1462-2920.16193>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.