

RESEARCH

Open Access



Genome analysis reveals *Erwinia amylovora* in China is closely related to those of Kyrgyzstan and Kazakhstan

Jianan Wang^{1,2}, Yuqiang Zhao³, Shitong Wu^{1,2}, Xinmin Gu^{1,2}, Xinran Liu^{1,2}, Yanli Tian^{1,2*} , Baishi Hu^{1,2,4,5*} and Youfu Zhao⁶

Abstract

Erwinia amylovora, the causal agent of fire blight disease, is a destructive agricultural plant pathogen in the *Rosaceae* family. The disease was first described in North America and has now spread to more than 50 countries. The disease has recently been reported in China's neighboring countries, including Russia, South Korea, Kyrgyzstan, and Kazakhstan. In 2016, the typical symptoms of fire blight in pear tree shoots and leaves were observed in Xinjiang province and later, it spread eastward to Gansu province, posing a threat on the national pear and apple industries. In this study, fire blight pathogen was isolated from diseased plants and the whole genome sequencing was performed for 14 (12 from Xinjiang and 2 from Gansu), 3, and 2 isolates from China, Kyrgyzstan, and Kazakhstan, respectively. Phylogenomic analysis showed that the Chinese isolates are closely related to those from Kyrgyzstan and Kazakhstan. Genome annotation reveals that most of the gene families and metabolic pathways are conserved in *E. amylovora* isolates. These results provide useful information in understanding how the fire blight occurs in China and what measures should be taken to eradicate the disease.

Keywords Fire blight, *Erwinia amylovora*, Comparative genomics, Phylogenetic analysis

Background

Fire blight, caused by the Gram-negative bacterium *Erwinia amylovora*, is a devastating disease of rosaceous plants (i.e., apple and pear). Since it was discovered, fire blight has been considered as the most destructive disease of pome fruit trees and a major threat to apple and pear production. The typical symptom of fire blight shows water soaking first, followed by wilting and rapid necrosis, leaving infected tissues with a scorched and blackened appearance (Eastgate 2000). These symptoms can be seen on all above-ground tissues, including blossoms, fruits, shoots, branches, limbs, and in the rootstock near the graft union on the lower trunk (Zhao et al. 2019). The quarantine of *E. amylovora* imposes additional economic costs by phytosanitary measures (Krissoff and Calvin 1998).

*Correspondence:

Yanli Tian
tianyanni@njau.edu.cn
Baishi Hu
hbs@njau.edu.cn

¹ College of Plant Protection and Key Laboratory of Integrated Management of Crop Diseases and Pests, Ministry of Education, Nanjing Agricultural University, Nanjing 210095, China

² Key Laboratory of Plant Quarantine Pests Monitoring and Control, Ministry of Agriculture and Rural Affairs, Nanjing 210095, China

³ Institute of Botany, Jiangsu Province and Chinese Academy of Sciences (Nanjing Botanical Garden Mem. Sun Yat-sen), Nanjing 210014, China

⁴ Xinjiang Key Laboratory of Agricultural Biosafety, Ürümqi 830013, China

⁵ College of Agriculture, Xinjiang Agricultural University, Ürümqi 830052, China

⁶ Plant Pathology Department, Endowed Chair in Bacterial Diseases of Tree Fruits, Washington State University, Prosser, WA 99350, USA



© The Author(s) 2025. **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/>.

Fire blight disease is indigenous to North America and has spread to more than 50 countries and regions, including countries of North America, Europe, North Africa, the Middle East, Oceania, and Asia. Currently, it has reached Central/East Asia, the native origin of apple germplasm resources, including Kyrgyzstan, Kazakhstan, and South Korea (Doolotkeldieva and Bobusheva 2016; Park et al. 2016). In 2016, fire blight was first observed on pear and apple trees in Yili, Xinjiang province of China (Wang et al. 2022). Since then, it has spread to most pear-producing regions in Xinjiang province and adjacent Gansu province, causing considerable yield reduction and economic losses to the apple and pear industries (Sun et al. 2023).

As of August 2024, 253 assembled *E. amylovora* genomes at different levels are publicly available in GenBank. The sequenced *Erwinia* genomes from different hosts enable the identification of virulence, host-specificity, and metabolic determinants involved in pathogenicity by comparative genomic analysis (Kamber et al. 2012). Comparative genomic analysis of *Spiraeoideae*- and *Rubus*-infecting strains of *E. amylovora* led to the identification of a pan-genome with the large conservative core genome (Mann et al. 2013). The genomic features of the Korean *E. amylovora* strains were also investigated, and evolutionary pathways were inferred based on comparative genome analysis and phylogenomic reconstruction (Song et al. 2021). The first reported genome of *E. amylovora* isolate in China was published in 2023 (Fei et al. 2023). In this study, the comparative genomes of nineteen isolates from China, Kyrgyzstan, and Kazakhstan were conducted, providing a clue for the origin and the outbreak of fire blight in China.

Results

Overview of the genomic features

The fourteen Chinese *E. amylovora* isolates, twelve from Xinjiang and two from Gansu province, three Kyrgyzstan, and two Kazakhstan isolates were obtained from symptomatic plant tissues, which were confirmed to be *E. amylovora* using the primers p29A/B that target the plasmid pEA29 (Additional file 1: Figure S1). The plasmid patterns of the 19 isolates were similar among the isolates except for XJ15 (Additional file 1: Figure S2). Genome sequences of these isolates were obtained by whole genome sequencing and were deposited in the National Center for Biotechnology Information (NCBI) GenBank database under the accession numbers listed in Table 1.

The basic genome features of the nineteen strains are shown in Additional file 2: Table S1 and Fig. 1. Genome size and gene numbers of the nineteen strains shared a high similarity among the isolates. The circular

chromosomes of the nineteen genomes range from 3.83 to 3.99 Mb with an average GC content of 53.5%. Among them, the genome size of GS01 is the smallest, whereas the size of HZ01 from Kazakhstan is the largest. The average predicted protein-coding numbers present in Chinese isolates were 3866, while those of Kyrgyzstan and Kazakhstan were 3839 and 4101, respectively. The number of tRNA genes ranges from 67 to 90, and the number of sRNA is between 25 and 54.

Phylogenetic analysis of *E. amylovora* isolates

Phylogenetic tree was constructed using the core genome of the nineteen strains and the previously reported *E. amylovora* isolates (Fig. 2). It showed close relationships and high sequence similarities within the isolates. The isolates from China, Kazakhstan, and Kyrgyzstan were clustered to be separate branches, suggesting that although their genome sequences are not identical, these isolates were originated from a single source. In addition, the large branch of the nineteen isolates is closest to the isolates from Britain, Switzerland, and France, and then to those from North America (Fig. 2).

To further confirm the genetic relationship between these isolates, synteny plots were generated using MUMmer version 3.22 (Kurtz et al. 2004). Results showed that the similarity between KG01 and other six genomes (KG02, KG03, XJ10, XJ12, XJ14, and GS01) of these 19 isolates was high (Fig. 3). These results further suggested that the 19 *E. amylovora* isolates are very closely related at genomic level.

Gene function analysis

About 3801–4187 protein-coding sequences were predicted in the 19 isolates. Genome function was annotated by blasting genes against different databases, including COG (Cluster of Orthologous Groups of proteins), KEGG (Kyoto encyclopedia of genes and genomes), and GO (Gene ontology) databases.

Based on GO database annotation, distribution of genes in cellular component, molecular function, and biological process can be retrieved. In the category of cellular component, the genes are mainly distributed in cell part, membrane, macromolecular complex and organelle category. In the category of molecular function, the genes are mainly concentrated in catalytic activity, binding, transporter activity, nucleic acid binding transcription factor activity, and structural molecule activity (Fig. 4a). In the category of biological process, the genes are primarily distributed in cellular process, metabolic process, single-organism process, localization, and regulation of biological process.

Figure 4b showed the statistics of KEGG annotation. Gene functions of *E. amylovora* are enriched in

Table 1 List of sequenced isolates from China and other countries

Strains	Area	Host	Accession numbers	References/source
KG01	Kyrgyzstan	Apple	JBBJYX000000000	This study
KG02	Kyrgyzstan	Hawthorn	JBBJYW000000000	This study
KG03	Kyrgyzstan	Pear	JBBJYV000000000	This study
XJ10	Yili Kazakh Autonomous Prefecture, Xinjiang Province, China	Crabapple	JBBJYU000000000	This study
XJ12	Yili Kazakh Autonomous Prefecture, Xinjiang Province, China	Apple	JBBJYT000000000	This study
GS01	Zhangye, Gansu Province, China	Pear	JBBJYS000000000	This study
XJ14	Altay, Xinjiang Province, China	Hawthorn	JBBJYR000000000	This study
XJ15	KuiTun, Xinjiang Province, China	Hawthorn	JBBJYQ000000000	This study
XJ19	Yining, Xinjiang Province, China	Pear	JBBJYP000000000	This study
XJ23	Tacheng, Xinjiang Province, China	Hawthorn	JBBJYO000000000	This study
GS05	Zhangye, Gansu Province, China	Pear	JBBJYN000000000	This study
XJ26	Changji, Xinjiang Province, China	Apple	JBBJYM000000000	This study
XJ30	Korla, Xinjiang Province, China	Apple	JBBJYL000000000	This study
XJ35	Korla, Xinjiang Province, China	Pear	JBBJYK000000000	This study
XJ51	Hotan, Xinjiang Province, China	Quince	JBBJYJ000000000	This study
XJ60	Aksu, Xinjiang Province, China	Duli	JBBJYI000000000	This study
XJ79	Korla, Xinjiang Province, China	Pear	JBBJYH000000000	This study
HZ01	Kazakhstan	Apple	JBBPFT000000000	This study
HZ02	Kazakhstan	Apple	JBBJYG000000000	This study
CFBP1430	France	Hawthorn (<i>Crataegus</i> sp.)	FN434113, FN434114	Smits et al. (2010)
ATCC49946	USA	Apple (<i>Malus domestica</i>)	FN666575, FN666576, FN666577	Sebahia et al. (2010)
E-2	Belarus	Crabapple (<i>Malus</i> sp.)	CP024970, CP024971	Lagonenko et al. (2008)
CFBP1232	UK	Pear (<i>Pyrus communis</i>)	CAPB000000000, HF560650	Mann et al. (2013)
Ea266	Canada	Crabapple (<i>Malus</i> sp.)	CAOY000000000, HF560644	Mann et al. (2013) and Zhang and Geider (1997)
ATCCBAA-2158	USA	<i>Rubus</i> sp. (Blackberry)	FR719181-FR719212	Powney et al. (2011)
ACW56400	Switzerland	Pear (<i>Pyrus communis</i>)	AFHN000000000, CP002951	Mann et al. (2013)
LA637	Mexico	N/A	CBVU000000000	Smits et al. (2014)
Ea356 (Ea 1/79)	Germany	<i>Cotoneaster</i> sp.	CAOX0000000, HF560643	Mann et al. (2013)
Ea1189	Germany	Pear	CP055227	Yang et al. (2023)
FB-20	South Korea	Pear	CP050240	Song et al. (2021)
TS3238	South Korea	Pear	CP050244	Song et al. (2021)

membrane transport, amino acid, carbohydrate, energy and nucleotide metabolisms, cell motility, signal transduction, and cellular community. By cluster analysis, genome sequences of the nineteen isolates were further grouped into 4016 gene clusters, each of which represents a collection of orthologous genes. As shown in Fig. 4c, 3095 gene clusters exist in the genomes of all nineteen isolates, which forms the core

genome. Roughly, 92% of the genes in the genomes are identical.

Among the core genome sequences, COG annotation showed that the number of genes involved in amino acid transport and metabolism is the largest cluster (average 7.5%), followed by cell wall/membrane/envelope biogenesis (average 5.8%), translation, ribosomal structure and biogenesis (average 5.7%), carbohydrate

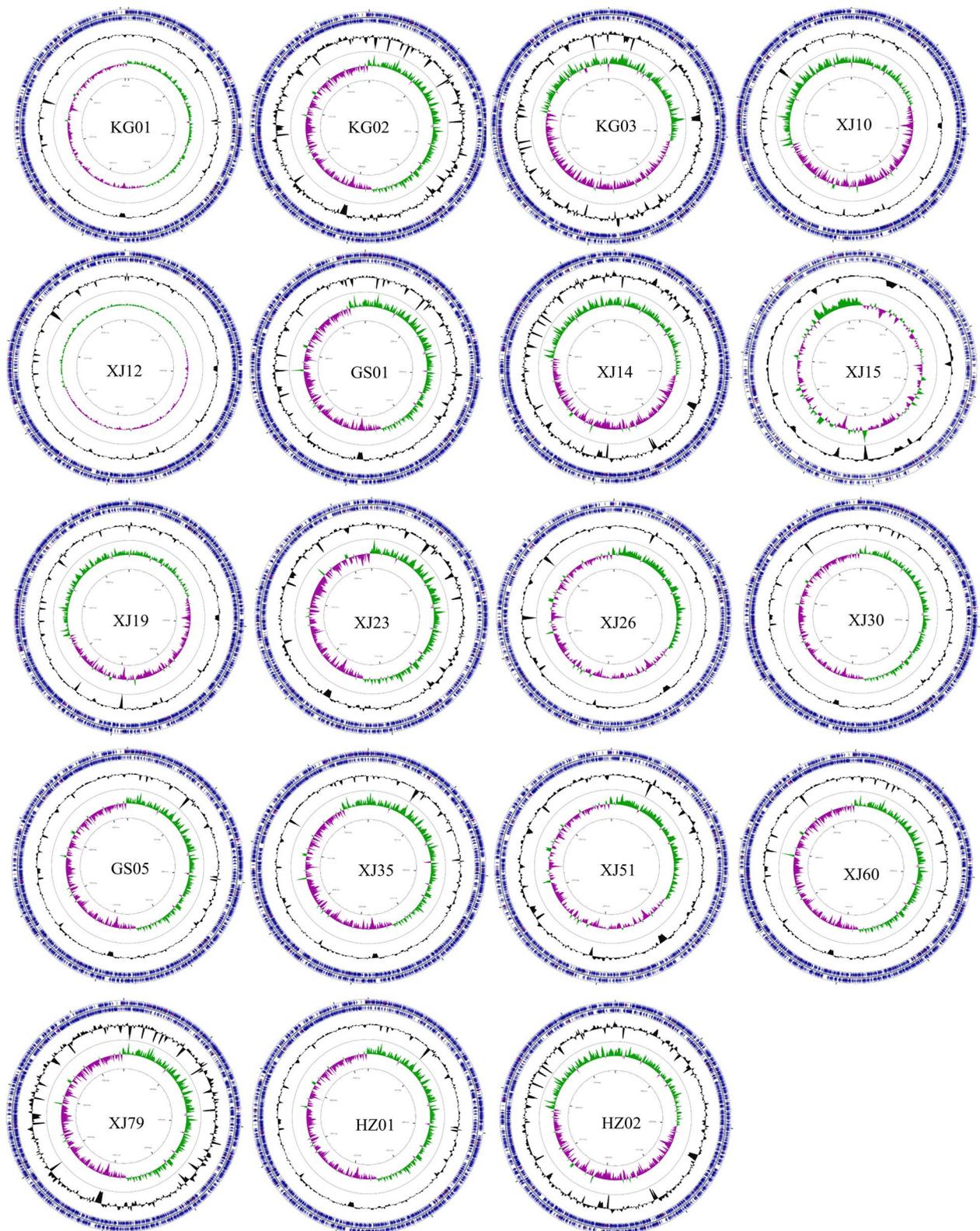


Fig. 1 Circular representation of genomes of 19 *Erwinia amylovora* isolates. From outside to the center, circles represent CDS on forward strand, CDS on reverse strand, G+C content, GC skew and genome size

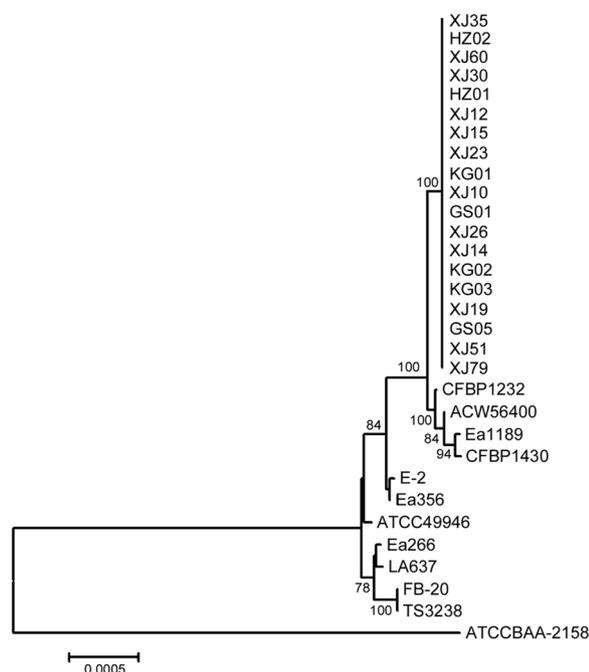


Fig. 2 Phylogenetic analysis of *E. amylovora* strains. The phylogenetic relationships were inferred with a core genome alignment of 31 *E. amylovora* strains

transport and metabolism (average 5.4%), and transcription (average 5.3%) (Fig. 4d).

Secretory systems

Plant pathogenic bacteria usually contain type I to type VI secretory systems, which play important roles in pathogenesis and environmental adaptation. Gram-negative bacteria like *Erwinia*, *Xanthomonas*, *Pectobacterium*, and *Ralstonia* species usually have the type II secretion system (T2SS), an important virulence factor (Smits et al. 2010). T2SS is mainly composed of the outCDEFHJKLMNOS-chiV proteins. T2SS is found in all nineteen isolates, and the sequence of genes are identical, indicating that T2SS is highly conserved.

Type III secretion system (T3SS) is one of the most important pathogenicity factors for *E. amylovora*, enabling them to successfully infect the host plants. Bacteria transport a large number of effectors into plants through T3SS (Khokhani et al. 2013; Vrancken et al. 2013). T3SS of plant pathogenic bacteria are mainly composed of Hrc and Hrp proteins. Among them, *hrpW*, *hrpN*, *hrpS*, *hrpY*, *hrpX*, *hrpJ*, *hrpK*, *hrcC*, *hrcJ*, *hrcN*, and *hrcS* are conserved (Fig. 5a, Additional file 2: Table S2). Meanwhile, all nineteen isolates exhibited virulence in young pear leaves and immature pear fruits (Additional file 1: Figure S3).

Cell wall degrading enzymes

Plant pathogens utilize cell wall degradation enzymes, such as cellulase and pectinase, to break down cellulose and pectin in the epidermis of plants. A large number of plant cell wall degrading enzyme genes are identified in the nineteen genomes (Additional file 2: Table S3 and Table S4) including polygalacturonase (GH28) and pectin lyase (PL1, PL3). Additionally, an Arabinan endo-1,5-Alpha-L-arabinosidase (GH43) is present in fourteen *E. amylovora* isolates, but not in the other five isolates (Additional file 2: Table S5).

Flagellum synthesis and chemotactic gene analysis

E. amylovora is a nonobligate plant pathogenic bacterium, which can be spread through different means to survive under unfavorable conditions (Santander et al. 2014). It has been reported that the motility mechanism has been proven to be closely related to bacterial transmission and host adhesion. It has been observed that flagella-mediated motility and chemotaxis is very important for the colonization on plants. Flagellum synthesis genes in the genomes of nineteen isolates were systematically analyzed. The complete flagellum synthesis gene clusters and 10–11 chemotactic proteins are identified in the genome of nineteen *E. amylovora* isolates.

Signal transduction system

Two-component system (TCS) is the most essential signal transduction system that mediates bacterial response to environmental signals. It was reported that the enterobacterial-specific Rcs phosphorelay system is essential for pathogenicity and amylovoran production in *E. amylovora* (Lee et al. 2018). As an unusual complex TCS, the Rcs system is composed of three core RcsBCD proteins and one auxiliary RcsA protein, and both of them do not carry the phosphorylation site (Ancona et al. 2015). The Rcs system exists in the nineteen *E. amylovora* genomes and is highly conserved.

The second messenger cyclic di-GMP (c-di-GMP)-mediated signaling plays a key role in the establishment and development of plant infections (Edmunds et al. 2013). C-di-GMP is synthesized by diguanylate cyclase (DGC) enzymes, which contains GGDEF domains and is degraded by phosphodiesterase (PDE) enzymes encoding either an EAL or an HD-GYP domain. We found that all three types of proteins can be identified in the genomes of the nineteen isolates (Fig. 5b), including seven GGDEF proteins that are mainly involved in c-di-GMP synthesis. In addition, eight EAL proteins are mainly involved in c-di-GMP degradation, and six HD-GYP proteins are mainly involved in c-di-GMP phosphorylation.

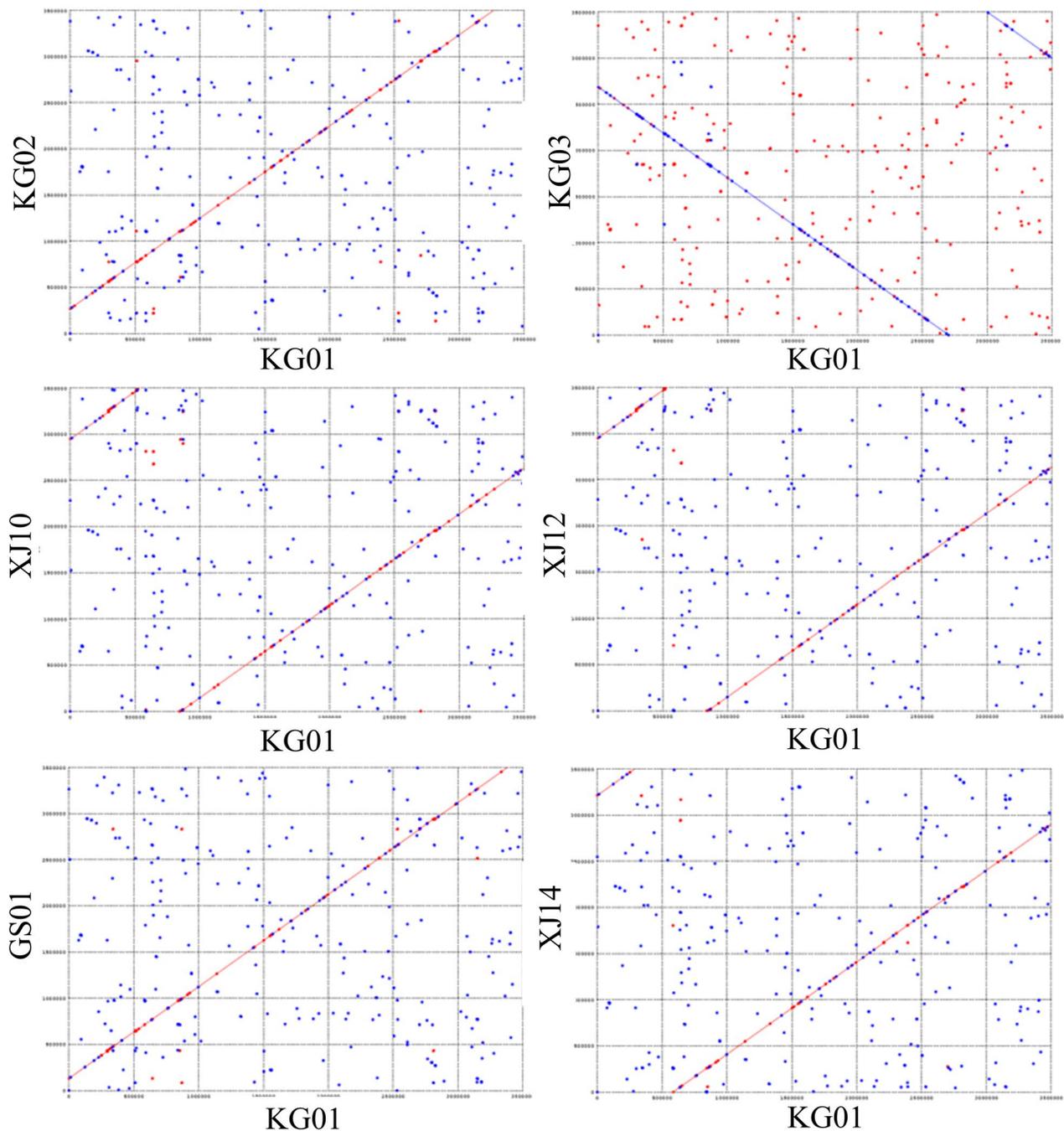


Fig. 3 Comparison of genomic sequences between isolates

Exopolysaccharide biosynthesis

It is well-known that bacterial exopolysaccharides (EPSs) can protect bacteria from hostile conditions (Ordax et al. 2010). Amylovoran and levan are two important exopolysaccharides produced by *E. amylovora*. Amylovoran contributes to pathogenesis and biofilm formation (Koczan et al. 2011). It is a

heteropolymer of glucuronic acid and galactose synthesized by the *ams* operon, which includes twelve genes from *amsA* to *amsL* presenting in a large gene cluster (Geider 2000). These genes are conserved in the nineteen genomes (Fig. 5c). The genes *rscB3*, *rscC*, and *yojN* control amylovoran synthesis, which are also very conserved. Furthermore, the two primary regulative

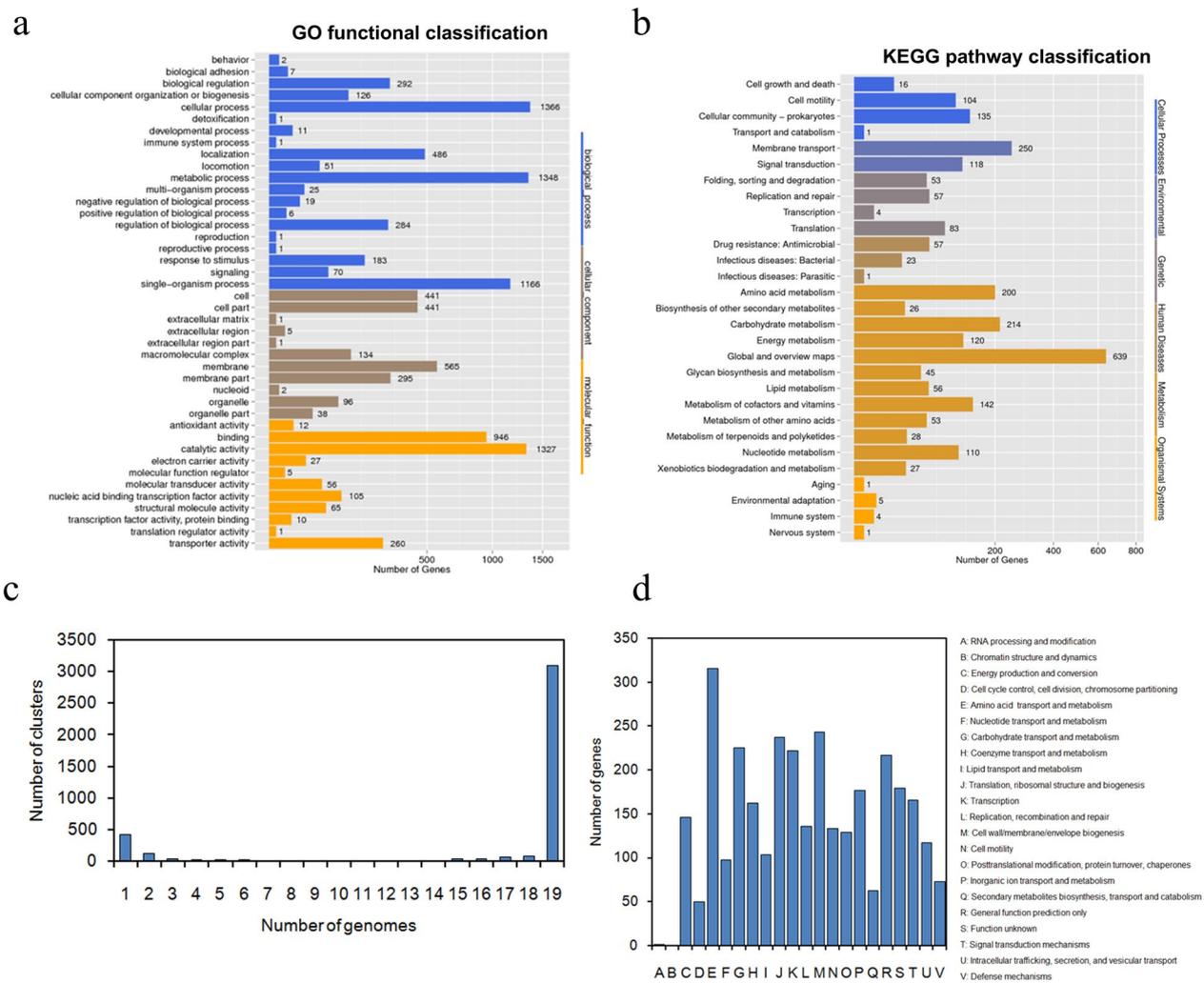


Fig. 4 Genome function annotation of 19 *E. amylovora* strains. **a** Distribution of GO annotation; **b** Distribution of KEGG annotation; **c** Frequency of cluster genes within 19 genomes; **d** COG functional categories

genes in levan production, *rlsA* and *rlsB*, are conserved in all nineteen strains.

Determinant of iron intake

Iron serves as a cofactor for many proteins, and is a fundamental nutritional component for all living organisms. In bacteria, siderophore production and uptake are regulated by iron uptake receptors, which control iron uptake globally (Dellagi et al. 1998). Ferric uptake regulator (Fur) exists in the nineteen isolates and is conserved, including the hydroxamate siderophore desferrioxamine E (DFO E) and the *dfoJAC* genes. (Additional file 2: Table S6).

The reabsorption of siderophore depends on TonB receptors, which help transport siderophore to the periplasmic space. Four TonB-dependent receptor genes are present in the genomes, including *OprC* which codes

the copper receptor protein and *FoxR* which codes the ferroxamine receptor protein. Both of the two receptor protein functions are unknown.

Comparative analysis of secondary metabolites

Plant pathogenic bacteria produce a series of toxins and other secondary metabolites to help infect host plants via the nonribosomal peptide synthetase (NRPS) and polyketide synthetase (PKS) pathways. A total of twelve distinct NRPS/PKS gene clusters were predicted in the nineteen isolate genomes using the antiSMASH software (Fig. 5d). These gene clusters shared high similarities with the previously identified secondary metabolite gene clusters (Additional file 2: Table S7).

Gene clusters 1, 2, and 3 were present in all nineteen strains and shared 75%, 9%, and 13% similarity with clusters desferrioxamine E, basiliskamide A/B, and

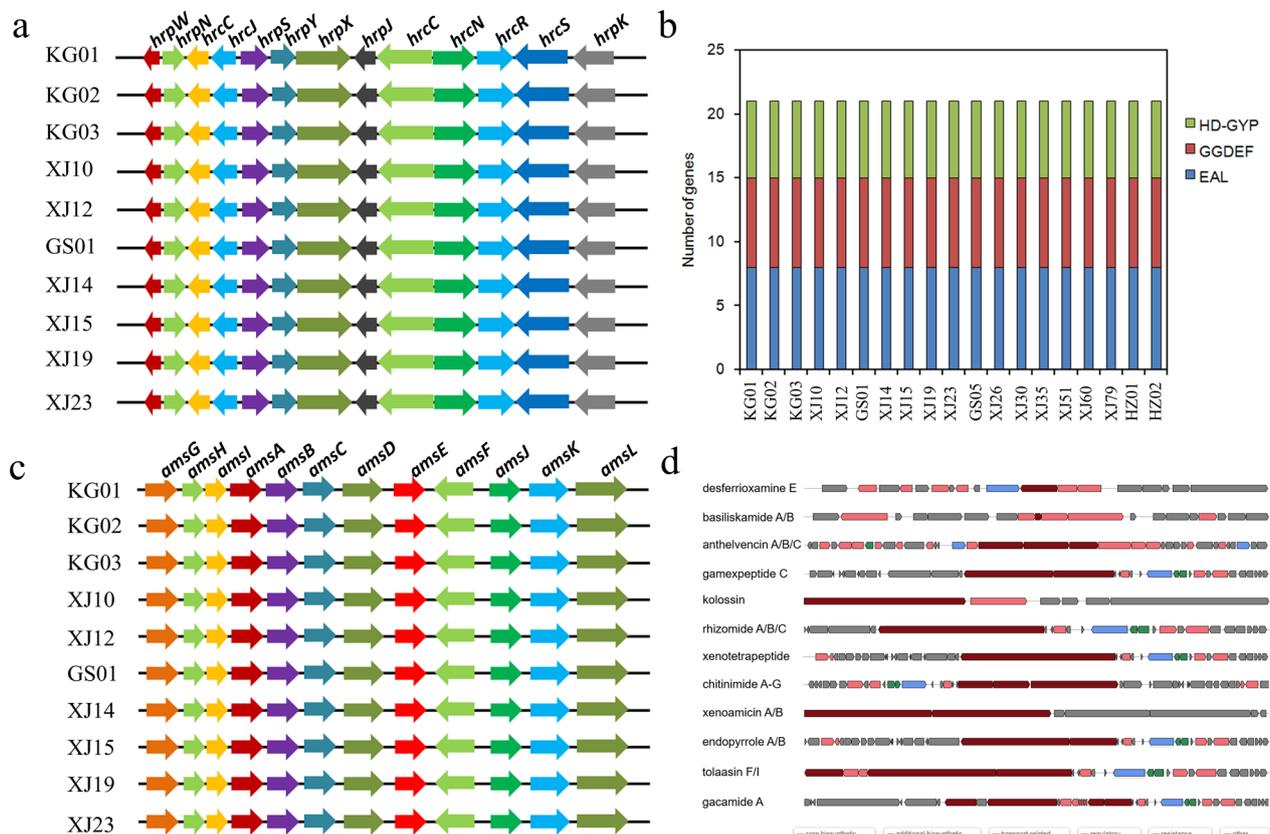


Fig. 5 **a** Physical map of type III secretion system; **b** Distribution of regulators of cyclic di-GMP; **c** Physical map of amylovoran biosynthesis gene cluster; **d** Gene clusters of secondary metabolites

anthelvincin A/B/C, respectively. Gene cluster 4 was present in 11 isolates and had 100% similarity with the gamexpeptide C gene cluster. Gene cluster 5 was identified in eight isolates and shared 100% similarity with kilossin gene cluster. Gene cluster 6 was present in four isolates and showed 100% similarity with the rhizomide A/B/C gene cluster. Gene cluster 7 only existed in XJ14, and the similarity to xenotetrapeptide gene cluster was 100%. Gene cluster 8 was only found in KG03, with a 13% similarity with the chitinimide A-G gene cluster. Gene cluster 9 only existed in strain XJ19, and its similarity to xenoamicin A/B gene cluster was 20%. Gene cluster 10 was only present in strain XJ23, sharing 33% similarity with endopyrrole A/B gene cluster. Gene cluster 11 was only found in HZ01, which shared 50% similar with tolaasin F/I gene cluster. Gene cluster 12 was only found in HZ02, which showed 42% similarity to gacamide A gene cluster (Additional file 2: Table S7).

Discussions

E. amylovora has been included in the “Catalogue 366 of quarantine pests for import plants to the People’s Republic of China” in the year of 2007. In 2020, fire blight was listed in the “National Class I pest list”. China has consistently carried out the strictest quarantine measures for the control of *E. amylovora*, including the examination of imported pear, apple, hawthorn, and other host plant seedlings and scions. Nonetheless, fire blight disease was first observed in China in 2016.

E. amylovora, one of the top ten important plant pathogenic bacteria, is indigenous in North America in 1780s and spread to the Eurasian continent in the later of the twentieth century. At the beginning of the twenty-first century, it has spread to China’s neighboring countries, including Russia (in 2003), Kazakhstan (in 2008), Kyrgyzstan (in 2008), and South Korea (in 2015). The first occurrence of fire blight was observed in Yili Autonomous Prefecture of Xinjiang province in 2016,

where it borders Kazakhstan and Russia. Three years later, the disease spread to most of the regions in Xinjiang province and the adjacent province Gansu. However, the outbreak of the disease in China and how it spreads is still unclear (Sun et al. 2023). To answer these questions, we sequenced 19 *E. amylovora* isolates, including 12 from Xinjiang province, two from Gansu province, three from Kyrgyzstan, and two from Kazakhstan. The comparative genomic analysis was performed for the isolates, including isolates from Korea (Song et al. 2021).

Genome comparison showed that genome sequences of the 19 isolates are highly similar, with about 92% of the genes that are identical. In addition, most of the gene families and metabolic pathways were very conserved. Phylogenetic tree based on comparative genomics showed that isolates from China are more closely related to those from Kyrgyzstan and Kazakhstan than those from Europe and North America. Moreover, when comparing the genomes of the 19 *E. amylovora* isolates with the strains involved in the genomic studies by Zeng et al. (2018), they fit in the “Widely-Prevalent” clade. It is thus speculated that fire blight pathogen, which is originated from North America and then spread to Europe, where it further spread from Kyrgyzstan and Kazakhstan to China, although the speculation needs further research to verify. Furthermore, since *E. amylovora* genomes are highly homogeneous, the genetic diversity among these isolates should be further investigated in the future.

Conclusion

In summary, our results of genome sequencing and annotation would help us to understand and explore the specific mechanisms of the interaction between *E. amylovora* and plants, and to study the whole metabolic pathway and molecular mechanism at genomic level. The genome sequencing and annotation for pear fire pathogen provide clues about the epidemics of the pathogen in China.

Methods

Bacterial isolates and growth conditions

The *E. amylovora* isolates used in this study were isolated from fruits, trunk exudates, and branches from symptomatic trees in Xinjiang and Gansu Province, China, Kyrgyzstan, and Kazakhstan (Table 1). Tissues with typical symptoms were surface sterilized with 75% ethanol and allowed to air dry. Small pieces of tissue from the margins of lesions were ground in sterilized water and then left at room temperature for 5 min. Aliquots of the macerated leaf tissue were streaked onto Luria–Bertani (LB) medium, incubated at 28°C for 24 h,

and then subcultured onto LB medium to obtain single colonies. Pure isolates of the bacteria were tested for induction of a hypersensitive response by infiltrating fully developed leaves of *Nicotiana tabacum* cv. Samsun plants with bacterial suspensions (approximately 1×10^8 colony-forming units (CFU)/mL) of 16-h-old cultures. Bacterial cells were cryopreserved in 25% glycerol at -80°C . All strains were routinely cultured in Luria–Bertani (LB) medium at 28°C.

Whole genome sequencing and assembly

Whole genome sequences of nineteen isolates (Table 1) of *E. amylovora* were determined using Illumina HiSeq 4000 system (BGI, Shenzhen, China). Genomic DNA was sheared to construct two read libraries with length of 300–400 bp and 5–6 kb by a Bioruptor ultrasonicator (Diagenode, Denville, NJ, USA) and physicochemical methods. Paired-end fragment libraries were sequenced according to the Illumina HiSeq 4000 system’s protocol. Raw reads of low quality from paired-end sequencing (those with a certain proportion of low quality (≤ 20) bases, a certain proportion of Ns (10% as default), adapter and duplication contamination) were discarded. Sequenced reads were assembled using SOAPdenovo v1.05 software (<http://soap.genomics.org.cn/soapdenovo.html>).

Genome annotation

Protein-coding gene sequences from the assembled *E. amylovora* genomes were predicted using RAST (Rapid Annotation using Subsystem Technology) software (Aziz et al. 2008). Gene prediction was performed for the *E. amylovora* genome assembled by glimmer3 (<http://www.cbc.umd.edu/software/glimmer/>) with Hidden Markov models. tRNA, rRNA, and sRNA recognition use tRNAscan-SE, RNAmmer, and the Rfam database. The best hit abstracted using Blast alignment tool for function annotation. The databases, KEGG, COG, NR (Non-Redundant Protein Database databases), Swiss-Prot, and GO, TrEMBL, and EggNOG, are used for general function annotation.

Gene family is constructed by the gene of reference, with integration of multi software: align the protein sequence in BLAST, eliminate the redundancy by solar, and carry out gene family clustering treatment for the alignment results with Hcluster_sg software. The antiSMASH program (Medema et al. 2011) was used for secondary metabolite gene cluster identification. The dbCAN software was used to predict the glycoside hydrolase, two-component systems, lipopolysaccharide biosynthesis, flagella, and chemotaxis genes.

Whole genome alignments analysis

The genomes of nineteen *E. amylovora* isolates were compared with *E. amylovora* strains as previously reported (Table 1), including CFBP 1430 from France, ATCC49946 from USA, E-2 from Belarus, CFBP 1232 from England, Ea266 from Canada, ATCCBAA-2158 from USA, ACW56400 from Switzerland, LA637 from Mexico, and Ea356 from Germany. BLAST was used to compare conserved genes of all strains according to pairwise comparison method. Based on genome sequences, MEGA7 was also used to generate the whole genome phylogenetic tree.

The synteny of the nineteen isolates with *E. amylovora* CFBP 1430 strain was performed using MUMmer v3.22. BLAST Core/Pan genes were clustered by the CD-HIT rapid clustering of similar proteins software with a threshold of 50% pairwise identity and 0.7 length difference cutoff in amino acid.

Abbreviations

COG	Cluster of Orthologous Groups of proteins
GO	Gene ontology
KEGG	Kyoto encyclopedia of genes and genomes

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s42483-025-00316-6>.

Additional file 1: Figure S1. Identification of *Erwinia amylovora* isolates by PCR, using specific primer pairs p29A/B. **Figure S2.** Plasmid profiling of *Erwinia amylovora* isolates. **Figure S3.** Virulence assays results.

Additional file 2: Table S1. Basic genomic features of the sequenced isolates. **Table S2.** Analysis of the T3SS components derived from 19 *Erwinia amylovora* isolates. **Table S3.** Analysis of carbohydrate esterases and polysaccharide lyases derived from 19 *Erwinia amylovora* isolates. **Table S4.** Analysis of glycoside hydrolases derived from 19 *Erwinia amylovora* isolates. **Table S5.** Distribution of plant cell wall degrading enzymes. **Table S5.** Analysis of hydroxamate siderophore desferrioxamine E biosynthesis genes derived from 19 *Erwinia amylovora* isolates. **Table S7.** Distribution of gene clusters of secondary metabolites among the 19 genomes.

Acknowledgements

Not applicable.

Author contributions

YT and BH designed the research; YZ, YT, and YZ prepared the materials, JW, SW, XL, and XG analyzed the data and wrote the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by Major Science and Technology Projects of Xinjiang (2023A02006).

Availability of data and materials

The datasets generated and/or analysed during the current study are available in the NCBI repository, <https://dataview.ncbi.nlm.nih.gov/object/PRJNA1087690?reviewer=4du0h9s0pan9dvrj71gk6alluk>.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 27 May 2024 Accepted: 14 January 2025

Published online: 08 April 2025

References

- Ancona V, Chatnaparat T, Zhao Y. Conserved aspartate and lysine residues of RcsB are required for amylovan biosynthesis, virulence, and DNA binding in *Erwinia amylovora*. *Mol Genet Genomics*. 2015;290(4):1265–76. <https://doi.org/10.1007/s00438-015-0988-8>.
- Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics*. 2008;9(1):75. <https://doi.org/10.1186/1471-2164-9-75>.
- Dellagi A, Brisset MN, Paulin JP, Expert D. Dual role of desferrioxamine in *Erwinia amylovora* pathogenicity. *Mol Plant Microbe Interact*. 1998;11(8):734–42. <https://doi.org/10.1094/MPMI.1998.11.8.734>.
- Doolotkeldieva T, Bobusheva S. Fire blight disease caused by *Erwinia amylovora* on rosaceae plants in Kyrgyzstan and biological agents to control this disease. *AIDS Patient Care STDs*. 2016;06(11):831–51. <https://doi.org/10.4236/aim.2016.611080>.
- Eastgate JA. *Erwinia amylovora*: the molecular basis of fireblight disease. *Mol Plant Pathol*. 2000;1(6):325–9. <https://doi.org/10.1046/j.1364-3703.2000.00044.x>.
- Edmunds AC, Castiblanco LF, Sundin GW, Waters CM. Cyclic Di-GMP modulates the disease progression of *Erwinia amylovora*. *J Bacteriol*. 2013;195(10):2155–65. <https://doi.org/10.1128/JB.02068-12>.
- Fei N, Song B, Yang Y, Zhu X, Guan W, Zhao T. Draft genome sequence data for *Erwinia amylovora* strain S618–2–2 isolated from *Pyrus Sinkiangensis* in China. *PhytoFrontiers*. 2023. <https://doi.org/10.1094/PHYTOFR-10-22-0122-A>.
- Geider K. Exopolysaccharides of *Erwinia amylovora*: structure, biosynthesis, regulation, role in pathogenicity of amylovan and levan. In: Vanneste JL, editor. Fire blight: the disease and its causative agent, *Erwinia amylovora*. Wallingford: CAB International; 2000. pp. 117–40. <https://doi.org/10.1079/9780851992945.0117>.
- Kamber T, Smits THM, Rezzonico F, Duffy B. Genomics and current genetic understanding of *Erwinia amylovora* and the fire blight antagonist *Pantoea vagans*. *Trees*. 2012;26(1):227–38. <https://doi.org/10.1007/s00468-011-0619-x>.
- Khokhani D, Zhang C, Li Y, Wang Q, Zeng Q, Yamazaki A, et al. Discovery of plant phenolic compounds that act as type III secretion system inhibitors or inducers of the fire blight pathogen, *Erwinia amylovora*. *Appl Environ Microbiol*. 2013;79(18):5424–36. <https://doi.org/10.1128/AEM.00845-13>.
- Koczan JM, Lenneman BR, McGrath MJ, Sundin GW. Cell surface attachment structures contribute to biofilm formation and Xylem colonization by *Erwinia amylovora*. *Appl Environ Microbiol*. 2011;77(19):7031–9. <https://doi.org/10.1128/AEM.05138-11>.
- Krissoff B, Calvin L. Technical barriers to trade: a case study of phytosanitary barriers and U.S.-Japanese apple trade. *J Agric Resour Econ*. 1998. <https://doi.org/10.22004/ag.econ.31191>.
- Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al. Versatile and open software for comparing large genomes. *Genome Biol*. 2004;5(2):1–9. <https://doi.org/10.1186/gb-2004-5-2-r12>.
- Lagonenko AL, Komardina VS, Nikolaichik YA, Evtushenkov AN. First report of *Erwinia amylovora* fire blight in Belarus. *J Phytopathol*. 2008;156(10):638–40. <https://doi.org/10.1111/j.1439-0434.2008.01420.x>.
- Lee JH, Ancona V, Zhao Y. Lon protease modulates virulence traits in *Erwinia amylovora* by direct monitoring of major regulators and indirectly

- through the Rcs and Gac-Csr regulatory systems. *Mol Plant Pathol*. 2018;19(4):827–40. <https://doi.org/10.1111/mpp.12566>.
- Mann RA, Smits TH, Bühlmann A, Blom J, Goesmann A, Frey JE, et al. Comparative genomics of 12 strains of *Erwinia amylovora* identifies a pan-genome with a large conserved core. *PLoS ONE*. 2013;8(2):e55644. <https://doi.org/10.1371/journal.pone.0055644>.
- Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, et al. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res*. 2011;39(Web Server issue):W339–346. <https://doi.org/10.1093/nar/gkr466>.
- Ordax M, Marco-Noales E, López MM, Biosca EG. Exopolysaccharides favor the survival of *Erwinia amylovora* under copper stress through different strategies. *Res Microbiol*. 2010;161(7):549–55. <https://doi.org/10.1016/j.resmic.2010.05.003>.
- Park DH, Yu J-G, Oh E-J, Han K-S, Yea MC, Lee SJ, et al. First report of fire blight disease on asian pear caused by *Erwinia amylovora* in Korea. *Plant Dis*. 2016;100(9):1946–1946. <https://doi.org/10.1094/PDIS-11-15-1364-PDN>.
- Powney R, Smits THM, Sawbridge T, Frey B, Blom J, Frey JE, et al. Genome sequence of an *Erwinia amylovora* strain with pathogenicity restricted to *Rubus* plants. *J Bacteriol*. 2011;193(3):785–6. <https://doi.org/10.1128/JB.01352-10>.
- Santander RD, Monte-Serrano M, Rodríguez-Herva JJ, López-Solanilla E, Rodríguez-Palenzuela P, Biosca EG. Exploring new roles for the *rpoS* gene in the survival and virulence of the fire blight pathogen *Erwinia amylovora*. *FEMS Microbiol Ecol*. 2014;90(3):895–907. <https://doi.org/10.1111/1574-6941.12444>.
- Sebahia M, Bocsanczy AM, Biehl BS, Quail MA, Perna NT, Glasner JD, et al. Complete genome sequence of the plant pathogen *Erwinia amylovora* Strain ATCC 49946. *J Bacteriol*. 2010;192(7):2020–1. <https://doi.org/10.1128/JB.00022-10>.
- Smits TH, Rezzonico F, Kamber T, Blom J, Goesmann A, Frey JE, et al. Complete genome sequence of the fire blight pathogen *Erwinia amylovora* CFBP 1430 and comparison to other *Erwinia* Spp. *Mol Plant Microbe Interact*. 2010;23(4):384–93. <https://doi.org/10.1094/MPMI-23-4-0384>.
- Smits TH, Guerrero-Prieto VM, Hernández-Escarcega G, Blom J, Goesmann A, Rezzonico F, et al. Whole-genome sequencing of *Erwinia amylovora* strains from Mexico detects single nucleotide polymorphisms in *rpsL* conferring streptomycin resistance and in the *avrRpt2* effector altering host interactions. *Genome Announc*. 2014;2(1):e01229–e1313. <https://doi.org/10.1128/genomeA.01229-13>.
- Song JY, Yun YH, Kim GD, Kim SH, Lee SJ, Kim JF. Genome analysis of *Erwinia amylovora* strains responsible for a fire blight outbreak in Korea. *Plant Dis*. 2021;105(4):1143–52. <https://doi.org/10.1094/PDIS-06-20-1329-RE>.
- Sun W, Gong P, Zhao Y, Ming L, Zeng Q, Liu F. Current situation of fire blight in China. *Phytopathology*. 2023;113(12):2143–51. <https://doi.org/10.1094/PHTO-05-23-0170-RVW>.
- Vrancken K, Holtappels M, Schoofs H, Deckers T, Valcke R. Pathogenicity and infection strategies of the fire blight pathogen *Erwinia amylovora* in Rosaceae: State of the art. *Microbiology*. 2013;159(5):823–32. <https://doi.org/10.1099/mic.0.064881-0>.
- Wang J, Gao J, Bayinkekike MM, Zhang J, Tian Y, Hu B. Blocking field spread of fire blight by electric heating automatic disinfection pruning scissors. *Plant Quarantine*. 2022;36:25–8. <https://doi.org/10.19662/j.cnki.issn1005-2755.2021.00.023>. (in Chinese).
- Yang HW, Thapa R, Johnson K, DuPont ST, Khan A, Zhao Y. Examination of large chromosomal inversions in the genome of *Erwinia amylovora* strains reveals worldwide distribution and North America-specific types. Examination of large chromosomal inversions in the genome of *Erwinia amylovora* strains reveals worldwide distribution and North America-specific types. *Phytopathology*. 2023;113(12):2174–86. <https://doi.org/10.1094/PHTO-01-23-0004-SA>.
- Zeng Q, Cui Z, Wang J, Childs KL, Sundin GW, Cooley DR, et al. Comparative genomics of Spiraeoideae-infecting *Erwinia amylovora* strains provides novel insight to genetic diversity and identifies the genetic basis of a low-virulence strain: comparative genomics of *Erwinia amylovora*. *Mol Plant Pathol*. 2018;19(7):1652–66. <https://doi.org/10.1111/mpp.12647>.
- Zhang Y, Geider K. Differentiation of *Erwinia amylovora* strains by pulsed-field gel electrophoresis. *Appl Environ Microbiol*. 1997;63(11):4421–6. <https://doi.org/10.1128/aem.63.11.4421-4426.1997>.
- Zhao Y, Tian Y, Wang L, Geng G, Zhao W, Hu B, et al. Fire blight disease, a fast-approaching threat to apple and pear production in China. *J Integr Agric*. 2019;18(4):815–20. [https://doi.org/10.1016/S2095-3119\(18\)62033-7](https://doi.org/10.1016/S2095-3119(18)62033-7).