Isolation, Identification, and Whole-Genome Sequence Analysis of Pantoea Agglomerans Causing Leaf Spot-Hole Disease in Wild Apricot in Xinjiang, China

Fangyuan Xu
Xinjiang Agricultural University

Liqiang Liu
Xinjiang Agricultural University

Jun Liu
Xinjiang Agricultural University

Wei He
Beijing Forestry University

Kang Liao (✉ liaokang01@163.com)
Xinjiang Agricultural University

Research Article

Abstract

Wild apricot in Yili wild fruit forest in Xinjiang have been seriously affected by leaf spot-hole disease, with the incidence reaching 100%. To identify the pathogen of apricot perforation in the Yili wild fruit forest, two bacterial strains with strong virulence were obtained by the dilution separation method. The bacterial strains were gram-negative bacteria with yellow colonies, smooth surfaces and neat edges. The results of the pathogenicity test showed that the bacteria could cause symptoms of leaf spot-hole disease in wild apricot, similar to the symptoms in the field, and could cause HR in tobacco. Based on the 16S rDNA gene sequence and multilocus sequence analysis of fusA, gyrB, leuS, pyrG, rpoB and rlpB, combined with the physiological and biochemical characteristics, the isolated strain was identified as *Pantoea agglomerans.* The pathogen causing bacterial leaf spot-hole disease in wild apricot was determined to be *P. agglomerans* in the wild fruit forest of Yili, Xinjiang. The whole genome of the pathogen strain GL9-2 was sequenced based on the Illumina HiSeq500 and PacBio RS platforms. The genome size was 4765392 bp, and the G+C value was 55.27%. There was one chromosome and two plasmids in the genome, and 4353 CDs were identified. The annotation results showed that 52 glycoside hydrolase-related genes, 38 bacterial secretory system-related genes and 600 toxin-related genes were predicted.

Introduction

Armeniaca vulgaris Lam.[1] is a precious wild fruit tree resource that is considered to be the original population of cultivated apricot and plays a decisive role in the domestication of cultivated apricot worldwide[2]. The Xinjiang wild fruit forest is an integral part of the Tianshan wild fruit forest on the border of Central Asia and is the largest distribution area of wild apricot in China[3-5]. In 1984, China listed the Yili wild apricot forest on an endangered plant protection list. Xinjiang has a wide variety of apricot genetic resources with rich diversity[6], among which wild apricot is an important genetic resource of characteristic fruit trees in Xinjiang[7]. This tree species likes light, is tolerant to drought and cold and can adapt to harsh climate environments and poor soil conditions[8]. Wild apricot generally grows on the hillside below 1500 m above sea level, with sufficient sunshine, abundant heat and much precipitation. In winter, wild apricot is in the inversion layer. The unique climatic conditions and geographical environment in the Yili area provide optimal conditions for the growth and development of wild apricot, making it an important area for deciduous broad-leaved tree species[9].

 However, due to the increase in human activities in wild fruit forests, the stability of natural ecosystems in these areas has been destroyed, which has caused the continuous outbreak of pests in wild fruit forests and resulted in tree death and a sharp decline in the distribution area. The most serious and common disease is leaf spot-hole disease, but there is little research on the disease of *A. vulgaris* Lam. in the Yili wild fruit forest. To date, research on this pathogen has not been reported in detail. Since 2018, the symptoms of leaf spot-hole disease in wild apricot in wild fruit forests have increased annually. According to the author's investigation, the rate of diseased plants has reached 100%. Leaf spot-hole

disease of apricot endangers not only leaves but also fruits and young shoots. In serious cases, this disease causes the abscission of leaves and fruits of wild apricot until the death of the plants.

Bacterial leaf spot-hole disease is also known as bacterial black spot. The host plants of this disease are mainly plum plants such as peach, mountain peach, plum, apricot and cherry, as well as ornamental plants such as cinnamon. Peach, nectarine, plum, and apricot were shown to be the most susceptible tree species to this disease[10]. At the beginning of the 20th century, bacterial leaf spot-hole disease was found and reported in plum trees in the United States for the first time. The pathogen was determined to be *Xanthomonas arborcola* pv*. Pruni*[11] and was shown to latently infect plant materials used for agricultural planting. With the continuous rise in international trade, the pathogen gained a global reach, but the disease is not evenly distributed worldwide. This disease mainly occurs in warm and humid areas, especially in Europe[12,13]. In 2018, M.T. Hulin et al. found that *Pseudomonas syringae* can cause perforation of leaves and fruits of cherry and plum[14], which mainly occurs in sandy loam with cold winters, poor water conditions and light soil[15]. In 2020, Tong Yaping and others found a new pathogen of bacterial perforation disease, *Pantoea agglomerans*, on peach trees in orchards in Fujian Province[16]. In 2020, Li Chaohui and others also isolated *P. agglomerans* on plum leaves in Dazhou, Sichuan Province, which can lead to perforation symptoms of plum leaves[17]. There are many kinds of pathogens that cause perforation symptoms in stone fruit trees, but there has been no clear conclusion on the types of pathogens causing wild apricot leaf spot-hole disease in the Xinjiang wild fruit forest.

In this study, the pathogens of bacterial perforation disease samples in the Yili area were isolated, and the morphology and pathogenicity of the isolated bacteria were observed. At the same time, the bacteria were identified in combination with the physical and chemical property determination, 16S rDNA gene sequence and multilocus sequence analysis (MLSA) of pathogenic bacteria to clarify the pathogenicity and taxonomic status of pathogenic bacteria. The genome sequence of the pathogen was determined, the genome characteristics were analyzed, and the disease-related genes or related signaling pathways of the pathogen were found by bioinformatics methods to lay a foundation for the further study of the pathogenesis of *P. agglomerans*.

Materials And Methods

Sample collection

The leaves and fruits of wild apricot with typical spot-hole disease symptoms were collected from Yili Botanical Garden in Gongliu County and Xinyuan County. The botanical Garden is the implementation site of the project, so the experiments carried out in the botanical garden are allowed. The fresh diseased tissues were packed in clean self-sealing bags and taken back to the laboratory Xinjiang Agriculture University Forestry Protection Lab, deposition numbers was in Table 1for storage at 4°C. The experiments conducted, including the collection of plant material, in compliance with relevant institutional, national, and international guidelines and legislation. The identification of wild apricot

samples in this experiment was completed by Professor Kang Liao, the chief expert of Xinjiang forestry and fruit industry.

Table 1 Collection site and sample information summary

Numbers	Longitude	Latitude	Altitude 〔m〕	Location	Region	Deposition NO.
1	83°36'25"	43°22'40"	1356	Leaf and Fruit	XinYuan	XJAU2019090401
2	83°36'27"	43°22'34"	1450	Leaf and Fruit	XinYuan	XJAU2019090402
3	83°36'26"	43°22'40"	1378	Leaf and Fruit	XinYuan	XJAU2019090403
4	83°36'29"	43°22'37"	1365	Leaf and Fruit	XinYuan	XJAU2019090404
5	83°36'29"	43°22'37"	1370	Leaf and Fruit	XinYuan	XJAU2019090405
6	82°11'03"	43°36'14"	1270	Leaf and Fruit	GongLiu	XJAU2019090411
7	82°11'03"	43°36'12"	1300	Leaf and Fruit	GongLiu	XJAU2019090412
8	82°10'99"	43°36'13"	1288	Leaf and Fruit	GongLiu	XJAU2019090413
9	82°10'99"	43°36'13"	1289	Leaf and Fruit	GongLiu	XJAU2019090414
10	82°10'99"	43°36'12"	1285	Leaf and Fruit	GongLiu	XJAU2019090415

Isolation and culture of pathogenic bacteria

When the diseased samples were back at the laboratory, the tissue at the junction of diseased and healthy regions was cut into 5 mm×5 mm pieces and washed with tap water. The surface was disinfected in 3% sodium hypochlorite solution for 3-5 min, rinsed with sterile water 3 times, and then dried with sterile absorbent paper. The diseased tissue was ground in a sterilized mortar and left to stand for approximately 10 min. A total of 100 μL of supernatant was coated on an NA plate, and the plate was inverted and cultured in a 28°C constant temperature incubator for 24 h. A single colony was isolated, purified, cultured, and stored at -20°C[18].

Pathogenicity determination of pathogenic bacteria

Preparation of inoculum

The isolated strains were cultured on NA plates at 28°C for 24 h. A single colony was selected and inoculated in liquid LB medium at 200 R/min and 28°C for 24 h and cultured until OD 0.5. The culture was then used for subsequent in vitro leaf inoculation.

HR of tobacco

The mature leaves of healthy Bunsen tobacco cultivated in the greenhouse for approximately 1 month were inoculated by injecting a bacterial suspension onto the back of the leaf with a sterile syringe. The

left side of each tobacco leaf was inoculated with sterile water as the control, and the right side was inoculated with the bacterial suspension. Three points were inoculated on each side, and 50 μL of suspension was injected at each point. After inoculation, the plants were watered and cultured at 28°C, and spots on the leaves were observed after 24 h.

Inoculation of wild apricot leaves

The pathogenicity was determined by injection inoculation. Fresh and healthy wild apricot leaves with consistent growth were collected, their surfaces were cleaned with sterile water and wiped with 75% alcohol cotton balls, and microwounds were made on the leaf surface with a disinfected inoculation needle. A sterile cotton ball was dipped in bacterial solution and inoculated on the wounds for 24 h. Sterile water was used as the control and, and plants were kept moist and at 28°C. The disease incidence was observed and recorded 24 h later.

Physiological and biochemical determination of pathogenic bacteria

The physiological and biochemical properties, Gram staining, indole production, methyl red test, gelatin liquefaction, starch hydrolysis, nitrate reduction test, oxidase and catalase activities, citrate utilization, malonate utilization and carbohydrate acid production were determined according to the eighth edition of Berger's bacterial identification manual.

Molecular biological identification of pathogenic bacteria

DNA extraction

The tested strains were cultured on NA medium plates at 28°C for 24 h. Single colonies were isolated, inoculated in liquid LB medium, and shaken at 200 rpm/min for 24 h. The genomic DNA of the tested strain was extracted by using the Tiangen bacterial genomic DNA extraction kit.

16S rDNA sequencing

16S rDNA sequence fragments of pathogenic bacteria were amplified by the 16S rDNA gene universal primer 27F/1492R. The primer sequences are shown in Table 1 and were synthesized by Beijing Liuhe Huada Biotechnology Co., Ltd. The PCR system (25 μL) contained the following: 12.5 μL of 2×Easy TaqTMpcr Supermix, 1 μL of 10 μMol/L primer 27F/1492R, and 8.5 μL to 25 μL of sterilized deionized water. A total of 5 μL of the amplified products was detected by electrophoresis on 1.5% agarose gel. The detected bands were purified and recovered, connected to the PMD19-T vector and transformed into competent DH5α cells. Positive clones were selected and sent to Beijing Liuhe Huada Biotechnology Co., Ltd. for sequencing. The sequencing results were BLAST analyzed in the NCBI database, and the neighbor-joining (NJ) method was used to construct a phylogenetic tree in MEGA X software.

MLSA

Six housekeeping genes, namely, fusA, gyrB, leuS, pyrG, rlpB and rpoB, were amplified by PCR. The primer sequences are shown in Table 2. The PCR system (25 μL) contained the following: 12.5 μL of 2×Easy TaqTMpcr Supermix, 10 1 μL of μMol/L primer, and 8.5 μL to 25 μL of sterilize deionized water. The amplified products were subjected to electrophoresis on 1.5% agarose gel and sent to Shanghai Meiji Biotechnology Co., Ltd. for sequencing. The sequencing results were BLAST analyzed in the NCBI database, and the NJ method was used to construct a phylogenetic tree in MEGA X software.

Table 2 Primers used for amplification

NO.	Gene	Primer	Sequence	PCR amplified procedure
1	16s rDNA	27F	5'-AGAGTTTGATCCTGGCTCAG-3'	94℃ 4min，94℃ 30s，
		1492R	5'-GGTTACCTTGTTACGACTT-3'	55℃ 30s，72℃ 1min，35 cycles，
				72℃ 10 min，
2	fusA	fusA 3	5'- CAT CGG TAT CAG TGC KCA CAT CGA-3	94℃2 min，
		fusA 4	5'- CAG CAT CGC CTG AAC RCC TTT GTT-3'	94℃1 min，
				60℃1min，
3	leuS	leuS3	5'-CAG ACC GTG CTG GCC AAC GAR CARGT-3'	72℃1min，10 cycles，
		leuS4	5'-CGG CGC GCC CCA RTA RCG CT-3'	94℃1 min，
4	pyrG	pyrG3	5'-GGG GTC GTA TCC TCT CTG GGT AAAGG-3'	50℃1 min，
		pyrG4	5'-GGA ACG GCA GGG ATT CGA TAT CNCCKA-3'	72℃1min，21 cycles，
5	gyrB	gyrB3	5'-GCG TAA GCG CCC GGG TAT GTA-3	72℃5 min，
		gyrB4	5'-CCG TCG ACG TCC GCA TCG GTC AT-3'	
6	rplB	rplB3	5'-CAG TTG AAC GTC TTG AGT ACGATC C-3	
		rplB4	5'-CAC CAC CAC CAT GYG GGT GRT C-3	
7	rpoB	Vic3	5´-GGCGAA ATGGCWGAGAACCA-3´	
		Vic2	5´-GAGTCTTCGAAGTTGTAACC-3´	

Whole-genome sequencing and sequence analysis of pathogenic bacteria

The whole genome of pathogenic strain GL9-2 was sequenced by the Illumina HiSeq500 and PacBio RS platforms and assembled by Soap Denovo, and the gene function was predicted by Glimmer. The common function database and specific function database were used for annotation. Public function

databases included the Gene Ontology (GO) database, Kyoto Encyclopedia of Genes and Genomes (KEGG) database and Clusters of Orthologous Groups of proteins (COGs) database. Specific function databases included the Transporter classification database (TCDB), Cazymes (CAZY) database and pathogen-host interactions (PHI) database. Genome sequencing was completed by Shanghai Meiji Bioinformatics Co., Ltd.

Results

Disease symptoms

An investigation in Gongliu County and Xinyuan County in the Yili region from 2017 to 2020 showed that bacterial perforation of wild apricot was serious, and the rate of diseased plants in the forest was 100%.

The symptoms of leaf spot-hole disease included damage to leaves, small spots in the shape of water stains in the initial stage, dark purple spots in the shape of water stains in the later stage, brown spots, light yellow-green halos at the junction of diseased and healthy tissues, and finally leaf perforation and loss. At the initial stage of damage, the surface of the fruit exhibited water stain-like, round brown spots, which began to yellow at the initial stage and expanded into brown spots at the later stage, with the edge of the spots curling up. There were obvious canker spots on the branches, and the bark at the disease spots was cracked (Fig. 1).

Morphological characteristics of isolates

A total of 97 leaf samples and 13 fruit samples were collected from Xinyuan County and Gongliu County in Yili. A total of 128 bacterial strains were obtained after isolation and purification. These strains can be divided into three types in terms of colony morphology. Class I colonies are white, round and opaque; Class II colonies are white, round, smooth and transparent; and Class III colonies are yellow, round, smooth and transparent. After the pathogenicity test, only the yellow strain was determined to be pathogenic. Representative strains GL9-2 and GL9-3 were selected for morphological observation. The colonies cultured on the NA plate for 24 h were yellow and round, with neat edges that were smooth and bright (Fig. 2-b). The colonies turned red after Gram staining and were determined to be a gram-negative bacterium (Fig. 2-a).

Pathogenicity of bacteria

The two representative strains were isolated, purified and preserved and then inoculated into leaves to observe their pathogenicity. The results of in vitro inoculation of wild apricot leaves showed that the two representative bacterial strains could cause disease in wild apricot leaves. One day after inoculation, water stains appeared on the leaf surface of wild apricot, and light-yellow spots the size of pinholes appeared at the inoculation point, which expanded in a circle. Seven days after inoculation, the lesion gradually expanded, and necrosis occurred in the center of the diseased site; 14 d after inoculation, the lesion continued to expand, the central necrotic part of the lesion fell off, and perforation began to occur

(Fig. 3). The disease spots of inoculated, diseased leaves were separated again, and colonies with similar morphology to the tested strains were obtained, and there were no symptoms in the sterile water control. It was further proven that the tested strain could cause necrotic spots and perforation of wild apricot leaves. The representative strains could cause tobacco HR(Fig. 4).

Physiological and biochemical characteristics of pathogenic bacteria

Physiological and biochemical test results of strains GL9-2 and GL9-3 are shown in Table 3. Strains GL9-2 and GL9-3 were determined to be gram-negative bacteria that can liquefy gelatin and hydrolyze starch. The indole test was negative and the VP test was positive. These bacteria can use malonate and citrate to reduce nitrate, which is consistent with the characteristics of *Pantoea* in Berger's Bacterial Identification Manual[19].

Table 3 Physiological and biochemical features of isolates GL9-2 and GL9-3

Test	Stain GL9-2	Stain GL9-3
Gram	-	-
Salt-tolerance	+	+
VP Test	+	+
Malonate utilization	+	+
Cysteine produce H_2S	+	+
Xylose	+	+
Inositol	+	+
Mannitol	+	+
Maltose	+	+
Gelatin liquefaction	+	+
Amylolysis	+	+
Lactose	+	+
Nitrate reduction	+	+
Indole test	-	-
Aerobism	±	±
PAL	+	+
Methyl red test	+	+
Citrate	+	+

Note: Symbol of "+"represents positive response, "-"represents negative response.

Molecular biological identification of pathogenic bacteria

16S rDNA sequence analysis

The bacterial 16S rDNA primers 27F and 1492R were used to amplify the total DNA of the pathogen, and a 1350 bp band was obtained. After sequencing, the data were uploaded to NCBI for BLAST comparison. The similarity between the bacterial 16S rDNA and *P. agglomerans* was 99.4%. The phylogenetic tree constructed based on the 16S rDNA sequence (Fig. 5-a) showed that the isolated strains GL9-2 and GL9-3 had the closest genetic relationship with the *P. agglomerans* SSH strain (accession No. kt075196.1).

MLSA

The housekeeping genes fusA, gyrB, leus, pyrG, rpoB and rlpB were used to construct multiple genes phylogenetic tree[20](Fig. 5-b). The multisite sequences phylogenetic tree of GL9-2 and GL9-3 and 15 *Pantoea* strains were constructed by MEGA-X software. The phylogenetic results based on multigene sequences showed that the tested strains clustered with the published *P. agglomerans* strain to form a group and had the highest homology with the *P. agglomerans* th81 strain.

Whole-genome sequence analysis

Genome-wide basic information

The whole genome of strain GL9-2 was sequenced using the Illumina NovaSeq pe150 sequencing platform, and 1,644,858,704 bp original data were obtained, of which the filtered effective data were 1,629,960,143 bp, accounting for 99.1% of the original data (Table 4). The Q20 of the effective data reached 98.2%, the Q30 reached 94.6%, the predicted genome size was 4,765,392 bp, the G+C content was 55.27%, and the genome was composed of one chromosome and two plasmids. The prediction results of coding genes showed that 4353 CDs were predicted, with a total length of 4,065,777 bp and an average length of 934.02 bp. The CDs accounted for 85.32% of the whole-genome sequence (Table 5).

Table 4 The results of assembly quality control.

Sample Name	GL9_2
Raw Bases (bp)	1644858704
Clean Bases (bp)	1629960143
Clean Q20(%)	98.2
Clean Q30(%)	94.58
Total Reads No.	191099
Total Bases (bp)	1516908664
Largest (bp)	196720
Average Len (bp)	7937.82

Table 5 Genome annotation results

Sample Name	GL9-2
Chromosome No.	1
Plasmid No.	2
Genome Size (bp)	4765392
G+C(%)	55.27
Depth	318.32
COD gene（bp）	4353
tRNA	76
rRNA	22

Genome assembly

The bacterial genome scanning map uses the short sequence assembly software soap denovo2 to splice multiple KMER parameters of the optimized sequence after second-generation sequencing to obtain the optimal contigs assembly results, compare the reads to the contigs, and locally assemble and optimize the assembly results according to the paired end and overlap relationship of reads to form scaffolds.

The bacterial genome completion map assembles three generations of sequences by using assembly software CANU and SPAdes(Fig. 6). If there was an overlap region of more than a certain length at both ends of the final assembly sequence, the sequence was looped, and the overlap sequence at one end was truncated. Finally, complete chromosome and plasmid sequences were obtained.

Gene function annotation

GO function annotation: The protein sequence of the predicted gene was compared with the GO database by BLASTP to obtain the GO annotation information, and blast2GO software was used for GO function cluster analysis. GO annotation includes three aspects: biological processes, cellular components and molecular functions (Fig. 7). According to the classification of cell components, the coding genes were enriched into 14 items, and the number of genes enriched by integral component of membrane and cytotasm was highest. In the biological process category, the items with the largest number of enriched genes were related to regulation of transcription and transmembrane transport. The two items with the largest number of enriched genes in the molecular function category was binding and catalytic activity.

COGs annotation: Coding genes can be divided into 23 categories according to their functions in the COG database. The four entries with the largest number of enriched genes were amino acid transport and metabolism, transcription, carbohydrate transport and metabolism, and inorganic ion transport and metabolism (Fig. 8).

KEGG annotation: The CDs identified to be *P. agglomerans* were enriched in multiple KEGG entries. In the category of cellular processes, the two items with the largest number of enriched genes were cellular community prokaryotes and cell mobility. In the environmental information processing category, membrane transport and signal transmission pathways were enriched. In the category of genetic information processing, the coding genes were enriched in four pathways, and that with the largest number of enriched genes was translation. In the category of human diseases, the coding genes were enriched in 11 pathways, and the two pathways with the largest number of genes were drug resistance (antimicrobial) and infectious diseases (bacterial). CDs were enriched in the category of metabolism, with 12 entries. The largest number of enriched genes was in global and overview map pathways, followed by carbohydrate metabolism and amino acid metabolism pathways. In the classification of organic systems, the coding genes were enriched into 7 entries, including the endocrine system, aging, environmental adaptation and the immune system (Fig. 9).

PHI annotation: The pathogens in the PHI database are mainly fungi, oomycetes and bacteria. The infected hosts include animals, plants, fungi and insects. The number of matched genes of *P. agglomerans* coding genes in the PHI database was counted. The results (Table 6) showed that a total of 1131 genes were identified in the PHI database, of which 123 genes were distributed in the increased virulence (hypervirulence) classification, 778 genes in the reduced virulence classification and 18 genes in the lethal classification.

Table 6 Distribution of phenotypic categories of GL9-2 gene orthologs using the PHI database

Phenotype	Gene No.
reduced virulence	778
unaffected pathogenicity	351
increased virulence (hypervirulence)	123
loss of pathogenicity	89
effector (plant avirulence determinant)	20
lethal	18
chemistry target: sensitivity to chemical	2
chemistry target: resistance to chemical	2

Metabolic system analysis: Carbohydrates play an important role in many biological functions. Much meaningful biological information can be obtained by studying carbohydrate-related enzymes. Carbohydrate-active enzymes (cazymes) are a large class of very important enzymes that are divided into glycoside hydrolases (GHs), glycosyltransferases (GTs), polysaccharide lyases (PLs) and glycoesterases. These enzymes have the functions of degrading, modifying and generating glycosidic bonds. In-depth analysis of the types and quantities of cazymes of pathogenic bacteria is of great significance to reveal the pathogenic mechanism of pathogenic bacteria .

CAZY is a professional database on enzymes that synthesize or decompose complex carbohydrates and glycoconjugates. According to the similarity of amino acid sequences in protein domains, cazymes from different species can be divided into GHs, GTs, PLs, carbohydrate esterases (CEs), carbohydrate binding modules (CBMs), and auxiliary oxidoreductases (AAs)(Table 7).

Table 7 annotation of CAZyme

Class Definition	Genes No.	Precent%
Glycoside Hydrolases	52	39.10%
Glycosyl Transferases	46	34.59%
Carbohydrate Esterases	20	15.04%
Auxiliary Activities	13	9.77%
Carbohydrate-Binding Modules	2	1.5%

Secretory system analysis: At present, there are 7 kinds of secretory systems, namely, types I ~ VII. Pathogenic bacteria control the immune response and cell death by secreting TNSS effector proteins into extracellular or host cells. Among them, the type III secretory system (T3SS) is usually used to study the infection mechanism and virulence of gram-negative bacteria at the molecular level. The coding genes of

P. agglomerans GL9-2 was annotated into the T1SS, T2SS and T6SS, the Tat secretory system and SEC SRP secretory system, of which the T6SS had the largest number of genes (Table 8).

Table 8   Secretion system type No.of Stain GL9-2

Type	Gene No.
Type I Secretion System	2
Type II Secretion System	1
Type VI Secretion System	19
Tat Secretion System	4
Sec-SRP Secretion System	12

Discussion

In this study, the bacterial isolate was isolated from wild apricot disease samples in Yili, Xinjiang. The pathogenicity test showed that the bacterial isolate could infect wild apricot leaves, cause leaf spot-hole disease symptoms consistent with field symptoms, and cause local wilting symptoms of tobacco. The observation of morphological characteristics, determination of physiological and biochemical characteristics and comparative analysis of 16S rDNA, gyrB, pyrG, Fusa, leus, rlpb and rpoB gene sequences showed that the bacterial isolate was highly consistent with the characteristics of the clustered *Pantoea* genus. Therefore, the pathogen causing the perforation of wild apricot in Yili, Xinjiang, China, was identified as *P. agglomerans*.

There are three common types of perforation: bacterial perforation (*Xanthomonas arboricola* PV. *Pruni*), mold perforation (*Wilsonmyces carpophilus*) and brown spot perforation (*Pseudocercospora circumcissa*). There were light green halos on the perforation edge caused by bacterial perforation, the edge was irregular, ulcer spots appeared on the branches, and the disease spots expanded and produced cracks of different depths. The perforation edges of mold spot perforation and brown spot perforation were neat, and there were no obvious symptoms on the branches. The symptoms of the three perforation diseases were similar, but their pathogens were completely different. Whether these diseases will occur simultaneously on the same host in the same environment has not been studied.

It is generally believed that the pathogen responsible for bacterial perforation of peaches and plums is *X. arboricola* PV. *Pruni*. In recent years, it has been gradually found that other bacteria can also cause perforation symptoms in stone fruit trees. For example, *Pseudomonas* can also cause perforation of leaves and fruits of stone fruit trees[14, 16, 17], as can *P. agglomerans* obtained in this study. This bacterium can also cause perforation symptoms of peach, plum and apricot and is considered to be a new pathogen that can cause bacterial perforation.

The pathogen *P. agglomerans* obtained in this study can cause allergic reactions to tobacco. In the study by Tong Yaping[16], the pathogen of bacterial perforation disease (*P. agglomerans*) isolated from peach and plum leaves in Fujian did not produce disease symptoms in tobacco. Whether pathogenicity of the isolated strain is stronger than that of the Fujian strain found in the large-scale outbreak of bacterial perforation in Tianshan wild fruit forests remains to be studied[18,19].

P. agglomerans has always been considered an endophytic bacterium, but recent studies have shown that agglomerated panbacteria are also pathogenic to many plants. This species can cause onion rot[21], cotton bacterial boll rot[22], corn bacterial stem rot[23], banana leaf sheath rot[24], sweet cherry bacterial fruit spot[25], corn and sorghum leaf blight and vascular wilt[26], walnut bacterial black spot[27], Pennisetum leaf blight[28], and cabbage soft rot[29]. The whole-genome sequence of *P. agglomerans* GL9-2 was determined by the combination of second-generation sequencing technology and third-generation sequencing technology, and the whole-genome sequence of *P. agglomerans* GL9-2 was obtained and analyzed by bioinformatics. After genome sequencing and comparison with the CAZY database, it was found that the number of GHs was the largest. GHs catalyze the cleavage of glycosidic bonds between sugar molecules and biomolecules and can be divided into exo- and endo-GHs. When plant pathogens invade the host, they will produce a large number of enzymes that degrade the plant cell wall. Most of these enzymes are GHs, PLs or glycoesterases, some of which are cell-wall degrading enzymes. For example, the lyase pectinase and the hydrolase xylanase are directly related to the pathogenicity of pathogenic bacteria.

Comparing the genomic data of the strain with the virulence factor database (VFDB), the results showed that most genes in the aggressive virulence factors were T6SS and flagella genes. The T6SS widely exists in *Proteus* and is composed of structural proteins, translocation proteins and helper proteins. The T6SS can mainly enhance the adaptability of pathogenic bacteria and make them pathogenic[30-32]. The flagellum is the mobile organ of bacteria; allows for bacteria to move rapidly in liquid environments or semisolid environments; and plays an important role in the growth, reproduction and infection of bacteria[33]. The flagellum has two main functions; one function is chemical tropism[34-35], which allows for bacteria to move to nutrient-rich places or help them escape harmful substances, and the other function is related to the toxicity of bacteria[36].

In recent years, *Pantoea vagans*[37], *Pantoea stewartii*, *P. agglomerans*[37,38] and *Pantoea anantis*[39,40] have undergone complete genome sequencing, but they have been found to have biocontrol effects and endophytic characteristics, and the whole genome of strains that can cause plant diseases has not been sequenced. Whole-genome sequencing is helpful to fully understand the pathogenic genes of bacterial perforation disease and the pathogenic variation in the process of interaction with the host. Whole-genome sequencing provides highly valuable basic data for the systematic study of the pathogenic mechanism and the evolution process of pathogens.

Conclusion

In this study, the Illumina HiSeq2500 and PscBio RS platforms were used to sequence the whole genome of *P. agglomerans*. The genes in the whole-genome data were annotated. In particular, annotations were made for effectors, such as secondary metabolite synthesis gene clusters, databases of pathogen-host interactions, and databases of virulence factors, for which preliminary pathogenicity analysis was performed. The whole-genome sequence of *P. agglomerans* provides substantial data and a theoretical basis for studies on the transcriptome, proteome, and metabonome of this bacterium. The results of this study enrich the knowledge base of *P. agglomerans*.

References

1. Rehder, A.. Manual of cultivated trees and shrubs hardy in North America. (The Macmillan Company, 1927).
2. Liao Kang. Study on wild fruit tree resources in Xinjiang. (Xinjiang science and Technology Press, 2013)
3. Gu, M. Apricot cultivars in China. Acta Hortic. 209, 63–67 (1979).
4. Zhang, J.Y. & Zhang, Z. In Flora of Chinese Fruit Trees(eds. Zhang, J.Y. & Zhang, Z.) 42–46 (China Forestry Press: Beijing, 2003).
5. Lin P.J., Cui N.R.. Tianshan wild fruit forest resources – a comprehensive study of Yili wild fruit forest(eds.Lin P.J., Cui N.R.) 45-52 (China Forestry Press, 2000).
6. Zhebentyayeva T., Reighard G., Gorina V., & Abbott A.. Simple sequence repeat (SSR) analysis for assessment of genetic variability in apricot germplasm. Theoretical and Applied Genetics, 106(3), 435-444(2003).
7. Yan G.R.. Fruit tree planting resources and biodiversity protection in Xinjiang. J. Research On Arid Areas, 13 (1), 64-65(1996).
8. Xu D.Y., Guo Q.Y., Yu S.L.. Eco climatic adaptability of apricot in Xinjiang. Xinjiang Meteorology, 26 (3), 37-39(2003).
9. Hou B., Xu Z.. Study on wild fruit trees and related species in Yili, China. J. Arid Area Research, 9(23), 455(2006).
10. Leyns F., Cleene M D., Swings J.G., & Ley J.D.. The host range of the genusxanthomonas. J. Botanical Review, 50(3), 308-356(1984).
11. Lamichhane, J.R.. Xanthomonas arboricola diseases of stone fruit, almond, and walnut trees: progress toward understanding and management. Plant Disease, 98(12), 1600-1610(2014).
12. Giovanardi D., Dallai D., & Stefani E. Population features of Xanthomonas arboricola pv. pruni from prunus spp. orchards in northern Italy. European Journal of Plant Pathology, 147(4), 1-11(2017).
13. Gerard M., Isidre L., Emilio M., Concepció M. & George-John N.. A model for predicting Xanthomonas arboricola pv. pruni growth as a function of temperature. Plos One, 12(5), (2017).
14. Hulin M. T., Mansfield J. W., Brain P., Xu X., Jackson R. W., & Harrison R. J.. Characterization of the pathogenicity of strains of Pseudomonas syringae towards cherry and plum. Plant Pathology, 67(5),

1177-1193(2018).

15. Lamichhane J. R., Varvaro L., Parisi L., Audergon J. M., Morris C. E.. Disease and frost damage of woody plants caused by Pseudomonas syringae: seeing the forest for the trees. Advances in Agronomy, 126, 235–295(2014).

16. Tong Y. P., Yang B. Y., Tian Q., Hu F. P., & Cai X. Q.. Isolation and identification of new pathogens of bacterial perforation of peach and plum. Journal of Fujian agriculture and Forestry University: Natural Science Edition, 49 (3), 7(2020).

17. Li, C. H., Jia, Y., Tian Y., Zhou L., & Liu F.. First report of necrotic disease caused by pantoea agglomerans on plum (prunus salicina) in china. Plant Disease, 104(4), 1248(2019).

18. Fang Z. D.. Research methods of plant diseases, Third Edition(ed.Fang Z. D.)179-180(China Agricultural Publishing House, 1998).

19. Bowman, J.P., & Mcmeekin, T.A.. Bergey's Manual of Systematic Bacteriology(eds. Bowman, J. P., & Mcmeekin, T. A.)465(Springer,2005)

20. Deletoile A., Decre D., Courant S., Passet V., Audo J., & Grimont P., et al. Phylogeny and identification of pantoea species and typing of Pantoea agglomerans strains by multilocus gene sequencing. Journal of Clinical Microbiology, 47(2), 300(2009).

21. Tho K. E., Wiriyajitsomboon P., & Hausbeck M.. First report of Pantoea agglomerans causing onion leaf blight and bulb rot in michigan. Plant Disease, 99(7), 1034-1038(2015).

22. Liu Y. Q., Ren Y. Z., Li G. Y., Ding S. L., Zhang H.. Identification of pathogens of cotton bacterial Boll Rot in Xinjiang. Journal of Plant Pathology, 2008, 38(3), 238-243

23. Cao H. Y.. Bacterial dry stalk rot in maizea, new disease. Beijing: Chinese Academy of Agricultural Sciences, 20-25(2010).

24. Yan Y. N., He H., Ye Y.J., Lu W.B., Wen S. H., Zhang X. Q.. Pathogen identification of banana leaf sheath rot. Journal of plant pathology, 41 (2), 124-130(2011)

25. Xu C. N., Chi F. M., Ji Z. R., Zhang J. X., Wang N., Liu Y. C., et al. Isolation and identification of pathogenic bacteria of bacterial fruit spot of sweet cherry. Fruit trees in southern China, 46 (4), 6-9(2017).

26. Morales V., Silva R., Ochoa M., Valadez M., & Farfán G.. First report of Pantoea agglomerans causing leaf blight and vascular wilt in maize and sorghum in Mexico. Plant Disease, 91(10), 1365-1367(2007).

27. Yang K. Q., Qu W. W., Liu X., et al. First report of Pantoea agglomerans causing brown apical necrosis of walnut in China. Plant Disease, 95(6), 773-773(2011)

28. Han Y. Z., Fan Z.W., Yao Y. T., et al. First report of leaf blight disease on Pennisetum sinese in China Caused by Pantoea agglomerans. Plant disease, 104(10), 272(2020).

29. Guo M., Liu Y., Liu S.N., et al. First report of bacterial soft rot caused by Pantoea agglomerans on Chinese Cabbage in China. Plant disease, 104(10), 277(2020).

30. Coulthurst, Sarah J.J.. The Type VI secretion system a widespread and versatile cell targeting system. Research in Microbiology, 164(6), 640-654(2013).

31. Bernal P., Llamas M.A., Filloux A.. Type VI secretion Systems in plant-associated bacteria. Environmental Microbiology, 20(1), 1-15(2018).

32. Shyntum D., Venter S, Moleleki L., et al. Comparative genomics of type VI secretion systems in strains of Pantoea ananatis from different environments. Bmc Genomics, 15(1), 163(2014).

33. Kirov S.. Bacteria that express lateral flagella enable dissection of the multifunctional roles of flagella in pathogenesis. Journal of Basic Microbiology, 224(2), 151-159(2003).

34. Baker M.D., Wolanin, P. M., & Stock, J. B.. Signal transduction in bacterial chemotaxis. Journal of Biological Chemistry, 28(1), 9-22(2010).

35. Matilla M. A., & Tino K.. The effect of bacterial chemotaxis on host infection and pathogenicity. Fems Microbiology Reviews, 42(1), 1-14(2017).

36. hua K. L., Chan Y. Y., & Gan Y. H.. Flagella are virulence determinants of burkholderia pseudomallei. Infection and Immunity, 71(4), 1622-1629(2003).

37. almer M., Maayer P. D., Poulsen M., Steenkamp E.T., & Venter S. N.. Draft genome sequences of pantoea agglomerans and pantoea vagans isolates associated with termites. Standards in Genomic Sciences, 11(1), 23(2016).

38. Matsuzawa T., Mori K., Kadowaki T., Shimada M., Tashiro K., & Kuhara S., et al. Genome sequence of pantoea agglomerans strain IG1. Journal of Bacteriology, 194(5), 1258(2012).

39. Maayer P. D.. Genome comparisons to identify selected pathogenicity factors of a plant- associated Pantoea ananatis strain. Molecular microbiology, 46(2), 427-437(2011).

40. Weller S. T., Toth I., Maayer P. D., & Coutinho T.. Swimming and twitching motility are essential for attachment and virulence of Pantoea ananatis in onion seedlings. Molecular Plant Pathology, 18(5), 185-196(2017).
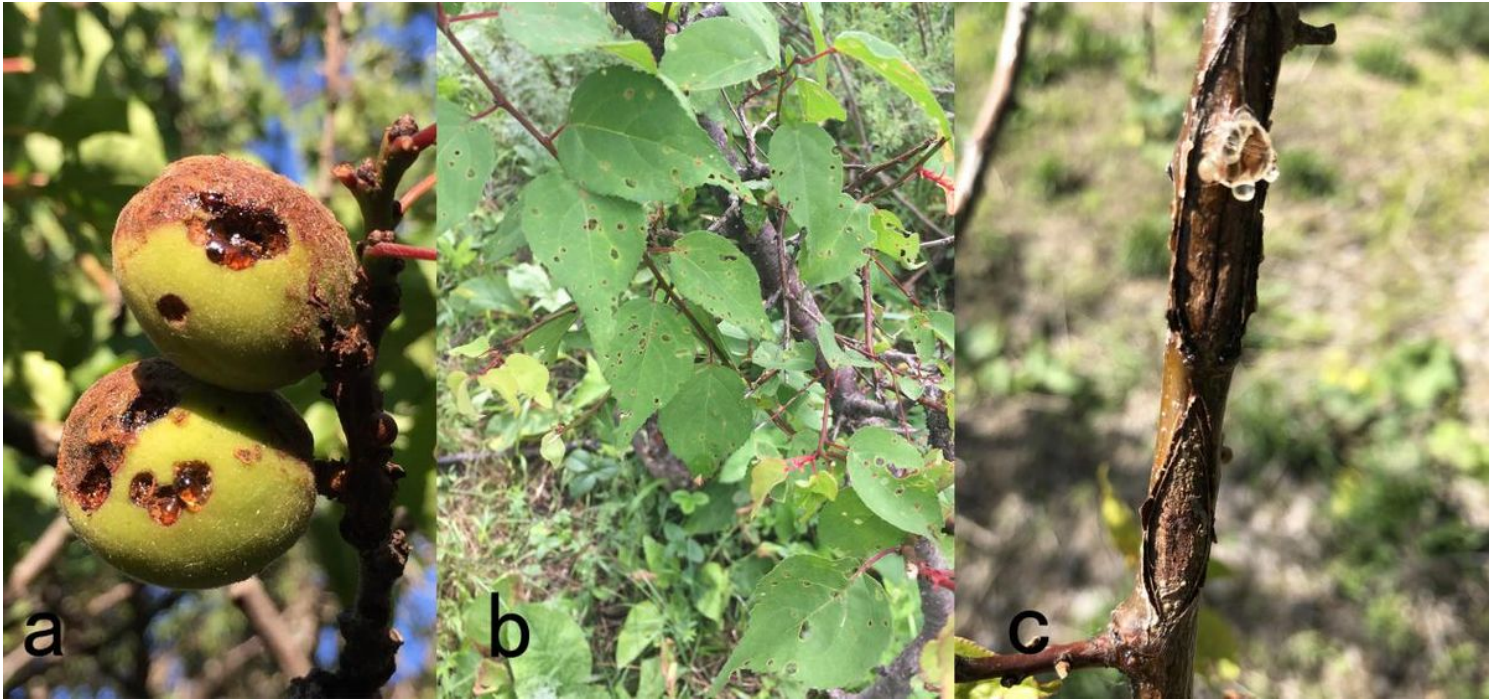
Figures

Figure 1

Symptoms of pathogen-infected apricot leaves, fruits and twigs a. A scar on apricot fruits; **b.** A spot-hole appeared on apricot leaves; **c.** A canker on the stems.
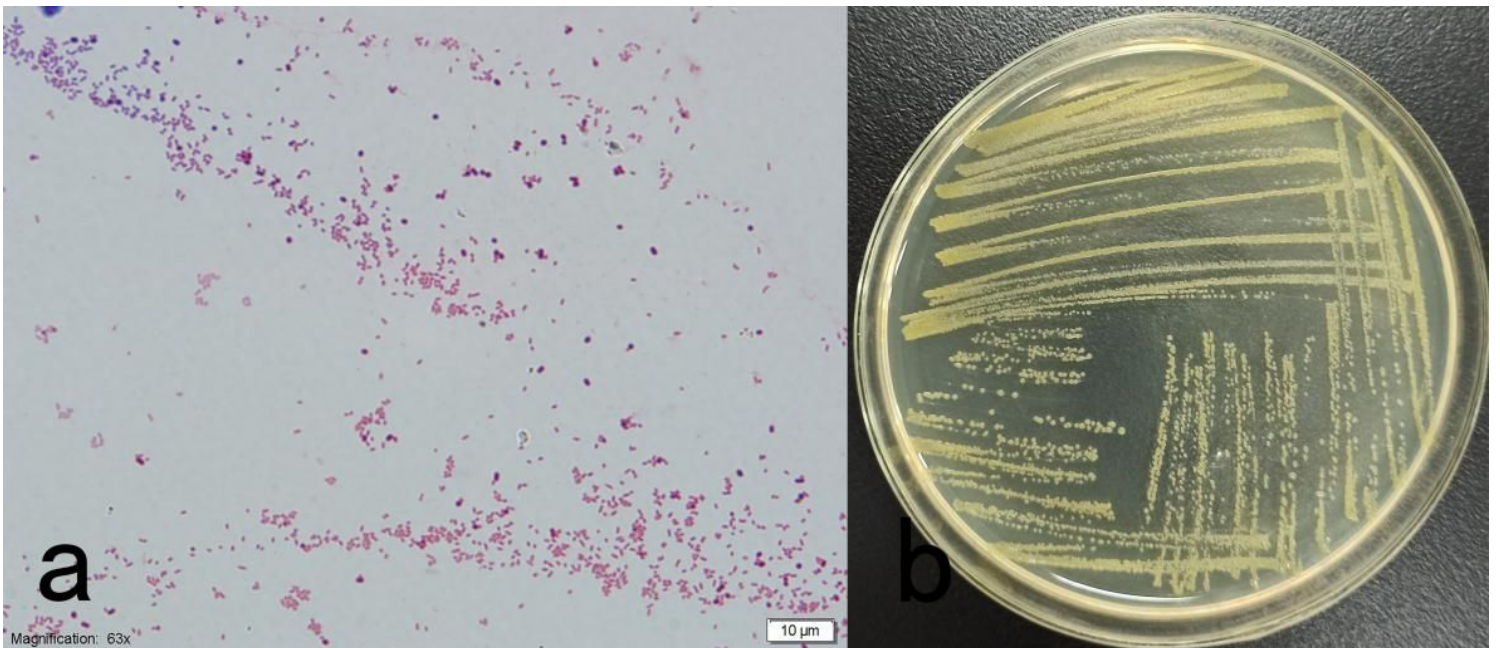

Figure 2

Morphological characteristics of the pathogen a. Gram staining; **b.** colony on NA.
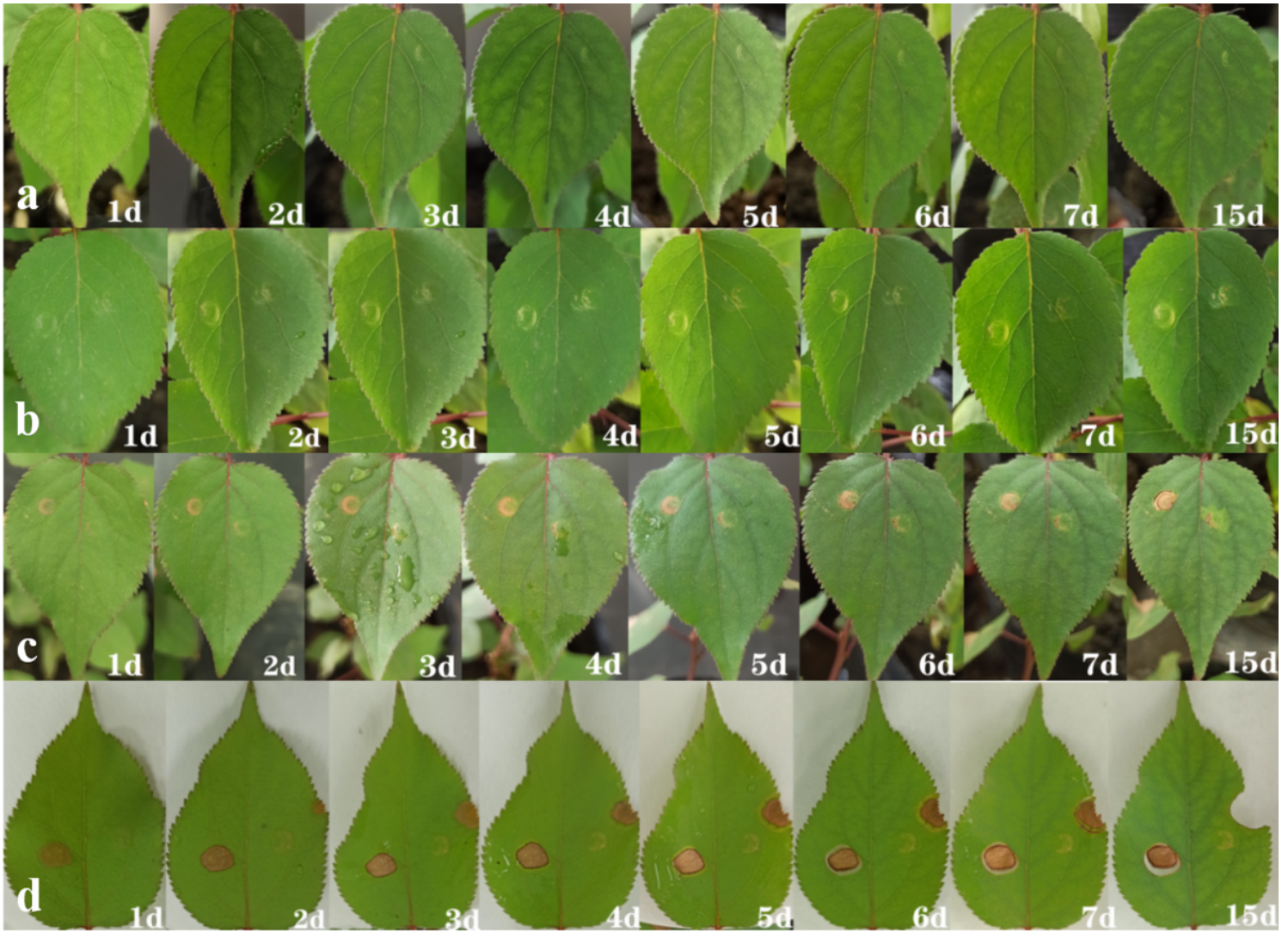
Figure 3

Pathogenicity of the tested strains. a. CK inoculated after 1-15 d; **b.** bacterial suspension of GL9-2 OD=0.1 inoculated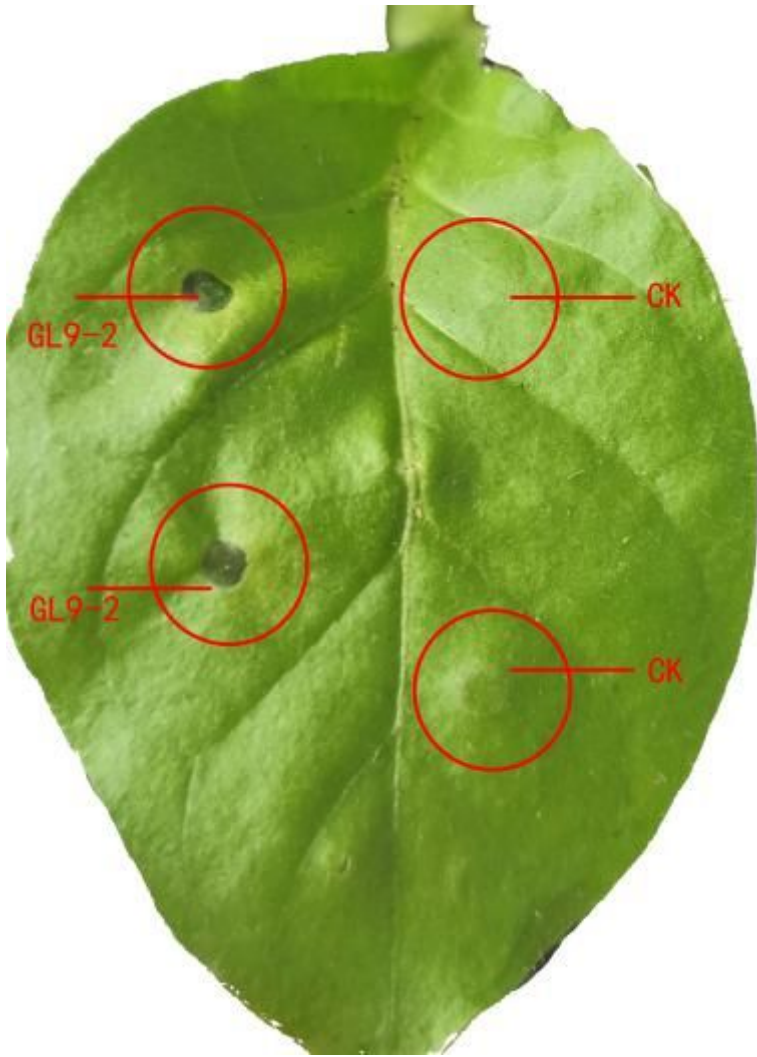 after 1-15 d; **c.** bacterial suspension of GL9-2 OD=0.3 inoculated after 1-15 d; **d.** bacterial suspension of GL9-2 OD=0.5 inoculated after 1-15 d.

Figure 4

Strains GL9-2 inoculated in tobacco

Figure 5

Phylogenetic tree. a. Phylogenetic tree based on 16S rDNA gene sequences; **b.** Phylogenetic tree based on *fusA, gyrB, leuS, pyrG, rpoB* and *rlpB* gene sequences.

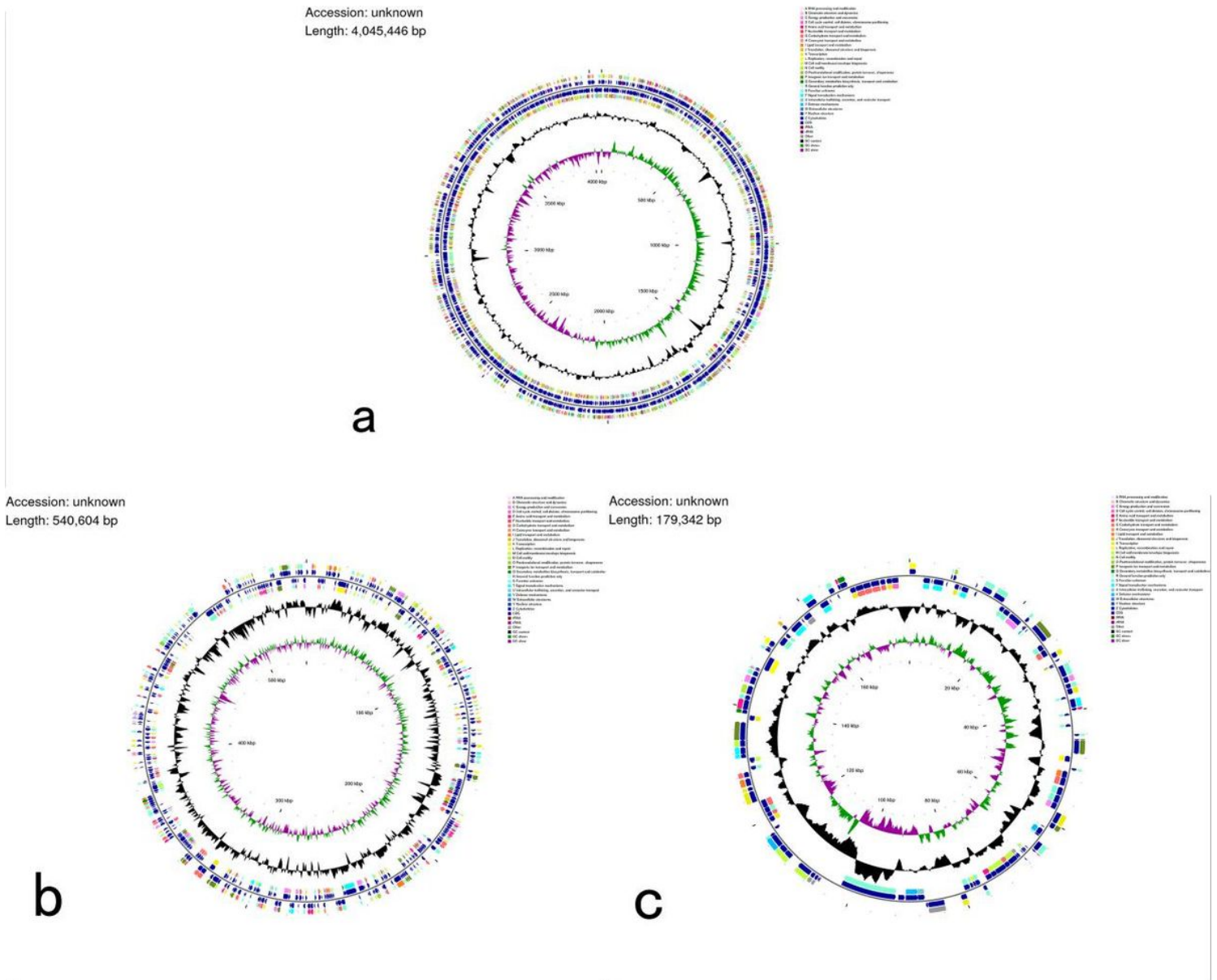# Figure 6

Circular genome diagram of the GL9-2 strain genome. a. Circular genome diagram of the chromosome; b. Circular genome diagram of the plasmid A; **c.** Circular genome diagram of the plasmid B.
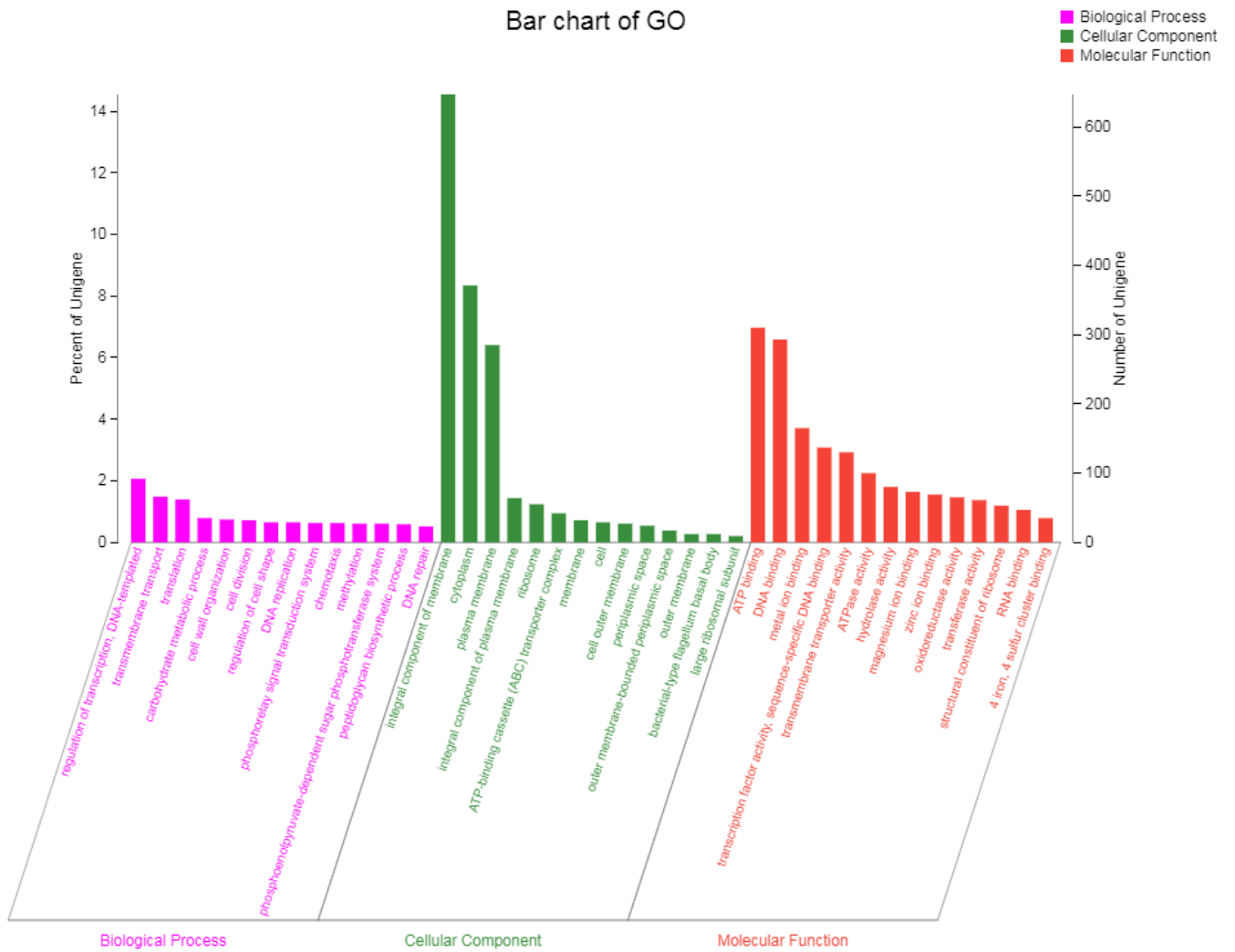
Figure 7

Gene distribution based on the Gene Ontology (GO) database of strain GL9-2
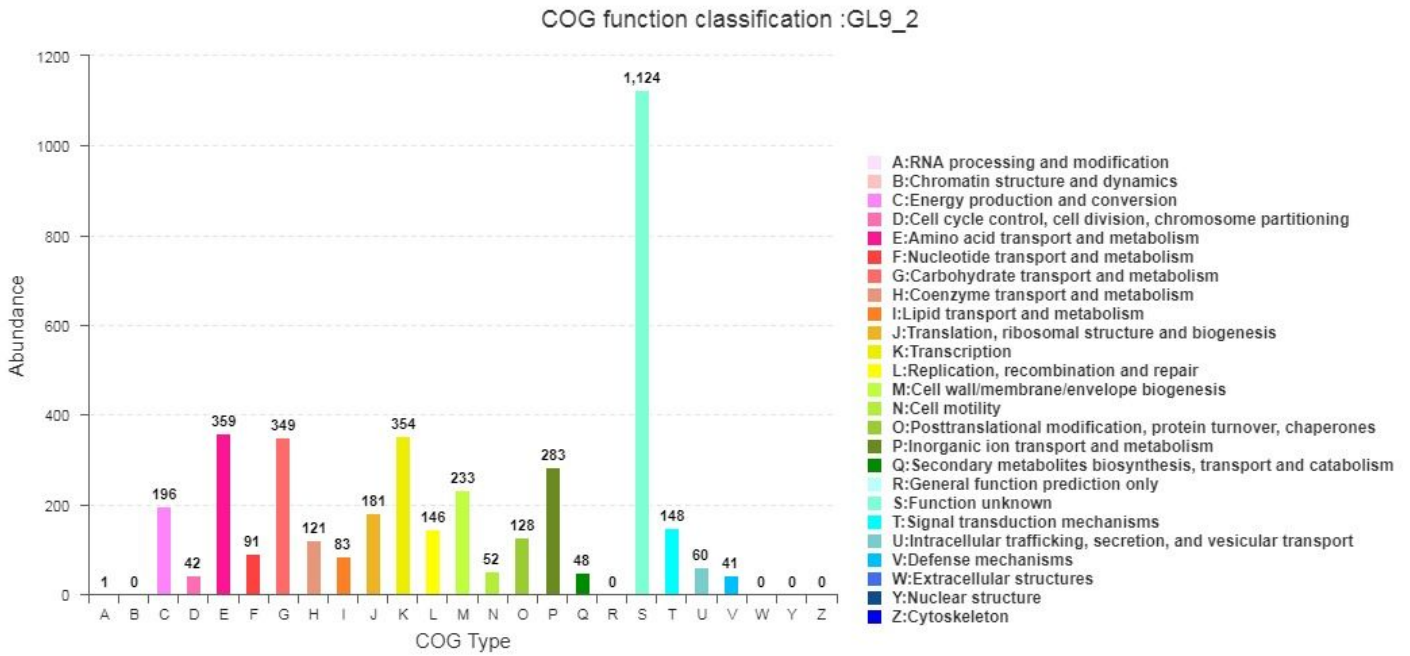
COG function classification : GL9_2

Figure 8

Gene distribution based on COGs classification of strain GL9-2

Histogram of KEGG

Figure 9

KEGG pathway annotation of stain GL9-2