High-Quality Genome Resource of Mango Bacterial Black Spot Pathogen *Xanthomonas citri* pv. *mangiferaeindicae* GXG07 Isolated from Guangxi, China

Fengzhi Bie,¹ Yiming Li,^{1,2} Zhibin Liu,¹ Meijing Qin,¹ Shuping Li,¹ Xue Dan,^{1,2} Sheng Huang,¹ Yong-Qiang He,^{1,2,†} and Wei Jiang^{1,†}

¹ State Key Laboratory for Conservation and Utilization of Subtropical Agro-bioresources and College of Life Science and Technology, 100 Daxue Road, Nanning, Guangxi 530004, China

² National Demonstration Center for Experimental Plant Science Education, College of Agriculture, Guangxi University, Nanning 530004, Guangxi, China

Funding

The work was supported by the National Key R&D Program of China (2018YFD0200302) and the Natural Science Foundation of China (31860033).

Keywords

mango bacterial black spot, *Xanthomonas citri* pv. *mangiferaeindicae*, PacBio sequencing, genome resource, transcription activator-like effectors

Genome Announcement

Xanthomonas citri pv. *mangiferaeindicae* is the causal agent of mango bacterial black spot (MBBS), which is present in many mango-growing regions and leads to great economic losses to the mango industry. Due to the limitation of high-quality genomic resources, little is known about the molecular pathogenesis of *X. citri* pv. *mangiferaeindicae*. Here, we used PacBio High Fidelity reads (HiFi) sequencing technology to sequence and analyze the whole genome of *X. citri* pv. *mangiferaeindicae* strain GXG07 isolated from Guangxi, the largest mango-growing region in China.

Mango (*Mangifera indica* L.), belonging to the Anacardiaceae family, is one of the major special flavor fruit in tropical and subtropical areas and one of the most important fruits in the Guangxi region, China. In recent years, MBBS, also called mango bacterial canker or angular leaf spot, was observed increasingly in most mango yards in Guangxi, which led to huge economic losses in both mango planting and processing. The causal agent of MBBS, *X. citri* pv. *mangiferaeindicae*, was first isolated in South Africa (Doidge 1915), then reported in tropical and subtropical mango-producing areas of India, Pakistan, Brazil, Philippines, China, and the United States (Gagnevin and Pruvost 2001; Gagnevin et al. 1997; Patel et al. 1948). *X. citri* pv. *mangiferaeindicae* can infect mango leaves, fruit, branches, and trunk, resulting in a significant reduction in mango yield and quality, and decline in vitality of the evergreen tree. Genomic analysis is currently the most powerful tool for genetic assays and biological research; however, few complete genome sequences are available in *X. citri* pv. *mangiferaeindicae*. To date, three strains of *X. citri* pv. *mangiferaeindicae* whose genomes were sequenced have been reported: the LMG 941 strain from India (Midha et al. 2012); XC01 strain from Guangxi, China (Liu et al. 2018); and L242 from the United States (Dommel et al. 2019). However, there are obvious contradictions and inconsistency among the three genomes, because the genome sizes of strains LMG 941, XC01, and L242 are 5.11 Mb (Midha et al. 2012), 3.86 Mb (Liu et al. 2018), and 5.30 Mb (Dommel et al. 2019), respectively. Strikingly, in the first two genomes, there is no information about the transcription activator-like effector (TALE) genes, the important virulence and avirulence genes in most pathovars of *X. citri* (An et al. 2020). For strain LM941, because the genome was sequenced on the Roche 454 GS (FLX Titanium) pyrosequencing platform, it is reasonable that the

[†]Corresponding authors: W. Jiang; weijiang@gxu.edu.cn, and Y.-Q. He; yqhe@gxu.edu.cn

F. Bie and Y. Li contributed equally to this work.

The author(s) declare no conflict of interest.

Accepted for publication 1 October 2021.

Table 1. Genomic features of *X. citri* pv. *mangiferaeindicae* strain GXG07 and other sequenced *X. citri* pv. *mangiferaeindicae* strains^a

Genome features	Strains					
	GXG07	LMG 941 ^b	XC01 ^c	LG56-10 ^d	LG81-27 ^d	DE0062 ^d
Genome size (bp)						
Chromosome	5,166,537	5,111,537	3,865,165	5,255,984	5,107,011	6,002,465
Plasmid-1	(pGXG07-1) 79,634	No	No	16,241	71,216	No
Plasmid-2	(pGXG07-2) 30,169	No	No	4,818	22,906	No
G+C content (%)	64.69	64.85	68.93	64.58	64.72	57.20
Coding density (%)	86.08	–	89.55	–	–	–
Genes	4,871	4,420	3,374	4,555	4,493	5,586
Coding sequences	4,533	4,113	3,173	4,464	4,402	5,502
Transfer RNA	54	50	52	50	50	73
Ribosomal RNA	6	4	6	41	41	8
Noncoding RNA	278	215	139	38	38	–
TALE genes	10	0	0	0	0	–
Average CDS length	1,001	–	1,006	–	–	–

^a Abbreviations: TALE = transcription activator-like effector, CDS = coding sequence, and – indicates data not available.

^b Data from Midha et al. (2012).

^c Data from Liu et al. (2018).

^d Genomic data from NCBI (<https://www.ncbi.nlm.nih.gov/genome/>). There is no detailed genome data of strain L242 published (Dommel et al. 2019).

highly repetitive TALE gene sequences were not correctly assembled. However, for strain XC01, the correctness of the extremely small genome in *X. citri* pv. *mangiferaeindicae* deserves further consideration. Here, we reported the complete genome sequence for strain GXG07 generated by long-read sequencing.

In summer 2020, *X. citri* pv. *mangiferaeindicae* strain GXG07 was isolated and purified from a diseased leaf of *M. indica* L. ‘Guire 7’, a major mango cultivar in Baise, Guangxi Autonomous Region, the largest mango-growing area in China. GXG07 can produce typical black necrosis on mango leaves by artificial inoculation. Molecular identification showed that the 16S ribosomal DNA sequence of GXG07 is 99.9% of that of strain LMG 941, the characterized *X. citri* pv. *mangiferaeindicae* strain (Midha et al. 2012), suggesting that GXG07 belongs to *X. citri* pv. *mangiferaeindicae*. To sequence the whole genome of GXG07, its total DNA was obtained and purified by using Qiagen Genomic-tip 20G. The average fragment length of the sequencing library is about 15 kb and the sequencing was performed using the PacBio Sequel II System (Pacific Biosciences) at Biomarker Technologies Co., Ltd. (Beijing, China). In total, 250,505 reads and 2,393,214,969 bp were obtained. N₅₀ length was 10,307 bp and mean coverage for the strain GXG07 genome was about 450x. The HiFi reads have been de novo assembled by using the Hifiasm software (Cheng et al. 2021). The Pilon v1.22 software was used to correct the assembly with Illumina data, and Circulator v1.5.5 software was used to cyclize and adjust the starting site. Three circular contigs were assembled and the two smaller contigs contained *parA*, *trwB*, and *trwC*, the plasmid-related genes, revealing their plasmid identity (Niu et al. 2015). CheckM v1.1.3 identified an average of 99.52% genome completeness and an average contamination of 0.00%, using the *Xanthomonas* genus-level taxonomic marker set (Parks et al. 2015). Thus, the genome of strain GXG07 contained a chromosome and two plasmids, with sizes of 5,166,537, 79,634, and 30,169 bp, respectively (Table 1). Phylogenetic analysis (Zuo and Hao 2015) based on genome sequences of some sequenced *Xanthomonas* strains indicated that GXG07 but not XC01 is clearly clustered with other *X. citri* pv. *mangiferaeindicae* strains in the clade of the species of *X. citri* (Fig. 1). Values of average nucleotide identity of strain GXG07 compared with LG81-27 and LMG 941 were 99.79 and 99.87%, respectively, which exceed the generally accepted 95% cut-off level for taxonomy affiliation of newly sequenced genomes (Richter et al. 2016). Functional annotation of the three assembled contigs was done using Prodigal v2.6.3 (Hyatt D et al. 2010). The predicted genes encode 4,871 proteins, 54 transfer RNAs, 6 ribosomal RNAs, and 278 other noncoding RNAs. The strain GXG07 genome encodes typical pathogenicity and virulence factors found in *Xanthomonas* spp. (An et al. 2020), including extracellular hydrolases, extracellular polysaccharides, adhesins, type II secretion system, type III secretion system, and type III-secreted effectors (T3SEs) (An et al. 2020; Dommel et al. 2019). Of the T3SEs, 10 TALEs were annotated to be encoded in the

Fig. 1. Phylogenetic analysis of *Xanthomonas citri* pv. *mangiferaeindicae* GXG07 based on genome sequences of representative *Xanthomonas* strains. Genome sequences were obtained from NCBI (<https://www.ncbi.nlm.nih.gov/genome/>). The tree was annotated and visualized by using the CVTtree3 web server (<http://cvttree.online/v3/cvttree/>) and phylogenetic trees by using the neighbor-joining method. K-tuple length is 6 (Zuo and Hao 2015). The “X” indicates the position of *X. citri* pv. *mangiferaeindicae* XC01 and the red dotted line corresponds to the *X. citri* pv. *mangiferaeindicae* clade.

genome, which belong to TalED, TalIL, TalIT, TalIU, TalIV TalIW, TalIX, TalIY, TalIZ, and TalJA classes based on the patterns of the repeat variable diresidue (RVD) in the central domain of each TALE (Grau et al. 2016) (Table 2). The strain GXG07 genome provides gene resources for elucidating the biology and pathogenicity of the phytopathogen and developing new strategies for controlling mango bacterial black spot.

This whole-genome PacBio sequencing project has been deposited in NCBI under the accession numbers CP073209 (chromosome), CP073210 (pGXG07-1), and CP073211 (pGXG07-2).

Gene	TALE class ^a	Repeat variable diresidues (RVDs)	Start ^b	End	Strand
Chromosome					
<i>tal1a</i>	TalIL	NI-NG-HD-NG-HD-HD-NG-NG-NG-NI-HD-HD-NG	1,362,518	1,365,503	+
<i>tal1b</i>	TalED	NI-NG-HD-NG-NI-NG-HD-NG-NG-HD-NI-NI-NG-HD-NI-NI-NI-NG	1,368,509	1,372,004	+
<i>tal2</i>	TalIT	NI-NG-HD-HD-NI-HD-NG-HD-NG-NG-HD-NG-NG-HD-NG-NI-NG-NG-HD-NG-NG	1,385,932	1,389,733	+
<i>tal3</i>	TalJA	NI-NG-NI-NI-NG-NG-HD-NI-NG-NI-NG-NI-NG-HD-NG	1,401,180	1,404,396	−
<i>tal4a</i>	TalIU	NI-NG-HD-NG-HD-NI-NI-HD-NI-HD-NG-NG-NG-HD-NI-NI-HD-NG	1,829,310	1,832,805	+
<i>tal4b</i>	TalIZ	NI-NG-NI-NG-HD-NI-HD-NI-HD-HD-HD-NI-NG-HD-NI-HD-NI-HD-HD-HD-HD-NI-HD-NG-NG	1,836,635	1,840,844	−
<i>tal5</i>	TalIV	NI-NG-NG-NG-HD-HD-NI-NG-HD-HD-NI-NG-HD-NI-NG-NG-NI-NI-NG-HD-NG-NG	2,846,486	2,850,389	+
pGXG07-1					
<i>tal6</i>	TalIW	NI-NG-NI-NG-NI-NG-HD-NG-HD-NI-NG-NG-NG-NI-HD-HD-NG	41,464	44,856	+
<i>tal7</i>	TalIX	NI-NS-NS-NI-NI-NS-HD-NS-NS-NG-NS-HD-NI-NI-NG-HD-NG-NG-HD-NI-NG	52,470	56,270	+
<i>tal8</i>	TalIY	NI-NG-HD-HD-NG-HD-NG-HD-NG-NG-NG-NI-NI-NG-NG-HD-NG-NG	61,374	64,868	+

^b Genomic coordinates of TALE genes in the genome of strain GXG07 (CP073209, CP073210, and CP073211).