مجله دنیای میکروب‌ها
سال یازدهم شماره سوم (پیاپی ۳۶) پاییز ۱۳۹۷
صفحات ۳۰۹-۲۹۴

Original Article

ویژگی های ژنومی و آنالیز فیلوژنتیک سویه ایرانی با دامنه میزبانی محدود *زانتوموناس سیتری* زیرگونه *سیتری* NIGEB-88

امیر جلالی[۱]، سید مهدی علوی*[۲]، محمد حسین سنگتراش[۳]

[۱]دکتری، پژوهشکده بیوتکنولوژی کشاورزی، پژوهشگاه ملی مهندسی ژنتیک و زیست فناوری، تهران، [۲]استادیار، پژوهشکده بیوتکنولوژی کشاورزی، پژوهشگاه ملی مهندسی ژنتیک و زیست فناوری، تهران، [۳]دانشیار، گروه زیست شناسی، دانشکده علوم پایه، دانشگاه سیستان و بلوچستان، زاهدان.

چکیده

سابقه و هدف: باکتری *زانتوموناس سیتری* زیرگونه سیتری (Xcc) عامل بیماری شانکر در مرکبات است. نتایج به دست آمده از آزمون های تشخیصی اولیه، نشان دهنده وجود سویه هایی با دامنه میزبانی گسترده (فرم بیماری زای A) و محدود (فرم های بیماری زای A^* و A^w) در ایران است. با این حال، آنالیزهای تفکیکی جدید انجام شده در سال ۲۰۱٤ بر مبنای نتایج MLVA-14 و MLVA-31 نشان می دهد که این سویه ها از نظر ژنتیکی تنها به فرم بیماری زای A^* تعلق دارند. این مطالعه با هدف بررسی ویژگی های ژنومی و آنالیز فیلوژنتیک سویه XccA* NIGEB-88 باکتری *زانتوموناس سیتری* زیرگونه *سیتری* با دامنه میزبانی محدود انجام شد.

مواد و روش ها: در این مطالعه به منظور به دست آوردن اطلاعات بیشتر درباره ویژگی های ژنتیکی و روابط فیلوژنتیک سویه های ایرانی باکتری *زانتوموناس سیتری*، با استفاده از روش Illumina ژنوم سویه ایرانی NIGEB-88 به طور کامل تعیین توالی و به وسیله نرم افزارهای مختلف بررسی شد.

یافته ها: مشخصات کلی ژنوم مانند اندازه، تعداد پلاسمیدها، میانگین محتوای GC، تعداد مناطق کد کننده پروتئینی و ساختار RNA سویه ایرانی مورد مطالعه، مشابه با سویه های فرم های بیماری زای دیگر باکتری Xcc بود. همچنین مطالعه عوامل بیماری زای بالقوه و عوامل شرکت کننده در تعیین دامنه میزبانی، مانند افکتورهای سیستم ترشحی تایپ ۳، اجزای سیستم ترشحی تایپ ٤ و لیپوپلی ساکاریدهای سطحی نشان داد که شباهت زیادی با سویه های دو فرم بیماری زای A و A^w وجود دارد.

نتیجه گیری: نتایج آنالیزهای فیلوژنتیکی و بررسی ژن های بیماری زایی و تعیین دامنه میزبانی نشان داد که سویه ایرانی XccA*NIGEB-88، شباهت زیادی به سویه ایرانی XccA*NIGEB-386 و سویه های فرم بیماری زای A^* با دامنه میزبانی محدود دارد.

واژگان کلیدی: شانکر باکتریایی مرکبات، فرم بیماری زای A^*، توالی یابی کامل ژنوم، *زانتوموناس سیتری* زیر گونه سیتری.

دریافت مقاله: اسفند ماه ۹٦ پذیرش برای چاپ: اردیبهشت ماه ۹۷

*) آدرس برای مکاتبه: تهران، پژوهشگاه ملی مهندسی ژنتیک و زیست فناوری.

تلفن: ۰۲۱٤٤۷۸۷٤۲۵

پست الکترونیک: mealavi@nigeb.ac.ir

Journal of Microbial World
Volume 11, No. 3, September 2018

Original Article

Genomic characterization and phylogenetic analysis of a narrow host-range Iranian strain of *Xanthmonas citri* sub. citri, NIGEB-88

Amir Jalali[1], Seyed Mehdi Alavi[2], Mohammad Hossein Sangtarash[3]

[1] Ph.D, Institute of Agriculture Biotechnology, National Institute for Genetic Engineering and Biotechnology, Tehran, Iran. [2]Assistant Professor, Institute of Agriculture Biotechnology, National Institute for Genetic Engineering and Biotechnology, Tehran, Iran. [3]Associate Professor, Department of Biology, Faculty of Sciences, University of Sistan and Baluchestan, Zahedan, Iran.

Abstract

Background & Objectives: *Xanthomonas citri* subsp. citri (Xcc) is the causal agent of citrus bacterial canker. In Iran, for the first time, Alizadeh and Rahimian reported the presence of Xcc on Mexican lime (*C. aurantifolia*), in 1990. Early characterization data suggested the presence of both wide (pathotype A) and narrow (pathotype A* and A^{w}) host range Xcc strains in Iran. However, more recently a discriminant analysis of MLVA-31 and MLVA-14 data in 2014 showed that Iranian strains of Xcc have genetically belonged to the host-restricted pathotype A* but not to the pathotype A. This study was aimed for genomic characterization and phylogenetic analysis of a narrow-host-rang Iranian strain of *X. citri* sub. citri, NIGEB-88.

Materials & Methods: To determine genetic characteristics and phylogenetic relationships of Iranian strains, Illumina sequencing method was used to obtain a draft genome sequence of Xcc strain NIGEB-88.

Results: General features of Iranian strain such as genome size, number of plasmids, average genomic GC content, number of CDS and structural RNA were similar to those of other Xcc pathotypes. Furthermore, studies of various potential virulence and host range determinants factors such as type III secretion system effectors, type IV secretion system and surface Lipopolysaccharides revealed that this pathogen is very close to XccA and XccA^{w} strains.

Conclusion: Phylogenetic and assessment of genes related to virulence and host specificity showed that XccA*NIGEB-88 is more closely related to other Iranian strain, XccA*NIGEB-386, and narrow-host-rang pathotype A*.

Keywords: Citrus bacterial canker, A*-type, Whole genome sequencing, *Xanthomonas citri* subsp. *citri.*

Received: April 2018 Accepted: May 2018

Correspondence to: Seyed Mehdi Alavi
Tel: +982144787425
E-mail: mealavi@nigeb.ac.ir
Journal of Microbial World 2018, 11(3): 294-309.

Introduction

Citrus culture in Iran is a great venture occupying ~1.5% of the arable lands, providing more than three million job opportunities in different sectors. Among all citrus species, Mexican lime (*Citrus aurantifolia*) has a great place with more than six hundred thousand tones ranks the sixth in the world (1).

Despite current strategic momentums in terms of citriculture, common pathogens pose a great threat, endangering citrus production. One of the most important citrus diseases which has affected citrus crop is Asiatic citrus canker that caused by a gram-negative pathogen, *Xanthomonas citri* subsp. citri (2).

This pathogen appears to be a polymorphic type of the disease, containing three different pathotypes: A, A* and Aw (3). Strains belonging to pathotype A show a broad-host-range and affect most commercial citrus varieties, including grapefruit (*C. paradisi*), sweet orange (*C. sinensis*), and Mexican lime (*Citrus aurantifolia*). In contrast, A* and Aw strains are narrow-host range pathotypes. Aw strains affect Mexican lime and alemow (Citrus macrophylla) (4), while A* strains only affect Mexican lime (5).

The results of comparative genomic analysis on Xcc strains showed that strains of XccA* are the most diverse with more genomic polymorphism compared to A and Aw pathotypes strains (6).

Disease symptoms include distinctively raised, necrotic lesions on fruits, twigs, and leaves (Figure 1) which in severe cases, the infection results in defoliation, blemished fruit, premature fruit drop, die-back of twigs and can eventually kill the tree (7).

In Iran, the presence of citrus bacterial canker reported for the first time in the Kerman province in 1990 (8). Primary studies, based on biochemical and physiological tests, suggested the existence of both A and A* pathotypes in Iran (9-11). However, more recent findings by Provost-Goodarzi et al. (12) showed all Iranian strains are related to Xcc A* pathotype. They were analyzed two hundred and thirty-nine Iranian strains by MLVA-31 and MLVA-14 analysis as well as pathogenicity tests.

Based on the results of these analyses, two geographically separate groups with different genetic and pathogenic characteristics were identified. One of the subgroups named cluster 4.4 and only includes the strains isolated from the Sistan and Baluchistan province. Strains of this cluster have some pathogenic properties that had never been previously reported in any strains of Xcc A* pathotype.

Figure 1. Asiaticus citrus canker symptoms are recognizable on leaf, stem and fruit. The symptoms are irregular to round, corky, raised pustules, water soaked margins, yellow halo in the center, or ring effect around the lesion.

All other Iranian strains were grouped in cluster 4.1 that also containing strains from Saudi Arabia. However, strains of both subgroups have moderate phenotypes and are distinguishable from A strains by delayed development of canker-like lesions (13).

In our previous study, we employed comparative genomic analysis on *Xanthomonas citri* subsp. citri strain A* NIGEB-386 and *Xanthomonas citri* subsp. citri strain Aw12879 and *Xanthomonas citri* subsp. citri strains 306 to identify any differences between three Xcc pathotypes. Comparison of candidate genes related to the virulence and host specificity such as type III effectors, type IV secretion system, and lipopolysaccharide synthesis, showed that XccA* NIGEB-386 shared the closest relationship with other narrow host range strain from pathotype Aw but not to the pathotype A strain (14).

Materials and methods

Organism information, genome sequencing, and annotation

The experimental Xcc strain in this study was originally isolated in 2009 from infected leaves of Mexican lime (*C. aurantifolia*) in Hasht Bandi city (Hormozgan province, Iran). For purification and isolation of genomic DNA, bacterial culture was grown overnight on nutrient broth medium (Merk KGaA, Darmstadt, Germany) at 28°C and AccuPrep® Genomic DNA Extraction Kit (Cat. No.; K-3032, Bioneer, Republic of Korea) was used according to the manufacturer's instructions. To check the condition of Genomic DNA, including the actual presence or absence of DNA in a sample, we used gel electrophoresis and the ratio of OD260/280, as an indicator of sample purity. The values of 1.8–2.0 for the ratio of OD260/280 are an indication of a good DNA sample. A ladder or smear below a band of interest may indicate the nicking or other damage to the DNA. In addition to proper quality, correct quantification of gDNA is essential for library construction. Therefore, we quantify the genomic dsDNA material by using a Nanodrop spectrophotometer (Nanodrop ND-1000; NanoDrop Tech. Inc., Wilmington, DE) (15).

Whole-genome sequencing performed using Illumina HiSeq 2000 platform with a paired-end module (Macrogen, Republic of Korea). The shotgun sequencing yielded 14,649,820 paired-end reads amounting to 1,479,631,820 bases (280-fold coverage) with an insert size of 419. de novo assembly using Velvet version 1.2.10 (16) performed for several times and obtained contigs from these assemblies aligned against each other by using BLASTn to close gaps and merge contigs together within the scaffolds. The scaffolding procedure leading to build up a total length of 5,236,943 bp (GC content 64.7%) placed in 16 scaffolds with a: N50 = 1,308,013; L50 = 2 and two plasmids named, pXCC_28 (28,978 bp) and pXCC_55 (55677 bp) that were submitted to GenBank. Automated annotation performed using the NCBI Prokaryotic Genome Automatic Annotation Pipeline (PGAAP) (17).

We used PSORTb 3.0.2 (18) to classified proteins based on their subcellular localization. To provide a complete and accurate classification of protein families and domains, assignment of genes to the Pfam domains database performed using EggNOG 4.1 (19).

For functional categorization of each protein-coding gene according to the Cluster of Orthologous Groups database (COGs) EMBL-EBI server used (20).

Phylogenetic analysis

The evolutionary position of XccA* NIGEB-88 relative to other strains and species of *Xanthomonas* was determined by multi-locus sequence analysis (MLSA) (21). Together with the two Iranian strains, partial and complete protein sequences of nine housekeeping genes (*uvrD, secA, carA, recA, groEL, dnaK, atpD, gyrB*, and *infB*) from 30 other *Xanthomonas* spp., and five *Xylella* fastidiosa (as out-group species) strains were selected from GenBank for phylogenetic tree construction.

The concatenated amino acid sequences aligned using Clustal Omega (22) and the phylogenetic tree constructed from sequences using CLC Genomics Workbench v7.5 (CLC Bio, Aarhus, Denmark) by the maximum likelihood method.

Additional genetic analysis

To identifying conserved syntenic segments and possible genome rearrangements such as deletions, inversions, duplications and translocations between two Iranian Xcc A* strains, genome sequences of XccA* NIGEB-88 (GenBank assembly accession: GCF_002742455.1) and XccA* NIGEB-386 (GenBank assembly accession: GCF_001956275.1) were aligned and visualized in progressive mode using MAUVE (23). To identify the dispersion of coding sequences, genomic regions with atypical GC content and GC skew, we used web-based tool

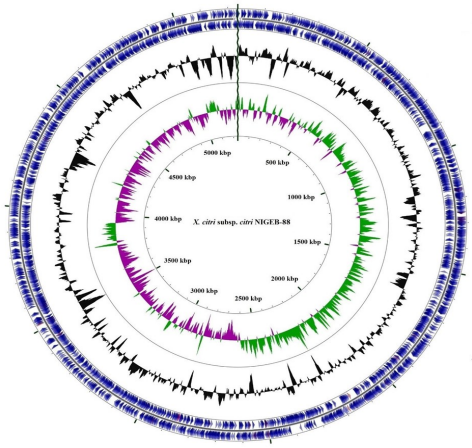

Figure 2. Circular representation of the genome of *Xanthomonas citri* subsp. *citri* strain *A** NIGEB-88. Rings from inside to outside: (1) GC skew$^-$ (violet), (2) GC skew$^+$ (green), (3) GC content (black) and (4) ORF sequences distribution on plus and minus strands (Blue).

GC-Profile (24). To display genomic features in a circular graphical representation, CG View Server used to create a custom circular map (25) (Figure 2).

Both nucleotide and protein BLASTs analysis against the *Xanthomonas* T3SS effector database (http://www.xanthomonas.org) used to identify candidate T3SS protein effector genes. In addition, whole genome sequence submitted to IS Finder website (26) to identifying Insertion sequences. Clustered regularly interspaced short palindromic repeat (CRISPR) is a family of DNA sequence in bacteria that act as a defense system against foreign genetic elements such as those present in plasmids and phages. To determine the characteristics and location of CRISPRs in XccA* NIGEB-88 genome, we used the CRISPRFinder (27).

Some bioinformatics software such as Geneious 11.0.4 (28) and online servers used to perform additional gene prediction analysis and manual annotation of other genes involved in virulence and host range specificity.

Nucleotide sequence accession number

The draft genome sequence of strain XccA* NIGEB-88 has been deposited at DDBJ/ EMBL/GenBank under the accession number LJGA00000000. The version described in this paper is the first version LJGA01000000.

Results

Sequencing and general features of the genome

The draft genome sequence XccA* NIGEB-88 strain comprised 5,236,943 bp and two circular plasmids named, pXCC_28 (28,978 bp) and pXCC_55 (55,677 bp). Annotation of the XccA* NIGEB-88 genome sequences resulted in the identification of 4,496 CDSs in 473 subsystems and 55 structural RNAs (2,287 genes on plus and 2,264 genes on minus strand).

Function of 1,440 CDSs, as hypothetical proteins are unknown, but for 3,056 other CDSs, a known function is recorded. The coding density of the genome is 84.03% and similar to the most of xanthomonads. CDSs that could be assigned to one or more COG functional classes, genes with Pfam domains and other XccA* NIGEB-88 genome statistics are shown in Table 1.

We found one set of 5S–16S-23S rRNA, organized in an operon cluster located in contig number 4 from 139,364 bp to 144,525 bp. In addition, 52 putative tRNA genes for the 20 standard amino acids also identified. In bacteria, replication origin includes several DnaA boxes and GC disparity (GC skew). So according to the results of cumulative GC skew, the dnaA gene location, and distribution of DnaA boxes the location of the bacterial DNA replication origin or terminus identified on contig-01 (29).

One of the valuable tools for genome analysis and annotation is the computational prediction of the subcellular localization of proteins. Presence of a signal peptide or membrane-spanning alpha helices are some of the features

Table 1. *Xanthomonas citri* subsp. *citri* strain A* NIGEB-88 genome statistics.

Attribute	Value	% of total [a]
Genome size (bp)	5,321,598	100
DNA coding (bp)	4,471,872	84.03
DNA G + C (bp)	3,443,400	64.70
Total operons	1018	
Total genes	4551	100
Protein coding genes	4496	98.79
RNA genes	55	1.21
Pseudo Genes	140	3.08
Genes for cytoplasmic proteins	1338	29.76
Genes for cytoplasmic membrane proteins	765	17.02
Genes for Gram-negative motifs	30	0.67
Genes for outer membrane proteins	154	3.42
Genes for Gram-negative outer membrane protein motifs	5	0.11
Genes for periplasmic proteins	55	1.22
Genes with signal peptides	857	19.06
Genes with transmembrane helices	1181	26.26
Genes for extracellular proteins	82	1.82
Genes with Pfam domains	3713	82.58
Genes assigned to COGs	3905	86.85
CRISPR repeats	2	-

[a] The total of RNA and protein coding genes is based on the total number of coding genes; the total of other protein feature is based on the total number of protein coding genes and total draft genome size.

present within the protein's primary structure that can influence the subcellular localization of proteins.

For bacterial pathogens, the prediction of proteins on the cell surface is of particular interest due to the the potential of such proteins to be primary drug or vaccine targets. We used PSORTb 3.0.2 which assigns a probable localization site to a protein given an amino acid sequence to find genes with signal peptides, genes with trans-membrane helices, genes for cytoplasmic proteins etc. (Table 2).

One of the major mechanisms for genome plasticity that led to diversification and speciation of the bacteria is horizontal gene transfer (HGT) or the transmission of DNA between different genomes (30).

To identify potential horizontally transferred genes in the XccA* NIGEB-88 genome, we look for regions with atypical G+C content using GC-Profile. The results of this analysis showed that the G+C content of XccA* NIGEB -88 genome ranged from 53% to 69.48% with an average of 64.7%.

There are at least seven regions of low GC content with different size from approximately 3 kb to 60 kb were determined as genomic islands (Figure 3). It was noteworthy that one of the genomic islands encodes for virB6, virB11, and virB9 proteins that are part of type IV secretion system. Furthermore, there is a large number of transposase and integrase genes in these regions, which is a specific feature of the genomic islands.

The genome of XccA* NIGEB-88 includes 61 insertion sequence (IS) elements that most of them originate from the genome of other strains of *Xanthomonas* genus. We found 28 IS elements from IS3 family (IS 51 and IS407), 21 IS elements from IS5 and five elements from

Table 2. Number of genes associated with general COG functional categories.

Code	Value	%age†	Description
A	0	0.0	RNA processing and modification
B	2	0.04	Chromatin structure and dynamics
C	194	4.31	Energy production and conversion
D	37	0.82	Cell cycle control, Cell division, chromosome partitioning
E	219	4.87	Amino acid transport and metabolism
F	79	1.76	Nucleotide transport and metabolism
G	221	4.91	Carbohydrate transport and metabolism
H	108	2.40	Coenzyme transport and metabolism
I	123	2.74	Lipid transport and metabolism
J	171	3.80	Translation, ribosomal structure and biogenesis
K	208	4.62	Transcription
L	238	5.29	Replication, recombination and repair
M	261	5.81	Cell wall/membrane biogenesis
N	57	1.27	Cell motility
O	164	3.64	Posttranslational modification, protein turnover, chaperones
P	225	5.00	Inorganic ion transport and metabolism
Q	51	1.13	Secondary metabolites biosynthesis, transport and catabolism
R	5	0.11	General function prediction only
S	1142	25.40	Function unknown
T	226	5.02	Signal transduction mechanisms
U	113	2.51	Intracellular trafficking and secretion
V	61	1.36	Defense mechanisms
-	591	13.15	Not in COGs

† The total is based on the total number of protein coding genes in the draft genome.

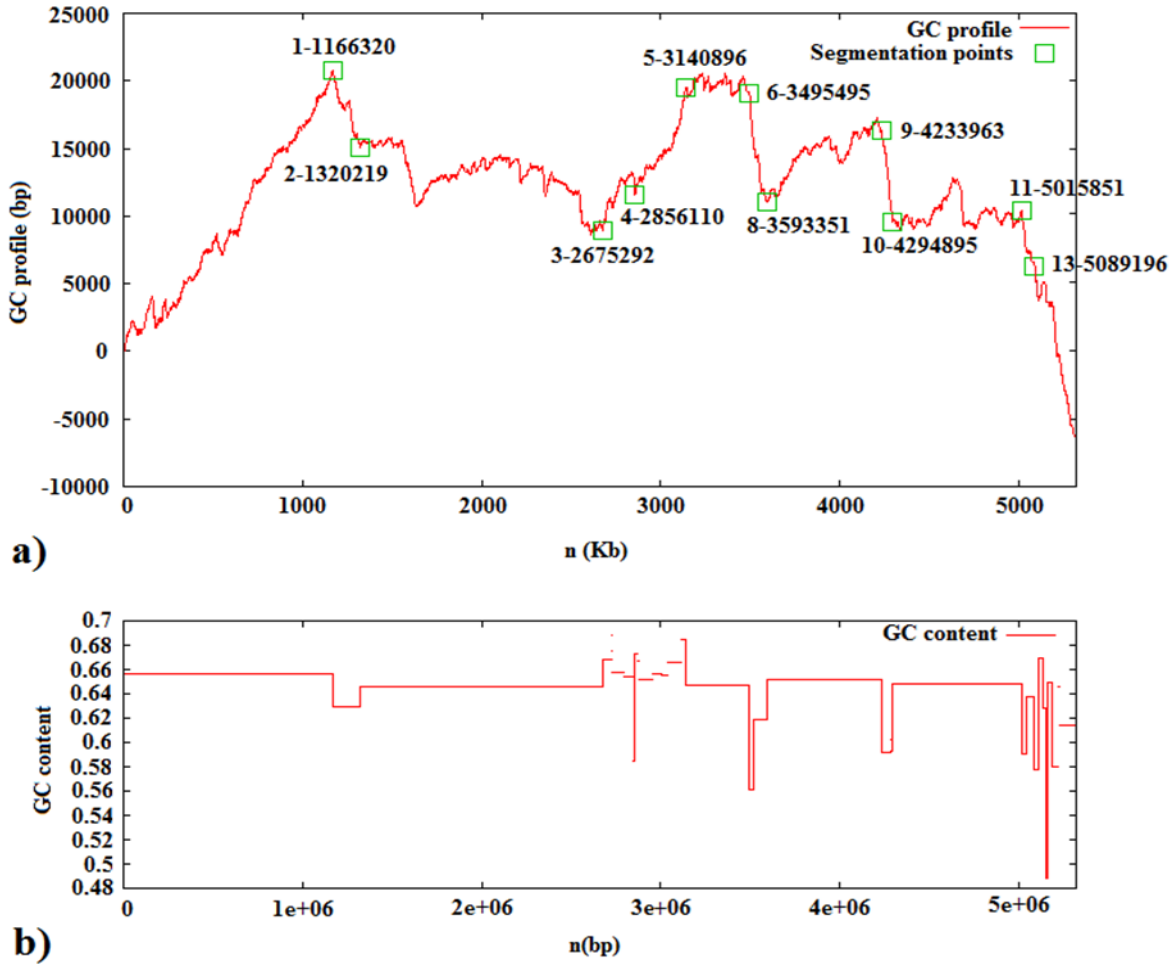

Figure 3. GC profile and GC content of *Xcc* strain *A** NIGEB-88 genome. a) The negative cumulative GC profile for the genome marked with the segmentation points. The segmentation points are obtained at t_0= 50. b) Plot representing the distributions of G+C content along the XACM genome. It shows at least seven regions of low GC content, which are recognized as genomic island (24).

Tn3 family. In order to determine gene loss, duplication and genome rearrangement between two Iranian strains, we used MAUVE in progressive mode to align whole genome sequences of XccA* NIGEB-88 and XccA* NIGEB-386.

Although the genome of these two strains has some rearrangements relative each other, zooming in on the alignment showed the most differences between these two strains are in the transposase coding genes and hypothetical proteins, but there is no significant difference between the pathogenicity and virulence genes (Figure 4).

Phylogenetic analysis

The phylogenetic tree of the MLSA based on nine concatenated housekeeping genes (Figure 5) showed that XccA* NIGEB-88 is closely related to other Iranian strain, XccA* NIGEB-386, and they form a clade. On the other hand, Xcaw12879 and XccA306 which form a clade, are closest strains to these Iranian A* strains. The three pathotypes of Xcc share a

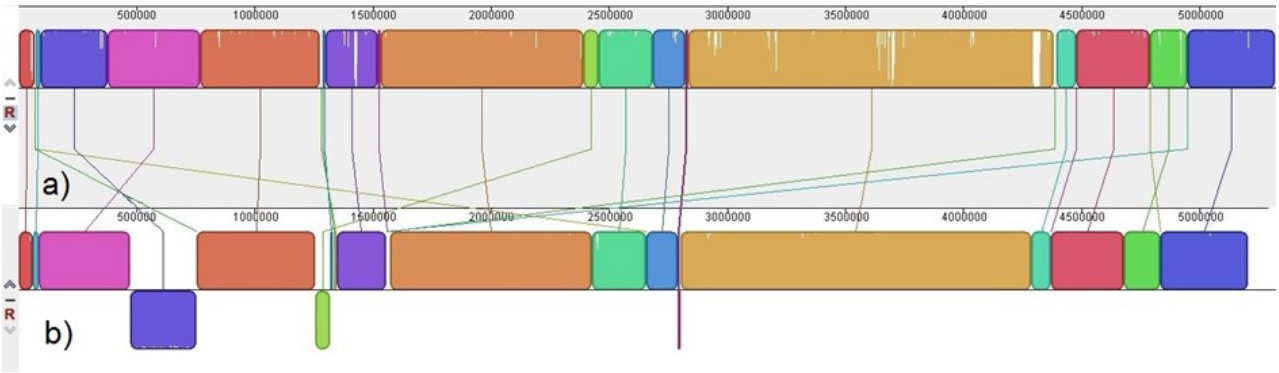

Figure 4. MAUVE alignment of the genome sequences of *Xanthomonas citri* subsp. *citri* strain A* NIGEB-88 (a) and *X. citri* subsp. *citri* strain A* NIGEB-386 (b). The same *color* boxes are *conserved and highly related* genomic *regions* and areas outside blocks (colorless areas) lack detectable homology among the three genomes. The colored lines indicate translocations of the homologous regions and identical colored blocks on opposite sides of the centerline indicate inversion.

close relationship with the other causal agents of citrus bacterial canker, *X. axonopodis* Xac29-1; bacterial black spot, *X. citri* pv. mangiferaeindicae LMG 941 and bacterial leaf blight, *X. axonopodis* pv. punicae LMG 859; which affect citrus, mango, and pomegranates respectively. This phylogenetic relationship is in agreement with the assumption that all of these pathogenic strains originated from the same ancestor in India (31) and evolved to adapt to different hosts. XccA* NIGEB-88 strain contains two plasmids. The pXCC_28 contained 30 CDSs including type IV secretion system components.

The sequence of this plasmid has about 91.4% and 45.6% identity to pXCC48 and pXCC55 of XccA* NIGEB-386, respectively. There are 65 CDSs on pXCC_55. The sequence of this plasmid has about 49.9% and 62.0% identity to pXCC48 and pXCC55 of XccA* NIGEB-386, respectively. The pXCC55 contains one copy of the pathogenicity and virulence gene pthA*, which is capable of conferring the ability to cause canker-like symptoms (32).

XccA* NIGEB-88 has two clustered regularly interspaced short palindromic repeats (CRISPRs). The CRISPR1 (DR length: 23 bp; Number of spacers: 5) located on contig 01 and is different with two CRISPRs of XccA* NIGEB-386. The CRISPR2 with the total length of 892 bp (DR length: 31 bp; Number of spacers: 13) located on contig 04 and is identical to CRISPR1 of XccA* NIGEB-386.

Comparison of candidate genes related to the virulence and host specificity

XccA* NIGEB-88, such as other Xcc strains has six specialized secretion systems, with a specific physiological process. All components of type I, II, IV, V and VI secretion systems are the same in two analyzed Iranian Xcc strains. Type III secretion system (T3SS) used to secret virulence type-III effector proteins into the host cells by many plant and animal pathogens, including Xanthomonads. Without the intact and complete T3SS, the pathogenic bacteria cannot defeat basal defenses, grow in plants, produce disease lesions in hosts, or elicit the hypersensitive response (HR) in nonhosts.

This secretion system can be further subdivided

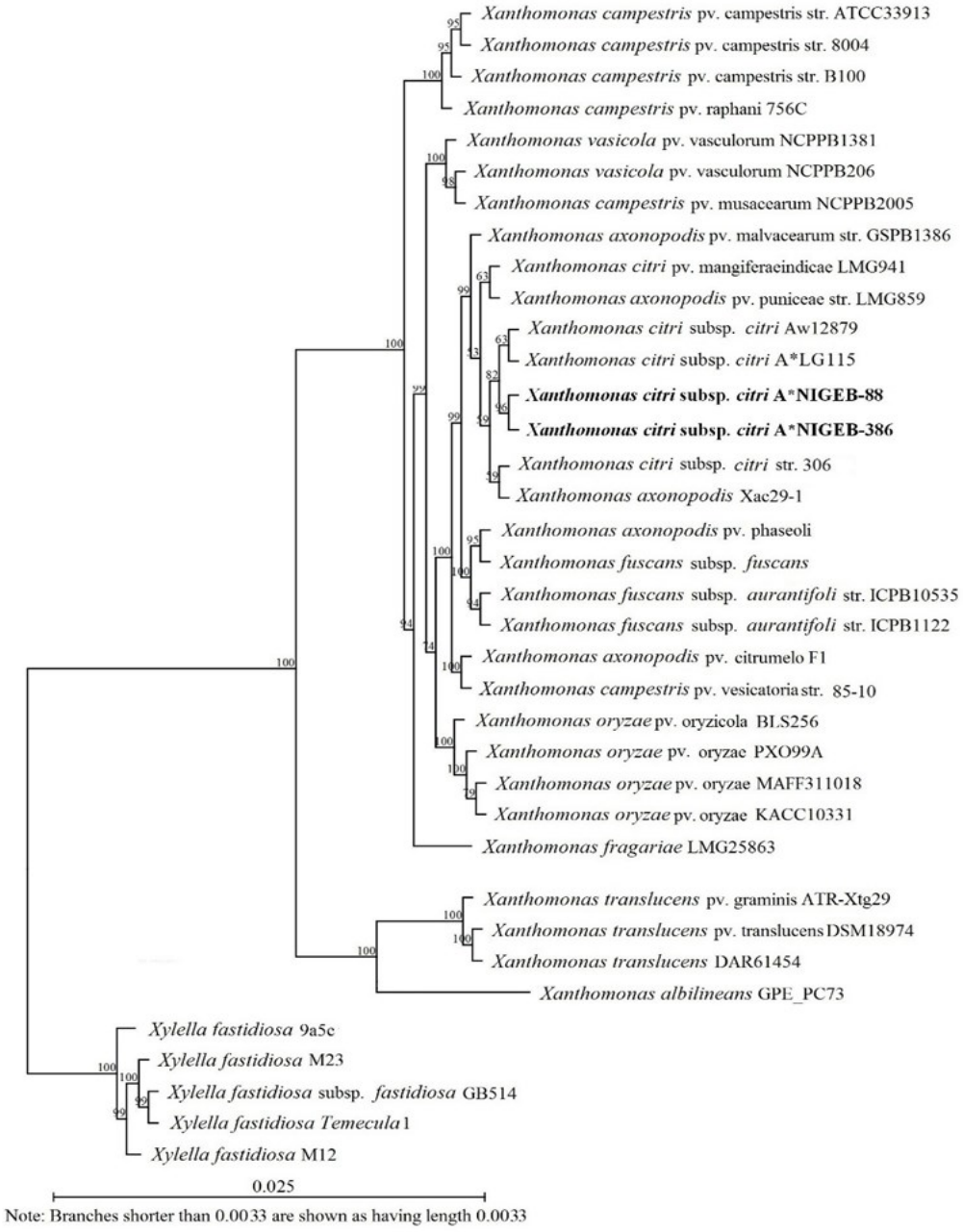

Figure 5. Maximum likelihood phylogenetic tree of the genome of *Xanthomonas citri* subsp. *citri* strain *A** NIGEB-88 showing the relationship to other *Xanthomonads* species. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) shown next to the branches.

into the non-flagellar T3SS (T3aS) and the flagellar T3SS (T3bS) subgroups.

The T3aS is involved in the formation of the injectisome or Hrp (hypersensitive response and pathogenicity) pilus whereas the subfamily b of the T3SS is responsible for assembly of the flagellum (33). In XccA* NIGEB-88 such as other Iranian strain, XccA* NIGEB-386 (15), there are twenty-eight effector genes. AvrGf1 effector protein was previously introduced by Jalan et al. (34) as one of the factors that contribute to the host range

limitation of Xcc Aw strains were not found in XccA* NIGEB-88.

To validate our results, we checked the sequence of avrGf1 in raw data reads sequence. Based on the MLVA-31 results of Provost-Goodarzi et al. (12), the T3SS effector xopC1was only detected from Iranian strains assigned to cluster 4.1 but not in those assigned to cluster 4.4 (i.e. strains from Sistan-Baluchistan). However, our results showed this gene does not exist in XccA* NIGEB-88 that assigned to cluster 4.1. The effector protein PthA identified to function for pathogenicity and utilized by all the three-pathotype strains of Xcc to modulate host transcription to promote citrus canker. The PthA protein belongs to the AvrBs3/PthA family and carries a repeat domain contain variable numbers of near-identical tandem repeats of 34 amino acids (32) that mediates protein-protein and protein-DNA interactions. In draft genome sequences of XccA* NIGEB-88, only one copy of pthA* was found on plasmid pXCC_55.

This PthA* such as another functional homolog of pthA in Xcc strains has 17.5 tandem repeats and exhibit 98.1% sequence identity with PthA*2 of XccA* NIGEB-386 at the amino acid level (Figure 6).

It is noteworthy that apart from the pthA*, other T3SS effectors genes have the same nucleotide sequences and length in two Iranian strains assigned to clusters 4.1 and cluster 4.4 (i.e. strains from Sistan-Baluchistan).

Lipopolysaccharide (LPS)

Lipopolysaccharides are the key component of the outer membrane present that in almost all of the Gram-negative bacteria and act as extremely strong stimulators of innate or natural immunity in diverse eukaryotic species. The LPS gene clusters in XccA* NIGEB-88 contain 19 genes that all of them are homologous with XccA* NIGEB-386.

The extracellular polysaccharides and the opsX (for outer-membrane polysaccharide) which plays roles in the production of EPS and LPS, flagellar gene cluster, the regulation of pathogenicity factors (rpf) genes that control the synthesis of the DSF molecule and signal transduction and plays a major role in quorum sensing, two potential secretion genes belonged to the RTX toxin family, hlyB and hlyD that are known to be pore-forming cytotoxins and acting as virulence factors are other strain-specific genes that might contribute to the difference in virulence and host specificity of Xcc pathotypes. Our results showed the presence of these genes in XccA* NIGEB-88 that have identical length and sequence with XccA* NIGEB-386 strain genes.

Discussion

Based on the MLVA-31 and MLVA-14 results, Provost-Goodarzi et al. (12) highlighted the

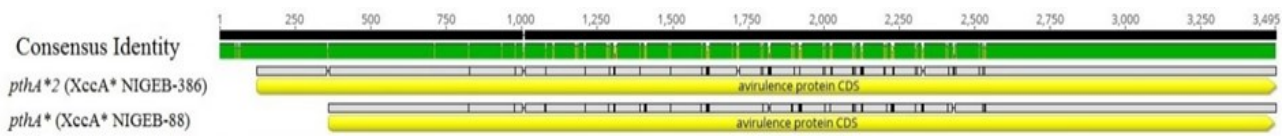

Figure 6. Comparison of the length and sequence of functional homolog of *pthA** in *X. citri* subsp. *citri* strain A* NIGEB-88 and *X. citri* subsp. *citri* strain A* NIGEB-386. The image was generated by *Geneious* software (28).

occurrence in Iran of two different groups of Xcc strains, which appear geographically isolated.

In this study, we focus on the differences in candidate genes and factors associated with pathogenicity and host range of two Iranian Xcc strains assigned to cluster 4.1 (XccA* NIGEB-88) and cluster 4.4 (XccA* NIGEB-386). In addition, data presented here shows the phylogenetic relationship of the Iranian strains with other strains and species of *Xanthomonas*.

In general, the results obtained in our study confirmed the finding of Gordon et al. (6) and Provost-Goodarzi et al. (12) that Iranian strains showing an intermediate pathogenicity phenotype and are genetically related to the host-restricted pathotype A* but not to the generalist pathotype A. The data shows close similarity of general features of XccA* NIGEB -88, such as genome size, the number of plasmids, GC content, number of CDS and structural RNA with other Iranian strain as well as other *Xanthomonas citri* subsp. citri strains.

Of course, there are some minors of genetic variation between two Iranian XccA* strains. All components of type I, II, IV, V and VI secretion systems are the same in two analyzed Iranian Xcc strains, and it does not seem to have a role in the difference of pathogenicity between them. According to Provost-Goodarzi et al. (12), the two groups of Iranian strains were different, not only in pathogenicity but also in their type III effector repertoire.

However, the only difference we have seen in type III effector repertoires was related to a functional homolog of PthA* and other effector proteins were similar in two strains. PthA effectors are transcription activator like (TAL) effectors and reprogram host cells by binding, specifically to the promoters. XccA* NIGEB-88 strain has one PthA* with 17.5 tandem repeats. Although PthA* effectors in both Iranian strains have 17th amino acids repeat that might be important for the pathogenicity on citrus plants, the length and sequence of amino acids in both proteins are not exactly the same. In general, the functional homolog of PthA* in XccA* NIGEB-88 has 97.9% similarity to XccA* NIGEB-386, while this similarity with the PthA of Xcaw12879 and XccA306 is 96.9% and 97.2% respectively.

Given that the little differences between the sequence of this effector may result in recognition of different target genes in host plant or differences in strength of induction of plant genes, it is necessary to determine the effects of PthA* in virulence and difference pathogenicity by more functional analysis at the laboratory. The AvrGf1 is another T3SS effector protein that contributes to host range limitation of Xcaw12879 (34), and the finding of this study showed is absent in two Iranian strains. The absence of this effector in the narrow host range A* strains might indicate that this effector is not a specific host range determining factor and that other factors are necessary to explain differing host ranges between these strains. Moreover, in contrast with the results of Provost-Goodarzi et al. (12), the xopC1 does not exist in both XccA* NIGEB-88 and XccA* NIGEB-386 strains and does not appear to affect their pathogenicity. Lipopolysaccharide (LPS) of XccA* NIGEB-88 is similar to one in XccA* NIGEB-386 and more related to other host-limited Xcc strain,

Xcaw12879, than wide host range XccA306 strain. CRISPR is a bacterial immunity system that requires a perfect sequence match between the CRISPR cassette spacer and a proto-spacer in invading DNA, such as plasmids and phages, to fight off a foreign genetic element infection. XccA* NIGEB-88 has two CRISPRs, one of them is different with two CRISPRs of XccA* NIGEB-386.

Due to the role of CRISPR in biological control of pathogenic bacteria, it is important to note the difference when choosing an appropriate phage to control different strains of Iranian Xcc strains. This study provided a description of genetic features of two Iranian Xcc strains in the whole genome scale that is a good source of background information and can be used for further studies on the causal agent of citrus bacterial canker in Iran.

Conclusion

In this study, we report a high-quality draft genome sequence of XccA* NIGEB-88, an Iranian strain isolated from an infected orchard in the Hormozgan province, to characterize its genomic and phylogenetic properties. Because the Xcc strains of this province were previously assigned to cluster 4.1 of pathotype A* by Provost-Goodarzi et al. (12), we also compared some of the candidate genes participate in virulence and pathogenicity in this strain and other Iranian strain, XccA* NIGEB-386, assigned to cluster 4.4. Although, there is a close similarity in general genomics feature of two Iranian strains, however, some differences between CDS with unknown function in two strains may require further study due to their potential role in pathogenicity. In the context of known virulence genes and pathogenic factors, there is no significant difference between the two Iranian strains. However, there are some minor genetic variations between the functional homolog of PthA* in two strains that may play a role on the strength of induction of plant genes and affect the virulence of two Iranian strains. Furthermore, phylogenetic analysis based on the protein sequences of nine housekeeping genes showed, XccA* NIGEB-88 is closely related to other Iranian A* strain, XccA* NIGEB-386 than other narrow host range strain from pathotype Aw, Xcaw12879, and this is in agreement with phylogeny inferred from the whole genome alignment by Gordon et al. (6). In general, a high degree of sequence similarity between the virulence genes of two Iranian Xcc strains, suggesting the presence of other unknown genes and factors that probably contribute to the the difference of pathogenicity strength between the two bacteria.

Acknowledgment

This study was supported by National Institute of Genetic Engineering and Biotechnology, Grant Project M-406.

References

1. Food and Agricultural Organization of The United Nations. http://faostat3.fao.org/browse/Q/QC/E. According to the latest update in 2015

2. Ference CM, Gochez AM, Behlau F, Wang N, Graham JH, Jones JB. Recent advances in the understanding of *Xanthomonas citri* ssp. citri pathogenesis and citrus canker disease management. Mol Plant Pathol. 2017. doi: 10.1111/mpp.12638.

3. Rodriguez-R LM, Grajales A, Arrieta-Ortiz ML, Salazar C, Restrepo S, Bernal A. Genomes-based phylogeny of the genus *Xanthomonas*. BMC Microbiol. 2012; 12: 43.

4. Sun X, Stall RE, Jones JB, Cubero J, Gottwald TR, Graham JH, Dixon WH, Schubert TS, Chaloux PH, Stromberg VK, Lacy GH and Sutton BD. Detection and characterization of a new strain of citrus canker bacteria from key/Mexican lime and alemow in south Florida. Plant Dis. 2004; 88: 1179-1188.

5. Verniere C, Hartung JS, Pruvost OP, Civerolo EL, Alvarez AM, Maestri P and Luisetti J. Characterization of phenotypically distinct strains of *Xanthomonas axonopodis* pv. *citri* from Southwest Asia. Eur J Plant Pathol. 1998; 104: 477-487.

6. Gordon JL, Lefeuvre P, Escalon A, Barbe V, Cruveiller S, Gagnevin L, Pruvost O. Comparative genomics of 43 strains of *Xanthomonas citri* pv. *citri* reveals the evolutionary events giving rise to pathotypes with different host ranges. BMC Genomics. 2015; 16: 1098.

7. Piazza A, Zimaro T, Garavaglia BS, Ficarra FA, Thomas L, Marondedze C, Feil R, Lunn JE, Gehring C, Ottado J, Gottig N. The dual nature of trehalose in citrus canker disease: a virulence factor for *Xanthomonas citri* subsp. *citri* and a trigger for plant defence responses. J Exp Bot. 2015; 66(9): 2795-2811.

8. Alizadeh A, Rahimian H. Citrus canker in Kerman province. Iran J Plant Pathol. 1990; 26: 42.

9. Khodakaramian G, Rahimian H, Mohamadi M, Allameh A. Phenotypic characteristics, host range and distribution of the strains of *Xanthomonas axonopodis* inducing citrus canker in southern Iran. Iran J Plant Pathol. 1999; 35: 40-43.

10. Mohammadi M, Mirzaee MR, Rahimian H. Physiological and biochemical characteristics of Iranian strains of *Xanthomonas axonopodis* pv. *citri*, the causal agent of citrus bacterial canker disease. J Phytopath. 2001; 149: 65-75.

11. Khodakaramian G, Swings J. Genetic diversity and pathogenicity of *Xanthomonas axonopodis* strains inducing citrus canker disease in Iran and South Korea. Indian J Microbiol. 2011; 51: 194-199.

12. Pruvost O, Goodarzi T, Boyer K, Soltaninejad H, Escalon A, Alavi SM, Javegny S, Boyer C, Cottyn B, Gagnevin L and Vernière C. Genetic structure analysis of strains causing citrus canker in Iran reveals the presence of two different lineages of *Xanthomonas citri* pv. *citri* pathotype A*. Plant Pathol. 2015; 64(4): 776-784.

13. Escalon A, Javegny S, Vernière C, Noël LD, Vital K, Poussier S, Hajri A, Boureau T, Pruvost O, Arlat M, Gagnevin L. Variations in type III effector repertoires, pathological phenotypes and host range of *Xanthomonas citri* pv. *citri* pathotypes. Mol Plant Pathol. 2013; 14(5): 483-496.

14. Jalali A, Alavi SM, Sangtarash MH. Comparative genomic analysis of wide and narrow host

range strains of *Xanthomonas citri* subsp. *citri*, showing differences in the genetic content of their pathogenicity and virulence factors. Australasian Plant Pathol. 2017; 46(1): 49-61.

15. Desjardins P, Conklin D. NanoDrop microvolume quantitation of nucleic acids. J Vis Exp. 2010; 45: 2565.

16. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 2008; 18: 821-829.

17. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res. 2016; 44(14): gkw569.

18. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, Dao P, Sahinalp SC, Ester M, Foster LJ, Brinkman FSL. PSORTb 3.0: Improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics. 2010; 26(13): 1608-1615.

19. Powell S, Forslund K, Szklarczyk D, Trachana K, Roth A, Huerta-Cepas J, Gabaldón T, Rattei T, Creevey C, Kuhn M, Jensen LJ, von Mering C, Bork P. eggNOG v4.0: nested orthology inference across 3686 organisms. Nucleic Acids Res. 2014; 42(D1): D231-D239.

20. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer ELL, Tate J, Punta M. The Pfam protein families database. Nucleic Acids Res Database Issue. 2014; 42: D222-D230.

21. Glaeser SP, Kämpfer P. Multilocus sequence analysis (MLSA) in prokaryotic taxonomy. Syst Appl Microbiol. 2015; 38(4): 237-245.

22. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011; 11(7): 539.

23. Darling AE, Mau B, Perna NT. Progressive Mauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One. 2010; 5(6): e11147.

24. Gao F, Zhang CT. GC-Profile: a web-based tool for visualizing and analyzing the variation of GC content in genomic sequences. Nucleic Acids Res. 2006; 34: W686-W691

25. Grant JR, Stothard P. The CGView Server: a comparative genomics tool for circular genomes. Nucleic Acids Res. 2008; 36: 181-184.

26. Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Research. 2006; 34: D32-D36.

27. Grissa I, Vergnaud G, Pourcel C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucl Acids Res. 2007; 35: W52-W57.

28. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Mentjies P, Drummond A. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012; 28(12): 1647-1649.

29. Frank AC, Lobry JR. Oriloc: prediction of replication boundaries in unannotated bacterial chromosomes. Bioinformatics. 2000; 16: 560-561.

30. Ochman H, Lawrence JG, Groisman EA. Lateral gene transfer and the nature of bacterial innovation. Nature. 2000; 405: 299-304.

31. Midha S, Patil PB. Genomic insights into the evolutionary origin of *Xanthomonas axonopodis* pv. *citri* and its ecological relatives. Appl Environ Microbiol. 2014; 80(20): 6266-6279.

32. Al-Saadi A, Reddy JD, Duan YP, Brunings AM, Yuan Q, Gabriel DW. All five host-range variants of *Xanthomonas citri* carry one *pthA* homolog with 17.5 repeats that determines pathogenicity on citrus, but none determine host-range variation. Mol Plant Microbe Interact. 2007; 20: 934-943.

33. Desvaux M, Hebraud M, Henderson IR, Pallen MJ. Type III secretion: what's in a name? Trends Microbiol. 2006; 14(4): 157-160.

34. Jalan N, Kumar D, Andrade MO, Yu F, Jones JB, Graham JH, White FF, Setubal JC, Wang N. Comparative genomic and transcriptome analyses of pathotypes of *Xanthomonas citri* subsp. *citri* provide insights into mechanisms of bacterial virulence and host range. BMC Genomics. 2013; 14: 551.