EXTENDED GENOME REPORT

The genome of the cotton bacterial blight pathogen Xanthomonas citri pv. malvacearum strain MSCT1

Kurt C. Showmaker1,2*, Mark A. Arick II1, Chuan-Yu Hsu1, Brigitte E. Martin3, Xiaoqiang Wang2, Jiayuan Jia2, Martin J. Wubben4, Robert L. Nichols5, Tom W. Allen6, Daniel G. Peterson1,7 and Shi-En Lu2

Abstract

Xanthomonas citri pv. malvacearum is a major pathogen of cotton, Gossypium hirsutum L. In this study we report the complete genome of the X. citri pv. malvacearum strain MSCT1 assembled from long read DNA sequencing technology. The MSCT1 genome is the first X. citri pv. malvacearum genome with complete coding regions for X. citri pv. malvacearum transcriptional activator-like effectors. In addition functional and structural annotations are presented in this study that will provide a foundation for future pathogenesis studies with MSCT1.

Keywords: Xanthomonas citri pv. malvacearum, Bacterial blight, TAL effectors, Cotton, Long read DNA sequencing

Introduction

Xanthomonas citri pv. malvacearum is the causal agent of bacterial blight of cotton (Gossypium hirsutum L.). Xanthomonas citri pv. malvacearum infects plant tissues and organs of cotton during all stages of development beginning with the seedling stage [1]. Typical disease symptoms caused by X. citri pv. malvacearum include cotyledon/seedling blight, angular leaf spot, systemic vein blight, black arm (of petioles and main stems), boll shedding, and internal boll rot [1]. Histology studies reported that the host cotton plant cells begin to degenerate 3 days post-infection [2]. Over the 3 day period the degradation of host cells begins by; first, the host tissue appearing to loosen, then the granal and stromal membranes of the chloroplasts disappear, followed by the degeneration of the chloroplast and other organelles [2, 3]. At 6 days post-infection, cellular degeneration along with the production of a hydrophilic extracellular polymeric substance by the bacterium, causes water to accumulate in the infected tissues forming lesions known as “water soaked spots”, a classical plant pathogen-associated symptom [2–4].

Resistance to X. citri pv. malvacearum has been identified in cotton, as well as additional Gossypium species. Currently, most lines resistant to X. citri pv. malvacearum exist in G. hirsutum cultivars since breeding for X. citri pv. malvacearum resistance has been ongoing since 1939 [5] and continues today as G. hirsutum cultivars and germplasm releases are screened for X. citri pv. malvacearum resistance [6–8]. At least 18 genes participate in resistance to X. citri pv. malvacearum [1, 9]. The ability of the X. citri pv. malvacearum strains to escape specific resistance genes resulted in a classification scheme of races. To date, 22 races have been reported and assigned numerical names (i.e. 1 to 22) [9]. Most races are geographically distinct. Of note, bacterial blight in the U.S. is predominantly caused by race 18. Genetic resistance within cotton cultivars is generally attributed to a certain race or multiple races of X. citri pv. malvacearum. The ability of G. hirsutum to mount a defense response to X. citri pv. malvacearum is, at least in some cases, dependent upon the transcription activator-like effector avrBs3/phA gene family in X. citri pv. malvacearum indicating the presence of a gene-for-gene relationship in X. citri pv. malvacearum-G. hirsutum interactions [9, 10]. With the ever increasing understanding of the importance of TAL effectors in
pathogenesis [11–13], the objective of this study was to generate the first genome sequence for a *X. citri* pv. *malvacearum* strain that contains the TAL effector complement to serve as a foundation for a better understanding of the *X. citri* pv. *malvacearum*-G. *hirsutum* interaction.

To date, four draft genomes of *Xanthomonas citri* pv. *malvacearum* have been published. However, all sequenced *X. citri* pv. *malvacearum* isolates were obtained from outside the United States [14, 15]. The diversity of the four previously reported draft genomes includes two race 18 isolates, one race 20 isolate, and a highly virulent strain. The project described here was undertaken to provide the first *X. citri* pv. *malvacearum* genome sequence from the Mid-South region of the United States, a major production area of upland cotton. The isolate, MSCT1, was isolated during the 2011 outbreak of *X. citri* pv. *malvacearum* in the Mississippi Delta (i.e. Mississippi river’s alluvial plain). This outbreak resulted in the greatest estimated *X. citri* pv. *malvacearum*-based losses (52,000 bales) in Arkansas and Mississippi as reported by the National Cotton Council Disease Database [16]. This study was undertaken to generate a genome sequence for the *X. citri* pv. *malvacearum* strain MSCT1 to identify protein candidates that may be involved in the pathogenesis of bacteria bight of cotton. The genome sequence will also serve as a template for which further studies of genetic diversity of *X. citri* pv. *malvacearum* in the United States can be conducted.

### Organism information

#### Classification and features

*Xanthomonas citri* pv. *malvacearum* has gone through a series of name changes over time as additional information has been learned about its biology and genetics. In chronological order, *X. citri* pv. *malvacearum* has previously been classified as *Pseudomonas malvacearum*, *Bacterium malvacearum*, and *Xanthomonas malvacearum* [9]. In 2009, Ah-You et al. assigned the *X. citri* pv. *malvacearum* moniker [9, 17]. *Xanthomonas citri* pv. *malvacearum* is a motile, Gram-negative, rod-shaped bacterium that produces yellow, copiously mucoid, wet, shining growth on 2% w/v peptone-sugar agar [1]. *Xanthomonas citri* pv. *malvacearum*, like other *Xanthomonas* species (xanthomonads), produces the heteropolysaccharide xanthan [4]. Additional characteristics of *X. citri* pv. *malvacearum* are provided in Table 1.

For specimen isolation, cotton leaves with the typical blight symptoms (Fig. 1) were collected from a field located north of Yazoo City, Mississippi in Yazoo County, during the 2011 growing season. Strain MSCT1 was isolated using a routine method for foliar bacterial pathogens. In brief, the disease lesions were cut into small pieces (3 × 3 mm) from the junction of diseased and healthy tissues. The cut pieces were transferred into a sterile 1.5 ml microcentrifuge tube and surface-sterilized using 10% sodium hypochlorite (bleach; Clorox) for 1 min. The sterilized tissues were washed twice using sterile water, and then stabbed with a sterile lab needle in 200 μl of sterile water. A full loop of the resulting bacterial suspension was streaked on nutrient broth-yeast extract plates [18]. The streaked nutrient broth-

### Table 1 Classification and general features of *Xanthomonas citri* pv. *malvacearum* strain: MSCT1 [75]

| MIGS ID | Property | Term | Evidence code |
|---------|----------|------|---------------|
| MIGS-6  | Habitat  | Plant-associated | TAS [1] |
| MIGS-6.3| Salinity | Not reported | TAS [1] |
| MIGS-22 | Oxygen requirement | Not reported | TAS [1] |
| MIGS-15 | Biotic relationship | Parasitic | TAS [1] |
| MIGS-14 | Pathogenicity | Pathogenic | IDA |
| MIGS-4  | Geographic location | Mississippi, USA | IDA |
| MIGS-5  | Sample collection | 2011 | IDA |
| MIGS-4.1| Latitude | Not Reported | TAS [1] |
| MIGS-4.2| Longitude | Not Reported | TAS [1] |
| MIGS-4.4| Altitude | Not Reported | TAS [1] |

*Evidence codes - IDA inferred from direct assay, TAS traceable author statement (i.e., a direct report exists in the literature), NAS non-traceable author statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project [82]*
yeast extract plates were incubated at 20 °C for 2 days under ambient laboratory temperatures and a 16:8 day:night photoperiod. Single colonies of the resulting bacterium were isolated in a sterilized loop and streaked onto fresh NBY plates for purification. The pathogenicity of MSCT1 to cause bacterial blight of cotton was confirmed by fulfilling Koch’s Postulates. Briefly, cotton seedlings (cotton cultivar PHY499WRF) were grown in the greenhouse until they reached the three-leaf growth stage. A vacuum system (20″ psi for 10 s) was used to inoculate the seedling leaves with a suspension of MSCT1 (OD 0.3 at 420 nm) suspended in sterile phosphate buffer (0.01 M; pH 7.0). After 10 days the characteristic symptoms of bacterial blight were observed on the inoculated leaf tissues. The X. citri pv. malvacearum strain MSCT1 that is described in this manuscript was deposited in the USDA Agricultural Research Service Culture Collection under deposition number NRRL B-65440. The isolate MSCT1 was confirmed to be Xanthomonas citri pv. malvacearum based on the 16S rRNA sequence analysis, as described previously [19]. Multilocus sequence typing was used to construct a phylogenetic tree for Xanthomonas strains based upon three genes from the MLST described by Ah-You et al. 2009 [17], and included; atpD coding ATP synthase β chain, dnaK coding heat shock protein 70, and gyrB coding the gyrase subunit β (Fig. 2). A transmission electron micrograph of MSCT1 was generated by the Mississippi State University’s Institute for Imaging & Analytical Technologies (Fig. 3).

**Genome sequencing information**

**Genome project history**
The MSCT1 sequencing project arose from the 2011 outbreak of bacterial blight in the cotton growing regions of the Mississippi Delta. Following MSCT1 isolation, additional testing determined that MSCT1 was capable of producing disease symptoms on several cultivars of upland cotton commonly grown in the Mid-South. Preliminary bioinformatics investigations determined X. citri pv. malvacearum assemblies, generated from short reads, lacked detectable TAL effectors in their genomes, although TAL effectors have been previously described as occurring in X. citri pv. malvacearum [20–22]. To better understand the pathology of X. citri pv. malvacearum, and more specifically of isolate MSCT1, we conducted a long read genome sequencing project to identify MSCT1’s effector complement, including the TAL effectors that do not assemble well with short read DNA sequencing technology. The resultant complete genome sequence has been deposited in the NCBI genome database under genome assembly accession GCF_001719155.1.

**Growth conditions and genomic DNA preparation**
An MSCT1 colony was isolated from a LB plate (pectone 10 g/L, yeast extract 5 g/L sodium chloride 10 g/ L, agarose 15 g/L) and used to inoculate 1.5 ml of LB medium (pectone 10 g/L, yeast extract 5 g/L sodium chloride 10 g/ L) in a sterile, plastic culture tube. The culture tube was placed at 25 °C with 200 rpm orbital shaking overnight. The resulting bacterial culture was pelleted by centrifugation at 5000 rpm for 10 min. The pellet was washed twice to remove LB medium; each wash consisted of resuspending the pellet in 1 ml of phosphate buffered saline (PBS; NaCl 8 g/L, KCl 0.2 g/L, Na2HPO4 1.42 g/L, KH2PO4 0.24 g/L), centrifuging the suspension at 5000 rpm for 10 min, and discarding the supernatant. Genomic DNA was isolated using a modified version of the method described in Chen and Kuo 1993 [23]. Briefly, the cell pellet was resuspended in 300 μl of extraction buffer (40 mM Tris-HCl, 1 mM EDTA, 1% w/v SDS, pH 7.8). After adding 50 μl of 10 mg/ml lysozyme (Sigma-Aldrich; St. Louis, MO, USA), the cell suspension was incubated at 37 °C for 30 min with occasional mixing until the cell suspension became clear. The bacterial nucleic acid sample was further purified using a series of phenol, phenol/chloroform, and chloroform extraction steps, then precipitated with two volumes of 100% ethanol. DNA was pelleted by centrifugation at 12,000 rpm for 10 min. After two washes with 70% ethanol (v/v), the nucleic acid pellet was air-dried (approximately 15 min). The pellet was then dissolved in 50 μl of 10 mM Tris buffer (pH 7.5). The bacterial nucleic acid sample was treated

![Fig. 1 Top (a) and bottom (b) of a cotton leaf displaying the bacterial blight disease symptom caused by Xanthomonas citri pv. malvacearum](image)
with 20 μl of 30 mg/ml RNase A (Sigma-Aldrich; St. Louis, MO, USA) at 37 °C for 20 min, followed by phenol/ chloroform and chloroform extraction steps to remove the enzyme. The DNA was precipitated with 100% ethanol and cleaned with 70% ethanol as described above. The air-dried genomic DNA pellet was dissolved in 50 μl of 10 mM Tris buffer (pH 7.5). The resultant DNA was visualized on a 0.8% w/v agarose gel.

**Genome sequencing and assembly**

Two long read technologies, PacBio (Pacific Biosciences of California, Melon Park, CA, USA) and Nanopore (Oxford Nanopore Technologies, Oxford, UK), were used to sequence MSCT1. A 20 kb PacBio library was prepared and sequenced on two P6-C4 SRMT cells at the University of Delaware Sequencing & Genotyping Center (Newark, DE, USA). Additionally, a Nanopore library was prepared and sequenced on a R9 Nanopore flowcell at the Mississippi State Institute for Genomics, Biocomputing, and Biotechnology (Mississippi State, MS, USA). The PacBio and Nanopore reads were assembled with the Canu long read assembler [24]. The resultant contigs from the assembly were aligned against themselves using the Maximum Likelihood, with 100 bootstraps, method based on the Tamura-Nei model [86] with MEGA6 [87] software. Sequence identifiers of each subunit are as reported by Ah-You et al. 2009 [17]. Type (T) and Pathovar Type (PT) strains are noted in superscript.
the molecule while allowing the new cut sight to fall outside a predicted ORF. To ensure the circulation was correct PacBio reads longer than 4000 bp were aligned to the circularized assembly with blasr [27] and manually checked with IGV [28, 29]. For additional error correction, an Illumina PCR-free DNA library with a DNA insert size of 416 bp was prepared at the Institute of Genomics, Biocomputing and Biotechnology (Mississippi State, MS, USA). The Illumina library was paired-end sequenced (2 × 300 bp) using the Illumina MiSeq. The short read pairs were trimmed with Trimmomatic [30] and subsequently used to error correct the Canu assembly with Pilon [31]. After Pilon error correction, the resultant assembly was polished with 20 kb PacBio reads using the Quiver algorithm within the PacBio SMRT Analysis software suite (version 2.3.0.140936, Pacific Biosciences of California). The Minimum Information about a Genome Sequence specification was used to report the MSCT1 genome sequencing and assembly methods (Table 2) [32].

**Genome annotation**

Proteins and noncoding RNAs (including rRNA, tRNA, ncRNA) were predicted with the NCBI Prokaryote Genome Annotation Pipeline [33]. Clusters of Orthologous Groups annotation of the predicted proteins against the COG position-specific scoring matrices downloaded from the NCBI Conserved Domain Database was conducted with RSP-BLAST [34–36]. InterProScan V51.0 was used to add Pfam annotations using the Pfam applet [37]. Signal peptides and transmembrane helices were predicted with SignalP [38] and TMHMM [39], respectively. Clustered regularly-interspaced short palindromic repeats sequences were identified using CRISPR-Finder [40]. Plant inducible promoter sequences in the promoter region (both strands) of genes were identified with the regular expression 'TTCGN [16] TTCG, where N is any nucleotide, as described by Lee et al. 2005 [41–43]. EffectiveDB was used to determine if MSCT1 contains functional T3SS, T4SS, and T6SS secretory systems. EffectiveDB also identified eukaryotic-like domains, potential T3SS, and potential T4SS secreted proteins in the MSCT1 predicted proteome. Additionally, blastp was used to align the proteins of the MSCT1 predicted proteome to the 502 proteins representing 53 effector families of *Xanthomonas* species found in the *Xanthomonas* effector database (*Xanthomonas.org*) [34]. Transcription activator-like effectors and Repeat Variable Diresidues were predicted with AnnoTALE [44]. TALgetter [45] was used to identify the DNA target domain on the *G. hirsutum* line TM – 1 promoterome [46].

**Genome properties**

The MSCT1 long read assembly had a sum length of 5,123,946 bp distributed along one large circular chromosome 5 Mb (Fig. 4) in length and 3 circular plasmids (60, 44, and 15 kb in length) (Table 3). Sequencing depth was 558.48 genome equivalents for the long read sequencing technology and 1820.26 genome equivalents for the Illumina PCR-free DNA library (Table 2). Dot plots determined the MSCT1 chromosome exhibited a high degree of sequence similarity to the circular chromosomes reported in previous *X. citri pv. malvacearum* assemblies (Fig. 5). A total of 4410 genes were predicted for MSCT1 including 4102 protein coding, 95 rRNA, and 213 pseudogenes (Table 4). The NCBI Prokaryotic
Genome Annotation Pipeline added functional annotation to 2843 proteins.

The predominate COG functional classifications were R (general function), E (amino acid transport and metabolism), M (cell wall/membrane biogenesis), and H (coenzyme transport and metabolism), representing 16.31, 11.68, 10.36, and 9.68% of the predicted proteome, respectively (Table 5). InterProScan identified 3302 proteins containing at least one Pfam domain. In total, 3375 proteins contained at least one functional annotation from either the Pfam or COG annotations (Table 4). The rRNA segments were comprised of two copies of each of the 23S, 5S, and 16S rRNA subunits. At least one tRNA for each of the 20 basic amino acids was identified in the 54 predicted tRNA loci. Transmembrane helices prediction identified 911 proteins with at least one predicted transmembrane helix. Signal peptides were identified on 553 proteins; of these, after in silico cleavage of the predicted signal peptide, 23 contained a predicted transmembrane helix leaving 530 proteins that can be secreted from the cell. A single CRISPR sequence with a sequence length of 298 bp was predicted in the genome assembly in the 27,394 to 27,692 bp region of the MSCT1 chromosome. As is common in species of *Xanthomonas* multiple copies of the transposase coding genes were identified dispersed throughout the genome [47]. In total 26 transpose genes were predicted in MSCT1, making it the fourth most abundant functional annotation in the proteome (Table 6).
Insights from the genome sequence

Functional T3SS, T4SS, and T6SS secretory systems were predicted in MSCT1. Comparison of the MSCT1 predicted proteins with previously described Xanthomonas effectors resulted in the identification of 7 families of effectors common among species of Xanthomonas (Table 7). These classes include AvrBs2, XopAG, XopK, XopP, XopR, XopT, and XopZ1. Effectors AvrBs2, XopK, XopP, XopR, and XopZ1, have been shown to suppress the host disease resistance response and immunity in other plant-Xanthomonas interactions [48–54]. XopAG effector family members have been shown to be responsible for eliciting the hyper-sensitive response in grapefruit [55]. The predicted protein sequence WP_033481547.1, predicted from the MSCT1 genome, exhibited homology to AvrBs2 effector proteins from

![Fig. 5 Dot plot of X. citri pv. malvacearum strain MSCT1 chromosome (NZ_CP017020.1) (X-Axis) compared to X. citri pv. malvacearum strain X18 (NZ_CM002136.1) (left, Y-Axis) and X. citri pv. malvacearum strain X20 (NZ_CM002029.1) (right, Y-Axis) Chromosomes. Dot plot produced with YASS web server using default settings [89]](image)

Table 4 Genome statistics

| Attribute                  | Value  | % of Total |
|----------------------------|--------|------------|
| Genome size (bp)           | 5,123,946 | 100.00    |
| DNA coding (bp)            | 4,365,468 | 85.20     |
| DNA G + C (bp)             | 3,313,791 | 64.67     |
| DNA scaffolds              | 4       | 100.00     |
| Total genes                | 4410    | 100.00     |
| Protein coding genes       | 4102    | 93.02      |
| RNA genes                  | 95      | 2.15       |
| Pseudo genes               | 213     | 4.83       |
| Genes in internal clusters | -       | -          |
| Genes with function prediction | 3375   | 76.53      |
| Genes assigned to COGs     | 3228    | 73.20      |
| Genes with Pfam domains    | 3302    | 74.88      |
| Genes with signal peptides | 553     | 12.54      |
| Genes with transmembrane helices | 911 | 20.66     |
| CRISPR repeats             | 1       | -          |

The total is based on the total number of protein coding genes in the genome.
several species of *Xanthomonas* and contained a predicted glycerolphosphoryl diester phosphodiesterase family (PF03009) domain characteristic of the AvrBs2 effector family [10]. AvrBs2 produced by *Xanthomonas campestris* pv. *vesicatoria* is recognized by a NBS-LRR in peppers containing the Bs2 resistance gene; however, field strains of *X. campestris* pv. *vesicatoria* have been identified that evade the recognition [56, 57].

EffectiveDB predicted 408 T3SS and 44 T4SS secreted proteins. MSCT1 predicted secreted proteins that have previously been associated with diseases in *G. hirsutum* and other plant systems include; endoglucanase [58], polygalacturonase [59], glutathione S-transferase [60],pectate lyase [61], glutathione peroxidase [62], as well as catabolic enzymes such as peptidases and lipases. These protein likely aid the mediation of the host disease response as well as the breaking down of host tissues. The PIP-box sequence was identified 78 bp up stream of the start codon for the *HrpB1* gene, that indicates gene regulation via PIP targeted transcription factors are present in the MSCT1 genome. EffectiveDB also identified 22 eukaryotic-like domains among 36 MSCT1 proteins. The most represented eukaryotic-like domains were the of M13 peptidase family (PF01431 and PF05649); however, M13 peptidases are commonly identified among bacteria [63].

**Extended insights**

AnnoTAL identified 8 potential CDS regions in the MSCT1 genome that could potentially code for TAL effectors (Table 8). AnnoTAL did not predict any TAL sequences in the other four draft *X. citri* pv. *malvacearum* genomes reported previously (GCF_000454505.1 (strain: X18), GCF_000454525.1 (strain: X20), GCF_00309925.1 (strain: GSPB22388) and GCF_000309915.1 (strain: GSPB1386)). Interestingly, 7 of the 8 TAL effectors in *X. citri* pv. *malvacearum* MSCT1 are located on plasmids. This arrangement is in contrast to other xanthomonads such as *Xanthomonas oryzae* pv. *oryzae* and *Xanthomonas oryzae* pv. *oryzicola* where the vast majority of TAL effectors are located on the large chromosome. The presence of the *X. citri* pv. *malvacearum* TAL effectors in *X. citri* pv. *malvacearum* plasmids can be traced back to the initial report by De Feyter et al. 1991 [64], that described 6 avirulence genes on a 90 kb *X. citri* pv. *malvacearum* plasmid [20–22]. However, the *X. citri* species and *X. oryzae* species such as *X. oryzae* pv. *oryzae* and *X. oryzae* pv. *oryzicola* exhibit evolutionarily divergence and fall into different clades among the other sequenced xanthomonads in phylogenic analysis [65]. Although, the overall total number of TAL effectors found in MSCT1 (n = 8) is less than what has been previously reported for some *X. oryzae* pv. *oryzae* and *X. oryzae* pv. *oryzicola* strains it is similar to strains of *X. translucens* [43, 47, 66].

The variable dinucleotide repeats were identified in the 8 MSCT1 TAL sequences for recognition of the TAL DNA target domain with the previously reported TAL code (Table 9). Due to the inherit degeneracy nature of TAL DTD prediction [12, 45, 67–70], potential TAL DTDs reported in this study are limited to the top 2 DTD site predictions for each TAL with the additional constraint of being within 150 bp of the gene start codon. Interestingly, MSCT1 TALs (MSCT1-TAL2 and

| Table 6 | Ten most represented functional annotations |
|---------|------------------------------------------|
| Annotation | Count |
| Membrane protein | 64 |
| TonB-dependent receptor | 42 |
| MFS transporter | 33 |
| Transposase | 26 |
| Transcriptional regulator | 25 |
| ABC transporter ATP-binding protein | 23 |
| Oxidoreductase | 19 |
| LysR family transcriptional regulator | 19 |
| GGDEF domain-containing protein | 16 |
| Two-component sensor histidine kinase | 15 |

| Table 7 | Xanthomad Non-TAL Effector families found in MSCT1 |
|---------|--------------------------------------------------|
| Family | Refseq_ID | BlastP HIT | Notes | REF |
| XopAG | WP_033479491.1 | CAP49915.1 | HR response in Grapefruit | [55] |
| XopK | WP_005915119.1 | CAP50604.1 | Unclear role in virulence | [52, 83, 84] |
| XopP | WP_069288200.1 | CAJ22867.1 | Suppresses immune response in rice | [49] |
| XopR | WP_005923840.1 | BAE70889.1 | Suppression of MAMP-triggered immune responses | [48, 53, 54] |
| XopT | WP_069288215.1 | BAE68965.1 | - | [83] |
| XopZ1 | WP_005914471.1 | BAE69157.1 | Contributes to virulence in rice | [51, 52] |
| AvrBs2 | WP_033481547.1 | CAJ21683.1 | Suppresses rice immunity | [50] |
Table 8: MSCT1 Potential TAL Effectors

| TAL      | Molecule    | Refseq_ID          | Starta  | Stopa  | Strandb |
|----------|-------------|--------------------|---------|--------|---------|
| MSCT1-TAL1 | pMSCT44kb   | WP_069288206.1     | 36,431  | 40,333 | 1       |
| MSCT1-TAL2 | pMSCT60kb   | WP_069288209.1     | 16,043  | 19,127 | 1       |
| MSCT1-TAL3 | Chromosome  | WP_069288181.1*b  | 2,568,181 | 2,571,268 | -1 |
| MSCT1-TAL4 | pMSCT44kb   | WP_069288204.1     | 15,111  | 19,626 | -1      |
| MSCT1-TAL5 | pMSCT60kb   | WP_069288212.1     | 41,404  | 44,689 | 1       |
| MSCT1-TAL6 | pMSCT60kb   | WP_069288211.1     | 44,689  | 34,259 | -1      |
| MSCT1-TAL7 | pMSCT60kb   | WP_069288210.1     | 21,870  | 21,870 | 1       |
| MSCT1-TAL8 | pMSCT60kb   | WP_069288208.1     | 3549    | 8064   | 1       |

*aStart, Stop, and Strand annotations by AnnoTAL
*bNCBI Annotation differs from AnnoTAL prediction, the MSCT1-TAL3 NCBI Start Codon begins at 2,570,908

MSCT1-TAL8) with a DTD prediction had predictions on corresponding sections of the A and D sub-genomes of the *G. hirsutum* TM-1 assembly [46]. However, these in silico predictions still need to be confirmed with RNA expression data from studies of *G. hirsutum* undergoing infection by MSCT1. Of note, no MSCT1 TAL DTD was predicted to target any promoter region on *G. hirsutum* chromosome 14 or 20 that contain the *B*2, *B*3 and *B*12 genes that are a major source of resistance to *X. citri pv. malvacearum* [71–73].

Of the predicted TALs only two, MSCT1-TAL2 and MSCT1-TAL8, had target sequences that fall within 100 bp of the start codon. MSCT1-TAL2 was predicted to target 21 bp from the start codon of the two paralogous proteins (Gh_A04G1143, Gh_D04G1757) found on chromosome 4 of each of the respective sub-genomes of tetraploid cotton. The proteins that MSCT1-TAL2 potential targets contain the ProSiteProfiles NAC domain profile (PS51005). The NAC domain has been reported to participate in both biotic and abiotic stress related responses [74]. MSCT1-TAL8 targeted 19 and 20 base pairs upstream of the paralogous proteins (Gh_A01G1702, Gh_D01G1952) in the A and D sub-genomes of *G. hirsutum*, respectively.

**Conclusions**

The MSCT1 genome reported in this study is the first *X. citri pv. malvacearum* genome to be completed with long read DNA sequencing technology. The long read sequencing and assembly strategy allowed for the identification of eight TAL effectors in *X. citri pv. malvacearum* and makes the MSCT1 genome assembly the only *X. citri pv. malvacearum* genome with assembled TAL effectors. In addition to the TAL effector identification, many T3SS effectors were identified in MSCT1 genome. The genome assembly, as outlined in this paper, provides a basis for future epidemiological and pathogenesis studies of the *X. citri pv. malvacearum-G. hirsutum* pathogen host complex.

**Abbreviations**

(DTD): DNA target domain; (HVS): Highly virulent strain; (MLST): Multilocus sequence typing; (PIP): Plant inducible promoter; (TAL): Transcription activator-like

**Acknowledgements**

We would like acknowledge Amanda Lawrence from the Mississippi State University Institute for Imaging & Analytical Technologies for assistance in the generation of the TEM imaging of MSCT1.

**Funding**

This work was funded by Cotton Incorporated 13-479 (to SL and DGP); USDA ARS 58-6402-1-644 (to DGP), and USDA ARS 58-6066-6-046 (to DGP).

**Authors’ contributions**

KCS, CH, and SL wrote the manuscript. TA provided the leaf sample and images of the diseased leaves. KCS, MAA, CH, BEM, SL, MJW, DGP edited the manuscript. MAA analysed raw sequence data, assembled, and circularized the genome and plasmids. XW conducted the pathology assays. JJ assisted in the generation of the TEM image. MAA analysed raw sequence data and assembled the genome. KCS, MAA, BEM, and MJW analysed functional sequence data. All authors read and approved the final manuscript.
Competing interests

There are no significant competing financial, professional or personal interests that might have influenced the performance or presentation of the research reported in this manuscript.

Publisher’s Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

1. Institute for Genomics, Biocomputing and Biotechnology, Mississippi State University, Mississippi State, MS 39762, USA. 2. Department of Biochemistry, Molecular Biology, Entomology and Plant Pathology, Mississippi State University, Mississippi State, MS 39762, USA. 3. Department of Basic Sciences, College of Veterinary Medicine, Mississippi State University, Mississippi State, MS 39762, USA. 4. USDA-ARS, Crop Science Research Lab, Genetics and Sustainable Agriculture Research Unit, Mississippi State, MS 39762, USA. 5. Cotton Incorporated, Cary, NC 27513, USA. 6. Mississippi State University, Delta Research and Extension Center, 82 Stoneville Rd, Stoneville, MS 38776, USA. 7. Department of Plant & Soil Sciences, Mississippi State University, Mississippi State, MS 39762, USA.

Received: 22 February 2017 Accepted: 12 July 2017

Published online: 24 July 2017

References

1. Hillocks R. Bacterial blight. In: Hillocks R, editor. Cotton diseases. Wallingford: CAB International; 1992. p. 39–64.
2. Morgham AT, Richardson PE, Eisenberg M, Cover EC. Effects of continuous cotton breeding. In: Cotton. Madison: American Society of Agronomy, Inc., Crop Science Society of America, Inc., and Soil Science Society of America, Inc.; 2013.
3. Bourland FM, Myers GO. Conventional cotton breeding. In: Cotton. Madison: American Society of Agronomy, Inc., Crop Science Society of America, Inc., and Soil Science Society of America, Inc.; 2013.
4. El, editors. Xanthomonas. London: Chapman & Hall; 1993. p. 193–204.
5. Al Mousawi A, Richardson P, Eisenberg M, Johnson W. Ultrastructural studies of a compatible interaction between Xanthomonas campestris pv. maltacearum and bacterial blight susceptible and resistant cotton. Physiol Mol Plant Pathol. 1988;32(1):141–62.
6. Al Mousawi A, Richardson P, Eisenberg M, Johnson W. Ultrastructural studies of a compatible interaction between Xanthomonas campestris pv. maltacearum and cotton [Gossypium hirsutum, bacterial blight]. Phytopathology. 1982;72:2122–30.
7. Rudolph K. Infection of the plant by Xanthomonas. In: Swings JG, Civerolo EL, editors. Xanthomonas. London: Chapman & Hall; 1993. p. 194–196.
8. Knight R, Clouton T. The genetics of bacterial resistance. J Genet. 1939;38(1):133–59.
9. Bourland FM, Myers GO. Conventional cotton breeding. In: Cotton. Madison: American Society of Agronomy, Inc., Crop Science Society of America, Inc., and Soil Science Society of America, Inc.; 2013.
10. Bourland FM, Jones DC. Registration of ‘UA103’ cotton cultivar. J Plant Registrations. 2013;7(2):135–9.
11. Bourland FM, Jones DC. Registration of Arkot 0305, Arkot 0306, Arkot 0309, and Arkot 0316 germplasm lines of cotton. J Plant Registrations. 2013;7(2):135–9.
12. Bourland FM, Jones DC. Registration of Arkot 0305, Arkot 0306, Arkot 0309, and Arkot 0316 germplasm lines of cotton. J Plant Registrations. 2015;9(1):94–8.
13. Delannoy E, Lyon BR, Marmey P, Jalloual A, Daniel JF, Montillet JL, Eisenberg M, Nicole M. Resistance of cotton towards Xanthomonas campestris pv. maltacearum. Annu Rev Phytopathol. 2005;43:63–82.
14. Buttnor D, Bonas U. Regulation and secretion of Xanthomonas avirulence/pathogenicity gene family encodes functional plant nuclear targeting signals. Mol Plant Microbe Interact. 1995;8(4):627–31.
15. Xu J, Deng P, Showmaker KC, Wang H, Baird SM, Lu S-E. The pqqC gene is essential for antifungal activity of Pseudomonas kloeseri. JX22 against Fusarium oxysporum f. sp. lycopersici. FEMS Microbiol Lett. 2014;353(2):98–105.
16. Yang Y, De Feyter R, Gabriel DW. Host-specific symptoms and increased release of Xanthomonas citri and X. campestris pv. maltacearum from leaves are determined by the 102-bp tandem repeats of pthA and avrB, respectively. Mol Plant Microbe Interact. 1994;7(2):345–55.
17. Gross DC, DeVay JE. Production and purification of syringomycin, a phytotoxin produced by Pseudomonas syringae. Physiol Plant Pathol. 1977;11:13–28.
18. Bourland FM, Jones DC. Registration of Arkot 0305, Arkot 0306, Arkot 0309, and Arkot 0316 germplasm lines of cotton. J Plant Registrations. 2013;7(2):135–9.
19. Xu J, Deng P, Showmaker KC, Wang H, Baird SM, Lu S-E. The pqqC gene is essential for antifungal activity of Pseudomonas kloeseri. JX22 against Fusarium oxysporum f. sp. lycopersici. FEMS Microbiol Lett. 2014;353(2):98–105.
20. Yang Y, De Feyter R, Gabriel DW. Host-specific symptoms and increased release of Xanthomonas citri and X. campestris pv. maltacearum from leaves are determined by the 102-bp tandem repeats of pthA and avrB, respectively. Mol Plant Microbe Interact. 1994;7(2):345–55.
21. Rice P, Longden I, Bleasby A. EMBOSS: the European molecular biology open software suite. Trends Genet. 2000;16(6):276–7.
22. Atschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–8.
23. Chaisson MJ, Tesler G. Mapping single molecule sequencing reads using basic local alignment with subsequent refinement (BLASR): application and theory. BMC Bioinformatics. 2012;13(1):238.
24. Robinson JT, Thomson N, Allen MJ, Angiuoli SV, et al. The minimum information about a genome sequence (MIGS) specification. Nat Biotechnol. 2008;26(5):541–8.
25. Robinson JT, Thomson N, Allen MJ, Angiuoli SV, et al. The minimum information about a genome sequence (MIGS) specification. Nat Biotechnol. 2008;26(5):541–8.
26. Bolger AM, Lohse M, Usadel B. Trimmmomatic: a flexible trimmer for illumina sequence data. Bioinformatics. 2014;30(15):2114–20.
27. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Quail MA, Sanchez J, Grimont PA, Brisse S, Nesme X, Chiroleu F, Bui A, Gross DC, DeVay JE. Production and purification of syringomycin, a phytotoxin produced by Pseudomonas syringae. Physiol Plant Pathol. 1977;11:13–28.
Grissa I, Vergnaud G, Proucel C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res. 2007;35(Web Server issue):W52–7.

Fenselau S, Bonas U. Sequence and expression analysis of the hrpB pathogenicity operon of Xanthomonas campestris pv. vesicatoria which encodes eight proteins with similarity to components of the Hrp, Ysc, spa, and Fli secretion systems. Mol Plant Microbe Interact. 1995;8(6):845–54.

Lee B-M, Park Y-J, Park D-S, Kang H-W, Kim J-G, Song E-S, Park I-C, Yoon U-H, Hahn J-H, Kim B-S, et al. The genome sequence of Xanthomonas oryzae pv. oryzae pathovar oryzae KACC10331, the bacterial blight pathogen of rice. Nucleic Acids Res. 2005;33(21):777–86.

Peng Z, Hu Y, Lei J, Potnis N, Athukorala A, Jones J, Liu Z, White FF, Liu S. Long read and single molecule DNA sequencing simplifies genome assembly and TAL effector gene analysis of Xanthomonas translucens. BMC Genomics. 2016;17(1):1–19.

Grau J, Reschke M, Eriks A, Streubel J, Morgan RG, Wilson GG, Koebnik R, Boch J. AnnoTALE: bioinformatics tools for identification, annotation, and nomenclature of TALs from Xanthomonas genomic sequences. Sci Rep. 2016;6:21077.

Grau J, Wolff A, Reschke M, Bonas U, Poisich S, Boch J. Computational predictions provide insights into the biology of TAL effector target sites. PLoS Comput Biol. 2013;9(3):e1002962.

Zhang T, Hu Y, Jiang W, Fang L, Guan X, Chen J, Zhang J, Sasaki CA, Scheffler BE, Stelly DM, et al. Sequencing of allotetraploid cotton (Gossypium hirsutum L. acc. TM–1) provides a resource for fiber improvement. Nat Biotechnol. 2015;33(5):537–7.

Booher NJ, Carpenter SCD, Sebra RP, Wang L, Salzberg SL, Leach JE, Bowler CJ. Single molecule real-time sequencing of Xanthomonas oryzae genome reveals a dynamic structure and complex TAL (transcription activator-like) effector gene relationships. Microb Genomics. 2015;1(4):1–22.

Akimoto-Tomiyama C, Furutani A, Tsuge S, Washington EJ, Nishizawa Y, Hirose N, Gassmann W, Dahlbeck D, Chesnokova O, Minsavage GV, Jones JB, Staskawicz BJ. Molecular evolution of virulence in natural field strains of Xanthomonas campestris pv. vesicatoria. Mol Microbiol. 2005;55:2235–46.
84. Mutka AM, Fentress SJ, Sher JW, Berry JC, Pretz C, Nusinow DA, Bart R. Quantitative, image-based phenotyping methods provide insight into spatial and temporal dimensions of plant disease. Plant Physiol. 2016;172(2):650–60.
85. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–80.
86. Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol. 1993;10(3):512–26.
87. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol. 2013;30(12):2725–9.
88. Grant JR, Stothard P. The CGView server: a comparative genomics tool for circular genomes. Nucleic Acids Res. 2008;36(Web Server issue):W181–4.
89. Noé L, Kucherov G. YASS: enhancing the sensitivity of DNA similarity search. Nucleic Acids Res. 2005;33(suppl 2):W540–3.