Phytopathology • 2018 • 108:172-185 • https://doi.org/10.1094/PHYTO-05-17-0171-R

Bacteriology

e-Xtra*

Comparative Genomic Analyses of *Clavibacter michiganensis* subsp. *insidiosus* and Pathogenicity on *Medicago truncatula*

You Lu, Carol A. Ishimaru, Jane Glazebrook, and Deborah A. Samac†

First and third authors: Department of Plant and Microbial Biology, second and fourth authors: Department of Plant Pathology, and first, second, third, and fourth authors: the Microbial and Plant Genomics Institute, University of Minnesota, St. Paul 55108; and fourth author: United States Department of Agriculture–Agricultural Research Service, Plant Science Research, St. Paul, MN 55108.
Accepted for publication 22 September 2017.

ABSTRACT

Clavibacter michiganensis is the most economically important gram-positive bacterial plant pathogen, with subspecies that cause serious diseases of maize, wheat, tomato, potato, and alfalfa. Much less is known about pathogenesis involving gram-positive plant pathogens than is known for gram-negative bacteria. Comparative genome analyses of *C. michiganensis* subspecies affecting tomato, potato, and maize have provided insights on pathogenicity. In this study, we identified strains of *C. michiganensis* subsp. *insidiosus* with contrasting pathogenicity on three accessions of the model legume *Medicago truncatula*. We generated complete genome sequences for two strains and compared these to a previously sequenced strain and genome sequences of four other subspecies. The three *C. michiganensis* subsp. *insidiosus* strains varied in gene content due to genome rearrangements, most likely facilitated by insertion elements, and plasmid number, which varied from one to three depending on strain. The core *C. michiganensis* genome consisted of 1,917 genes, with 379 genes unique to *C. michiganensis* subsp. *insidiosus*. An operon for synthesis of the extracellular blue pigment indigoidine, enzymes for pectin degradation, and an operon for inositol metabolism are among the unique features. Secreted serine proteases belonging to both the *pat-1* and *ppa* families were present but highly diverged from those in other subspecies.

The genus *Clavibacter* currently consists of only one species, *Clavibacter michiganensis*, which is further divided into nine subspecies recognized as plant pathogens or plant- associated bacteria (Eichenlaub and Gartemann 2011; Gonzalez and Trapiello 2014; Oh et al. 2016; Yasuhara-Bell and Alvarez 2015). The molecular biology of gram-positive phytopathogenic bacteria such as *C. michiganensis* has not been as well characterized as their gram-negative counterparts, partially due to the limited genetic tools available for research. Early studies of *C. michiganensis* subsp. *michiganensis* and *C. michiganensis* subsp. *sepedonicus* relied on naturally occurring or induced variants to infer the roles of the underlying genomic regions (Meletzus et al. 1993), followed by extensive mapping to identify the responsible genes (Dreier et al. 1997; Jahr et al. 2000). Forward-genetic screening of transposon mutant populations provided a powerful alternative approach to isolate genes of interest without the prerequisite of an available whole-genome sequence (Kirchner et al. 2001). Though there are successful examples of studies applying this strategy (Nissinen et al. 2009), transposon mutagenesis has a number of drawbacks such as biases in insertion sites and reliance on high transformation efficiency. Reverse-genetic approaches can overcome most issues associated with transposon mutagenesis by targeting insertions into genes of interest through homologous recombination, thus creating desirable mutants for further functional characterization (Stork et al. 2008). However, a reverse-genetic approach cannot be applied unless the sequences of the genes of interest are known. Hampered by several such technical limitations, research on *C. michiganensis* subsp. *insidiosus*, the causal agent of bacterial wilt of alfalfa (*Medicago sativa* L.), has focused primarily on plant resistance and diagnostic methods (Marefat et al. 2007; Samac et al. 1998). Resistance to bacterial wilt in alfalfa was first identified in Ladak and Turkestan germplasm (Peltier and Schroeder 1932) and has been incorporated into most alfalfa cultivars currently grown in the United States. However, the autotetraploid genome and complex inheritance of traits limits use of alfalfa for understanding host–pathogen interactions. The related diploid model plant, *M. truncatula*, has proven highly amenable to research on interactions with pathogens and symbionts (Kouchi et al. 2010; Tivoli et al. 2006) but a pathosystem with *C. michiganensis* has been lacking.

Traditionally, complete microbial genomes were sequenced from DNA shotgun clone libraries using the Sanger chain-termination sequencing method (Sanger et al. 1977). This approach successfully produced high-quality, fully assembled genomes, including *C. michiganensis* subsp. *michiganensis* NCPPB 382 (Gartemann et al. 2008), *C. michiganensis* subsp. *sepedonicus* ATCC 33113 (Bentley et al. 2008), and *C. michiganensis* subsp. *nebraskensis* NCPPB 2581 (Eichenlaub and Gartemann 2011), but is expensive and time consuming. One of the recent technical advances for whole genome sequencing is the PacBio Single-Molecule Real-Time (SMRT) sequencing technology, which can directly sequence DNA samples following the SMRTbell template preparation without the need for amplification and produces very long sequences in a high-throughput fashion (Eid et al. 2009). Taking advantage of the SMRT sequencing technology, the complete genomes of *C. michiganensis* subsp. *capsici* PF008 and *C. michiganensis* subsp. *insidiosus* R1-1 were published recently (Bae et al. 2015; Lu et al. 2015).

The availability of complete genomes not only enables comparative analyses such as conservation of proteomes and presence of known virulence factors but also elucidates the structural differences between the genomes of these subspecies. The chromosomes of *C. michiganensis* subsp. *michiganensis* NCPPB 382 and *C. michiganensis* subsp. *nebraskensis* NCPPB 2581 show colinearity over most of their lengths (Załuga et al. 2014), while the chromosome of *C. michiganensis* subsp. *sepedonicus* ATCC 33113 has undergone extensive rearrangement relative to the former two, partially due to transpositions of the insertion element IS*1121*,

†Corresponding author: D. A. Samac; E-mail: debby.samac@ars.usda.gov

*The *e*-**X**tra logo stands for "electronic extra" and indicates that three supplementary figures and two supplementary tables are published online.

which is present in many copies (Bentley et al. 2008). Limited by only one complete genome each for subspecies *michiganensis, sepedonicus,* and *nebraskensis*, the extent of chromosomal rearrangement among strains within the same subspecies is not known.

The objectives of this study were to identify *C. michiganensis* subsp. *insidiosus* strains varying in pathogenicity on accessions of *M. truncatula* and to generate de novo assemblies of the complete genome sequences of three *C. michiganensis* subsp. *insidiosus* strains contrasting in pathogenicity. Genomic comparisons among the three strains revealed common structural features as well as variations in chromosomal organization and gene contents. Genomic comparisons with other subspecies of *C. michiganensis* identified a set of candidate genes potentially involved in pathogenicity.

MATERIALS AND METHODS

Bacterial strains and culture conditions. The *C. michiganensis* subsp. *insidiosus* strains used in this research were isolated from infected *M. truncatula* plants. Seed of *M. truncatula* accession R108 were sandpaper scarified, planted in steam-pasteurized sand, and grown in a greenhouse. At 10 days after planting, the plants were removed from the sand, roots were rinsed briefly to remove sand particles, and the distal 1 cm of the taproot was excised. The remaining roots were soaked for 30 min in an inoculum prepared from 10 g of crushed alfalfa roots with symptoms of bacterial wilt in 100 ml of sterile water. The alfalfa roots had been harvested from alfalfa plants grown in a plant disease nursery at the University of Minnesota Research and Outreach Center in Rosemount, MN, and frozen at −20°C until used as inoculum. Inoculated *M. truncatula* R108 plants were replanted in sand and grown for 30 days. Stems were removed from plants, surface sterilized, cut into 1-cm pieces, and soaked in sterile water. Serial dilutions were plated on nutrient-broth yeast extract (NBY) agar plates (Vidaver 1967) and incubated at 25°C. After 7 days, colonies with the culture characteristics of *C. michiganensis* subsp. *insidiosus* were removed and streaked on NBY agar plates. Strains CmiR1-1 and CmiR1-3 isolated from R108 were confirmed to be *C. michiganensis* subsp. *insidiosus* by polymerase chain reaction (PCR), as previously described (Samac et al. 1998). *C. michiganensis* subsp. *insidiosus* ATCC 10253 (CmiATCC 10253) is the type strain and was obtained from the American Type Culture Collection (ATCC, Manassas, VA).

Pathogenicity assays. Cultures of CmiR1-1, CmiR1-3, and CmiATCC 10253 were initiated from frozen glycerol stocks and were grown on NBY agar plates for 7 days at 25°C. Seed of *M. truncatula* accessions A17, DZA315.16, and F83005.5 were sandpaper scarified and planted in steam-pasteurized sand in a growth chamber at day and night temperatures of 24 and 19°C, respectively, with a 16-h photoperiod. Roots of 7-day-old plants were soaked in a bacterial inoculum for 30 min, then replanted in sand. The inoculum consisted of CmiR1-1, CmiR1-3, or CmiATCC 10253 removed from NBY plates, suspended in sterile distilled water, and adjusted to a concentration of optical density at 600 nm (OD_{600}) = 1.0 (approximately 2.5×10^9 CFU/ml). Plants were watered daily and received a complete fertilizer at 14 and 21 days after planting.

Pathogenicity of each strain was determined by measuring CFU in roots and shoots at 5, 10, and 20 days after inoculation by a quantitative PCR (qPCR) assay using the primers CmiF241005 (5′-GTCAGGCG TTTGTCCTGGT-3′) and CmiR241005 (5′-CCACCACCATTCAC TCCG-3′) (Marefat et al. 2007). A standard curve was generated using serial dilutions of CmiR1-1 cells in sterile water. Dilutions were plated on NBY to determine CFU per milliliter. Bacterial suspensions from inoculated plants were diluted 1:10 with sterile water and 5 µl of the dilution was used in the qPCR. Each 25-µl PCR consisted of SYBR Green Supermix with ROX (Bio-Rad Laboratories, Hercules, CA), 40 pmol of each primer, and 5 µl of bacterial suspension. All reactions were run in duplicate using an ABI Prism7000 Sequence Detection System (Applied Biosystems, Foster City, CA). The PCR conditions were 95°C for 10 min followed by 40 cycles of 95°C for 10 s and 60°C for 1 min. A melting curve analysis was conducted according to the ABI software. A linear relationship existed between cycle threshold values and the log CFU between 10^1 to 10^6 CFU/ml. Five plants from each accession inoculated with each strain were assayed at each time point.

Genomic DNA extraction and whole genome sequencing. *C. michiganensis* subsp. *insidiosus* cultures were grown in 50 ml of yeast extract glucose mineral salts (YGM) broth (De Boer and Copeman 1980) at 25°C to OD_{600} = 1.0, then centrifuged to collect bacterial cells. Preparation of genomic DNA of CmiR1-3 and CmiATCC 10253 for whole-genome sequencing was performed with a modified protocol based on the cetyltrimethylammonium bromide method developed for gram-negative bacterial genomic DNA extraction (Wilson 2001). A step involving incubation with lysozyme (Sigma-Aldrich, St. Louis) at a final concentration of 2.5 µg/ml at 37°C for 1 h was added prior to the lysis step in the original protocol. The resulting DNA pellets were dissolved in 0.5× Tris-EDTA buffer (5 mM Tris-HCl and 0.5 mM EDTA, pH 8.0) and adjusted to concentrations specified by the downstream sequencing facilities.

CmiR1-3 and CmiATCC 10253 were sequenced using both Illumina MiSeq and the PacBio SMRT sequencing technology. For Illumina MiSeq, each sample was prepared using the Illumina TruSeq DNA library preparation kit with an average 300-bp insert size, and was sequenced as paired ends with a 150-bp read length. For PacBio SMRT sequencing, each sample was prepared as a 20-kb insert library for PacBio P6-C4 chemistry followed by BluePippin size selection at a 15-kb cut-off. Each library was then sequenced using one SMRT cell in the PacBio RSII sequencing system. The SMRT sequencing data were quality filtered with a read quality score cut-off of 0.80 in the PacBio SMRT portal (version 2.2.0).

De novo genome assembly. The de novo genome assembly pipeline used was the same as described for the published genome of CmiR1-1 (Lu et al. 2015). The initial assembly of the PacBio sequencing reads was conducted using the hierarchical genome-assembly process 3 algorithm (Chin et al. 2013) in the PacBio SMRT portal (version 2.2.0). The minimum subread length cut-off value was set at 4,000 bp for the genome assembly of CmiR1-3 and CmiATCC 10253. The polymerase reads were then used to preassemble 48,349 subreads with a mean length of 8,970 bp for CmiR1-3 and 47,528 subreads with a mean length of 8,533 bp for CmiATCC 10253. The mean fold-coverage obtained was 120.11× and 117.1× for CmiR1-3 and CmiATCC 10253, respectively. The resulting assemblies were corrected with the Quiver consensus algorithm (Chin et al. 2013) implemented in the Resequencing Pipeline in the PacBio SMRT portal using the same set of PacBio sequencing data with the default subread length cut-off setting, in order to obtain a high-accuracy genome assembly. Further improvement of the quality of the genome sequences was performed with Pilon (version 1.10) (Walker et al. 2014) using the sequencing data generated by the Illumina MiSeq platform for CmiR1-3 and CmiATCC 10253. The MiSeq reads of CmiR1-3 and CmiATCC 10253 were aligned to their corresponding PacBio genome assembly scaffolds using Bowtie2 (version 2.2.4) (Langmead and Salzberg 2012) and SAMtools (version 1.2) (Li et al. 2009), and the aligned reads were then used as input for Pilon. This process resolved 89 insertion and deletion (INDEL) errors in CmiR1-3 and 114 INDEL errors in CmiATCC 10253. All the contigs showed greater than 5-kb overlapping sequences at both ends and were manually closed to circular contigs. These were the final versions of the de novo assemblies used in the downstream analyses. Genome assemblies were deposited at the National Center for Biotechnology Information (NCBI) under accession numbers CP021034 to CP021037 for CmiR1-3 and CP021038 to CP021039 for CmiATCC 10253. Analyses of these genomes are based on the first versions of the assemblies.

Genome annotation. Genome annotation of CmiR1-3 and CmiATCC 10253 was performed with the RAST Server (Aziz et al. 2008) and manually curated based on the current version of the published CmiR1-1 genome. The annotation for the CmiR1-1 genome

has been updated relative to the original publication (Lu et al. 2015), which was performed using the NCBI Prokaryotic Genome Annotation Pipeline (version 2.10; https://www.ncbi.nlm.nih.gov/genome/annotation_prok/). For the analysis presented here, annotation based on the NCBI Prokaryotic Genome Annotation Pipeline (version 4.2) was used. Signal peptides for protein secretion were predicted with the SignalP 4.1 server (Petersen et al. 2011).

Functions of the *C. michiganensis* subsp. *insidiosus* genes were predicted by searching the NCBI Conserved Domain Database (CDD) (https://www.ncbi.nlm.nih.gov/cdd/) with the default parameter setting. The *C. michiganensis* subsp. *insidiosus* genes associated with a clusters of orthologous groups (COG) hit from the CDD search results were grouped based on the COG functional categories (Galperin et al. 2015; Tatusov et al. 2000). COG annotation was performed using the NCBI Web CD-Search Tool with the COG database (version 3.16) at the default E-value cut-off (0.01). Metabolic pathways were analyzed using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (http://www.genome.jp/kegg/). The predicted protein sequences from *C. michiganensis* subsp. *insidiosus* were annotated using the BlastKOALA tool (http://www.kegg.jp/blastkoala/) and the resulting KEGG annotation was used for pathway mapping and module mapping with the KEGG Mapper tools (http://www.genome.jp/kegg/mapper.html). Visualization of the data were done with DNAPlotter (version 10.2) (Carver et al. 2009).

PCR test for plasmids in *C. michiganensis* subsp. *insidiosus*. The presence of plasmids pCI1, pCI2, and pCI3 was tested by colony PCR using primers specifically designed to amplify unique regions on each plasmid. Bacterial cells were boiled in water for 10 min and used as PCR templates. pCI1 was detected with primers pCI1_F (5′ CTACTCGTTCACGGACTGTGC 3′) and pCI1_R (5′ GATACCGG CTTGCGACTTCAA 3′), with an expected PCR product size of 663 bp. pCI2 was detected with primers pCI2_F1 (5′ GGACCGTC ACGAGGATCACTA 3′) and pCI2_R1 (5′ ACAGCGGTCATCTAC GAGGAA 3′), or with pCI2_F2 (5′ AGAGGGTCGGCTAGATGTG TC 3′) and pCI2_R2 (5′ CCAGAACTCGTACTCGGGTCT 3′), with expected PCR product sizes of 506 and 611 bp, respectively. Positive amplification with either pair of pCI2 PCR primers was considered positive for the presence of pCI2. pCI3 was detected with primers pCI3_F (5′ GATTCTGTTGCCGATCGGTGT 3′) and pCI3_R (5′ CGTCCACTTGCGCAAAATAGC 3′), with an expected PCR product size of 699 bp. PCR products were analyzed with DNA gel electrophoresis.

Comparative genome analysis. The sequence similarity of plasmids was visualized using dot plots drawn with Dotter (version 4.34.3) (Sonnhammer and Durbin 1995), with the sliding window size set to 200. Alignment of chromosomes of *C. michiganensis* subsp. *insidiosus* strain CmiATCC 10253, CmiR1-1, and CmiR1-3 was visualized with the Artemis Comparison Tool (version 13.0.0) (Carver et al. 2005).

Colinearity among the chromosomes of CmiR1-1, *C. michiganensis* subsp. *capsici* strain PF008 (CmcPF008), CmmNCPPB 382, CmnNCPPB 2581, and CmsATCC 33113 was analyzed and visualized with Mauve (version 2.4.0) (Darling et al. 2004). The alignment between chromosomal sequences was performed with the built-in function "progressive Mauve" using the default settings for all parameters except the seed weight, which was set to 20.

To analyze the presence and absence of orthologous genes among the subspecies of *C. michiganensis*, the predicted protein sequences from CmiR1-1, CmcPF008, CmmNCPPB 382, CmnNCPPB 2581, and CmsATCC 33113 were first clustered for orthologous genes using the software package GET_HOMOLOGS (version 30122016) (Contreras-Moreira and Vinuesa 2013), with the following parameters set differently from the default: t = 0 (report all clusters), C = 80 (80% minimum coverage in BLAST pairwise alignment), E = 0.001 (maximum E-value cut-off at 0.001), S = 60 (60% minimum sequence identity in BLAST pairwise alignment), and D = 1 (requiring equal Pfam domain composition required for consideration as orthologous genes). The COGtriangle algorithm (Kristensen et al. 2010), which is set as one option in the get_homologs.pl as implemented in GET_HOMOLOGS, was used to compute the clustering of orthologous genes. To analyze the genome compositions from the established orthologous clusters, a pangenome matrix was computed using the compare_clusters.pl included in GET_HOMOLOGS. The information from this matrix was then extracted using the script parse_pangenome_matrix.pl included in GET_HOMOLOGS to compute the number of clusters present or absent in each input genome. A Venn diagram was drawn using the R package VennDiagram (version 1.6.9) to visualize the result.

Phylogenetic tree construction. Protein sequences of members of the Pat-1 and Ppa family from the genomes of CmmNCPPB 382 and CmsATCC 33113 were used to query the genome of CmiR1-1 using BLAST. This process resulted in four *pat-1* and five *ppa* homologs from CmiR1-1. For each family, the protein sequences of all the members were aligned using MUSCLE (Edgar 2004). The aligned protein sequences were used for constructing a phylogenetic tree using the neighbor-joining method (Saitou and Nei 1987) with 1,000 bootstraps. The resulting phylogenetic tree was visualized using CLC Main Workbench (version 7.6).

RESULTS

Pathogenicity of CmiR1-1, CmiR1-3, and CmiATCC 10253 on *M. truncatula* accessions. Mild symptoms of disease were observed after inoculation with CmiR1-3 in accessions DZA315.16 and F83005.5 but not inoculated plants of accession A17. At 5 days postinoculation (dpi), cotyledons of inoculated DZA315.16 plants were flaccid and yellow, at 10 dpi cotyledons were necrotic, and at 20 dpi cotyledons had fallen from plants (Supplementary Fig. S1). Cotyledons of inoculated A17 plants remained green and did not differ from mock-inoculated plants (Supplementary Fig. S2). Stunting of stems and roots occurred in inoculated DAZ315.16 plants compared with inoculated A17 plants and mock-inoculated plants. The fresh weight of inoculated DZA315.16 plants was significantly different at 20 dpi from that of mock-inoculated plants (Supplementary Fig. S3).

The three strains were detected in roots and stems of the three accessions at 5 dpi (Table 1). However, neither CmiATCC 10253 nor CmiR1-1 was detected at 10 or 20 dpi in any of the accessions. In contrast, populations of CmiR1-3 remained constant in roots and stems of DZA315.16 and F83005.5 through 20 dpi. Populations of CmiR1-3 dropped below the level of detection in A17 stems at 10 dpi and the pathogen was not detected in roots or stems at 20 dpi. The same patterns were observed after inoculation of the three accessions with strains CmiR1-5 and CmiR1-8 isolated from *M. truncatula* R108 (data not shown). These results indicate that CmiR1-3 is pathogenic whereas CmiATCC 10253 and CmiR1-1 are not pathogenic on *M. truncatula* and that A17 expresses defense mechanisms against *C. michiganensis* subsp. *insidiosus* that result in elimination of the pathogen below the level of detection.

General features of the chromosomes of *C. michiganensis* subsp. *insidiosus* CmiR1-1, CmiR1-3, and CmiATCC 10253. All three *C. michiganensis* subsp. *insidiosus* genomes contained one circular chromosome of approximately 3.2 Mb in size with a very high G+C content (73%), comparable with the genomes of other subspecies of *C. michiganensis* (Table 2). The zero point of the three chromosomes was assigned at the start codon of the *dnaA* gene, for consistency with other published *C. michiganensis* genomes. The chromosome of CmiR1-1 contained 2,968 protein coding genes (CDS), two ribosomal RNA (rRNA) operons (16S, 23S, and 5S), and 46 transfer RNA (tRNA) genes (Supplementary Table S1; Fig. 1). One tRNA with anticodon TGC coding for alanine was not present in the published genomes of other *C. michiganensis* subspecies. Its location in a 3-kb region conserved only in the *C. michiganensis* subsp. *insidiosus* genomes indicates that it may have been acquired by horizontal gene transfer. This 3-kb region is located from 3,130,922 to 3,133,941 bp on the chromosome of CmiR1-1. Searching the NCBI Nucleotide Collection database (nr) with BLASTN identified only a

subregion of approximately 600 bp from the 3-kb query with hits to the bacterial genera *Tessaracoccus, Alloactinosynnema*, and *Rhodococcus*. However, the highest nucleotide identity of the alignment is 68%, which is insufficient to infer the possible donor of the sequence. The chromosomes of CmiR1-3 and CmiATCC 10253 have sets of rRNA and tRNA identical to those of CmiR1-1, while the number of predicted CDS varies due to genomic INDEL variation in the copy number of insertion elements (Table 2).

Inspection of the three genome sequences revealed a high level of occurrence of the insertion element IS*1122,* which can lead to genomic rearrangements due to recombination between these elements. To determine whether the chromosomes of the three strains are organized in the same order, the sequences of the three chromosomes were aligned (Fig. 2). Eight recombination breakpoints between the chromosomes of CmiR1-1 and CmiATCC 10253 were identified. Among these, four breakpoints coincided with insertion sites of IS*1122* in both genomes and four breakpoints coincided with insertion sites in one genome. Between CmiR1-1 and CmiR1-3, all six recombination breakpoints detected were flanked by IS*1122* insertions in both genomes. Thus, chromosomal organization differs among the three strains, and the rearrangements are likely due to homologous recombination events between the IS*1122* sites. The differences in chromosomal organization between CmiR1-1 and CmiR1-3 derived from infected alfalfa roots in a plant disease nursery indicate heterogeneity of *C. michiganensis* subsp. *insidiosus* populations in the field in Minnesota.

The colinearity among the genomes of *C. michiganensis* subsp. *insidiosus* and other subspecies of *C. michiganensis* was visualized by determining the organization of locally colinear blocks (LCB) (Fig. 3). The alignment of the chromosomes was broken down into many LCB, mostly separated by the presence of insertion elements, as well as rearrangements in the genome of *C. michiganensis* subsp. *sepedonicus*. The majority of the LCB were present in all six aligned genomes, indicating that the gene contents of the LCB are preserved. The order of the LCB in the *C. michiganensis* subsp. *insidiosus* genomes was mostly consistent with the order of LCB in the genomes of *C. michiganensis* subsp. *michiganensis* and *C. michiganensis* subsp. *nebraskensis*, while the genome of *C. michiganensis* subsp. *sepedonicus* showed the most reordering of the LCB. Given that 59 of 102 copies of the IS elements are located at the boundary of LCB in *C. michiganensis* subsp. *sepedonicus* (Bentley et al. 2008), as well as similar incidents observed in the genomes of *C. michiganensis* subsp. *insidiosus* (Fig. 2), it seems likely that the copy number of IS elements in the genomes of *C. michiganensis* subsp. *sepedonicus* and *C. michiganensis* subsp. *insidiosus* correlates with the frequency of recombination events causing chromosome rearrangements.

The predicted functions of the CDS were categorized based on the COG terms (Galperin et al. 2015; Tatusov et al. 2000), as described in Methods and Materials. In total 1,648 CDS are associated with at least one specific COG term hit (Table 3; Fig. 1). In the following description of predicted CDS, if the described feature is conserved among all three *C. michiganensis* subsp. *insidiosus* genomes, only the representative gene from CmiR1-1 is listed, followed by the NCBI locus tags in parentheses. The CDS involved in metabolism and metabolic pathways were analyzed using the KEGG database. *C. michiganensis* subsp. *insidiosus* lacks enzymes to reduce nitrate, nitrite, sulfate, or sulfite and relies on reduced forms of nitrogen and sulfur, similar to *C. michiganensis* subsp. *michiganensis* (Gartemann et al. 2008). Ammonium may be taken up by CmiR1-1_1451 (VO01_RS07490), which encodes an ammonium channel protein, and used for glutamine biosynthesis by CmiR1-1_1416 to 1418 (VO01_RS07315 to 07325). Sulfide can be used for synthesizing cysteine from O-acetyl-L-serine by CmiR1-1_0578 (VO01_RS03020) and CmiR1-1_0579 (VO01_RS03025), which encode a cysteine synthase and a serine O-acetyltransferase.

Characteristics of plasmids in *C. michiganensis* subsp. *insidiosus*. Plasmids are widely prevalent in *C. michiganensis*. The published genomes of CmmNCPPB 382 and CmsATCC 33113 both contain two plasmids (Bentley et al. 2008; Gartemann et al. 2008), and a previous report of the presence of plasmids in *C. michiganensis* subspecies revealed that most isolates possess two plasmids while, rarely, some isolates contain three (Gross et al. 1979). In the genomes of CmiR1-1 and CmiR1-3, we detected three circular plasmids (pCI1, pCI2, and pCI3) with G+C contents slightly lower than that of the

TABLE 1. Average $\log_{10}$ CFU in roots and stems of three accessions of *Medicago truncatula* inoculated with *Clavibacter michiganensis* subsp. *insidiosus* strains CmiATCC 10253, CmiR1-1, and CmiR1-3 at 5, 10, and 20 days postinoculation (dpi)[a]

	Average $\log_{10}$ CFU at 5, 10, or 20 dpi								
	CmiATCC 10253			CmiR1-1			CmiR1-3		
Sample	5 dpi	10 dpi	20 dpi	5 dpi	10 dpi	20 dpi	5 dpi	10 dpi	20 dpi
A17 root	5.35 (0.33)	0	0	4.59 (0.57)	0	0	5.28 (0.45)	4.86 (0.79)	0
A17 stem	4.42 (0.53)	0	0	4.85 (0.32)	0	0	4.66 (1.0)	0	0
DZA315.16 root	4.11 (0.23)	0	0	5.92 (0.64)	0	0	6.27 (0.45)	5.44 (0.59)	5.43 (0.32)
DZA315.16 stem	4.50 (0.16)	0	0	4.15 (0.14)	0	0	5.41 (0.18)	5.33 (0.11)	5.50 (0.10)
F83005.5 root	4.81 (0.36)	0	0	4.58 (0.68)	0	0	5.54 (0.35)	5.54 (0.16)	5.46 (0.78)
F83005.5 stem	5.66 (0.20)	0	0	4.10 (0.45)	0	0	5.13 (0.30)	4.84 (0.51)	4.88 (0.71)

[a] Standard deviation of the mean is shown in parentheses (n = 5).

TABLE 2. Summary of general characteristics of the genomes of three strains of *Clavibacter michiganensis* subsp. *insidiosus*[a]

	CmiR1-1	CmiR1-3	CmiATCC 10253
Chromosome	3,207,520 bp, 73.0% G+C	3,193,464 bp, 73.0% G+C	3,186,916 bp, 72.9% G+C
pCI1	47,690 bp, 66.5% G+C	47,291 bp, 67.8% G+C	48,869 bp, 67.7% G+C
pCI2	49,401 bp, 67.6% G+C	48,316 bp, 67.6% G+C	ND
pCI3	103,451 bp, 66.2% G+C	102,376 bp, 66.1% G+C	ND
CDS	3,082	3,114	3,011
CDS on chromosome	2,906	2,945	2,961
Ribosomal RNA	6	6	6
Transfer RNA	46	46	46
IS*1122*	27	22	38
Secreted proteins	260	240	227

[a] ND = not detected and CDS = protein coding sequence.

chromosome (Table 2). pCI1 and pCI2 were named for homology to pCM1/pCS1 and pCM2, respectively (Fig. 4). The homology was determined based on sequence similarity and alignment with other *C. michiganensis* plasmids. The start points of pCI1 and pCI2 plasmids were aligned to the start points of pCM1/pCS1 and pCM2, respectively. The start points of these plasmids are within the origins of replication, which display very high similarities between homologous groups (pCM1/pCS1/pCI1 and pCM2/pCI2). On the other hand, pCI3, the largest plasmid (approximately 103 kb), is not homologous to any known *C. michiganensis* plasmid. Interestingly, only pCI1 was detected in CmiATCC 10253. To assess whether pCI1, pCI2, and pCI3 are present in other *C. michiganensis* subsp. *insidiosus* isolates, PCR primers developed specifically for detecting each of the three plasmids were used to test 30 additional isolates (Table 4). pCI1 is present in all 33 strains. However, pCI2 and pCI3 are not conserved in all the strains. Simultaneous presence of pCI2 and pCI3 was only

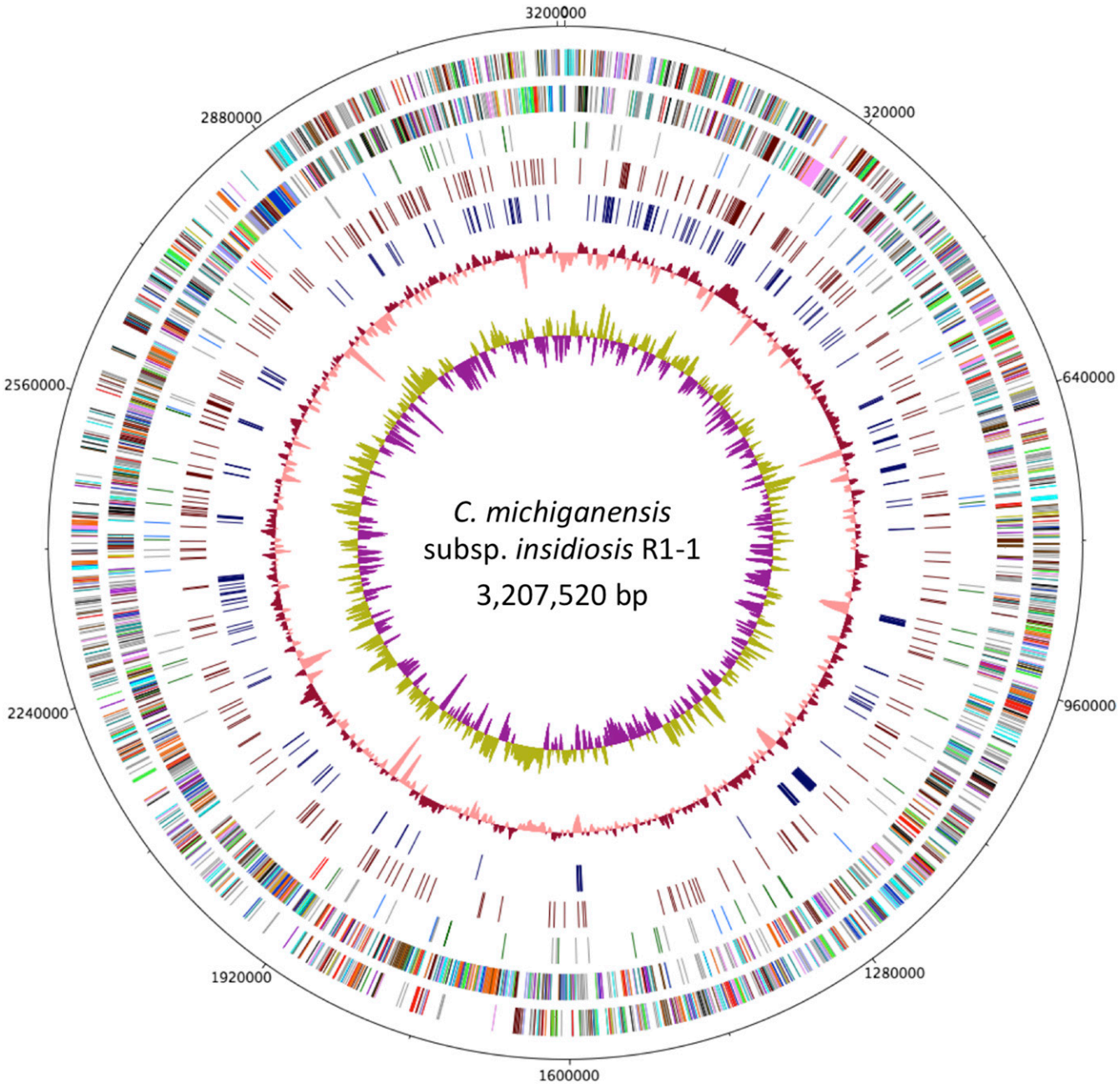

Fig. 1. Circular plot of the chromosome of CmiR1-1 with relevant features. From the outermost circle to the innermost circle: circle 1, nucleotide position (bp); circles 2 and 3, predicted protein coding sequences (CDS) in forward (circle 2) and reverse (circle 3) directions, with the colors indicating assigned clusters of orthologous groups (COG) categories; circle 4, predicted transfer RNA (in green), ribosomal RNA (in red), insertion element IS*1122* (in blue), and pseudogenes (in gray); circle 5, predicted CDS containing signal peptides for secretion (in red); circle 6, CDS (in blue) that do not have orthologs in the genomes of *Clavibacter michiganensis* subsp. *michiganensis*, *C. michiganensis* subsp. *sepedonicus*, or *C. michiganensis* subsp. *nebraskensis*; circle 7, G+C content, above average shown in dark red and below average shown in light red (window size = 5,000 bp); and circle 8, G+C skew, higher values shown in yellow and lower values shown in purple (window size = 5,000 bp). Color keys for COG categories in circle 2 and 3: red, = energy production and conversion; green = cell cycle control, cell division, chromosome partitioning; orange = amino acid transport and metabolism; yellow = nucleotide transport and metabolism; maroon = carbohydrate transport and metabolism; pink = coenzyme metabolism, secondary metabolites biosynthesis, transport, and catabolism; blue = translation, ribosomal structure, and biogenesis; teal = transcription; azure = replication, recombination, and repair; brown = cell wall, membrane, and envelope biogenesis; light blue = posttranslational modification, protein turnover, and chaperones; light green = inorganic ion transport and metabolism; black = signal transduction mechanisms and defense mechanisms; and gray = other. Data visualization was done with DNAPlotter (version 10.2) (Carver et al. 2009).

detected in strains isolated in Minnesota. The strains ISTA-6 and CS86, which contain pCI1 and pCI2, are clonal descendants of CmiATCC 10253 maintained in different culture collections. Thus, CmiATCC 10253, the type strain we obtained from ATCC, seems to have lost pCI2.

All three plasmids in *C. michiganensis* subsp. *insidiosus* contain genes whose functions are predicted to be involved in conjugal transfer of DNA. For example, CmiR1-1_pCI2_0012 and 0014 (VO01_RS15635 and 15640) encode proteins homologous to

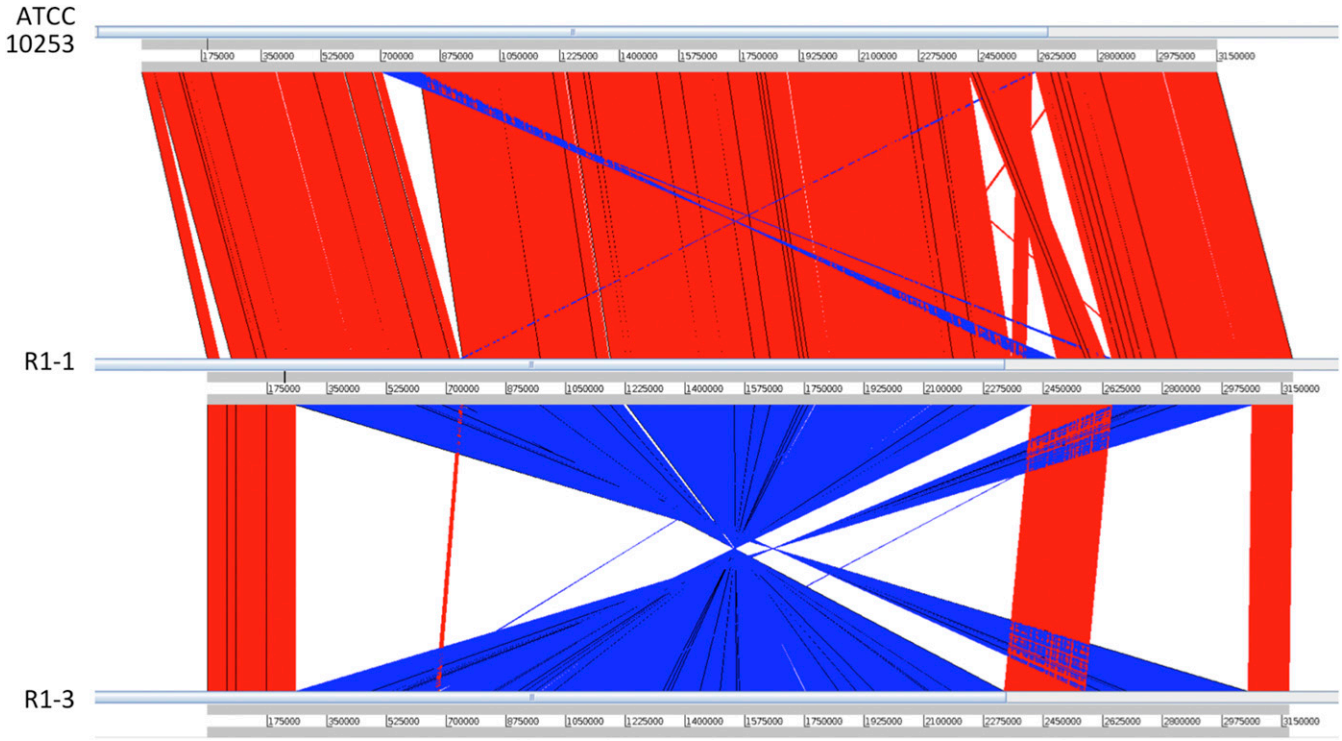

Fig. 2. Alignment of chromosomes of *Clavibacter michiganensis* subsp. *insidiosus* strain CmiATCC 10253, CmiR1-1, and CmiR1-3. The start point of the chromosomal sequences of CmiATCC 10253 (top), CmiR1-1 (middle), and CmiR1-3 (bottom) are set at the start codon of *dnaA*. Sequences of IS*1122* in CmiR1-1 are removed to avoid alignments between copies of IS*1122*. Homologous regions larger than 1,000 bp in collinear or inverted direction are indicated by red or blue lines, respectively. Visualization of the alignment was performed with the Artemis Comparison Tool (version 13.0.0) (Carver et al. 2005).

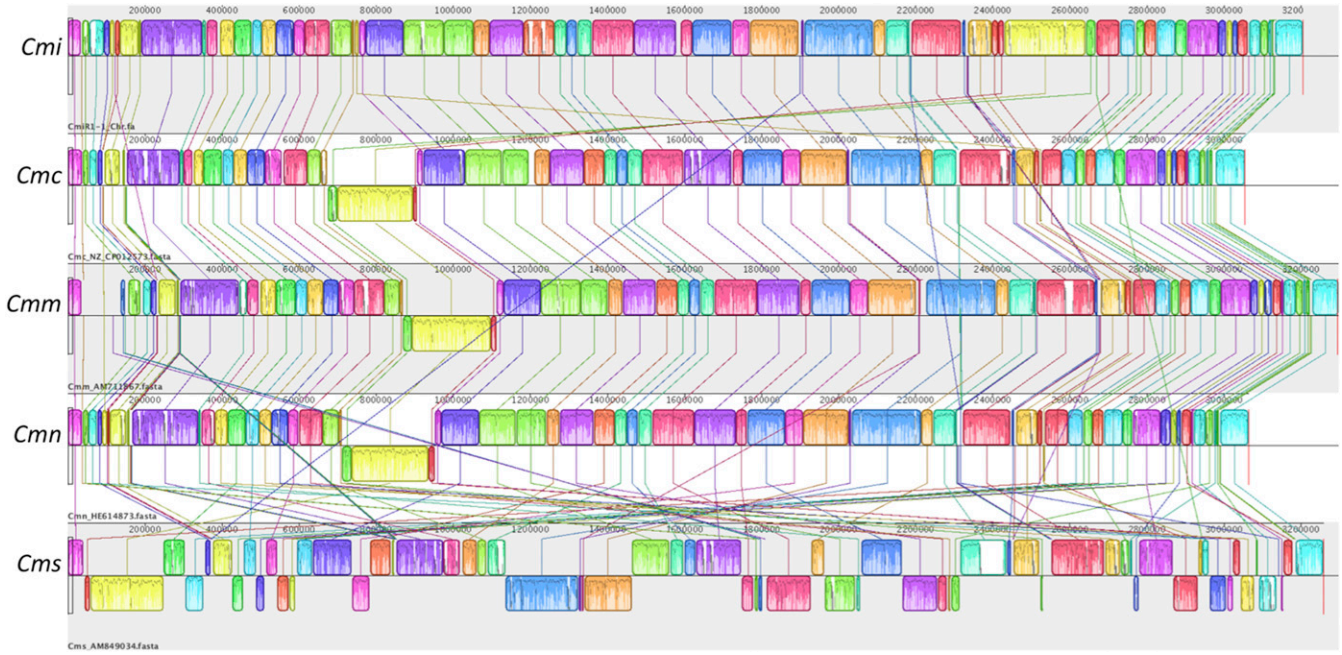

Fig. 3. Alignment of the chromosome of *Clavibacter michiganensis* subsp. *insidiosus* with four other subspecies. The chromosomal sequences are shown as horizontal lines. From top to bottom: CmiR1-1 (as reference), CmcPF008, CmmNCPPB 382, CmnNCPPB 2581, and CmsATCC 33113. Blocks with matching colors and connected with lines are considered locally collinear blocks (LCB). LCB appearing above each horizontal line are in the same orientation as the reference genome, and the ones below each horizontal line are in the reverse orientation relative to the reference genome. The height of the plots within each LCB depicts the level of sequence homology relative to the reference genome. Visualization of the alignment was done using Mauve (version 2.4.0) (Darling et al. 2004).

TrwC/TraI and TraG, respectively. The product of CmiR1-1_pCI3_0072 (VO01_RS16185) appears to be a TrwC-like relaxase fused with an AAA domain. Several genes on pCI3 are similar to sequences in the pathogenicity island (PAI) in the genome of CmmNCPPB 382. For example, CmiR1-1_pCI3_0001 (VO01_RS15830) and CmiR1-1_pCI3_0003 (VO01_RS15840) are homologous to *parX* (CMM_0066) and its neighboring gene encoding a putative ATPase (CMM_0067), respectively. In addition, the *C. michiganensis* subsp. *insidiosus* homologs of *ppaB1/B2* (CmiR1-1_pCI3_0056; VO01_RS16105) and pectate lyase *pelA1/A2* (CmiR1-1_pCI3_0057; VO01_RS16110) are organized the same way as in the *C. michiganensis* subsp. *michiganensis* PAI. There are several plasmid-encoded genes in the chromosome, including copies of the transposase encoded in the IS*1122* transposon. In addition to the transposons, for example, VO01_RS15550 on pCI1 is highly similar to VO01_RS00250 on the chromosome. However, the majority of the genes encoded on plasmids do not have homologs on the chromosome.

Genetic disruptions associated with insertion elements and prophage in *C. michiganensis* subsp. *insidiosus*. The insertion element IS*1122* was detected in all three *C. michiganensis* subsp. *insidiosus* genomes, and its copy number varied from 22 copies in CmiR1-3 to 38 copies in CmiATCC 10253 (Table 2). The 22 copies in CmiR1-3 are all present in the same genomic context as in CmiR1-1, whereas only 16 of the 22 are conserved in CmiATCC 10253. Therefore, only a subset of the insertion positions of IS*1122* are conserved among the three genomes, suggesting that this transposon remains active.

TABLE 3. Clusters of orthologous groups (COG) categories of predicted genes in *Clavibacter michiganensis* subsp. *insidiosus* strain CmiR1-1[a]

COG categories and descriptions	Number of genes[b]
Information storage and processing	
[J] Translation, ribosomal structure and biogenesis	172
[A] RNA processing and modification	0
[K] Transcription	197
[L] Replication, recombination and repair	85
[B] Chromatin structure and dynamics	0
Cellular processes and signaling	
[D] Cell cycle control, cell division, chromosome partitioning	18
[Y] Nuclear structure	0
[V] Defense mechanisms	37
[T] Signal transduction mechanisms	80
[M] Cell wall/membrane/envelope biogenesis	123
[N] Cell motility	10
[Z] Cytoskeleton	0
[W] Extracellular structures	20
[U] Intracellular trafficking, secretion, and vesicular transport	18
[O] Posttranslational modification, protein turnover, chaperones	79
[X] Mobilome: prophages, transposons	1
Metabolism	
[C] Energy production and conversion	97
[G] Carbohydrate transport and metabolism	225
[E] Amino acid transport and metabolism	172
[F] Nucleotide transport and metabolism	64
[H] Coenzyme transport and metabolism	116
[I] Lipid transport and metabolism	91
[P] Inorganic ion transport and metabolism	100
[Q] Secondary metabolites biosynthesis, transport and catabolism	58
Poorly characterized	
[R] General function prediction only	177
[S] Function unknown	34

[a] COG annotation was performed using the National Center for Biotechnology Information Web CD-Search Tool with the COG database (version 3.16) at the default E-value cut-off (0.01).

[b] Some genes are associated with more than one COG category and are counted in both.

Although most of the insertion sites of IS*1122* are in intergenic regions, insertions into an operon or into the coding region of a gene do occur. For example, one copy of IS*1122* (CmiR1-1_0252; VO01_RS01325) is located between two genes encoding a sugar ABC transporter substrate-binding protein and a sugar ABC transporter permease within one operon in CmiR1-1. The insertion may not completely block the expression of the downstream genes within the operon but it may reduce the abundance of the proteins encoded by the downstream genes. The same insertion exists in the genome of CmiR1-3 but not in CmiATCC 10253. In another case, one copy of IS*1122* (CmiR1-1_0372; VO01_01945) was inserted in the middle of a gene encoding a phosphoketolase, resulting in disruption of the open reading frame. This was only observed in the genome of CmiR1-1, because the corresponding phosphoketolase genes in CmiR1-3 and CmiATCC 10253, CmiR1-3_2696 (BEH61_13735) and CmiATCC_0345 (BEH62_01740), respectively, remain intact.

The chromosome of CmiATCC 10253 has an approximately 35-kb deletion, corresponding to base pair position 34,802 to 69,531 on chromosome of CmiR1-1 and 34,763 to 69,505 on chromosome of CmiR1-3, replaced by one copy of IS*1122*, which may be the cause of the large deletion. The genes within the 35-kb region are from CmiR1-1_0031 (VO01_RS00180) to CmiR1-1_0060 (VO01_RS16240) (Table 5). Most of these genes are present in the other *C. michiganensis* genomes, excluding the possibility that this region was specifically acquired by CmiR1-1 and CmiR1-3.

The genomes of *C. michiganensis* subsp. *insidiosus* contain regions that resemble a prophage or remnant of previous phage insertion. The first region is a 10-kb low G+C region bordered by CmiR1-1_0856 (VO01_RS04425) and CmiR1-1_0865 (VO01_RS04470), containing three genes likely of phage origin, including a DNA replication initiation protein, a XerC-like integrase, and a PinR-like recombinase. The two genes flanking each border of the putative prophage, CmiR1-1_0855 (VO01_RS04420) and CmiR1-1_0866 (VO01_RS04475), both encode a NAD(P)H-dependent aldo-keto reductase, with approximately 85% nucleotide sequence similarity, which might be a duplication created during the phage insertion. The second phage-related region is between CmiR1-1_1150 (VO01_RS05925) and CmiR1-1_1163 (VO01_RS05990), containing genes encoding a phage tail tape measure protein and a DNA recombinase. This phage insertion might have disrupted one gene originally encoding a Rossmann fold nucleotide-binding protein, DprA, because the two flanking genes can be aligned to the N and C terminus of CMN_01351, respectively. Interestingly, this phage insertion is present in the genomes of CmiR1-1 and CmiATCC 10253 but not in CmiR1-3, and the *DprA* homologous gene is intact in CmiR1-3. Another possible phage insertion region conserved in all three genomes is bounded by CmiR1-1_1847 (VO01_RS09565) and CmiR1-1_1850 (VO01_RS09580), and the DNA sequence within has weak similarity to the coding sequence of a viral tegument protein. This potential phage insertion has possibly caused the deletion of three genes (a pair of two-component sensor kinase and response regulators and a D-alanyl-D-alanine carboxypeptidase) within a region that is otherwise collinear with the genomes of CmnNCPPB 2581 and CmmNCPPB 382.

Genes in *C. michiganensis* subsp. *insidiosus* known to be important in virulence of other subspecies. The known virulence factors of *C. michiganensis* are all secreted proteins. Therefore, we performed functional annotations on the predicted secretome of CmiR1-1 (Supplementary Table S2). Secreted serine proteases belonging to both the *pat-1* and *ppa* families have been shown to be important for colonization and/or symptom development for *C. michiganensis* subsp. *michiganensis* and *C. michiganensis* subsp. *sepedonicus* (Eichenlaub and Gartemann 2011). The genome of CmiR1-1 contains four homologs of *pat-1* and five homologs of *ppa* family genes (Fig. 5). Based on phylogenetic analyses of their amino acid sequences, all of these homologs in *C. michiganensis* subsp. *insidiosus* are divergent from members of the *pat-1* and *ppa* families known to be important in virulence on tomato or potato such as Pat-1, ChpC, PpaA, and PpaC in *C. michiganensis* subsp.

michiganensis (Burger et al. 2005; Dreier et al. 1997; Eichenlaub and Gartemann 2011) and Chp-7 in *C. michiganensis* subsp. *sepedonicus* (Nissinen et al. 2009). However, CmiR1-1_pCI1_0036 (VO01_RS15527), CmiR1-1_pCI3_0055 (VO01_RS16100), and CmiR1-1_pCI3_0056 (VO01_RS16105) clustered closely with ChpE, PpaJ, and PpaB1/2 of *C. michiganensis* subsp. *michiganensis*, respectively (Fig. 5). Although the exact roles of the three serine proteases in *C. michiganensis* subsp. *michiganensis* are not clear, each was detected in xylem sap of infected tomato and was highly induced during infection or by culturing in medium supplemented with tomato sap (Savidor et al. 2012), suggesting an involvement in interactions with hosts.

Certain cell-wall-degrading enzymes are secreted proteins known or predicted to affect virulence of *C. michiganensis* subsp. *michiganensis* and *C. michiganensis* subsp. *sepedonicus*. All three genomes of *C. michiganensis* subsp. *insidiosus* encode homologs of two cellulases, CelA and CelB, present in *C. michiganensis* subsp. *michiganensis* and *C. michiganensis* subsp. *sepedonicus* (Jahr et al. 2000; Laine et al. 2000). CmiR1-1_pCI1_0025 (VO01_RS15470) encodes a homolog of CelA, with 89% amino acid identity, and CmiR1-1_ 2201 (VO01_RS11370) encodes a homolog of CelB, with 91% amino acid identity. All three *C. michiganensis* subsp. *insidiosus* genomes encode a putative pectinesterase (CmiR1-1_0242; VO01_RS16260), which seems to be a fusion between its homologs CMN_00249 and CMN_00250 in CmnNCPPB 2581. Pectinesterase is not found in the genomes of *C. michiganensis* subsp. *michiganensis* or *C. michiganensis* subsp. *sepedonicus.* The plasmid pCI3 also encodes one pectate lyase gene (CmiR1-1_pCI3_0057; VO01_RS16110), homologous to *pelA1* and *pelA2* in *C. michiganensis* subsp. *michiganensis*. Although none of the *C. michiganensis* subsp. *insidiosus* genomes contains a functional copy of *xysA*, the genomes of CmiR1-1 and CmiR1-3 contain one gene (CmiR1-1_1455; VO01_RS07515) homologous to *xysB* of *C. michiganensis* subsp. *michiganensis* and *C. michiganensis* subsp. *nebraskensis*, encoding a β-xylanase. Other potential secreted cell-wall-degrading enzymes present in the genomes of *C. michiganensis* subsp. *insidiosus* are a polygalacturonase (CmiR1-1_2859; VO01_RS14785) and an endoglucanase (CmiR1-1_2678; VO01_RS13855).

Comparative genomic analyses of five *C. michiganensis* subspecies. To identify genes that may have functions in specific host–pathogen interactions at the subspecies level, the predicted protein sequences from the genomes of CmiR1-1, CmmNCPPB 382, CmsATCC 33113, CmnNCPPB 2581, and CmcPF008 were analyzed. In total, 4,805 clusters of orthologous genes were identified. Although the majority of the clusters contain genes that are single copy, some clusters contain genes that appear to be paralogs. For example, one cluster encoding subtilisin-like serine proteases consisted of three copies in CmmNCPPB 382 and two copies in each of the other analyzed genomes, for a total of 11 genes in the cluster.

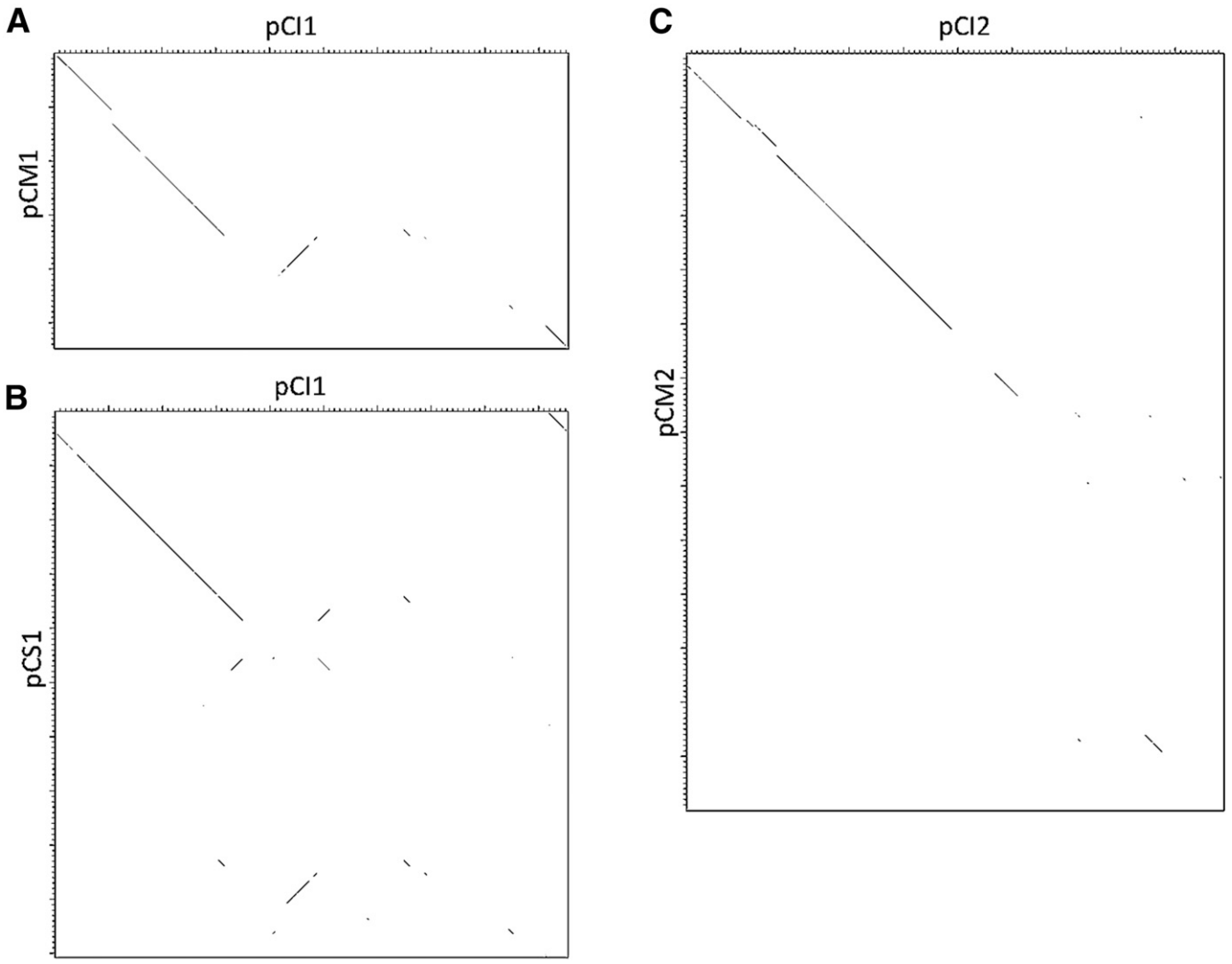

Fig. 4. Dot-plot of alignments of plasmids in the genomes of CmiR1-1, CmmNCPPB 382, and CmsATCC 33113. **A,** pCI1 (x-axis) versus pCM1 (y-axis); **B,** pCI1 (x-axis) versus pCS1 (y-axis); and **C,** pCI2 (x-axis) versus pCM2 (y-axis). Visualization of the alignments was done using Dotter (version 4.34.3) (Sonnhammer and Durbin 1995) with sliding window size set to 200. Major ticks on the axis represent 5,000 bp and minor ticks represent 500 bp.

TABLE 4. Plasmid distribution among strains of *Clavibacter michiganensis* subsp. *insidiosus*[a]

Strain	Alternative strain designation	Location	Year isolated	pCI1	pCI2	pCI3
ATCC 10253[b]		Kansas	1955	+	–	–
ISTA-1	LNPV 0.51	Europe	NA	+	–	–
ISTA-3	NCPPB 1110	United States	1943	+	–	–
ISTA-6	CFBP 2404[c]	United States	1955	+	+	–
048A		California	2012	+	–	+
114		California	2012	+	–	+
253		California	2012	+	–	–
CS80	DAR26781	New South Wales, Australia	1973	+	–	–
CS82	DAR34844	New South Wales, Australia	1980	+	–	+
CS83	DAR35660a	New South Wales, Australia	1981	+	+	–
CS84	DAR26653/PDDCC4190	New Zealand	1975	+	+	–
CS85	DAR26656/PDDCC3570	New Zealand	1972	+	+	–
CS86	DAR26658/PDDCC2621[c]	United States	1955	+	+	–
CS87	DAR26661/PDDCC2613	United Kingdom	1964	+	+	–
C1-5	…	Minnesota	1994	+	-	+
C1-6	…	Minnesota	1994	+	+	+
C1-7	…	Minnesota	1994	+	+	+
C1-9	…	Minnesota	1994	+	+	+
ND2-1	…	Minnesota	1995	+	+	+
N3-3	…	Minnesota	1995	+	+	–
N1313	…	Minnesota	1995	+	–	+
FL1-20	…	Minnesota	1995	+	–	+
ND1-6	…	Minnesota	1995	+	–	+
ND3163	…	Minnesota	1995	+	–	+
ND3142B	…	Minnesota	1995	+	–	+
ND1-3	…	Minnesota	1995	+	+	+
FL1-3	…	Minnesota	1995	+	+	+
N214	…	Minnesota	1995	+	+	+
R1-1	…	Minnesota	2009	+	+	+
R1-2	…	Minnesota	2009	+	+	+
R1-3	…	Minnesota	2009	+	+	+
R1-4	…	Minnesota	2009	+	+	+
R1-8	…	Minnesota	2009	+	+	+

[a] Symbols: + and – indicate detected and not detected, respectively, with polymerase chain reaction; NA = not available.
[b] Type strain of *C. michiganensis* subsp. *insidiosus*, obtained from the American Type Culture Collection.
[c] Type strain of *C. michiganensis* subsp. *insidiosus*, clone of ATCC 10253.

TABLE 5. Chromosomal genes in a 35-kb region present in *Clavibacter michiganensis* subsp. *insidiosus* strains CmiR1-1 and CmiR1-3 and absent in strain CmiATCC 10253[a]

Locus in CmiR1-1	Locus in CmiR1-3	Identified domains	Predicted gene function
CmiR1-1_0031; VO01_RS00180	CmiR1-3_0031; BEH61_00170	pfam12697	α/β Hydrolase family protein
CmiR1-1_0032; VO01_RS00185	CmiR1-3_0032; BEH61_00175	pfam12681	Glyoxalase-like domain-containing protein
CmiR1-1_0033; VO01_RS00190	CmiR1-3_0033; BEH61_00180	pfam01261	Xylose isomerase-like protein
CmiR1-1_0034; VO01_RS00195	CmiR1-3_0034; BEH61_00185	pfam01408; pfam02894	Oxidoreductase
CmiR1-1_0035; VO01_RS00200	CmiR1-3_0035; BEH61_00190	pfam12697	α/β Hydrolase family protein
CmiR1-1_0036; VO01_RS00205	CmiR1-3_0036; BEH61_00195	pfam13646	HEAT repeats containing protein
CmiR1-1_0037; VO01_RS00210	CmiR1-3_0037; BEH61_00200	pfam13411	MerR HTH family regulatory protein
CmiR1-1_0038; VO01_RS00215	CmiR1-3_0038; BEH61_00205	pfam01738	Dienelactone hydrolase family protein
CmiR1-1_0039; VO01_RS00220	CmiR1-3_0039; BEH61_00210	pfam00440	TetR family regulatory protein
CmiR1-1_0040; VO01_RS00225	…	pfam02682	ATP-dependent urea carboxylase
CmiR1-1_0041; VO01_RS00235	CmiR1-3_0040; BEH61_00215	ND	Hypothetical protein
CmiR1-1_0042; VO01_RS00240	CmiR1-3_0041; BEH61_00220	ND	Hypothetical protein
CmiR1-1_0043; VO01_RS00245	CmiR1-3_0042; BEH61_00225	ND	Hypothetical protein
CmiR1-1_0044; VO01_RS00250	CmiR1-3_0043; BEH61_00230	ND	Hypothetical protein
CmiR1-1_0046; VO01_RS00260	CmiR1-3_0045; BEH61_00240	pfam00501; pfam00550; pfam07993	Long-chain acyl-CoA synthetases with a putative dehydrogenase domain
CmiR1-1_0047; VO01_RS00265	CmiR1-3_0046; BEH61_00245	ND	Putative ketoreductase, subgroup 1
CmiR1-1_0048; VO01_RS00270	CmiR1-3_0047; BEH61_00250	pfam13561	3-oxoacyl-ACP reductase
CmiR1-1_0049; VO01_RS00275	CmiR1-3_0048; BEH61_00255	ND	Hypothetical protein
CmiR1-1_0050; VO01_RS00280	CmiR1-3_0049; BEH61_00260	ND	Hypothetical protein
CmiR1-1_0051; VO01_RS00285	CmiR1-3_0050; BEH61_00265	ND	Hypothetical protein
CmiR1-1_0052; VO01_RS00290	CmiR1-3_0051; BEH61_00270	pfam01470	Pyroglutamyl peptidase
CmiR1-1_0053; VO01_RS00295	CmiR1-3_0052; BEH61_00275	pfam01494	Monooxygenase
CmiR1-1_0055; VO01_RS00305	CmiR1-3_0054; BEH61_00285	ND	Hypothetical protein
VO01_RS16230	CmiR1-3_0055; BEH61_00290	ND	Hypothetical protein
CmiR1-1_0056; VO01_RS00315	CmiR1-3_0056; BEH61_00295	ND	Hypothetical protein
VO01_RS00320	CmiR1-3_0057; BEH61_00300	pfam00561	α/β Hydrolase family protein
CmiR1-1_0057; VO01_RS00325	CmiR1-3_0058; BEH61_00305	ND	Hypothetical protein
CmiR1-1_0058; VO01_RS00330	CmiR1-3_0059; BEH61_00310	pfam00202	Adenosylmethionine-8-amino-7-oxononanoate aminotransferase
CmiR1-1_0059; VO01_RS16235	CmiR1-3_0060; BEH61_00315	pfam00465	Iron-containing alcohol dehydrogenase
CmiR1-1_0060; VO01_RS16240	CmiR1-3_0061; BEH61_00320	pfam13649	Methyltransferase
VO01_RS16245	CmiR1-3_0062; BEH61_00325	cd05709	Zinc metallorotease

[a] Symbols: … for locus indicates a pseudogene and ND = not detected.

The distribution of the 4,833 clusters among the five genomes was further dissected and visualized as a Venn diagram (Fig. 6). The intersect of all five genomes consists of a cluster of 1,917 genes, which forms the core genome of the five analyzed subspecies. There are 379 clusters unique to CmiR1-1, with 298 clusters consisting of chromosomal genes and 81 consisting of plasmid-encoded genes. Colonies of *C. michiganensis* subsp. *insidiosus* are often associated with indigo pigments, a unique characteristic within *C. michiganensis*. An 8.5-kb region consisting of seven genes that may be involved in the biosynthesis of indigoidine was identified among the *C. michiganensis* subsp. *insidiosus*-unique clusters. This region is conserved among the three sequenced strains. These seven genes include one MarR family transcriptional regulator (CmiR1-1_1123; VO01_RS05795), and the remaining six are organized as an operon (Fig. 7). CmiR1-1_1122 (VO01_RS05790) encodes a pseudouridine-5′-phosphate glycosidase and is homologous to *Dickeya dadantii indA*, with 64.7% amino acid sequence similarity. CmiR1-1_1121 (VO01_RS05785) encodes a phosphoglycolate phosphatase and is similar to *D. dadantii indB*, with 57.4% similarity. CmiR1-1_1120 (VO01_RS05780) encodes a phosphoribosyl transferase. CmiR1-1_1119 (VO01_RS05775) encodes a transporter, which may export the blue pigment. CmiR1-1_1118 (VO01_RS05770) encodes a nonribosomal peptide synthetase (NRPS) of a single module type, similar to *indC* in *D. dadantii* and *bpsA* in *Streptomyces lavendulae*, consisting of an N-terminal adenylation domain, an oxidase domain, a center-located acyl carrier domain, and a C-terminal thioesterase domain. CmiR1-1_1117 (VO01_RS05765) encodes a phosphopantetheinyl transferase with the center region showing 47.6% similarity to *Vogesella indigofera igiA*.

Many of the genes on pCI3 are unique to *C. michiganensis* subsp. *insidiosus* among the sequenced genomes of *C. michiganensis*. CmiR1-1_pCI3_0015 to -0023 (VO01_RS15910 to -15950) encode genes likely involved in inositol metabolism, and are homologous

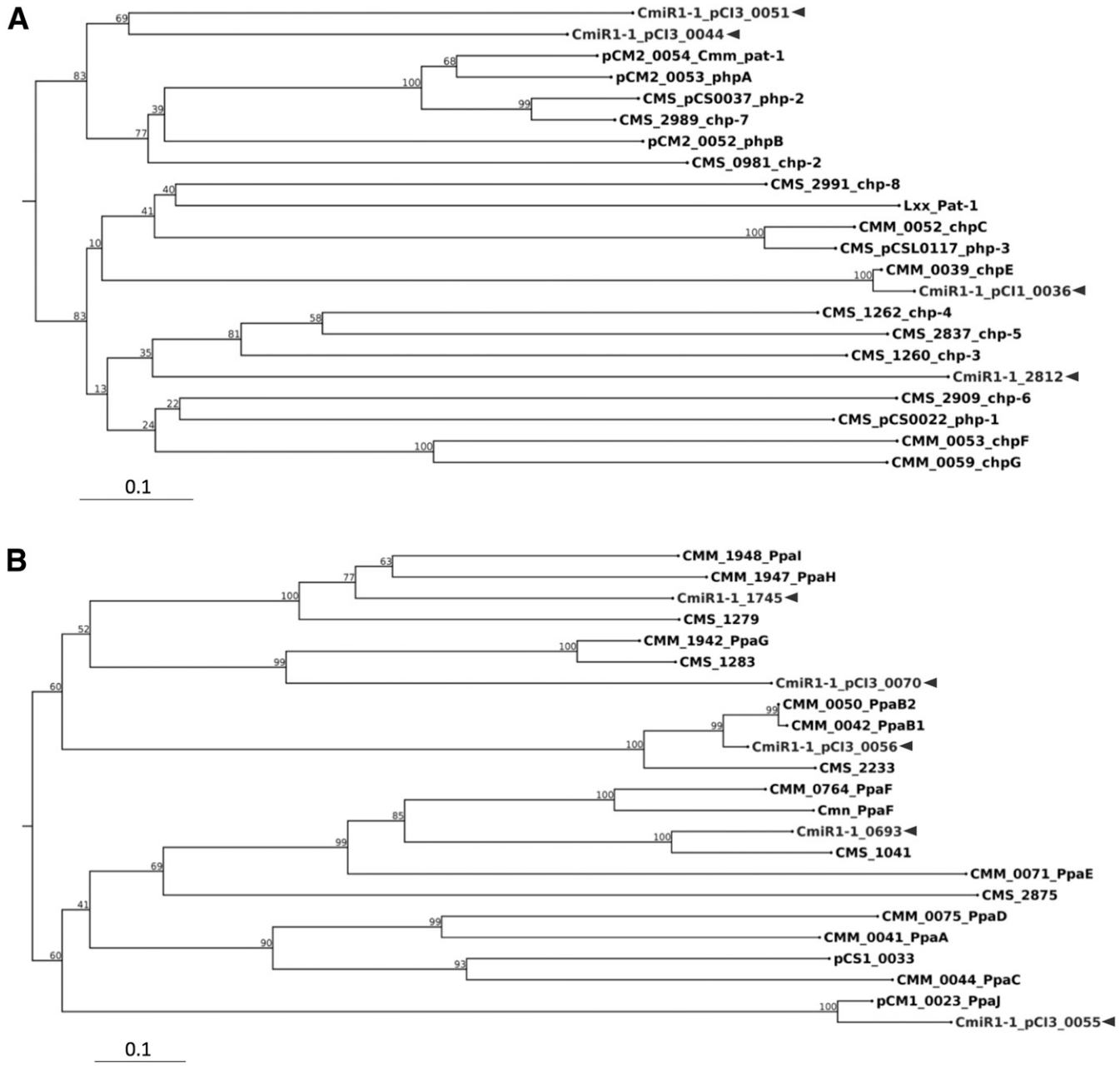

Fig. 5. Maximum-likelihood phylogenetic trees of the *pat-1* family and *ppa* family serine proteases. The full-length amino acid sequences of the gene products of **A,** the *pat-1* family and **B,** the *ppa* family were aligned with Multiple Sequence Comparison by log-expectation. Phylogenetic trees were then created using the neighbor-joining method with the WAG protein substitution model and 1,000× bootstrap analysis in CLC Main Workbench (version 7.6). Percentage bootstrap values are shown at the branching positions. The *Clavibacter michiganensis* subsp. *insidiosus* gene products are highlighted with triangles.

to *iolT*, *iolJ*, *iolC*, *iolA*, *iolB*, *iolD*, *iolE*, *iolG*, and *iolH* of *Bacillus subtilis*, respectively. These genes are transcribed in the same direction and, therefore, may be organized as one operon. No transcription factors were found in the upstream region of these genes, although one GntR family transcriptional regulator (CmiR1-1_pCI3_0025; VO01_RS15960) is located approximately 580 bp downstream and is in reverse orientation to the putative *iol* operon.

DISCUSSION

Availability of three complete genome sequences of *C. michiganensis* subsp. *insidiosus* enabled comparative genomic analyses to identify genes potentially involved in the interactions with its legume host, as well as genes involved in general pathogenicity of *C. michiganensis*. This study also identified accessions of *M. truncatula* with differential responses to *C. michiganensis* subsp. *insidiosus.* The bacterium initially colonized roots of the three accessions but was not maintained in roots or shoots of accession A17, suggesting that this accession has effective defenses against the pathogen; whereas, in accessions DZA315.16 and F83005.5, the bacterium continued to colonize the host and caused mild symptoms of disease. Ongoing research with green fluorescent protein-marked strains will help to clarify the population dynamics and movement of the bacteria within the plant at a finer timescale. In contrast to the extensive knowledge of host–pathogen interactions of gram-negative bacteria, little is known about defense responses against gram-positive bacteria. Identification of resistant and susceptible accessions of *M. truncatula*, a model host with extensive genomic resources, and a complete genome sequence of the pathogen, should help accelerate identification of genes important in host–pathogen interactions. Although symptoms of disease in *M. truncatula* were mild, they occurred more rapidly than in susceptible alfalfa plants, which may take 3 months or more to show symptoms of discoloration in the interior tissues of the taproot, leaf yellowing, and stunting. The genome sequences are also valuable for further mining to detect differences among the subspecies of *C. michiganensis* in order to develop molecular diagnostic tools with precise discrimination capabilities.

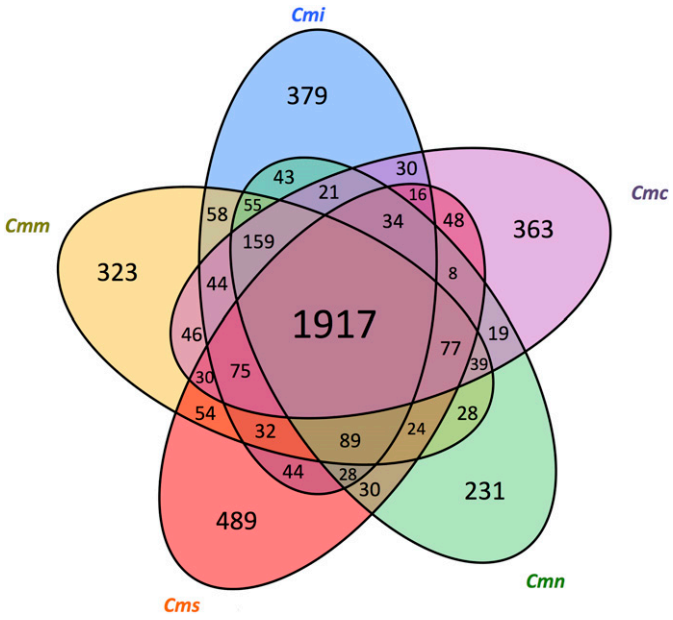

Fig. 6. Venn diagram depicting the number of clusters of orthologous genes among *Clavibacter michiganensis* subsp. *insidiosus, C. michiganensis* subsp. *capsici, C. michiganensis* subsp. *michiganensis, C. michiganensis* subsp. *nebraskensis*, and *C. michiganensis* subsp. *sepedonicus*. Numbers in the Venn diagram represent the number of clusters of orthologous genes analyzed with the software package GET_HOMOLOGS (version 30122016) (Contreras-Moreira and Vinuesa 2013). The input protein sequences from the five genomes are CmiR1-1, CmcPF008, CmmNCPPB 382, CmnNCPPB2581, and CmsATCC33113.

An earlier effort at de novo assembly of the genomes of CmiR1-1, CmiR1-3, and CmiATCC 10253 using paired-end sequencing reads from Illumina GAIIx and Miseq systems resulted in hundreds of contigs (unpublished), which is not sufficient to study the structure of the genomes. To obtain complete genome sequences for the three strains of *C. michiganensis* subsp. *insidiosus*, we utilized PacBio SMRT sequencing for its advantages of long read length and unbiased read quality of GC-rich sequences (Quail et al. 2012). The combination of Illumina sequencing and PacBio SMRT sequencing enabled de novo assembly of three complete genomes of *C. michiganensis* subsp. *insidiosus.*

Previous comparative genomic analyses of *Clavibacter* strains were conducted with one whole-genome sequence representative of each subspecies or phylogenetic group (Bentley et al. 2008; Gartemann et al. 2008; Tambong 2017; Załuga et al. 2014). Although these studies provided valuable insights about each strain, the features of the one particular sequenced genome may not reflect the genome structure and content of the subspecies due to recombination or loss or gain of plasmids and plasmid-encoded genes (Gross et al. 1979). These issues can be addressed with analyses of multiple complete genomes for one subspecies or phylogenetic group. In this study, the characterization of *C. michiganensis* subsp. *insidiosus* based on three complete genomes indeed revealed differences between strains while providing greater confidence for the common features of this subspecies.

The chromosomes of CmiR1-1, CmiR1-3, and CmiATCC 10253 have undergone chromosomal rearrangements likely mediated through homologous recombination between the repetitive sequences of insertion elements. Similar observations have been reported at the subspecies level for *C. michiganensis* (Bentley et al. 2008) but not at the strain level. The variation in IS*1122* copy number and insertion position among the three genomes of *C. michiganensis* subsp. *insidiosus* strongly suggests that this transposon remains active, which could potentially create many sites for recombination, adding flexibilities to this organism that might enable adaptation to new environmental conditions. Assessing genome sequences and pathogenic aggressiveness of natural variants may identify candidate genes involved in pathogenicity.

The genomes of CmiR1-1 and CmiR1-3 possess three different circular plasmids, named pCI1, pCI2, and pCI3, whereas the genome

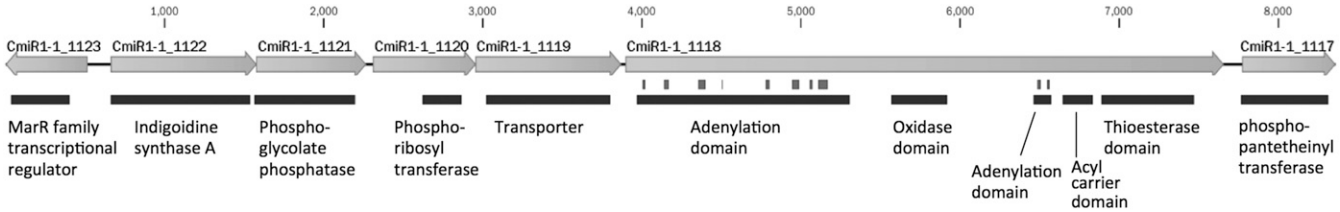

Fig. 7. Cluster of genes in *Clavibacter michiganensis* subsp. *insidiosus* potentially involved in indigoidine biosynthesis. Predicted protein-coding genes are shown as arrows in the direction of transcription. Identified domains or predicted functions are marked in solid black below the genes at corresponding positions. Predicted AMP binding sites and conserved motifs were shown as gray bars above the adenylation domain. The scale on top is in base pairs.

of CmiATCC 10253 only contains pCI1. The heterogeneity of plasmids present in different strains of *C. michiganensis* subsp. *insidiosus* has been documented previously (Gross et al. 1979). In this study, we developed PCR primers based on the genome sequence of CmiR1-1 to specifically detect the presence of pCI1, pCI2, and pCI3 in an additional 30 isolates from various geographic origins. We discovered that pCI1 is highly conserved among isolates. All of the 33 tested strains were positive for pCI1, indicating that this plasmid may provide sufficient fitness advantages to the bacterium to prevent loss, at least in its natural context. Moreover, pCI1 shares high similarities with pCM1 and pCS1, both of which encode *celA* and are required for virulence (Jahr et al. 2000; Laine et al. 2000). Altogether, this could suggest that pCI1, pCM1, and pCS1 were derived from a similar origin and contributed to the evolution of an ancestral *Clavibacter* strain toward a pathogenic lifestyle. On the other hand, although pCI2 and pCI3 are not conserved in all *C. michiganensis* subsp. *insidiosus* strains, some genes encoded by the plasmids may improve fitness or may be involved in pathogenicity. For example, the predicted inositol metabolism operon encoded by pCI3 likely allows *C. michiganensis* subsp. *insidiosus* to utilize inositol, which is present in soil in large quantities, as an alternative carbon source.

Serine proteases play important roles in the pathogenicity of *C. michiganensis*. The genes homologous to the *pat-1* family and *ppa* family are present in the genomes of *C. michiganensis* subsp. *insidiosus.* Some members of these families seem to have nonredundant functions, because single gene mutants of certain homologs led to measurable decreases in virulence or loss of a hypersensitive response on nonhost plants (Nissinen et al. 2009; Stork et al. 2008). The homologs of *pat-1* and *ppa* families in *C. michiganensis* subsp. *insidiosus* are quite divergent from the ones in *C. michiganensis* subsp. *michiganensis* and *C. michiganensis* subsp. *sepedonicus* known for virulence functions (Fig. 5). This suggests that these might have been under selection pressure to target divergent plant proteins and, thus, may play different roles than those known for *C. michiganensis* subsp. *michiganensis* and *C. michiganensis* subsp. *sepedonicus*. These genes are not clustered in the genome of *C. michiganensis* subsp. *insidiosus*, in contrast to the organization of *chp* genes in the *C. michiganensis* subsp. *michiganensis* PAI (Gartemann et al. 2008). Several of these homologs, together with the pectate lyase gene, are located on pCI3 in a pattern similar to the PAI in *C. michiganensis* subsp. *michiganensis*. Given that the *C. michiganensis* subsp. *michiganensis* PAI and pCI3 were both likely acquired through horizontal gene transfer, it is possible that they originated from a similar, yet-to-be-discovered source. It further raises an interesting possibility that these genes may exist in relatively high frequency in the genomes of microorganisms that associate with plants and coordinately function in promoting microbial fitness during interactions with plants.

Pectinesterase catalyzes the de-esterification of pectin, which is a major component of the plant cell wall and synthesized in highly methyl-esterified forms. It is thought that de-esterified pectin is more accessible to pectin-degrading enzymes, including pectin lyases and polygalacturonases (Willats et al. 2001). The simultaneous presence of a pectinesterase (CmiR1-1_0242; VO01_RS16260), a polygalacturonase (CmiR1-1_2859; VO01_RS14785), and a pectate lyase (CmiR1-1_pCI3_0057; VO01_RS16110) in *C. michiganensis* subsp. *insidiosus* suggests that it may be more efficient in degrading pectin than *C. michiganensis* subsp. *michiganensis* and *C. michiganensis* subsp. *sepedonicus*, which lack pectinesterases. The *C. michiganensis* subsp. *insidiosus* pectate lyase, which is encoded on pCI3, shares 94.0% amino acid sequence identity with *C. michiganensis* subsp. *michiganensis* PelA1, which is encoded on the pathogenicity island. Interestingly, *C. michiganensis* subsp. *nebraskensis* possesses a different type of pectate lyase (CMN_02654), indicating that it was acquired from a different origin than the ones in *C. michiganensis* subsp. *insidiosus* and *C. michiganensis* subsp. *michiganensis*.

The two previously described transcriptional regulators, Vatr1 and Vatr2 (Savidor et al. 2014), involved in pathogenicity of *C. michiganensis* subsp. *michiganensis* are highly conserved among all five analyzed genomes. Therefore, although mutants with defects in either gene result in decreased levels of virulence gene induction without affecting bacterial growth in culture (Savidor et al. 2014), it is unlikely that Vatr1 and Vatr2 specifically contribute to pathogenicity of *C. michiganensis* subsp. *michiganensis*. It is possible that other genes functioning downstream of Vatr1 or Vatr2 play more direct roles in regulating expression of virulence genes. For example, CMM_0089, which encodes a CatR LacI-family transcriptional regulator, and CMM_0678, which encodes a PknC family serine/threonine-protein kinase, were both downregulated by Vatr1 when *C. michiganensis* subsp. *michiganensis* was grown in an infection-mimicking medium (Savidor et al. 2014), and are only found in the genome of *C. michiganensis* subsp. *michiganensis*.

The blue pigmentation of *C. michiganensis* subsp. *insidiosus* colonies is a distinctive feature among *C. michiganensis* subspecies and is caused by production of a water-insoluble extracellular indigoidine (Starr 1958). Gene clusters involved in biosynthesis of indigoidine have also been reported in *D. dadantii* (previously *Erwinia chrysanthemi*) (Reverchon et al. 2002), *S. lavendulae* (Takahashi et al. 2007), *S. chromofuscus* (Yu et al. 2013), and *V. indigofera* (unpublished, GenBank AF088856). Indigoidine is synthesized by the single NRPS gene in each of these identified gene loci. These NRPS, including CmiR1-1_1118 (VO01_RS05770) (Fig. 7), all share the same set of modules and have an oxidation domain integrated into the N-terminal adenylation domain (Takahashi et al. 2007). However, organization of the gene clusters varies among different bacterial species. The gene cluster in *C. michiganensis* subsp. *insidiosus* is most similar to the one in *S. chromofuscus* (Yu et al. 2013), both containing one gene encoding a transmembrane transporter and one gene encoding a phosphoribosyl transferase. In addition, *C. michiganensis* subsp. *insidiosus* has one phosphoglycolate phosphatase gene in this operon. The production of indigoidine could confer tolerance to oxidative stresses and has been shown to promote virulence of *D. dadantii* (Reverchon et al. 2002). It is possible that indigoidine production in *C. michiganensis* subsp. *insidiosus* also provides similar advantages, and may be a contributing factor in the virulence of *C. michiganensis* subsp. *insidiosus* in its hosts.

Loss of virulence after passage in culture has been documented for *C. michiganensis* subsp. *insidiosus* (Carroll and Lukezic 1971). The basis for attenuation remains unknown. Our study was not designed to investigate this phenomenon; however, results from the comparative genomic analyses provide a basis for future hypothesis-driven studies. In particular, we found that (i) pCI2 was absent in the type strain, CmiATCC 10253 obtained from ATCC, but present in CmiATCC 10253 from two other sources; (ii) pCI3 was absent in all three representatives of CmiATCC 10253; (iii) CmiATCC 10253 was less virulent in *M. truncatula* relative to CmiR1-3; and (iv) the genome of CmiATCC 10253 harbors apparent deletions and rearrangements relative to CmiR1-1. Future studies on pathogenicity, similar to those conducted with *C. michiganensis* subsp. *michiganensis* and *C. michiganensis* subsp. *sepedonicus* (Eichenlaub and Gartemann 2011), are needed to advance an understanding of the phenomenon of attenuation in *C. michiganensis* subsp. *insidiosus* and to formulate generalized models of pathogenicity and subspeciation in *C. michiganensis*.

ACKNOWLEDGMENTS

This project was supported by funds from the Minnesota Agricultural Experiment Station/Hatch and National Science Foundation grant IOS-1353854 awarded to J. Glazebrook and the United States Department of Agriculture (USDA) Agricultural Research Service Current Research Information System project 5062-12210-002D. Y. Lu was supported by a Monsanto/University of Minnesota Multi-functional Agriculture Initiative Graduate Student Fellowship from Monsanto Co. We thank K. Ophel-Keller, D. Caffier, and D. Opgenorth for providing strains of *C. michiganensis* subsp. *insidiosus*; M. Dornbusch for excellent technical help; the University of Minnesota Genomics Center (Saint Paul, MN) for performing the genome sequencing with MiSeq; the Molecular Biology Core at Mayo Clinic

(Rochester, MN) for performing the PacBio sequencing; and the Minnesota Supercomputing Institute at the University of Minnesota (Saint Paul, MN) for facilitating the data analysis. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. USDA is an equal opportunity provider and employer.

LITERATURE CITED

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kubal, M., Meyer, F., Olsen, G. J., Olson, R., Osterman, A. L., Overbeek, R. A., McNeil, L. K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G. D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., and Zagnitko, O. 2008. The RAST Server: Rapid annotations using subsystems technology. BMC Genomics 9:75.

Bae, C., Oh, E. J., Lee, H. B., Kim, B. Y., and Oh, C. S. 2015. Complete genome sequence of the cellulase-producing bacterium *Clavibacter michiganensis* PF008. J. Biotechnol. 214:103-104.

Bentley, S. D., Corton, C., Brown, S. E., Barron, A., Clark, L., Doggett, J., Harris, B., Ormond, D., Quail, M. A., May, G., Francis, D., Knudson, D., Parkhill, J., and Ishimaru, C. A. 2008. Genome of the actinomycete plant pathogen *Clavibacter michiganensis* subsp. *sepedonicus* suggests recent niche adaptation. J. Bacteriol. 190:2150-2160.

Burger, A., Grafen, I., Engemann, J., Niermann, E., Pieper, M., Kirchner, O., Gartemann, K. H., and Eichenlaub, R. 2005. Identification of homologues to the pathogenicity factor Pat-1, a putative serine protease of *Clavibacter michiganensis* subsp. *michiganensis*. Microbiol. Res. 160:417-427.

Carroll, R. B., and Lukezic, F. L. 1971. Preservation of *Corynebacterium insidosum* in a sterile soil mix without loss of virulence. Phytopathology 61: 688-690.

Carver, T., Thomson, N., Bleasby, A., Berriman, M., and Parkhill, J. 2009. DNAPlotter: Circular and linear interactive genome visualization. Bioinformatics 25:119-120.

Carver, T. J., Rutherford, K. M., Berriman, M., Rajandream, M. A., Barrell, B. G., and Parkhill, J. 2005. ACT: The Artemis Comparison Tool. Bioinformatics 21:3422-3423.

Chin, C. S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E. E., Turner, S. W., and Korlach, J. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat. Methods 10:563-569.

Contreras-Moreira, B., and Vinuesa, P. 2013. GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. Appl. Environ. Microbiol. 79:7696-7701.

Darling, A. C., Mau, B., Blattner, F. R., and Perna, N. T. 2004. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 14:1394-1403.

De Boer, S. H., and Copeman, R. J. 1980. Bacterial ring rot testing with the indirect fluorescent antibody staining procedure. Am. Potato J. 57:457-465.

Dreier, J., Meletzus, D., and Eichenlaub, R. 1997. Characterization of the plasmid encoded virulence region pat-1 of phytopathogenic C*lavibacter michiganensis* subsp. *michiganensis*. Mol. Plant-Microbe Interact. 10:195-206.

Edgar, R. C. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792-1797.

Eichenlaub, R., and Gartemann, K. H. 2011. The *Clavibacter michiganensis* subspecies: Molecular investigation of gram-positive bacterial plant pathogens. Annu. Rev. Phytopathol. 49:445-464.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., Bibillo, A., Bjornson, K., Chaudhuri, B., Christians, F., Cicero, R., Clark, S., Dalal, R., Dewinter, A., Dixon, J., Foquet, M., Gaertner, A., Hardenbol, P., Heiner, C., Hester, K., Holden, D., Kearns, G., Kong, X., Kuse, R., Lacroix, Y., Lin, S., Lundquist, P., Ma, C., Marks, P., Maxham, M., Murphy, D., Park, I., Pham, T., Phillips, M., Roy, J., Sebra, R., Shen, G., Sorenson, J., Tomaney, A., Travers, K., Trulson, M., Vieceli, J., Wegener, J., Wu, D., Yang, A., Zaccarin, D., Zhao, P., Zhong, F., Korlach, J., and Turner, S. 2009. Real-time DNA sequencing from single polymerase molecules. Science 323:133-138.

Galperin, M. Y., Makarova, K. S., Wolf, Y. I., and Koonin, E. V. 2015. Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Res. 43:D261-D269.

Gartemann, K. H., Abt, B., Bekel, T., Burger, A., Engemann, J., Flugel, M., Gaigalat, L., Goesmann, A., Grafen, I., Kalinowski, J., Kaup, O., Kirchner, O., Krause, L., Linke, B., McHardy, A., Meyer, F., Pohle, S., Ruckert, C., Schneiker, S., Zellermann, E. M., Puhler, A., Eichenlaub, R., Kaiser, O., and Bartels, D. 2008. The genome sequence of the tomato-pathogenic actinomycete *Clavibacter michiganensis* subsp. *michiganensis* NCPPB382 reveals a large island involved in pathogenicity. J. Bacteriol. 190:2138-2149.

Gonzalez, A. J., and Trapiello, E. 2014. *Clavibacter michiganensis* subsp. *phaseoli*, a new subspecies pathogen in bean. Int. J. Syst. Evol. Microbiol. 64:1752-1755.

Gross, D. C., Vidaver, A. K., and Keralis, M. B. 1979. Indigenous plasmids from phytopathogenic *Corynebacterium* species. J. Gen. Microbiol. 115: 479-489.

Jahr, H., Dreier, J., Meletzus, D., Bahro, R., and Eichenlaub, R. 2000. The endo-beta-1,4-glucanase CelA of *Clavibacter michiganensis* subsp. *michiganensis* is a pathogenicity determinant required for induction of bacterial wilt of tomato. Mol. Plant-Microbe Interact. 13:703-714.

Kirchner, O., Gartemann, K. H., Zellermann, E. M., Eichenlaub, R., and Burger, A. 2001. A highly efficient transposon mutagenesis system for the tomato pathogen *Clavibacter michiganensis* subsp. *michiganensis*. Mol. Plant-Microbe Interact. 14:1312-1318.

Kouchi, H., Imaizumi-Anraku, H., Hayashi, M., Hakoyama, T., Nakagawa, T., Umehara, Y., Suganuma, N., and Kawaguchi, M. 2010. How many peas in a pod? Legume genes responsible for mutualistic symbioses underground. Plant Cell Physiol. 51:1381-1397.

Kristensen, D. M., Kannan, L., Coleman, M. K., Wolf, Y. I., Sorokin, A., Koonin, E. V., and Mushegian, A. 2010. A low-polynomial algorithm for assembling clusters of orthologous groups from intergenomic symmetric best matches. Bioinformatics 26:1481-1487.

Laine, M. J., Haapalainen, M., Wahlroos, T., Kankare, K., Nissinen, R., Kassuwi, S., and Metzler, M. C. 2000. The cellulase encoded by the native plasmid of *Clavibacter michiganensis* ssp. *sepedonicus* plays a role in virulence and contains an expansin-like domain. Physiol. Mol. Plant Pathol. 57: 221-233.

Langmead, B., and Salzberg, S. L. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9:357-359.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078-2079.

Lu, Y., Samac, D. A., Glazebrook, J., and Ishimaru, C. A. 2015. Complete genome sequence of *Clavibacter michiganensis* subsp. *insidiosus* R1-1 using PacBio Single-Molecule Real-Time technology. Genome Announc. 3: e00396-15.

Marefat, A., Ophel-Keller, K., and McKay, A. 2007. A real-time PCR assay for detection of *Clavibacter michiganensis* subsp. *insidiosus* in lucerne. Australas. Plant Pathol. 36:262-269.

Meletzus, D., Bermphol, A., Dreier, J., and Eichenlaub, R. 1993. Evidence for plasmid-encoded virulence factors in the phytopathogenic bacterium *Clavibacter michiganensis* subsp. *michiganensis* NCPPB382. J. Bacteriol. 175: 2131-2136.

Nissinen, R., Xia, Y., Mattinen, L., Ishimaru, C. A., Knudson, D. L., Knudson, S. E., Metzler, M., and Pirhonen, M. 2009. The putative secreted serine protease Chp-7 is required for full virulence and induction of a nonhost hypersensitive response by *Clavibacter michiganensis* subsp. *sepedonicus*. Mol. Plant-Microbe Interact. 22:809-819.

Oh, E.-J., Bae, C., Lee, H.-B., Hwang, I. S., Lee, H.-I., Yea, M. C., Yim, K.-O., Lee, S., Heu, S., Cha, J.-S., and Oh, C.-S. 2016. *Clavibacter michiganensis* subsp. *capsici* subsp. nov., causing bacterial canker disease in pepper. Int. J. Syst. Evol. Microbiol. 66:4065-4070.

Peltier, G. L., and Schroeder, F. R. 1932. The nature of resistance in alfalfa to wilt (*Aplanobacter insidiosum* L. Mc.). Nebr. Agric. Exp. Stn. Res. Bull. Vol. 63. University of Nebraska, College of Agriculture, Agricultural Experiment Station.

Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. 2011. SignalP 4.0: Discriminating signal peptides from transmembrane regions. Nat. Methods 8:785-786.

Quail, M., Smith, M., Coupland, P., Otto, T., Harris, S., Connor, T., Bertoni, A., Swerdlow, H., and Gu, Y. 2012. A tale of three next generation sequencing platforms: Comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics 13:341.

Reverchon, S., Rouanet, C., Expert, D., and Nasser, W. 2002. Characterization of indigoidine biosynthetic genes in *Erwinia chrysanthemi* and role of this blue pigment in pathogenicity. J. Bacteriol. 184:654-665.

Saitou, N., and Nei, M. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4:406-425.

Samac, D. A., Nix, R. J., and Oleson, A. E. 1998. Transmission frequency of *Clavibacter michiganensis* subsp. *insidiosus* to alfalfa seed and identification of the bacterium by PCR. Plant Dis. 82:1362-1367.

Sanger, F., Nicklen, S., and Coulson, A. R. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463-5467.

Savidor, A., Chalupowicz, L., Teper, D., Gartemann, K. H., Eichenlaub, R., Manulis-Sasson, S., Barash, I., and Sessa, G. 2014. *Clavibacter michiganensis* subsp. *michiganensis* Vatr1 and Vatr2 transcriptional regulators are required for virulence in tomato. Mol. Plant-Microbe Interact. 27: 1035-1047.

Savidor, A., Teper, D., Gartemann, K. H., Eichenlaub, R., Chalupowicz, L., Manulis-Sasson, S., Barash, I., Tews, H., Mayer, K., Giannone, R. J., Hettich, R. L., and Sessa, G. 2012. The *Clavibacter michiganensis* subsp.

michiganensis-tomato interactome reveals the perception of pathogen by the host and suggests mechanisms of infection. J. Proteome Res. 11: 736-750.

Sonnhammer, E. L., and Durbin, R. 1995. A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. Gene 167:GC1-GC10.

Starr, M. 1958. The blue pigment of *Corynebacterium insidiosum*. Arch. Microbiol. 30:325-334.

Stork, I., Gartemann, K. H., Burger, A., and Eichenlaub, R. 2008. A family of serine proteases of *Clavibacter michiganensis* subsp. *michiganensis*: chpC plays a role in colonization of the host plant tomato. Mol. Plant Pathol. 9: 599-608.

Takahashi, H., Kumagai, T., Kitani, K., Mori, M., Matoba, Y., and Sugiyama, M. 2007. Cloning and characterization of a *Streptomyces* single module type non-ribosomal peptide synthetase catalyzing a blue pigment synthesis. J. Biol. Chem. 282:9073-9081.

Tambong, J. T. 2017. Comparative genomics of *Clavibacter michiganensis* subspecies, pathogens of important agricultural crops. PLoS One 12: e0172295.

Tatusov, R. L., Galperin, M. Y., Natale, D. A., and Koonin, E. V. 2000. The COG database: A tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. 28:33-36.

Tivoli, B., Baranger, A., Sivasithamparam, K., and Barbetti, M. J. 2006. Annual Medicago: From a model crop challenged by a spectrum of necrotrophic pathogens to a model plant to explore the nature of disease resistance. Ann. Bot. (Lond.) 98:1117-1128.

Vidaver, A. K. 1967. Synthetic and complex media for the rapid detection of fluorescence of phytopathogenic pseudomonads: Effect of the carbon source. Appl. Microbiol. 15:1523-1524.

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., and Earl, A. M. 2014. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9:e112963.

Willats, W. G. T., Orfila, C., Limberg, G., Buchholt, H. C., van Alebeek, G.-J. W. M., Voragen, A. G. J., Marcus, S. E., Christensen, T. M. I. E., Mikkelsen, J. D., Murray, B. S., and Knox, J. P. 2001. Modulation of the degree and pattern of methyl-esterification of pectic homogalacturonan in plant cell walls: Implications for pectin methyl esterase action, matrix properties, and cell adhesion. J. Biol. Chem. 276:19404-19413.

Wilson, K. 2001. Preparation of genomic DNA from bacteria. Unit 2.4, pages 2.4.1-2.4.5 in: Current Protocols in Molecular Biology. John Wiley & Sons Hoboken, NJ.

Yasuhara-Bell, J., and Alvarez, A. M. 2015. Seed-associated subspecies of the genus *Clavibacter* are clearly distinguishable from *Clavibacter michiganensis* subsp. *michiganensis*. Int. J. Syst. Evol. Microbiol. 65:811-826.

Yu, D., Xu, F., Valiente, J., Wang, S., and Zhan, J. 2013. An indigoidine biosynthetic gene cluster from *Streptomyces chromofuscus* ATCC 49982 contains an unusual IndB homologue. J. Ind. Microbiol. Biotechnol. 40:159-168.

Załuga, J., Stragier, P., Baeyen, S., Haegeman, A., Van Vaerenbergh, J., Maes, M., and De Vos, P. 2014. Comparative genome analysis of pathogenic and non-pathogenic *Clavibacter* strains reveals adaptations to their lifestyle. BMC Genomics 15:392.