e-**X**tra*

Genetic Characterization of *Clavibacter michiganensis* subsp. *michiganensis* Population in Turkey

Yusuf Sen, Wageningen University and Research, Plant Breeding, 6700 AJ Wageningen, The Netherlands; **Yesim Aysan,** Cukurova University, Faculty of Agriculture, Department of Plant Protection, 01330 Adana, Turkey; **Mustafa Mirik,** Namik Kemal University, Department of Plant Protection, TR-59030 Tekirdag, Turkey; **Duygu Ozdemir** and **Fien Meijer-Dekens,** Wageningen University and Research, Plant Breeding, 6700 AJ Wageningen, The Netherlands; **Jan M. van der Wolf,** Wageningen University and Research, Bio-interactions and Plant Health, 6700 AB, Wageningen, The Netherlands; and **Richard G. F. Visser** and **Sjaak van Heusden,**[†] Wageningen University and Research, Plant Breeding, 6700 AJ Wageningen, The Netherlands

Abstract

The pathogenic gram-positive bacterium *Clavibacter michiganensis* subsp. *michiganensis* (Smith) Davis et al. is the most harmful bacterium to tomatoes in many countries with a cooler climate. Multilocus sequence analysis was performed on five housekeeping genes (*bipA*, *gyrB*, *kdpA*, *ligA*, and *sdhA*) and three virulence-related genes (*ppaA*, *chpC*, and *tomA*) to determine evolutionary relationships and population structure of 108 *C. michiganensis* subsp. *michiganensis* strains collected from Turkey between 1996 and 2012. Based on these analyses, we concluded that *C. michiganensis* subsp. *michiganensis* in Turkey is highly uniform. However, at least four novel *C. michiganensis* subsp. *michiganensis* strains were recently introduced, possibly at the beginning of the 1990s. The singletons might point to additional sources or to strains that have evolved locally in Turkey.

Clavibacter michiganensis subsp. *michiganensis* (Smith) Davis et al. is a gram-positive bacterium belonging to the Actinobacteria phylum. *C. michiganensis* subsp. *michiganensis* causes one of the most dangerous bacterial diseases of tomato: bacterial canker (Basim et al. 2004; Şahin et al. 2002; Sen et al. 2012). *C. michiganensis* subsp. *michiganensis* is a quarantine organism in the European Union and some other countries (Anonymous 2000; Eichenlaub et al. 2006). Turkey is a major tomato producing country and is third in the world after the U.S.A. and China. Outbreaks of bacterial canker regularly take place in Turkey with high costs for growers (Basim et al. 2004; Baysal et al. 2011; Şahin et al. 2002).

C. michiganensis subsp. *michiganensis* is a genetically and phenotypically diverse subspecies (Eichenlaub et al. 2006). Its virulence level and ability to spread throughout the ecosystem is variable (Yazdankhah et al. 2004). It can survive for more than 2 years in crop residues on the soil surface (Chang et al. 1992; Gleason et al. 1993; Sen et al. 2015). Trevors and Finnen (1990) found that survival potential can be strain-dependent. Long distance transmission of the bacteria occurs mainly via seed or planting material. A few infected seeds (under 5 per 10,000 seeds) can result in a local outbreak (Chang et al. 1991).

Multilocus sequence typing (MLST) is a method that makes it possible to characterize bacterial species using sequences of six to 10 housekeeping genes (Maiden 2006) and is used in epidemiological studies (Jones et al. 2003). The accumulation of mutations in housekeeping genes is a relatively slow process, and this makes it a valuable tool for studying genetic relationships between strain collections on a global level (Enright and Spratt 1999). In *C. michiganensis* subsp. *michiganensis*, differences in virulence have been assessed in conjunction with an MLST analysis of housekeeping genes (Jacques et al. 2012; Milijašević-Marčić et al. 2012; Quesada-Ocampo et al. 2012; Waleron et al. 2011). However, the most suitable genes for studying the degree of clonality have not yet been determined.

In this study, we studied MLST profiles of 108 strains originating from Turkey to measure the degree of clonality using core and flexible sequences. A network was constructed that identified the relationships between our collection and global strains. Additionally, the selective forces that affect the diversification of the *C. michiganensis* subsp. *michiganensis* population were determined.

Materials and Methods

Bacterial strains. *C. michiganensis* subsp. *michiganensis* strains (*n* = 378) were collected between 1996 and 2012 from different parts of Turkey. Of those, 108 strains representing different years and locations were selected for further studies. Sixteen strains from other countries, hereafter called "global strains," were included. Global strains included one pathogenic strain from the Wageningen UR collection, four pathogenic strains from Israel (Kleitman et al. 2008), seven Serbian strains (one from each of the seven described MLST groups) (Milijašević-Marčić et al. 2012), and reference strain NCPPB382 and NCPPB2979 strain (NCBI database). *C. michiganensis* subsp. *sepedonicus* (*Cms*) and *C. michiganensis* subsp. *nebraskensis* (*Cmn*) were used as outgroups.

DNA isolation. The Qiagen QIAamp DNA Mini Kit was used for genomic DNA isolation from the bacteria. The purity and quantity of DNA were measured by Nanodrop spectrophotometer analysis and agarose gel electrophoresis.

Strain identification tests. Pathogenicity measurements and strain identification were performed on the 108 strains from Turkey. Pathogenicity was tested using a tomato-stem inoculation assay, a tomato-cotyledon leaf assay, and measurement of hypersensitive reactivity in *Mirabilis jalapa*. The strains were identified with Gram staining, an oxidase test, ELISA, and PCR with *C. michiganensis* subsp. *michiganensis*-specific primers.

For the stem inoculation tests, three replicates of five tomato plants of young tomato seedlings (*Solanum lycopersicum* Mill 'H2274') at 3 to 5 true leaf stage were inoculated with 100 μl of bacterial suspension into the stem (10^8 cfu/ml). An injection of sterile distilled water was the negative control. After inoculation, tomato plants were covered with clear polyethylene bags for 24 h at 25°C. After removing the bags, the plants were maintained in a controlled climate at 25°C at 70% relative humidity (RH) and a 16/8-h day/night cycle. Disease development was evaluated 8 to 10 days after inoculation, and reisolation of strains was carried out with diseased material (EPPO 2013). For the tomato cotyledon leaf tests, three replicates of five plants of freshly germinated 'H2274' tomato seedlings were inoculated by putting the

tip of a cotton swab, dipped in the bacterial suspension (10^8 cfu/ml), on the cotyledon leaves. After inoculation, the tomato seedlings were incubated in the same controlled climate as described above. Disease development was evaluated 3 to 4 days after inoculation (Lelliott and Stead 1987). Wilting symptoms were recorded using the following scale: 0 = no disease symptom; 1 = 1 to 25% wilting and considered low susceptibility; 2 = 26 to 50% wilting and considered medium susceptibility; 3 = 51 to 75% wilting considered high susceptibility; and 4 = 76 to 100% wilting and considered very high susceptibility.

The Gram reaction (Klement et al. 1990), oxidase reaction, and measurements of hypersensitivity to tobacco (Schaad et al. 2001) and *Mirabilis jalapa* leaves (Gitaitis 1990) were replicated three times.

To confirm the identity of *C. michiganensis* subsp. *michiganensis* strains, an indirect ELISA method was used based on a *C. michiganensis* subsp. *michiganensis*-specific monoclonal antibody (BRA 44001, Agdia; Coligan et al. 1991). The strain identity was further confirmed by a PCR assay using the *C. michiganensis* subsp. *michiganensis*-specific primers *C. michiganensis* subsp. *michiganensis* 5 and *C. michiganensis* subsp. *michiganensis* 6 (Pat-1 gene) according to Dreier et al. (1995).

Gene selection. Five housekeeping genes were chosen, namely *bipA* encoding the GTP-binding typA/bipA-like protein, *gyrB* encoding the DNA gyrase subunit B, *kdpA* encoding the potassium-transporting ATPase subunit A, *ligA* encoding the NAD-dependent DNA ligAse, and *sdhA* encoding the succinate dehydrogenase flavoprotein subunit (Milijašević-Marčić et al. 2012). In addition to these sequences, three pathogenicity-related genes were selected: *ppaA*, which encodes a putative extracellular serine protease; *chpC*, which encodes a serine protease; and *tomA*, which encodes tomatinase, an endo-1,4-beta-glycosidase (Kleitman et al. 2008). Primer sets for each gene were chosen based on previous studies (Table 1). After PCR amplification (Applied Biosystems GeneAmp 2700 system), the fragments were sequenced. Chromatograms were analyzed using LASERGENE DNAStarSeqMan Pro version (DNAStar Inc.). Mega5 (Tamura et al. 2011) was used to align sequences using the ClustalW algorithm and were then further manually edited. The borders of sequences were trimmed according to the coding region for each gene using the sequence of the reference strain NCPPB382 (NCBI database), hereafter referred to simply as "Reference."

Statistical analysis of data. The GC content, total number of segregating sites (i.e., polymorphic sites), nucleotide diversity, number of haplotypes, haplotype diversity, and minimum number of recombination events were calculated using DnaSP version 5.10 (Rozas et al. 2003). The synonymous/nonsynonymous mutation ratio was calculated using the Tajima D model (Tajima 1989). Three different tests were performed to measure the diversity of genes: the Tajima D, Fu and Li's D, and Fu and Li's F outgroup tests. To estimate the effect of mutation versus recombination events, a plot of the nucleotide diversity versus the haplotype diversity across all loci was constructed.

Allele assignments were made using the MLST nonredundant database (NRDB), which determined an allele profile (i.e., sequence type [ST]) for each strain. STs were named according to strain number. Clonal complex structure analysis of STs was performed using the eBURST v3 program (Feil et al. 2004). Clonal complexes were named according to the founder ST. The eBURST analysis detects single-locus variants (SLV) and double-locus variants (DLV). Bootstrap analysis with 1,000 replicates was performed to assess the ancestral and subgroup founder ST in a clonal complex. Strains that do not fit any ST or eBURST grouping are considered as singletons (S).

The eBURST grouping was used to identify groups of related STs (clonal complexes), in which all assigned group members have identical alleles for at least six of the eight loci tested and have at least one nonidentical allele. Polymorphisms between a clonal complex founder and its SLVs were checked to determine whether the difference was due to a single nucleotide change, in which case the polymorphism was designated a point mutation. When the difference was due to more than one nucleotide change, the polymorphism was designated a recombination event. Recombination among loci was estimated using the Maynard Smith method (Smith et al. 1993) implemented in the Start2 package (Jolley et al. 2001).

Split network analysis was carried out by the SplitTree 4.9 program (Huson and Bryant 2006) using neighbor-net analysis with the Jukes Cantor distance correction method.

Population structure analysis was investigated using a Bayesian clustering approach implemented in the STRUCTURE 2.3.4 software program (Pritchard et al. 2000). The program was run using an admixture model with a burn-in period of 30,000 iterations followed by 300,000 Markov Chain Monte Carlo (MCMC) repeats. The optimal number of populations (K) was set at 1 to 10 with 10 replications for each K. The LOCPRIOR model (Hubisz et al. 2009) was implemented and strains were grouped according to overlapping year and location, which was used as the prior grouping for structure analysis. The final number of subgroups was assessed according to L(K) and Δ(K), an ad hoc quantity related to the second-order rate change of the log probability of data with respect to the number of clusters (Evanno et al. 2005).

Maximum likelihood (ML) analysis was performed using RAxML 8.0 (Stamatakis 2014) using the CIPRES gateway platform (Miller et al. 2010). Concatenated data for genes were obtained using Mesquite 2.74 (Maddison and Maddison 2010). Then, the data were transformed to PHYLIP format for ML analysis. Phylogenetic analysis was performed for individual genes, concatenated housekeeping genes, concatenated pathogenicity-related genes, and concatenated total genes. Concatenated gene data sets were handled using a gene partition concept. The missing values were treated as missing characters in the phylogenetic analysis. Tree visualization was achieved with the FigTree v1.4.0 program.

All data, including sequence information and allele profile for each strain, are available in the PAMDB (www.pamdb.org) database (Almeida et al. 2010).

Results

Identification of strains. All 108 strains from Turkey were pathogenic to tomato (both stem as cotyledon assay) (Supplementary Table S1). They all induced a hypersensitive response on *Mirabilis japala* and were positive in ELISA using *C. michiganensis* subsp. *michiganensis*-specific monoclonal antibodies and gave positive PCR results using the *C. michiganensis* subsp. *michiganensis* 5-*C. michiganensis* subsp. *michiganensis* 6 primers (Dreier et al. 1995). These tests confirmed the identity of all isolates as *C. michiganensis* subsp. *michiganensis*.

Sequence polymorphism. The overall size of the concatenated sequence of the eight genes was 4,472 base pairs. Polymorphisms were found in all genes with the lowest values in *tomA* and the highest in *sdhA* and *kdpA*. The number of alleles ranged from 10 to 18 per locus. Recombination events were detected within housekeeping genes but none were detected in pathogenicity-related genes (data not shown). The GC content was higher in the housekeeping genes than in the pathogenicity-related genes (Table 2). An uneven positive correlation

Table 1. Primers used to amplify parts of the selected genes

Gene	Primers	References
TomA	F:CGAACTCGACCAGGTTCTCG R: GGTCTCACGATCGGATCC	Kleitman et al. 2008
ChpC	F:GCTCTTGGGCTAATGGCCG R:GTCAGTTGTGGAAGATGCTG	Kleitman et al. 2008
PpaA	F:CATGATATTGGTGGGGAAAG R: CCCCGTCTTTGCAAGACC	Kleitman et al. 2008
LigA	F: GTTCGACGAGCTGAATGC R: CTCGACCTTCTCCATGAC	Milijašević-Marčić et al. 2012
BipA	F:GCATGATGGACTCGAACG R:GATCTTCACGTTCTTGACG	Milijašević-Marčić et al. 2012
SdhA	F:CCTGGATGTTCGTGTACC R:GAGGACATGGAGTTCTTC	Milijašević-Marčić et al. 2012
KdpA	F:GTGCAGAACTTCGTCTCGG R:GAGCATCATGTTGATCATCG	Milijašević-Marčić et al. 2012
GyrB	F:GACATCCAGATCACCATC R:GCTGATCTTCTTGACCGTG	Milijašević-Marčić et al. 2012

was detected between haplotype diversity and nucleotide diversity as an indicative of the evolutionary mechanism (Fig. 1). Three independent neutrality tests showed significantly negative values for pathogenicity related genes (Table 2). The Ka/Ks ratio (the non-synonymous to synonymous substitution ratio) was higher for *ppaA*, *chpC*, and *tomA* than for the housekeeping genes, except for *kdpA* (Table 3).

Sequence type analysis. To determine the clonality of the population, two ST analyses were carried out: the clonal complex and eBURST grouping methods. Three of the 108 strains had missing data in four or more genes and were therefore not included in the ST analysis. Forty-three strains were found to be identical to at least one other strain, and all but one of the identical strains were excluded from ST analysis. Among the 105 strains, 62 unique STs were identified.

Clonal complex analysis grouped 62 STs (considering SLVs) into nine different clonal complexes and 24 singletons (Fig. 2A). Clonal complexes were named according to the name of the founder strain. Strains not belonging to any clonal complex were designated as singletons and indicated with S. Allelic differences between clonal complex founders and their satellites (SLVs and DLVs) were inspected. At least 14 recombination events and 10 mutation events were detected. Recombination among loci resulted in an index of association value (I_A) of 0.87, which indicates significant linkage disequilibrium among loci (*P* = 0.000).

The eBURST group analysis using the criterion of six of eight shared alleles (instead of seven shared alleles, which was used for clonal complex analysis) indicated the presence of two major and two minor groups within the population (Fig. 2B). Twelve strains did not fit in any group and were therefore considered singletons. Several clonal complexes ended up into group 1, whereas other groups (groups 2, 3, and 4) were based on individual clonal complexes. Group 1 was the largest and comprised 66 strains. Groups 2, 3, and 4 comprised 15, 10, and two strains, respectively. Upon considering the year and location of the collected strains, representative members of group 1 were isolated in different years and from a variety of locations. Representatives of group 2 were initially discovered in 1996, but until 2004, no other members of this group were found. Members of group 3 were only detected in the Erdemli-Mersin region of Turkey in 2010. The members of group 4, which was distinctly separated from the other groups in the ML analysis, were only isolated in 2010 in the Tokat region. Singletons were mainly observed in the strains that were collected after 2004 in several locations.

Population structure. To understand the presence of subpopulations within the whole population, structure analysis was performed on the 62 isolates from Turkey (excluding identical strains and strains with missing data). The outgroup strains, *Cms* and *Cmn*, were added as controls. Based on the L(K) and Δ(K) method, two major structures were identified (Fig. 3), in which outgroup strains represent one group and all the Turkish strains represent the second group. When we used only data of housekeeping genes instead of concatenated data, group 2 strains could be separated from other Turkish strains and structured into different subpopulations.

Table 3. Types of polymorphisms in the eight studied genes

Gene	Total segregating sites	Parsimony informative sites	Synonymous changes	Non-synonymous changes	Ka/Ks ratio
TomA	10	2	5	6	0.132
ChpC	34	3	10	15	0.119
PpaA	24	5	7	18	6.737
LigA	18	6	2	17	0.098
BipA	30	22	20	10	0.056
SdhA	34	32	36	1	0.038
KdpA	53	21	29	22	0.350
GyrB	20	10	6	10	0.037

Table 2. Sequence variation and neutrality tests of all genes used in this study[a]

Gene	FS (bp)	GC (%)	NH	HD	MR	ND	K	Tajima's D[b]	Fu and Li's D[b]	Fu and Li's F[b]
TomA	471	61	10	0.612	0	0.0017	0.778	−1.531	−4.432**	−4.060**
ChpC	564	53	10	0.184	0	0.0013	0.729	−2.751***	−7.627**	−6.712**
PpaA	519	58	12	0.711	0	0.0024	1.246	−2.248**	−8.168**	−7.057**
LigA	468	72	10	0.537	2	0.0024	1.113	−2.015*	−4.651**	−4.486**
BipA	633	69	10	0.466	1	0.0089	5.595	−0.274	0.948	−0.805
SdhA	658	69	14	0.775	3	0.0174	11.454	1.871	1.294	1.727
KdpA	598	69	12	0.814	7	0.0132	0.013	−1.456	−1.239	−1.270
GyrB	561	69	18	0.663	2	0.0028	1.573	−2.098*	−4.073**	−3.954**

[a] FS: Fragment size, GC: GC content, NH: number of haplotypes, HD: haplotype diversity, MR: minimum recombination events, ND: nucleotide diversity, K: average number of nucleotide differences.
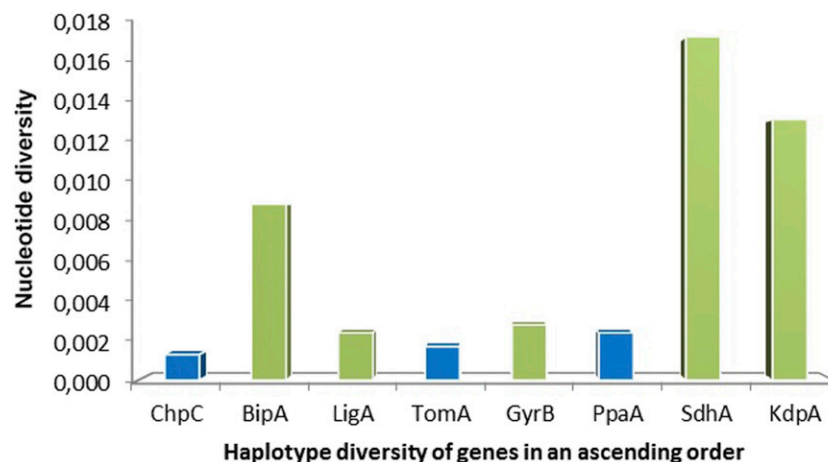[b] *** (*P* < 0.001), **(*P* < 0.02), *(*P* < 0.05).

Fig. 1. Nucleotide diversity versus haplotype diversity of the eight analyzed genes. Pathogenicity-related genes are blue, and housekeeping genes are green.

Reticulation network. To visualize and confirm the presence of recombination within the genes, a network analysis was performed. Reticulation network is a consequence of recombination events. No reticulation was found in the pathogenicity related gene network but it was detected in the housekeeping gene network. In the *bipA* and *gyrB* network, some degree of reticulation was detected in groups 2 and 3 strains (data not shown). An example of reticulation of concatenated data are shown in Figure 4.

Phylogenetic relationships. Phylogenetic signals were checked in the individual gene data, concatenated pathogenicity gene data or housekeeping gene data, and concatenated all loci data. The phylogenetic tree of individual genes resulted in a partly incongruent structure (data not shown). The visibility of clonal complexes at the edges of lineages can be considered as an informative unit of each gene. We observed that more clonal complexes were visible at the edges of lineages in *kdpA* and *sdhA* phylogenetic trees than in other gene trees (data not shown). The non-tomato host group strains (strains from eggplant and outgroup strains) were clearly separated from all *C. michiganensis* subsp. *michiganensis* strains. A phylogenetic

signal from the concatenated pathogenicity-related genes (*chpC*, *tomA*, and *ppaA*) tree was also strong (data not shown). Stronger phylogenetic signals were obtained from the concatenated data of housekeeping genes than from the pathogenicity-related genes. All non-tomato host strains were separated with a very high bootstrap value (data not shown).

For the Turkish strains, the final tree was constructed with concatenated data of all loci using the ML algorithm (Fig. 5). ML analysis was performed on 16 global strains (including *Cms*, *Cmn*, the reference strain NCPPB382, and strain NCPPB2979) and the 62 strains from Turkey. Identical strains were excluded from ML analysis to prevent statistical bias in the bootstrap analysis. The ML tree of the concatenated eight genes showed a separation of the non-tomato *C. michiganensis* subspecies (*Cms* and *Cmn*) and strain 46 (isolated from eggplant) from the other members. Surprisingly, group 4 strains (CC number 92) were also separated from the other *C. michiganensis* subsp. *michiganensis* strains. Major and minor groups (Fig. 2A) detected by eBURST were visible at the edge of lineages (shown as colored groups, Fig. 5).
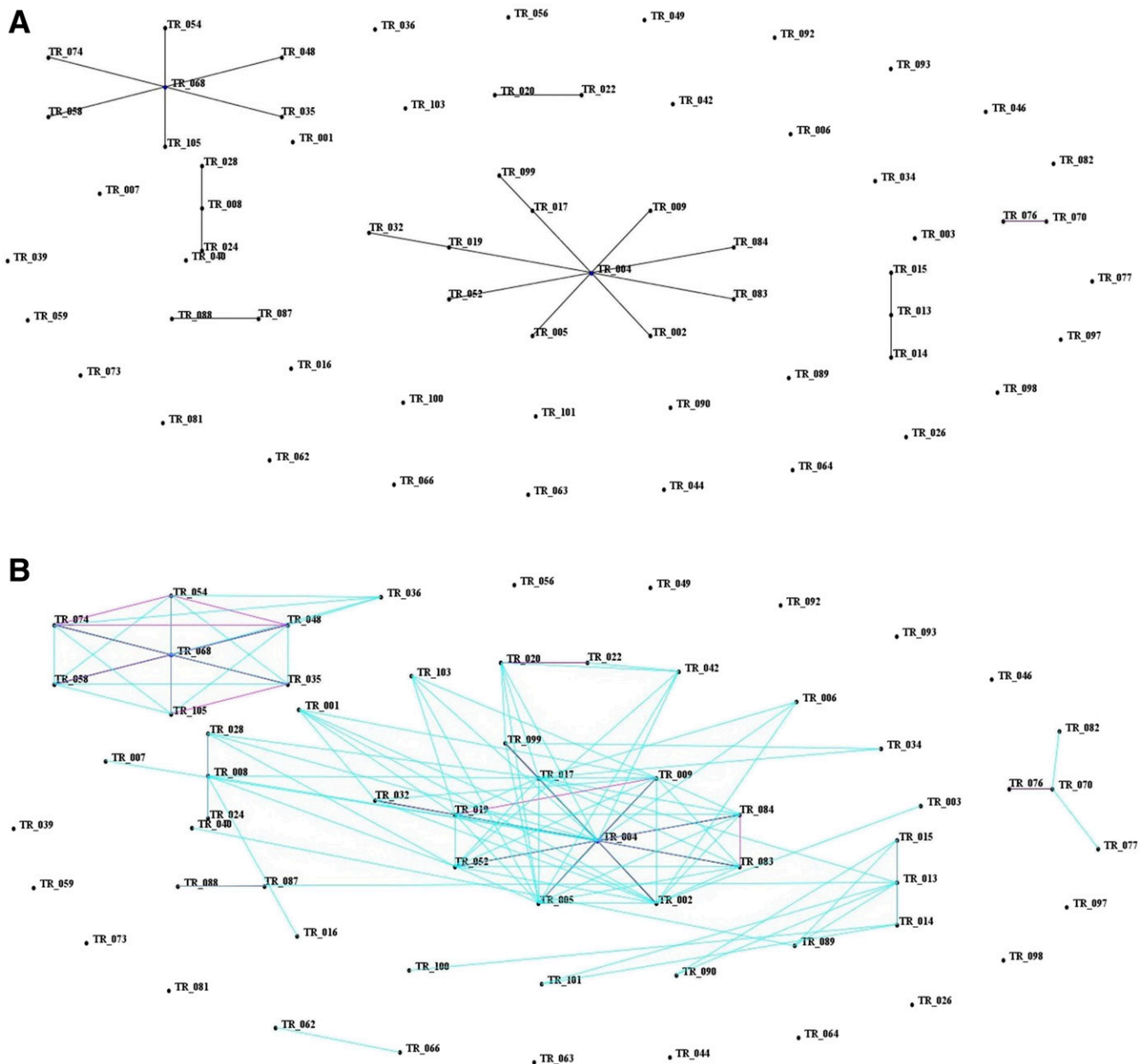

Fig. 2. A, Clonal complex analysis of 62 Turkish and 16 global strains. **B,** eBURST grouping of 62 strains. Pink lines are single-locus variants (SLVs) within clonal complexes, and blue lines indicate double locus variant (DLV) relationships between and within sequence types (STs).

When we examine the relationship between global and Turkish strains, Serbian strains (P121, P520) grouped with NCPPB2979 in group 1. The Serbian strains P64, P137, and P01 grouped with reference strain NCPP382. The remaining Serbian strains P140 and P70 were grouped together within one part of group 1. Those strains were grouped similarly as indicated by the original study. Two strains from Israel (46 and 402) belonged in group 1 together with *C. michiganensis* subsp. *michiganensis* 542. The other strains from Israel (18 and 42) belonged in group 2. Twelve singletons were spread throughout the ML tree and are indicated in red (Fig. 5).

Discussion

Identification of strains. *C. michiganensis* subsp. *michiganensis* comprises strains with different virulence levels (avirulent, virulent, and hyper virulent). In breeding for *C. michiganensis* subsp. *michiganensis* resistance, the screening should be done with virulent or hyper virulent strains. The identity of all *C. michiganensis* subsp. *michiganensis* strains used in our study was confirmed. The strains from Turkey formed an interesting panel because they were assembled after a series of major outbreaks in a period of 20 years. All strains were highly aggressive after inoculation of stems or cotyledons.

Genetic diversity of the population. A split network analysis didn't detect any recombination events in the pathogenicity-related genes. This supports the hypothesis that those genes are transmitted via horizontal transfer. The *pat-1* and *ceIA* genes (Dreier et al. 1997; Meletzus et al. 1993) were not selected for MLST or MSLA analysis because they are located on the plasmids of *C. michiganensis* subsp. *michiganensis*. The presence of these genes does not always correlate with virulence (Jacques et al. 2012; Kleitman et al. 2008) and plasmid

exchange is very frequent. Our gene selection for the population genetic study based on chromosome located virulence genes is more appropriate than a selection of plasmid-originated virulence genes.

The genetic diversity within the 108 *C. michiganensis* subsp. *michiganensis* strains was relatively high compared with other studies (Croce et al. 2016; Quesada-Ocampo et al. 2012). However, *C. michiganensis* subsp. *michiganensis* has a low genetic diversity worldwide compared with other plant bacteria (Andreani et al. 2014; Castillo and Greenberg 2007; Sarkar and Guttman 2004).

An assumption in population genetics is that housekeeping genes are under a stabilizing selection and pathogenicity-related genes are under positive selection. A Tajima's D test can be used to test neutrality of genes (i.e., the presence of selection forces), but the assumption made in the test is not always accurate for conditions of neutrality. To determine the selection acting on genes, the average frequency of synonymous substitutions per potential synonymous sites (Ks) and non-synonymous substitutions per potential non-synonymous sites (Ka) was measured. The Ka/Ks ratio indicated that the housekeeping genes in our study have a K-value lower than 1, which either indicates a stabilizing selection acting on these genes or a population expansion event. The Ka/Ks ratio for the disease-related genes indicated that only the *ppaA* gene has a value higher than 1, indicating positive selection. However, the pathogenicity-related genes *chpC* and *tomA*, which were expected to show positive selection, did not give values indicative of positive selection. This result might be explained by the fact that these pathogenicity-related genes are involved but are not required for pathogenicity (Jacques et al. 2012). The association of virulence with these pathogenicity genes was not always absolute (Tancos et al. 2015). To test whether a significant
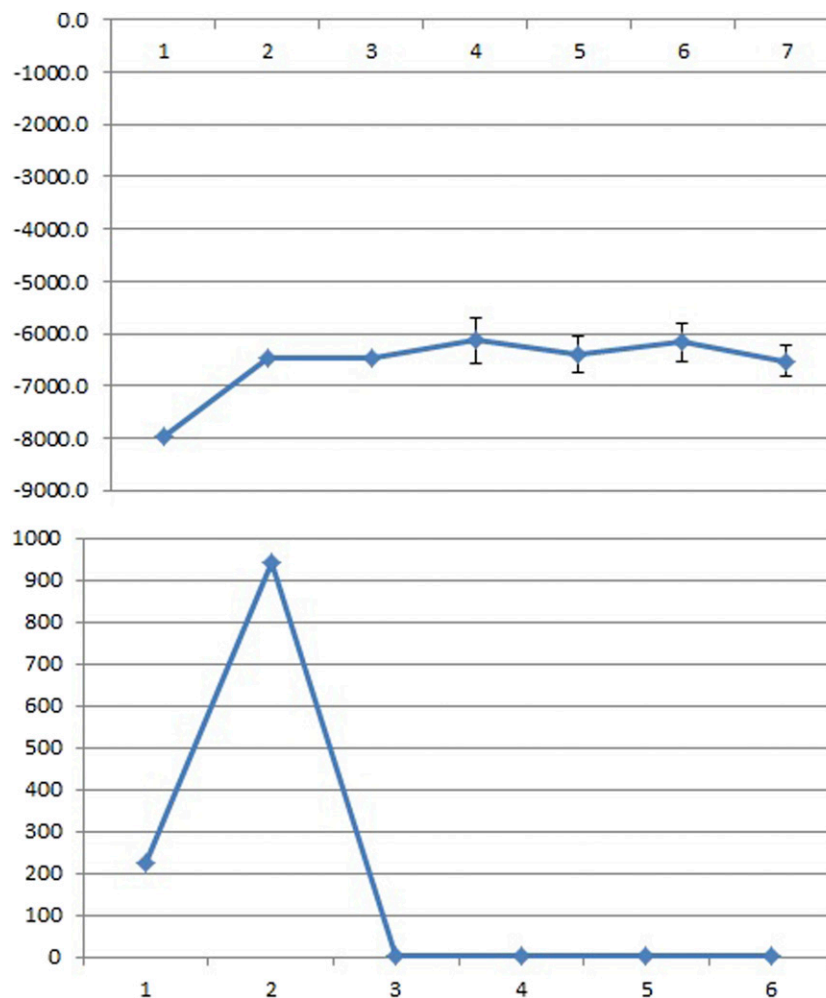

Fig. 3. Structure analysis result. Mean L(K) (± SD) over 10 runs for each K value (top). ΔK calculated as $\Delta K = m|L''(K)|/ s[L(K)]$ (bottom).

genetic reduction occurred in our *C. michiganensis* subsp. *michiganensis* population, three independent tests (Table 2) were used. The tests indicated a low level of genetic diversity for all disease-related genes. These results are likely due to selective sweep, purifying selection, or population expansion (Tajima 1989).
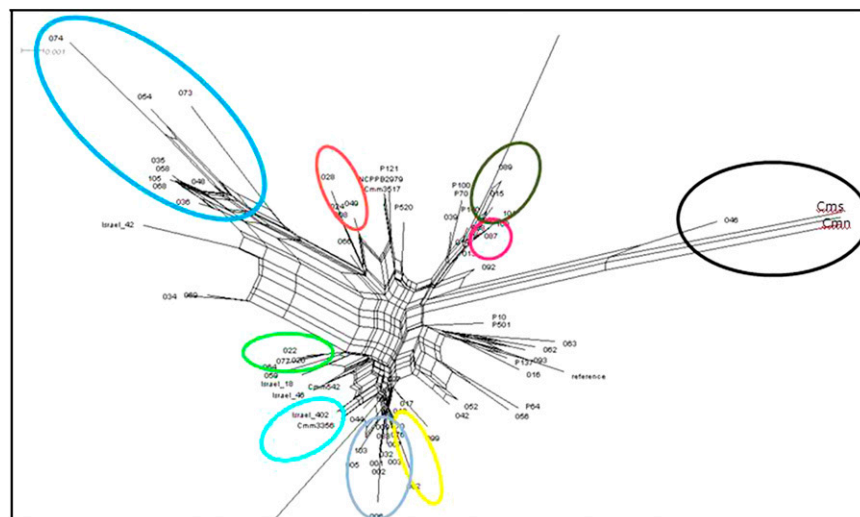
It is important to choose the right genes for an intraspecies genetic analysis with sufficient levels of polymorphism because in some studies, only three polymorphic sites were found in seven housekeeping genes (Kidgell et al. 2002). In general, the housekeeping genes had a higher genetic diversity than the pathogenicity-related genes. But overall, the genes that we selected for genetic analysis of the bacterial population had a higher genetic diversity compared with genes that have been used to characterize *C. michiganensis* subsp. *michiganensis* in other studies (Croce et al. 2016; Quesada-Ocampo et al. 2012).

MLST analysis. MLST analysis is a very informative tool to monitor population dynamics. This analysis can determine allelic combinations of strains and therefore their clonality. eBURST analysis detected a few clonal complexes and many singletons; similar results were found in other *C. michiganensis* subsp. *michiganensis* characterization studies (Croce et al. 2016; Ialacci et al. 2016; Quesada-Ocampo et al. 2012). Although the 62 unique STs show a high heterogeneity, most STs are related by means of DLVs. However, the intermediate units (SLVs) were missing, making it impossible to connect them to the same clonal complex. Based on the eBURST group definition, most strains can be grouped into five units. This model of strain relation is consistent with a population in which a selective sweep or rapid population expansion has been a diversifying force (Spratt and Maiden 1999). Most groups in this study were related to a specific year or location, except for group 1. Strains in this group are widespread in Turkey. These strains were the first observed and were collected between 1996 and 2010. Strains within this group might have survived in debris and became active again in the following years. The other groups and 12 singletons were isolated after 2004 and may represent newly introduced strains. Unfortunately, information is lacking on the precise origin of the isolated strains. We speculate that in the 1990s, at least four novel *C. michiganensis* subsp. *michiganensis* strains were introduced. The existence of 12 singletons might be indicative of additional introductions but they may also have been evolved from other strains. Thus, based on the eBURST grouping, we speculate that only limited new introductions occurred in Turkey.

Multilocus sequence analysis (MLSA). Phylogenetic signals can be quantified by the number of visible clonal complexes or eBURST groups at the edges of lineages. This shows how informative selected genes are and whether gene selection is appropriate for a phylogenetic analysis. Phylogenetic signals were found for all genes, but the signals varied. A weak phylogenetic signal can be the result of a recombination event (Feil and Spratt 2001; Spratt 2004) because recombination can interfere with the phylogenetic signal between genes resulting in incongruent signals from different gene trees (Feil and Spratt 2001). Reticulation networks were found in some genes, which is also an indication of recombination events. The relationship between nucleotide diversity and haplotype diversity is shown in Figure 1. Assuming a mutational model, nucleotide diversity and allelic diversity should show a positive correlation. We have detected an uneven positive correlation between these parameters, which supports the possibility of both recombination and mutation events in pathogen evolution as also found in other bacterial species (Feil et al. 2003). The diversification ratio of clones from founder strains within clonal complexes was 1.4 (14 recombination events and 10 mutation events). This ratio is very low compared with that of other bacteria (Feil et al. 1999, 2000), in which this ratio is usually at least 15. Based on the partial congruence of the genetic tree, the relationship between nucleotide and haplotype diversity, and the recombination ratio, we conclude that recombination and mutation have played a near-equal role in pathogen evolution. However, we should not ignore occurrence of recombination within the population containing single base differences in case recombination occurs with the donor and recipient, for which DNA are very similar. In that case, true differences cannot be inferred with statistical significance.

The phylogenetic relation between STs (or eBURST groups), singletons, and external strains gives information about genetic relations between strains and groups. MLSA by ML analysis of concatenated loci separated the non-tomato host group from the other strains. Strain 46, which was isolated from eggplant, was separated from the tomato strains. This suggests that there is a significant genetic difference between *C. michiganensis* subsp. *michiganensis* strains from tomato strains and other hosts. All eBURST groups were visible at the edges of lineages in the ML tree. The phylogenetic relation of the Serbian groups was similar to that observed in a previous study (Milijašević-Marčić et al. 2012). Both Serbian and Turkish isolates were distributed throughout an ML phylogenetic tree with no distinct population clusters, which indicates that both populations shared a common origin and recent introduction event. A similar conclusion was drawn between the Serbian collection and American collection (Tancos et al. 2015). These parameters indicate that the ML tree gives a reliable representation. Bootstrap values supporting tree branching were low (below 50%) in the ML tree, which can result from a low diversity or from recombination events that are shorter than the lengths of the genes used for the construction of the genetic trees. As a result, the genetic trees may lack statistical significance

Fig. 4. Split network analysis of concatenated data from the eight analyzed genes.

because different parts of genes have different evolutionary histories (Feil and Spratt 2001). Although the ML tree was constructed using concatenated data, only the *sdhA* and *kdpA* genes would have been useful to construct a similar tree. Tancos et al. (2015) indicated that both genes had a strong discriminatory power as concatenated data in terms of inferring genetic relations within the population.

MLSA can be used to reveal evolutionary relationships on a species or subspecies level. Adaptation of microbial organisms to a particular host requires metabolic changes in which housekeeping genes are involved. This form of analysis is suitable for studying long-term epidemiological processes, but MLSA might be unsuitable for short-term epidemiological studies due to the slow accumulation of genetic variation and, thus, strains are often indistinguishable (Spratt 2004). In line, MLSA has successfully revealed relations on species and subspecies level of global populations (Waleron et al. 2011).

Population structure. We applied a population structure analysis to determine whether subpopulations were found in our sampled population due to geographic or time isolation. Despite the clear ST clonal complexes determined by the clonal and ML tree analyses, the structure analysis identified only two major populations, *C. michiganensis* subsp. *michiganensis* strains and *C. michiganensis* spp. The results show a high genetic diversity between *C. michiganensis* subsp. *michiganensis* and other *C. michiganensis* spp., but a low genetic diversity within *C. michiganensis* subsp. *michiganensis* strains. Although some clonal complexes were structured into different subpopulations, the genetic diversity among the different subgroups was not high enough to support a further subdivision. Subpopulations were observed when taking the housekeeping genes in consideration, and group 2 strains were then substructured. This might be an indication of how important the flexible and core genome effects on organism evolution are. Therefore, population structure analysis might be more appropriate only on core genome genes.

Reticulation network. To visualize the recombination events within selected genes, we applied a network analysis. Reticulation

Fig. 5. Maximum likelihood of concatenated data of the eight analyzed genes. Coloring indicates separate eBURST groups. Blue text indicates group_5, and red text indicates singletons.

was detected in the central network of concatenated data (Fig. 4), which is indicative of past recombination events. The relative lack of reticulation in the split decomposition graphs, particularly when individual loci are examined, further supports a relatively limited role of recombination. Split network analysis is very like PCA analysis. This latter analysis is used for visualization of genetic relationships of organisms in which recombination has had a strong effect on gene evolution, and therefore tree construction using a bifurcating method is not appropriate (Spratt and Maiden 1999). In our study, however, recombination and mutation showed almost equal impacts on *C. michiganensis* subsp. *michiganensis* evolution, and therefore the split network analysis gave similar results to the bifurcating tree construction. This result provides additional proof of similar effects of recombination and mutation on pathogen evolution.

Conclusions

We conclude that the *C. michiganensis* subsp. *michiganensis* population in Turkey is highly uniform. The main diversifying forces are recombination and mutation at a nearly equal level. We also conclude that sequencing housekeeping genes gives more information than pathogenicity-related genes and are therefore more appropriate in MLST studies. The genetic diversity of the studied *C. michiganensis* subsp. *michiganensis* isolates was not related to time and place, showing that after its introduction, *C. michiganensis* subsp. *michiganensis* was further spread due to agricultural practices. Overall, the Turkish strains are quite like global strains.

Literature Cited

Almeida, N. F., Yan, S., Cai, R., Clarke, C. R., Morris, C. E., Schaad, N. W., Schuenzel, E. L., Lacy, G. H., Sun, X., Jones, J. B., Castillo, J. A., Bull, C. T., Leman, S., Guttman, D. S., Setubal, J. C., and Vinatzer, B. A. 2010. PAMDB, a multilocus sequence typing and analysis database and website for plant-associated microbes. Phytopathology 100:208-215.

Andreani, N. A., Martino, M. E., Fasolato, L., Carraro, L., Montemurro, F., Mioni, R., Bordin, P., and Cardazzo, B. 2014. Tracking the blue: a MLST approach to characterise the *Pseudomonas fluorescens* group. Food Microbiol. 39:116-126.

Anonymous. 2000. European Union Council directive 2000/29/EC on protective measures against the introduction into the Community of organisms harmful to plants or plant products and against their spread within the Community. Off. J. Eur. Communities L 169:33.

Basim, E., Basim, H., Dickstein, E. R., and Jones, J. B. 2004. Bacterial canker caused by *Clavibacter michiganensis* subsp. *michiganensis* on greenhouse-grown tomato in the western Mediterranean region of Turkey. Plant Dis. 88:1048.

Baysal, Ö., Mercati, F., İkten, H., Yıldız, R. Ç., Carimi, F., Aysan, Y., and Teixeira da Silva, J. A. 2011. *Clavibacter michiganensis* subsp. *michiganesis*: Tracking strains using their genetic differentiations by ISSR markers in Southern Turkey. Physiol. Mol. Plant Pathol. 75:113-119.

Castillo, J. A., and Greenberg, J. T. 2007. Evolutionary dynamics of *Ralstonia solanacearum*. Appl. Environ. Microbiol. 73:1225-1238.

Chang, R. J., Ries, S. M., and Pataky, J. K. 1991. Dissemination of *Clavibacter michiganensis* subsp. *michiganensis* by practices used to produce tomato transplants. Phytopathology 81:1276-1281.

Chang, R. J., Ries, S. M., and Pataky, J. K. 1992. Local sources of *Clavibacter michiganensis* ssp. *michiganensis* in the development of bacterial canker on tomatoes. Phytopathology 82:553-560.

Coligan, J. E., Kruisbeek, A. M., Margulies, D. H., Shevach, E. M., and Strober, W. 1991. Current Protocols in Immunology. John Wiley & Sons, Hoboken, NJ.

Croce, V., Pianzzola, M. J., Durand, K., González-Arcos, M., Jacques, M.-A., and Siri, M. I. 2016. Multilocus sequence typing reveals high variability among *Clavibacter michiganensis* subsp. *michiganensis* strains affecting tomato crops in Uruguay. Eur. J. Plant Pathol. 144:1-13.

Dreier, J., Bermpohl, A., and Eichenlaup, R. 1995. Southern hybridization and PCR for specific detection of phytopathogenic *Clavibacter michiganensis* subsp. *michiganensis*. Phytopathology 85:462-468.

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

Eichenlaub, R., Gartemann, K. H., and Burger, A. 2006. *Clavibacter michiganensis*, a group of gram-positive phytopathogenic bacteria. Pages 385-421 in: Plant-Associated Bacteria. S. S. Gnanamanickam, ed. Springer, Dordrecht, The Netherlands.

Enright, M. C., and Spratt, B. G. 1999. Multilocus sequence typing. Trends Microbiol. 7:482-487.

EPPO. 2013. EPPO Standards Diagnostic PM 7/42 (2) Clavibacter michiganensis subsp. michiganensis. EPPO Bull. 43:46-47.

Evanno, G., Regnaut, S., and Goudet, J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Mol. Ecol. 14:2611-2620.

Feil, E. J., Cooper, J. E., Grundmann, H., Robinson, D. A., Enright, M. C., Berendt, T., Peacock, S. J., Smith, J. M., Murphy, M., Spratt, B. G., Moore, C. E., and Day, N. P. J. 2003. How clonal is *Staphylococcus aureus?* J. Bacteriol. 185:3307-3316.

Feil, E. J., Li, B. C., Aanensen, D. M., Hanage, W. P., and Spratt, B. G. 2004. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J. Bacteriol. 186:1518-1530.

Feil, E. J., Maiden, M. C., Achtman, M., and Spratt, B. G. 1999. The relative contributions of recombination and mutation to the divergence of clones of *Neisseria meningitidis*. Mol. Biol. Evol. 16:1496-1502.

Feil, E. J., Smith, J. M., Enright, M. C., and Spratt, B. G. 2000. Estimating recombinational parameters in *Streptococcus pneumoniae* from multilocus sequence typing data. Genetics 154:1439-1450.

Feil, E. J., and Spratt, B. G. 2001. Recombination and the population structures of bacterial pathogens. Annu. Rev. Microbiol. 55:561-590.

Gitaitis, R. D. 1990. Induction of a hypersensitive-like reaction in four-o'clock by *Clavibacter michiganensis* subsp. *michiganensis*. Plant Dis. 74:58-60.

Gleason, M. L., Gitaitis, R. D., and Ricker, M. D. 1993. Recent progress in understanding and controlling bacterial canker of tomato in eastern North America. Plant Dis. 77:1069-1076.

Hubisz, M. J., Falush, D., Stephens, M., and Pritchard, J. K. 2009. Inferring weak population structure with the assistance of sample group information. Mol. Ecol. Resour. 9:1322-1332.

Huson, D. H., and Bryant, D. 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23:254-267.

Ialacci, G. M., Bella, P., Licciardello, G., Strano, C. P., Eichenlaub, R., Gartemann, K. H., La Rosa, R., and Catara, V. 2016. Clonal populations of *Clavibacter michiganensis* subsp. *michiganensis* are responsible for the outbreaks of bacterial canker in greenhouse tomatoes in Italy. Plant Pathol. 65:484-495.

Jacques, M.-A., Durand, K., Orgeur, G., Balidas, S., Fricot, C., Bonneau, S., Quillévéré, A., Audusseau, C., Olivier, V., Grimault, V., and Mathis, R. 2012. Phylogenetic analysis and polyphasic characterization of *Clavibacter michiganensis* strains isolated from tomato seeds reveal that nonpathogenic strains are distinct from *C. michiganensis* subsp. *michiganensis*. Appl. Environ. Microbiol. 78:8388-8402.

Jolley, K. A., Feil, E. J., Chan, M. S., and Maiden, M. C. 2001. Sequence type analysis and recombinational tests (START). Bioinformatics 17:1230-1231.

Jones, N., Bohnsack, J. F., Takahashi, S., Oliver, K. A., Chan, M. S., Kunst, F., Glaser, P., Rusniok, C., Crook, D. W., Harding, R. M., Bisharat, N., and Spratt, B. G. 2003. Multilocus sequence typing system for group B streptococcus. J. Clin. Microbiol. 41:2530-2536.

Kidgell, C., Reichard, U., Wain, J., Linz, B., Torpdahl, M., Dougan, G., and Achtman, M. 2002. *Salmonella typhi*, the causative agent of typhoid fever, is approximately 50,000 years old. Infect. Genet. Evol. 2:39-45.

Kleitman, F., Barash, I., Burger, A., Iraki, N., Falah, Y., Sessa, G., Weinthal, D., Chalupowicz, L., Gartemann, K. H., Eichenlaub, R., and Manulis-Sasson, S. 2008. Characterization of a *Clavibacter michiganensis* subsp *michiganensis* population in Israel. Eur. J. Plant Pathol. 121:463-475.

Klement, Z., Rudolph, K., and Sands, D. C. 1990. Methods in phytobacteriology. Akadémiai Kiadó, Budapest, Hungary.

Lelliott, R. A., and Stead, D. E. 1987. Methods for the diagnosis of bacterial diseases of plants. Blackwell Scientific Publications, Oxford, U.K.

Maddison, W. P., and Maddison, D. R. 2010. Mesquite: a modular system for evolutionary analysis, version 2.74. http://mesquiteproject.org

Maiden, M. C. J. 2006. Multilocus sequence typing of bacteria. Annu. Rev. Microbiol. 60:561-588.

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.

Milijašević-Marčić, S., Gartemann, K.-H., Frohwitter, J., Eichenlaub, R., Todorović, B., Rekanović, E., and Potočnik, I. 2012. Characterization of *Clavibacter michiganensis* subsp *michiganensis* strains from recent outbreaks of bacterial wilt and canker in Serbia. Eur. J. Plant Pathol.: 1-15.

Miller, M. A., Pfeiffer, W., and Schwartz, T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Pages 1-8 in: Proceedings of the Gateway Computing Environments Workshop (GCE), 14 Nov 2010, New Orleans, LA.

Pritchard, J. K., Stephens, M., and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945-959.

Quesada-Ocampo, L. M., Landers, N. A., Lebeis, A. C., Fulbright, D. W., and Hausbeck, M. K. 2012. Genetic structure of *Clavibacter michiganensis* subsp. *michiganensis* populations in Michigan commercial tomato fields. Plant Dis. 96:788-796.

Rozas, J., Sánchez-DelBarrio, J. C., Messeguer, X., and Rozas, R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496-2497.

Şahin, F., Uslu, H., Kotan, R., and Donmez, M. F. 2002. Bacterial canker, caused by *Clavibacter michiganensis* ssp. *michiganensis*, on tomatoes in eastern Anatolia region of Turkey. Plant Pathol. 51:399.

Sarkar, S. F., and Guttman, D. S. 2004. Evolution of the core genome of *Pseudomonas syringae*, a highly clonal, endemic plant pathogen. Appl. Environ. Microbiol. 70:1999-2012.

Schaad, N. W., Jones, J. B., Chun, W., and Committee, A. P. S. B. 2001. Laboratory guide for identification of plant pathogenic bacteria. APS Press, St. Paul, MN.

Sen, Y., van der Wolf, J., Visser, R. G. F., and van Heusden, S. 2015. Bacterial canker of tomato: current knowledge of detection, management, resistance, and interactions. Plant Dis. 99:4-13.

Sen, Y., Feng, Z., Vandenbroucke, H., van der Wolf, J., Visser, R., and van Heusden, A. 2012. Screening for new sources of resistance to *Clavibacter michiganensis* subsp. *michiganensis* (*C. michiganensis* subsp. *michiganensis*) in tomato. Euphytica 190:309-317.

Smith, J. M., Smith, N. H., O'Rourke, M., and Spratt, B. G. 1993. How clonal are bacteria? Proc. Natl. Acad. Sci. USA 90:4384-4388.

Spratt, B. 2004. Exploring the Concept of Clonality in Bacteria. Pages 323-352 in: Genomics, Proteomics, and Clinical Bacteriology, Vol. 266. N. Woodford and A. Johnson, eds. Humana Press, New York.

Spratt, B. G., and Maiden, M. C. 1999. Bacterial population genetics, evolution and epidemiology. Philos. Trans. R. Soc. Lond. B Biol. Sci. 354:701-710.

Stamatakis, A. 2014. RAxML Version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312-1313.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585-595.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28:2731-2739.

Tancos, M. A., Lange, H. W., and Smart, C. D. 2015. Characterizing the genetic diversity of the *Clavibacter michiganensis* subsp. *michiganensis* population in New York. Phytopathology 105:169-179.

Trevors, J. T., and Finnen, R. L. 1990. Introduction and recovery of *Clavibacter michiganensis* subsp. *michiganensis* from agricultural soil. Plant Soil 126:141-143.

Waleron, M., Waleron, K., Kamasa, J., Przewodowski, W., and Lojkowska, E. 2011. Polymorphism analysis of housekeeping genes for identification and differentiation of *Clavibacter michiganensis* subspecies. Eur. J. Plant Pathol. 131:341-354.

Yazdankhah, S. P., Kriz, P., Tzanakaki, G., Kremastinou, J., Kalmusova, J., Musilek, M., Alvestad, T., Jolley, K. A., Wilson, D. J., McCarthy, N. D., Caugant, D. A., and Maiden, M. C. J. 2004. Distribution of serogroups and genotypes among disease-associated and carried isolates of *Neisseria meningitidis* from the Czech Republic, Greece, and Norway. J. Clin. Microbiol. 42:5146-5153.