



Comparative Genomics and Phylogenetic Analyses Suggest Several Novel Species within *Clavibacter* sp. Including Non-Pathogenic Tomato-Associated Strains

Ebrahim Osdaghi, Touraj Rahimi, S. Mohsen Taghavi, Maryam Ansari, Sadegh Zarei, Perrine Portier, Martial Briand, Marie Agnès Jacques

► To cite this version:

Ebrahim Osdaghi, Touraj Rahimi, S. Mohsen Taghavi, Maryam Ansari, Sadegh Zarei, et al.. Comparative Genomics and Phylogenetic Analyses Suggest Several Novel Species within *Clavibacter* sp. Including Non-Pathogenic Tomato-Associated Strains. *Applied and Environmental Microbiology*, 2020, 10.1128/AEM.02873-19 . hal-02624788

HAL Id: hal-02624788

<https://hal.inrae.fr/hal-02624788>

Submitted on 26 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

1 Running Title: Comparative genomics of *Clavibacter* spp.

2

3 **Comparative Genomics and Phylogenetic Analyses Suggest Several Novel Species**
4 **within *Clavibacter* sp. Including Non-Pathogenic Tomato-Associated Strains**

5

6 Ebrahim Osdaghi^{1*}, Touraj Rahimi², S. Mohsen Taghavi¹, Maryam Ansari¹, Sadegh Zarei¹, Perrine
7 Portier^{3,4}, Martial Briand³, Marie-Agnes Jacques^{3*}

8

9 ¹ Department of Plant Protection, College of Agriculture, Shiraz University, Shiraz 71441-65186,
10 Iran

11 ² Department of Agronomy and Plant breeding, Shahr-e-Qods Branch, Islamic Azad University,
12 Tehran, Iran

13 ³ INRA, Agrocampus-Ouest, Université d'Angers, SFR 4207 QUASAV, 49071 Beaucouzé, France

14 ⁴ CIRM-CFBP, French Collection of Plant-Associated Bacteria, INRA, 49071 Beaucouzé, France

15

16 *For correspondence:

17 E Osdaghi eosdaghi@shirazu.ac.ir

18 MA Jacques marie-agnes.jacques@inra.fr

19

20

21 **Abstract**

22 Members of *Clavibacter* spp. are economically important bacterial plant pathogens infecting a set
 23 of diverse agricultural crops (e.g. alfalfa, corn, potato, tomato, and wheat). Tomato-associated
 24 *Clavibacter* spp. strains occupy a great portion of genetic diversity of the genus, and *C.*
 25 *michiganensis sensu stricto* (formerly *C. michiganensis* subsp. *michiganensis*) causing bacterial
 26 canker disease considered one of the destructive seed-borne agents of the crop worldwide.
 27 However, current taxonomic descriptions of the genus do not reflect the existing diversity of the
 28 strains, resulting in unsatisfactory consequences in quarantine surveys for the pathogens. In this
 29 study, we used all the available genome sequences of *Clavibacter* spp. strains – including the type
 30 strains of newly described subspecies – to provide a precise insight into the diversity of tomato-
 31 associated members of the genus, and further clarify taxonomic status of the strains using
 32 genotypic and phenotypic features. Results of phylogenetic analyses revealed the existence of
 33 nine hypothetical new species among the investigated strains. None of the three new subspecies
 34 (i.e. *C. michiganensis* subsp. *californiensis*, *C. michiganensis* subsp. *chilensis* and *C. michiganensis*
 35 subsp. *phaseoli*) is included within the tomato-pathogenic *C. michiganensis sensu stricto* lineage.
 36 Although comparative genomics revealed the lack of *chp* and *tomA* pathogenicity determinant
 37 gene clusters in the non-pathogenic strains, a number of pathogenicity related genes were noted
 38 to be present in all the strains regardless of their pathogenicity characteristics. Altogether, our
 39 results advocate a need for a formal taxonomic reconsideration of tomato-associated *Clavibacter*
 40 spp. strains to facilitate differentiation of the lineages in quarantine inspections.

41

42 **Importance**

43 *Clavibacter* spp. are economically important bacterial plant pathogens infecting a set of diverse
 44 agricultural crops such as alfalfa, corn, pepper, potato, tomato, and wheat. A number of plant
 45 pathogenic members of the genus (*e.g. C. michiganensis sensu stricto* and *C. sepedonicus* infecting
 46 tomato and potato plants, respectively) are included in the A2 (high risk) list of quarantine
 47 pathogens by the European and Mediterranean Plant Protection Organization (EPPO). Although
 48 tomato-associated members of *Clavibacter* spp. occupy a significant portion of the genetic
 49 diversity in the genus, only the strains belonging to *C. michiganensis sensu stricto* (formerly *C.*
 50 *michiganensis* subsp. *michiganensis*) cause bacterial canker disease of tomato and subjected to
 51 the quarantine inspections. Hence, discrimination of the pathogenic and non-pathogenic
 52 *Clavibacter* spp. strains associated with tomato seeds and transplants plays a pivotal role in the
 53 accurate detection and cost efficiently management of the disease. On the other hand, detailed
 54 information on the genetic contents of different lineages of the genus would lead to the
 55 development of genome-informed specific detection techniques. In this study, we have provided
 56 an overview on the phylogenetic and genomic differences between the pathogenic and non-
 57 pathogenic tomato-associated *Clavibacter* spp. strains. We have also noted that the taxonomic
 58 status of newly introduced subspecies of *C. michiganensis* (*i.e. C. michiganensis* subsp.
 59 *californiensis*, *C. michiganensis* subsp. *chilensis* and *C. michiganensis* subsp. *phaseoli*) should be
 60 reconsidered.

61 **Keywords:** Actinobacteria, Bacterial canker of tomato, Bacterial Taxonomy, *Clavibacter*
 62 *michiganensis sensu stricto*, Quarantine pathogen

64 Introduction

65 A number of plant-pathogenic bacterial species are reported to have non-pathogenic lineages
 66 which usually exist in the same ecological niche with their pathogenic counterparts (1-3). Non-
 67 pathogenic lineages or strains typically have similar genetic contents to their pathogenic relatives,
 68 but lack some of the key pathogenicity determinants (*e.g.* pathogenicity islands, virulence genes,
 69 and plasmids) (4-6). As far as economically important quarantine plant pathogenic bacteria are
 70 concerned, presence of non-pathogenic strains in commercial seeds or propagative parts of plants
 71 will be interfering in the accurate detection of the pathogens, leading to false positive results in
 72 the quarantine inspections, and unsatisfaction of seed producers and traders (7). This is due in
 73 part to the fact that most of the non-pathogenic bacterial strains are phenotypically similar to
 74 their pathogenic relatives; therefore, not differentiable from each other on the culture media (7).
 75 Furthermore, most of the molecular detection protocols (*e.g.* PCR primers, probes, and
 76 antibodies) are designed based on the general features of a given species/subspecies/pathovar
 77 rather than focusing on the pathogenicity determinants of the pathogen (7). As a consequence,
 78 contradictions in the results of quarantine inspections will be leading to economic losses of seed
 79 producers and will have negative impact on transportation of plant materials in a global scale (7,
 80 8).

81 *Clavibacter* spp. are economically important Gram-positive bacterial plant pathogens infecting a
82 set of diverse agricultural crops *e.g.* alfalfa, corn, pepper, potato, tomato, and wheat (9). Until
83 recently, the genus *Clavibacter* was considered to include only one species *C. michiganensis*,
84 comprising five plant pathogenic subspecies *i.e.* *C. michiganensis* subsp. *insidiosus*; *C.*
85 *michiganensis* subsp. *michiganensis*; *C. michiganensis* subsp. *nebraskensis*; *C. michiganensis* subsp.
86 *sepedonicus*; and *C. michiganensis* subsp. *tessellarius* (9, 10). Furthermore, all the tomato- and
87 pepper-associated *Clavibacter* sp. strains were classified as members of *C. michiganensis* subsp.
88 *michiganensis* regardless of being pathogenic or non-pathogenic on the host of isolation.
89 However, using the multilocus sequence analysis and typing (MLSA/MLST) Jacques et al. (11)
90 showed that tomato-associated non-pathogenic *Clavibacter* spp. strains are phylogenetically
91 distinct from the pathogenic counterparts in the species. Differentiation of the pathogenic and
92 non-pathogenic strains of *C. michiganensis* has always been an ongoing challenge for official
93 sanitary agencies, quarantine inspectors and seed providers (12), since the false positive results
94 would lead to the rejection of seed/seedling lots in an economically significant scale (7, 8). This led
95 to the assumption that a comprehensive complete genome sequence-based reconsideration in the
96 *C. michiganensis sensu lato* (all the former members of *C. michiganensis* according to Davis et al.
97 (9)) is warranted to shed a light on the genetic diversity, genomic repertoires and taxonomic
98 status of the pathogenic and non-pathogenic tomato-associated strains of the species (11).

99 Following the emergence of high throughput molecular-phylogenetic techniques many *Clavibacter*
100 spp. strains, which have often previously been misidentified based on phenotypic features, were
101 assigned into novel taxa. For instance, tomato-associated non-pathogenic members of *C.*
102 *michiganensis sensu lato* were assigned into two new subspecies *C. michiganensis* subsp.

103 *californiensis* and *C. michiganensis* subsp. *chilensis* (13). Additionally, *C. michiganensis* subsp.
104 *phaseoli* and *C. michiganensis* subsp. *capsici* were identified as the causal agents of bacterial bean
105 leaf yellowing on common bean (*Phaseolus vulgaris*) and bacterial canker of pepper (*Capsicum*
106 *annuum*), respectively (14, 15). Furthermore, non-pathogenic peach-colored strains isolated from
107 tomato phyllosphere were reported to be distinct from the tomato-pathogenic members of
108 *Clavibacter* spp. (16).

109 Recently, a re-classification of *Clavibacter* spp. into five new species and a new combination was
110 proposed based on the genomic information *e.g.* average nucleotide identity (ANI) and digital
111 DNA-DNA hybridization (dDDH) indices (17-19). The original subspecies of *C. michiganensis sensu*
112 *lato* were elevated at the species level and designated as *C. michiganensis* (hereafter referred to
113 as *C. michiganensis sensu stricto*: formerly *C. michiganensis* subsp. *michiganensis*), *C. tessellarius*,
114 *C. insidiosus*, *C. nebraskensis*, *C. capsici*, as well as *C. sepedonicus* as a new combination (17, 19).
115 However, due to the lack of genomic information from the newly proposed subspecies (*C.*
116 *michiganensis* subsp. *californiensis*, *C. michiganensis* subsp. *chilensis*, and *C. michiganensis* subsp.
117 *phaseoli*) as well as several taxonomically undetermined strains, additional investigations are
118 warranted to further clarify the taxonomy of the genus. Moreover, strains associated with
119 solanaceous vegetables contain a large fraction of diversity within the *Clavibacter* spp. members,
120 and much of the molecular, phylogenetic, and genomic information for these strains remain
121 unexplored. As for the tomato-associated strains of *Clavibacter* spp., comparative genomics on a
122 wide collection of non-pathogenic and pathogenic strains would further elucidate the genetic
123 diversity of these bacteria, resulting in the development of genome-informed specific molecular
124 markers (*e.g.* specific conventional PCR and real-time PCR primers, as well as loop-mediated

125 isothermal amplification) for the detection and differentiation of the pathogenic and non-
 126 pathogenic strains in the quarantine posts.

127 The objectives of the present study were to I) investigate genetic diversity of tomato-associated
 128 *Clavibacter* spp. strains using the genome sequences of all available non-pathogenic and
 129 pathogenic strains and II) provide a novel taxonomic overview into the status of tomato-
 130 pathogenic and non-pathogenic strains within the genus. For this aim, we used the genome
 131 sequences of 40 *Clavibacter* spp. strains including the type strains of three newly described
 132 subspecies (*C. michiganensis* subsp. *californiensis*, *C. michiganensis* subsp. *chilensis* and *C.*
 133 *michiganensis* subsp. *phaseoli*), as well as additional atypical non-pathogenic strains isolated from
 134 tomato plants around the globe (20). Draft genome sequence-based phylogenetic analyses
 135 revealed a higher diversity among the non-pathogenic strains of *Clavibacter* spp. than that has
 136 previously been reported, delineating them into several new species. On the other hand, our data
 137 revealed that the two individual subspecies *C. michiganensis* subsp. *chilensis* and *C. michiganensis*
 138 subsp. *phaseoli* need to be considered as the members of one species according to the 99%
 139 genome similarity among the type strains. Furthermore, comparative genomics among the
 140 pathogenic and non-pathogenic strains of tomato-associated strains, as well as the type strains of
 141 the remaining species/subspecies within the genus revealed several pathogenicity determinant
 142 genes presenting only in *C. michiganensis sensu stricto*, which could be considered as suitable
 143 genomic targets for the development of specific detection methods for the tomato pathogen.

144

145 **Results**

146 **Pathogenicity and Host Range**

147 Tomato and pepper plants inoculated with the standard strain of *C. michiganensis sensu stricto*
148 (ICMP 22049) showed the expected disease symptoms 10-12 days post inoculation. Although
149 tomato plants inoculated with the strain ICMP 22049 showed wilting and plant death (Figure S1a),
150 pepper plants inoculated using the same strain showed only stem canker symptoms on the site of
151 inoculation with no wilting nor plant death in the same timeframe (Figure S1b). However, neither
152 *C. michiganensis* subsp. *phaseoli* nor *C. michiganensis* subsp. *chilensis* did induce symptoms on the
153 inoculated plant species *i.e.* common bean, cowpea, pepper, mung bean, and tomato (Table 1).
154 Furthermore, we could not re-isolate *C. michiganensis* subsp. *phaseoli* and *C. michiganensis* subsp.
155 *chilensis* from the stem, petiole, and leaf tissues 5-10 cm above the inoculation site on the stem.
156 This could be an indication of the fact that *C. michiganensis* subsp. *phaseoli* and *C. michiganensis*
157 subsp. *chilensis* were unable to endophytically colonize the evaluated plant species. As for the
158 orange-pigmented tomato-associated strains CFBP 8615 and CFBP 8616, although no symptoms
159 were observed on common bean, cowpea, pepper, mung bean, and tomato plants, bacterial
160 colonies similar to those originally inoculated were consistently re-isolated from the leaf tissues of
161 common bean cv. Navy plants inoculated with CFBP 8616 (Table 1). Furthermore the standard
162 strain ICMP 22049 was consistently re-isolated from the symptomatic pepper and tomato plants
163 on YPGA medium and their identity was confirmed using the genus-specific primer pair
164 CMR16F1/CMR16R1 (data not shown). Similar results were obtained in both the replications of
165 the experiments, while the negative control plants remained healthy.

166

167 **Phylogenetic Analyses**

168 Neighbor joining phylogenetic tree constructed using the genome sequences of 40 *Clavibacter*
 169 spp. strains (Table 2) via ANI Calculator online service with all-vs.-all strategy revealed high genetic
 170 diversity among tomato-associated non-pathogenic strains of the genus (Figure 1). ANI values
 171 between different pairs of strains varied from 87% to 100% among the *Clavibacter* spp. strains
 172 (Table 3). While all tomato-pathogenic strains of *C. michiganensis sensu stricto* clustered in a
 173 monophyletic clade showing 99-100% ANI with one another, non-pathogenic strains isolated from
 174 tomato were scattered in several clades, most of which had <96% ANI values with the other clades
 175 (Table 3). The closest non-pathogenic clade to *C. michiganensis sensu stricto* group consisted of
 176 three strains *i.e.* type strain of *C. michiganensis* subsp. *californiensis* (CFBP 8216^T), and the non-
 177 pathogenic strains LMG 26808 and CFBP 7493. ANI value between the type strain of *C.*
 178 *michiganensis sensu stricto* (LMG 7333^T) and the type strain of *C. michiganensis* subsp.
 179 *californiensis* was 95% in all the calculating strategies, retaining them at the threshold of species
 180 definition (21). Nevertheless, the dDDH value (57.70%) between the type strains of *C.*
 181 *michiganensis sensu stricto* and *C. michiganensis* subsp. *californiensis* was far below the threshold
 182 for species delineation (70%) with this method (Table 3). Altogether, given the differences in their
 183 pathogenicity and biochemical characteristics (13) as well as bellow-threshold genomic similarity,
 184 the two taxa *C. michiganensis sensu stricto* and *C. michiganensis* subsp. *californiensis* could be
 185 considered as separate species. Furthermore, ANI between the type strains of *C. michiganensis*
 186 *sensu stricto*, *C. michiganensis* subsp. *californiensis*, and the cluster which included the non-
 187 pathogenic strains LMG 26808 and CFBP 7493 was 94-95% (Table 3). The dDDH values between

188 these type strains and LMG 26808 and CFBP 7493 were also 57-58% indicating them as separate
189 species (Figure 1).

190 *Clavibacter sepedonicus* strains formed a monophyletic cluster separate from all the other lineages
191 by ANI values <93%, which is in coherence with its elevation at the species level (17, 18).

192 *Clavibacter insidiosus* and *C. nebraskensis* strains clustered in a monophyletic group showing 95%
193 ANI between the type strains of the species (Table 3). Here also, the dDDH value (59.90%)
194 between the type strains of these two taxa was far below the threshold for species definition
195 (70%) with this method (Table 3) supporting their elevation into separate species (17-19). The two
196 taxa are also different in their host of isolation and pathogenicity pattern. While the strain CFBP
197 7494 showed only 64.60% dDDH value with the type strain of *C. insidiosus*, 96% ANI (on the upper
198 edge of species definition) prevents differentiation of this strain from the *C. insidiosus* species. The
199 strain CFBP 7494 was isolated from tomato but was non-pathogenic on this plant species, while it
200 has been shown that induces disease symptoms on wheat plants in greenhouse conditions (22).
201 Further evidences – including a comprehensive field survey and host range assay – are needed to
202 elucidate the prevalence and exact taxonomic status of the strain CFBP 7494. Surprisingly, type
203 strains of *C. michiganensis* subsp. *phaseoli* and *C. michiganensis* subsp. *chilensis* shared 99% ANI
204 with one another and 98% ANI with CFBP 7491 isolated from tomato seeds. These three strains
205 had ANIs below 93% with all the remaining clades, suggesting a novel species within the genus.
206 High dDDH value (87.50%) also confirmed the close relationships between the type strains of *C.*
207 *michiganensis* subsp. *phaseoli* and *C. michiganensis* subsp. *chilensis* (Table 3).

208 Two peach-colored strains CFBP 8615 and CFBP 8616 shared 100% ANI with one another, while
209 they differed from all the remaining clades with the ANI values <93%. Furthermore, non-

210 pathogenic strain CFBP 8019 was determined as the phylogenetically closest strain to the peach-
 211 colored strains with 93% ANI value between the two clades. These ANI values are far below the
 212 accepted threshold (95-96%) for the definition of prokaryotic species (21), suggesting that the
 213 strains CFBP 8615 and CFBP 8616 could be defined as forming a new species separated from CFBP
 214 8019, while the strain CFBP 8019 itself belong to a new stand-alone specie (Figure 1). Low ANI
 215 values were also confirmed by dDDH which were <48% between the peach-colored strains and all
 216 the remaining clades (Table 3). The strain CF11 – isolated from soil in a tomato growing
 217 greenhouse (23) – as well as the type strain of the pepper pathogen *C. capsici* (PF008^T) clustered in
 218 a monophyletic clade, while differed from one another with 95% ANI and 58.50% dDDH values.
 219 Hence, CF11 could be proposed as forming a new species within the genus, while the elevation of
 220 ex-*C. michiganensis* subsp. *capsici* into the species level (*C. capsici*) was confirmed as proposed by
 221 Li et al. (17). Type strain of *C. tessellarius* showed <93% ANI with the type strains of all the other
 222 subspecies/species confirming the wheat pathogen as a stand-alone species. However, none of
 223 the two strains CFBP 8017 and DOAB 609, which were clustered in a shared clade with the type
 224 strain of *C. tessellarius* could be included within this species. The ANIs of CFBP 8017 and DOAB 609
 225 with the type strain of *C. tessellarius* were 95% and 93%, respectively, while the dDDH values
 226 between the same strains were 57.70% and 49.00%, respectively (Table 3). Thus, each of the CFBP
 227 8017 and DOAB 609 strains could be defined as representing novel species. The non-pathogenic
 228 strain CASJ009 also had ANI values <90% with all the *Clavibacter* spp. strains evaluated in this
 229 study, indicating that this strain also represents a novel species within the genus (Figure 1; Table
 230 3).

231

Comparative Genomics

Comparative genomics data obtained using the RAST online service revealed that the genome size among the *Clavibacter* spp. strains varied between 3,024 kbp in CFBP 8019 and 3,420 kbp in LMG 26808 with the G+C% content of 72.0% in LMG 26808 to 73.7% in *C. tessellarius* ATCC 33566^T. Furthermore, the number of coding sequences (CDS) varied from 2,629 in *C. michiganensis* subsp. *chilensis* (CFBP 8217^T) to 3,181 in DOAB 609. Genomic characteristics of *Clavibacter* spp. in a panel of 20 representative strains, which were selected on the basis of the phylogenetic analyses (as detailed above) to cover all lineages/clades of the genus are shown in Table 4. The number of subsystems varied from 260 in CF11 to 345 in the reference strain of *C. michiganensis sensu stricto* NCPB 382 and type strain of *C. sepedonicus* ATCC 33113^T. Although the feature counts were similar in most of the subsystems among the pathogenic and non-pathogenic *Clavibacter* spp. strains, differences in the siderophore producing subsystems were observed, and siderophore assembly kit was detected only in the non-pathogenic strain CFBP 8616 (Table 4).

One-vs.-one BLASTn- and BLASTp-based explorations using the complete genome sequence of *C. michiganensis sensu stricto* NCPB 382 as the reference genome vs. the individual *Clavibacter* spp. strains revealed the lack of pathogenicity determinant genes/clusters in all the tomato-associated non-pathogenic strains evaluated in this study (Table 5). For the *chp* gene cluster (*i.e.* loci CMM_0034 to CMM_0077 in NCPB 382 genome sequence: AM711867.1) only a fraction of the genes were detected in the non-pathogenic strains (Figure 2). For instance, a sugar phosphate isomerase (CMM_0034) was present in all the non-pathogenic strains. A putative phosphotransferase (CMM_0065) and ATPase (*parX*=CMM_0066) were found in CFBP 7491, CFBP 7493, and LMG 26808. A hypothetical protein produced by CMM_0054 locus and a transcriptional

regulator protein secreted by CMM_0055 were found only in the type strain of *C. michiganensis* subsp. *californiensis*. A serine protease (*ppaD*=CMM_0075) and a putative ATPase (CMM_0067) were found only in CASJ009 and CFBP 7491, respectively, while a putative DNA invertase (CMM_PS_07) was found in both the last strains. Among the pathogenicity determinant genes inside the *chp* gene cluster, *ppaA* (CMM_0041), *pelA1* (CMM_0043), *pelA2* (CMM_0051), *chpC* (CMM_0052), and *chpG* (CMM_0059) were found in none of the evaluated non-pathogenic strains. Non-*chp* pathogenicity determinant genes *clvG* (CMM_1963), *clvF* (CMM_1964), *clvA* (*mica*=CMM_1967), and *perF* (CMM_2382) were also not detected in the evaluated non-pathogenic bacterial strains. Interestingly, a subtilisin-like serine proteases (*sbtA*=CMM_0070) was found in a number of non-pathogenic strains (Table 5), while the nucleotides 1-600 were missing in all the non-pathogenic members. The expansin encoding gene *expA* (CMM_1480) was found in CFBP 8017, DOAB 609, CFBP 7493 and LMG 26808.

As for the *tomA* gene cluster (CMM_0078 to CMM_0112 in the genome sequence of NCPPB 382) a β -glucosidase related gene (*bglC*=CMM_0083) was found in the type strains of *C. michiganensis* subsp. *chilensis*, *C. capsici*, *C. michiganensis* subsp. *phaseoli*, as well as the strains CFBP 8615, CFBP 8616, CFBP 7493, LMG 26808, and CFBP 7494. A putative Alpha-glucosidases gene (*aglA*=CMM_0106) was found in CFBP 7491, CFBP 8019, and CASJ009. Furthermore, a putative ABC-type sugar transport permease (CMM_0108) was found in CFBP 7491 and CASJ009, while the *srtA* gene (CMM_0013) which encodes a putative sortase enzyme was found in all the evaluated non-pathogenic strains in this study (Table 5).

We have also assessed the presence of a set of eight virulence genes *i.e.* *celB* (CMM_2443), *pelA1* (CMM_0043), *pelA2* (CMM_0051), *xysA* (CMM_1673), *xysB* (CMM_1674), CMM_2691,

276 CMM_2692, and CMM_2871, which are responsible for cell wall degradation at the later stages of
277 tomato infection by *C. michiganensis sensu stricto*. The polygalacturonase encoding locus
278 CMM_2871 was found in tomato-associated strains CFBP 7493 and LMG 26808, as well as the type
279 strains of *C. insidiosus* (LMG 3663^T), *C. nebraskensis* (NCPBP 2581^T) and *C. tessellarius* (ATCC
280 33566^T). Furthermore, *celB* (CMM_2443) which is a homologue of plasmid-born *celA* gene (24)
281 was present in tomato-associated non-pathogenic strains CFBP 8017, DOAB 609, CFBP 7493, LMG
282 26808, and CFBP 7494. The four loci CMM_1673, CMM_1674, CMM_2691, and CMM_2692 were
283 present in all tomato-associated strains except for CMM_2692 which was absent in CASJ009, while
284 the query coverage varied between 30-100%, and sequence similarity varied from 75-100% among
285 the strains. Surprisingly, the virulence-associated transcriptional regulator genes *vatr1*
286 (CMM_2645) and *vatr2* (CMM_2969), which regulate *C. michiganensis sensu stricto* virulence
287 during the infection (25) were present in all the strains evaluated in this study regardless of their
288 pathogenicity status (Table 5).

289 Orthologous gene clusters were determined using OrthoVenn online service through four-vs.-four
290 and five-vs.-five designations of the representative strains from different phylogenetic lineages
291 (Figure 3A-D). Type strains of the five ex-*C. michiganensis sensu lato* subspecies shared 2,157
292 proteins in their genome sequences (Figure 3A). Although none of the type strains of *C.*
293 *michiganensis sensu stricto* and *C. nebraskensis* showed unique proteins in their sequences, type
294 strains of *C. sepedonicus*, *C. tessellarius* and *C. insidiosus* showed four, 12, and 16 unique proteins
295 among their genome sequences, respectively. Furthermore, when the two phylogenetic
296 neighboring clades of *C. michiganensis sensu stricto* (i.e. *C. michiganensis* subsp. *californiensis* and
297 CFBP 7493) were compared with the type/reference strains of *C. michiganensis sensu stricto* (LMG

7333^T and NCPPB 382), seven and 11 unique proteins were detected in the genome sequences of CFBP 7493 and *C. michiganensis* subsp. *californiensis* strain CFBP 8216^T, respectively (Figure 3B). Unique and shared proteins in the type strains *C. michiganensis* subsp. *californiensis*, *C. michiganensis* subsp. *phaseoli* and *C. michiganensis* subsp. *chilensis* as well as the atypical peach-colored strain CFBP 8615 are depicted in the Figure 3C and D.

Plasmids, Phages, and Bacteriocins

No integrative plasmid (episome) was detected using the PlasmidFinder online service in the draft genome sequences of bacterial strains investigated in this study, except for LMG 26808 in which two *Enterobacteriaceae* plasmids IncL/M(pOXA-48) and IncR were identified. Surprisingly homologues sequences to the plasmid-born *celA* gene were detected in the sequences of the type strains of *C. insidiosus* (LMG 3663^T) and *C. nebraskensis* (NCPPB 2581^T) with the query coverage of respectively 97% and 76%, and sequence similarity of 90% and 84% to the sequence of the reference strain NCPPB 382 (Table 5). On the other hand, plasmid profiling detected the two expected plasmids pCM1 (~27 kb) and pCM2 (~70 kb) in tomato-pathogenic strain of *C. michiganensis sensu stricto* ICMP 22049 (data not shown). However, type strains of *C. michiganensis* subsp. *phaseoli* and *C. michiganensis* subsp. *chilensis*, as well as the two peach-colored strains CFBP 8615 and CFBP 8616 did not carry any detectable plasmid (data not shown).

The PHASTER online service was used to detect prophage sequences within the bacterial genomes. Altogether five hypothetical prophage groups *i.e.* Gordon_Schwabeltier (NC_031255), Gordon_Smoothie (NC_030696), N15 (NC_001901), P1 (NC_005856), and Phi92 (NC_023693) were

319 detected in the *Clavibacter* spp. strains investigated in this study (Table 6). Prophages were
320 detected in tomato-associated strains CFBP 8615, CFBP 7491, LMG 26808, and CASJ009 as shown
321 in Table 6. While the strain LMG 26808 contained three prophages, only one prophage per strain
322 was detected in the strains CFBP 8615, CFBP 7491, and CASJ009. The phage Phi92, which has
323 originally been isolated from a pathogenic *Escherichia coli* strain was detected in CFBP 8615, LMG
324 26808, CASJ009, *C. sepedonicus* (ATCC 33113^T) and *C. tessellarius* (ATCC 33566^T) while each of the
325 remaining four prophages was detected in only one strain (Table 6).

326 *In silico* screening for bacteriocins and antibiotic peptides revealed distinct differences between
327 the pathogenic and non-pathogenic tomato-associated *Clavibacter* spp. strains. The lantibiotic
328 “Michiganin A” was detected in all the 12 *C. michiganensis sensu stricto* strains (data not shown)
329 but was not detected in the non-pathogenic tomato-associated strains nor in the pathogenic
330 strains on other plant species. Furthermore, sactipeptides (peptides with cysteine sulfur to α -
331 carbon crosslinks) were the most common group of bacteriocins among all the *Clavibacter* spp.
332 strains. Indeed, except for tomato-pathogenic *C. michiganensis sensu stricto* strains, all the strains
333 which contained bacteriocins has had at least one type of sactipeptides (Table 7). Linear azol(in)e-
334 containing peptides (LAPs) were detected in both the pathogenic and non-pathogenic strains,
335 while thiopeptides which are commonly produced by Actinobacteria found in the two
336 phylogenetically closely related non-pathogenic strains CFBP 7493 and LMG 26808.
337 Enterocin_AS_48, a circular bacteriocin produced by *Enterococcus* sp. was exclusively detected in
338 the strain CF11 which was originally isolated from soil in a tomato-growing greenhouse (23).

339

340 Discussion

341 In this study, using phylogenetic analyses, comparative genomics, and pathogenicity assays, we
 342 provide a novel insight into the diversity of *Clavibacter* spp. strains with a special focus on tomato-
 343 associated members of the genus. Phylogenetic analyses accomplished with ANI and dDDH
 344 calculations revealed a higher genetic diversity of *Clavibacter* spp. strains than that has so far been
 345 assumed (17). We also aim to decipher phylogenetic position of the three newly described
 346 subspecies of *C. michiganensis sensu lato* (i.e. *C. michiganensis* subsp. *californiensis*, *C.*
 347 *michiganensis* subsp. *chilensis* and *C. michiganensis* subsp. *phaseoli*). Although our results confirm
 348 that the three mentioned subspecies are no longer included in the *C. michiganensis sensu stricto*,
 349 we still used their original name in this study to avoid confusions. A formal taxonomic study would
 350 provide appropriate epithet for these taxa. On the other hand, BLAST-based comparative
 351 genomics revealed that several genes (i.e. *vatr1*, *vatr2*, *xysA*, *xysB*, and *srtA*) which had previously
 352 been identified as pathogenicity determinants, were detected in all the pathogenic and non-
 353 pathogenic tomato-associated strains, indicating further complexities in the functions of these
 354 genes (25, 26). Non-pathogenic counterparts of actinobacterial plant pathogens have frequently
 355 been isolated from a set of taxonomically diverse plant species, which were distant from the main
 356 host of the pathogen. For instance, both the pathogenic and non-pathogenic *Curtobacterium*
 357 *flaccumfaciens* strains phylogenetically closely related to the common bean pathogen *C.*
 358 *flaccumfaciens* pv. *flaccumfaciens* were isolated from solanaceous annual crops i.e. eggplant,
 359 pepper, and tomato (2). While the pathogenic and non-pathogenic strains of *C. flaccumfaciens* are
 360 not differentiable using the routine molecular techniques e.g. MLSA (27), all the non-pathogenic

361 members of *C. michiganensis sensu lato* could be differentiated from the tomato-pathogenic *C.*
 362 *michiganensis sensu stricto* strains (11).

363 Non-pathogenic strains of *Clavibacter* spp. were consistently reported to associate with seeds,
 364 transplants, and aerial portions of tomato plants (11, 13, 22, 28). However, until recently, only a
 365 few number of genome sequences from non-pathogenic *Clavibacter* spp. strains were available
 366 (22, 28) limiting our understanding of the putative role of these bacteria on the host plants and
 367 their environment. In a preliminary complete genome sequence-based comparative study, Zaluga
 368 et al. (28) investigated the genome of tomato-associated non-pathogenic strain LMG 26808 and
 369 provided initial insights into the genetic bases of differences between the pathogenic and non-
 370 pathogenic members of *C. michiganensis sensu lato*. However, it has been noted that LMG 26808
 371 is phylogenetically very close to the *C. michiganensis sensu stricto* clade leaving a greater portion
 372 of non-pathogenic *Clavibacter* spp. diversity uninvestigated (28). Our results revealed that the
 373 strain LMG 26808 as well as two other strains *i.e.* CFBP 7493 and CFBP 8216^T are phylogenetically
 374 closely related, and fall into a monophyletic clade along with the pathogenic members of *C.*
 375 *michiganensis sensu stricto* (Figure 1). Genomic contents of the two clades represented by LMG
 376 26808/CFBP 7493 as “hypothetical new species I” and CFBP 8216^T as “hypothetical new species II”
 377 varied in the pathogenicity-related genes *sbtA*, *expA*, *celB*, and the locus CMM_2871, while there
 378 was no difference between the strains LMG 26808 and CFBP 7493 in the evaluated genomic areas
 379 (Table 5). More specifically, the *expA* gene (CMM_1480) which is responsible for expansin
 380 production (29), as well as a polygalacturonase encoded by CMM_2871 locus at the final stages of
 381 infection were found in CFBP 7493 and LMG 26808 strains but not in the type strain of *C.*
 382 *michiganensis* subsp. *californiensis* (Table 5).

383 Recently, Li and his colleagues (17) re-evaluated the taxonomy of *C. michiganensis sensu lato* and
384 proposed the re-classification of each of the ex-*C. michiganensis* subspecies into the species
385 status. Since only two genome sequences of *C. insidiosus* were available at that time (both isolated
386 in the USA; 29), we sequenced two further “Old World” strains (*i.e.* CFBP 1195 and CFBP 6488
387 isolated in the UK and Czech Republic, respectively) to gain a precise vision from the intra-species
388 diversity of the alfalfa pathogen (20). Our analyses confirm the existence of *C. capsici*, *C.*
389 *insidiosus*, *C. michiganensis sensu stricto*, *C. nebraskensis*, *C. sepedonicus* and *C. tessellarius* as
390 stand-alone species. While the 95% ANI did not solely support the separation of the alfalfa and
391 maize pathogens, 59.90% dDDH between the type strains of the two taxa, as well as their distinct
392 host plants could be considered as the evidences for the separation of *C. insidiosus* and *C.*
393 *nebraskensis* (Tables 2, 3). The strain CFBP 7494 which was isolated from tomato seeds and causes
394 disease symptoms on wheat plants in greenhouse conditions (22) clustered in a monophyletic
395 clade with the alfalfa pathogenic strains and still fall into *C. insidiosus* species with 96% ANI and
396 64.60% dDDH. While the phylogenetic position of the strain CFBP 7494 was clarified in these
397 analyses, only further investigations using a larger collection of strains will shed a light on the
398 genetic content, biological characteristics and taxonomic status of tomato-associated wheat-
399 pathogenic members of *Clavibacter* spp.

400 Draft genome sequences of the three new subspecies of *C. michiganensis sensu lato* revealed their
401 phylogenetic position, highlighting inaccuracy in the nomenclature of *C. michiganensis* subsp.
402 *chilensis* and *C. michiganensis* subsp. *phaseoli*. Type strains of these two subspecies shared 99%
403 ANI and 87.50% dDDH with one another, indicating a synonymy and orientating to the proposal of
404 a novel unique new species (Figure 1, Table 3). Type strain of *C. michiganensis* subsp. *phaseoli*

(CFBP 8627^T) was isolated from common bean seeds in Spain and reported to cause bacterial bean leaf yellowing disease in greenhouse assays (14). However, we could not observe any symptoms on the inoculated common bean plants even when three different cultivars were evaluated in the pathogenicity tests (Table 1). This could be attributed to differences in the environmental conditions between the two assays, and probably also differences in the susceptibility of common bean cultivars used in the two studies. Further field surveys are needed to decipher potential natural occurrence in the field conditions, and putative frequency and prevalence of the common bean-associated *Clavibacter* spp. strains.

Comparative genomics revealed that tomato-associated *C. michiganensis sensu lato* strains are adapted to a non-pathogenic lifestyle, which is reflected by the lack of pathogenicity gene clusters present in the pathogenic members (Table 5; Figure 2). Although the absence of almost all 129-kb *chp/tomA* region was common among all the non-pathogenic strains, some of the putative virulence factors were present in the non-pathogenic strains, suggesting contribution of these genes in the endophytic lifestyle of the bacteria. An in-depth comparative analysis with newly sequenced *Clavibacter* spp. genomes allowed us to illustrate a more precise insight underlying genetic contents of these bacteria. For instance the *expA* gene was detected in the non-pathogenic strains CFBP 7493 and LMG 26808 as well as the wheat-pathogenic strains CFBP 8017 and DOAB 609, but not in the type strain of *C. tessellarius* (Table 5). Microbial expansins are found in the genomes of several plant pathogenic bacteria, and it is assumed that they provide particular advantages to xylem-dwelling phytopathogens (29, 30). Expansin enhances cellulose breakdown by cellulase enzymes in the later stages of pathogen invasion (31). These observations correlate with the initial assumptions that non-pathogenic *Clavibacter* spp. strains must have lost or never

427 contained prominent virulence determinants (e.g. 129-kb *chp/tomA* region) responsible for
428 disease induction in tomato plants. With the availability of genome sequences covering a broader
429 diversity of non-pathogenic *Clavibacter* spp. strains one would assume the gene flow and
430 evolutionary pathways of pathogenicity determinants in the genus similar to that has previously
431 been estimated for plant pathogenic xanthomonads (4).

432 In conclusion, our results obtained from the analyses of 40 genome sequences provide a
433 comprehensive insight into the genetic diversity of *Clavibacter* spp., and confirm the recent
434 taxonomic revision of the genus. However, phylogenetic analyses suggest that the recently
435 described subspecies *C. michiganensis* subsp. *chilensis* and *C. michiganensis* subsp. *phaseoli* should
436 be classified as members of the same novel species (13, 14). Taking together all the phylogenetic,
437 genomics and pathogenicity data, nine hypothetical novel species could be identified within
438 *Clavibacter* spp., seven of which (i.e. hypothetical new species I, II, III, IV, V, VIII, and IX as shown in
439 Figure 1) were isolated from asymptomatic tomato tissues or seed lots. These findings raise a
440 question whether current taxonomy of tomato-associated *Clavibacter* spp. strains is technically
441 applicable to the quarantine purposes, and emphasize at the same time the need for more
442 detailed taxonomic investigations among the phylogenetically diverse tomato-associated
443 *Clavibacter* spp. strains. Indeed, the only pathogenic lineage of tomato-associated strains is *C.*
444 *michiganensis sensu stricto* while the seven non-pathogenic lineages need to designate into novel
445 formal taxa. This would help the plant pathology agencies and tomato seed industry inspectors to
446 specifically target the enemy and neglect the non-pathogenic lineages. Only a formal taxonomic
447 study would address this issue with delineation of appropriate epithet and species description for
448 these new taxa. On the other hand, the nine pathogenicity determinant genes (Table 5) would be

449 appropriate targets for the development of novel genome-informed detection methods for
 450 differentiation of tomato-pathogenic and non-pathogenic strains.

451

452 **Materials and Methods**

453 **Bacterial Strains and Genome Sequences**

454 Draft genome sequences of 10 *Clavibacter* spp. strains (Table 2) were prepared using the shotgun
 455 genome sequencing facility of Illumina HiSeq X platform. Culture media, bacterial growth
 456 conditions, genomic DNA preparation, sequencing procedure and genome annotation were
 457 described previously (20). In this framework, we investigated the type strains of *C. michiganensis*
 458 subsp. *californiensis* (CFBP 8216^T), *C. michiganensis* subsp. *chilensis* (CFBP 8217^T) and *C.*
 459 *michiganensis* subsp. *phaseoli* (CFBP 8627^T), two *C. insidiosus* (CFBP 1195 and CFBP 6488), one *C.*
 460 *nebraskensis* (CFBP 7577), as well as the non-pathogenic peach colored (*i.e.* CFBP 8615 and CFBP
 461 8616) and yellow-pigmented (*i.e.* CFBP 7491 and CFBP 7493) strains. Furthermore, all the publicly
 462 available genome sequences of *Clavibacter* spp. – until April 2019 – were retrieved from the NCBI
 463 GenBank database and included in the phylogenetic analysis and comparative genomics. Table 2
 464 describes the 40 *Clavibacter* spp. strains used in this study, their origin of isolation and
 465 pathogenicity features.

466

467 **Pathogenicity Tests and Host Range**

Due to the close phylogenetic relationships between the type strains of *C. michiganensis* subsp. *phaseoli* (CFBP 8627^T) and *C. michiganensis* subsp. *chilensis* (CFBP 8217^T), these two strains as well as the atypical peach-colored strains recently isolated from tomato (16) were subjected to the pathogenicity tests and host range assays in greenhouse conditions. Pathogenicity tests were performed on bell pepper (cv. Sereno), chili pepper (cv. Aziz), common bean (cvs. Red kidney, Pinto, and Navy), cowpea (*Vigna unguiculata* cv. Partow), mung bean (*Vigna radiata* cv. Mashhad), and tomato (cv. Sunseed 6189) plants. Plant growth conditions, inoculation procedure, and incubation environment were the same as detailed previously (32, 33). Inoculated plants were periodically monitored for the appearance of disease symptoms up to 30 days post inoculation. Positive and negative control plants were treated in the same manner using the standard strain of *C. michiganensis sensu stricto* (ICMP 22049; isolated from symptomatic tomato plant in Iran in 2015; 16) and sterile distilled water, respectively. Koch's postulates were accomplished by re-isolating the inoculated strains on yeast-extract peptone glucose agar (YPGA) medium from all inoculated plants. Confirmation of the identity of the re-isolated bacteria was made by determining Gram reaction and colony characteristics on yeast extract-dextrose-calcium carbonate (YDC) agar medium as well as by using the genus-specific primer pair CMR16F1/CMR16R1 (34) as described previously (16). The pathogenicity tests were conducted twice.

Phylogenetic Analyses

488 Average Nucleotide Identity (ANI) was calculated among all the *Clavibacter* spp. genome
 489 sequences included in this study. The ANI was estimated using both one-vs.-one and all-vs.-all
 490 strategies via different algorithms *i.e.* JSpeciesWS (<http://jspecies.ribohost.com/jspeciesws/>; 35),
 491 ANI calculator (<http://enve-omics.ce.gatech.edu/g-matrix/>; 36), and OrthoANlu
 492 (<https://www.ezbiocloud.net/tools/orthoani/>; 37). An ANI-based Neighbor Joining phylogenetic
 493 tree was constructed using the ANI calculator online service, and genome sequence of *Leifsonia*
 494 *xyli* subsp. *cynodontis* (DSM 46306; NC_022438.1) was used as an out-group in the tree.
 495 Additionally, Genome-to-Genome Distance Calculator online service
 496 (<http://ggdc.dsmz.de/distcalc2.php>) was used to calculate digital DNA-DNA hybridization (dDDH)
 497 value which infers to the genome-to-genome distances between pairs of genomes based on the
 498 Genome Blast Distance Phylogeny (38). A combination of ANI and dDDH indices was used to
 499 designate a taxonomic status to a given phylogenetic clade, where the “new species” status was
 500 assigned to a clade only when both ANI and dDDH values were below the accepted threshold
 501 ($\leq 95\%$ and $\leq 70\%$ for ANI and dDDH, respectively, 21)

502

503 **Comparative Genomics**

504 Twenty strains representing the entire genetic diversity of *Clavibacter* spp. based on the
 505 ANI/dDDH data, host of isolation and pathogenicity characteristics were subjected to the
 506 comparative genomics analyses. Type strains of all the *C. michiganensis sensu lato*
 507 species/subspecies, as well as all the individual strains sharing $\leq 95\%$ and $\leq 70\%$ ANI and dDDH
 508 values, respectively, with the other taxa were selected for comparative genomics analyses.

509 Genome length (bp), G + C content (%), total number of protein-coding sequences (CDS), RNA
 510 genes, and pseudogenes were determined for all the genomes.

511 The online annotating service RAST (Rapid Annotations using Subsystems Technology;
 512 <http://rast.nmpdr.org/>; 39) was used for fully automated annotation of the bacterial genomes and
 513 the obtained information was used to reconstruct metabolic networks and subsystems. A
 514 subsystem is a set of functional roles that the annotator considers as related categories.
 515 Subsystems represent a collection of functionally related protein families that make up a
 516 metabolic pathway (*e.g.* Iron acquisition and metabolism), a complex (*e.g.* the ribosome), or a
 517 class of proteins (*e.g.* bacteriocins) (40). Subsequently, the genomes were transferred to the
 518 comparative environment of the SEED-Viewer (http://www.theseed.org/wiki/Main_Page; 41) for
 519 comparative genomics analyses. The SEED-Viewer was used for the identification of protein-
 520 encoding sequences (CDS), assigning functions to the genes, and prediction of represented gene
 521 clusters in the genomes. Distribution of the genes among various clusters, and specific protein
 522 encoding genes within each cluster was estimated using the same service. Furthermore,
 523 BLASTn/BLASTp-based investigation was performed to decipher whether the pathogenicity
 524 determinant genes/clusters are present in the genomes (26). Using the complete genome of *C.*
 525 *michiganensis sensu stricto* NCPPB 382, one-vs.-one BLASTn/BLASTp search was accomplished
 526 against the sequences of the pathogenicity island (a 129-kb low G+C region which includes *chp* and
 527 *tomA* clusters) as well as several individual genes proposed to have effective contribution to the
 528 virulence of *C. michiganensis sensu stricto* (22, 25, 26, 28, 42, 43). Proteins with amino acid
 529 sequence similarities higher than 50% and with a query coverage higher than 70% were
 530 considered homologs (28).

On the other hand, we screened the genome sequences for the presence of hypothetical bacteriocin-encoding genes/clusters using the web-based tool BAGEL4 (<http://bagel4.molgenrug.nl/>, 44). BAGEL4 combines direct mining for the structural genes with indirect mining for bacteriocin-associated genes. Furthermore, the online service PlasmidFinder 2.0 (<https://cge.cbs.dtu.dk/services/PlasmidFinder/>; 45) was used for the screening of all the genomic sequences for presence of integrative plasmids/ episomes. Identification and annotation of prophage sequences within bacterial genomes were performed using the online service PHASTER (PHAge Search Tool Enhanced Release; <http://phaster.ca/>; 46). Given to the fact that identification of overlaps among the orthologous clusters can enable us to elucidate the function and evolution of proteins across multiple species, genome wide comparisons and visualization of orthologous clusters were performed using the online service OrthoVenn (<http://www.bioinfogenome.net/OrthoVenn/>; 47). The analyses were conducted on the “bacteria” section of the platform using default settings (E-value: 1e-5 and inflation value: 1.5). Regarding the numeric limitation of OrthoVenn in the handling of bacterial genomes (up to six genomes per run) different series of the strains were evaluated using the same parameters.

Plasmid Profiling

In order to further investigate the genetic contents of non-pathogenic tomato-associated *Clavibacter* spp. strains, we evaluated the plasmid profile of the type strains of *C. michiganensis* subsp. *chilensis* (CFBP 8217^T) and *C. michiganensis* subsp. *phaseoli* (CFBP 8627^T) as well as the two peach-colored strains CFBP 8615 and CFBP 8616 isolated from tomato. Plasmids were isolated

552 according to the procedure described by Kotchoni et al. (48) with minor modifications. Tomato-
 553 pathogenic strain of *C. michiganensis sensu stricto* ICMP 22049 was used as positive control (49).
 554 Bacterial strains were grown in 50 ml Luria-Bertani (LB) medium on a 150 rpm shaker at 27 °C for
 555 48 h. Bacterial cells were harvested by centrifugation at 13,000 g for 1 min at room temperature,
 556 the pellet was re-suspended in 200 µl of “solution I” of Kotchoni et al. (48) protocol, mixed well,
 557 and incubated at 37 °C for 20 min. Subsequently, 400 µl of freshly prepared “solution II” was
 558 added into the microtubes and mixed well by inverting gently four to six times to avoid breaking
 559 the plasmid(s). Immediately 200 µl of “solution III” was added, mixed very gently, and incubated
 560 at 4 °C for 15 min without any intervention. The mixture was centrifuged at 10,000 g for 5 min,
 561 supernatant was transferred to a new microtube, 0.6 volume of isopropanol was added to the
 562 supernatant, mixed gently by inverting four to six times, and kept at room temperature for 10 min.
 563 Then, centrifuged at 10,000 g for 5 min and the supernatant was discarded. The pellet containing
 564 precipitated plasmid DNA was washed with 400 µl of 70 % (v/v) ethanol and centrifuged at 10,000
 565 g for 3 min at room temperature. Supernatant was removed and the pellet was air-dried. Finally,
 566 plasmid DNA was re-suspend in 50 µl sterile distilled water containing 10 mg/ml RNase A.
 567 Presence of plasmids was analyzed on 0.6 % agarose gel as described previously (27).

568 **Acknowledgment**

570 We thank the French Collection of Plant-Associated Bacteria (CIRM-CFBP;
 571 http://www6.inra.fr/cirm_eng/CFBP-Plant-Associated-Bacteria) for strain conservation and supply,
 572 CATI BBRIC for bioinformatics facilities, and the French Network on Xanthomonads (FNX)

573 (<https://www.reseau-xantho.org/>) for recurrent scientific exchanges. Authors benefited from
574 interactions promoted by COST Action CA16107 EuroXanth (<https://euroxanth.eu/>).

575

576 **Conflict of Interest Statement**

577 The authors declare that the research was conducted in the absence of any commercial or
578 financial relationships that could be construed as a potential conflict of interest.

579

580 **Data availability**

581 The datasets generated for this study can be found in the NCBI GenBank Database.

582

583 **Author Contribution**

584 EO and MAJ conceived and designed the study, with assistance from TR. EO, MA, SMT, and SZ
585 carried out the experiments. MB and PP performed the genome sequencing and annotation. EO
586 analyzed and interpreted the data with assistance from MAJ, TR, and MB. EO prepared the
587 manuscript with assistance from MAJ. All the co-authors revised the final manuscript, while EO
588 and MAJ acted as the corresponding authors.

589

590 **Funding**

591 Financial support for this study was co-provided by Shiraz University (Iran) and INRA, Agrocampus-
592 Ouest, Université d'Angers (France).

593

594 References

595 1- Jacques MA, Arlat M, Boulanger A, Boureau T, Carrère S, Cesbron S, Chen NW, Cociancich S,
596 Darrasse A, Denancé N, Fischer-Le Saux M, Gagnevin L, Koebnik R, Lauber E, Noël LD,
597 Pieretti I, Portier P, Pruvost O, Rieux A, Robène I, Royer M, Szurek B, Verdier V, Vernière C.
598 (2016). Using Ecology, Physiology, and Genomics to Understand Host Specificity in
599 *Xanthomonas*. Annu Rev Phytopathol. 54:163-187.

600 2- Osdaghi E, Taghavi SM, Hamzehzarghani H, Fazliarab A, Harveson RM, Tegli S, Lamichhane JR
601 (2018). Epiphytic *Curtobacterium flaccumfaciens* strains isolated from symptomless
602 solanaceous vegetables are pathogenic on leguminous but not on solanaceous plants.
603 Plant Pathology, 67: 388-398.

604 3- Mafakheri H, Taghavi SM, Puławska J, de Lajudie P, Lassalle F, Osdaghi E (2019). Two Novel
605 Genomospecies in the *Agrobacterium tumefaciens* Species Complex Associated with Rose
606 Crown Gall. Phytopathology. (In Press) DOI: 10.1094/PHYTO-05-19-0178-R

607 4- Merda D, Briand M, Bosis E, Rousseau C, Portier P, Barret M, Jacques MA, Fischer-Le Saux M.
608 (2017). Ancestral acquisitions, gene flow and multiple evolutionary trajectories of the type
609 three secretion system and effectors in *Xanthomonas* plant pathogens. Molecular Ecology.
610 26(21):5939-5952.

- 611 5- Triplett LR, Verdier V, Campillo T, Van Malderghem C, Cleenwerck I, Maes M, Deblais L, Corral R,
612 Koita O, Cottyn B, Leach JE. (2015). Characterization of a novel clade of *Xanthomonas*
613 isolated from rice leaves in Mali and proposal of *Xanthomonas maliensis* sp. nov. *Antonie*
614 *van Leeuwenhoek* 107(4): 869-881
- 615 6- Eichenlaub R, Gartemann KH, 2011. The *Clavibacter michiganensis* Subspecies: Molecular
616 Investigation of Gram-Positive Bacterial Plant Pathogens. *Annual Review of Phytopathology*
617 49, 445-464.
- 618 7- Gitaitis R, Walcott R. (2007). The epidemiology and management of seedborne bacterial
619 diseases. *Annu Rev Phytopathol.* 45: 371-97.
- 620 8- Schaad NW, Abrams J, Madden LV, Frederick RD, Luster DG, Damsteegt VD, and Vidaver AK
621 (2006). An assessment model for rating high-threat crop pathogens. *Phytopathology*
622 96:616-621.
- 623 9- Davis M.J., Gillaspie A.G., Vidaver A.K., Harris R.W., 1984. *Clavibacter*: a new genus containing
624 some phytopathogenic coryneform bacteria, including *Clavibacter xyli* subsp. *xyli* sp. nov.,
625 subsp. nov. and *Clavibacter xyli* subsp. *cynodontis* subsp. nov., pathogens that cause ratoon
626 stunting disease of sugarcane and Bermudagrass stunting disease. *International Journal of*
627 *Systematic Bacteriology* 34: 107-117.
- 628 10- Dye DW, and Kemp WJ. 1977. A taxonomic study of plant pathogenic *Corynebacterium*
629 species. *New Zeal J Agr Res*, 20:563-582.
- 630 11- Jacques MA, Durand K, Orgeur G, Balidas S, Fricot C, Bonneau S, Quillévéré A, Audusseau C,
631 Olivier V, Grimault V, Mathis R, 2012. Phylogenetic analysis and polyphasic characterization

- 632 of *Clavibacter michiganensis* strains isolated from tomato seeds reveal that nonpathogenic
633 strains are distinct from *C. michiganensis* subsp. *michiganensis*. Applied and Environmental
634 Microbiology 78, 8388-8402.
- 635 12- EPPO (2016). PM 7/42 (3) *Clavibacter michiganensis* subsp. *michiganensis*. Bulletin
636 OEPP/EPPO Bulletin.46, 202–225.
- 637 13- Yasuhara-Bell J. and Alvarez A.M. (2015). Seed-associated subspecies of the genus *Clavibacter*
638 are clearly distinguishable from *Clavibacter michiganensis* subsp. *michiganensis*.
639 *International Journal of Systematic and Evolutionary Microbiology* **65**, 811–826.
- 640 14- Gonzalez AJ, Trapiello E, 2014. *Clavibacter michiganensis* subsp. *phaseoli* subsp. nov.,
641 pathogenic in bean. *International Journal of Systematic and Evolutionary Microbiology* **64**,
642 1752–1755.
- 643 15- Oh EJ, Bae C, Lee HB, Hwang IS, Lee HI, Yea MC, Yim KO, Lee S, Heu S, Cha JS, Oh CS,
644 2016. *Clavibacter michiganensis* subsp. *capsici* subsp. nov., causing bacterial canker disease
645 in pepper. *International Journal of Systematic and Evolutionary Microbiology* **66**, 4065-
646 4070.
- 647 16- Osdaghi E, Ansari M, Taghavi SM, Zarei S, Koebnik R, Lamichhane JR. (2018). Pathogenicity and
648 phylogenetic analysis of *Clavibacter michiganensis* strains associated with tomato plants in
649 Iran. Plant pathol, 67:957-970.
- 650 17- Li X, Tambong J, Yuan KX, Chen W, Xu H, Lévesque CA, De Boer SH. 2018. Re-classification of
651 *Clavibacter michiganensis* subspecies on the basis of whole-genome and multi-locus
652 sequence analyses. Int J Syst Evol Microbiol. 68:234-240.

- 653 18- Nouiou I, Carro L, García-López M, Meier-Kolthoff JP, Woyke T, Kyrpides NC, Pukall R, Klenk H-
654 P, Goodfellow M and Göker M (2018) Genome-Based Taxonomic Classification of the
655 Phylum *Actinobacteria*. Front. Microbiol. 9:2007. doi: 10.3389/fmicb.2018.02007
- 656 19- Li X and De Boer SH. (2019). Descriptions of *Clavibacter insidiosus* sp. nov. and *Clavibacter*
657 *tessellarius* sp. nov. Int J Syst Evol Microbiol DOI 10.1099/ijsem.0.003439
- 658 20- Osdaghi E, Portier P, Briand M, Taghouthi G, Jacques M-A (2018). Draft genome sequences of
659 the type strains of three *Clavibacter* subspecies, and atypical peach-colored strains isolated
660 from tomato. Microbiology Resource Announcements. 7:e01357-18.
- 661 21- Kim M, Oh HS, Park SC, Chun J, 2014. Towards a taxonomic coherence between average
662 nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of
663 prokaryotes. *International Journal of Systematic and Evolutionary Microbiology* 64, 346–
664 351.
- 665 22- Thapa SP, Pattathil S, Hahn MG, Jacques MA, Gilbertson RL, Coaker G, 2017. Genomic Analysis
666 of *Clavibacter michiganensis* Reveals Insight Into Virulence Strategies and Genetic Diversity
667 of a Gram-Positive Bacterial Pathogen. *Molecular Plant-Microbe Interactions*, 30(10):786-
668 802.
- 669 23- Du Y, Yuan B, Zeng Y, Meng J, Li H, Wang R, Li G, Feng F. (2015). Draft Genome Sequence of the
670 Cellulolytic Bacterium *Clavibacter* sp. CF11, a Strain Producing Cold-Active Cellulase.
671 Genome Announc. 3(1). pii: e01304-14.

- 672 24- Jahr H, Dreier J, Meletzus D, Bahro R, Eichenlaub R, 2000. The endo- β -1,4-glucanase *CelA* of
673 *Clavibacter michiganensis* subsp. *michiganensis* is a pathogenicity determinant required for
674 induction of bacterial wilt of tomato. *Molecular Plant-Microbe Interactions* 13, 703–714.
- 675 25- Savidor, A., Chalupowicz, L., Teper, D., Gartemann, K. H., Eichenlaub, R., Manulis-Sasson, S.,
676 Barash, I., and Sessa, G. 2014. *Clavibacter michiganensis* subsp. *michiganensis* Vatr1 and
677 Vatr2 transcriptional regulators are required for virulence in tomato. *Mol. Plant-Microbe*
678 *Interact* 27:1035-1047.
- 679 26- Gartemann KH, Abt B, Beke T, Burger A, Engemann J, Flugel M, Gaigalat L, Goesmann A, Grafen
680 I, Kalinowski J, Kaup O, Kirchner O, Krause L, Linke, B, McHardy A, Meyer F, Pohle S,
681 Ruckert C, Schneiker S, Zellermann E, Puhler A, Eichenlaub R, Kaiser O, Bartels D, 2008. The
682 genome sequence of the tomato-pathogenic actinomycete *Clavibacter michiganensis*
683 subsp *michiganensis* NCPPB 382 reveals a large island involved in pathogenicity. *Journal of*
684 *Bacteriology* 190, 2138–2149.
- 685 27- Osdaghi E, Taghavi SM, Calamai S, Biancalani C, Cerboneschi M, Tegli S, Harveson RM (2018).
686 Phenotypic and Molecular-Phylogenetic Analysis Provide Novel Insights into the Diversity
687 of *Curtobacterium flaccumfaciens*. *Phytopathology*. 108: 1154-1164
- 688 28- Zaluga J, Stragier P, Baeyen S, Haegeman A, Van Vaerenbergh J, Maes M, De Vos P. (2014).
689 Comparative genome analysis of pathogenic and non-pathogenic *Clavibacter* strains
690 reveals adaptations to their lifestyle. *BMC Genomics* 2014 15:392.

- 691 29- Lu Y, Samac DA, Glazebrook J, Ishimaru CA (2015) Complete Genome Sequence of *Clavibacter*
692 *michiganensis* subsp. *insidiosus* R1-1 Using PacBio Single-Molecule Real-Time Technology.
693 Genome Announc. 3(3). pii: e00396-15. doi: 10.1128/genomeA.00396-15.
- 694 30- Tancos MA, Lowe-Power TM, Peritore-Galve FC, Tran TM, Allen C, Smart CD (2018). Plant-like
695 bacterial expansins play contrasting roles in two tomato vascular pathogens. Mol Plant
696 Pathol. 19(5):1210-1221. doi: 10.1111/mpp.12611
- 697 31- Georgelis N, Nikolaidis N, and Cosgrove DJ. (2015). Bacterial expansins and related proteins
698 from the world of microbes. Appl Microbiol Biotechnol. 99(9): 3807–3823.
- 699 32- Yaripour Z, Taghavi SM, Osdaghi E, Lamichhane JR (2018). Host range and phylogenetic
700 analysis of *Xanthomonas alfalfae* causing bacterial leaf spot of alfalfa in Iran. European
701 Journal of Plant Pathology, 150: 267-274.
- 702 33- Osdaghi E, Taghavi SM, Hamzehzarghani H, Fazliarab A, Harveson RM, Lamichhane JR (2016).
703 Occurrence and characterization of a new red-pigmented variant of *Curtobacterium*
704 *flaccumfaciens*, the causal agent of bacterial wilt of edible dry beans in Iran. European
705 Journal of Plant Pathology, 146: 129-145.
- 706 34- Lee IM, Bartoszyk IM, Gundersen DE, Mogen B, Davis RE, 1997. Nested PCR for ultrasensitive
707 detection of the potato ring rot bacterium, *Clavibacter michiganensis* subsp. *sepedonicus*.
708 *Applied and Environmental Microbiology* 63, 2625–2630.
- 709 35- Richter M, Rosselló-Móra R, Oliver Glöckner F, Peplies J. 2016. JSpeciesWS: a web server for
710 prokaryotic species circumscription based on pairwise genome comparison. Bioinformatics,
711 32:929-931.

- 712 36- Rodriguez-R LM, Konstantinidis KT. 2016. The enveomics collection: a toolbox for specialized
713 analyses of microbial genomes and metagenomes. PeerJ Preprints 4:e1900:v1.
714 <https://doi.org/10.7287/peerj.preprints.1900v1>.
- 715 37- Yoon, S. H., Ha, S. M., Lim, J. M., Kwon, S.J. & Chun, J. (2017). A large-scale evaluation of
716 algorithms to calculate average nucleotide identity. *Antonie van Leeuwenhoek*. 110:1281–
717 1286.
- 718 38- Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M. (2013). Genome sequence-based species
719 delimitation with confidence intervals and improved distance functions. *BMC*
720 *Bioinformatics*. 21;14:60. doi: 10.1186/1471-2105-14-60.
- 721 39- Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM,
722 Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D,
723 Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A,
724 Zagnitko O. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC*
725 *Genomics*. 9:75. doi: 10.1186/1471-2164-9-75.
- 726 40- Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B,
727 Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. 2014. The SEED and the Rapid
728 Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res*
729 42:D206–D214.
- 730 41- Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, Cohoon M, de Crécy-Lagard V,
731 Diaz N, Disz T, Edwards R, Fonstein M, Frank ED, Gerdes S, Glass EM, Goesmann A, Hanson
732 A, Iwata-Reuyl D, Jensen R, Jamshidi N, Krause L, Kubal M, Larsen N, Linke B, McHardy AC,

- 733 Meyer F, Neuweiger H, Olsen G, Olson R, Osterman A, Portnoy V, Pusch GD, Rodionov DA,
734 Rückert C, Steiner J, Stevens R, Thiele I, Vassieva O, Ye Y, Zagnitko O, Vonstein V. (2005).
735 The subsystems approach to genome annotation and its use in the project to annotate
736 1000 genomes. *Nucleic Acids Res.* 33(17):5691-702.
- 737 42- Yasuhara-Bell, J., Marrero, G. & Alvarez, A. M. (2014). Genes *clvA*, *clvF* and *clvG* are unique to
738 *Clavibacter michiganensis* subsp. *michiganensis* and highly conserved. *Eur J Plant Pathol*
739 140, 655–664.
- 740 43- Chalupowicz, L., Cohen-Kandli, M., Dror, O., Eichenlaub, R., Gartemann, K. H., Sessa, G.,
741 Barash, I., and Manulis-Sasson, S. 2010. Sequential expression of bacterial virulence and
742 plant defense genes during infection of tomato with *C. michiganensis* subsp.
743 *michiganensis*. *Phytopathology* 100:252-261.
- 744 44- de Jong A, van Hijum SA, Bijlsma JJ, Kok J, Kuipers OP. (2006). BAGEL: a web-based bacteriocin
745 genome mining tool. *Nucleic Acids Res.* 34(Web Server issue):W273-9.
- 746 45- Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, Villa L, Møller Aarestrup
747 F, Hasman H. (2014). In silico detection and typing of plasmids using PlasmidFinder and
748 plasmid multilocus sequence typing. *Antimicrob Agents Chemother.* 58(7):3895-3903. doi:
749 10.1128/AAC.02412-14. Epub 2014 Apr 28.
- 750 46- Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS (2016). PHASTER: a better,
751 faster version of the PHAST phage search tool. *Nucleic Acids Res.* 44(W1):W16-21.

- 752 47- Wang Y, Coleman-Derr D, Chen G, Gu YQ (2015). OrthoVenn: a web server for genome wide
753 comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids*
754 *Res.* 43(W1):W78-84.
- 755 48- Kotchoni, S. O., Gachomo, E. W., Betiku, E., Shonukan O. O. 2003. A home made kit for plasmid
756 DNA mini-preparation. *African Journal of Biotechnology.* 2: 88–90.
- 757 49-Ansari M, Taghavi SM, Hamzehzarghani H, Valenzuela M, Siri MI, Osdaghi E (2019). Multiple
758 Introductions of Tomato Pathogen *Clavibacter michiganensis* subsp. *michiganensis* into
759 Iran as Revealed by a Global-Scale Phylogeographic Analysis. *Applied and Environmental*
760 *Microbiology.* 85:e02098-19
- 761 50- Tambong JT, Xu R, Daayf F, Brière S, Bilodeau GJ, Tropiano R, Hartke A, Reid LM, Cott M, Cote
762 T, Agarkova I. (2016). Genome Analysis and Development of a Multiplex TaqMan Real-Time
763 PCR for Specific Identification and Detection of *Clavibacter michiganensis* subsp.
764 *nebraskensis*. *Phytopathology.* 106(12):1473-1485.
- 765 51- Li X, Yuan X. 2017. Genome sequences for multiple *Clavibacter* strains from different
766 subspecies. *Genome Announc* 5:e00721-17. DOI: 10.1128/genomeA.00721-17
- 767 52- Vasilenko OV, Starodumova IP, Dorofeeva LV, Tarlachkov SV, Prisyazhnaya NV, Chizhov VN,
768 Subbotin SA, Huntemann M, Clum A, Duffy K, Pillay M, Palaniappan K, Varghese N, Chen I-
769 MA, Stamatis D, Reddy TBK, O'Malley R, Daum C, Shapiro N, Ivanova N, Kyrpides NC,
770 Woyke T, Whitman WB, Evtushenko LI. 2018. Draft genome sequences of new isolates and
771 the known species of the family *Microbacteriaceae* associated with plants. *Microbiol*
772 *Resour Announc* 7:e01051-18. <https://doi.org/10.1128/MRA.01051-18>

53- Tambong JT, Xu R, Adam Z, Cott M, Rose K, Reid LM, Daayf F, Brière S, Bilodeau GJ. 2015. Draft genome sequence of *Clavibacter michiganensis* subsp. *nebraskensis* strain DOAB 397, isolated from an infected field corn plant in Manitoba, Canada. Genome Announc 3(4):e00768-15. doi:10.1128/genomeA.00768-15.

54- Bentley SD, Corton C, Brown SE, Barron A, Clark L, Doggett J, Harris B, Ormond D, Quail MA, May G, Francis D, Knudson D, Parkhill J, Ishimaru CA (2008). Genome of the actinomycete plant pathogen *Clavibacter michiganensis* subsp. *sepedonicus* suggests recent niche adaptation. J Bacteriol, 190(6):2150–2160

Legend of Figures and Tables

Figure 1: Average Nucleotide Identity (ANI)-based Neighbor Joining phylogenetic tree of 40 *Clavibacter* spp. strains constructed using the ANI calculator online service. Different colors represent hypothetical new species (I-IX). Seven hypothetical novel species were determined among tomato-associated non-pathogenic *Clavibacter* spp. strains. Furthermore, based on the ANI/dddH indices, type strains of *C. michiganensis* subsp. *chilensis* and *C. michiganensis* subsp. *phaseoli* belong to the same novel species. **C. michiganensis* subsp. *californiensis*

Figure 2: Results of comparative genomics on the two main pathogenicity determinant regions of *Clavibacter michiganensis sensu stricto* NCPPB 382 (i.e. *chp* and *tomA*) against the non-pathogenic tomato-associated strains, as well as the type strains of five species pathogenic

794 on other plants. While only a fraction (<10%) of the clusters were detected in the non-
 795 pathogenic strains (A), variations were observed in the patterns of the genes (B).

796

797 **Figure 3:** Venn diagrams constructed using the OrthoVenn online service showing the distribution
 798 of shared gene families (orthologous clusters) among different sets of *Clavibacter* spp.
 799 strains.

800

801 **Table 1:** Results of pathogenicity tests and host range assays of the type strains of *Clavibacter*
 802 *michiganensis* subsp. *phaseoli* and *C. michiganensis* subsp. *chilensis* as well as two atypical
 803 peach-colored *Clavibacter* sp. strains on different annual crops including their host of
 804 isolation in greenhouse conditions. None of the evaluated bacterial strains was pathogenic
 805 on the tested plants, while the non-pathogenic strain CFBP 8616 was re-isolated from
 806 common bean cv. Navy tissues.

807 **Table 2:** *Clavibacter* spp. genome sequences used for the comparative genomics and phylogenetic
 808 analyses. The first ten sequences were obtained in this study and announced previously
 809 (20), while the remaining ones were retrieved from the NCBI GenBank database. The
 810 leftmost column represents the original nomenclature of the taxa, while the second
 811 column indicates either revised taxonomy of the strains (17) or their new taxonomic status
 812 as proposed in this study.

813

814 **Table 3:** Average nucleotide identity (ANI; lower diagonal) and digital DNA–DNA hybridization
 815 (dDDH; upper diagonal) values among the type and/or representative strains of different
 816 lineages defined within the genus *Clavibacter* spp. ANI values were calculated using three
 817 different algorithms, *i.e.* JSpeciesWS, ANI calculator, and OrthoANlu, and presented
 818 respectively in a left to right order. A combination of ANI and dDDH indices was used to
 819 designate a taxonomic status to a given phylogenetic clade, where the “new species”
 820 status was assigned to a clade only when both the ANI and dDDH values were below the
 821 accepted threshold ($\leq 95\%$ and $\leq 70\%$ for ANI and dDDH, respectively; 21).

822

823 **Table 4:** Genomic characteristics of *Clavibacter* spp. strains used in this study. Individual genomes
 824 were analyzed using the online annotating service RAST, and protein-encoding sequences
 825 (CDS), functions of the genes, and represented subsystems in the genomes were
 826 determined for each genome using the SEED-Viewer comparative environment. Feature
 827 corresponding to cell wall and capsule, DNA metabolism, dormancy and sporulation,
 828 membrane transport, respiration, RNA metabolism, as well as miscellaneous groups have
 829 not been included in the table since they were common among all the evaluated taxa.

830

831 **Table 5:** Results of one-vs.-one BLASTn/BLASTp searches using the genome sequence of
 832 *Clavibacter michiganensis sensu stricto* NCPPB 382 (AM711867.1) against all other genome
 833 sequences shown in Table 2. Putative pathogenicity determinant genes/regions described
 834 in the literature were subjected to the analyses. While nine chromosomal genes (*i.e.* *chpC*,

835 *chpG*, *clvA* (*micA*), *clvF*, *clvG*, *pelA1*, *pelA2*, *perF*, and *ppaA*) as well as the pCM2 plasmid-
 836 borne pathogenicity-associated gene *pat-1* were determined as exclusively present in *C.*
 837 *michiganensis sensu stricto*, the remaining genes were detected in different phylogenetic
 838 lineages regardless of their pathogenicity and host range.

839
 840 **Table 6:** Prophages within the genome sequences of *Clavibacter* spp. strains detected using the
 841 online service PHASTER. Five prophage groups *i.e.* Gordon_Schwabeltier,
 842 Gordon_Smoothie, N15, P1, and Phi92 were detected in CFBP 8615, CFBP 7491, LMG
 843 26808, CASJ009, LMG 3663, and ATCC 33113 strains. No prophage was detected within the
 844 genome sequences of the strains ATCC 33566T, CF11, CFBP 7493, CFBP 7494, CFBP 8017,
 845 CFBP 8019, CFBP 8216T, CFBP 8217T, CFBP 8616, CFBP 8627T, DOAB 609, NCPPB 2581T,
 846 NCPPB 382, and PF008T.

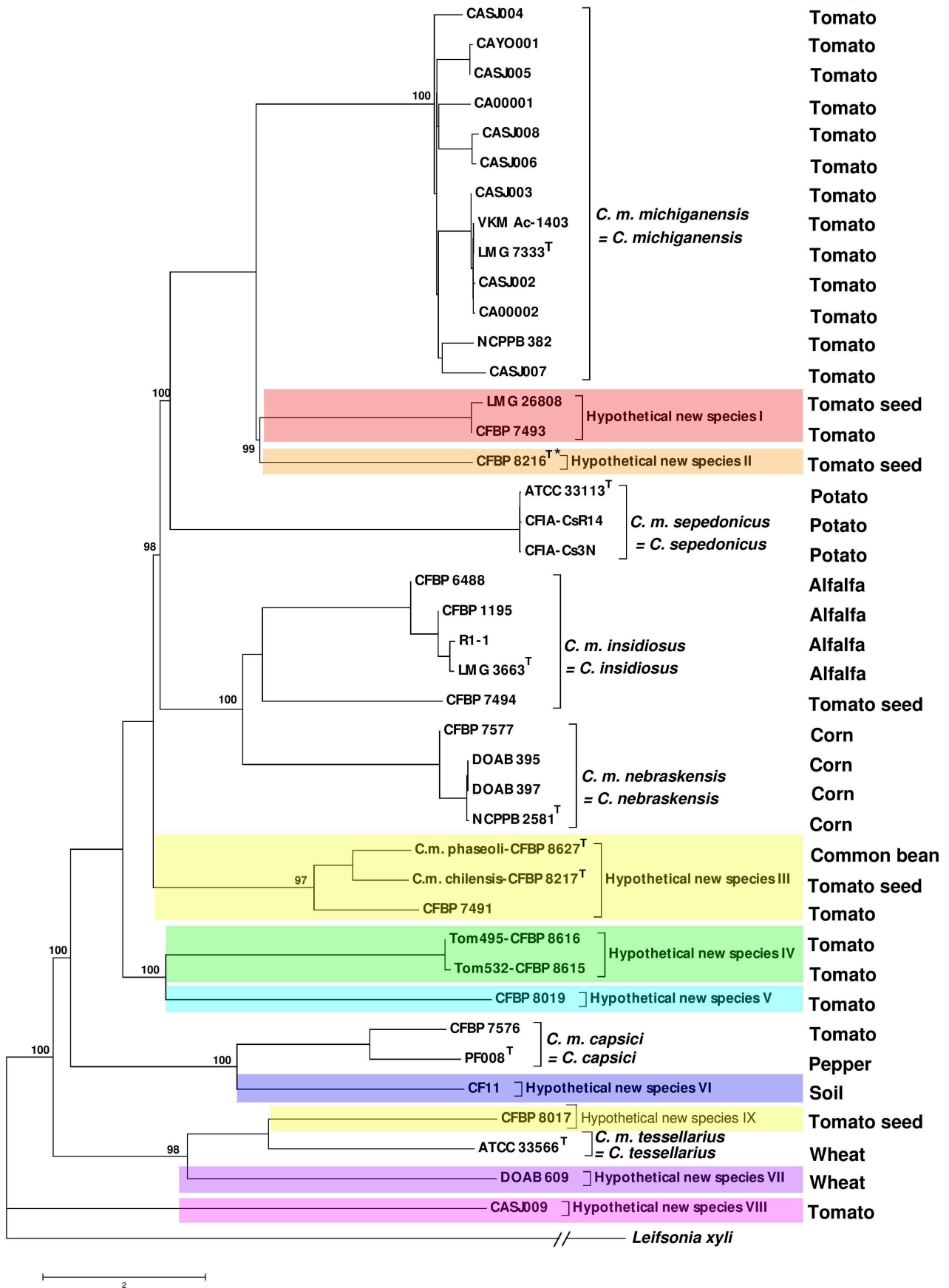
847
 848 **Table 7:** In silico screening for bacteriocins and antibiotic peptides among the *Clavibacter* spp.
 849 genome sequences analyzed in this study. The lantibiotic "Michiganin A" was detected in
 850 all the 12 *C. michiganensis sensu stricto* strains (Figure 1) but was not found in non-
 851 pathogenic tomato-associated strains nor in the pathogenic strains on other plant species.

852
 853 **Figure S1:** Tomato and pepper plants inoculated with *Clavibacter michiganensis sensu stricto*
 854 (ICMP 22049). Tomato plants inoculated with the strain ICMP 22049 showed wilting and
 855 plant death 10-12 days post inoculation (a, right side plant), while pepper plants inoculated

856 using the same strain showed only stem canker symptoms on the site of inoculation with
857 no wilting nor plant death in the same timeframe (b). Control tomato plant remained
858 healthy (a, left side plant).

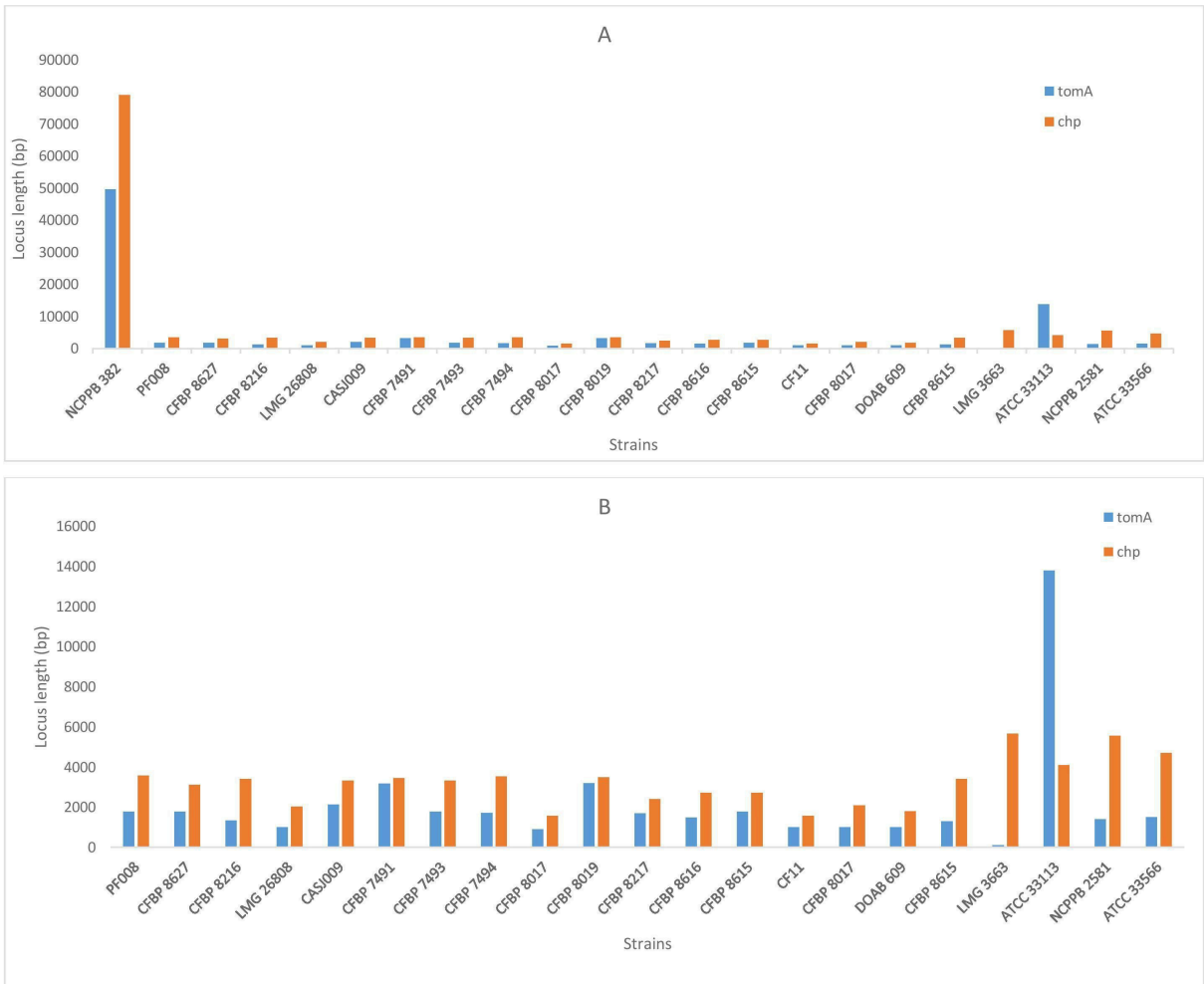
859

860



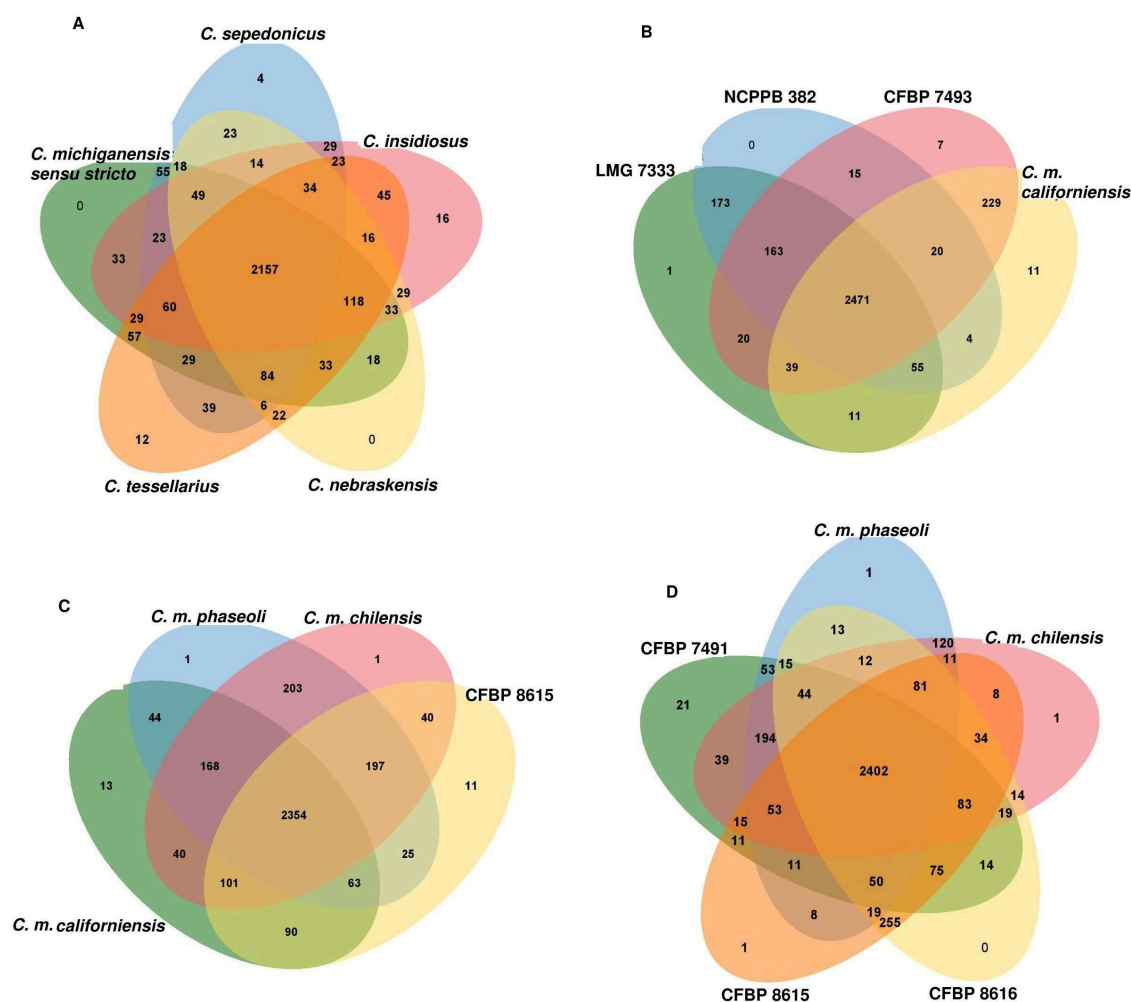
Comment citer ce document :

Osdaghi, E., Rahimi, T., Taghavi, Ansari, Zarei, Portier, P., Briand, M., Jacques, M. A. (2020). Comparative Genomics and Phylogenetic Analyses Suggest Several Novel Species within *Clavibacter* sp. Including Non-Pathogenic Tomato-Associated Strains. *Applied and Environmental Microbiology*. preprint. . DOI : 10.1128/AEM.02873-19



Comment citer ce document :

Osdaghi, E., Rahimi, T., Taghavi, Ansari, Zarei, Portier, P., Briand, M., Jacques, M. A. (2020). Comparative Genomics and Phylogenetic Analyses Suggest Several Novel Species within *Clavibacter* sp. Including Non-Pathogenic Tomato-Associated Strains. *Applied and Environmental Microbiology*. preprint. . DOI : 10.1128/AEM.02873-19



Comment citer ce document :

Osdaghi, E., Rahimi, T., Taghavi, Ansari, Zarei, Portier, P., Briand, M., Jacques, M. A. (2020). Comparative Genomics and Phylogenetic Analyses Suggest Several Novel Species within *Clavibacter* sp. Including Non-Pathogenic Tomato-Associated Strains. *Applied and Environmental Microbiology*. preprint. . DOI : 10.1128/AEM.02873-19

Table 1: Results of pathogenicity tests and host range assays of the type strains of *Clavibacter michiganensis* subsp. *phaseoli* and *C. michiganensis* subsp. *chilensis* as well as two atypical peach-colored *Clavibacter* sp. strains on different annual crops including their host of isolation in greenhouse conditions. None of the evaluated bacterial strains was pathogenic on the tested plants, while the non-pathogenic strain CFBP 8616 was re-isolated from common bean cv. Navy tissues.

Taxon	Strain	Pepper		Common bean				Cowpea	Mung bean	Tomato
		cv. Sereno	cv. Aziz	cv. Red Kidney	cv. Pinto	cv. Navy				
<i>C. m. phaseoli</i>	CFBP 8627 ^T	-	-	-	-	-	-	-	-	-
<i>C. m. chilensis</i>	CFBP 8217 ^T	-	-	-	-	-	-	-	-	-
<i>Clavibacter</i> sp.	CFBP 8615	-	-	-	-	-	-	-	-	-
<i>Clavibacter</i> sp.	CFBP 8616	-	-	-	-	- ^a	-	-	-	-
<i>C. michiganensis sensu stricto</i>	ICMP 22049	+ ^b	+ ^b	-	-	-	-	-	-	+

C. m.: *Clavibacter michiganensis* subsp.

cv.: Cultivar

-: Negative

+: Positive

T: Type strain

a: The inoculated bacterial strain was re-isolated from the asymptomatic leaf tissues of the test plants.

b: These plants showed only stem canker symptoms on the site of inoculation with no systemic wilting and plant death.

Comment citer ce document :

Osdaghi, E., Rahimi, T., Taghavi, Ansari, Zarei, Portier, P., Briand, M., Jacques, M. A. (2020). Comparative Genomics and Phylogenetic Analyses Suggest Several Novel Species within *Clavibacter* sp. Including Non-Pathogenic Tomato-Associated Strains. Applied and Environmental Microbiology. preprint. . DOI : 10.1128/AEM.02873-19

Table 2: *Clavibacter* spp. genome sequences used for the comparative genomics and phylogenetic analyses. The first ten sequences were obtained in this study and announced previously (20), while the remaining ones were retrieved from the NCBI GenBank database. The leftmost column represents the original nomenclature of the taxa, while the second column indicates either revised taxonomy of the strains (17) or their new taxonomic status as proposed in this study.

Taxonomic position		Strain	Host of Isolation	Pathogenic	Date	Country of Isolation	GenBank Accession No.	Reference
Previous name	New name ¹							
<i>Clavibacter</i> sp.	new species ² IV	CFBP 8615*	<i>Solanum lycopersicum</i>	NP	2015	Iran	QWGT01000000	20
<i>Clavibacter</i> sp.	new species IV	CFBP 8616*	<i>Solanum lycopersicum</i>	NP	2015	Iran	QWGU01000000	20
<i>C. m. californiensis</i>	new species II	CFBP 8216 ^T *	<i>Solanum lycopersicum</i>	NP	2000	USA	QWEE01000000	20
<i>C. m. chilensis</i>	new species III	CFBP 8217 ^T *	<i>Solanum lycopersicum</i>	NP	2007	Netherlands	QWGS01000000	20
<i>Clavibacter</i> sp.	new species III	CFBP 7491*	<i>Solanum lycopersicum</i>	NP	ND	ND	QWEB01000000	20
<i>Clavibacter</i> sp.	new species I	CFBP 7493*	<i>Solanum lycopersicum</i>	NP	ND	ND	QWEC01000000	20
<i>C. m. insidiosus</i>	<i>C. insidiosus</i>	CFBP 1195	<i>Medicago sativa</i>	alfalfa	1964	UK	QWDZ01000000	20
<i>C. m. insidiosus</i>	<i>C. insidiosus</i>	CFBP 6488	<i>Medicago sativa</i>	alfalfa	1998	Czech Republic	QWEA01000000	20
<i>C. m. nebraskensis</i>	<i>C. nebraskensis</i>	CFBP 7577	<i>Zea mays</i>	corn	ND	ND	QWED01000000	20
<i>C. m. phaseoli</i>	new species III	CFBP 8627 ^T	<i>Phaseolus vulgaris</i>	NP	2009	Spain	QWGV01000000	20
<i>Clavibacter</i> sp.	new species VIII	CASJ009*	<i>Solanum lycopersicum</i>	NP	2011	USA	MDHJ01000000	22
<i>Clavibacter</i> sp.	new species V	CFBP 8019*	<i>Solanum lycopersicum</i>	NP	2011	USA	MDHJ01000000	22
<i>Clavibacter</i> sp.	new species IX	CFBP 8017*	<i>Solanum lycopersicum</i>	wheat	2006	Netherlands	MDJY00000000	22
<i>Clavibacter</i> sp.	new species VI	CF11	soil	NP	2011	China	JROD00000000	23
<i>Clavibacter</i> sp.	new species I	LMG 26808*	<i>Solanum lycopersicum</i>	NP	ND	Netherlands	AZQZ00000000	28
<i>Clavibacter</i> sp.	<i>C. insidiosus</i>	CFBP 7494*	<i>Solanum lycopersicum</i>	wheat	1999	Chile	MDJW00000000	22
<i>Clavibacter</i> sp.	new species VII	DOAB 609	<i>Triticum aestivum</i>	ND	1976	USA	LQXA00000000	50
<i>C. m. michiganensis</i>	<i>C. michiganensis</i>	LMG 7333 ^T	<i>Solanum lycopersicum</i>	tomato	1957	Hungary	NZ_MZMP00000000	51
<i>C. m. michiganensis</i>	<i>C. michiganensis</i>	NCPBP 382	<i>Solanum lycopersicum</i>	tomato	1956	UK	AM711867.1	26
<i>C. m. michiganensis</i>	<i>C. michiganensis</i>	CASJ004	<i>Solanum lycopersicum</i>	tomato	1999	USA	MDHE00000000	22

Comment citer ce document :

Osdaghi, E., Rahimi, T., Taghavi, Ansari, Zarei, Portier, P., Briand, M., Jacques, M. A. (2020). Comparative Genomics and Phylogenetic Analyses Suggest Several Novel Species within *Clavibacter* sp. Including Non-Pathogenic Tomato-Associated Strains. *Applied and Environmental Microbiology*. preprint. DOI : 10.1128/AEM.02873-19

<i>C. m. michiganensis</i>	<i>C. michiganensis</i>	CAYO001	<i>Solanum lycopersicum</i>	tomato 2001	USA	MDHL00000000	22
<i>C. m. michiganensis</i>	<i>C. michiganensis</i>	CASJ005	<i>Solanum lycopersicum</i>	tomato 2001	USA	MDHF00000000	22
<i>C. m. michiganensis</i>	<i>C. michiganensis</i>	CA00001	<i>Solanum lycopersicum</i>	tomato 2000	USA	MDHK00000000	22
<i>C. m. michiganensis</i>	<i>C. michiganensis</i>	CASJ008	<i>Solanum lycopersicum</i>	tomato 2002	USA	MDHI00000000	22
<i>C. m. michiganensis</i>	<i>C. michiganensis</i>	CASJ006	<i>Solanum lycopersicum</i>	tomato 2002	USA	MDHG00000000	22
<i>C. m. michiganensis</i>	<i>C. michiganensis</i>	CASJ003	<i>Solanum lycopersicum</i>	tomato 1999	USA	MDHD00000000	22
<i>C. m. michiganensis</i>	<i>C. michiganensis</i>	VKM Ac-1403	<i>Solanum lycopersicum</i>	tomato 2017	USA	FVZG00000000	52
<i>C. m. michiganensis</i>	<i>C. michiganensis</i>	CASJ002	<i>Solanum lycopersicum</i>	tomato 1999	USA	MDHC00000000	22
<i>C. m. michiganensis</i>	<i>C. michiganensis</i>	CA00002	<i>Solanum lycopersicum</i>	tomato 2000	USA	MDHM00000000	22
<i>C. m. michiganensis</i>	<i>C. michiganensis</i>	CASJ007	<i>Solanum lycopersicum</i>	tomato 2011	USA	MDHH00000000	22
<i>C. m. capsici</i>	<i>C. capsici</i>	PF008 ^T	<i>Capsicum</i> sp.	pepper ND	Korea	NZ_CP012573	15
<i>C. m. capsici</i>	<i>C. capsici</i>	CFBP 7576*	<i>Solanum lycopersicum</i>	pepper 1997	ND	MDJX00000000	22
<i>C. m. insidiosus</i>	<i>C. insidiosus</i>	LMG 3663 ^T	<i>Medicago sativa</i>	alfalfa 1955	USA	MZMO00000000	51
<i>C. m. insidiosus</i>	<i>C. insidiosus</i>	R1-1	<i>Medicago truncatula</i>	alfalfa 2009	USA	NZ_CP011043	29
<i>C. m. nebraskensis</i>	<i>C. nebraskensis</i>	NCPPB 2581 ^T	<i>Zea mays</i>	corn 1971	USA	NC_020891.1	50
<i>C. m. nebraskensis</i>	<i>C. nebraskensis</i>	DOAB 395	<i>Zea mays</i>	corn 2014	Canada	LSOE00000000	50
<i>C. m. nebraskensis</i>	<i>C. nebraskensis</i>	DOAB 397	<i>Zea mays</i>	corn 2014	Canada	LAKL00000000	53
<i>C. m. sepedonicus</i>	<i>C. sepedonicus</i>	ATCC 33113 ^T	<i>Solanum tuberosum</i>	potato ND	Canada	NC_010407.1	54
<i>C. m. sepedonicus</i>	<i>C. sepedonicus</i>	CFIA-CsR14	<i>Solanum tuberosum</i>	potato ND	Canada	MZMN00000000	51
<i>C. m. sepedonicus</i>	<i>C. sepedonicus</i>	CFIA-Cs3N	<i>Solanum tuberosum</i>	potato ND	Canada	MZMM00000000	51
<i>C. m. tessellarius</i>	<i>C. tessellarius</i>	ATCC 33566 ^T	<i>Triticum aestivum</i>	wheat 1978	USA	MZMQ01000000	51

1: Although new names of six species (*i.e.* *C. capsici*, *C. insidiosus*, *C. michiganensis*, *C. nebraskensis*, *C. sepedonicus* and *C. tessellarius*) were formally described previously (Li et al. 2018), a formal taxonomic description is needed for the hypothetical new species I to IX.

2: Hypothetical new species based on the ANI/dddH values (Table 3) and comparative genomics (Table 4).

*These strains were isolated from tomato plant but were non-pathogenic on the host of isolation.

C. m.: *Clavibacter michiganensis* subsp.

Comment citer ce document :

Osdaghi, E., Rahimi, T., Taghavi, Ansari, Zarei, Portier, P., Briand, M., Jacques, M. A. (2020). Comparative Genomics and Phylogenetic Analyses Suggest Several Novel Species within *Clavibacter* sp. Including Non-Pathogenic Tomato-Associated Strains. Applied and Environmental Microbiology. preprint. . DOI : 10.1128/AEM.02873-19

T: Type strain

ND: Not determined

NP: Non-pathogenic

Comment citer ce document :

Osdaghi, E., Rahimi, T., Taghavi, Ansari, Zarei, Portier, P., Briand, M., Jacques, M. A. (2020). Comparative Genomics and Phylogenetic Analyses Suggest Several Novel Species within *Clavibacter* sp. Including Non-Pathogenic Tomato-Associated Strains. *Applied and Environmental Microbiology*. preprint. . DOI : 10.1128/AEM.02873-19

Table 3: Average nucleotide identity (ANI; lower diagonal) and digital DNA–DNA hybridization (dDDH; upper diagonal) values among the type and/or representative strains of different lineages defined within the genus *Clavibacter* spp. ANI values were calculated using three different algorithms, *i.e.* JSpeciesWS, ANI calculator, and OrthoANIu, and presented respectively in a left to right order. A combination of ANI and dDDH indices was used to designate a taxonomic status to a given phylogenetic clade, where the “new species” status was assigned to a clade only when both the ANI and dDDH values were below the accepted threshold ($\leq 95\%$ and $\leq 70\%$ for ANI and dDDH, respectively; 21).

Taxon	Strain	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1 <i>C. michiganensis sensu stricto</i>	LMG 7333 ^T		57.70	58.70	48.00	48.60	49.70	48.00	49.80	49.50	44.80	43.10	40.20	40.00	37.50	37.20	36.80	34.60
2 <i>C. m. californiensis</i>	CFBP 8216 ^T	95/95/9		57.80	46.40	47.90	48.60	47.40	48.10	47.90	44.80	43.70	40.50	40.40	38.50	38.40	38.00	35.70
3 <i>Clavibacter</i> sp.	LMG 26808	95/94/9	94/95/9		45.80	47.80	48.40	47.10	47.90	47.70	44.30	42.60	39.60	39.50	37.10	37.00	36.40	34.50
4 <i>C. sepedanicus</i>	ATCC 33113 ^T	92/92/9	92/92/9	92/92/9		45.10	46.20	45.20	47.00	46.60	43.90	42.30	39.10	39.10	36.40	36.20	36.00	34.00
5 <i>C. insidiosus</i>	LMG 3663 ^T	93/93/9	93/93/9	92/92/9	92/92/9		64.60	59.90	51.20	51.00	47.10	43.30	40.50	40.20	37.90	37.40	37.20	34.80
6 <i>Clavibacter</i> sp.	CFBP 7494	93/93/9	92/93/9	93/92/9	92/92/9	96/96/9		60.20	52.20	52.10	46.00	44.10	40.70	40.60	38.10	37.60	37.50	34.90
7 <i>C. nebraskensis</i>	NCPBP 2581 ^T	92/93/9	92/93/9	92/92/9	92/92/9	95/95/9	95/95/9		51.40	51.10	45.30	43.30	40.70	40.70	37.70	37.20	36.90	34.90
8 <i>C. m. chilensis</i>	CFBP 8217 ^T	93/93/9	93/93/9	92/93/9	92/93/9	93/93/9	93/93/9	93/93/9		87.50	47.00	45.80	43.40	42.90	40.20	39.50	39.20	37.10
9 <i>C. m. phaseoli</i>	CFBP 8627 ^T	93/93/9	93/93/9	92/92/9	92/93/9	93/93/9	93/93/9	93/93/9	99/99/9		47.10	45.70	43.30	42.90	40.00	39.50	39.10	37.00
10 <i>Clavibacter</i> sp.	CFBP 8615	92/92/9	92/92/9	91/92/9	91/92/9	92/92/9	92/92/9	92/92/9	92/92/9	92/92/9		48.00	40.90	40.50	38.20	37.60	37.70	35.80
11 <i>Clavibacter</i> sp.	CFBP 8019	91/92/9	91/92/9	91/91/9	91/92/9	91/92/9	91/91/9	91/91/9	91/91/9	92/92/9	92/92/9	93/93/9		40.10	40.00	36.90	36.80	34.40
12 <i>C. capsici</i>	PF008 ^T	90/91/9	90/91/9	90/90/9	90/91/9	90/91/9	90/90/9	90/91/9	91/92/9	91/92/9	90/91/9	90/91/9	90/91/9		58.50	38.40	38.50	34.50

Comment citer ce document :

Osdaghi, E., Rahimi, T., Taghavi, Ansari, Zarei, Portier, P., Briand, M., Jacques, M. A. (2020). Comparative Genomics and Phylogenetic Analyses Suggest Several Novel Species within *Clavibacter* sp. Including Non-Pathogenic Tomato-Associated Strains. *Applied and Environmental Microbiology*. preprint. . DOI : 10.1128/AEM.02873-19

		0	1	0	0	0	0	0	1	1	1	0							
13	<i>Clavibacter</i> sp.	CF11	90/90/9	90/90/9	90/90/9	90/90/9	90/90/9	91/90/9	91/90/9	91/91/9	91/91/9	90/90/9	90/90/9	95/95/9	38.10	38.50	37.60	34.40	
			0	0	0	0	0	0	0	1	1	0	0	5					
14	<i>C. tessellarius</i>	ATCC 33566 ^T	90/90/8	90/91/9	89/89/8	89/90/8	90/90/8	90/90/9	90/90/8	91/91/9	91/91/9	90/90/9	89/90/8	90/91/9	49.00	57.70	33.50		
			9	0	9	9	9	0	9	0	0	0	9	0					
														90/90/9					
15	<i>Clavibacter</i> sp.	DOAB 609	89/89/8	89/90/9	89/89/8	89/89/8	89/89/8	90/90/8	89/89/8	90/90/9	90/90/9	89/90/8	89/89/8	90/90/9	90/90/9	93/93/9	47.20	33.60	
			9	0	9	9	9	9	9	0	0	9	9	0	0	3			
16	<i>Clavibacter</i> sp.	CFBP 8017	89/89/8	89/90/9	89/89/8	89/89/8	89/89/8	90/90/8	89/89/8	90/90/9	90/90/9	89/90/8	89/89/8	90/90/8	90/90/8	95/95/9	92/92/9	33.30	
			9	0	9	9	9	9	9	0	0	9	9	9	5	2			
17	<i>Clavibacter</i> sp.	CASJ009	88/89/8	89/90/8	88/88/8	88/89/8	88/89/8	88/88/8	88/88/8	89/89/8	89/90/8	89/90/9	89/90/8	88/89/8	88/89/8	88/88/8	88/89/8	88/88/8	87/88/8
			9	8	8	8	8	8	8	9	0	9	8	9	8	8	7	7	

C. m.: *Clavibacter michiganensis* subsp.

T: Type strain

Comment citer ce document :

Osdaghi, E., Rahimi, T., Taghavi, Ansari, Zarei, Portier, P., Briand, M., Jacques, M. A. (2020). Comparative Genomics and Phylogenetic Analyses Suggest Several Novel Species within *Clavibacter* sp. Including Non-Pathogenic Tomato-Associated Strains. *Applied and Environmental Microbiology*. preprint. . DOI : 10.1128/AEM.02873-19

Table 4: Genomic characteristics of *Clavibacter* spp. strains used in this study. Individual genomes were analyzed using the online annotating service RAST, and protein-encoding sequences (CDS), functions of the genes, and represented subsystems in the genomes were determined for each genome using the SEED-Viewer comparative environment. Feature corresponding to cell wall and capsule, DNA metabolism, dormancy and sporulation, membrane transport, respiration, RNA metabolism, as well as miscellaneous groups have not been included in the table since they were common among all the evaluated taxa.

Subsystem feature	<i>C. michiganensis sensu stricto</i> (NCPBP 382)	<i>C. capsici</i> (PF008 [†])	<i>Clavibacter</i> sp. (CFBP 8615)	<i>Clavibacter</i> sp. (CFBP 8616)	<i>C. m. californiensis</i> (CFBP 8216 [†])	<i>Clavibacter</i> sp. (CFBP 7491)	<i>Clavibacter</i> sp. (CFBP 7493)	<i>Clavibacter</i> sp. (CFBP 7494)	<i>Clavibacter</i> sp. (CFBP 8019)	<i>Clavibacter</i> sp. (CFBP 8017)	<i>Clavibacter</i> sp. (DOAB 609)	<i>Clavibacter</i> sp. (CF11)	<i>Clavibacter</i> sp. (LMG 26808)	<i>C. m. phaseoli</i> (CFBP 8627 [†])	<i>C. m. chilensis</i> (CFBP 8217 [†])	<i>Clavibacter</i> sp. (CASI009)	<i>C. insidiosus</i> (LMG 3663 [†])	<i>C. nebraskensis</i> (NCPBP 2581 [†])	<i>C. sepedonicus</i> (ATCC 33113 [†])	<i>C. tessellarius</i> (ATCC 33566 [†])
Genome size (bp)	3,297	3,056	3,129	3,094	3,193	3,288	3,275	3,313	3,024	3,172	3,296	3,118	3,420	3,052	3,044	3,268	3,387	3,063	3,258	3,318
GC content (%)	72.7	73.6	73.2	73.2	72.7	73.0	72.9	73.3	73.5	73.5	73.2	73.6	72.0	73.5	73.5	73.6	72.7	73.0	72.6	73.7
Number of coding sequences (CDS)	2979	2725	2807	2730	2784	2917	2897	2956	2676	3014	3181	3002	3097	2642	2629	3054	3091	2739	3047	2956
Number of subsystems	345	326	323	319	338	316	342	330	326	263	266	260	340	315	311	341	332	325	345	317
RNAs	51	51	47	47	48	50	48	50	48	48	53	49	57	50	49	51	52	51	51	52
Cofactors, vitamins, pigments	177	158	169	157	165	139	173	161	158	107	116	99	175	167	130	181	165	139	169	176
Virulence, disease and defense	34	38	37	36	38	36	36	32	38	26	25	25	40	35	34	42	36	30	33	22
Resistance to antibiotics/toxic compounds	19	23	22	21	23	21	21	17	24	17	16	15	25	21	19	27	21	17	20	18
Invasion and intracellular resistance	15	15	15	15	15	15	15	15	14	9	9	10	15	14	15	15	15	13	13	4
Potassium metabolism	13	12	12	12	13	12	13	13	16	5	5	5	9	9	13	8	13	9	10	11
Phages, prophages, transposable elements	0	0	0	0	2	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Iron acquisition and metabolism	16	17	13	25	16	8	19	18	13	5	10	5	17	10	10	12	15	14	17	16
Siderophores*:																				

Comment citer ce document :

Osdaghi, E., Rahimi, T., Taghavi, Ansari, Zarei, Portier, P., Briand, M., Jacques, M. A. (2020). Comparative Genomics and Phylogenetic Analyses Suggest Several Novel Species within *Clavibacter* sp. Including Non-Pathogenic Tomato-Associated Strains. Applied and Environmental Microbiology. preprint. . DOI : 10.1128/AEM.02873-19

Siderophore Yersiniabactin	2	0	0	0	2	0	2	0	0	0	5	0	2	0	0	0	2	2	0	0
Siderophore Aerobactin	3	3	3	3	3	3	4	3	3	0	0	0	3	3	3	3	3	3	3	3
Nucleosides and nucleotides	87	72	93	92	89	76	100	76	81	80	79	75	74	98	68	90	79	80	99	65
Protein metabolism	211	201	209	192	189	185	209	204	208	166	162	167	205	178	198	192	205	178	213	185
Cell division and cell cycle	26	21	8	8	23	24	24	23	21	0	0	0	8	23	8	22	24	20	22	23
Motility and chemotaxis	1	2	2	2	2	3	1	2	2	0	0	0	3	3	3	3	2	1	2	2
Regulation and cell signaling	33	30	19	20	22	22	25	30	33	14	17	18	36	28	17	29	31	33	35	24
Secondary metabolism	6	18	5	5	4	5	11	12	11	9	9	9	12	0	1	11	5	5	13	11
Fatty acids, lipids, and isoprenoids	88	78	84	103	82	108	78	88	59	44	44	42	79	85	76	82	65	65	80	101
Nitrogen metabolism	7	10	6	6	10	6	7	6	6	6	6	6	7	6	6	11	7	7	6	11
Stress response	67	73	76	69	74	67	64	70	78	27	27	28	68	71	61	82	74	71	78	64
Metabolism of aromatic compounds	10	8	9	8	9	14	10	10	8	4	3	3	10	10	10	18	10	9	10	10
Amino acids and derivatives	224	190	223	212	233	221	210	192	199	217	221	203	236	193	195	224	219	229	232	220
Sulfur metabolism	20	16	16	11	9	12	10	16	17	6	5	5	13	10	11	26	17	9	22	18
Phosphorus metabolism	38	26	25	27	28	26	27	23	25	28	27	27	36	25	26	25	23	23	25	24
Carbohydrates	285	266	255	268	267	283	247	268	237	181	194	175	278	263	261	382	266	266	269	281

C. m.: *Clavibacter michiganensis* subsp.

T: Type strain

*Siderophore assembly kit was detected only in *Clavibacter* sp. CFBP 8616.

Comment citer ce document :

Osdaghi, E., Rahimi, T., Taghavi, Ansari, Zarei, Portier, P., Briand, M., Jacques, M. A. (2020). Comparative Genomics and Phylogenetic Analyses Suggest Several Novel Species within *Clavibacter* sp. Including Non-Pathogenic Tomato-Associated Strains. *Applied and Environmental Microbiology*. preprint. . DOI : 10.1128/AEM.02873-19

Table 5: Results of one-vs.-one BLASTn/BLASTp searches using the genome sequence of *Clavibacter michiganensis sensu stricto* NCPPB 382 (AM711867.1) against all other genome sequences shown in Table 2. Putative pathogenicity determinant genes/regions described in the literature were subjected to the analyses. While nine chromosomal genes (*i.e.* *chpC*, *chpG*, *clvA* (*micA*), *clvF*, *clvG*, *pelA1*, *pelA2*, *perF*, and *ppaA*) as well as the pCM2 plasmid-borne pathogenicity-associated gene *pat-1* were determined as exclusively present in *C. michiganensis sensu stricto*, the remaining genes were detected in different phylogenetic lineages regardless of their pathogenicity and host range.

Locus tag on NCPPB 382 genome	Gene/cluster	Length (bp)	<i>C. michiganensis sensu stricto</i> (NCPPB 382)	<i>Clavibacter</i> sp. (CFBP 8615)	<i>Clavibacter</i> sp. (CF11)	<i>Clavibacter</i> sp. (CFBP 8017)	<i>Clavibacter</i> sp. (DOAB 609)	<i>C. m. californiensis</i> (CFBP 8216 ¹)	<i>C. m. chiliensis</i> (CFBP 8217 ¹)	<i>C. capsici</i> (PF008 ¹)	<i>Clavibacter</i> sp. (CFBP 7491)	<i>Clavibacter</i> sp. (CFBP 7493)	<i>Clavibacter</i> sp. (LMG 26808)	<i>Clavibacter</i> sp. (CFBP 7494)	<i>Clavibacter</i> sp. (CFBP 8019)	<i>Clavibacter</i> sp. (CAS1009)	<i>C. m. phaseoli</i> (CFBP 8627 ¹)	<i>C. insidiosus</i> (LMG 3663 ¹)	<i>C. nebraskensis</i> (NCPPB 2581 ¹)	<i>C. sepedonicus</i> (ATCC 33113 ¹)	<i>C. tessellarius</i> (ATCC 33566 ¹)	Reference
CMM_0070	<i>sbtA</i>	3102	[100]100	[75]75	[98]82	ND	[97]81	[82]82	[76]94	[97]82	[97]89	ND	ND	ND	[98]82	[77]82	[96]89	ND	[99]80	ND	ND	43
CMM_0013	<i>srtA</i>	801	[100]100	[100]92	[100]87	[100]87	[100]87	[100]97	[100]91	[100]87	[100]91	[100]97	[100]97	[100]90	[100]89	[100]89	[100]91	[100]92	[100]92	[100]93	[100]88	43
CMM_1480	<i>expA</i>	762	[100]100	ND	ND	[77]81	[77]82	ND	ND	ND	ND	[100]99	[100]99	ND	ND	ND	ND	ND	ND	ND	ND*	26
CMM_1673	<i>xysA</i>	1296	[100]100	[98]90	[99]87	[99]92	[97]91	[99]95	[81]76	[100]88	[83]76	[99]94	[99]94	[99]93	[99]89	[48]80	[81]75	[99]95	[99]92	ND	[99]92	43
CMM_1674	<i>xysB</i>	2010	[100]100	[92]93	[100]89	[100]89	[54]75	[100]93	[100]93	[100]90	[100]93	[94]95	[99]94	[100]92	[100]92	[95]90	[93]93	[100]90	[100]91	ND	[99]89	43
CMM_2443	<i>celB</i>	1608	[100]100	ND	ND	[100]89	[100]87	ND	ND	[100]88	ND	[100]91	[100]91	[100]91	ND	ND	ND	[100]92	[64]85	[100]91	[100]85	26

Comment citer ce document :

Osdaghi, E., Rahimi, T., Taghavi, Ansari, Zarei, Portier, P., Briand, M., Jacques, M. A. (2020). Comparative Genomics and Phylogenetic Analyses Suggest Several Novel Species within *Clavibacter* sp. Including Non-Pathogenic Tomato-Associated Strains. Applied and Environmental Microbiology. preprint. . DOI : 10.1128/AEM.02873-19

CMM_2691	NA	1002	[100]	[100]	[99]	[89]	[100]	[89]	[99]	[90]	[100]	[91]	[100]	[89]	[99]	[89]	[100]	[89]	[100]	[94]	[100]	[94]	[89]	[92]	[99]	[90]	[97]	[83]	[100]	[88]	[100]	[94]	[98]	[92]	[99]	[89]	[100]	[90]	26		
CMM_2692	NA	1410	[100]	[100]	[90]	[93]	[100]	[86]	[100]	[86]	[100]	[86]	[91]	96	[77]	90	[100]	[89]	[30]	[86]	[100]	[91]	[100]	[91]	[100]	[90]	[100]	[87]	ND	[90]	90	[100]	[92]	[100]	[91]	[100]	[86]	[100]	[88]	26	
CMM_2871	NA	1491	[100]	[100]	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	[100]	[90]	ND	ND	ND	ND	ND	ND	ND	ND	ND	[92]	[88]	[98]	[92]	ND	[100]	[92]	26			
CMM_2645	<i>vatr1</i>	624	[100]	[100]	[100]	[94]	[100]	[92]	[100]	[92]	[100]	[91]	[85]	98	[100]	[93]	[100]	[92]	[100]	[94]	[100]	[98]	[100]	[98]	[100]	[96]	[100]	[93]	[100]	[89]	[100]	[94]	[100]	[96]	[100]	[96]	[100]	[94]	[100]	[92]	25
CMM_2969	<i>vatr2</i>	1323	[100]	[100]	[100]	[92]	[100]	[92]	[99]	[92]	[100]	[92]	[100]	[96]	[100]	[94]	[100]	[92]	[100]	[94]	[100]	[95]	[100]	[95]	[100]	[94]	[100]	[93]	[100]	[90]	[100]	[94]	[100]	[93]	[100]	[93]	[100]	[93]	[100]	[91]	25
pCM1_0020	<i>celA</i>	2241	ND	ND	ND	[46]	[84]	[46]	[75]	[47]	75	ND	ND	[45]	77	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	[97]	90	[76]	[84]	ND	[45]	75	26			

1: The values in brackets refer to the query coverage (%) of the locus while the values at the right side of the brackets refer to the sequence similarity (%) between the reference strain (NCCPB 382^T) and the strain in question. The query coverage of <50% were not considered as reliable data.

NA: Not Assigned

ND: Not Detected (Absence of the target gene/cluster)

C. m.: *Clavibacter michiganensis* subsp.

T: Type strain

*: A putative expansin protein was found in this species (GenBank: OQJ63896.1) in which the nucleotide sequence was different from CMM_1480: *expA*.

Comment citer ce document :

Osdaghi, E., Rahimi, T., Taghavi, Ansari, Zarei, Portier, P., Briand, M., Jacques, M. A. (2020). Comparative Genomics and Phylogenetic Analyses Suggest Several Novel Species within *Clavibacter* sp. Including Non-Pathogenic Tomato-Associated Strains. Applied and Environmental Microbiology. preprint. . DOI : 10.1128/AEM.02873-19

Table 6: Prophages within the genome sequences of *Clavibacter* spp. strains detected using the online service PHASTER. Five prophage groups *i.e.* Gordon_Schwabeltier, Gordon_Smoothie, N15, P1, and Phi92 were detected in CFBP 8615, CFBP 7491, LMG 26808, CASJ009, LMG 3663, and ATCC 33113 strains. No prophage was detected within the genome sequences of the strains ATCC 33566T, CF11, CFBP 7493, CFBP 7494, CFBP 8017, CFBP 8019, CFBP 8216T, CFBP 8217T, CFBP 8616, CFBP 8627T, DOAB 609, NCPPB 2581T, NCPPB 382, and PF008T.

Taxon	Strain	Region Length (kb)	Completeness	Number of proteins	Position	phage (GenBank accession number)	GC%
<i>Clavibacter</i> sp.	CFBP 8615	6.9	incomplete	8	38: 3643-10586	Phi92 (NC_023693)	71.28
<i>Clavibacter</i> sp.	CFBP 7491	6.8	incomplete	14	4: 13543-20391	Gordon_Schwabeltier (NC_031255)	66.29
<i>Clavibacter</i> sp.	LMG 26808	9.9	incomplete	10	13: 128032-137979	Phi92 (NC_023693)	71.79
		7	incomplete	11	15: 712-7762	N15 (NC_001901)	46.49
		23.7	incomplete	10	15: 5546-29310	P1 (NC_005856)	48.19
		7.7	incomplete	9	1009722-1017492	Phi92 (NC_023693)	71.29
<i>Clavibacter</i> sp.	CASJ009	7.7	incomplete	9	1009722-1017492	Phi92 (NC_023693)	71.29
<i>C. insidiosus</i>	LMG 3663 ^T	6.7	incomplete	16	scaffold2: 5055-11844	Gordon_Smoothie (NC_030696)	66.95
<i>C. sepedonicus</i>	ATCC 33113 ^T	7.1	incomplete	10	42745-49937	Phi92 (NC_023693)	70.71
—	—	7.7	incomplete	8	660037-667813	Phi92 (NC_023693)	71.20

C. m.: *Clavibacter michiganensis* subsp.

ND: Not detected

T: Type strain

Comment citer ce document :

Osdaghi, E., Rahimi, T., Taghavi, Ansari, Zarei, Portier, P., Briand, M., Jacques, M. A. (2020). Comparative Genomics and Phylogenetic Analyses Suggest Several Novel Species within *Clavibacter* sp. Including Non-Pathogenic Tomato-Associated Strains. *Applied and Environmental Microbiology*. preprint. . DOI : 10.1128/AEM.02873-19

Table 7: In silico screening for bacteriocins and antibiotic peptides among the *Clavibacter* spp. genome sequences analyzed in this study. The lantibiotic “Michiganin A” was detected in all the 12 *C. michiganensis sensu stricto* strains (Figure 1) but was not found in non-pathogenic tomato-associated strains nor in the pathogenic strains on other plant species.

Taxon	Strain	#Contig	Start codon	End codon	Class
<i>C. michiganensis sensu stricto</i>	NCPBP 382	1	2211104	2233501	63.1; Michiganin A (Lantibiotic)
<i>C. capsici</i>	PF008 ^T	1	2406476	2426476	Sactipeptides
<i>Clavibacter</i> sp.	CFBP 8615	14	7079	27079	LAPs
		91	1570	18430	Sactipeptides
<i>Clavibacter</i> sp.	CFBP 8616	166	4888	15112	Sactipeptides
<i>C. m. californiensis</i>	CFBP 8216 ^T	201	9178	10822	Sactipeptides
<i>Clavibacter</i> sp.	CFBP 7491	224	8131	11869	Sactipeptides
<i>Clavibacter</i> sp.	CFBP 7493	1	25925	45925	Thiopeptide; LAPs
		226	-8455	11545	Sactipeptides
<i>Clavibacter</i> sp.	CFBP 7494	9	267758	287758	LAPs
		15	89075	109075	Sactipeptides
<i>Clavibacter</i> sp.	CFBP 8019	11	186443	206443	Sactipeptides
		17	190604	210604	LAPs
<i>Clavibacter</i> sp.	LMG 26808	6	43154	63154	Thiopeptide; LAPs
		10	130865	150865	Sactipeptides
<i>Clavibacter</i> sp.	CF11	17	47765	67765	Sactipeptides
		2	247508	267508	LAPs
		13	138056	158368	150.1; Enterocin_AS_48
<i>Clavibacter</i> sp.	DOAB 609	20	4781	24781	Sactipeptides
		30	31400	51400	LAPs
<i>Clavibacter</i> sp.	CFBP 8017	33	39236	59236	LAPs
		46	11546	31546	Sactipeptides
<i>C. m. phaseoli</i>	CFBP 8627 ^T	212	-9916	10084	Sactipeptides
<i>C. m. chilensis</i>	CFBP 8217 ^T	ND	ND	ND	ND
<i>Clavibacter</i> sp.	CASJ009	ND	ND	ND	ND
<i>C. insidiosus</i>	LMG 3663 ^T	1	2292719	2312719	Sactipeptides

Comment citer ce document :

Osdaghi, E., Rahimi, T., Taghavi, Ansari, Zarei, Portier, P., Briand, M., Jacques, M. A. (2020). Comparative Genomics and Phylogenetic Analyses Suggest Several Novel Species within *Clavibacter* sp. Including Non-Pathogenic Tomato-Associated Strains. *Applied and Environmental Microbiology*. preprint. . DOI : 10.1128/AEM.02873-19

<i>C. nebraskensis</i>	NCPBPB 2581 ^T	1	2422475	2442475	Sactipeptides
<i>C. sepedonicus</i>	ATCC 33113 ^T	1	2663669	2683669	Sactipeptides
			3119744	3139861	294.1; Plantathiazolicin (Plantazolicin)
<i>C. tessellarius</i>	ATCC 33566 ^T	1	1820699	1840699	Sactipeptides

C. m.: *Clavibacter michiganensis* subsp.

ND: Not detected

T: Type strain

Comment citer ce document :

Osdaghi, E., Rahimi, T., Taghavi, Ansari, Zarei, Portier, P., Briand, M., Jacques, M. A. (2020). Comparative Genomics and Phylogenetic Analyses Suggest Several Novel Species within *Clavibacter* sp. Including Non-Pathogenic Tomato-Associated Strains. *Applied and Environmental Microbiology*. preprint. . DOI : 10.1128/AEM.02873-19