Genome of the Actinomycete Plant Pathogen *Clavibacter michiganensis* subsp. *sepedonicus* Suggests Recent Niche Adaptation[†]

Stephen D. Bentley,¹ Craig Corton,¹ Susan E. Brown,² Andrew Barron,¹ Louise Clark,¹ Jon Doggett,¹ Barbara Harris,¹ Doug Ormond,¹ Michael A. Quail,¹ Georgiana May,³ David Francis,⁴ Dennis Knudson,² Julian Parkhill,¹ and Carol A. Ishimaru^{2,5*}

Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton CB10 1SA, United Kingdom¹; Department of Biological Sciences and Pest Management, Colorado State University, Fort Collins, Colorado 80523²; Department of Ecology, Evolution and Behavior, University of Minnesota, St. Paul, Minnesota 55108³; Department of Horticulture and Crop Science Ohio Agricultural Research and Development Center, The Ohio State University, Wooster, Ohio 44691⁴; and Department of Plant Pathology, University of Minnesota, St. Paul, Minnesota 55108⁵

Received 2 October 2007/Accepted 1 January 2008

Clavibacter michiganensis subsp. *sepedonicus* is a plant-pathogenic bacterium and the causative agent of bacterial ring rot, a devastating agricultural disease under strict quarantine control and zero tolerance in the seed potato industry. This organism appears to be largely restricted to an endophytic lifestyle, proliferating within plant tissues and unable to persist in the absence of plant material. Analysis of the genome sequence of *C. michiganensis* subsp. *sepedonicus* and comparison with the genome sequences of related plant pathogens revealed a dramatic recent evolutionary history. The genome contains 106 insertion sequence elements, which appear to have been active in extensive rearrangement of the chromosome compared to that of *Clavibacter michiganensis* subsp. *michiganensis*. There are 110 pseudogenes with overrepresentation in functions associated with carbohydrate metabolism, transcriptional regulation, and pathogenicity. Genome comparisons also indicated that there is substantial gene content diversity within the species, probably due to differential gene acquisition and loss. These genomic features and evolutionary dating suggest that there was recent adaptation for life in a restricted niche where nutrient diversity and perhaps competition are low, correlated with a reduced ability to exploit previously occupied complex niches outside the plant. Tolerant of factors such as multiplication and integration of insertion sequence elements, genome rearrangements, and functional disruption of many genes and operons seems to indicate that there has been general relaxation of selective pressure on a large proportion of the genome.

High-G+C-content gram-positive coryneform bacteria cause economic losses in several crops worldwide, yet the relatively slow in vitro and in planta growth and general genetic intractability of these bacteria have long been deterrents to successful identification of the specific molecular mechanisms by which they cause diseases in plants. Consequently, there is a clear disparity between the amount of scientific research on plant-pathogenic coryneform bacteria and the amount of scientific research on their gram-negative counterparts. Recent advances in this field have coincided with the availability of transformation systems and complete genome sequences for representatives of the genera *Clavibacter* and *Leifsonia*, two of the major coryneform plant-pathogenic genera (14, 32, 48, 56, 74). Importantly, these advances allowed breakthrough identification of a novel set of pathogenicity-related genes in the tomato pathogen *Clavibacter michiganensis* subsp. *michiganensis* and identification of homologues in other coryneform plant pathogens (15, 27, 32; see the accompanying paper). Because methodologies created for functional analysis of *C.*

michiganensis subsp. *michiganensis* are generally applicable to other members of the genus, additional advances in this field can be expected in the near future (40, 46).

The genus *Clavibacter* provides an excellent resource for obtaining a more comprehensive understanding of plant-microbe interactions. *Clavibacter* is a member of the family *Microbacteriaceae* in the *Actinomycetales* (60). The related plant pathogens include *Leifsonia*, *Curtobacterium*, and *Rathayibacter* and the more distantly related genera *Rhodococcus* and *Streptomyces*, (29, 47, 64, 78, 84). *Clavibacter* is generally considered a genus of plant pathogens, but recent ecological surveys suggested that environmental, nonpathogenic isolates occur more commonly than was previously thought (21, 25, 36, 85). Members of *C. michiganensis* can usually be further classified to the subspecies level. A cornerstone of subspecies identification in *C. michiganensis* is the striking host specificity of its plant-pathogenic members. Polyphasic schemes also support the current subspecies classification, but it is noteworthy that the genetic basis for subspecies identification remains unknown (3, 16, 21, 50).

To provide genetic resources that could lead to a better understanding of pathogenicity and host specificity for the genus *Clavibacter* specifically and for coryneform plant pathogens in general, our studies were focused on obtaining the complete genome sequence of *Clavibacter michiganensis* subsp. *sepedonicus* (Spieckermann and Kotthoff 1914) Davis et al. 1984 comb. nov. (21). This international and national quarantine pest

* Corresponding author. Mailing address: 495 Borlaug Hall, 1991 Upper Buford Circle, University of Minnesota, St. Paul, MN 55108. Phone: (612) 625-9736. Fax: (612) 625-9728. E-mail: cishimar@umn.edu.

† Supplemental material for this article may be found at <http://jb.asm.org/>.

‡ Published ahead of print on 11 January 2008.

causes bacterial ring rot, a devastating disease of potato. *C. michiganensis* subsp. *sepedonicus* spreads easily within potato farms during seed cutting and can be readily disseminated in latently infected tubers, by infested farm equipment, in storage facilities, and in packing materials. Infections can result in crop losses in the fresh and processed potato industries, but the main economic losses occur in the seed industry, where there is strict zero tolerance for the disease (23). Bacterial ring rot is usually associated with the temperate climates of North America, Asia, Scandinavia, and Northern Europe (73). Crop losses are due to colonization of the tuber vascular system and surrounding tissues, which can lead to extensive secondary breakdown in storage (43, 70).

A specific objective of this work was to conduct whole-genome comparisons between *C. michiganensis* subsp. *sepedonicus* and its plant-pathogenic relatives *C. michiganensis* subsp. *michiganensis* and *Leifsonia* as a means of identifying putative pathogenicity-related genes in the ring rot pathogen. Because *C. michiganensis* subsp. *sepedonicus* thrives almost exclusively as a plant endophyte, while *C. michiganensis* subsp. *michiganensis* is both an endophyte and an epiphyte, the comparative genome studies should also provide insight into aspects of niche adaptation (17, 18, 51, 77). Whole-genome comparisons also were used to identify genomic events associated with the evolution of host specificity and therefore subspeciation within *C. michiganensis*.

MATERIALS AND METHODS

The type strain of *C. michiganensis* subsp. *sepedonicus*, strain ATCC 33113 (= NCPPB 2137 = PDDCC 2535 = LMG 2889), was chosen for sequencing because it is virulent and representative of the subspecies. This strain was originally isolated from infected potato. Many *C. michiganensis* subsp. *sepedonicus* strains have a genome structure similar to that of the type strain, which contains a circular chromosome, one linear plasmid, and one circular plasmid (12, 35, 53). Strain ATCC 33113 also has the largest genome, as estimated by contour-clamped homogeneous electric field analysis, and therefore likely contains the majority of the genetic content of the taxon (13). Purified total genomic DNA (approximately 100 µg) of ATCC 33113 was prepared in agarose blocks as previously described (13). To improve the representation of chromosomal sequences in genomic libraries for sequencing, linear plasmid DNA was separated from high-molecular-weight DNA by gel electrophoresis of agarose plugs as previously described (13).

Genome sequencing. An approximately 8× shotgun sequence was produced from a total of 49,536 end sequences from pUC clones with 2.0- to 2.8-kb inserts using a Big Dye terminator cycle sequencing kit from Applied Biosystems. Reactions were performed with Applied Biosystems 3700 sequencers. Approximately 0.1× sequence coverage (4.7× clone coverage) was obtained from 768 end sequences from 40-kb inserts cloned into fosmid pFOS1 and was used to scaffold contigs and bridge repeat sequences. The sequence was finished to standard criteria (61). Sequence assembly, visualization, and finishing were performed using PHRAP (www.phrap.org; P. Green, unpublished data) and Gap4 (10). All insertion element sequences were individually verified.

Annotation and genome comparison. Coding sequences were initially identified using a combination of Glimmer 2 (24) and Orpheus (31) and then manually curated using Artemis (61) and Frameplot (7). All genes were manually annotated in Artemis using standard criteria (6). Genome comparisons were visualized using the Artemis comparison tool (19). Putative orthologs were identified by reciprocal best-match FASTA searches of the *C. michiganensis* subsp. *sepedonicus*, *C. michiganensis* subsp. *michiganensis*, and *Leifsonia xyli* subsp. *xyli* protein sequences with the following cutoffs: 80% sequence length and 30% identity.

Time of divergence. Orthologous sequences were obtained for the following seven protein-encoding genes: *atpG* (921 nucleotides), *dnaK* (1,889 nucleotides), *fadA* (1,185 nucleotides), *gcpE* (1,107 nucleotides), *purM* (860 nucleotides), *rpoA* (944 nucleotides), and *trmU* (869 nucleotides). Sequences were aligned using MUSCLE v3.6 (28) and were edited using Se-AI v2.0a9 (http://tree.bio.ed.ac.uk

FIG. 1. Circular diagram of the chromosome of *C. michiganensis* subsp. *sepedonicus*. The circles provide the following information (from the outermost circle to the innermost circle): circles 1 and 2, all CDSs (transcribed clockwise and anticlockwise) (dark blue, pathogenicity and adaptation; black, energy metabolism; red, information transfer; dark green, surface associated; cyan, degradation of large molecules; magenta, degradation of small molecules; yellow, central and intermediary metabolism; pale green, unknown; pale blue, regulators; orange, conserved hypothetical; brown, pseudogenes; pink, phage and IS elements; gray, miscellaneous); circle 3, putative laterally acquired CDSs; circle 4, CDSs not present in *C. michiganensis* subsp. *michiganensis* or *L. xyli*; circle 5, pseudogenes; circle 6, IS element transposases; circle 7, G+C content (window size, 10,000 bp); and circle 8, GC deviation ($G - C/G + C$; window size, 10,000 bp).

/software/seal/), and maximum likelihood trees and branch lengths were obtained using GARLI v0.951 (http://www.molecularrevolution.org/software/garli/). Lacking external calibration dates (i.e., a fossil record), we employed a relatively simple distance estimation method of Kumar et al. (44): $t = d/2r$. The divergence time (t) at nodes was estimated from the nucleotide distance (d) calculated by adding the branch lengths obtained in maximum likelihood trees, and a mutation rate (r) of 5×10^{-10} mutation/bp/generation (26) and a generation time of 1 h were assumed. To take into consideration the slow in vitro and in planta growth of *C. michiganensis* subsp. *sepedonicus*, calculations were also made with an assumed generation time of 0.5 generation per day (5, 22). Divergence times were estimated for each gene individually, and standard deviations were calculated using individual estimates.

Nucleotide sequence accession numbers. The sequence and annotation of strain ATCC 33113 have been deposited in the EMBL database under the following accession numbers: chromosome, AM849034; plasmid pCS1, AM849035; and plasmid pCSL1, AM849036.

RESULTS AND DISCUSSION

General features of the *C. michiganensis* subsp. *sepedonicus* genome. The genome of *C. michiganensis* subsp. *sepedonicus* comprises a circular chromosome and two previously described plasmids, one circular (pCS1) and one linear (pCSL1) (Fig. 1 and Table 1). The genome size of ATCC 33113 as determined by contour-clamped homogeneous electric field analysis was previously estimated to be about 2.6 Mb (13). The actual size,

TABLE 1. General characteristics of the *C. michiganensis* subsp. *sepedonicus* ATCC 33113 genome

Characteristic	Chromosome	pCS1	pCSL1
Size (bp)	3,258,645	50,350	94,791 (including two 4,846-bp inverted repeats)
Geometry	Circular	Circular	Linear
G+C content (%)	72.56	67.46	68.84
Coding %	88.4	82.0	84.8
No. of CDSs	3,058	67	117
No. of CDSs which are pseudogenes	106	3	1
No. of rRNA operons	2	0	0
No. of tRNAs	45		
No. of IS elements	102	2	2
No. of IS1121	68	2	1
No. of ISCMi2	24	0	1
No. of ISCMi3	9	0	0
No. of IS1121/ISCMi2 chimeras	1	0	0

based on the genome sequence, is 3,258,645 bp. The relatively high chromosomal G+C content is typical of free-living actinomycetes, as is the slightly lower G+C content of the plasmids. The coding capacity of the *C. michiganensis* subsp. *sepedonicus* genome is reduced due to the presence of 110 pseudogenes (106 of which are chromosomal), which make up 3.4% of the predicted coding sequences (CDSs). This high level of nonfunctional genes suggests that there has been genome decay, which is often associated with bacterial lineages that appear to have recently acquired a new niche, rendering certain genes dispensable or disadvantageous, allowing or selecting for their functional ablation (8, 63, 75). The evolutionary bottleneck associated with niche adaptation may lead to increased fixation of deleterious mutations and expansion of insertion sequence (IS) elements (see below), both of which are a consequence of the reduced selective pressure associated with the bottleneck (9, 57, 62, 63).

Time of divergence. The assumptions about generation time greatly affected estimates of divergence. Based on a generation time of 1 h, which is reasonable for many plant-pathogenic bacteria, the divergence of *C. michiganensis* subsp. *sepedonicus* and *C. michiganensis* subsp. *michiganensis* occurred as few as 1,100 to 7,800 years ago (standard deviation, 6,800 years). However, when a more realistic, longer generation time, 0.5 generation/day, was used, the divergence of these pathogens was estimated to have occurred much longer ago: as few as 53,000 years ago but as many as 1,120,000 years ago (standard deviation, 330,000 years). Using either generation time placed the divergence of *C. michiganensis* subsp. *michiganensis* and *C. michiganensis* subsp. *sepedonicus* after the speciation dates for tomato (*Solanum esculentum*) and potato (*Solanum tuberosum*), ca. 4 to 5 million years ago (59, 83). Although the exact time of domestication of potato has not been established, this domestication is generally assumed to have taken place in the Bolivian-Peruvian Andes as early as 8,000 years ago (72). Based on the most conservative estimate of *C. michiganensis* subsp. *sepedonicus*-*C. michiganensis* subsp. *michiganensis* divergence (53,000 to 1,120,000 years ago), our findings suggest that

subspeciation within *C. michiganensis* predated known domestication events.

Chromosomal rearrangements and IS elements. The recent evolutionary pathway followed by *C. michiganensis* subsp. *sepedonicus* (see above) appears to have led to the expansion of IS elements. The *C. michiganensis* subsp. *sepedonicus* genome contains 106 IS elements, which fall into three groups, 71 IS1121 elements (68 chromosomal, 2 on pCS1, and 1 on pCSL1), 25 ISCMi2 elements (24 chromosomal and 1 on pCSL1), and nine ISCMi3 elements, as well as one element that appears to be a chimera between IS1121 and ISCMi2. IS1121 and ISCMi2 are members of the IS481 family, and ISCMi3 is related to the IS30 family (Table 1). IS1121 is widespread among strains of *C. michiganensis* subsp. *sepedonicus* (54). ISCMi2 is related to IS1122, which is repeated many times in the genome of the alfalfa pathogen *Clavibacter michiganensis* subsp. *insidiosus* (67). In terms of chromosomal coordinates, the IS elements appear to be randomly distributed. However, the majority are located in non-protein-encoding DNA, and only five are inserted directly into CDSs where they are likely to have caused loss of function (see Table S1 in the supplemental material). Alignment with the chromosome of *C. michiganensis* subsp. *michiganensis* revealed high levels of sequence identity, typically 90 to 100% (median, 95%) DNA

FIG. 2. Alignment of chromosomes of *C. michiganensis* subsp. *sepedonicus* (Cms), *C. michiganensis* subsp. *michiganensis* (Cmm), and *L. xyli* subsp. *xyli* (Lxx). The diagram shows forward and reverse DNA strands (gray bars), and base coordinates with the positions of *pat-1* homologues indicated by vertical black lines. Similar regions that are more than 1,000 bases long are indicated by red (colinear) and blue (inverted) lines.

identity for orthologous genes, and extensive rearrangements in *C. michiganensis* subsp. *sepedonicus*, mostly associated with recombination between IS elements in the genome of *C. michiganensis* subsp. *sepedonicus* (Fig. 2). Fifty-nine IS elements in *C. michiganensis* subsp. *sepedonicus* lie at the boundary of a region of synteny between *C. michiganensis* subsp. *sepedonicus* and *C. michiganensis* subsp. *michiganensis* and are likely to have been the foci of large-scale genomic recombination events. In three cases an IS element appears to have inserted within a CDS and a subsequent recombination has moved the two parts of the CDS to distant locations on the chromosome (see Table S1 in the supplemental material). The chromosome of the related actinomyete phytopathogen *L. xyli* subsp. *xyli* also contains large numbers of IS elements which appear to have generated extensive rearrangements (Fig. 2) (56). Although *C. michiganensis* subsp. *sepedonicus* IS1121 and ISCmi3 elements are related to IS elements found in *L. xyli* subsp. *xyli*, the levels of sequence identity are low and there are no syntenic/orthologous occurrences, indicating that all IS insertions occurred independently and after divergence of these species. Furthermore, the rarity of IS elements in the chromosome of *C. michiganensis* subsp. *michiganensis* suggests that the IS expansion in *C. michiganensis* subsp. *sepedonicus* was specific to this subspecies and may have coincided with or closely followed establishment of IS elements in the *C. michiganensis* subsp. *sepedonicus* genome. The high levels of sequence identity between *C. michiganensis* subsp. *sepedonicus* and *C. michiganensis* subsp. *michiganensis* suggest that subspeciation was relatively recent. The average levels of sequence identity between members of the IS element families (IS1121, 99.8%; ISCmi2, 97.9%; ISCmi3, 99.9%), along with the intact inverted repeat sequences, support the hypothesis that the acquisition and expansion of IS elements followed subspeciation and that the IS elements may still be functional (79). The hypothesis that the expansion of IS elements was a relatively recent event is also supported by the estimated time of divergence between *C. michiganensis* subsp. *sepedonicus* and *C. michiganensis* subsp. *michiganensis*, given the relatively long generation times and corresponding limited number of generations since divergence. Furthermore, the genomic variability among strains of *C. michiganensis* subsp. *sepedonicus* is low, as is the local variation associated with IS1121 (13, 30, 53).

Loss of gene function in the *C. michiganensis* subsp. *sepedonicus* chromosome. Although the *C. michiganensis* subsp. *sepedonicus* chromosome appears to have undergone extensive expansion of IS elements, only 5 of the 106 pseudogenes detected on the chromosome were due to IS insertion; the remainder were due to nonsense mutation, frameshift mutation, and partial deletion (see Table S1 in the supplemental material). This indicates that the mechanism driving formation of pseudogenes is only indirectly associated with IS expansion. However, it is notable that 39 of the 94 chromosomal IS insertions occurring in noncoding DNA are located directly upstream of a CDS whose expression they may affect (see Table S2 in the supplemental material). These genes include genes whose inactivation would be expected to have significant effects, such as CMS0434 encoding a LacI family transcriptional regulator, CMS0645 encoding a WhiB family transcriptional regulator, CMS1765 encoding a cytochrome transporter, and CMS2042 encoding the 3-hydroquinone dehydratase AroQ.

Moreover, there are incidents where insertion and subsequent recombination between IS elements appear to have segregated two parts of a gene cluster likely to constitute an operon without actually interrupting any CDSs. Possible examples include CMS0785- CMS0786 and CMS1044 to CMS1046, which form two parts of a glycogen metabolism operon whose orthologues are adjacent in both *C. michiganensis* subsp. *michiganensis* and *L. xyli* subsp. *xyli*, and CMS2725 to CMS2727 and CMS0565, which form two parts of the four-gene ABC phosphate transport operon conserved in *C. michiganensis* subsp. *michiganensis*, *L. xyli* subsp. *xyli*, and other actinomycetes (11, 80). Since genes in operons are cotranscribed and coregulated, it is possible that such rearrangements would disrupt or ablate the collective functions. It seems likely, therefore, that the disruption of gene function in *C. michiganensis* subsp. *sepedonicus* extends far beyond that of the identified pseudogenes.

The distribution of pseudogenes (and other possibly inactivated loci) across functional categories shows that genes for transport and degradation of carbohydrates, regulation, and specialized functions related to pathogenicity and adaptation are overrepresented (Fig. 3; see Table S1 in the supplemental material). Many pseudogenes encode enzymes likely to affect the ability of *C. michiganensis* subsp. *sepedonicus* to utilize carbohydrate nutrients, including cellulase CelB (encoded by CMS0045; cellulose utilization); glycerol kinase (encoded by CMS0701; glycerol utilization); *N*-acetylglucosamine-6-phosphate deacetylase (encoded by CMS0914; *N*-acetylglucosamine utilization, essential for growth of *Mycobacterium tuberculosis*) (68); three glycosyl hydrolases, one of which appears to be secreted (encoded by CMS0959, CMS1694, and CMS1700; carbohydrate degradation and utilization); glycogen-debranching enzyme TreX (encoded by CMS1527; glycogen utilization and trehalose synthesis); and tandem polysaccharide hydrolases (encoded by CMS2666 and CMS2667; carbohydrate degradation and utilization). Cellulase is an important determinant in *C. michiganensis* pathogenicity, so the disruption of *celB* is intriguing, although it should be noted that the plasmid-borne cellulase gene (*celA*), which is known to be involved in virulence in *C. michiganensis* subsp. *michiganensis*, appears to be intact (32, 45, 58). A gene encoding a 2,5-diketo-D-gluconic acid reductase is also disrupted. This enzyme is involved in the utilization of ketogluconates as a source of carbon and energy. It is also involved in the biosynthesis of ascorbate and has attracted much attention for its potential use in production of vitamin C.

Together with losses of peptidase (encoded by CMS1285) and lipase (encoded by CMS1291) activities, these disruptions in catabolic functions suggest a narrowing of nutrient utilization for *C. michiganensis* subsp. *sepedonicus*. A reduction in nutrient utilization is consistent with the fact that *C. michiganensis* subsp. *sepedonicus* is restricted to an endophytic niche in which the environmental conditions and carbohydrate supply are expected to be less varied than those experienced by plant epiphytic or soil-inhabiting bacteria (49, 55, 66). This hypothesis is further supported by the fact that all the *C. michiganensis* subsp. *sepedonicus* carbohydrate metabolism pseudogenes are intact and apparently functional in *C. michiganensis* subsp. *michiganensis*, which can multiply on a variety of plant surfaces. Curiously, the *C. michiganensis* subsp. *michiganensis* orthologue of the peptidase gene CMS1285

FIG. 3. Bar chart of functional classes of all CDSs (open bars) and pseudogenes (shaded bars). Note the overrepresentation of pseudogenes in CDSs encoding transport and binding proteins, macromolecule degradation, small-molecule degradation, pathogenicity, and regulation.

(CMM1338) is also disrupted, although by a different mechanism; CMS1285 contains a nonsense mutation, while in CMM1338 an IS element is inserted in the 3' region. Although clearly due to independent events, the loss of these genes in the two subspecies could reflect adaptation to a common niche in which peptidase activity is not required.

Genes for extracellular polysaccharide (EPS) biosynthesis have also been affected by genome decay in *C. michiganensis* subsp. *sepedonicus*. The *C. michiganensis* subsp. *michiganensis* genome has four gene clusters for production of EPS, and orthologous gene clusters are present in *C. michiganensis* subsp. *sepedonicus*. IS insertion in the gene for the polysaccharide polymerase Wzy (CMS2263) is likely to ablate the function of the EPS biosynthesis operon to which it belongs, and other mutations are likely to have inactivated at least one, and possibly two, of the remaining three clusters. The loss of the ability to produce an EPS coat suggests that *C. michiganensis* subsp. *sepedonicus* occupies a niche in which the production of such a coat is no longer advantageous or essential. It is tempting to speculate that IS insertions may play a role in generating naturally occurring mucoid and nonmucoid variants of *C. michiganensis* subsp. *sepedonicus* or the reported change from mucoid to nonmucoid morphology triggered by heat or nutrient stress (4, 41). Aromatic amino acid biosynthesis may be affected by an IS element insertion directly upstream of CMS2042 (*aroQ*; encoding 3-hydroquinate dehydrogenase [EC 4.2.1.10]).

The high proportion of pseudogenes in regulatory genes is expected to have had cascade effects on the global transcriptome and proteome and is likely to amplify the differences in the phenotypes of *C. michiganensis* subsp. *michiganensis* and *C. michiganensis* subsp. *sepedonicus* (52). Ten regulators, representing 10% of all pseudogenes and 5% of all regulators, are disrupted.

Agar plate-grown colonies of *C. michiganensis* subsp. *sepedonicus* and *C. michiganensis* subsp. *michiganensis* can be distinguished by color; the colonies are white or faint yellow and yellow, respectively (21). The pigmentation is thought to be due to production of carotenoids. Both genomes have a complete carotenoid biosynthesis gene cluster (CMS2604 and CMS2609 and CMM2884 to CMM2889) with no apparent pseudogenes to account for the phenotypic difference. One possible explanation may be the presence of an extra pair of genes for carotenoid cyclases (CMS0965 and CMS0966), which may modulate the final product in *C. michiganensis* subsp. *sepedonicus*. Other unknown regulatory differences may also be important.

FIG. 4. Venn diagram showing numbers of shared and unique genes in the genomes of *C. michiganensis* subsp. *sepedonicus* (Cms), *C. michiganensis* subsp. *michiganensis* (Cmm), and *L. xyli* subsp. *xyli* (Lxx). The red numbers do not include the number of IS element transposases.

Three-way coding sequence comparison and laterally acquired DNA. The predicted proteomes of *C. michiganensis* subsp. *sepedonicus*, *C. michiganensis* subsp. *michiganensis*, and *L. xyli* subsp. *xyli* were compared by three-way reciprocal FASTA analysis to assess the numbers of orthologous and unique CDSs (Fig. 4; see Table S3 in the supplemental material). The genome sizes and numbers of CDSs of *C. michiganensis* subsp. *sepedonicus* and *C. michiganensis* subsp. *michiganensis* are similar, so although the *C. michiganensis* subsp. *sepedonicus* genome appears to have undergone some decay, there does not appear to have been genome reduction in *C. michiganensis* subsp. *sepedonicus* compared to *C. michiganensis* subsp. *michiganensis*. However, given that these organisms are considered subspecies of the same species, they have surprisingly large numbers of subspecies-specific CDSs (12 to 16%), suggesting that there may have been significant differential gene acquisition or loss since divergence from the common ancestor. These proportions are equivalent to those seen in comparisons of *Escherichia coli* and *Salmonella enterica* genomes, where many of the unique genes are associated with horizontally acquired islands or prophages (81). Clearly, any subspecies-specific CDSs may be related to host-specific recognition, so it is notable that for both *C. michiganensis* subsp. *sepedonicus* and *C. michiganensis* subsp. *michiganensis* these CDSs include several CDSs encoding surface-exposed or secreted proteins and proteins involved in production and modification of surface polysaccharides (see Tables S3 and S7 in the supplemental material).

Excluding IS element transposases, CDSs present only in *C. michiganensis* subsp. *sepedonicus* often occur in clusters or islands, some of which have features characteristic of mobile islands, such as low-G+C-content IS elements, putative bacteriophage genes, putative plasmid genes, and/or flanking repeat sequences (Fig. 1; see Tables S3 and S4 in the supplemental material). Furthermore, at least seven islands are adjacent to tRNA genes, a frequent insertion site for mobile genetic elements. Many islands are discrete insertions in one genome compared to the other, but there are several gene clusters present only in *C. michiganensis* subsp. *sepedonicus* whose equivalent genomic locations in *C. michiganensis* subsp. *michiganensis* are occupied by alternative gene clusters present only in *C. michiganensis* subsp. *michiganensis*. These regions are often flanked by inverted repeat sequences, suggesting that they could be sites for future recombination.

The *C. michiganensis* subsp. *michiganensis* genome contains a large island (130 kb) known as the *chp/tom* region, which encodes known and putative virulence determinants (see the accompanying paper). The *C. michiganensis* subsp. *sepedonicus* genome does not contain an equivalent single large island, although it does share much of the gene content (Fig. 2). One *C. michiganensis* subsp. *sepedonicus* island (CmsPI) has significant synteny with the *tom* region, suggesting that either a mobile element integrated into the common ancestral genome and has since diverged or related mobile genetic elements have been independently introduced since divergence of the two lineages. Other regions of the *C. michiganensis* subsp. *sepedonicus* genome have significant matches with the *chp/tom* region CDSs but no other part of the *C. michiganensis* subsp. *michiganensis* genome. These regions include the divergently transcribed gene pair CMS2233 and CMS2234, which encode a

putative exported protein and putative secreted pectate lyase, respectively. It is also notable that *C. michiganensis* subsp. *michiganensis* chromosomal *pat-1* homologous genes (considered to be potential virulence genes) are located exclusively in the *chp/tom* region, while in *C. michiganensis* subsp. *sepedonicus* they are scattered throughout the chromosome (Fig. 2). Although *pat-1* genes have diverse sequences, making orthologue assignment impossible, it seems feasible that the *C. michiganensis* subsp. *sepedonicus*-*C. michiganensis* subsp. *michiganensis* common ancestral genome contained an island analogous to the *chp/tom* region which has remained largely intact in *C. michiganensis* subsp. *michiganensis* but has been dissipated throughout the *C. michiganensis* subsp. *sepedonicus* genome, possibly in association with IS-related recombination events. All but one of the *C. michiganensis* subsp. *sepedonicus* *pat-1* genes is located within four CDSs of an IS element. An alternative explanation for the differential distribution of *pat-1* homologues may be that *pat-1* genes have been acquired on multiple occasions as discrete insertions. Indeed, of the eight *pat-1* genes present on the *C. michiganensis* subsp. *sepedonicus* chromosome, six are present as pairs on three separate islands, one has inserted into and disrupted CDS CMS2908, and one (CMS2837) has inserted between two CDSs. There are no obvious repeat sequences flanking putative *pat-1* insertions, and the mechanism of insertion is unclear.

Genes present in any of the three genomes may have been present in the common ancestor; therefore, genes present in *C. michiganensis* subsp. *michiganensis* but not present in *C. michiganensis* subsp. *sepedonicus* may have been lost from *C. michiganensis* subsp. *sepedonicus*, although clearly they could also have been acquired by *C. michiganensis* subsp. *michiganensis*. Accepting this caveat, it is interesting that the genes present in *C. michiganensis* subsp. *michiganensis* but not present in *C. michiganensis* subsp. *sepedonicus* have a distribution of functions similar to that seen for *C. michiganensis* subsp. *sepedonicus* pseudogenes, with a high frequency of catabolic functions such as degradation and transport of carbohydrates and peptides (see Table S5 in the supplemental material). Gene loss may therefore have been under the same selective influences as pseudogene formation.

Pathogenicity determinants and host adaptation. The relative genetic intractability of the *Clavibacter* species has meant that there has been little correlation of genes with pathogenicity. The major candidate functions so far are exopolysaccharides and secreted enzyme activities, such as endocellulase, xylanase, polygalacturonase, and serine protease. The clearest demonstrations of *Clavibacter* pathogenicity genes have been for the cellulase-encoding *celA* gene and the serine protease-encoding *pat-1* gene, both carried on plasmids in *C. michiganensis* subsp. *michiganensis* (32). The *C. michiganensis* subsp. *michiganensis* *celA* gene is on plasmid pCM1, and an intact orthologue is present on *C. michiganensis* subsp. *sepedonicus* plasmid pCS1. However, a second cellulase gene, *celB* (CMS0045 and CMM2443), present on the chromosome of both subspecies, has been inactivated in *C. michiganensis* subsp. *sepedonicus* by a nonsense mutation at codon 192. The *C. michiganensis* subsp. *michiganensis* *pat-1* gene is present on plasmid pCM2. Homologues of *pat-1* are referred to as *chp* (for chromosomal homologue of *pat-1*) or *php* (for plasmid homologue of *pat-1*). For a phylogenetic analysis of *pat-1* ho-

mologues in *C. michiganensis* subsp. *sepedonicus* and *C. michiganensis* subsp. *michiganensis*, see the accompanying paper. *C. michiganensis* subsp. *sepedonicus* has 11 *pat-1* homologues; 8 are chromosomal, 2 are on plasmid pCS1, and 1 is on pCSL1 (see Table S6 in the supplemental material). Of the eight *pat-1* homologues on the *C. michiganensis* subsp. *sepedonicus* chromosome, six appear to be intact with N-terminal signal sequences, one lacks a signal sequence (CMS1260), and one has a frameshift mutation (CMS0980). Alignments of the *C. michiganensis* subsp. *sepedonicus* and *C. michiganensis* subsp. *michiganensis* *pat-1* homologues suggest that there are distinct lineages within *pat-1* homologues, and the lineages are generally analogous in *C. michiganensis* subsp. *sepedonicus* and *C. michiganensis* subsp. *michiganensis*, with *C. michiganensis* subsp. *sepedonicus* *chp-3*, *chp-4*, and *chp-5* representing a lineage distinct from the *C. michiganensis* subsp. *michiganensis* lineage (see the accompanying paper). Only *C. michiganensis* subsp. *sepedonicus* *chp-7* and *php-2* contain the LPGSG sortase signal for cell wall anchoring of the protein. *C. michiganensis* subsp. *sepedonicus* *chp-7* is most like the *C. michiganensis* subsp. *michiganensis* *pat-1* gene, with 82% amino acid identity. In comparison, *C. michiganensis* subsp. *michiganensis* has three *pat-1* homologues (including *pat-1* itself) on pCM2, and seven homologues are clustered within the *C. michiganensis* subsp. *michiganensis* chromosomal *chp/tom* region. All 10 homologues have an N-terminal signal sequence, but only 7 appear to be intact and 3 of the chromosomal genes contain frameshift mutations.

The *C. michiganensis* subsp. *michiganensis* *tomA* subregion of the *chp/tom* region contains a gene, *tomA* (CMM0090), which encodes an exported endo-1,4-beta-glycosidase involved in the detoxification of the saponin α -tomatine, a plant defense and antimicrobial compound produced by tomato and other members of the Solanaceae (38). TomA is a member of a family of glycoside hydrolases which match the Pfam domain model PF00331. Bacterial proteins matching this domain model generally have a single domain and an N-terminal signal sequence, and characterized examples are involved with plant-specific glycans, primarily xylan (2, 33). They are relatively rare, and the encoding genes tend to occur only once in environmental bacteria likely to be associated with plants or algae. The *C. michiganensis* subsp. *sepedonicus* genome includes one CDS (CMS0087) with a match to PF00331. Although the similarity between CMS0087 and *tomA* from *C. michiganensis* subsp. *michiganensis* is low (23.9% identity and 41.8% similarity over 201 residues), they are both the only members of the PF00331 group in their respective genomes, and dot plot alignment showed that they are clearly related (Fig. 5). This suggests that they may have analogous, if not orthologous, functions, and CMS0087 may be involved in degradation of potato-produced glycoalkaloids present during infection with *C. michiganensis* subsp. *sepedonicus* (65). However, CMS0087 is likely to be disabled, and its genomic status is notable. It lies directly upstream of an IS element (CMS0086) which appears to have truncated the 3' end of the gene. The 5' region may also have been lost, as it does not encode the N-terminal signal sequence present in similar proteins (34). CMS0086 and CMS0087 lie within an exopolysaccharide biosynthesis gene cluster (EPS2), and alignment with the *C. michiganensis* subsp. *michiganensis* genome showed that the current genomic ar-

FIG. 5. Dot plot amino acid alignment of the CMS0087 with CMM0090 (TomA) proteins. The axes show the amino acid residue, and the dots and lines indicate amino acid identity between the protein sequences. The plot was generated using Dotter (71). The Karlin/Altschul statistics for the sequences and score matrix are as follows: $K = 0.129$; $\Lambda = 0.301$; expected MSP score in a 100×100 matrix, 23.817; expected residue score in MSP, 1.186; expected MSP length, 20.

angement is likely to be the result of recombination between IS elements. The potential loss of the ability to degrade glycoalkaloids in *C. michiganensis* subsp. *sepedonicus* may indicate that either this organism does not encounter such growth inhibitors in its current niche or it has adapted to a slow-growth lifestyle to avoid such plant defense mechanisms.

Other genes encoding proteins with a potential impact on pathogenicity include CMS0584 (encoding a putative siderophore-binding protein), CMS0653 and CMS0654 (encoding putative heavy metal detoxification), CMS0682 (encoding a putative peroxidase), CMS0930 (encoding iron uptake permease), CMS0960 (encoding a putative secreted glycosyl hydrolase), CMS0974 (encoding a putative hydroperoxide resistance protein), CMS1135 (encoding a putative siderophore biosynthesis protein), CMS1296 (encoding nonheme haloperoxidase), CMS1306 (encoding a putative gamma-glutamyltransferase), CMS1449 (encoding a putative siderophore-interacting protein), CMS1551 (encoding a putative heme-binding protein), CMS1668 (encoding a putative undecaprenyl diphosphatase), CMS1881 (encoding a putative iron-chelating protein), CMS1989 (encoding superoxide dismutase), CMS2178 (encoding the endo-polygalacturonase Peh), CMS2234 (encoding a putative secreted pectate lyase), CMS2235 (encoding the catalase KatA), CMS2291 (encoding a putative sortase-sorted copper resistance surface protein), CMS2719 (encoding a putative quaternary ammonium compound efflux protein), CMS2835 (encoding a putative heme oxygenase), CMS3013 (encoding a putative salicylate biosynthesis isochorismate synthase),

CMS3048 (encoding a putative manganese catalase), and CMS3063 to CMS3066 (encoding a putative iron-siderophore uptake system). Thus, *C. michiganensis* subsp. *sepedonicus* has the genetic capacity to withstand low iron and oxidative stresses, which may be present during the infection process (20).

There are also several genes with the potential to encode resistance to antibiotics, such as CMS0149 (encoding a putative aminoglycoside phosphotransferase), CMS0172 (encoding a putative VanZ-like membrane protein) (3), CMS0216 (encoding a putative cytidine deaminase with 55% amino acid identity over the full length to blasticidin S deaminase from *Aspergillus terreus*) (39), CMS0694 (encoding a putative macrolide resistance protein), CMS0862 (encoding a putative multidrug efflux protein), CMS961 (encoding a putative drug efflux protein), CMS1440 (encoding a putative toxin resistance acetyltransferase), CMS1893 (encoding a putative macrolide phosphotransferase), CMS2286 (encoding a putative resistance protein), CMS2306 (encoding a putative dimethyladenosine transferase), CMS2483 (encoding a putative drug efflux protein), CMS2903 (encoding a putative drug resistance dioxygenase), CMS2936 (encoding a putative multiple-antimicrobial-agent extrusion protein), and CMS3023 and CMS3049 (encoding a putative beta-lactamase). *C. michiganensis* subsp. *sepedonicus* growth in culture is often inhibited by the presence of other microbes in plant samples, making disease diagnosis by pathogen cultivation especially challenging and necessitating other, less-culture-dependent approaches (69, 76, 77). Thus, finding several antibiotic resistance genes was unexpected.

Exopolysaccharide production. The *C. michiganensis* subsp. *sepedonicus* and *C. michiganensis* subsp. *michiganensis* chromosomes each contain four gene clusters for Wzx/Wzy-dependent biosynthesis of exported polysaccharide (designated EPS1 to EPS4 according to their order in *C. michiganensis* subsp. *michiganensis* [see Table S7 in the supplemental material]). Such gene clusters generally encode glycosyl transferases necessary for linking of sugars to form the oligosaccharide repeat unit, a Wzx flippase required for transport of the repeat unit across the cytoplasmic membrane, and a Wzy polymerase responsible for linking repeat units to form the polysaccharide. The gene clusters are broadly syntenic, and there are some notable differences between the subspecies. All four EPS clusters in *C. michiganensis* subsp. *michiganensis* appear to be intact and are likely to be functional. In *C. michiganensis* subsp. *sepedonicus* three of the four clusters appear to have been disrupted, and at least two are likely to have been inactivated.

The *C. michiganensis* subsp. *sepedonicus* EPS1 repeat unit polymerase gene (*wzyI* [CMS2263]) has been interrupted, and probably inactivated, by insertion of an *IS1121* element. The divergence in Wzy sequences suggests that there are differences in substrate specificity, making it unlikely that the polymerase encoded by one of the other three EPS clusters could compensate. The mutation in *wzyI* is likely to ablate the function of the entire gene cluster due to the central role played by Wzy. Also, EPS2 in *C. michiganensis* subsp. *sepedonicus* has been grossly disrupted by recombination between IS elements. The central portion of the gene cluster, including genes encoding two glycosyl transferases (CMS2390 and CMS2391), Wzx flippase (CMS2389), and a candidate Wzy polymerase (CMS2400), has been translocated to a

distant region of the genome flanked by a pair of *IS1121* elements and replaced by another *IS1121* and a CDS (CMS0087) encoding a putative glycoside hydrolase. This rearrangement is likely to have disrupted the regulation of the gene cluster, probably rendering it nonfunctional. Further disruption of the *C. michiganensis* subsp. *sepedonicus* EPS2 gene cluster is evident in CMS0084, which has a nonsense mutation at codon 871 and deletions in the 3' region compared to the equivalent CDS in *C. michiganensis* subsp. *michiganensis*.

EPS3 and EPS4 seem to be largely intact and possibly functional in *C. michiganensis* subsp. *sepedonicus*, although an *IS1121* element in the 3' end of the *galE* gene in EPS3 may disrupt the function of the protein product and may have polar effects on the expression of the rest of the gene cluster. Comparison of the EPS4 clusters of *C. michiganensis* subsp. *sepedonicus* and *C. michiganensis* subsp. *michiganensis* showed that although they are clearly related, they have different complements of genes in the central region, including the genes for Wzx flippase, Wzy polymerase, and alternative glycosyl transferases, indicating that they are likely to produce grossly different polysaccharides. Altogether, it appears that *C. michiganensis* subsp. *sepedonicus* has lost at least half of its ability to produce EPS due to genome degradation. Such polysaccharides are located at the cell surface, and their involvement in interactions with the environment and particularly host organisms has been well documented. This correlates well with the notion that *C. michiganensis* subsp. *sepedonicus* recently adapted to a narrowed niche where such interactions are less variable. Additional genetic studies are needed to demonstrate the specific gene set required for EPS biosynthesis and to reconcile reports on sugar composition and the contribution of EPS to virulence in *C. michiganensis* subsp. *sepedonicus* (37, 82).

Conclusions. While other members of the species *C. michiganensis* can grow in a variety of environmental and plant-associated niches, *C. michiganensis* subsp. *sepedonicus* is almost entirely restricted to the vascular system of its host plant. Analysis of the *C. michiganensis* subsp. *sepedonicus* genome revealed numerous correlations with this endophytic lifestyle and suggested that there was recent specialization for life in this restricted niche and that the organism has a reduced ability to exploit previously occupied complex niches outside the plant.

Tolerance to generation of pseudogenes, expansion of IS elements, genome rearrangements, and the associated disruption of operons suggest that there was relaxation of selective pressure or an increase in fixation of mutations due to genetic drift during the recent evolutionary history of *C. michiganensis* subsp. *sepedonicus*. These events may have been due to passage through an evolutionary bottleneck associated with niche acquisition and adaptation in *C. michiganensis* subsp. *sepedonicus* (62, 63). The bottleneck may have involved a change from a plant-associated generalist lifestyle, where the organism thrived in soil and on various plant surfaces and was able to opportunistically exploit plant wounds, to a lifestyle where the bacterium could live successfully within the plant vascular system without a need for movement outside that niche. This change may have been triggered by loss or acquisition of a function leading to enhanced success within the newly acquired niche. One possible trigger is the observed disruption of *C.*

michiganensis subsp. *sepedonicus*'s ability to produce surface polysaccharides. EPS is considered a virulence determinant in many bacterial plant pathogens because of its central role in wilt induction, host colonization, and biofilm formation, and it is thus likely to be important in physical interactions with the host (1, 42). It is feasible that alteration of *C. michiganensis* subsp. *sepedonicus* surface polysaccharides or some other genetic change altered the host interaction such that reducing the host response to invasion by the pathogen enhanced survival within the vascular system. In this regard, it is interesting that *C. michiganensis* subsp. *sepedonicus* appears to have lost the ability to produce a plant defense detoxification enzyme.

Adaptation to the narrower niche of the vascular system would have allowed disruption of genes whose products are no longer required. Bacteria on leaf surfaces must adapt to an ever-changing set of environmental signals by modulating their own gene expression (49). The abundance of pseudogenes in regulatory genes suggests that *C. michiganensis* subsp. *sepedonicus* lost some of its capacity to adapt to such harsh or varied environments. Pseudogene functions indicate that there was a reduction in nutrient diversity, allowing loss of catabolic enzymes and nutrient transporters. This generally reduced constraint on genome disruption would have also allowed the multiplication of IS elements and associated chromosomal rearrangements. The fact that pseudogenes have not been removed from the genome suggests that the bottleneck and subsequent events were relatively recent. This suggestion is supported by the intact status of the IS elements and the fact that the genome size is not reduced.

ACKNOWLEDGMENTS

We acknowledge the use of core facilities at the Wellcome Trust Sanger Institute. We thank Karl-Heinz Gartemann, Rudolf Eichenlaub, and Alfred Pühler for helpful discussions and for sharing information prior to publication.

This project was supported by Initiative for Future Agriculture and Food Systems grant 2001-52100-11428 from the USDA Cooperative State Research, Education, and Extension Service, the Colorado Experiment Station, and the Minnesota Experiment Station.

REFERENCES

- Abramovitch, R. B., J. C. Anderson, and G. B. Martin. 2006. Bacterial elicitation and evasion of plant innate immunity. *Nat. Rev. Mol. Cell Biol.* 7:601-611.
- Adelsberger, H., C. Hertel, E. Glawischnig, V. V. Zverlov, and W. H. Schwarz. 2004. Enzyme system of *Clostridium stercorarium* for hydrolysis of arabinoxylan: reconstitution of the *in vivo* system from recombinant enzymes. *Microbiology* 150:2257-2266.
- Arthur, M., F. Depardieu, C. Molinas, P. Reynolds, and P. Courvalin. 1995. The *vanZ* gene of *Tn1546* from *Enterococcus faecium* BM4147 confers resistance to teicoplanin. *Gene* 154:87-92.
- Baer, D., and N. C. Gudmestad. 1993. Serological detection of nonmucoid strains of *Clavibacter michiganensis* subsp. *sepedonicus* in potato. *Phytopathology* 83:157-163.
- Baer, D., A. R. White, and N. C. Gudmestad. 1998. Partial characterization of an extracellular beta-fructofuranosidase from *Clavibacter michiganensis* subspecies *sepedonicus*. *Can. J. Microbiol.* 44:852-865.
- Berriman, M., and K. Rutherford. 2003. Viewing and annotating sequence data with Artemis. *Brief Bioinform.* 4:124-132.
- Bibb, M. J., P. R. Findlay, and M. W. Johnson. 1984. The relationship between base composition and codon usage in bacterial genes and its use for the simple and reliable identification of protein-coding sequences. *Gene* 30:157-166.
- Blanc, G., H. Ogata, C. Robert, S. Audic, K. Suhre, G. Vestris, J. M. Claverie, and D. Raoult. 2007. Reductive genome evolution from the mother of *Rickettsia*. *PLoS Genet.* 3:e14.
- Bolotin, A., B. Quinquis, P. Renault, A. Sorokin, S. D. Ehrlich, S. Kulakauskas, A. Lapidus, E. Goltsman, M. Mazur, G. D. Pusch, M. Fonstein, R. Overbeek, N. Kyprides, B. Purnelle, D. Prozzi, K. Ngui, D. Masuy, F. Hancy, S. Burteau, M. Boutry, J. Delcour, A. Goffeau, and P. Hols. 2004. Complete sequence and comparative genome analysis of the dairy bacterium *Streptococcus thermophilus*. *Nat. Biotechnol.* 22:1554-1558.
- Bonfield, J. K., K. Smith, and R. Staden. 1995. A new DNA sequence assembly program. *Nucleic Acids Res.* 23:4992-4999.
- Braibant, M., L. Dewit, P. Peirs, M. Kalai, J. Ooms, A. Drowart, K. Huygen, and J. Content. 1994. Structure of the *Mycobacterium tuberculosis* antigen-88, a protein related to the *Escherichia coli* PstA periplasmic phosphate permease subunit. *Infect. Immun.* 62:849-854.
- Brown, S. E., D. L. Knudson, and C. A. Ishimaru. 2002. Linear plasmid in the genome of *Clavibacter michiganensis* subsp. *sepedonicus*. *J. Bacteriol.* 184:2841-2844.
- Brown, S. E., A. A. Reilley, D. L. Knudson, and C. A. Ishimaru. 2002. Genomic fingerprinting of virulent and avirulent strains of *Clavibacter michiganensis* subspecies *sepedonicus*. *Curr. Microbiol.* 44:112-119.
- Brumbley, S. M., L. A. Petrasovits, S. R. Hermann, A. J. Young, and B. J. Croft. 2006. Recent advances in the molecular biology of *Leifsonia xyli* subsp. *xyli*, causal organism of ratoon stunting disease. *Aust. Plant Pathol.* 35:681-689.
- Burger, A., I. Grafen, J. Engemann, E. Niermann, M. Pieper, O. Kirchner, K. H. Gartemann, and R. Eichenlaub. 2005. Identification of homologues to the pathogenicity factor Pat-1, a putative serine protease of *Clavibacter michiganensis* subsp. *michiganensis*. *Microbiol. Res.* 160:417-427.
- Carlson, R. R., and A. K. Vidaver. 1982. Taxonomy of *Corynebacterium* plant pathogens, including a new pathogen of wheat, based on polyacrylamide gel electrophoresis of cellular proteins. *Int. J. Syst. Bacteriol.* 32:315-326.
- Carlton, W. M., E. J. Braun, and M. L. Gleason. 1998. Ingress of *Clavibacter michiganensis* subsp. *michiganensis* into tomato leaves through hydathodes. *Phytopathology* 88:525-529.
- Carlton, W. M., M. L. Gleason, and E. J. Braun. 1994. Effects of pruning on tomato plants supporting epiphytic populations of *Clavibacter michiganensis* subsp. *michiganensis*. *Plant Dis.* 78:742-745.
- Carver, T. J., K. M. Rutherford, M. Berriman, M. A. Rajandream, B. G. Barrell, and J. Parkhill. 2005. ACT: the Artemis comparison tool. *Bioinformatics* 21:3422-3423.
- Coaker, G. L., B. Willard, M. Kinter, E. J. Stockinger, and D. M. Francis. 2004. Proteomic analysis of resistance mediated by Rcm 2.0 and Rcm 5.1, two loci controlling resistance to bacterial canker of tomato. *Mol. Plant-Microbe Interact.* 17:1019-1028.
- Davis, M. J., A. G. Gillaspie, A. K. Vidaver, and R. W. Harris. 1984. *Clavibacter*: a new genus containing some phytopathogenic coryneform bacteria, including *Clavibacter xyli* subsp. *xyli* sp. nov., subsp. nov. and *Clavibacter xyli* subsp. *cynodontis* subsp. nov., pathogens that cause ratoon stunting disease of sugarcane and bermudagrass stunting disease. *Int. J. Syst. Bacteriol.* 34:107-117.
- De Boer, S. H., and M. McCann. 1989. Determination of population densities of *Corynebacterium sepedonicum* in potato stems during the growing season. *Phytopathology* 79:946-951.
- De Boer, S. H., and S. A. Slack. 1984. Current status and prospects for detecting and controlling bacterial ring rot of potatoes in North America. *Plant Dis.* 68:841-844.
- Delcher, A. L., D. Harmon, S. Kasif, O. White, and S. L. Salzberg. 1999. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res.* 27:4636-4641.
- de Souza, M. L., D. Newcombe, S. Alvey, D. E. Crowley, A. Hay, M. J. Sadowsky, and L. P. Wackett. 1998. Molecular basis of a bacterial consortium: interspecies catabolism of atrazine. *Appl. Environ. Microbiol.* 64:178-184.
- Drake, J. W. 1991. A constant rate of spontaneous mutation in DNA-based microbes. *Proc. Natl. Acad. Sci. USA* 88:7160-7164.
- Dreier, J., D. Meletzus, and R. Eichenlaub. 1997. Characterization of the plasmid encoded virulence region *pat-1* of phytopathogenic *Clavibacter michiganensis* subsp. *michiganensis*. *Mol. Plant-Microbe Interact.* 10:195-206.
- Edgar, R. C. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinform.* 5:1-19.
- Evtushenko, L., L. Dorofeeva, S. Subbotin, J. Cole, and J. Tiedje. 2000. *Leifsonia poae* gen. nov., sp. nov., isolated from nematode galls on *Poa annua*, and reclassification of '*Corynebacterium aquaticum*' Leifson 1962 as *Leifsonia aquatica* (ex Leifson 1962) gen. nov., nom. rev., comb. Nov. and *Clavibacter xyli* Davis et al. 1984 with two subspecies as *Leifsonia xyli* (Davis et al. 1984) gen. nov., comb. nov. *Int. J. Syst. Evol. Microbiol.* 50:371-380.
- Fousek, J., I. Mraz, and K. Petrzik. 2002. Comparison of genetic variability between Czech and foreign isolates of phytopathogenic bacteria *Clavibacter michiganensis* subsp. *sepedonicus* by Rep-PCR technique. *Folia Microbiol.* 47:450-454.
- Frishman, D., A. Mironov, H. W. Mewes, and M. Gelfand. 1998. Combining diverse evidence for gene recognition in completely sequenced bacterial genomes. *Nucleic Acids Res.* 26:2941-2947.
- Gartemann, K. H., O. Kirchner, J. Engemann, I. Grafen, R. Eichenlaub, and A. Burger. 2003. *Clavibacter michiganensis* subsp. *michiganensis*: first steps in

- the understanding of virulence of a Gram-positive phytopathogenic bacterium. *J. Biotechnol.* **106**:179–191.
33. Gat, O., A. Lapidot, I. Alchanati, C. Regueros, and Y. Shoham. 1994. Cloning and DNA sequence of the gene coding for *Bacillus stearothermophilus* T-6 xylanase. *Appl. Environ. Microbiol.* **60**:1889–1896.
 34. Gibbs, M. D., R. A. Reeves, and P. L. Bergquist. 1995. Cloning, sequencing, and expression of a xylanase gene from the extreme thermophile *Dictyoglomus thermophilum* Rt46b.1 and activity of the enzyme on fiber-bound substrate. *Appl. Environ. Microbiol.* **61**:4403–4408.
 35. Gross, D. C., A. K. Vidaver, and M. B. Keralis. 1979. Indigenous plasmids from phytopathogenic *Corynebacterium* species. *J. Gen. Microbiol.* **115**:479–489.
 36. Hahn, M. W., H. Lunsdorf, Q. Wu, M. Schauer, M. G. Hofle, J. Boenigk, and P. Stadler. 2003. Isolation of novel ultramicrobacteria classified as actinobacteria from five freshwater habitats in Europe and Asia. *Appl. Environ. Microbiol.* **69**:1442–1451.
 37. Henningson, P. J., and N. C. Gudmestad. 1993. Comparison of exopolysaccharides from mucoid and nonmucoid strains of *Clavibacter michiganensis* subspecies *sepedonicus*. *Can. J. Microbiol.* **39**:291–296.
 38. Kaup, O., I. Grafen, E. M. Zellermann, R. Eichenlaub, and K. H. Gartemann. 2005. Identification of a tomatinase in the tomato-pathogenic actinomycete *Clavibacter michiganensis* subsp. *michiganensis* NCPPB382. *Mol. Plant-Microbe Interact.* **18**:1090–1098.
 39. Kimura, M., S. Sekido, Y. Isogai, and I. Yamaguchi. 2000. Expression, purification, and characterization of blasticidin S deaminase (BSD) from *Aspergillus terreus*: the role of catalytic zinc in enzyme structure. *J. Biochem.* **127**:955–963.
 40. Kirchner, O., K. H. Gartemann, E. M. Zellermann, R. Eichenlaub, and A. Burger. 2001. A highly efficient transposon mutagenesis system for the tomato pathogen *Clavibacter michiganensis* subsp. *michiganensis*. *Mol. Plant-Microbe Interact.* **14**:1312–1318.
 41. Kokoskova, B., and V. Kudela. 2002. Induction of nonmucoid variants of *Clavibacter michiganensis* subsp. *sepedonicus* and comparison of their immunochromatological and biochemical characteristics. *Z. Pflanzenkr. Pflanzenschutz.* **109**:630–638.
 42. Koutsoudis, M. D., D. Tsaltas, T. D. Minogue, and S. B. von Bodman. 2006. Quorum-sensing regulation governs bacterial adhesion, biofilm development, and host colonization in *Pantoea stewartii* subspecies *stewartii*. *Proc. Natl. Acad. Sci. USA* **103**:5983–5988.
 43. Kreutzer, W. A., D. P. Glick, and J. G. McLean. 1941. Bacterial ring rot of potato. *Colo. Expt. Stn. Press Bull.* **94**:1–11.
 44. Kumar, S., K. A. Balczarek, and Z. C. Lai. 1996. Evolution of the hedgehog gene family. *Genetics* **142**:965–972.
 45. Laine, M. J., M. Haapalainen, T. Wahlroos, K. Kankare, R. Nissinen, S. Kassuwi, and M. C. Metzler. 2000. The cellulase encoded by the native plasmid of *Clavibacter michiganensis* ssp. *sepedonicus* plays a role in virulence and contains an expansin-like domain. *Physiol. Mol. Plant Pathol.* **57**:221–233.
 46. Laine, M. J., H. Nakhei, J. Dreier, K. Lehtila, D. Meletus, R. Eichenlaub, and M. C. Metzler. 1996. Stable transformation of the gram-positive phytopathogenic bacterium *Clavibacter michiganensis* subsp. *sepedonicus* with several cloning vectors. *Appl. Environ. Microbiol.* **62**:1500–1506.
 47. Lee, I.-M., I. M. Bartoszyk, D. E. Gundersen-Rindal, and R. E. Davis. 1997. Phylogeny and classification of bacteria in the genera *Clavibacter* and *Rathayibacter* on the basis of 16S rRNA gene sequence analyses. *Appl. Environ. Microbiol.* **63**:2631–2636.
 48. Li, T. Y., H. L. Zeng, Y. Ping, H. Lin, X. L. Fan, Z. G. Guo, and C. F. Zhang. 2007. Construction of a stable expression vector for *Leifsonia xyli* subsp. *cynodontis* and its application in studying the effect of the bacterium as an endophytic bacterium in rice. *FEMS Microbiol. Lett.* **267**:176–183.
 49. Lindow, S. E., and M. T. Brandl. 2003. Microbiology of the phyllosphere. *Appl. Environ. Microbiol.* **69**:1875–1883.
 50. Louws, F. J., J. Bell, C. M. Medina-Mora, C. D. Smart, D. Ogenorth, C. A. Ishimaru, M. K. Hausbeck, F. J. de Bruijn, and D. W. Fulbright. 1998. rep-PCR-mediated genomic fingerprinting: a rapid and effective method to identify *Clavibacter michiganensis*. *Phytopathology* **88**:862–868.
 51. Medina-Mora, C. M., M. K. Hausbeck, and D. W. Fulbright. 2001. Bird's eye lesions of tomato fruit: production by aerosol and direct application of *Clavibacter michiganensis* subsp. *michiganensis*. *Plant Dis.* **85**:88–91.
 52. Mira, A., and R. Pushker. 2005. The silencing of pseudogenes. *Mol. Biol. Evol.* **22**:2135–2138.
 53. Mogen, B. D., A. E. Oleson, R. B. Sparks, N. C. Gudmestad, and G. A. Secor. 1988. Distribution and partial characterization of pCS1, a highly conserved plasmid present in *Clavibacter michiganensis* subsp. *sepedonicus*. *Phytopathology* **78**:1381–1386.
 54. Mogen, B. D., H. R. Olson, R. B. Sparks, N. C. Gudmestad, and A. E. Oleson. 1990. Genetic variation in strains of *Clavibacter michiganensis* subsp. *sepedonicus*: polymorphisms in restriction fragments containing a highly repeated sequence. *Phytopathology* **80**:90–96.
 55. Mongodin, E. F., N. Shapir, S. C. Daugherty, R. T. Deboy, J. B. Emerson, A. Shvartzbeyn, D. Radune, J. Vamathevan, F. Riggs, V. Grinberg, H. Khouri, L. P. Wackett, K. E. Nelson, and M. J. Sadowsky. 2006. Secrets of soil survival revealed by the genome sequence of *Arthrobacter aurescens* TC1. *PLoS Genet.* **2**:2094–2106.
 56. Monteiro-Vitorello, C. B., L. E. A. Camargo, M. A. Van Sluys, J. P. Kitajima, D. Truffi, A. M. do Amaral, R. Harakava, J. C. F. de Oliveira, D. Wood, M. C. de Oliveira, C. Miyaki, M. A. Takita, A. C. R. da Silva, L. R. Furlan, D. M. Carraro, G. Camarotte, N. F. Almeida, H. Carrer, L. L. Coutinho, H. A. El-Dorry, M. I. T. Ferro, P. R. Gagliardi, E. Gigliotti, M. H. S. Goldman, G. H. Goldman, E. T. Kimura, E. S. Ferro, E. E. Kuramae, E. G. M. Lemos, M. V. F. Lemos, S. M. Z. Mauro, M. A. Machado, C. L. Marino, C. F. Menck, L. R. Nunes, R. C. Oliveira, G. G. Pereira, W. Siqueira, A. A. de Souza, S. M. Tsai, A. S. Zanca, A. J. G. Simpson, S. M. Brumley, and J. C. Setubal. 2004. The genome sequence of the gram-positive sugarcane pathogen *Leifsonia xyli* subsp. *xyli*. *Mol. Plant-Microbe Interact.* **17**:827–836.
 57. Nierman, W. C., D. DeShazer, H. S. Kim, H. Tettelin, K. E. Nelson, T. Feldblyum, R. L. Ulrich, C. M. Ronning, L. M. Brinkac, S. C. Daugherty, T. D. Davidsen, R. T. Deboy, G. Dimitrov, R. J. Dodson, A. S. Durkin, M. L. Gwinn, D. H. Haft, H. Khouri, J. F. Kolonay, R. Madupu, Y. Mohamoud, W. C. Nelson, D. Radune, C. M. Romero, S. Sarria, J. Selengut, C. Shamblyn, S. A. Sullivan, O. White, Y. Yu, N. Zafar, L. W. Zhou, and C. M. Fraser. 2004. Structural flexibility in the *Burkholderia mallei* genome. *Proc. Natl. Acad. Sci. USA* **101**:14246–14251.
 58. Nissinen, R., S. Kassuwi, R. Peltola, and M. C. Metzler. 2001. *In planta* complementation of *Clavibacter michiganensis* subsp. *sepedonicus* strains deficient in cellulase production or HR induction restores virulence. *Eur. J. Plant Pathol.* **107**:175–182.
 59. Olmstead, R. G., and J. A. Sweere. 1994. Combining data in phylogenetic systematics—an empirical approach using 3 molecular data sets in the Solanaceae. *Syst. Biol.* **43**:467–481.
 60. Park, Y.-H., K.-I. Suzuki, D.-G. Yim, K.-C. Lee, E. Kim, J.-S. Yoon, S.-J. Kim, Y.-H. Kho, M. Goodfellow, and K. Komagata. 1993. Suprageneric classification of peptidoglycan group B actinomycetes by nucleotide sequencing of 5S ribosomal RNA. *Antonie van Leeuwenhoek* **64**:307–313.
 61. Parkhill, J., M. Achtman, K. D. James, S. D. Bentley, C. Churcher, S. R. Klee, G. Morelli, D. Basham, D. Brown, T. Chillingworth, R. M. Davies, P. Davis, K. Devlin, T. Feltham, N. Hamlin, S. Holroyd, K. Jagels, S. Leather, S. Moule, K. Mungall, M. A. Quail, M. A. Rajandream, K. M. Rutherford, M. Simmonds, J. Skelton, S. Whitehead, B. G. Spratt, and B. G. Barrell. 2000. Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491. *Nature* **404**:502–506.
 62. Parkhill, J., M. Sebaihia, A. Preston, L. D. Murphy, N. Thomson, D. E. Harris, M. T. Holden, C. M. Churcher, S. D. Bentley, K. L. Mungall, A. M. Cerdono-Tarraga, L. Temple, K. James, B. Harris, M. A. Quail, M. Achtman, R. Atkin, S. Baker, D. Basham, N. Bason, I. Cherevach, T. Chillingworth, M. Collins, A. Cronin, P. Davis, J. Doggett, T. Feltham, A. Goble, N. Hamlin, H. Hauser, S. Holroyd, K. Jagels, S. Leather, S. Moule, H. Norberczak, S. O'Neil, D. Ormond, C. Price, E. Rabinowitsch, S. Rutter, M. Sanders, D. Saunders, K. Seeger, S. Sharp, M. Simmonds, J. Skelton, R. Squares, S. Squares, K. Stevens, L. Unwin, S. Whitehead, B. G. Barrell, and D. J. Maskell. 2003. Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. *Nat. Genet.* **35**:32–40.
 63. Parkhill, J., B. W. Wren, N. R. Thomson, R. W. Titball, M. T. G. Holden, M. B. Prentice, M. Sebaihia, K. D. James, C. Churcher, K. L. Mungall, S. Baker, D. Basham, S. D. Bentley, K. Brooks, A. M. Cerdono-Tarraga, T. Chillingworth, A. Cronin, R. M. Davies, P. Davis, G. Dougan, T. Feltham, N. Hamlin, S. Holroyd, K. Jagels, A. V. Karlyshev, S. Leather, S. Moule, P. C. F. Oyston, M. Quail, K. Rutherford, M. Simmonds, J. Skelton, K. Stevens, S. Whitehead, and B. G. Barrell. 2001. Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature* **413**:523–527.
 64. Rainey, F., N. Weiss, H. Prauser, and E. Stackebrandt. 1994. Further evidence for the phylogenetic coherence of actinomycetes with Group B-peptidoglycan and evidence for the phylogenetic intermixing of the genera *Microbacterium* and *Aurobacterium* as determined by 16S rDNA analysis. *FEMS Microbiol. Lett.* **118**:135–140.
 65. Rokka, V. M., J. Laurila, A. Tauriainen, I. Laakso, J. Larkka, M. Metzler, and L. Pietila. 2005. Glycoalkaloid aglycone accumulations associated with infection by *Clavibacter michiganensis* ssp. *sepedonicus* in potato species *Solanum acaule* and *Solanum tuberosum* and their interspecific somatic hybrids. *Plant Cell Rep.* **23**:683–691.
 66. Rosenbluth, M., and E. Martinez-Romero. 2006. Bacterial endophytes and their interactions with hosts. *Mol. Plant-Microbe Interact.* **19**:827–837.
 67. Samac, D. A., R. J. Nix, and A. E. Oleson. 1998. Transmission frequency of *Clavibacter michiganensis* subsp. *insidiosus* to alfalfa seed and identification of the bacterium by PCR. *Plant Dis.* **82**:1362–1367.
 68. Sasseti, C. M., D. H. Boyd, and E. J. Rubin. 2003. Genes required for mycobacterial growth defined by high density mutagenesis. *Mol. Microbiol.* **48**:77–84.
 69. Schaad, N. W., Y. Berthier-Schaad, A. Sechler, and D. Knorr. 1999. Detection of *Clavibacter michiganensis* subsp. *sepedonicus* in potato tubers by BIO-PCR and an automated real-time fluorescence detection system. *Plant Dis.* **83**:1095–1100.
 70. Slack, S. A. 1987. Biology and ecology of *Corynebacterium sepedonicum*. *Am. Potato J.* **64**:665–701.

71. **Sonnhammer, E. L. L., and R. Durbin.** 1995. A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis (reprinted from *Gene Combis*, vol 167, pg GC1-GC10, 1996). *Gene* **167**:GC1-GC10.
72. **Spooner, D. M., K. McLean, G. Ramsay, R. Waugh, and G. J. Bryan.** 2005. A single domestication for potato based on multilocus amplified fragment length polymorphism genotyping. *Proc. Natl. Acad. Sci. USA* **102**:14694–14699.
73. **Stead, D.** 1999. Bacterial diseases of potato: relevance to *in vitro* potato seed production. *Potato Res.* **42**:449–456.
74. **Suzuki, K., M. Suzuki, J. Sasaki, Y. H. Park, and K. Komagata.** 1999. *Leifsonia* gen. nov., a genus for 2,4-diaminobutyric acid-containing actinomycetes to accommodate “*Corynebacterium aquaticum*” Leifson 1962 and *Clavibacter xyli* subsp. *cynodontis* Davis et al. 1984. *J. Gen. Appl. Microbiol.* **45**:253–262.
75. **Tyagi, J. S., and D. K. Saini.** 2004. Did the loss of two-component systems initiate pseudogene accumulation in *Mycobacterium leprae*? *Microbiology* **150**:4–7.
76. **van der Wolf, J. M., and J. van Beckhoven.** 2004. Factors affecting survival of *Clavibacter michiganensis* subsp. *sepedonicus* in water. *J. Phytopathol.* **152**:161–168.
77. **van der Wolf, J. M., J. van Beckhoven, A. Hukkanen, R. Karjalainen, and P. Muller.** 2005. Fate of *Clavibacter michiganensis* ssp. *sepedonicus*, the causal organism of bacterial ring rot of potato, in weeds and field crops. *J. Phytopathol.* **153**:358–365.
78. **Vidaver, A. K., and M. P. Starr.** 1981. Phytopathogenic coryneform and related bacteria, p. 1879–1887. *In* M. P. Starr, H. Stolp, H. G. Trüper, A. Balows, and H. G. Schlegel (ed.), *The prokaryotes: a handbook on habitats, isolation, and identification of bacteria*. Springer-Verlag, Berlin, Germany.
79. **Wagner, A.** 2006. Periodic extinctions of transposable elements in bacterial lineages: evidence from intragenomic variation in multiple genomes. *Mol. Biol. Evol.* **23**:723–733.
80. **Webb, D. C., H. Rosenberg, and G. B. Cox.** 1992. Mutational analysis of the *Escherichia coli* phosphate-specific transport system, a member of the traffic ATPase (or ABC) family of membrane transporters. A role for proline residues in transmembrane helices. *J. Biol. Chem.* **267**:24661–24668.
81. **Welch, R. A., V. Burland, G. Plunkett, P. Redford, P. Roesch, D. Rasko, E. L. Buckles, S. R. Liou, A. Boutin, J. Hackett, D. Stroud, G. F. Mayhew, D. J. Rose, S. Zhou, D. C. Schwartz, N. T. Perna, H. L. T. Mobley, M. S. Donnenberg, and F. R. Blattner.** 2002. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **99**:17020–17024.
82. **Westra, A. A. G., and S. A. Slack.** 1992. Isolation and characterization of extracellular polysaccharide of *Clavibacter michiganensis* subsp. *sepedonicus*. *Phytopathology* **82**:1193–1200.
83. **Wikstrom, N., V. Savolainen, and M. W. Chase.** 2001. Evolution of the angiosperms: calibrating the family tree. *Proc. R. Soc. Lond. Ser. B* **268**:2211–2220.
84. **Zgurskaya, H. I., L. I. Evtushenko, V. N. Akimov, and L. V. Kalakoutskii.** 1993. *Rathayibacter* gen. nov., including the species *Rathayibacter rathayi* comb. nov., *Rathayibacter tritici* comb. nov., *Rathayibacter iranicus* comb. nov., and six strains from annual grasses. *Int. J. Syst. Bacteriol.* **43**:143–149.
85. **Zinniel, D. K., P. Lambrecht, N. B. Harris, Z. Feng, D. Kuczmariski, P. Higley, C. A. Ishimaru, A. Arunakumari, R. G. Barletta, and A. K. Vidaver.** 2002. Isolation and characterization of endophytic colonizing bacteria from agronomic crops and prairie plants. *Appl. Environ. Microbiol.* **68**:2198–2208.