

Genetic diversity of ‘*Candidatus Liberibacter solanacearum*’ strains in the United States and Mexico revealed by simple sequence repeat markers

Hong Lin · Md Sajedul Islam · Yang Bai ·
Amin Wen · Suque Lan · Neil C. Gudmestad ·
Edwin L. Civerolo

Accepted: 7 September 2011 / Published online: 21 September 2011
© KNPV 2011

Abstract ‘*Candidatus Liberibacter solanacearum*’ is associated with the Zebra Chip (ZC) disorder of potatoes. A panel of eight simple sequence repeat (SSR) markers was developed and used to genetically characterize ‘*Ca. L. solanacearum*’ strains obtained from ZC-affected potato plants in the United States

and Mexico. The multilocus SSR markers in this study effectively differentiated genotypes and estimated genetic diversity of ‘*Ca. L. solanacearum*’ strains. Genotype assignment analyses identified two major lineages of ‘*Ca. L. solanacearum*’ in the North American populations while only one lineage type was identified in Mexican population. No clear genetic structure was found among haplotypes based on geographical proximity or host. The high resolution power of the SSR marker system developed in this study provides a useful tool for genotyping closely related strains and tracking sources of the pathogen. Genotype information combined with epidemiological data will advance knowledge of ZC disease and will facilitate development of effective disease management.

H. Lin · M. S. Islam · Y. Bai · S. Lan
Crop Diseases, Pests and Genetics research Unit,
USDA-ARS,
9611 S. Riverbend Avenue,
Parlier, CA 93648, USA

Y. Bai
Guangxi Citrus Research Institute,
Guilin, China

A. Wen · N. C. Gudmestad
Department of Plant Pathology,
North Dakota State University,
Fargo, ND 58108, USA

S. Lan
Institute of Cereal and Oil Crops,
Hebei Academy of Agricultural and Forestry Sciences,
Shijiazhuang, China

H. Lin (✉)
USDA-ARS-Crop Diseases,
Pests and Genetics Research Unit, Parlier,
California, CA 93648, USA
e-mail: Hong.Lin@ars.usda.gov

E. L. Civerolo
USDA-ARS,
9611 S. Riverbend Avenue,
Parlier, CA 93648, USA

Keywords Zebra chip (ZC) · ‘*Candidatus Liberibacter solanacearum*’ · Simple sequence repeat (SSR) markers · Genetic diversity

Introduction

Zebra chip (ZC) of potato (*Solanum tuberosum*) has emerged as a disease with significant economic impact on the potato industry in the United States, Mexico, and New Zealand. The disease is characterized by discolouration or striped pattern of necrosis of the tubers from infected plants. The fried chips from these infected tubers are commercially unacceptable (Gudmestad and Secor 2007). ZC was first reported in Mexico in 1994 (Munyaneza et al. 2007, 2009). By

2004, this disease was confirmed in the United States and caused significant economic losses in the potato-growing regions of Texas as well as several states including Arizona, California, Colorado, Kansas, Nebraska, and New Mexico (Gudmestad and Secor 2007; Munyaneza et al. 2009). It was also reported in New Zealand in 2008 (Liefting et al. 2008). Recently, molecular and phylogenetic evidence based on cloned 16S rRNA gene sequences from ZC symptomatic potato tubers and the potato psyllid (*Bactericera cockerelli* Sulc) confirmed that ZC is associated with the presence of a putative phloem-limited alpha-*Proteobacterium* called ‘*Candidatus Liberibacter solanacearum*’ (syn. ‘*Candidatus Liberibacter psyllaurosus*’) (Hansen et al. 2008; Liefting et al. 2009; Lin et al. 2009; Secor et al. 2009; Wen et al. 2009). ‘*Ca. L. solanacearum*’ was also associated with disease in other solanaceous plants such as in pepper (*Capsicum sp.*), tamarillo (*S. betaceum*), tomato (*S. lycopersicum*) (Liefting et al. 2009) and in a non-solanaceous host, carrot (*Daucus carota*) (Munyaneza et al. 2010). It also occurs in solanaceous weeds, such as wolfberry (*Lycium barbarum*), black nightshade (*S. ptychanthum*) and silverleaf nightshade (*S. elaeagnifolium*) (Wen et al. 2009). ‘*Ca. L. solanacearum*’ belongs to the *Rhizobiaceae* family and is closely related to four other ‘*Candidatus Liberibacter*’ species. Three of them, ‘*Ca. L. asiaticus*’ (Las), ‘*Ca. L. africanus*’ (Laf) and ‘*Ca. L. americanus*’ (Lam), are associated with citrus huanglongbing (HLB) (Bové 2006), and another one, ‘*Ca. L. europaeus*’ is found as non-pathogenic in pear plants (Raddadi et al. 2011).

While ‘*Ca. L. solanacearum*’ is confirmed to be associated with ZC, information regarding genetic diversity, structure and adaptation, as well as the epidemiological relationships of this bacterium across different geographical regions and hosts is limited. The recent completion of the ‘*Ca. L. solanacearum*’ genome sequence (Lin et al. 2011) has allowed for the development of molecular genetic markers to facilitate genotyping and genetic analyses of ‘*Ca. L. solanacearum*’.

The simple sequence repeat (SSR) markers, also known as variable number tandem repeat or microsatellite markers, are useful for genotyping and the genetic study of bacteria. The multilocus SSR markers have been used to differentiate strains of *Mycobacterium avium* (Amonsin et al. 2004), and for molecular epidemiology and genetic studies in *M. tuberculosis* (Sola et al.

2003). This marker system possesses adequate polymorphic power to differentiate closely related bacterial strains such as ‘*Ca. L. asiaticus*’ (Kato et al. 2011) and *Xylella fastidiosa* (Coletta-Filho et al. 2001; Lin et al. 2005; Almeida et al. 2008; Coletta-Filho et al. 2011).

In this study, we developed a panel of eight SSR markers and used this marker system for analyzing the genetic diversity and genetic relationships of ‘*Ca. L. solanacearum*’ across different geographical locations and hosts in Northern America, including the United States and Mexico.

Materials and methods

Plant materials

From 2007 to 2009, potato plants with symptoms of ZC, psyllid yellows (PY) and haywire (HW) were collected from commercial potato fields in Texas, Nebraska, Kansas, Colorado and California, and shipped on ice to North Dakota State University for processing. ZC symptomatic potato plants and tubers were also collected from commercial potato fields in Mexico (USDA APHIS permit: PDEP-07-00193) in 2008. In addition, potato plants displaying HW symptoms originating from seed pieces with ZC symptoms (designated as seedborne ZC, SBZC) were collected in O’Neill, Nebraska and Dalhart, Texas (Table 1). Plants showing phytoplasma-like foliar symptoms with tubers displaying ZC symptoms were also collected (designated as unknown, UK). Detailed sample information is provided in Table 1.

DNA extraction and PCR confirmation

Petiole, mid-vein, above-ground stem, stolon, and tuber tissue was ground in liquid nitrogen with a mortar and pestle. Ground plant tissues were freeze-dried, and stored at -20°C for future analyses. Sixty-two freeze-dried tissue samples, previously determined to be ‘*Ca. L. solanacearum*’ positive (Wen et al. 2009), were selected based on different geographical locations, host cultivars (variety) and symptom types. Twenty mg of freeze-dried tissues from each sample was used for DNA extraction using DNeasy Plant Mini Kit (Qiagen, Valencia, CA). DNA was quantified using Nanodrop (ThermoScientific). Sixty two DNA samples were

Table 1 Sampling information of '*Candidatus Liberibacter solanacearum*' strains collected from different geographic locations in the United States and Mexico

Sample ID	Sample location	Host cultivar	Collection date	Symptom type ^a
Iso1	Dalhart, Texas	FL 1291	7/12/2007	ZC
Iso2	Dalhart, Texas	FL 1867	7/12/2007	ZC
Iso3	Dalhart, Texas	FL 2048	7/12/2007	ZC
Iso4	Dalhart, Texas	FL 2101	8/3/2007	ZC
Iso5	Dalhart, Texas	FL1867	6/16/2009	SBZC
Iso6	Dalhart, Texas	FL1867	6/16/2009	SBZC
Iso7	Dalhart, Texas	FL1867	6/16/2009	SBZC
Iso8	Dalhart, Texas	FL1867	7/8/2009	ZC
Iso9	Dalhart, Texas	FL1867	7/21/2009	ZC
Iso10	Dalhart, Texas	FL1867	8/9/2009	ZC
Iso11	Dalhart, Texas	FL1922	6/30/2009	HW
Iso12	Dalhart, Texas	Russet Norkotah	7/24/2008	ZC
Iso13	Dalhart, Texas	Snowden	7/12/2007	ZC
Iso14	Dalhart, Texas	Unknown	7/24/2008	ZC
Iso15	Dalhart, Texas	Unknown	7/24/2008	ZC
Iso16	Olton, Texas	FL1867	6/2/2009	SBZC
Iso17	Olton, Texas	FL1867	6/16/2009	ZC
Iso18	Olton, Texas	FL1867	6/16/2009	ZC
Iso19	Olton, Texas	FL1867	6/25/2009	ZC
Iso20	Olton, Texas	FL1867	6/25/2009	ZC
Iso21	Olton, Texas	Russet Norkotah	6/9/2008	ZC
Iso22	Lubbock, Texas	Unknown	7/30/2009	UK
Iso23	Pearsall, Texas	unknown	4/15/2009	ZC
Iso24	Pearsall, Texas	unknown	4/28/2009	ZC
Iso25	Pearsall, Texas	FL1867	5/19/2009	ZC
Iso26	O'Neill, Nebraska	Russet Norkotah Texas Strain 296	6/19/2008	SBZC
Iso27	O'Neill, Nebraska	Russet Norkotah Texas Strain 296	6/19/2008	SBZC
Iso28	O'Neill, Nebraska	Russet Norkotah, WY seed	6/19/2008	SBZC
Iso29	Cody, Nebraska	FL1833	6/22/2009	HW
Iso30	Cody, Nebraska	FL1867	6/22/2009	HW
Iso31	Cody, Nebraska	FL2084	6/22/2009	HW
Iso32	Cody, Nebraska	FL1867, FG-2	9/22/2008	ZC
Iso33	Cody, Nebraska	FL1867, FG-3	9/30/2008	ZC
Iso34	Cody, Nebraska	FL2048.	8/11/2009	ZC
Iso35	Cody, Nebraska	Unknown	8/18/2008	ZC
Iso36	Linton, Cody, Nebraska	Burbank, FG-2	9/25/2008	PY
Iso37	Linton, Cody, Nebraska	Burbank, FG-2	9/25/2008	ZC
Iso38	Linton, Cody, Nebraska	Blazer, FG-2	9/25/2008	ZC
Iso39	Columbus, Nebraska	FL1867	1/17/2009	ZC
Iso40	KRNY, Nebraska	Sierra Gold	8/18/2008	PY
Iso41	Minden, Nebraska	FL2053	6/2/2009	SBZC
Iso42	Minden, Nebraska	FL2053	6/2/2009	SBZC
Iso43	Garden City, Kansas	Atlantic	8/6/2008	ZC
Iso44	Garden City, Kansas	Russet Norkotah	8/6/2008	ZC

Table 1 (continued)

Sample ID	Sample location	Host cultivar	Collection date	Symptom type ^a
Iso45	Garden City, Kansas	FL1867	8/6/2008	ZC
Iso46	Southwestern, Nebraska	Russet Norkotah	7/23/2008	ZC
Iso47	Southwestern, Nebraska	Russet Norkotah Strain Colorado 3	7/23/2008	ZC
Iso48	Southwestern, Nebraska	Russet Norkotah	7/23/2008	ZC
Iso49	Southwestern, Nebraska	Russet Norkotah Strain Colorado 3	7/23/2008	ZC
Iso50	Wray, Colorado	Russet Norkotah	7/23/2008	ZC
Iso51	Wray, Colorado	Russet Norkotah	7/23/2008	SBZC
Iso52	CSS Farms, California	FL795, FG-2	5/19/2009	PY
Iso53	Bakersfield, California	Unknown	6/10/2009	ZC
Iso54	Bakersfield, California	Unknown	6/10/2009	ZC
Iso55	Mexico	Unknown	8/12/2008	ZC
Iso56	Mexico	Unknown	8/27/2008	ZC
Iso57	Mexico	Unknown	8/27/2008	ZC
Iso58	Mexico	Unknown	8/27/2008	ZC
Iso59	Mexico	Unknown	8/27/2008	ZC
Iso60	Mexico	Unknown	8/27/2008	ZC
Iso61	Mexico	Unknown	8/27/2008	ZC
Iso62	Mexico	Unknown	9/26/2008	ZC

^a ZC symptoms include necrotic flecking and medullary ray discoloration in afflicted tubers. Foliar symptoms of ZC include swollen nodes, axillary bud development, reddening of the foliage, leaf scorching, and eventual plant death. PY foliar symptoms include pronounced yellowing of the foliage and resulting in a dwarf rosette appearance. SBZC symptoms include haywire on the foliage and medullary ray discoloration in seed piece tuber; HW symptoms are shortening of the internodes, haywire; UK infers phytoplasma-like symptoms on foliage.

verified using a quantitative real time PCR-based '*Ca. L. solanacearum*' specific primers (Lin et al. 2011).

SSR marker development and genotyping

A genome-wide sequence search was conducted to identify simple sequence repeat (SSR) loci using complete '*Ca. L. solanacearum*' genome sequences (GenBank accession # CP002371). The search included various perfect, imperfect and compound types of simple repeat motifs. The criteria to identify SSR loci for designing primers were: 1) type of repeat loci containing at least five or more of SSR motifs, and 2) at least one copy of each locus per genome. The results led to the identification of 85 loci that met the criteria and were potentially useful for designing SSR markers. To ensure that sequence loci selected for SSR marker development were unique to '*Ca. L. solanacearum*', *in-silico* sequencing analyses were carried out to compare candidate loci against all the available microbe sequences with BLASTn analysis

in the NCBI databases. The final set of SSR loci was selected for SSR primer design using Molecular Beacon software (version 7.0). Forty designed SSR primer pairs were then screened for the power of polymorphic detection. This screening procedure identified eight SSR primer pairs useful for this study (Table 2).

To facilitate high through-put genotyping assay, the 5' end of each of the SSR forward primers was labeled with fluorescent dyes with FAM, NED, VIC or PET, respectively (Table 2) (Applied Biosystems, Foster City, CA). PCR was performed in 20 μ l containing 2 μ l of 10X reaction buffer, 1.0 mM MgCl₂, 0.2 mM dNTPs, 0.25 U *Taq* polymerase, 2.5 pmole of each of SSR primer pairs and 2 μ l of diluted DNA sample. Temperature settings for PCR were as follows: initial denaturation at 95°C for 6 min, followed by 35 cycles of 95°C for 30 s, 58°C for 30 s and 72°C for 1 min, with a 72°C final extension step for 7 min. 5 μ l of amplification product was electrophoresed by a 1.5% agarose gel to confirm

Table 2 Characteristics of eight simple sequence repeat (SSR) markers developed from the ‘*Candidatus* *Liberibacter solanacearum*’ genome sequence

Locus	Primer sequence	Repeat	Location in genome	ORF	Size range (bp)	Na	H
Lso-SSR-1F	VIC- 5'-TTATTTTGAGATGTTTGTAAATG-3'	(TTTATT) ₅	20984–21161	Phosphatidylserine synthase	169–237	4	0.465
Lso-SSR-1R	5'-TATTAATCAITCTAATGCCTAATTCG-3'						
Lso-SSR-2F	VIC- 5'-AAACGCTACTACAACCTTGAC-3'	(TTGAD) ₇	196497–196704	Hypothetical protein	148–237	8	0.713
Lso-SSR-2R	5'-AGAGAACGGACATAGAACACCC-3'						
Lso-SSR-3F	FAM- 5'-CTCCGACTGATATACGAATCC-3'	(AATCTG) ₆	331770–332009	Translation initiation factor 1 (IF-1)	214–257	4	0.643
Lso-SSR-3R	5'-TTCTGAATTAATGCCGAATGC-3'						
Lso-SSR-4F	NED- 5'-TCCAAGGAAGATAGTAGCC-3'	(GATGAT) ₅	352719–352868	Hypothetical protein	128–147	3	0.562
Lso-SSR-4R	5'-TTAAACGCTAACTTGACTCTCC-3'						
Lso-SSR-6F	PET-5'-CCAAACCAATCCITTAACC-3'	(AGATGG) ₇	642572–642812	Rare lipoprotein A	214–247	4	0.621
Lso-SSR-6R	5'-ACAGCTATTCCTATTCCTACC-3'						
Lso-SSR-7F	NED-5'-AGCCGTTAGTACAATGATTAAG-3'	(AAAAAT) ₁₃	709412–709698	Hypothetical protein	240–280	5	0.628
Lso-SSR-7R	5'-CGCGAAATAATATTACATCCG-3'						
Lso-SSR-8F	VIC- 5'-GTTTGATTAICTTTAICTTGGGTTG-3'	(CTT) ₆	854298–854848	Hypothetical protein	254–263	2	0.422
Lso-SSR-8R	5'-GCTAGAAATGATACAACCTGATGG-3'						
Lso-SSR-9F	FAM- 5'-AATAGCGTCAGCGTGTTC-3'	(GAAGAACCA) ₆	1248159–1248311	Hypothetical protein	143–162	3	0.610
Lso-SSR-9R	5'-ACGGATGGAAGAGATGAGG-3'						

Na number of alleles; H haploid genetic diversity

the successful SSR products. For multiplex analysis, 2 μ l of each SSR product labeled with different fluorescent dye was pooled. From 8 μ l of pooled product, 2.5 μ l was mixed with 0.25 μ l of GeneScan-500 Liz molecular size standard (Applied Biosystems Cat #4322682A) and 7.25 μ l of Hi-Di Formamide (Applied Biosystems Cat. #4311320). The mixture of SSR products was then loaded onto a 3130 Genetic Analyzer (Applied Biosystems, Foster City, CA) equipped with the 36-cm 16-capillary array filled with POP-6 polymer (Applied Biosystems, Foster City, CA). Data acquisition and fragment size determinations were carried out by GeneMapper v4.0 software (Applied Biosystems, Foster City, CA).

Data analysis

Allelic data obtained from eight SSR loci were combined, and genotypes (haplotypes) were identified for each strain. Since '*Ca. L. solanacearum*' is primarily a clonally reproduced organism, a clonal lineage is expected within the group of strains. Therefore, a clonal-corrected (a single copy of each multilocus genotype) data set was used to estimate genetic diversity, evaluate genetic relationships and identify the genetic structures of '*Ca. L. solanacearum*' strains from different geographical locations and hosts.

GenAlEx Version 6.3 (Peakall and Smouse 2006) was used to calculate the average number of alleles (N_a), average number of effective alleles (N_e) and haploid genetic diversity (H) at each locus, and across all loci for the groups of strains at different geographical locations. A global test (Fisher's method) implemented in GENEPOP web version 4.0.10 (Raymond and Rousset 1995) was used to test for the genotyping linkage disequilibrium between each of the pairs of loci across the clonal corrected data set.

To evaluate genetic relationships among '*Ca. L. solanacearum*' strains, software START2 was used (Jolley et al. 2001). The lineage assignment was determined by UPGMA dendrogram (Prager and Wilson 1978) constructed from genotypic profiles of multilocus SSR data.

The program STRUCTURE 2.3.1 (Pritchard et al. 2000) was used as a clustering approach based on the analysis of Bayesian modelling. The analysis is based on assigning all individuals to specified numbers of clusters (K). Individuals in the sample are assigned probabilistically to these K clusters, or jointly to two

or more clusters if their genotypes indicate that they are admixed. STRUCTURE analyses in this study were performed for the group of strains from different geographical areas as well as from different host cultivars without consideration of their region or host of origin. To estimate the number of clusters (K), we set 20 independent runs of $K=1-10$. For each run, a burn-in period of 25,000 iterations was used followed by a run length of 50,000 Markov chain Monte Carlo iterations and a model with correlated allele frequencies and admixture among the group of strains. The number of clusters that best represented the observed data was determined by maximizing the estimated ln likelihood of the data for different values of K , and the ΔK index which is based on the rate of change in the ln likelihood of the data between successive K (1–10) (Evanno et al. 2005). The optimal probabilities for all individuals were estimated from 20 replicate runs at $K=2$ with permutation analysis using CLUMPP version 1.1.2 (Jakobsson and Rosenberg 2007), and the output of genetic clustering was visualized using software DISTRUCT version 1.1 (Rosenberg 2004).

Results

Characteristics of SSR loci and PCR amplification

Eight SSR markers capable of detecting polymorphism in '*Ca. L. solanacearum*' strains were selected for this study. Sixty-two '*Ca. L. solanacearum*' strains from different geographic locations in North America were amplified by 8 SSR markers. The number of alleles observed ranged from 2 to 8 per locus. Locus-based haploid genetic diversity ranged from 0.422 to 0.713 (Table 2). In clonal corrected data set, no genotypic linkage disequilibrium was observed between any pair of the loci across all samples ($P>0.01$).

Genotype and genetic diversity

Based on the combination of allelic data obtained from 8 SSR loci, 33 genotypes (haplotypes) were identified in the total of 62 strains, of which 26 haplotypes were found to be unique, and remaining 36 strains shared a total of 7 haplotypes. Overlapping of identical haplotypes was found in different geographical locations, such as haplotypes 3, 6 and 13 occurring among the strains from Texas and

Nebraska. Overlapping haplotypes (23 and 27) were also found among some strains from Texas, Kansas and Mexico (Fig. 1). When haplotype distributions were compared with respect to host cultivar, no identical haplotypes appeared to be restricted to any particular type of host cultivar. Though some identical

haplotypes (6, 23 and 27) overlapped within a potato cultivar (FL1867) from different locations, these haplotypes also overlapped with different host cultivars (Fig. 1).

The average number of alleles and the number of effective alleles for the group of strains at different

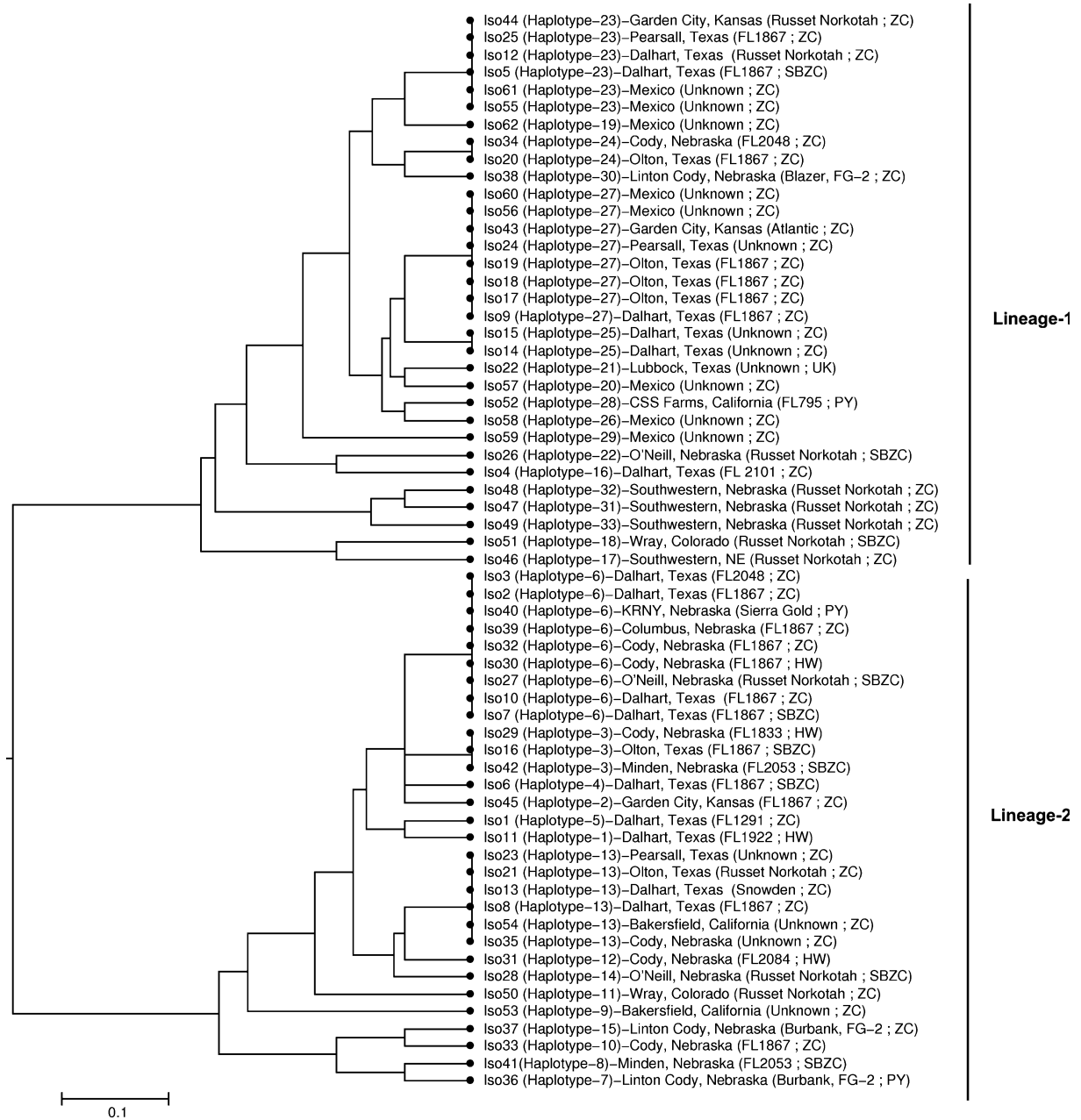


Fig. 1 Lineage assignment of ‘*Candidatus Liberibacter solanacearum*’ strains in the United States and Mexico as determined by UPGMA-generated dendrogram. Each label

contains sample ID, haplotype, geographical location, host cultivar and disease symptom type

geographic areas ranged from 1.4 to 2.6 (overall, 4.1) and 1.3 to 2.1 (overall, 2.5), respectively. Haploid genetic diversity in different groups of strains ranged from 0.167 (Minden, Nebraska area) to 0.550 (Wray, Colorado area). Overall haploid genetic diversity of ‘*Ca. L. solanacearum*’ in the North America (United States and Mexico) populations at eight SSR loci was 0.583 (Table 3).

Genetic relationship and structure

The UPGMA-generated dendrogram indicated two major lineages among ‘*Ca. L. solanacearum*’ strains analyzed in this study (Fig. 1). Strains from Texas were found in both lineages. However, all of the strains from Mexico, and the majority of the strains from Nebraska and California were distributed in lineage 2 whereas the majority of the strains from Wray, Colorado area and Kansas were found in lineage 1 (Fig. 1).

To further analyze the possible genetic structures of ‘*Ca. L. solanacearum*’ strains, a clustering approach was used to infer the most likely group of origin. The STRUCTURE analysis, without prior information of geography or host cultivar, placed all strains into two major clusters (Fig. 2a and b), which is consistent with the lineage assignment as determined by UPGMA (Fig. 1). Distributions of clusters showed similar patterns as observed in lineage assignment.

Discussion

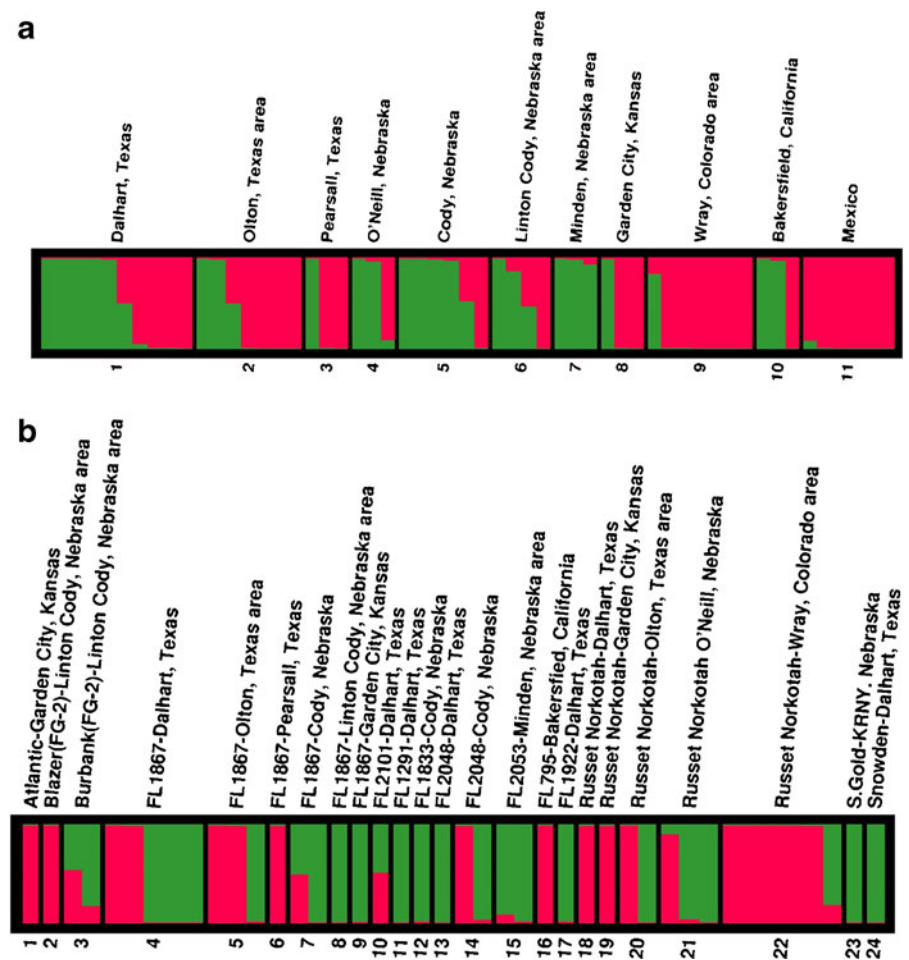
Genetic diversity in microbial populations has been investigated by multilocus molecular markers including Random Amplified Polymorphic DNA (RAPD) (Chelossia et al. 2004), Restriction Fragment Length Polymorphism (RFLP) (Mushegian et al. 2011) and Amplified Fragment Length polymorphism (AFLP) (Wittwer et al. 2005). However, these molecular marker techniques cannot be simply applied for non-culturable bacteria because they require the use of pure target genomic DNA. Due to the fastidious nature of ‘*Ca. L. solanacearum*’, this bacterium is non-culturable at this time. Therefore, pure genomic DNA for this prokaryote is not available. With the completion of genome sequence of ‘*Ca. L. solanacearum*’ (Lin et al. 2011), it is now possible to gain insights into the genome to identify and develop ‘*Ca. L. solanacearum*’ species specific molecular markers for disease diagnosis and for genetic characterization of ‘*Ca. L. solanacearum*’ populations. In this study, a panel of eight polymorphic SSR markers was developed based on ‘*Ca. L. solanacearum*’ genome sequence data. We developed this discriminatory marker system to conduct genetic analysis of ‘*Ca. L. solanacearum*’ strains collected from various potato cultivation regions in North America. Polymorphism analyses at the SSR loci indicate that as many as eight

Table 3 Descriptive statistics and diversity of ‘*Candidatus Liberibacter solanacearum*’ at eight SSR loci across 62 strains from different geographical areas in the United States and Mexico

Geographical areas	Total number of strains	Number of strains in clonal corrected data	Na	Ne	H
Dalhart, Texas	15	9	2.3	2.0	0.497
Olton, Texas area (Olton; Springlake; Lubbock)	7	5	2.3	2.1	0.500
Pearsall, Texas	3	3	2.0	1.8	0.444
O'Neill, Nebraska	3	3	2.0	1.9	0.417
Cody, Nebraska	7	6	2.1	1.7	0.403
Linton Cody, Nebraska area (Linton Cody; Columbus)	4	4	2.3	1.9	0.453
Minden, Nebraska area (Minden; KRNY)	3	3	1.4	1.3	0.167
Garden City, Kansas	3	3	2.0	1.8	0.444
Wray, Colorado area (Wray, Colorado; Southwestern, Nebraska)	6	6	2.6	2.1	0.550
Bakersfield, California	3	3	2.3	2.1	0.500
Mexico	8	6	1.8	1.4	0.215
Overall (United States and Mexico)	62	51	4.1	2.5	0.583

Na number of alleles per locus; Ne number of effective alleles per locus; H haploid genetic diversity; Clonal-corrected data (a single copy of each multilocus genotype within a group of strains)

Fig. 2 STRUCTURE analysis showing individual assignments of ‘*Candidatus Liberibacter solanacearum*’ strains: **a.** based on different geographical areas in the United States (1–10) and Mexico (11), where two clusters were identified ($K=2$) (red: Cluster 1 and green: Cluster 2), and **b.** based on different host cultivars (1–24), where two clusters were found ($K=2$) (red: Cluster 1 and green: Cluster 2). *Black lines* within the squares distinguish group of strains



unique alleles were detected at some loci and with high haploid genetic diversity (0.713) (Table 2). These multilocus markers demonstrate their practical utility for genotyping, genetic diversity and structure analyses of ‘*Ca. L. solanacearum*’.

Multilocus SSR analysis in this study showed high level of genetic diversity in ‘*Ca. L. solanacearum*’ in North America (United States and Mexico). Genetic analyses indicated that the occurrence of identical haplotypes were not restricted to any particular geographical area or host cultivar. That is, overlapping of identical haplotypes was observed among different geographical locations including Texas and Nebraska, and among Texas, Kansas and Mexico. Likewise, no haplotypes were found to be restricted to particular potato cultivar or solanaceous plant. In contrast to a natural population, in agricultural practice, it is expected that frequent exchange of plant materials could facilitate the movement of

ZC-affected materials which could largely account for such non-patterned geographical distribution among identical haplotypes.

In general, the genetic diversity was higher among ‘*Ca. L. solanacearum*’ strains across the different regions of United States (except Minden, Nebraska area) than among those from Mexico. However, the small numbers of strains from Mexico in this study are insufficient to make an accurate estimation of genetic diversity. The main mode of gene transfer among ‘*Ca. L. solanacearum*’ in agricultural ecosystems is not known. However, the large numbers of unique haplotypes identified and the high level of genetic diversity among ‘*Ca. L. solanacearum*’ strains observed in this study suggest that this pathogen has much higher diversity in North America than we expected. A high level of genetic diversity was found in the closely related citrus HLB-associated ‘*Ca. L. asiaticus*’ using similar DNA markers (VNTR) (Kato

et al. 2011). Similar results were also reported based on SSR analyses of another phytopathogenic bacterium, *Xylella fastidiosa* (Coletta-Filho et al. 2011).

The efficiency of using SSR markers over other conserved gene-based markers in understanding genetic differentiations and determining the population genetic structures for plant pathogens at the local level over a short divergence of time has been described (Almeida et al. 2008; Yuan et al. 2010). The multilocus SSR marker system we developed in this study demonstrates adequate discriminatory power for genotyping and genetic analysis of ‘*Ca. L. solanacearum*’ strains. Lineage assignment from genotype profiles and structure analysis from correlated allele frequencies consistently placed all of the ‘*Ca. L. solanacearum*’ strains into two major lineages or clusters regardless of host or geographic origin (Figs. 1 and 2). The results from this study agree with the studies of ‘*Ca. L. solanacearum*’ based on the 16 rRNA, 16S/23S ISR gene sequences of which strains were clustered into clade 1 and clade 2 (Wen et al. 2009). Similar results were recently reported based on SNPs analyses of 16S rRNA, 16s/23S ISR, and 50S rplJ and rplL ribosomal gene regions (Nelson et al. 2011) where two clusters of haplotypes (A and B) were observed among strains from different regions in the United States. Genetic analyses in this study did not reveal any distinct genetic structure among ‘*Ca. L. solanacearum*’ strains related to the geographical proximity or to the host cultivars. Strains throughout Texas were distributed primarily in two lineages or clusters. There was some grouping among the strains in some regions, such as in Mexico, Colorado, Nebraska, Kansas, California; however, overlapping of haplotypes and clusters among these regions was also observed (Figs. 1 and 2a).

Identification of ZC-associated ‘*Ca. L. solanacearum*’ sparked research interest in understanding the origin and the evolutionary relationships of this bacterium with other *Liberibacter* species. Previously, all three known species of ‘*Ca. Liberibacter*’ (Las, Laf, and Lam) were associated with citrus HLB. A phylogenetic tree constructed using the neighbour joining method clustered all three HLB-associated ‘*Ca. Liberibacter*’ species along with ‘*Ca. L. solanacearum*’ into a monophyletic group based on 16S rRNA gene sequences (Lin et al. 2009). Interestingly, ‘*Ca. L. solanacearum*’ was positioned at a node proximal to the basal node shared by other three ‘*Ca. Liberibacter*’ species, inferring early

divergence of Lam followed by ‘*Ca. L. solanacearum*’ (Lin et al. 2009). Shortly after the report of ‘*Ca. L. solanacearum*’ in New Zealand (Liefing et al. 2008), a new strain of ‘*Ca. L. solanacearum*’ that is associated with carrot, a non-solanaceous crop was identified outside of North, Central America and New Zealand (Munyaneza et al. 2010). This was followed by another report of non-pathogenic *Liberibacter* species named as ‘*Ca. L. europaeus*’ naturally vectored by the pear psyllid pest *Cacopsylla pyri* (L.) in Italy (Raddadi et al. 2011). Based on these research findings, it is reasonable to assume that ‘*Ca. Liberibacter*’ species have undergone relatively rapid evolutionary divergence, speciation and adaptation. Genetic diversity of ‘*Ca. L. solanacearum*’ in agricultural ecosystems is probably much larger than previously expected. The two lineages identified in North America populations are likely derived from two origins that were descended from a common ancestor. Whereas genetic analysis differentiates ‘*Ca. L. solanacearum*’ into two lineages, the biological implication of evolutionary divergence is not known.

The multilocus SSR marker system developed in this study is a prominent genotyping method to resolve closely related strains of ‘*Ca. L. solanacearum*’. Since pathogenic bacteria exist as populations, the individuals of which could exhibit varying degrees of virulence. SSR genotyping in combination with population genetic and epidemiological studies can provide critical insights into the sources and spread of ZC diseases. In addition to improving our understanding of the spatial and temporal dynamics of the disease, these insights are helpful in designing effective integrated pest and disease management for ZC problem.

Acknowledgements This work was supported by 2009 USDA National Institute of Food and Agriculture (NIFA) Specialty Crops Research Initiative (SCRI). Trade names or commercial products in this publication are mentioned solely for the purpose of providing specific information and does not imply recommendation or endorsement by the United States Department of Agriculture.

References

- Almeida, R. P. P., Nascimento, F. E., Chau, J., Prado, S. S., Tsai, C. W., Lopes, S. A., et al. (2008). Genetic structure and biology of *Xylella fastidiosa* strains causing disease in citrus and coffee in Brazil. *Applied and Environmental Microbiology*, 74, 3690–3701.

- Amonsin, A., Li, L. L., Zhang, O., Bannantine, J. P., Motiwala, A. S., Sreevatsan, S., et al. (2004). Multilocus short sequence repeat sequencing approach for differentiating among *Mycobacterium avium* subsp. *paratuberculosis* strains. *Journal of Clinical Microbiology*, *42*, 1694–1702.
- Bové, J. M. (2006). Huanglongbing: a destructive, newly-emerging, century-old disease of citrus. *Journal of Plant Pathology*, *88*, 7–37.
- Chelossi, E., Milanese, M., Milano, A., Pronzato, R., & Riccardi, G. (2004). Characterisation and antimicrobial activity of epibiotic bacteria from *Petrosia ficiformis* (Porifera, Demospongiae). *Journal of Experimental Marine Biology and Ecology*, *309*, 21–33.
- Coletta-Filho, H. D., Takita, M. A., De Souza, A. A., Aguilard-Vildoso, C. I., & Machado, M. A. (2001). Differentiation of strains of *Xylella fastidiosa* by a variable number of tandem repeat analysis. *Applied and Environmental Microbiology*, *67*, 4091–4095.
- Coletta-Filho, H. D., Bittleston, L. S., & Almeida, R. P. P. (2011). Spatial genetic structure of a vector-borne generalist pathogen. *Applied and Environmental Microbiology*, *77*, 2596–2601.
- Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, *14*, 2611–2620.
- Gudmestad, N. C., & Secor, G. A. (2007). Zebra chip: a new disease of potato. *Nebraska Potato Eyes*, *19*, 1–4.
- Hansen, A. K., Trumble, J. T., Stouthamer, R., & Paine, T. D. (2008). A new huanglongbing species, “*Candidatus liberibacter psyllauros*” found to infect tomato and potato, is vectored by the psyllid *Bactericera cockerelli* (sulc). *Applied and Environmental Microbiology*, *74*, 5862–5865.
- Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, *23*, 1801–1806.
- Jolley, K. A., Feil, E. J., Chan, M.-S., & Maiden, M. C. J. (2001). Sequence type analysis and recombinational tests (START). *Bioinformatics*, *17*, 1230–1231.
- Katoh, H., Subandiyah, S., Tomimura, K., Okuda, M., Su, H. J., & Iwanami, T. (2011). Differentiation of “*Candidatus liberibacter asiaticus*” isolates by variable-number tandem-repeat analysis. *Applied and Environmental Microbiology*, *77*, 1910–1917.
- Liefting, L. W., Perez-Egusquiza, Z. C., Clover, G. R. G., & Anderson, J. A. D. (2008). A new ‘*Candidatus liberibacter*’ species in *Solanum tuberosum* in New Zealand. *Plant Disease*, *92*, 1474.
- Liefting, L. W., Sutherland, P. W., Ward, L. I., Paice, K. L., Weir, B. S., & Clover, G. R. G. (2009). A new ‘*Candidatus liberibacter*’ species associated with diseases of solanaceous crops. *Plant Disease*, *93*, 208–214.
- Lin, H., Civerolo, E. L., Hu, R., Barros, S., Francis, M., & Walker, M. A. (2005). Multilocus simple sequence repeat markers for differentiating strains and evaluating genetic diversity of *Xylella fastidiosa*. *Applied and Environmental Microbiology*, *71*, 4888–4892.
- Lin, H., Doddapaneni, H., Munyaneza, J. E., Civerolo, E. L., Sengoda, V. G., Buchman, J. L., et al. (2009). Molecular characterization and phylogenetic analysis of 16s rRNA from a new “*Candidatus liberibacter*” strain associated with zebra chip disease of potato (*Solanum tuberosum* L.) and the potato psyllid (*Bactericera cockerelli* Sulc). *Journal of Plant Pathology*, *91*, 215–219.
- Lin, H., Lou, B. H., Glynn, J. M., Doddapaneni, H., Civerolo, E. L., Chen, C. W., et al. (2011). The complete genome sequence of ‘*Candidatus liberibacter solanacearum*’, the bacterium associated with potato zebra chip disease. *PLoS One*, *6*, e19135.
- Munyaneza, J. E., Crosslin, J. M., & Upton, J. E. (2007). Association of *Bactericera cockerelli* (Homoptera: Psyllidae) with “zebra chip,” a new potato disease in southwestern United States and Mexico. *Journal of Economic Entomology*, *100*, 656–63.
- Munyaneza, J. E., Sengoda, V. G., Crosslin, J. M., De La Rosa-Lozano, G., & Sanchez, A. (2009). First report of ‘*Candidatus liberibacter psyllauros*’ in potato tubers with zebra chip disease in Mexico. *Plant Disease*, *93*, 552.
- Munyaneza, J., Fisher, T., Sengoda, V., Garczynski, S., Nissinen, A., & Lemmetty, A. (2010). First report of “*Candidatus liberibacter solanacearum*” associated with psyllid-affected carrots in Europe. *Plant Disease*, *94*, 639.
- Mushegian, A. A., Peterson, C. N., Baker, C. C., & Pringle, A. (2011). Bacterial diversity across individual lichens: thallus centers house richer, more uniform communities than recently formed edges. *Applied and Environmental Microbiology*. doi:10.1128/AEM.02850-10.
- Nelson, W. R., Fisher, T. W., & Munyaneza, J. E. (2011). Haplotypes of “*Candidatus liberibacter solanacearum*” suggest long-standing separation. *European Journal of Plant Pathology*, *130*, 5–12.
- Peakall, R., & Smouse, P. (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, *6*, 288–295.
- Prager, E. M., & Wilson, A. C. (1978). Construction of phylogenetic trees for proteins and nucleic acids: empirical evaluation of alternative matrix methods. *Journal of Molecular Evolution*, *11*, 129–142.
- Pritchard, J., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*, 945–959.
- Raddadi, N., Gonella, E., Camerota, C., Pizzinat, A., Tedeschi, R., Crotti, E., et al. (2011). ‘*Candidatus liberibacter europaeus*’ sp. nov. that is associated with and transmitted by the psyllid *Cacopsylla pyri* apparently behaves as an endophyte rather than a pathogen. *Environmental Microbiology*, *13*, 414–426.
- Raymond, M., & Rousset, F. (1995). GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, *86*, 248–249.
- Rosenberg, N. A. (2004). DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes*, *4*, 137–138.
- Secor, G. A., Rivera, V. V., Abad, J. A., Lee, I. M., Clover, G. R. G., Liefting, L. W., et al. (2009). Association of ‘*Candidatus liberibacter solanacearum*’ with zebra chip disease of potato established by graft and psyllid transmission, electron microscopy, and PCR. *Plant Disease*, *96*, 574–583.

- Sola, C., Filliol, I., Legrand, E., Lesjean, S., Locht, C., Supply, P., et al. (2003). Genotyping of the *Mycobacterium tuberculosis* complex using MIRUs: association with VNTR and spoligotyping for molecular epidemiology and evolutionary genetics. *Infection, Genetics and Evolution*, 3, 125–33.
- Wen, A., Mallik, I., Alvarado, V. Y., Pasche, J. S., Wang, X., Li, W., et al. (2009). Detection, distribution, and genetic variability of ‘*Candidatus liberibacter*’ species associated with zebra complex disease of potato in North America. *Plant Disease*, 93, 1102–1115.
- Wittwer, M., Keller, J., Wassenaar, T. M., Stephan, R., Howald, D., Regula, G., et al. (2005). Genetic diversity and antibiotic resistance patterns in a campylobacter population isolated from poultry farms in Switzerland. *Applied and Environmental Microbiology*, 71, 2840–2847.
- Yuan, X., Morano, L., Bromley, R., Spring-Pearson, S., Stouthamer, R., & Nunney, L. (2010). Multilocus sequence typing of *Xylella fastidiosa* causing Pierce’s disease and oleander leaf scorch in the United States. *Phytopathology*, 100, 601–611.