Ann Microbiol (2015) 65:1195–1199
DOI 10.1007/s13213-014-0959-z

SHORT COMMUNICATION

Variability and genetic structure of a natural population of *Citrus psorosis virus*

Asmae Achachi · Franck Curk · Mohamed Haïssam Jijakli · Fatima Gaboun · Elmostafa El Fahime · Abdelmajid Soulaymani · Mohammed El Guilli · Mohammed Ibriz

Received: 14 May 2014 / Accepted: 5 August 2014 / Published online: 22 August 2014

Abstract In this study, we examined the population structure and genetic diversity of *Citrus psorosis virus* (CPsV) in Morocco. Analysis of the coat protein partial sequences of 34 isolates collected in the three main citrus-growing areas of Morocco showed that CPsV grouped in three major groups, with low within-group nucleotide diversity. Analyses indicated that CPsV genetic diversity is not structured by the geographic origin of the CPsV isolates. The genetic variation resulting from mutations depends on evolutionary forces that have contributed to the shaping of the genetic structure and diversity of the CPsV populations analyzed here: negative selective pressure for amino acid variation, recombination between variants or mixed infection, genetic drift induced by the founder effect associated with the transmission process, and migration explained by the exchange of infected propagative plant material.

Keywords *Ophiovirus* · Phylogeny · Recombination · Selection · Genetic drift · Gene flow

The nucleotide sequences obtained in this work have been deposited in the GenBank database under accession numbers KF857180- KF857213.

A. Achachi (✉) · F. Gaboun · A. Soulaymani · M. Ibriz
Laboratoire de Génétique et Biométrie, Département de Biologie, Faculté des Sciences, Université Ibn Tofaïl, 14000 Kénitra, Maroc
e-mail: asmaeachachi@yahoo.fr

F. Curk
UMR AGAP, Institut National de la Recherche Agronomique (INRA), Centre INRA de Corse, 20230 San Giuliano, France

M. H. Jijakli
Unité de phytopathologie, Faculté Universitaire des Sciences Agronomiques, Passage des déportés 2, 5030 Gembloux, Belgique

F. Gaboun
Institut National de la Recherche Agronomique (INRA) URAPVCRP, Guich, 10000 Rabat, Maroc

E. El Fahime
Plateforme Génomique Fonctionnelle, UATR?S, Centre National pour la Recherche Scientifique et Technique (CNRS?T), 10000 Rabat, Maroc

M. El Guilli
Laboratoire de Phytopathologie, Unité de Recherche sur la Protection des Plantes, Centre Régional de la Recherche Agronomique, Station El Menzeh, 14000 Kénitra, Maroc

Findings

Citrus psorosis virus (CPsV) is the presumed causal agent of psorosis disease (Martín et al. 2004), an important graft transmissible disease of citrus worldwide. CPsV, the type species of the genus *Ophiovirus*, has a genome consisting of three single-stranded RNAs of negative polarity that have been totally sequenced (Martín et al. 2005).

Genomes of RNA viruses are expected to rapidly accumulate mutations due to the error-prone nature of RNA polymerases (Domingo and Holland 1994). Such diversification may lead to loss of fitness, due to the build up of deleterious mutations or to rapid and unpredictable fitness gains (Clarke et al. 1993; Roossinck 1997). Genetic variation can occur in CPsV, being that the virus is transmissible by grafting to different citrus hosts. However, other evolutionary factors affecting this variability include genetic recombination, genome reassortment, natural selection, genetic drift and migration (Roossinck 1997; Moya et al. 2004). The relative contribution of these evolutionary forces can be studied by analyzing the genetic structure of virus populations.

In this work, we studied the genetic structure and diversity of a natural population of *Citrus psorosis virus* (CPsV), using sequence analysis of region R3 located on genomic RNA3 of 34 CPsV isolates collected from different citrus areas of Morocco. This study provides new insights into the population structure and genetic diversity within CPsV isolates, which are

relevant for designing alternative control strategies against this widespread citrus virus.

Thirty-four isolates of CPsV were collected in a field survey in 2011 and 2012 in three main citrus-growing areas of Morocco: the Gharb region (Northwest Morocco), the Moulouya region (Northeast; 469 km from the Gharb region) and the Tadla region (Central Morocco; 703 km from Moulouya, 419 km from the Gharb region). Isolates are denoted as follows: first letter, citrus geographic region (G, Gharb; M, Moulouya, and T, Tadla); different numbers from the same geographic region correspond to different isolates, including 31 from sweet orange plants [*C. sinensis* (L.) Osb.] of different varieties and three from Clementine plants (*C. clementina* Hort. ex Tan.) (isolates M17, M18 and M19).

Sequencing data associated with the 34 CPsV isolates analyzed in this study included: 11 from the Gharb region (GenBank accessions nos KF857180−KF857190), 11 from the Tadla region (GenBank accessions nos KF857191 −KF857201) and 12 from the Moulouya region (GenBank accessions nos KF857202−KF857213). Other CPsV CP nucleotides sequences used in our analyses were from the following GenBank entries: strain 90-1-1 (AR, FJ495195) from Argentina, isolates CA1 and CA2 (CA1, AM235980; CA2, AM235981) from California, isolate CPV-4 (FL, AF036338) from Florida, isolates IT1 and IT2 (IT1, AM159538; IT2, AM409317) from Italy. Isolate P-121 (SP, NC_006316) from Spain (Martín et al. 2005) was used as sequence reference in this study.

Total RNA was extracted from healthy or CPsV-infected leaf tissue with the RNeasy Plant Mini Kit (QIAGEN, Hilden, Germany) according to the manufacturer's protocol, and used as template for reverse transcription and PCR amplification (RT-PCR) using the QIAGEN® One Step RT-PCR Kit (QIAGEN, Hilden, Germany). The thermocycling conditions included 30 min at 50 °C for RT, 15 min at 95 °C as heating step of HotStar Taq DNA polymerase, 40 cycles consisting of 30 s at 94 °C, 1 min at 58 °C and 1 min at 72 °C, and a final elongation step of 10 min at 72 °C. Region R3 was amplified with primers Ps66 (5'-TCGAAGCTGTATGATGGTGA, positions 768–787 on RNA 3) and Ps65 (5'-TGCCATCTGGAG TGAGGCT, positions 1182–1200 on RNA3) (Martín et al. 2004). All CPsV isolates yielded PCR products of 430 bp, which corresponded to the size expected for the region R3, whereas no amplification was obtained using equivalent tissues from healthy plants.

The purified RT-PCR products were sequenced on both strands in an automatic sequencer (Applied Biosystems 3130XL sequencer) by priming the sequencing reactions with the same oligonucleotides used for cDNA synthesis. Sequences were analyzed using the BioEdit program (Hall 1999). Ambiguous (heterozygous) sites were coded using the standard IUPAC codes for combinations of two or more bases. Multiple alignment was performed with the algorithm CLUSTAL W program (Larkin et al. 2007), implemented in the program BioEdit.

The software DARwin v5 (Perrier and Jacquemoud-Collet 2006) was used to calculate dissimilarity from allelic data, and to build a phylogenetic tree applying a weighted neighbor-joining topology method (Saitou and Nei 1987) to visualize genetic relationships among the analyzed isolates. The significance of each node was evaluated with bootstrapping data for 500 replications of the original matrix. Phylogenetic relationships were also determined with the maximum-likelihood (ML) algorithm of MEGA 5.05 (Tamura et al. 2011). The substitution model of Tamura and Nei (1993) was used and the shape parameters, including the transition to transversion ratio with an eight category gamma distribution, were estimated during tree reconstruction.

The program MEGA 5.05 (Tamura et al. 2011) was used for: (1) estimation of nucleotide distances between sequences pairs and diversities (mean nucleotide distances) using Kimura-two-parameter as the nucleotide substitution model (Kimura 1980); (2) estimation of nucleotides diversities within-population and between populations from genetic distances, as described by Nei (1987); (3) estimation of the ratio between nonsynonymous and synonymous substitution (d_N/d_S) by the Kumar's method using P distance (Nei and Kumar 2000) to study the role of natural selection at the protein level. $d_N/d_S \approx 0$ indicates neutral evolution, $d_N/d_S < 1$ negative or purifying selection, and $d_N/d_S > 1$ positive or adaptive selection.

The program DnaSP 5.10 (Librado and Rozas 2009) was used to assess genetic differentiation and the gene flow level between different countries or geographic areas with the statistic F_{st} (Weir and Cockerham 1984). F_{st} can assume values from 0, no genetic differentiation and complete gene flow, to 1, complete genetic differentiation as a consequence of null gene flow. Only countries or geographical areas with more than four isolates of the virus were taken into account. DnaSP 5.10 was also used to estimate the minimum number of recombination events (Rm) in the history of the sample using the four-gamete test (Hudson and Kaplan 1985).

To assess the diversity within Moroccan CPsV isolates, nucleotide sequences corresponding to region R3 located on genomic RNA3 were aligned with respect to the CPsV reference sequence (Martín et al. 2005). The number of polymorphic sites for the 34 isolates was 62 (15 %). The nucleotide diversity between Moroccan isolates was 0.020±0.003 (Table 1). No correlation was observed between the genetic distance and the geographic origin or the host species from which the isolates were obtained.

Four isolates revealed double peaks in the chromatograms of both fragments sequenced. The presence of ambiguous sites and possible parental alleles among the isolates strongly suggests real heterozygosis and also the occurrence of some level of recombination. Ambiguous sites were encountered in

Table 1 Genetic diversity and population genetic parameters of Moroccan CPsV isolates

Population parameters[a]							Diversity within genetic groups			
M	D	d_N	d_S	d_N/d_S	Rm	F_{st}	I	II	III	IV
0.052	0.020±0.003	0.009±0.002	0.044±0.009	0.204	8	0.108	0.007±0.002	0.020±0.006	0.021±0.008	0.106±0.017

[a] Population parameters: M, maximum nucleotide distance between isolate pairs; D, diversity (mean nucleotide distance between isolates pairs); d_N, average number of nonsynonymous substitutions per nonsynonymous site; d_S, average number of synonymous substitutions per synonymous site; d_N/d_S, average ratio between nonsynonymous and synonymous substitutions for each pair of comparisons; Rm, minimum number of recombination events; F_{st}, measure of gene flow

seven positions in the sequences analyzed, and putative heterozygous parental were observed. Heterozygosis can be caused by mutation in one allele, but it can also be due to genetic exchange within isolate population of virus variants. Mutation is likely to occur in single heterozygous sites, but recombination is more likely to be associated to two or more sites (Cuervo et al. 2010). Three isolates presented two ambiguous sites; an exception was encountered in one isolate (T20) with just one ambiguous site. This observation reinforces that real recombination might be occurring, because in sequence alignments, at least two ambiguous sites were observed in three isolates. Moreover, after determination of the minimum number of recombination events by using DnaSP, eight recombinations events between sites were detected and involved regions between nucleotides 54–87, 87–179, 183–200, 200–210, 210–281, 281–282, 282–291, and 291–393 (Table 1).

Recombinant sites should be confirmed through biological cloning of isolates, to distinguish between mixed infection and quasispecies, and other possible causes such as disparate secondary structures that could bias certain nucleotide positions.

RNA viruses are known to have a high variation potential, because of their high mutation and recombination rates, and since no proofreading correction mechanism is associated with their RNA-dependent RNA polymerase. Consequently, each viral isolate is expected to consist of a population of genetically related variants, termed quasispecies (García-Arenal et al. 2001). Analysis of nucleotide sequences of clones obtained from RT-PCR products showed that most viral isolates had populations composed of a predominant sequence variant and some minor variants that are genetically close (Martín et al. 2006). Thus, CPsV isolates seem to develop more quasispecies than mixed infections in this case.

Phylogenetic relationships were inferred using the NJ and ML methods for the 34 sequences, together with seven sequences available in GenBank. Trees calculated with the NJ and ML methods were closely similar, with the 41 CPsV CP partial sequences falling into four divergent genetic groups, designated I–IV. Branch lengths were generally short, although they were supported by low bootstrap values (Fig. 1). The Moroccan isolates clustered into three of these groups (I – II and III), with group I diverging much more than the other three groups, but each group being very homogeneous, with a maximum nucleotide distance between isolates of 0.052 (Table 1). Isolates from group II were closely related at the nucleotide level to isolates from Spain (P-121) (95–98 % identity) and Italy (IT1) (96–99 % identity). The isolates from Argentina, California, Florida and Italy fell into group IV. There was no association between the group and the geographic origin of CPsV isolates. Genetically distant groups

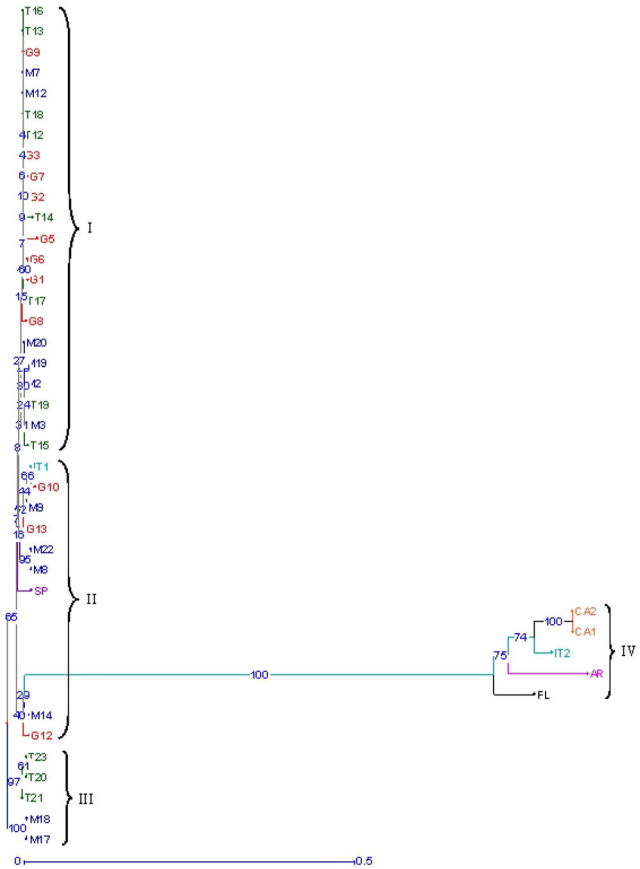

Fig. 1 Dendrogram representing the genetic relationships between the studied isolates generated by cluster analysis using the NJ method (Gharb isolates: *red*, Moulouya isolates: *blue*, Tadla isolates: *green*, Argentina isolate: *magenta*, California isolates: *orange*, Florida isolate: *black*, Italy isolates: *light blue* and Spain isolate: *purple*). Bootstrap probabilities are shown above the branch

Table 2 Within-population and between-populations nucleotide diversity in region R3 of RNA3 for CPsV subpopulations grouped by geographic origin

Subpopulation	Gharb	Moulouya	Tadla
Gharb	0.013±0.003	0.019±0.003	0.019±0.003
Moulouya		0.023±0.004	0.023±0.004
Tadla			0.022±0.004

A subpopulation was considered here as the set of isolates from the same geographic region (Gharb, Moulouya or Tadla). Nucleotide diversities were computed by using the Kimura two-parameter method (Kimura 1980)

of isolates were also observed in the same country (Callaghan and Dietzgen 2005). In addition, there was no apparent correlation between the geographic origin of the CPsV isolates and their nucleotide distance. For example, isolate SP from Spain and isolate IT1 from Italy fell in with isolates from Morocco in phylogenetic group II (Fig. 1), isolate FL from Florida fell in with isolates AR from Argentina, CA1 and CA2 from California, and IT2 from Italy in phylogenetic group IV (Fig. 1). Interestingly, a lack of correlation between genetic and geographic proximity was also reported for the closterovirus CTV (Rubio et al. 2001), the citrivirus CLBV (Vives et al. 2002), and another ophiovirus, MiLBVV (Navarro et al. 2005).

Phylogenetic groups of CPsV isolates from Morocco can be distinguished by informative changes (data not shown). Sequences from group I are characterized by two of these changes. Six specific point mutations are present in isolates M17 and M18 of group II and five specific point mutations are also present in isolates T20, T21 and T23 of group II, whereas six other changes are detected in the sequences of this group. Finally, other nucleotides variations are present in sequences of group III.

In order to confirm genetic structure and relationships between the studied isolates, nucleotide diversity values were estimated between and within CPsV subpopulations, considering a subpopulation as a group of isolates that were originally collected from the same geographical region. Between-subpopulation diversity values were smaller or similar in their order of magnitude to those corresponding to within-subpopulation diversity values (Table 2). This indicates that there is no differentiation of the population according to geographic origin. The genetic structure described here suggests a displacement of some CPsV isolates among geographically isolated subpopulations, since isolates from various regions separated by several hundreds of kilometers showed identical genetic structure. The same range of diversity was obtained among isolates from the Gharb region and other citrus regions in Morocco (Table 1). These findings can probably be explained by the intense exchange of citrus propagation material, including CPsV-infected material, between distant geographic regions of Morocco, and lead us to question the interpretation of genetic differentiation (F_{st}) estimates, in terms of effective dispersal.

The statistic F_{st} was calculated. The global gene flow was moderate (Table 1), and a more detailed analysis was performed by comparing virus subpopulations of different areas that did not correspond to geographic proximity, two by two (data not shown). A high gene flow occurred between the Moulouya and Tadla regions, whereas other subpopulations showed moderate gene flow.

Gene flow could have also contributed to shaping the genetic structure of population, and could also be responsible for the low genetic variability observed for the CPsV population analyzed here.

The selective constraint for amino acid change was estimated by computing nonsynonymous and synonymous substitutions. The number of synonymous substitutions per synonymous site (d_S) was 0.044±0.009 (Table 1), whereas the number of nonsynonymous substitutions per nonsynonymous site (d_N) was 0.009±0.002, smaller than d_S (Table 1). The ratio (d_N/d_S) was low, suggesting a negative selective pressure for amino acid changes (Nei 1987), and indicating functional or structural constraints for amino acid changes. Indeed, 12 out of the 136 amino acid positions (8.82 %) were polymorphic, but only five of them potentially altered amino acid polarity; particularly, Gly^2 was changed to Cys^2 in isolate G6, Ser^7 was changed to Pro^7 in isolate M2, Pro^{17} was changed to Gln^{17} in isolate M14, Gly^{72} was changed to Ser^{72} in isolate M8 and in isolate M22, and Asn^{107} was changed to His^{107} in isolate T20, T21 and in isolate T23 (data not shown).

The observed genetic stability of CPsV population could be due to the apparent absence of a natural vector, which might have prevented population changes induced by the founder effect often associated with the transmission process, as reported for CLBV (Vives et al. 2002). Indeed, no evidence of the natural vector of CPsV exists so far.

Taking all the above data together, a low degree of genetic variability depends on evolutionary forces, recombination between variants or mixed infections, selection pressure, founder effect and gene flow. The relative contributions of these evolutionary forces play a critical role in shaping virus evolution and the ensuing virus population (Grenfell et al. 2004). This knowledge will help us to understand the molecular structure of CPsV populations in order to use pathogen-mediated strategies for CPsV protection purposes, since these methods use virus sequences homologous to their targets.

Acknowledgments This work was supported by the Belgium development agency, BTC (project FRAB II).

The manuscript does not contain clinical studies or patient data.

References

Callaghan B, Dietzgen RG (2005) Nucleocapsid gene variability reveals two subgroups of *Lettuce necrotic yellows virus*. Arch Virol 150: 1661–1667

Clarke D, Duarte E, Moya A, Elena S, Domingo E, Holland J (1993) Genetic bottleneck and population passages cause profound fitness differences in RNA viruses. J Virol 67:222–228

Cuervo P, Domont GB, De Jesus JB (2010) Proteomics of trypanosomatids of human medical importance. J Proteomics 73:845–867

Domingo E, Holland JJ (1994) Mutation rates and rapid evolution of RNA viruses. In: Morse SS (ed) The evolutionary biology of viruses. Raven Press, New York, pp 161–183

García-Arenal F, Fraile A, Malpica JM (2001) Variability and genetic structure of plant virus populations. Annu Rev Phytopathol 39:157–186

Grenfell BT, Pybus OG, Gog JR, Wood JLN, Daly JM, Mumford JA, Holmes EC (2004) Unifying the epidemiological and evolutionary dynamics of pathogens. Science 303:327–332

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41:95–98

Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics 111:147–164

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–2948

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452

Martín S, Alioto D, Milne RG, Garnsey SM, García ML, Grau O, Guerri J, Moreno P (2004) Detection of *Citrus psorosis virus* by ELISA, molecular hybridization, RT-PCR and immunosorbent electron microscopy and its association with citrus psorosis disease. Eur J Plant Pathol 110:747–757

Martín S, López C, García ML, Naum-Onganía G, Grau O, Flores R (2005) The complete nucleotide sequence of a Spanish isolate of *Citrus psorosis virus*: comparative analysis with other ophioviruses. Arch Virol 150(1):167–176

Martín S, García ML, Troisi A, Rubio L, Legarreta G, Grau O, Alioto D, Moreno P, Guerri J (2006) Genetic variation of populations of *Citrus psorosis virus*. J Gen Virol 87:3097–3102

Moya A, Holmes EC, Gonzalez-Candelas F (2004) The population genetics and evolutionary epidemiology of RNA viruses. Nat Rev Microbiol 2:279–288

Navarro JA, Torok VA, Vetten HJ, Pallás V (2005) Genetic variability in the coat protein genes of *Lettuce big-vein associated virus* and *Mirafiori lettuce big-vein virus*. Arch Virol 150:681–694

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Nei M, Kumar S (2000) Molecular evolution and phylogenetics. Oxford University Press, New York

Perrier X, Jacquemoud-Collet JP (2006) DARwin software. http://darwin.cirad.fr/darwin

Roossinck MJ (1997) Mechanisms of plant virus evolution. Annu Rev Phytopathol 35:191–209

Rubio L, Ayllón MA, Kong P, Fernández A, Polek M, Guerri J, Moreno P, Falk BW (2001) Genetic variation of *Citrus tristeza virus* isolates from California and Spain: evidence for mixed infections and recombination. J Virol 75:8054–8062

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

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10:512–526

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA 5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739

Vives MC, Rubio L, Galipienso L, Navarro L, Moreno P, Guerri P (2002) Low genetic variation between isolates of *Citrus leaf blotch virus* from different host species and of different geographical origins. J Gen Virol 83:2587–2591

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370