

Research

Open Access

## Genetic variation and recombination of RdRp and HSP 70h genes of *Citrus tristeza virus* isolates from orange trees showing symptoms of citrus sudden death disease

Clarissa PC Gomes<sup>1</sup>, Tatsuya Nagata\*<sup>1</sup>, Waldir C de Jesus Jr<sup>2,5</sup>, Carlos R Borges Neto<sup>3</sup>, Georgios J Pappas Jr<sup>1</sup> and Darren P Martin<sup>4</sup>

Address: <sup>1</sup>Programa de Pós-Graduação em Ciências Genômicas e Biotecnologia, Universidade Católica de Brasília, SGAN, Quadra 916, Módulo B, Av. W5 Norte, 70.790-160, Brasília-DF, Brazil, <sup>2</sup>Fundecitrus, Av. Adhemar Pereira de Barros, 201, 14807-040, São Paulo, SP, Brazil, <sup>3</sup>CENARGEN, Parque Estação Biológica, Av. W5 Norte, 70770-900, Brasília, DF, Brazil, <sup>4</sup>Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Observatory, Cape Town, 7000, South Africa and <sup>5</sup>Universidade Federal do Espírito Santo, Centro de Ciências Agrárias, Alto Universitário, S/N, 29500-000, ES, Brazil

Email: Clarissa PC Gomes - clarissagomes@yahoo.com.br; Tatsuya Nagata\* - tatsuya@pos.ucb.br; Waldir C de Jesus - wcintra@yahoo.com; Carlos R Borges Neto - cborges@cnpq.br; Georgios J Pappas - gpappas@pos.ucb.br; Darren P Martin - Darrin.Martin@uct.ac.za

\* Corresponding author

Published: 16 January 2008

Received: 17 December 2007

*Virology Journal* 2008, **5**:9 doi:10.1186/1743-422X-5-9

Accepted: 16 January 2008

This article is available from: <http://www.virologyj.com/content/5/1/9>

© 2008 Gomes et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** Citrus sudden death (CSD), a disease that rapidly kills orange trees, is an emerging threat to the Brazilian citrus industry. Although the causal agent of CSD has not been definitively determined, based on the disease's distribution and symptomatology it is suspected that the agent may be a new strain of *Citrus tristeza virus* (CTV). CTV genetic variation was therefore assessed in two Brazilian orange trees displaying CSD symptoms and a third with more conventional CTV symptoms.

**Results:** A total of 286 RNA-dependent-RNA polymerase (RdRp) and 284 heat shock protein 70 homolog (HSP70h) gene fragments were determined for CTV variants infecting the three trees. It was discovered that, despite differences in symptomatology, the trees were all apparently coinfecting with similar populations of divergent CTV variants. While mixed CTV infections are common, the genetic distance between the most divergent population members observed (24.1% for RdRp and 11.0% for HSP70h) was far greater than that in previously described mixed infections. Recombinants of five distinct RdRp lineages and three distinct HSP70h lineages were easily detectable but respectively accounted for only 5.9 and 11.9% of the RdRp and HSP70h gene fragments analysed and there was no evidence of an association between particular recombinant mosaics and CSD. Also, comparisons of CTV population structures indicated that the two most similar CTV populations were those of one of the trees with CSD and the tree without CSD.

**Conclusion:** We suggest that if CTV is the causal agent of CSD, it is most likely a subtle feature of population structures within mixed infections and not merely the presence (or absence) of a single CTV variant within these populations that triggers the disease.

## Background

Diseases caused by the aphid-borne *Closterovirus*, *Citrus tristeza virus* (CTV), are among the greatest threats to citrus production worldwide. CTV has a c.a. 19.3 Kb single stranded positive sense RNA genome encoding 19 proteins expressed from 12 open reading frames (ORFs) that is the largest amongst the known closteroviruses [1]. A large amount of phenotypic diversity has been detected amongst CTV isolates characterized from around the world, with the identified strains differing from one another in the symptoms they induce in different citrus genotypes or their aphid transmissibility [2-9].

Since 2001, an emergent disease, called citrus sudden death (CSD), has caused the death or eradication of four million orange trees throughout the Brazilian states of Minas Gerais and São Paulo [10]. The disease is continuing to spread and epidemiological studies indicate that the casual agent is possibly distributed by an air-borne vector [11]. One curiosity of the disease is that it has almost exclusively been reported in trees comprising a sweet orange graft onto a Rangpur lime (*Citrus limonia* Osb.) rootstock. Neither Rangpur lime nor sweet orange appear sensitive to the disease in isolation and diseased trees can sometimes be salvaged by removing the sweet orange graft.

There is some circumstantial evidence implicating CTV as a casual agent of CSD: (1) Trees presenting with CSD symptoms are invariably infected with CTV (Fundecitrus, personal communication); (2) CSD and CTV have similar epidemiological properties [11] and (3) CTV induces symptoms that present and develop in a similar manner to those occurring in trees with CSD [12]. It remains to be confirmed, however, whether CSD is caused by some novel CTV variant, another citrus-infecting virus (for example a marafivirus) [13] or some synergistic interaction between CTV and one or more other infectious agents. Due to its unknown etiology, no serological or molecular tools are available for diagnosing CSD and currently the only reasonably reliable indicator of the disease is yellowing of the phloem vessel system in the tissues of rootstocks.

In an attempt to identify genetic features of CTV that may be associated with CSD, two loci within the RNA-dependent RNA polymerase (RdRp) and the HSP70 homologue (HSP70h) genes were extensively sampled from CTV populations infecting three trees – two with CSD and one without the disease. While significant differences in the population structures of the viruses in the three trees were observed, it was not possible to definitively associate any of these with CSD. While our results demonstrate extraordinary Brazilian CTV genetic diversity and intraspecific recombination within individual plants they also indicate

that it is unlikely to simply be the occurrence of a single CTV genetic variant within these mixed CTV infections that causes the disease.

## Results and Discussion

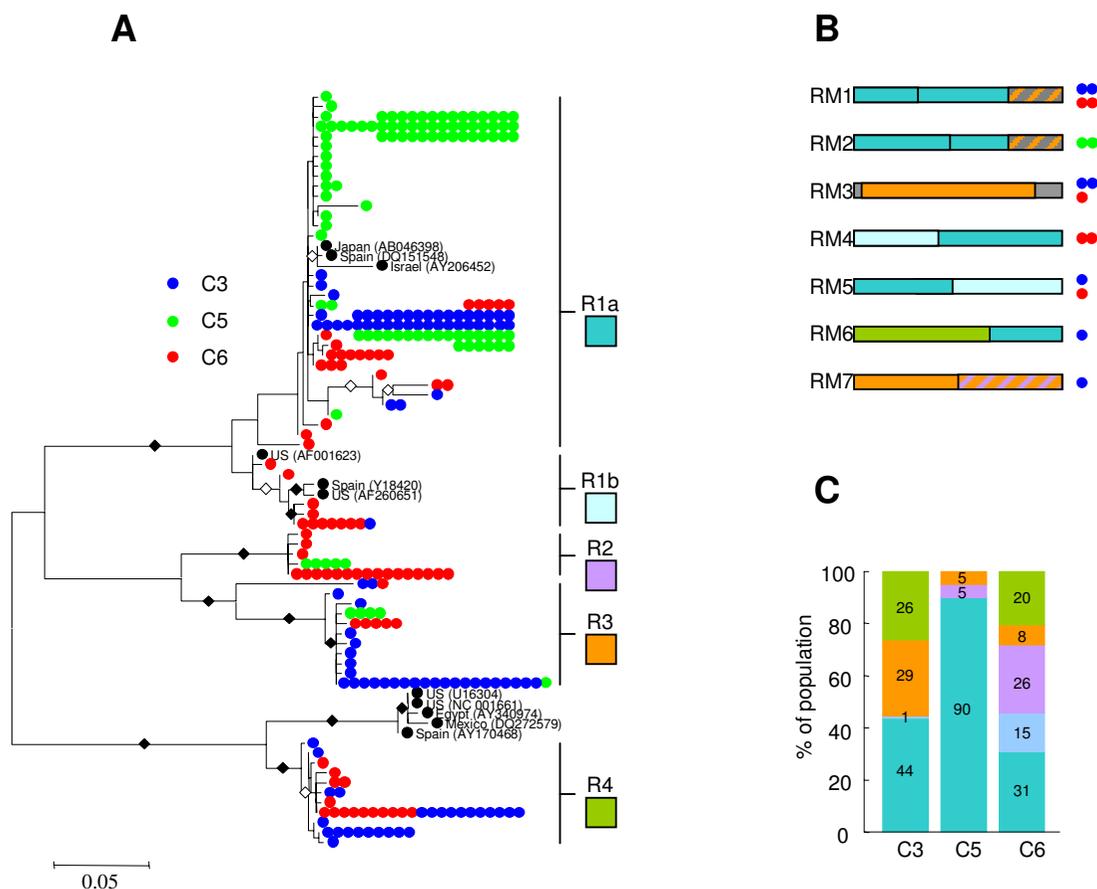
### **Within tree CTV populations are unexpectedly diverse**

Preliminary analyses of the pairwise genetic distances between the 286 RdRp and 284 HSP70h sequences from the three trees indicated that all were infected with a diverse range of CTV genotypes. Previous analyses of CTV mixed infections have revealed within host CTV population structures that are expected for RNA virus quasispecies: One or two predominant genotypes (>30% of the population) and a number of closely related variants of these [2,6,9]. The CTV populations in the three Brazilian trees analyzed here are extraordinarily diverse by comparison with these previously described populations. For example, the most diverse population of CTV sequences in an individual tree described by Sentandreu et al. [9] was for isolate T385 at the p20 locus. Out of 30 sequences sampled from T385, ten unique haplotypes were identified. The total number of haplotypes identified amongst 451 p20 and 274 p18 sequences sampled from six trees (originating in Spain and Japan) were 35 and 23 respectively. By contrast sequence analysis of the 286 RdRp and 284 HSP70h gene fragments sampled from amongst the three Brazilian orange trees here revealed a total of 76 RdRp and 153 HSP70h haplotypes.

The minimum pairwise sequence identities observed in individual Brazilian trees were 75.9% for two RdRp sequences from tree C3 and 89.0% for two HSP70h sequences from tree C5. In all the trees we observed that RdRp diversity was greater than that of HSP70h. Although the HSP70h gene, which has an important function in virion assembly [14,15], is characteristically more conserved than the RdRp gene, we have paradoxically sampled approximately twice as many HSP70h haplotypes.

### **Potential associations between specific CTV variants CSD**

We therefore used a mixture of recombination and phylogenetic analyses to characterize the complex populations of CTV RdRp and HSP70h sequences found in the three trees. Since recombination has been previously reported in CTV [2,16] and is a process that violates the assumptions of conventional phylogenetic reconstruction methods (such as neighbor joining, and maximum likelihood) [17], we sought to remove all obviously recombinant sequences from our dataset prior to phylogenetic analysis. We used a battery of recombination analysis methods implemented in RDP3 to identify and remove 17 RdRp and 32 HSP70h sequences that were obviously recombinant.

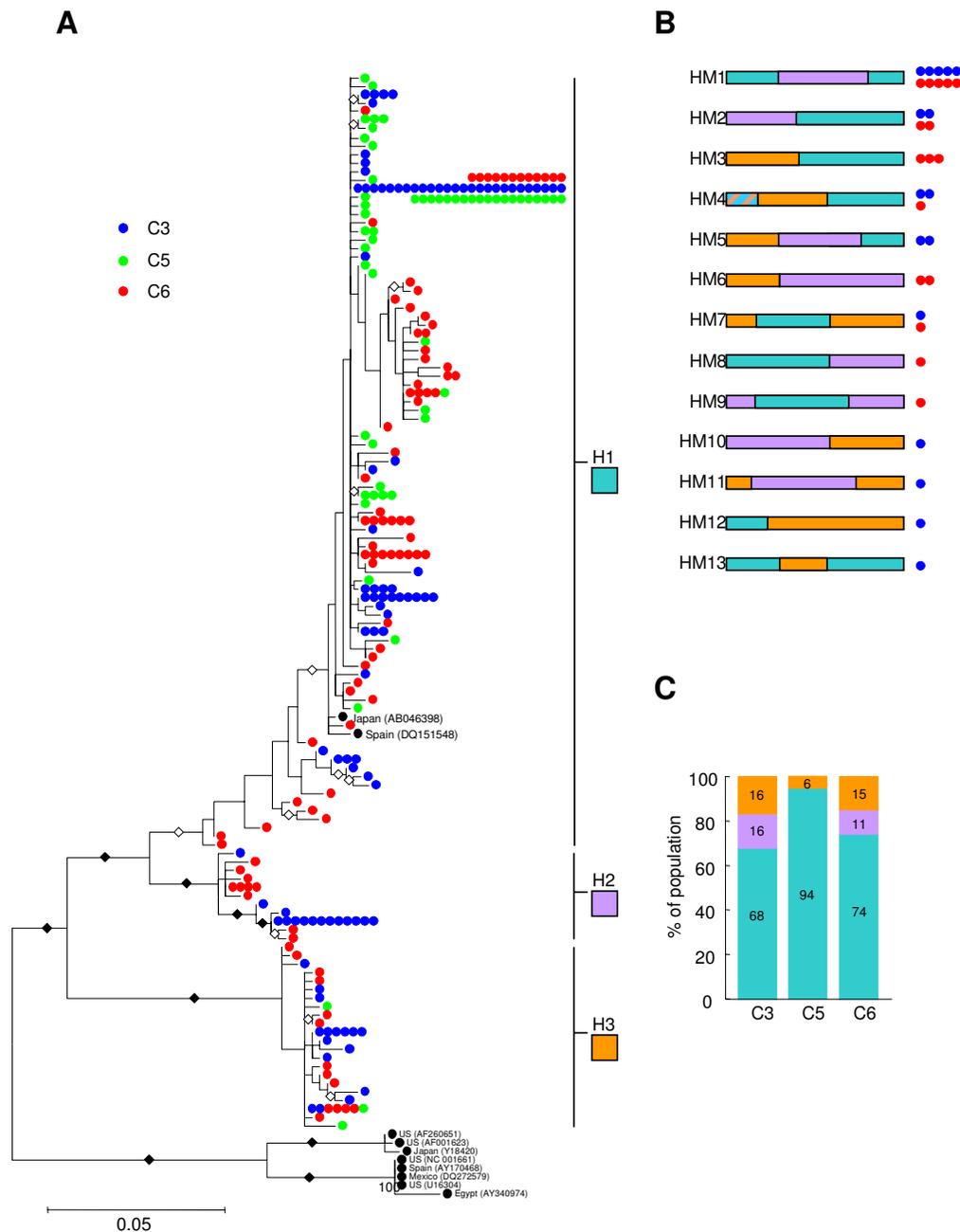


**Figure 1**  
**Phylogenetic and recombination analysis of 286 RdRp gene fragments sampled from three Brazilian orange trees.** (A) Largely recombination free maximum likelihood phylogeny of 269 RdRp gene fragment sequences. Sequences sampled from trees C3, C5 (both with CSD) and C6 (without CSD) are represented by blue, green and red dots respectively. Multiple dots on particular branches represent identical sequences. Whereas branches with more than 75% bootstrap support are labeled with filled black diamonds, those with between 50 and 74% support are labeled with unfilled black diamonds. Sequence lineages R1a, R1b, R2, R3 and R4 are indicated. (B) The mosaic structures of obviously recombinant sequences excluded from the phylogenetic analysis. The mosaics are colour coded according to the main sequence lineages from which they have probably been derived. Grey areas represent sequence tracts that were not clearly derived from one of the five identified sequence lineages in A. Coloured dots beside the schematic representations of the recombinant sequences represent the numbers and origins of recombinant sequences sharing similar mosaics. (C) The relative population representation of different major RdRp lineages identified in A.

Maximum likelihood trees constructed using the remaining 269 RdRp and 252 HSP70h sequences (Figures 1A and 2A) indicated that, with a few exceptions, the sequences obtained from all three orange trees respectively fell into four and three distinct (>75% bootstrap support) RdRp and HSP70h lineages. The lineages were arbitrarily named R1 through R4 for the RdRp sequences and H1 through H3 for the HSP70h sequences. The R1 lineage was further divided into R1a and R1b groups on the basis of there being >50% bootstrap support for a small but distinct R1b lineage. Whereas the R1b, R4 and H2 lineages contained

only sequences sampled from trees C3 and C6, the R2 lineage contained only sequences sampled from trees C5 and C6. There was no lineage containing only sequences sampled from the trees with CSD (C3 and C5).

We examined the inferred mosaic structures of the recombinant sequences identified previously and determined that most were clearly recombinants between the major lineages (Figs 1B and 2B). While we cannot rule out the possibility that some of these recombinants might be RT-PCR artifacts, the occurrence of nearly identical recom-



**Figure 2**  
**Phylogenetic and recombination analysis of 284 HSP70h gene fragments sampled from three Brazilian orange trees.** (A) Largely recombination free maximum likelihood phylogeny of 252 HSP70h gene fragment sequences. Sequences sampled from trees C3, C5 and C6 are represented by blue, green and red dots respectively. Multiple dots on particular branches represent identical sequences. Whereas branches with more than 75% bootstrap support are labeled with filled black diamonds, those with between 50 and 74% support are labeled with unfilled black diamonds. Sequence lineages H1, H2, and H3 are indicated. (B) The mosaic structures of obviously recombinant sequences excluded from the phylogenetic analysis. The mosaics are colour coded according to the main sequence lineages from which they have probably been derived. Coloured dots beside the schematic representations of the recombinant sequences represent the numbers and origins of recombinant sequences sharing similar mosaics. (C) The relative population representation of different major HSP70h lineages identified in A.

binant mosaics in different trees strongly suggests that at least three of the RdRp recombinants (RM1, RM3, and RM5) and four of the HSP70h recombinants (HM1, HM2, HM4, and HM7) are genuine. As none of these recombinants were found in both of the trees with CSD, they can also be excluded as causal agents of the disease.

Interestingly, all seven of these "shared" recombinants were detected in the tree without CSD, C6, and one of the trees with CSD (C3). Also, no HSP70h recombinants and only one class of RdRp recombinant (RM2) was found in C5, the other tree with CSD. One obvious reason for the lack of recombinants in tree C5 is that it was a much younger tree than the other two at the time of sampling. Analysis of the relative population representations of the two CTV gene lineages in the different trees (Figs 1C and 2C) also indicated that the CTV populations in the two older trees, C6 and C3 were substantially more similar to one another than either was to the CTV population in C5.

#### **There is no obvious association between CSD and CTV population structure**

Genetic differences between CTV population structures in the three trees were quantified using Hudson, Boos and Kaplan's  $K_{st}^*$  statistic [18]. Although CTV population structures were found to be significantly different between all trees (permutation test  $p$ -value < 0.001 for all tree and loci comparisons), the populations of both the RdRp and HSP70h sequences in trees C3 (with CSD) and C6 (without CSD) were substantially more similar to one another ( $K_{st}^* = 0.037$  for RdRp and 0.015 for HSP70h) than the populations in either tree were to populations sampled from tree C5 (for RdRp  $K_{st}^* = 0.15$  and 0.16 for comparisons of C5 with C3 and C6 respectively; for HSP70h  $K_{st}^* = 0.042$  and 0.051 for comparisons of C5 with C3 and C6 respectively). Unlike the association between CTV virulence and population structures observed by Sentandreu et al. [9], there therefore appears to be no simple association between population structures and CSD in these three Brazilian CTV trees.

#### **Conclusion**

We have determined that the two sampled trees with CSD (C3 and C5) have more divergent population structures than one of these trees (C3) has compared to a tree without the disease (C6). Our analysis therefore suggests that, rather than a recombinant CTV variant or obvious differences between the population structures of mixed CTV infections, the most likely hypothesis relating CTV infections and CSD is that it may be caused by the occurrence and interaction of two or more specific CTV sequence variants within these mixed infections. If nothing else this study indicates that it may be extremely difficult to fulfill Koch's postulates for potential causes of CSD. Properly demonstrating that the existence of specific sequence var-

iants within a mixed CTV infection are key to the development of CSD will require the construction of numerous infectious CTV full genome clones and a mechanism for artificially recreating complex mixed infections such as we've described here.

#### **Methods**

##### **Virus sample collection**

Three natural Brazilian CTV populations were surveyed. These were obtained from: A tree without CSD symptoms, C6 (12 years old), collected at Araraquara in São Paulo (SP) state and two trees with CSD symptoms, C3 (6 years old) and C5 (2 years old), collected at Colômbia, in SP state and at Comendador Gomes in Minas Gerais (MG) state, respectively. The three sampling locations were on an approximately 300 Km long straight line with C3 in the middle, approximately 200 Km from C6 and 100 Km from C5. Rootstock (Rangpur lime) phloem, specifically that showing yellowing symptom characteristic of CSD, was collected, initially stored on dry ice, and then frozen at  $-80^{\circ}\text{C}$  until further processing.

##### **RNA extraction and reverse transcription-polymerase chain reaction (RT-PCR)**

CTV genomic RNA was extracted using the RNeasy Plant Mini Kit (QUIAGEN, Valencia, CA) from triturated phloem parts with liquid nitrogen in a coffee grinder. Two pairs of specific RT-PCR primers were designed for amplification of a 542 bp RdRp fragment and a 564 bp HSP70h fragment using six complete CTV genome sequences [GenBank:U116304, GenBank:AF260651, GenBank:AB046398, GenBank:NC\_001661, GenBank:AF001623, GenBank:Y18420]; RdRp-forward = GAA CCG GCT CGY GTT CGG CGT; RdRp-reverse = TTC CGC YAA CCC AGC GTT CGT CAT; HSP70h-forward = CTG GAG TTA TAT GTT CGG TAC C; HSP70h-reverse = ACG AGC TTC CAC CGA CTA CCA CTA. Total extracted RNA was denatured at  $90^{\circ}\text{C}$  for five minutes and reverse transcribed by incubation at  $42^{\circ}\text{C}$  for one hour in a reaction mixture with a total volume of 20  $\mu\text{l}$  containing first strand buffer, 0.5 mM of the four deoxynucleoside triphosphates (dNTPs), 2.5  $\mu\text{M}$  of reverse primer, 40 U of RNaseOUT (Invitrogen) and 200 U of reverse transcriptase SuperScript II (Invitrogen). An aliquot of 3  $\mu\text{l}$  of this product was amplified in a 50  $\mu\text{l}$  reaction mixture containing 1 $\times$  buffer, 0.2 mM dNTPs, 2 mM  $\text{MgCl}_2$ , 200 pM primers, and 1 U Platinum *Taq* DNA Polymerase High Fidelity (Invitrogen). The PCR conditions were:  $94^{\circ}\text{C}$  for 2 minutes, 35 cycles of each  $94^{\circ}\text{C}$  (1 minute),  $55^{\circ}\text{C}$  (1 minute), and  $68^{\circ}\text{C}$  (1 minute), and last extension of  $72^{\circ}\text{C}$  for 7 minutes.

##### **Cloning and sequencing of amplified fragments**

RT-PCR amplified DNA fragments were separated by agarose gel, and DNA fraction was cut and purified using Per-

fectprep gel cleanup kit (Eppendorf). Adenosine residues were added to the 3' ends of purified DNA fragments using *Taq* Polymerase (Invitrogen) with dATP for 30 min at 70°C. Then, the DNA was cloned into pGEM-T (Promega) and transformed into *Escherichia coli* DH5 $\alpha$  by electroporation [19]. For each of the two amplified gene regions plasmid DNA was purified from 300 transformed colonies using the Wizard Plus SV Minipreps DNA Purification System kit (Promega). Nucleotide sequences of these clones were determined in both directions using SP6 and T7 primers and an ABI Prism DNA sequencer 377 (Perkin-Elmer). Sequence quality control and individual sequence assemblies using reads from both strands were performed using the Staden package [20].

### Sequence analyses

RdRp and HSP70h fragments were aligned along with corresponding sequences from ten full length CTV genomes sampled from around the world ([GenBank:AF260651, GenBank:U116304, GenBank:NC\_001661, GenBank:AF001623] from the United States, [GenBank:AY206452] from Israel, [GenBank:AB046398] from Japan, [GenBank:Y18420, GenBank:AY170468, GenBank:DQ151548] from Spain, [GenBank:DQ272579] from Mexico and [GenBank:AY340974] from Egypt) using the ClustalW method (with default settings) [21] implemented in Mega (version 3.1) [22]. Columns of the alignment containing indels or missing data for any of the sequences were excluded from analyses leaving indel-free alignments containing 499 and 411 columns of HSP70h and RdRp sequence respectively. Pair-wise Hamming distances between the sequences in each alignment were calculated using DNAMAN (Lynnon Corporation, Canada).

Recombination amongst the sequences in the two alignments was analyzed using the RDP [23], GENECONV [24], BOOTSCAN [25], MAXCHI [26], CHIMAERA [27], SISCAN [28] and 3SEQ [29] methods implemented in RDP3 [30]. Default settings were used throughout and only potential recombination events detected by two or more of the above methods coupled with phylogenetic evidence of recombination were considered significant. Also, the severity of Bonferroni correction during detection was minimized by only searching for recombinant signals in a single sequence within groups of sequences sharing >99.7% sequence identity.

Phylogenetic trees were constructed using the neighbor joining (JC distances, 1000 bootstrap replicates) and maximum likelihood methods (HKY model + Gamma with four substitution rates, transition/transversion ratios inferred from the data and 100 bootstrap replicates) implemented in MEGA and PHYML [31], respectively.

Genetic distances between populations sampled from different trees were estimated with DnaSP [32] using Hudson, Boos and Kaplan's Kst\* statistic [18]. Significant differences between estimated Kst\* values were determined with a permutation test (1000 replicates) also implemented in DnaSP.

### Competing interests

The author(s) declare that they have no competing interests.

### Authors' contributions

TN is the corresponding author and coordinated this project. DM contributed and TN, were involved in the statistical and recombination analysis of the data. Sample collection, gene cloning and sequence alignments were carried out by CG, WJ, CBN and GP. All authors have read and approved this manuscript.

### Acknowledgements

This study was supported by Fundo da Defesa de Citricultura (Fundecitrus). The first author was supported by a grant of Universidade Católica de Brasília. The last author is supported by the Wellcome Trust. We appreciate the excellent technical support and discussion for Dr. Alice K. Inoue-Nagata. We also thank Dr. Antonio Juliano Ayres and Dr. Renato B. Bassanezi of Fundecitrus for the management of the project.

### References

- Karasev AV, Boyko VP, Gowda S, Nikolaeva OV, Hilf ME, Koonin EV, Niblett CL, Cline K, Gumpf DJ, Lee RF, Garnsey MS, Dawson WO: **Complete sequence of the citrus tristeza virus RNA genome.** *Virology* 1995, **208**:511-520.
- Rubio L, Ayllón MA, Kong P, Fernández A, Polek M, Guerri J, Moreno P, Falk BW: **Genetic variation of *Citrus tristeza virus* isolates from California and Spain: evidence for mixed infections and recombination.** *J Virol* 2001, **75**:8054-8062.
- Niblett CL, Genc H, Cevik B, Halbert S, Brown L, Nolasco G, Bonacalza B, Manjunath L, Febres VJ, Pappu HR, Lee RF: **Progress on strain differentiation of *Citrus tristeza virus* and its application to the epidemiology of citrus tristeza disease.** *Virus Res* 2000, **71**:97-106.
- Ayllon MA, Rubio L, Moya A, Guerri J, Moreno P: **The haplotype distribution of two genes of citrus tristeza virus is altered after host change or aphid transmission.** *Virology* 1999, **255**:32-39.
- Gago-zachert SP, Costa NB, Semorile L, Grau O: **Sequence variability in p27 gene of *Citrus tristeza virus* (CTV) revealed by SSCP analysis.** *Electron J Biotechnol* 1999, **2**:16-25.
- Kong P, Rubio L, Polek M, Falk BW: **Population structure and genetic diversity of California *Citrus tristeza virus* (CTV) field isolates.** *Virus Genes* 2000, **21**:139-145.
- López C, Ayllon MA, Navas J, Guerri J, Moreno P, Flores R: **Molecular variability of 5' and 3' terminal regions of the citrus tristeza virus RNA.** *Phytopathology* 1998, **88**:685-691.
- Roy A, Manjunath KL, Brlansky RH: **Assessment of sequence diversity in the 5'-terminal region of *Citrus tristeza virus* from India.** *Virus Res* 2005, **113**:132-142.
- Sentandreu V, Castro JA, Ayllon MA, Rubio L, Guerri J, Gonzalez-Candela F, Moreno P, Moya A: **Evolutionary analysis of genetic variation observed in citrus tristeza virus (CTV) after host passage.** *Arch Virol* 2006, **151**:875-894.
- Bassanezi RB, Fernandes NG, Yamamoto PT: *Morte Súbita dos Citrus. Relatório técnico* Araraquara, SP, Brazil: Fundecitrus; 2003.
- Bassanezi RB, Bergamin Filho A, Amorim L, Fernandes NG, Gottwald TR, Bové JM: **Spatial and temporal analyses of citrus sudden death as a tool to generate hypotheses concerning its etiology.** *Phytopathology* 2003, **93**:502-512.

12. Román MP, Cambra M, Juárez J, Moreno P, Vila ND, Tanaka FAO, Alves E, Kitajima EW, Yamamoto PT, Bassanezi RB, Teixeira DC, Jesus WC Junior, Ayres AJ, Fernandes NG, Rabenstein F, Giroto LF, Bové JM: **Sudden death of citrus in Brazil: a graft-transmissible bud union disease.** *Plant Dis* 2004, **88**:453-467.
13. Maccheroni WV, Alegria MC, Greggio CC, Piazza JP, Kamla RF, Zacharias PRA, Bar-Joseph M, Kitajima EW, Assumpção RC, Camarotte G, Cardozo J, Casagrande EC, Ferrari F, Franco SF, Giachetto PF, Girasol A, Jordão H Jr, Silva VHA, Souza LCA, Aguilar-Vildoso CI, Zanca AS, Arruda P, Kitajima JP, Reinach FC, Ferro JA, Silva ACR: **Identification and Genomic Characterization of a New Virus (Tymoviridae Family) Associated with Citrus Sudden Death Disease.** *J Virol* 2005, **79**:3028-3037.
14. Satyanarayana T, Gowda S, Mawassi M, Albiach-Martí MR, Ayllón MA, Robertson C, Garnsey SM, Dawson WO: **Closterovirus encoded HSP70 homolog and p61 in addition to both coat proteins function in efficient virion assembly.** *Virology* 2000, **278**:253-265.
15. Alzhanova DV, Napuli AJ, Creamer R, Dolja VV: **Cell-to-cell movement and assembly of a plant closterovirus: roles for the capsid proteins and Hsp70 homolog.** *EMBO J* 2001, **20**:6997-7007.
16. Weng Z, Barthelson R, Gowda S, Hilf ME, Dawson WO, Galbraith DW, Xiong Z: **Persistent infection and promiscuous recombination of Multiple genotypes of an RNA virus within a single host generate extensive diversity.** *PLoS One* 2007, **2**(9):e917.
17. Posada D, Crandall KA: **The effect of recombination on the accuracy of phylogeny estimation.** *J Mol Evol* 2002, **54**:396-402.
18. Hudson RR, Boos DD, Kaplan NL: **A statistical test for detecting geographic subdivision.** *Mol Biol Evol* 1992, **9**(1):138-151.
19. Sambrook J, Russell DW: *Molecular cloning: a laboratory manual* 3rd edition. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 2001.
20. Bonfield JK, Smith KF, Staden R: **A new DNA sequence assembly program.** *Nucleic Acids Res* 1995, **24**:4992-4999.
21. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acids Res* 1994, **22**:4673-4680.
22. Kumar S, Tamura K, Nei M: **MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment.** *Briefings Bioinform* 2004, **5**:150-163.
23. Martin DP, Rybicki E: **RDP: detection of recombination amongst aligned sequences.** *Bioinformatics* 2000, **16**:562-563.
24. Padidam M, Sawyer S, Fauquet CM: **Possible emergence of new geminiviruses by frequent recombination.** *Virology* 1999, **265**:218-225.
25. Martin DP, Posada D, Crandall KA, Williamson C: **A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints.** *AIDS Res Hum Retroviruses* 2005, **21**:98-102.
26. Smith JM: **Analyzing the mosaic structure of genes.** *J Mol Evol* 1992, **34**:126-129.
27. Posada D, Crandall KA: **Evaluation of methods for detecting recombination from DNA sequences: Computer simulations.** *Proc Natl Acad Sci US* 2001, **98**:13757-13762.
28. Gibbs MJ, Armstrong JS, Gibbs AJ: **Sister-Scanning: a Monte Carlo procedure for assessing signals in recombinant sequences.** *Bioinformatics* 2000, **16**:573-582.
29. Boni MF, Posada D, Feldman MW: **An exact nonparametric method for inferring mosaic structure in sequence triplets.** *Genetics* 2007, **176**:1035-1047.
30. Martin DP, Williamson C, Posada D: **RDP2: Recombination detection and analysis from sequence alignments.** *Bioinformatics* 2005, **21**:260-262.
31. Guindon S, Gascuel O: **A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood.** *Syst Biol* 2003, **52**:696-704.
32. Rozas J, Sanchez-Belbarrio JC, Meusseguer X, Rozas R: **DnaSP, DNA polymorphism analysis by the coalescent and other methods.** *Bioinformatics* 2003, **19**:2496-2497.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

