Archives of Virology
https://doi.org/10.1007/s00705-018-4024-2

BRIEF REPORT

Cucurbit aphid-borne yellows virus from melon plants in Brazil is an interspecific recombinant

Thiago Marques Costa[1] · Rosana Blawid[1] · Miguel A. Aranda[2] · Débora Maria Sansini Freitas[3] · Genira Pereira Andrade[4] · Alice Kazuko Inoue-Nagata[5] · Tatsuya Nagata[1]

Received: 25 April 2018 / Accepted: 23 July 2018

Abstract

Melon plants with severe yellowing symptoms from in Brazil were analyzed by high-throughput sequencing. Sequences homologous to the genome of the polerovirus cucurbit aphid-borne yellows virus (CABYV) were frequently retrieved. Two draft CABYV genomes were assembled from two pooled melon samples that contained an identical putative recombinant fragment in the 3′ region with an unknown polerovirus. The complete genomes of these isolates revealed by Sanger sequencing share 96.8% nucleotide identity, while both sequences share 73.7% nucleotide identity with a CABYV-N isolate from France. A molecular-clock analysis suggested that CABYV was introduced into Brazil ~ 68 years ago.

A disease known as "Amarelão do meloeiro", meaning melon severe yellowing, is a huge problem for the melon industry in northeastern Brazil. This disease was first reported in Brazil in 1997 (cited in [1]), with an extremely high incidence and a presumed viral etiology, but, the causative agent has not yet been elucidated. Due to heavy whitefly infestations in melon crops and limited numbers of other potential viral vectors, whiteflies are believed to be the most likely vectors of the disease. Criniviruses were not detected in diseased melon plants, but the presence of a whitefly-transmissible carlavirus was later confirmed and named "melon yellowing-associated virus" (MYaV) [2], and it is considered as one of the causes of the disease.

We applied high-throughput sequencing (HTS) technology to identify viruses in melon plants with severe yellowing disease to analyze the viruses present in these plants and one with sequence similarity to cucurbit aphid-borne yellows virus (CABYV) was found. CABYV was first reported in France [3] and later detected in many other countries. This virus is often found in mixed infections in cucurbit plants with other viruses, especially with the potyviruses that share common insect vectors [4, 5]. Here, we describe the genome of CABYV isolates from diseased melon plants collected in two regions of Brazil.

Samples of melon plants with severe yellowing symptoms were collected in 2015 from two sites more than 800 km apart in two states in north-eastern Brazil: Ceará (CE, near the city of Mossoró), and Bahia (BA, Juazeiro county). Thirty-two samples of melon plants from Mossoró and six samples from Juazeiro were separately pooled. A partial purification procedure for virions and RNA extraction was performed as described by [6]. Two cDNA libraries of these pooled samples were constructed by Macrogen Inc. (Seoul, South Korea) and were sequenced using the Illumina HiSeq 2000 platform (100 bp paired-end and 5G scale). Sequence reads were analyzed as described by [6] using Geneious R8.1 package (Biomatters, Auckland, New Zealand).

The analyses identified sequences similar to CABYV and assembled near full-length viral genomes with a divergent region within the genome, suggesting a chimeric

Handling Editor: Robert H. A. Coutts.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00705-018-4024-2) contains supplementary material, which is available to authorized users.

✉ Tatsuya Nagata
tatsuya@unb.br

[1] Campus Darcy Ribeiro, Departamento de Biologia Celular, Universidade de Brasília, Brasília, DF 70910-900, Brazil

[2] Centro de Edafología y Biología Aplicada del Segura (CEBAS), Consejo Superior de Investigaciones Científicas (CSIC), 30100 Murcia, Spain

[3] Embrapa Semiárido, Petrolina, PE 56302-970, Brazil

[4] Departamento de Agronomia, Universidade Federal Rural de Pernambuco, Recife, PE 52171-900, Brazil

[5] Embrapa Hortaliças, Brasília, DF 70275-970, Brazil

Published online: 19 September 2018

genome assembly. We thus designed primers that annealed before and after the possible breakpoint with primers Polero_2769_For and Polero_3834_Rev (Table S1) and performed RT-PCR using total RNA preparations of the pooled samples and the selected plants from the fields.

Two new isolates collected in 2016, M3 (Icapuí, CE) and JMB1 (Mandacaru, Juazeiro, BA), were selected for complete genome sequencing using the Sanger method. Total RNA was extracted and RT-PCR was performed using Superscript IV reverse transcriptase (Thermo Fisher Scientific, Waltham, USA) with the reverse primer Polero_5596_Rev (Table S1). Two PCR primer pairs spanning the nearly complete genome in overlapping fragments were designed (Table S1) based on the analysis of the HTS data. The two amplicons (3.2 and 1.9 kbp) with an overlap of approximately 300 bp were obtained by PCR using the same cDNA synthesized with Polero_5596_Rev primer and using LongAmp *Taq* DNA polymerase (New England Biolabs, Ipswich, USA). The amplicons were gel-purified using an Illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare Life Sciences, Pittsburgh, USA), and the purified cDNAs were directly sequenced by the Sanger method at Macrogen Inc.

The assembled CABYV sequences lacked both genome ends, so primers for 3′ and 5′ RACE protocols were designed based on the HTS data (Table S1). 5′ and 3′ RACE was applied as described by [7]. Amplified cDNA fragments were gel-purified and sequenced by the Sanger method at Macrogen Inc.

The genome sequences were initially analyzed by tBlastX search implemented by Geneious R8.1. Phylogenetic trees were constructed with MrBayes3.2.6 [8] using complete genome sequences as well as the P1-P2 and the P3-P5 fusion proteins. jMODELTEST2 [9] and PROTTEST3 [10, 11] were used to estimate the best-fit models of nucleotide substitution and amino acid replacement, respectively. The MCMC chains were run until reaching convergence with the following parameters: 2 million generations using the model type VT for the P3-P5 taxon data set; 20 million generations with the WAG model for the P1-P2 data set and 10 million generations with the model GTR + I for the complete nucleotide sequence. We identified and characterized the potential recombination events using Recombination Detection Program (RDP) [12], Maximum chi-square (MaxChi2) and Geneconv implemented in RDP4, which are sensitive methods with less false positive detections [13]. To reduce false positive recombination signals, the complete genome sequences of poleroviruses that infect cucurbit plants, and phasey bean mild yellows virus were used as outlier sequences. The distance plot method implemented in RDP4 [12] was also used for selected genome sequences to support the recombination analysis.

A spatiotemporal analysis was then performed using nucleotide sequences of the P0 gene from 37 CABYV isolates. The sequences were analyzed with TemPest [14] to identify the temporal signal in the data, and we then proceeded with molecular-clock (BEAST v1.8.4, Tracer v1.6) [15] and phylogeographic (SPREAD, [16]) analyses. JModelTest2 was used to select the best-fit nucleotide substitution model for the alignment: HKY with G + I. BEAST was run. Chain length was 20 million and 10% were discarded as burn-in. Results were examined using Tree Annotator followed by TRACER v1.6 from the BEAST package. Statistical evidence was represented by the 95% HDP values. All effective sample size values of parameters were > 300. FigTree v1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/) was used to visualize and edit the maximum-credibility tree. A possible CABYV spatiotemporal migration around the world is animated in Fig. S1.

The HTS analysis detected CABYV as the major virus in both libraries. Two draft genome sequences were assembled with 1,276,525 (Mossoró/2015) and with 4,434,827 (Juazeiro/2015) reads. The assembled consensus sequences were ~ 5700 nucleotides long and shared 96.8% nucleotide identity to each other, and 76.3% identity with CABYV-N from France (X76931). Interestingly, the CABYV sequences from Brazil are unique at the 3′ region with no obvious identity with other CABYV genomes, suggesting a recombinant origin. To exclude a possible artificial chimeric genome assembly, the homologous sequencing reads to the 3′ region of CABYV-N (3505-5508 nt) were searched in Mossoró and Juazeiro HTS data, however, no single read was obtained in this "map to reference" search showing the absence of the homologous sequences to CABYV-N 3′ region in the HTS data. Next, RT-PCR was performed using a primer pair that anneals upstream and inside this unique sequence (Polero_2769_For/Polero_3834_Rev) using total RNA preparations from the pooled and several field samples, resulting in amplifications of cDNA fragments of the expected size (1066 bp) in all samples.

Two plant samples collected in 2016, M3 (Icapuí, near Mossoró) and JMB1 (Juazeiro), were selected for full-genome sequencing by the Sanger method (CABYV-M3 and -JMB1 isolate, respectively). The genome sequences of the CABYV variants from Brazil were deposited in the GenBank/EMBL/DDBJ database, LC217993 (CABYV-M3) and LC217994 (CABYV-JMB1). The full-length genome sequences of the M3 and JMB1 isolates were 5736 and 5737 nucleotides (nt), respectively. The JMB1 isolate had one nt indel in the 3′ UTR. The complete genome sequences of these two isolates differed at 181 nt, preferentially distributed in the 5′ region, sharing 96.8% nt identity.

A typical polerovirus genome organization was predicted for both genome sequences, including the P1-P2 and P3-P5 fusion proteins with the -1 ribosomal frameshift and

the readthrough domain, respectively. The motifs found in ORF0, ORF1 and ORF2 were common between CABYV from Brazil and the reported isolates [17, 18]. The non-AUG triplet (AUA, in both isolates) initiates the P3a ORF, which is thought to be essential for long distance movement [19]. Furthermore, the 3′ end of the coat protein (P3) gene (AAA UAGGUAGAC, leaky stop codon underlined) contains the conserved leaky UAG sequence of poleroviruses. The GLLAAFHQ motif is present in the CP of the CABYV from Brazil, instead of the typical G(IM)LK(AS)YHE motif conserved in poleroviruses [18]. The proline hinge, a conserved domain at the beginning of the readthrough domain (P5) of the minor CP, also differed (PPPIPPPPPP) from the typical CABYV hinge (PPPPSPSPTPPPPPPP) [18].

In the phylogenetic analysis, all CABYV and melon aphid-borne yellows virus (MABYV) (Fig. 1A) clustered together, but the CABYV sequences from Brazil were clearly distant from other CABYV sequences. One CABYV isolate from Taiwan (JQ700306) was also distant from the main CABYV cluster because it was demonstrated to be a recombinant between CABYV and MABYV isolates [20]. The CABYV isolates from Brazil were most closely related to the isolate N from France with 73.7% pairwise nt identity. The phylogenetic tree of P1-P2 fusion protein (Fig. 1B) showed closer relationship of the CABYV-M3 and JMB1 isolates from Brazil to other CABYV isolates. However, the position of the M3 and JMB1 isolates in the phylogenetic tree of P3-P5 was completely different (Fig. 1C). The P3-P5 protein sequences of CABYV-M3 and JMB1 cluster with those of African eggplant yellowing virus, sauropus yellowing virus and luffa aphid-borne yellows virus and are distantly related to other cucurbit-infecting poleroviruses. The recombination analysis suggested that the M3 and JMB1 isolates were recombinants with the CABYV-N isolate as the major parent, with the beginning breakpoint at nt 3376 in the intergenic region (IR) and the ending breaking point at

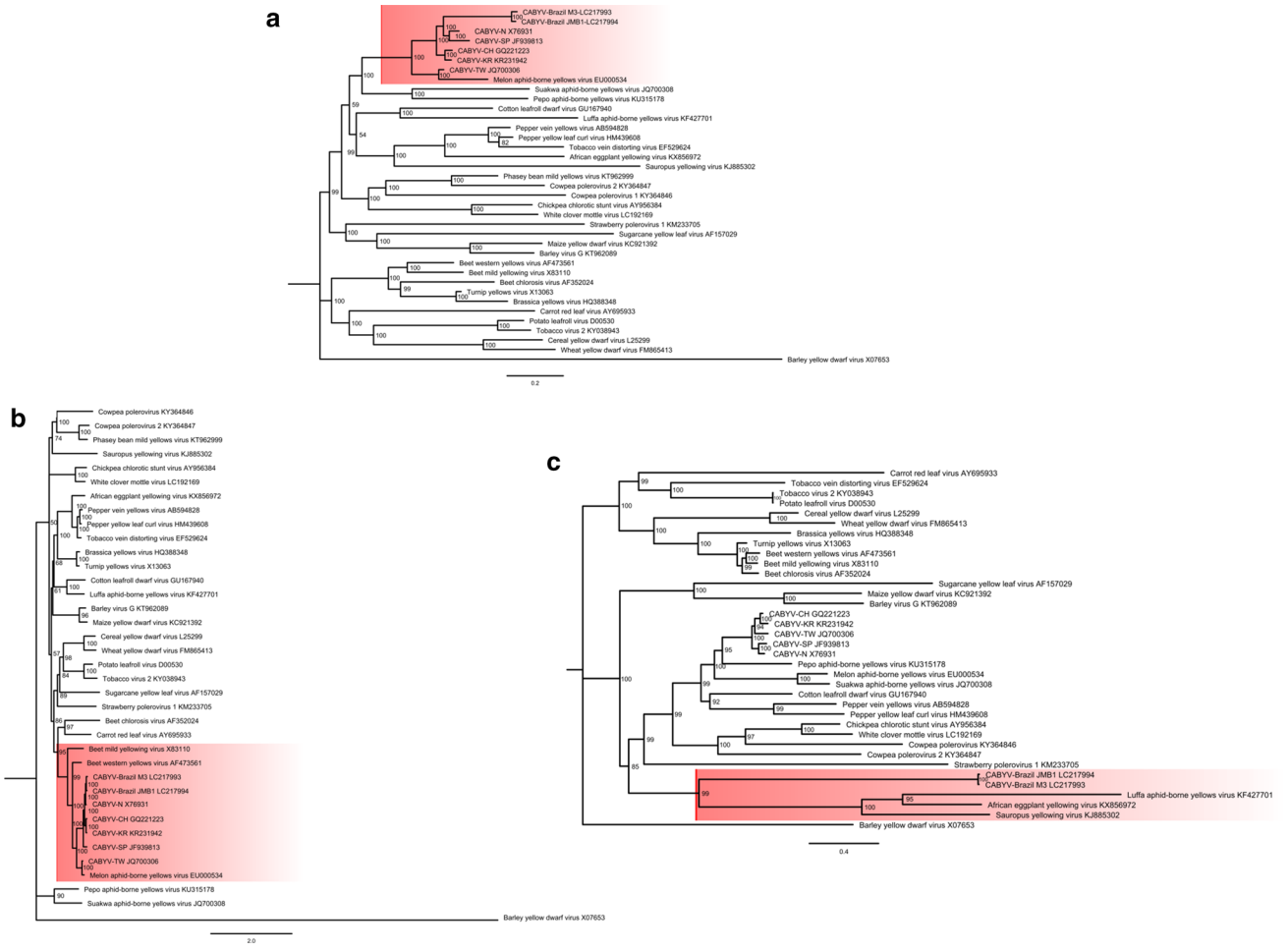

Fig. 1 Phylogenetic tree of poleroviruses. Tree based on complete genome nucleotide sequences (a). Barley yellow dwarf virus (BYDV) is used as outgroup. Trees based on the amino acid sequences of the fusion proteins P1-P2 (b) and P3-P5 (c). The amino acid sequences were obtained from the same accession numbers as those used in (a). The clusters in which cucurbit aphid-borne yellows virus (CABYV) M3 and JBM1 sequences are highlighted in red color. Scale bar = expected number of substitutions per site

nt 5678 in the 3′ UTR (Fig. 2). The putative recombination region represents ~ 40% of the genome. Due to this recombination, the isolates from Brazil had a longer 3′ UTR. The last 59 nt had a common origin, sharing 89.8% nt identity with the CABYV-N sequence (Fig. 2A). This recombination event was detected by all methods as RDP (p-value of 7.216×10^{-104}), GENECONV (p-value of 2.210×10^{-65}) and MaxChi (p-value of 7.311×10^{-56}). The JMB1 isolate shows similar results (p-value of 7.216×10^{-104} for RDP, GENECONV 2.210×10^{-65}, and MaxChi 1.665×10^{-56}). The minor parent was not identified by Blast search. The 5′ region of the genome (nt 1-3376, position in the M3 sequence) shares 93.6% identity with CABYV-N, whereas the 3′ recombinant region (nt 3377–5677) shares 48.9% with CABYV-N (the highest identity of 50.3% with pepper yellow leaf curl virus), and the 3′ end (nt 5678-5736) shares 89.8% with CABYV-N (Fig. 2A).

CABYV P0 gene was used for BEAST analysis. The correlation between root-to-tip divergence and sampling time inferred by TemPest for the nucleotide data set generated an r^2 of 0.30, a correlation coefficient of 0.55, and a residual mean r^2 of $1.13E^{-3}$, indicating a detectable signal of sequence divergence during the sampling time. The BEAST analysis found that the mean evolutionary rates for amino acids varied from $3.81E^{-3}$ to $4.54E^{-4}$ substitutions/site/year (mean = $2.081E^{-3}$), within the same range for nucleotides ($2.59E^{-3}$ to $2.79E^{-4}$, mean = $1.28E^{-3}$). The maximum clade credibility tree for the P0 protein of CABYV identified two clusters that emerged ~ 70 years ago (Fig. 3). The analysis indicated that the CABYV isolates from Brazil possibly came from a Spanish (location probability of 0.6) ancestral viral population that emerged ~ 68 years ago and the CABYV M3 and JMB1 isolates diverged ~ 10 years ago. However, the low number of P0 sequences may have hindered the accurate identification of geographic origins. The most recent common ancestor proposed for the P0 protein originated 125.5 (mean) years ago in Spain. These results are in agreement with the proposed diversification within the *Luteoviridae* during the last 500 years [21]. The evolutionary reconstruction with spatiotemporal diffusion was visualized using SPREAD (Fig. S1).

When phylogenetic and molecular-clock analyses were performed, it became clear that the 5′ genome region possibly originated from CABYV isolates from Europe. As already mentioned, the most recent common ancestor of two Brazilian isolates occurred ~ 10 years ago. By this time, the recombination had likely already occurred, because these two isolates possessed the same recombination event and showed high nucleotide identities in the minor parent region of the genome (3′ region). Interspecific recombination is common in poleroviruses, and examples have been described in the cucurbit-infecting poleroviruses. A

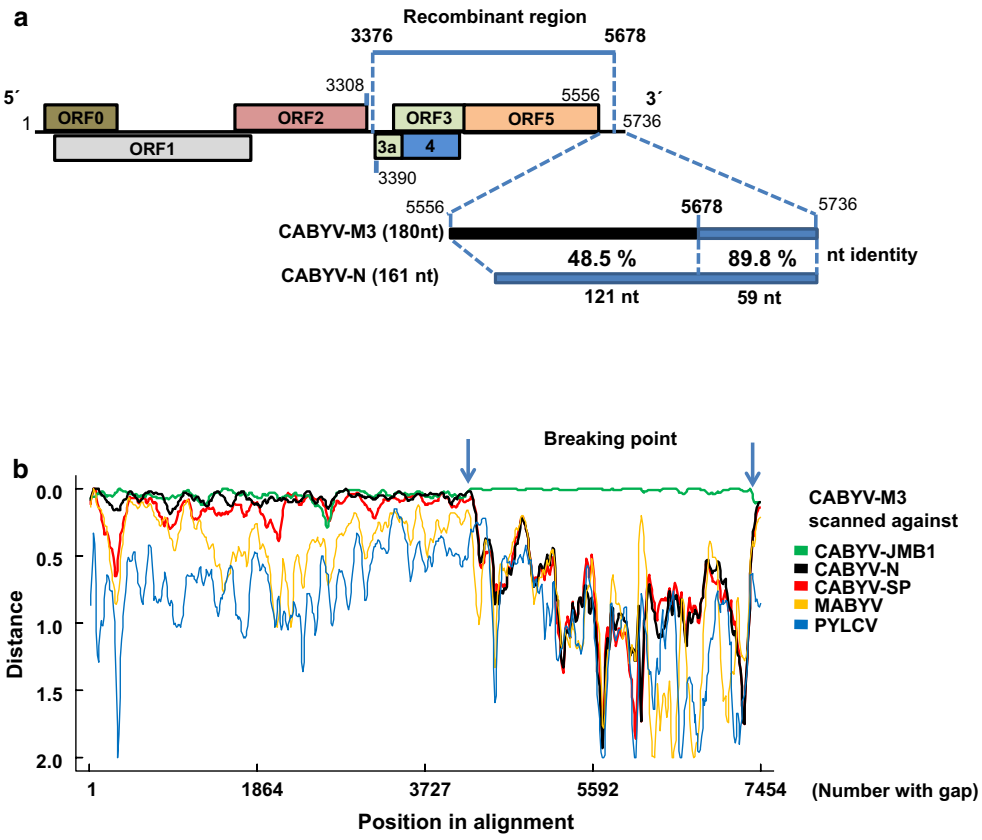

Fig. 2 The genome organization of the cucurbit aphid-borne yellows virus (CABYV) M3 with recombination sites (a) and the distance plot to selected poleroviruses (b). pepper yellow leaf curl virus (PYLCV) was selected as a outlier sequence in Fig 2b

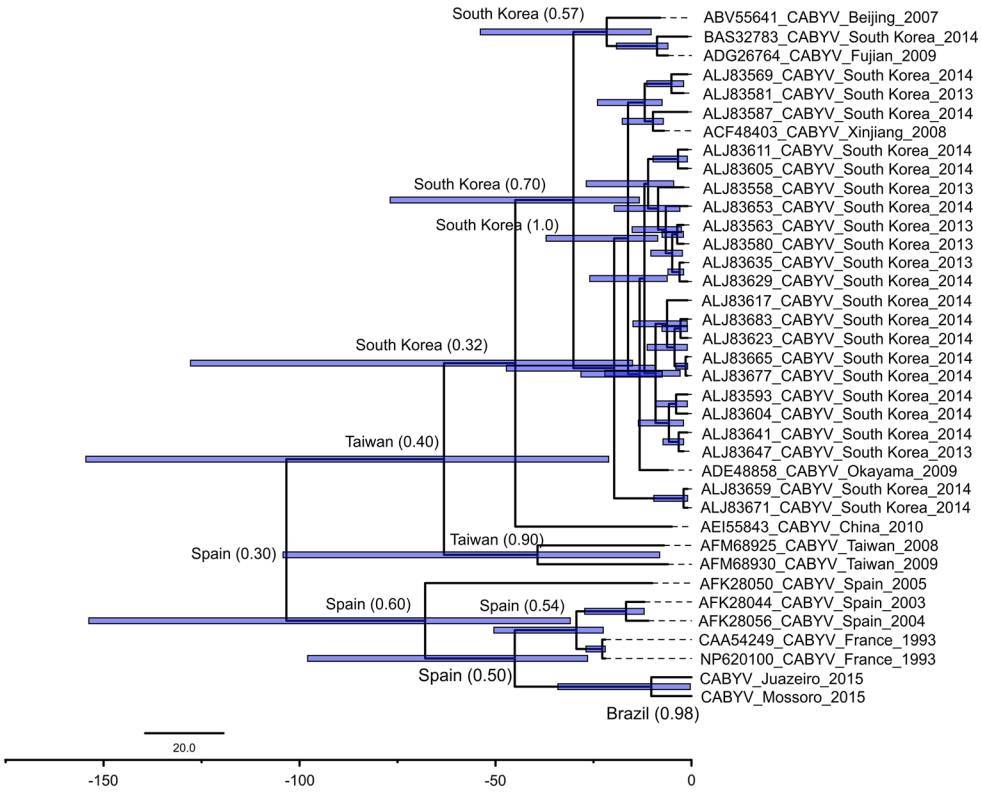

Fig. 3 Bayesian phylogeny resulting from the molecular-clock BEAST analysis. Number at the base of nodes denote the location probabilities. Node bars (blue) illustrate the extent of the 95% highest posterior density intervals (height-95%-HPD). Scale bar corresponds to the expected number of substitutions per site. The scale time is in years

recombinant between CABYV and MABYV in the intergenic region (IR) between ORF2 and ORF3 was reported in Taiwan [20]. Recombinations of SABYV with MABYV [22] and PABYV with CABYV [23], both in the intergenic region, were later reported. Furthermore, LABYV may have originated from a recombination between SABYV and cotton leafroll dwarf-like virus based on our analysis (data not shown), although it is not clearly stated in the original report [24]. Both inter- and intraspecific recombination have been reported for CABYV isolates [25]. Hence, it seems the recombination is an important evolutionary force for poleroviruses. Whether the recombination occurred before or after the virus entered Brazil is not clear, because this type of recombination has not been reported outside this country.

We now think that this CABYV strain may be the main cause of the melon severe yellowing disease in Brazil, because only a few reads corresponding to MYaV were detected in the HTS analysis of the samples (8,624 reads of MYaV, compared to 1,276,525 reads of CABYV in Mossoró samples, and none for Juazeiro samples). We are presently conducting a detailed study of the occurrence of CABYV in melons with the yellowing disease.

Acknowledgements This work was supported by CNPq with the project number of 401755/2013-4. MAA work was supported by grant AGL2015-65838 (Ministerio de Economía, Industria y Competitividad, Spain). AKIN and TN are CNPq fellows.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflicts of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Santos AA, Cardoso JE, Bezerra MA et al (2008) Progress analysis and damages due to melon yellowing-associated virus. Summa Phytopathol. 34:359–360
2. Costa TM, Blawid R, Costa AC Jr. et al (2017) Complete genome sequence of melon yellowing-associated virus from melon plants with the severe yellowing disease in Brazil. Arch Virol 162:3899–3901
3. Lecoq H, Bourdin D, Wipf-Scheibel C (1992) A new yellowing disease of cucurbits caused by a luteovirus, curcubit aphid-born yellow virus. Plant Pathol 41:749–761
4. Mnari-Hattab M, Kummert J, Roussel S et al (2005) First report of Cucurbit aphid-borne yellows virus in Tunisia causing yellows on five cucurbitacious species. Plant Dis 89:776
5. Kassem MA, Sempere RN, Juarez M et al (2007) Cucurbit aphid-borne yellows virus is prevalent in field-grown cucurbit crops of southeastern Spain. Plant Dis 91:232–238
6. Blawid R, Silva JMF, Nagata T (2017) Discovering and sequencing new plant viral genomes by next-generation sequencing: description of a practical pipeline. Ann Appl Biol 170:301–314

7. Nicolini C, Inoue-Nagata AK, Nagata T (2015) Complete genome sequence of a proposed new tymovirus, tomato blistering mosaic virus. Arch Virol 160:609–612
8. Ronquist F, Teslenko M, van der Mark P et al (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol 61:539–542
9. Darriba D, Taboada GL, Doallo R et al (2012) jModelTest 2: more models, new heuristics and parallel computing. Nat Methods 9:772
10. Darriba D, Taboada GL, Doallo R et al (2011) ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27:1164–1165
11. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52:696–704
12. Martin DP, Murrel B, Golden M et al (2015) RDP4: detection and analysis of recombination patterns in virus genomes. Virus Evol 1:vev003
13. Posada D (2002) Evaluation of methods for detecting recombination from DNA sequences: empirical data. Mol Biol Evol 19:708–717
14. Rambaut A, Lam TT, Carvalho LM et al (2016) Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol 2:vew007
15. Drummond AJ, Suchard MA, Xie D et al (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol 29:1969–1973
16. Bielejec F, Rambaut A, Suchard MA et al (2011) SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. Bioinformatics 27:2910–2912
17. Csorba T, Lózsa R, Hutvágner G et al (2010) *Polerovirus* protein P0 prevents the assembly of small RNA-containing RISC complexes and leads to degradation of ARGONAUTE1. Plant J 62:463–472
18. Guilley H, Wipf-Scheibel C, Richards K et al (1994) Nucleotide sequence of cucurbit aphid-borne yellows luteovirus. Virology 202:1012–1017
19. Smirnova E, Firth AE, Miller WA et al (2015) Discovery of a small non-AUG-initiated ORF in poleroviruses and luteoviruses that is required for long-distance movement. Plos Pathog 11:e1004868
20. Knierim D, Deng TC, Tsai WS et al (2010) Molecular identification of three distinct Polerovirus species and a recombinant Cucurbit aphid-borne yellows virus strain infecting cucurbit crops in Taiwan. Plant Pathol 59:991–1002
21. Pagán I, Holmes EC (2010) Long-term evolution of the Luteoviridae: time scale and mode of virus speciation. J Virol 84:6177–6187
22. Knierim D, Tsai WS, Deng TC et al (2013) Full-length genome sequences of four polerovirus isolates infecting cucurbits in Taiwan determined from total RNA extracted from field samples. Plant Pathol 62:633–641
23. Ibaba JD, Laing MD, Gubba A (2016) Pepo aphid-borne yellows virus: a new species in the genus *Polerovirus*. Virus Genes 53:134–136
24. Knierim D, Maiss E, Kenyon L et al (2015) First full-length genome sequence of the polerovirus luffa aphid-borne yellows virus (LABYV) reveals the presence of at least two consensus sequences in an isolate from Thailand. Arch Virol 160:2633–2636
25. Kassem MA, Juarez M, Gómez P et al (2013) Genetic diversity and potential vectors and reservoirs of Cucurbit aphid-borne yellows virus in southeastern Spain. Phytopathology 103:1188–1197