

## Genetic Diversity and Potential Vectors and Reservoirs of *Cucurbit aphid-borne yellows virus* in Southeastern Spain

Mona A. Kassem, Miguel Juarez, Pedro Gómez, Carmen M. Mengual, Raquel N. Sempere, María Plaza, Santiago F. Elena, Aranzazu Moreno, Alberto Fereres, and Miguel A. Aranda

First, third, fourth, fifth, and tenth authors: Centro de Edafología y Biología Aplicada del Segura (CEBAS), Consejo Superior de Investigaciones Científicas (CSIC), apdo. correos 164, 30100 Espinardo, Murcia, Spain; second author: Escuela Politécnica Superior de Orihuela, Universidad Miguel Hernández de Elche, Ctra. de Beniel km 3.2, 03312 Orihuela, Alicante, Spain; seventh author: Instituto de Biología Molecular y Celular de Plantas (IBMCP), CSIC-Universidad Politécnica de Valencia, Ingeniero Fausto Elio s/n, 46022 Valencia, Spain, and The Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501; and sixth, eighth, and ninth authors: Instituto de Ciencias Agrarias (ICA), CSIC, Serrano 115, 28006, Madrid.

Current address of P. Gómez: Centre for Ecology and Conservation, School of Biosciences, University of Exeter, Cornwall Campus, Penryn TR10 9EZ, UK.

Accepted for publication 23 May 2013.

### ABSTRACT

Kassem, M. A., Juarez, M., Gómez, P., Mengual, C. M., Sempere, R. N., Plaza, M., Elena, S. F., Moreno, A., Fereres, A., and Aranda, M. A. 2013. Genetic diversity and potential vectors and reservoirs of *Cucurbit aphid-borne yellows virus* in southeastern Spain. *Phytopathology* 103:1188-1197.

The genetic variability of a *Cucurbit aphid-borne yellows virus* (CABYV) (genus *Polerovirus*, family *Luteoviridae*) population was evaluated by determining the nucleotide sequences of two genomic regions of CABYV isolates collected in open-field melon and squash crops during three consecutive years in Murcia (southeastern Spain). A phylogenetic analysis showed the existence of two major clades. The sequences did not cluster according to host, year, or locality of collection, and nucleotide similarities among isolates were 97 to 100 and 94 to 97% within and between clades, respectively. The ratio of nonsynonymous to synonymous nucleotide substitutions reflected that all open reading

frames have been under purifying selection. Estimates of the population's genetic diversity were of the same magnitude as those previously reported for other plant virus populations sampled at larger spatial and temporal scales, suggesting either the presence of CABYV in the surveyed area long before it was first described, multiple introductions, or a particularly rapid diversification. We also determined the full-length sequences of three isolates, identifying the occurrence and location of recombination events along the CABYV genome. Furthermore, our field surveys indicated that *Aphis gossypii* was the major vector species of CABYV and the most abundant aphid species colonizing melon fields in the Murcia (Spain) region. Our surveys also suggested the importance of the weed species *Echallium elaterium* as an alternative host and potential virus reservoir.

*Additional keywords:* *Cucumis melo*, *Cucurbita pepo*, gene flow, luteovirus.

*Cucurbit aphid-borne yellows virus* (CABYV) is a member of the genus *Polerovirus* within the family *Luteoviridae*. CABYV was first described in 1992 in France (26) but it was later detected infecting cucurbits in many other countries in the world (2,23,26,28,54). In addition to cucurbits, CABYV can infect other crop species, including lettuce (*Lactuca sativa*) and fodder beet (*Beta vulgaris*), as well as common weeds of the species *Capsella bursa-pastoris*, *Lamium amplexicaule*, *Senecio vulgaris*, *Echallium elaterium*, *Bryonia dioica*, *Papaver rhoeas*, *Montia perfoliata*, and *Crambre abyssinica*, which are thought to be virus reservoirs. CABYV is a phloem-limited virus that is transmitted in a persistent, nonpropagative mode by aphids, including *Aphis gossypii* and *Myzus persicae* (26). Typical symptoms of CABYV infection include yellowing and thickening of basal and older leaves. In melon and cucumber, a high percentage of flower abortion was detected, with no effects on fruit quality (26).

CABYV particles are isometric, ≈25 nm in diameter, and encapsidate the CABYV genome, which consists of a single-stranded positive-sense RNA molecule 5.7 kb in length (10). The CABYV genome contains at least six open reading frames (ORFs) organized into two clusters (33). The first three (ORFs 0, 1, and 2) are expressed from the genomic RNA, whereas the other three (ORFs 3, 4, and 5) are expressed from a sub-genomic RNA (sgRNA1). Two additionally predicted ORFs (ORFs 6 and 7) might be expressed from a second sgRNA (sgRNA2) (1). The protein encoded by ORF 0 (P0) is a strong silencing suppressor and an enhancer of pathogenicity (42,43). The proteins, putatively encoded by ORF 1 (P1) and ORF 2 (P2 domain in P1-2; see below), have regions of amino acid sequence similarity with serine proteases and genome-linked viral proteins (VPgs) of other poleroviruses, as well as amino acid motifs typical of RNA-dependent RNA polymerases (RdRps) (15), respectively. By analogy with other poleroviruses, it is likely that CABYV ORF 2 is translated as a result of a ribosomal frameshift in the C-terminal portion of ORF 1 to generate a P1-2 fusion protein (15). ORFs 3, 4, and 5 encode proteins P3 (the coat protein [CP]), which is involved in viral transmission, particle packaging, and viral accumulation within the plant; P4 (the movement protein [MP]); and P3-5. P3-5 is expressed only as a read-through protein, and is needed for aphid transmission of the virus (6,7,14,56).

Corresponding author: M. A. Aranda; E-mail address: m.aranda@cebas.csic.es

\*The e-Xtra logo stands for "electronic extra" and indicates that the online version contains one supplemental figure and one supplemental table.

<http://dx.doi.org/10.1094/PHYTO-11-12-0280-R>

This article is in the public domain and not copyrightable. It may be freely reprinted with customary crediting of the source. The American Phytopathological Society, 2013.

Based on phylogenetic analysis, CABYV isolates have been classified into the Asian and Mediterranean groups (46). Manarri-Hattab et al. (34) further distinguished Tunisian, Italian, and French isolates from Spanish and Chinese isolates, and suggested that gene flow could have had a fundamental role in this geographically based distribution. Among poleroviruses, very low genetic diversity was described for *Potato leafroll virus* (16) and *Cotton leafroll dwarf virus* (47), whereas *Barley yellow dwarf virus* (BYDV)-PAV (4,31,52) and viruses belonging to the beet polerovirus complex (18) are known to be more variable. Recombination may have had a fundamental role in maintaining luteovirus diversity, as appears to be the case for CABYV, where the continuous generation of recombinant strains has led to great variability in the Asian group (24,46,53). Recombination within members of the family *Luteoviridae* has been extensively analyzed, showing that recombination break points are distributed along various regions of the virus genomes, although the main putative recombination hotspot seems to lie between the end of ORF 2 and the beginning of ORF 3 (24,38). Pagán and Holmes (38) have shown that luteovirus genes appear to be subjected to purifying selection, except for the *MP* gene that seems to be evolving neutrally in the majority of cases. Torres et al. (50) have shown that the *CP* gene of luteoviruses appears to be under heterogeneous selective forces; furthermore, highly conserved sites under purifying selection were identified mainly on the CP surface, perhaps associated with constraints imposed by the conservation of receptors in the aphid and the host plant (50).

CABYV has been shown to be the prevalent virus in cucurbit crops in southeastern Spain (23). With this study, our goal was to understand the epidemiology and population genetics of CABYV. We focused on an ecologically confined area, Campo de Cartagena (Murcia, Spain), where cucurbit cultivation has high economic importance. Thus, we examined the genetic variability of a CABYV population sampled in this area during three consecutive years by sequencing two genomic regions of CABYV isolates; the full-length sequence of three of these isolates was also determined to identify occurrence and location of recombination events along the CABYV genome. In addition, we studied the dynamics of aphid populations and sampled aphids landing on melon crops (vector activity) in this same geographical area to determine the proportion of viruliferous aphids and also which species were potentially acting as major vectors of CABYV (vector propensity) (20). Finally, we also sampled weeds and surrounding crops and determined the presence of CABYV to identify potential alternative hosts and virus reservoirs.

## MATERIALS AND METHODS

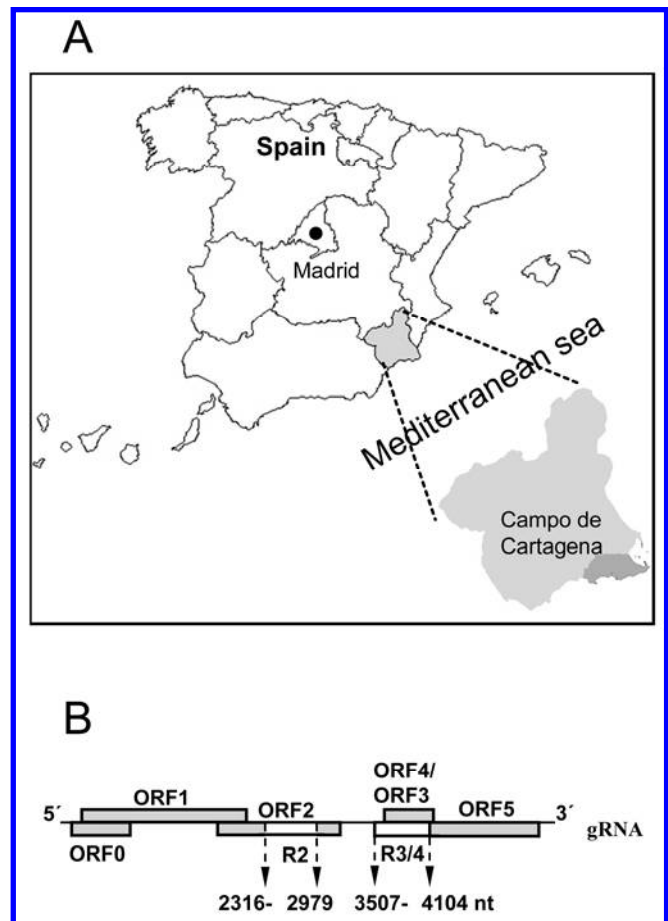
**CABYV surveys, virus detection and isolates.** Surveys were performed in open-field melon (*Cucumis melo*) and squash (*Cucurbita pepo*) crops of the Murcia (Spain) province during 2003, 2004, and 2005, as described by Kassem et al. (23). During these 3 years, >1,000 samples were obtained from symptomatic plants from 48 randomly distributed fields in a geographical area of ≈300 km<sup>2</sup> (Fig. 1A). All samples were analyzed to detect the presence of CABYV by dot-blot hybridization. Of the total, ≈80% of samples were positive for CABYV. CABYV isolates analyzed in this study were chosen randomly from this collection. The same geographical area was surveyed during 2009 to identify alternative hosts and potential virus reservoirs. In this case, 239 samples were collected from 28 weed and 6 alternative crop species (see below).

**Partial sequencing of CABYV isolates.** Complementary DNAs (cDNAs) to two regions of the CABYV genome (Fig. 1B) were synthesized by reverse-transcription polymerase chain reaction (RT-PCR). The primer pairs CE-124 (5'-CTCCTTCGATA TTGGCTCG-3') and CE-123 (5'-CCCATTCTGCGCCGAGT GG-3'), and CE-9 (5'-GAATACGGTCGCGGCTAGAAATC-3')

and CE-10 (5'-C TATTCGGGTTCTGGACCTGGC-3') (23) were designed based on the nucleotide sequence of isolate CABYV-N (GenBank accession number NC003688) to amplify fragments of nucleotides 2,316 to 2,979 of ORF 2 and nucleotides 3,507 to 4,104 of ORF 3, respectively (Fig. 1B). Note that ORF 3 overlaps ORF 4 for most of its sequence (nucleotides 3,534 to 4,104). RT and PCR were carried out using Expand Reverse Transcriptase (Roche, Mannheim, Germany) and Expand High Fidelity PCR System (Roche), respectively, following the recommendations of the manufacturer. RT-PCR products were purified from agarose gels after electrophoresis by using GeneClean turbo columns (MP Biomedicals, Illkrich, France) and then sequenced (Secugen, Madrid) using the same primers.

Sequences of ORF 3 from worldwide isolates were retrieved from GenBank (<http://www.ncbi.nlm.nih.gov/>). Accession numbers of the corresponding sequences are listed in the corresponding figures or tables.

**Phylogenetic and diversity analysis.** Nucleic and amino acid sequences were aligned using ClustalX and BioEdit software ([www.mbio.ncsu.edu/bioedit](http://www.mbio.ncsu.edu/bioedit)). To calculate the percentages of nucleotide and amino acid identity and similarity, GeneDoc Editor v 2.7.000 software was used. Further sequence and cluster analyses were performed with the MEGA4 package. Phylogenetic trees were constructed by using the minimum-evolution (ME) method, using distance matrices based on the Kimura 2-parameter (K2P) method. A bootstrap value for each internal node of the tree was calculated by using 1,000 random pseudoreplicates. Maximum



**Fig. 1.** Sampling area and regions sequenced in the *Cucurbit aphid-borne yellows virus* (CABYV) genome. **A**, Map of the area where CABYV isolates were sampled. **B**, Schematic representation of the CABYV genome; the two regions sequenced are indicated by R2, a fragment of open reading frame (ORF) 2 located between nucleotides 2,316 and 2,979, and R3/4, corresponding to the ORF 3 and its overlapping part of ORF 4, located between nucleotides 3,507 and 4,104.

likelihood (ML) and maximum parsimony trees were constructed using programs from the phylogeny inference package PHYLIP. The mean genetic distance for each gene was calculated using the K2P distance as implemented in the MEGA4 package. The ratio of nonsynonymous ( $d_N$ ) to synonymous ( $d_S$ ) nucleotide substitutions ( $\omega = d_N/d_S$ ) for each ORF was calculated using the Pamilo-Bianchi-Li method. The MEGA4 package was also used to perform Tajima's  $D$  statistical tests and to compute the total nucleotide diversity  $\pi$ . Nucleotide diversity was estimated using the K2P distance estimator and was expressed as the average number of nucleotide substitutions per site in each pair of sequences. To assess selection pressures, an ML approach was used. In this case,  $d_N$  and  $d_S$  differences that correlated with phylogenetic relationships were estimated using the HyPhy package (<http://www.hyphy.org>). The  $\omega$  ratio was calculated by the MG94 model as implemented in HyPhy.

A Bayesian Markov Chain Monte Carlo approach, implemented in the Bayesian tip-associated significance testing (BaTS) software, was employed to explore the overall structure of the sequence dataset in relation with host adaptation. This analysis was based on the trees that were constructed by using the BEAST program, version 1.4.8 (<http://beast.bio.ed.ac.uk>), removing 10% of burn-ins and using 1,000 replicates. From these trees, the significance of the parsimony score and association index statistics was computed to assess the strength of host clustering in the phylogeny. The Path-O-Gen software (<http://tree.bio.ed.ac.uk/software/pathogen/>) was employed to explore the potential correlation between genetic structure and year of sampling. ML trees were randomly generated using the best substitution models HKY+G+I (for ORFs 2 and 3) and GTR+G (for ORF 4). A Mantel test was used to examine the correlation between geographical and genetic distances. Geographical distances were determined from the geographical coordinates of each sample by using the Google Earth software, version 5 (<http://www.google.com/earth/index.html>). Mantel's test was performed with the ZT software using 50,000 nonparametric permutations for significance testing.

**Complete sequencing of CABYV isolates.** To determine the full-length genome sequence of CABYV isolates, five overlapping cDNA fragments covering the entire length of the genome were amplified from three randomly chosen isolates. The five primer pairs used were CE-56 (5'-ACAAAAGATACGAGCGG GTGATG-3')/CE-92 (5'-GCATACCCGGCATGCGATCCGTC-3') (nucleotides 1 to 823), CE-93 (5'-GGCGACTTGCCAATT TCCCT-3')/CE-95 (5'-GAACTGGGGCCACCCGAAGC-3') (nucleotides 646 to 2,016), CE-94 (5'-GCCATGAAGAAACCG AGGTCGCG-3')/CE-96 (5'-GGCGTCGGCGGTTTCTTCG-3') (nucleotides 1,762 to 3,649), CE-9/CE-10 (nucleotides 3,507 to 4,104) (23), and CE-97 (5'-GGACCCCTGCAAGCTTAGC-3')/CE-57 (5'-ACACCGAAACGCCAGGGGGAATC-3') (nucleotides 3,885 to 5,669). All cDNAs were sequenced in both directions using the adequate primers. The assembly of the CABYV genomic sequences was carried out using BioEdit.

**Recombination and other sequence analyses.** Putative recombinant genomes were evaluated using the RDP3 package (<http://web.cbio.uct.ac.za/~darren/rdp.html>). RDP3 implements six recombination detection programs: RDP, GENECONV, Maximum  $\chi^2$  (MaxChi), Chimera, BootScan, and SisterScan (SiScan). The default detection thresholds were used. Events detected by more than two methods were retained.

RNA secondary structure predictions were performed using the RNAalifold program included in the Vienna RNA package v1.8.2 (<http://rna.tbi.univie.ac.at>). The transmembrane regions were predicted by using the TMHMM (<http://www.cbs.dtu.dk/services>) and TMPred (<http://www.ch.embnet.org>) programs. Protein motifs and domains were identified by using motif libraries included in the Prosite, PRINTS, ProDom, and Pfam databases (<http://motif.genome.jp>).

**Aphid population dynamics and estimation of CABYV-viruliferous aphids landing in melon fields.** Aphids were sampled on six randomly distributed melon fields in the geographical area where the virus surveys were carried out. The sampling procedure was designed to estimate the number of aphids alighting on the crop (vector activity) and the number of aphids colonizing the crop (population density) at different time intervals during the melon growth cycle. Aphids alighting on the selected experimental plots were monitored weekly with a horizontal mosaic-green tile trap (19) during the spring-summer melon-growing season of two consecutive years, 2010 and 2011. Aphids from trap samples were separated in the laboratory using a stereomicroscope (Nikon Systems Europe, Amstelveen, The Netherlands). All aphid specimens were preserved in 70% ethanol and they were prepared and identified following the protocol described by Pérez et al. (41). Additionally, aphid populations colonizing the crop (alate adults, apterae adults, and nymphs) were monitored weekly in the six selected melon fields during summer 2010. Four rows were selected in the sampled plots and 5 plants were chosen within each of the rows for a complete total of 20 plants sampled each time. All leaves of each sampled plant were visually inspected and the numbers and morph of aphids of each species were recorded.

A sample of the aphid species caught in the field from the horizontal mosaic-green tile traps in 2011 was used to estimate the number of CABYV-viruliferous aphids visiting melon fields. Individual aphids were analyzed by quantitative real-time RT-PCR according to Moreno et al. (35).

**CABYV transmission from *E. elaterium* to melon.** To confirm the capacity of *E. elaterium* to act as a virus reservoir and infective source for the transmission of CABYV to melon crops, eight *E. elaterium* plants were agroinoculated with CABYV (6). Twelve days post inoculation, plants were visually inspected and tested (see above) to detect CABYV. Transmission assays were carried out using *E. elaterium*-infected plants placed inside insect-proof boxes as a virus source. Three to eight 2-week-old melon seedlings were introduced inside the boxes together with one CABYV-source *E. elaterium* plant. Then,  $\approx 100$  virus-free *A. gossypii* adults were released into the boxes, which were then maintained for >2 weeks in an insect-proof glasshouse. Virus-free *A. gossypii* colonies were reared as described by Moreno et al. (35). Test melon seedlings were visually inspected for CABYV symptoms appearance, and presence of CABYV (see above) was determined 12 days after aphid release.

## RESULTS

**Phylogenetic relationships among CABYV isolates.** Surveys were carried out in field plots used for commercial intensive production during the 2003, 2004, and 2005 growing seasons (23) (Fig. 1A). We sequenced two genomic fragments, one in ORF 2 and another in the overlapping ORFs 3 and 4 (Fig. 1B), for 50 isolates (Table 1), covering  $\approx 21\%$  of the CABYV genome. To analyze the phylogenetic relationships among CABYV isolates, a matrix of genetic distances among isolates was computed after the sequences were aligned. Each genomic region was used independently to infer phylogenetic trees by various methods, including ME, maximum parsimony, and ML; topologies inferred by the different methods were all similar. For both regions, two major clades were identified, dividing the population into groups Ia and Ib for ORF 2 (Fig. 2A) and IIa and IIb for ORFs 3 and 4 (Fig. 2B). Nucleotide similarities among isolates of each group were 97 to 100%, whereas nucleotide similarities among isolates from different groups were 94 to 97%. These phylogenies did not reveal any obvious association of host (i.e., melon or squash) or year of collection with phylogenetic grouping (Fig. 2). Indeed, the statistical methods implemented in BaTS and Path-O-Gen found no significant association between host species and phylogenetic

groups, or divergence time and genetic diversity, respectively (data not shown). To analyze whether the geographic origin of isolates could possibly be associated with genetic grouping, we looked for potential correlations between geographic and genetic distances among pairs of isolates; however, no significant correlation was detected (Mantel test,  $P > 0.1$ ). We then tested whether the inclusion of worldwide CABYV isolates would influence the resulting tree topology. This was done only for the region

covering ORFs 3 and 4, because databases contained a significant number of sequences for this region but not for other CABYV genomic regions. In the corresponding deduced ME tree, Asian and Mediterranean isolates clustered in different groups (Fig. 3). The Spanish isolates were in the two clades containing Mediterranean isolates, one with Spanish but also with isolates from other Mediterranean regions (Fig. 3, group IIa) and the other with Spanish isolates only (Fig. 3, group IIb). Therefore, the CABYV genetic structure suggested by phylogenetic analyses appeared to be related to geography only when the sampling scale was large enough but not at a local-scale.

TABLE 1. *Cucurbit aphid-borne yellows virus* isolates analyzed in this study, grouped by host and year of collection

Host	Year			Total
	2003	2004	2005	
Melon	12	11	11	34
Squash	7	7	6	20
Total	19	18	17	54

**Genetic diversity and selective pressures in non-overlapping and overlapping CABYV genes.** Values of nucleotide diversity were estimated independently for each of the ORFs analyzed and were 0.00 to 0.05 (Table 2). A pairwise comparison among means showed that nucleotide diversity estimated for ORF 2 was significantly higher than for ORF 3 ( $t = 10.897$ ,  $P < 0.0001$ ) and ORF 4 ( $t = 10.263$ ,  $P < 0.0001$ ), and that there were no differ-

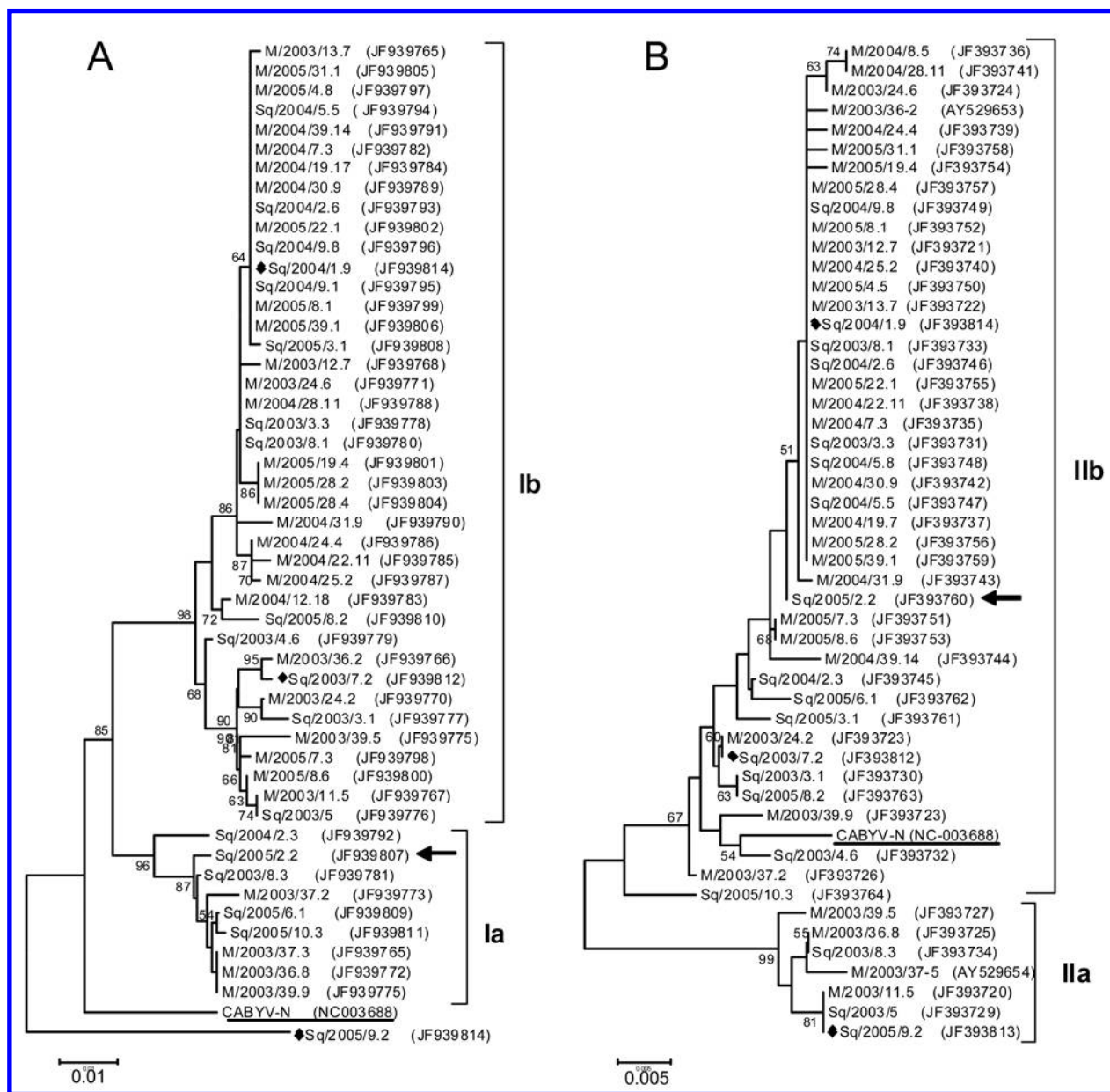


Fig. 2. Phylogenetic relationships of 50 Spanish *Cucurbit aphid-borne yellows virus* (CABYV) isolates inferred from sequences from A, open reading frame (ORF) 2 and B, ORFs 3 and 4. Minimum evolution phylogenetic trees were inferred using the Mega 4 program. Bootstrap values (1,000 pseudoreplicates)  $> 50\%$  are shown. The sequence of a French isolate was introduced as an outgroup and appears underlined in each tree. Isolate Sq/2005/2.2 (marked with an arrow) belongs to group Ia for the ORF 2 but to group IIb for the ORFs 3 and 4. Isolates fully sequenced are marked with a diamond.



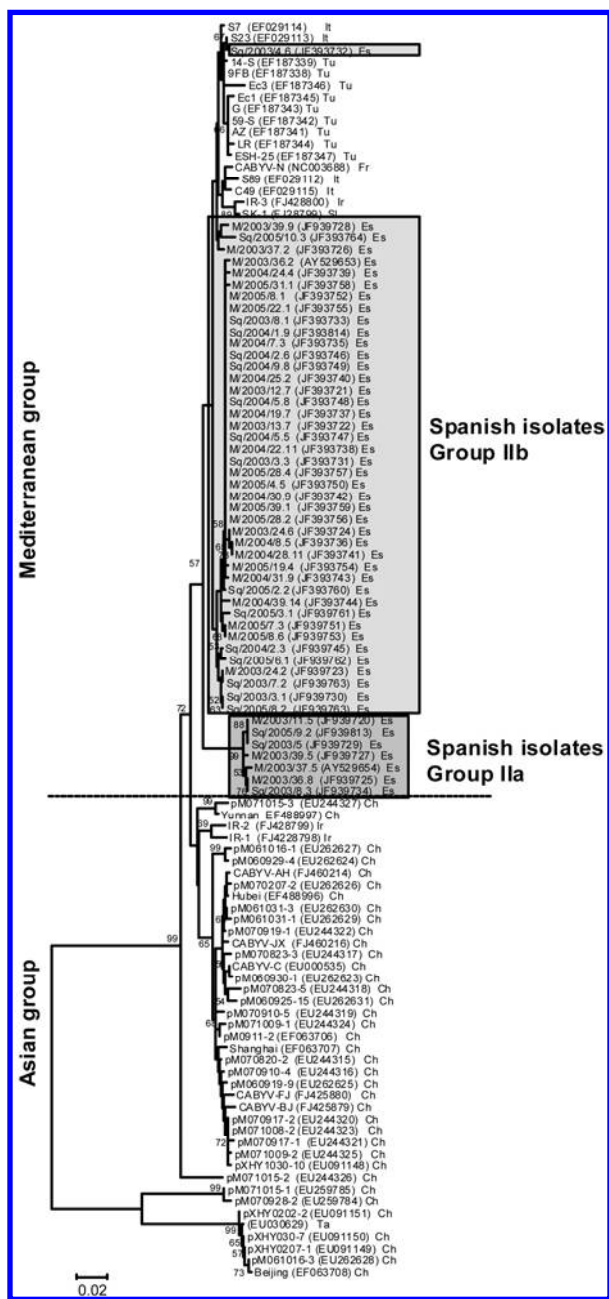
ences between ORFs 3 and 4 ( $t = 0.122$ ,  $P < 0.906$ ), indicating that variability was not evenly distributed throughout the viral genome. Furthermore, we also estimated the relative rates of change at  $d_N$  and  $d_S$  positions for each ORF by the use of  $\omega = d_N/d_S$ , in

order to assay the direction and strength of the selection pressure acting on coding regions within CABYV populations. We found a low ratio ( $\omega < 1$ ) for the three ORFs (Table 2), indicating that purifying selection was restricting variability for all three ORFs. Moreover, the  $\omega$  ratio of ORF 2 was approximately one order of magnitude smaller than for the other two ORFs, suggestive of purifying selection being stronger for this ORF than for the other two. After further evaluating the selective constraints operating in each region, codons under selection were detected using the three methods implemented in HyPhy: SLAC, FEL, and REL. The best substitution models were K81 for ORF 2 and HKY85 for ORFs 3 and 4. We found no significant evidence of positive selection on codons in ORFs 2 and 3, and only one codon (position 145) was found to be significantly under positive selection in ORF 4 (Fig. 4). In contrast, several codons were found to be under negative selection but only six codons (positions 356, 409, 417, 429, 432, and 445) on ORF 2, five codons (positions 16, 64, 109, and 175) on ORF 3, and two codons (positions 12 and 99) on ORF 4 were detected by all three methods (Fig. 4). Alignment of CABYV amino acid sequences around the only codon subjected to positive selection (Supplemental Figure 1) showed that proline (P) co-existed with glutamine (Q) in this position in 2003, becoming prevalent in the following years.

In order to further analyze whether host adaptation could have contributed to genetic differentiation of the CABYV population, nucleotide diversity values were estimated within and between CABYV subpopulations, considering a subpopulation as the group of isolates that were originally collected from melon or squash. Between-subpopulation diversity values were of the same order of magnitude as within-subpopulation diversity values, and the  $\omega < 1$  in all cases (Table 3). The Tajima's  $D$  test was applied for each ORF, showing that  $D$  values were not significantly different than zero ( $P > 0.1$ ), indicating that the hypotheses of selective neutrality cannot be rejected and further suggesting that there is no differentiation of the population according to the host species from which the isolates were sampled.

**Complete genome sequencing of three CABYV isolates: recombination events in CABYV genomes.** Recombination has been shown to significantly contribute to luteovirus diversity. An analysis of potential recombination events within the two regions sequenced of the isolates characterized here did not reveal any significant result (data not shown). However, when comparing phylogenies for the two regions sequenced, a phylogenetic incongruence was detected for some isolates. For instance, isolate Sq/2005/9.2 was included in group Ia for the ORF 2 tree but in group IIb for the ORFs 3 and 4 tree (Fig. 2). To analyze this aspect in more detail and along the whole CABYV genome, we took three isolates at random and determined their full sequences.

Genomic RNAs of isolates Sq/2003/7.2 (GenBank accession number JF939812) and Sq/2004/1.9 (GenBank accession number JF939814) were both 5,672 nucleotides (nt) long, whereas genomic RNA of isolate Sq/2005/9.2 (GenBank accession number JF939813) contained a 3-nt insertion resulting in 5,675 nt in length. The genomic structure of the three isolates did not differ from that described for the other two, completely sequenced CABYV isolates, one from France (CABYV-N) (15) and the other from China (CABYV-C) (53). Similarly, all motifs previously identified in the CABYV putative proteins were also identi-



**Fig. 3.** Phylogenetic tree constructed based on sequences of the open reading frames (ORFs) 3 and 4 of 107 *Cucurbit aphid-borne yellows virus* isolates from all over the world, including the 50 Spanish isolates characterized here (gray shading) and using the minimum evolution method. Bootstrap values (1,000 pseudoreplicates) >50% are shown. Abbreviations: Ch = China, Sp = Spain, Fr = France, Sl = Slovakia, It = Italy, Tu = Tunisia, Ir = Iran, and Ta = Taiwan.

**TABLE 2.** Nucleotide diversities for different regions of the *Cucurbit aphid-borne yellows virus* genome<sup>a</sup>

Open reading frame (ORF)	$\pi$	$d_N$	$d_S$	$d_N/d_S$
ORF 2	0.0232	0.00288 ± 0.00110	0.04591 ± 0.00918	0.063
ORF 3	0.0152	0.00933 ± 0.00273	0.02233 ± 0.00595	0.417
ORF 4	0.0155	0.01105 ± 0.00274	0.02097 ± 0.00665	0.526

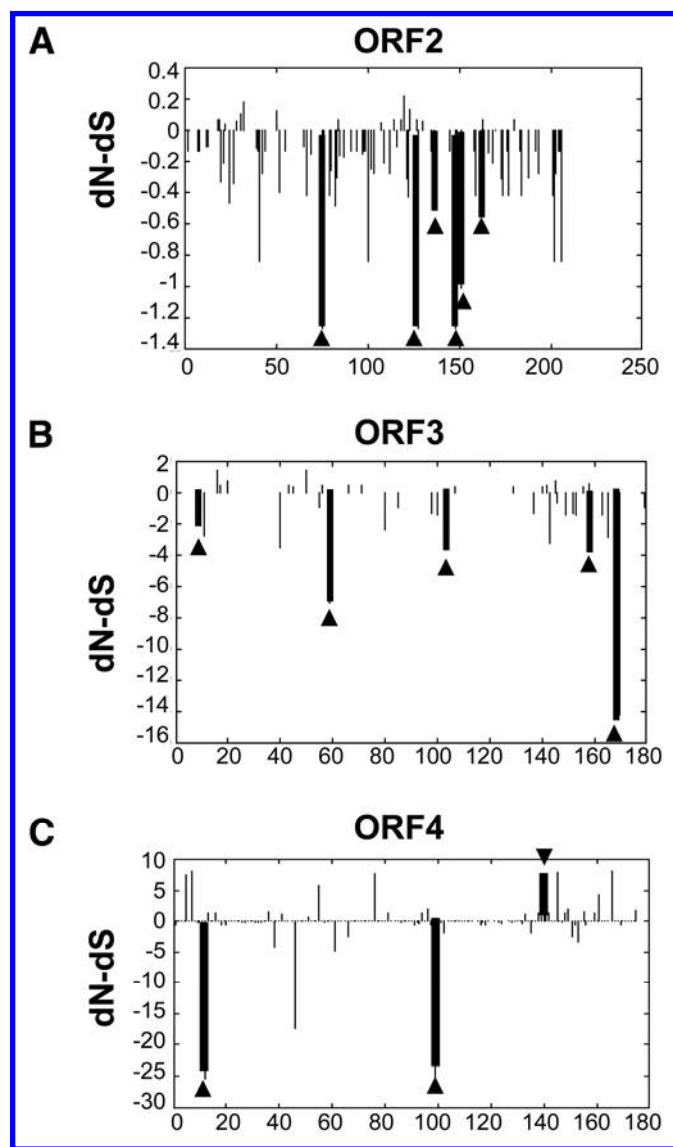
<sup>a</sup> Nucleotide diversity is defined here as the mean number of nucleotide substitutions per site ( $\pi$ ). Nucleotide diversities were computed separately for synonymous ( $d_S$ ) and nonsynonymous ( $d_N$ ) positions using the PBL method. Standard error of means was calculated by using the bootstrap method implemented in the Mega4 program.

fied in these three Spanish isolates, including the short motif LPLLI in P0 (39), the chymotrypsin-related serine protease and VPg motifs in P1 (45,51), three transmembrane regions in the amino-terminus of the P1 (32), the GDD motif in P1-2 (33), the arginine-rich RNA binding motif and epitopes 5 and 10 in P3 (CP) (5), and a proline-rich motif in RT (15). Isolates Sq/2003/7.2 and Sq/2004/1.9 shared 98% of their nucleotide sequence, whereas isolate Sq/2005/9.2 showed only a 91% nucleotide sequence identity with the other two isolates (Table 4). Nucleotide and amino acid identities among Sq/2003/7.2 and other poleroviruses were computed for the whole genome and independently for the six ORFs integrating the CABYV genome, showing that the less conserved is ORF 0 and the most conserved are ORFs 3 and 5; comparing different *Polerovirus* spp., the least conserved are ORFs 0 and 5 and the most conserved are ORFs 2 and 3 (Table 4).

Recombination events were identified using the RDP3 program, because it implements several recombination-detection methods. The sequences of CABYV-N and CABYV-C (15,53)

were also included in the analysis. In total, 29 potential recombination events were detected by at least one of the methods but only 3 events (Fig. 5A) were accepted based on a triple criteria: detection by more than one method, consensus score in RDP3 >40, and  $P < 0.01$ . In CABYV-C, the region between nucleotides 4,469 and 5,669, which forms part of ORF 5, may have originated by recombination of Sq/2004/1.9 (major parent) and an unknown minor parent, with the event detected by methods RDP ( $P = 9.627 \times 10^{-26}$ ) and BootScan ( $P = 1.602 \times 10^{-32}$ ). In Sq/2005/9.2, the region between nucleotides 1,296 and 1,617, which forms part of ORF 1, may have been the result of recombination of the major parent CABYV-N with an unknown minor parent, this event being detected by RDP ( $P = 5.76 \times 10^{-3}$ ), BootScan ( $P = 2.26 \times 10^{-2}$ ), and Chimaera ( $P = 2.46 \times 10^{-4}$ ) methods. Also, in Sq/2005/9.2, the region between nucleotides 4,785 and 5,669, which includes parts of ORF 5 and the 3'-noncoding region, is closely related to Sq/2004/1.9; whereas, for the rest of the genome, Sq/2005/9.2 is distantly related to Sq/2004/1.9; this event was detected by six methods: RDP ( $P = 3.74 \times 10^{-19}$ ), GENCONV ( $P = 9.85 \times 10^{-17}$ ), BootScan ( $P = 4.18 \times 10^{-22}$ ), MaxChi ( $P = 5.96 \times 10^{-13}$ ), Chimaera ( $P = 1.09 \times 10^{-13}$ ), and SisCan ( $P = 3.19 \times 10^{-16}$ ). For this latter case, assignment of parental and daughter sequences was illustrated by comparing the ME trees for the coding regions upstream and downstream of the putative recombination point (Fig. 5B). This recombination point is in close proximity to the putative transcription origin of the sgRNA 2, a genomic location that has been proposed to be a recombination hotspot for poleroviruses (18). Interestingly, a conserved secondary RNA structure can be predicted in this region, with the recombinant isolate Sq/2005/9.2 having an insertion of 3 nt in a loop of this structure (Fig. 5C), suggesting a potential role of this structure in the recombination mechanism.

**Aphid population dynamics and identification of CABYV-viruliferous aphid species.** *A. gossypii* was the most abundant



**Fig. 4.** Genetic diversity of *Cucurbit aphid-borne yellows virus* isolates in synonymous ( $d_S$ ) and nonsynonymous ( $d_N$ ) positions at each codon of the open reading frames (ORFs) **A**, 2; **B**, 3; and **C**, 4. Differences between the estimated diversity values ( $d_N - d_S$ ) are represented in the vertical axis and the codon positions are represented in the horizontal axis. Bold black lines marked with arrows represent sites under selection detected by the three methods implemented in HyPhy: SLAC, FEL, and REL.

**TABLE 3.** Within- and between-subpopulation nucleotide diversities in the *Cucurbit aphid-borne yellows virus* (CABYV) open reading frames (ORFs) 2, 3, and 4<sup>a</sup>

ORF	Squash	Melon
<b>ORF2</b>		
Squash		
$d_N$	$0.0037 \pm 0.0015$	$0.0031 \pm 0.0013$
$d_S$	$0.0648 \pm 0.0118$	$0.0499 \pm 0.0095$
$d_N/d_S$	$0.0571 \pm 0.0621$	...
Melon		
$d_N$	...	$0.0023 \pm 0.0009$
$d_S$	...	$0.0340 \pm 0.0075$
$d_N/d_S$	...	0.0676
<b>ORF3</b>		
Squash		
$d_N$	$0.0119 \pm 0.0034$	$0.0102 \pm 0.0032$
$d_S$	$0.0273 \pm 0.0074$	$0.0234 \pm 0.0066$
$d_N/d_S$	0.4359	0.4359
Melon		
$d_N$	...	$0.0072 \pm 0.0023$
$d_S$	...	$0.0190 \pm 0.0052$
$d_N/d_S$	...	0.3789
<b>ORF4</b>		
Squash		
$d_N$	$0.0136 \pm 0.0034$	$0.0102 \pm 0.0032$
$d_S$	$0.0270 \pm 0.0081$	$0.0234 \pm 0.0066$
$d_N/d_S$	0.5037	0.4373
Melon		
$d_N$	...	$0.0096 \pm 0.0023$
$d_S$	...	$0.0160 \pm 0.0056$
$d_N/d_S$	...	0.6000

<sup>a</sup> Nucleotide diversity was computed separately for nonsynonymous ( $d_N$ ) and for synonymous ( $d_S$ ) sites by using the PBL method. Between subpopulations, diversity values correspond to net nucleotide substitutions. Sequences were grouped according to the host (squash or melon) from which CABYV was originally isolated.

aphid species caught on the green tile traps located in the open melon fields sampled in both 2010 and 2011, with 41.56 and 28.57%, respectively, of the total cumulative aphid captures. *A. spiraeicola* was the second-most abundant aphid species that landed on melon fields, with 12.18% of the total aphid catches in 2010 and 10.20% in 2011. Other abundant species that landed on melon were *Acyrtosiphon pisum* (7.47%) in 2010, *Aploneura lentisci* (10.20%) in 2011, and *Hyalopterus pruni*, found in the green tile traps both in 2010 and 2011 (2.92 and 12.24%, respectively). Other less abundant species found in the 2010 and 2011 surveys were *Aphis fabae*, *Brachycaudus helichrysi*, *Macrosiphum euphorbiae*, and *Uroleucon* sp. Only a few individuals belonging to two aphid species contained CABYV when aphids caught in the green tile traps in 2011 were analyzed by quantitative real-time RT-PCR. The main aphid species positive for CABYV was again *A. gossypii*. The percentage of potentially CABYV-viruliferous *A. gossypii* was 69.23 % (9 positive aphids per 13 analyzed aphids). The only one individual of the species *M. euphorbiae* analyzed was CABYV-positive.

In addition to the sampling using green tile traps, the aphid species colonizing melon plants were also estimated in surveys carried out during 2010. In agreement with the above data, the most abundant species was *A. gossypii*, with its population peak occurring approximately mid-June, coinciding with the middle of the melon-growing season (data not shown).

**Alternative CABYV hosts and potential reservoirs.** In all, 28 weed and 6 alternative crop species, belonging to 20 different families, were sampled during 2009 to identify alternative hosts and potential CABYV reservoirs. In total, 239 samples were tested for CABYV infection. Out of these, 19 were infected, and belonged to the species *Abutilon theophrasti*, *Malva parviflora*, *Chenopodium murale*, *Ecballium elaterium*, *S. vulgaris*, *Sinapis arvensis*, and *Sonchus oleraceus* (Supplemental Table 1). Interestingly, >40% of the *E. elaterium* (family *Cucurbitaceae*) samples were infected by CABYV, and frequent yellowing symptoms were observed in *E. elaterium* plants close to commercial cucurbit fields. Given the potential importance of plants of this species as alternative or reservoir hosts, transmission experiments from *E. elaterium* to melon plants were performed. *E. elaterium* plants were infected with CABYV through agroinoculation (6); three of eight plants were infected and showed yellowing symptoms 12 days after inoculation. CABYV-infected *E. elaterium* plants were placed together with 32 healthy melon seedlings inside insect-proof boxes. Adult *A. gossypii* aphids were then released inside the boxes. Symptoms typically induced by CABYV appeared in

six melon plants 2 weeks after the release of the aphids. CABYV infection of melon plants was confirmed by molecular hybridization (data not shown), indicating that *A. gossypii* can transmit CABYV from *E. elaterium* to melon, at least under our experimental conditions.

## DISCUSSION

In this report, we analyzed the genetic structure of a CABYV population sampled within a limited geographical area and during a relatively short period of time. We found that, in spite of these restrictions, the virus population appeared to be genetically quite diverse. In comparison with two cases of virus populations sampled in southeastern Spain using similar spatial and temporal scales, the values of total nucleotide diversity estimated for *Cucumber vein yellowing virus* (CVYV) and *Cucurbit yellow stunting disorder virus* Spanish populations were  $\approx 20$  times lower than those estimated for CABYV (21,30) alone. Indeed, CABYV nucleotide diversity values were of the same order of magnitude as those found, for instance, for *Watermelon mosaic virus* populations sampled during a longer period of time and at a relatively much larger geographical scale (36). This is remarkable, because CABYV was detected for the first time in 1992 in France (26) and described in cucurbit crops of the Murcia region rather recently (22). Thus, the diversity observed in the CABYV population does not agree with a recent introduction in the area under study, such as in the case with CVYV mentioned above; rather, it fits with an older presence of CABYV or with multiple introduction events. Alternatively, CABYV might have been introduced recently in Murcia and then it may have evolved rapidly, a hypothesis that is congruent with the high substitution rates reported by Pagán and Holmes (38) for the CP of CABYV ( $>1 \times 10^{-2}$  substitution/site/year). Further studies, expanding the spatial and temporal scales of sampling, are required to clarify this issue.

As is evident after inspection of some branches of the phylogenetic trees deduced in this work, mutation accumulation has been a source of CABYV diversification. However, recombination events may have also played a role. Recombination events were not detected within ORF 2 or ORFs 3 and 4, in agreement with data deduced by Pagán and Holmes (38). Potential recombination events in the noncoding region between these two genes have been reported for many polero- and luteoviruses, including CABYV (24,38,46). Indeed, the topological incongruence of ORF 2 and ORFs 3 and 4 trees for some isolates in this study suggests that this is a likely possibility. However, this hypothesis has to be

TABLE 4. Nucleotide and amino acid sequence comparisons of Sq/2003/7.2 open reading frames (ORFs) versus the corresponding ORFs of other *Cucurbit aphid-borne yellows virus* (CABYV) isolates and other poleroviruses

Virus <sup>a</sup>	NI (%) <sup>b</sup>	Amino acid identity (similarity) (%) <sup>c</sup>					
		ORF 0	ORF 1	ORF 2	ORF 3	ORF 4	ORF 5
Sq/2004/1.9	98	98 (98)	98 (98)	99 (99)	97 (97)	98 (98)	98 (99)
Sq/2005/9.2	91	77 (86)	82 (90)	92 (96)	95 (99)	90 (95)	97 (98)
CABYV-N	95	97 (97)	96 (98)	96 (97)	98 (99)	94 (97)	95 (97)
CABYV-C	88	80 (89)	86 (93)	91 (95)	94 (95)	91 (93)	90 (95)
MABYV	69	69 (82)	61 (74)	74 (82)	81 (89)	66 (76)	51 (63)
BWYV	60	48 (67)	55 (69)	72 (82)	63 (72)	48 (56)	33 (49)
BMVYV	56	41 (61)	45 (60)	65 (77)	62 (72)	47 (55)	33 (49)
TuYV	52	23 (44)	30 (47)	57 (68)	64 (73)	47 (56)	34 (49)
BChV	51	18 (33)	26 (43)	57 (69)	64 (73)	46 (55)	32 (49)
CpCSV	48	18 (30)	25 (42)	55 (67)	67 (75)	56 (65)	39 (52)
TVDV	45	31 (47)	27 (45)	56 (66)	57 (69)	37 (48)	17 (30)
WYDV	45	16 (35)	26 (43)	54 (67)	60 (71)	43 (53)	18 (30)

<sup>a</sup> *Melon aphid-borne yellows virus* (MABYV; GenBank accession number NC\_010809), *Beet western yellows virus* (BWYV; NC\_004756), *Wheat yellow dwarf virus-GPV* (WYDV-GVP; NC\_012931), *Tobacco vein distorting virus* (TVDV; NC010732), *Chickpea chlorotic stunt virus* (CpCSV; NC\_008249), *Turnip yellows virus* (TuYV; NC\_003743), *Beet mild yellowing virus* (BMVYV; NC\_003491), *Beet chlorosis virus* (BChV; NC\_002766).

<sup>b</sup> Nucleotide identity.

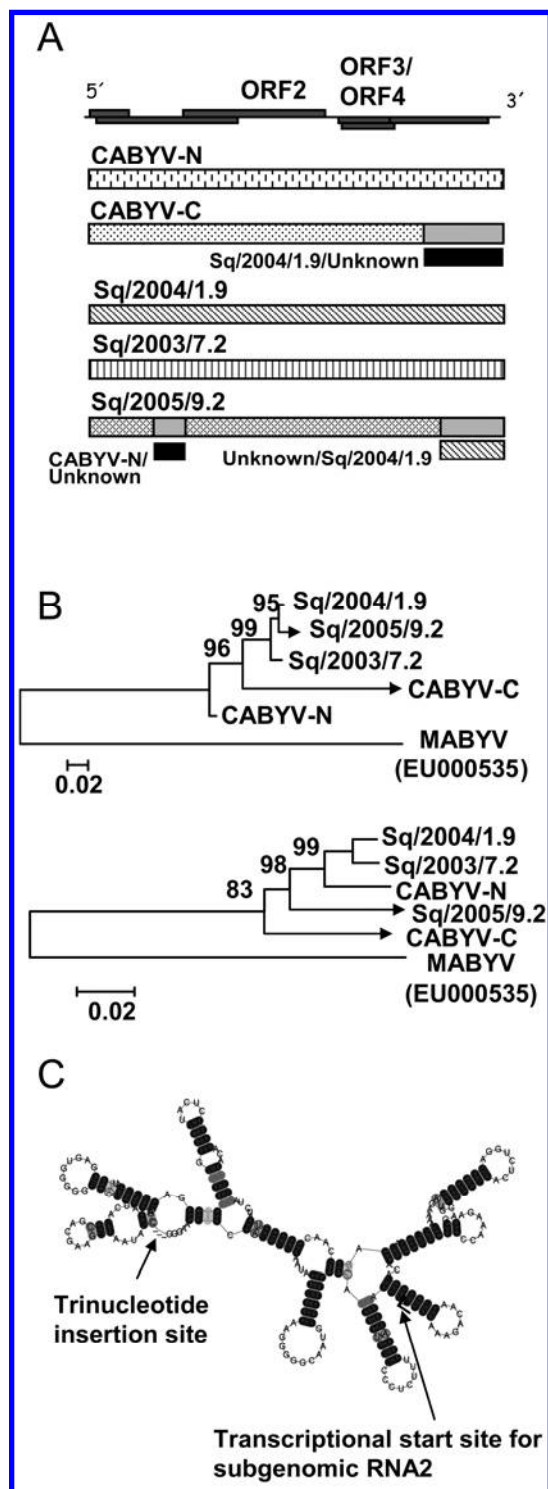
<sup>c</sup> Percentages of identity (similarity) between aligned sequences were calculated as the number of residues identical (similar) shared by compared sequences multiplied by 100 and divided by the number of residues of the shorter sequence excluding insertions or deletions.



taken with caution, because ORF 2 and ORFs 3 and 4 sequences come from independent PCRs from field samples, and mixed infections have been reported in samples of the same collection (23). To expand the genomic information available for CABYV, and to explore whether recombination may have taken place in the CABYV population, sequencing of three complete genomes was carried out. Thus, recombination events were detected, although the putative recombination break points did not correspond to the above-mentioned hotspot. Interestingly, the putative recombination point described by Gibbs and Cooper (13) at the 3'-terminus of the RT domain of three poleroviruses (including CABYV) coincides with a trinucleotide insertion in the recombinant isolate Sq/2005/9.2, which could be the result of a recombination event that occurred at this point. This hypothesis is strongly supported by the close presence of the sgRNA2 initiation site and the detection of a conserved secondary RNA structure in this region, two features which have been described in many studies associated with the occurrence of RNA recombination (18,25).

Regardless of how genetic diversity was generated, it is evident that purifying selection influenced the genetic structure of the CABYV population, though with different intensities for different genomic regions. ORF 2 showed the lowest  $\omega$  value, indicating that strong purifying selection is acting on it, and fits with the identification of RdRp as a functionally important protein of RNA viruses, probably the least prone to fix changes (52). The highest  $\omega$  value was observed for ORF 4, which codes for the MP, in agreement with results obtained for the MP genes of two other luteoviruses, BYDV and *Cereal yellow dwarf virus* (17,38). Here, it is important to comment upon the overlapping nature of the CP and MP genes, where any point mutation would have an impact in both genes simultaneously, possibly constraining the functional evolution of such genes (55). ORF 2 also had several codons under negative selection, in agreement with the presence of conserved motifs in viral RdRps (15,33). Negative selection was also detected acting on the arginine-rich N-terminal domain (R domain) and the shell central domain (S domain) (44) of the CP (ORF 3). Indeed, most residues under purifying selection were identified in close proximity to epitopes 5 and 10 of the S domain (5,49), in analogy with results obtained by Torres et al. (50). Regions located on the surface of the CP protein are presumably critical for virus assembly and stability, plant systemic infection, and virus transmission (27) by means of interactions with viral and nonviral factors (3,9). In contrast, we only detected one positively selected amino acid mutation, located in the MP gene, for which no specific function has been identified yet for CABYV. Further studies for the identification of the biological functions associated with this particular change could help us to understand the process of selection on the MP gene.

Our study suggests that the CABYV population is genetically structured; this could be explained by constraints intrinsic to the viral genome (37). However, ecological factors that influence long- and short-range transmission, persistence in the field, and so on could play an important role in the structure of the plant virus population (12). Here, we analyzed the potential association of genetic structure with host (melon or squash), location of virus isolation, and sampling season but we did not identify any specific relationship. This is in contrast with the identification of CABYV genetic groups associated with broad geographical regions (i.e., Mediterranean and Asian clades in tree of Supplemental Figure 1). Therefore, barriers to CABYV gene flux do not appear to exist in the area where the CABYV isolates come from, although they seem to play a role in CABYV diversity at a global scale. This raises the question of what factors are responsible for the observed genetic structure of our local CABYV population. To try to answer this question, we studied two additional aspects of CABYV epidemiology in the surveyed region: alternative hosts and vector transmission. The alternative host with the highest



**Fig. 5.** Recombination in *Cucurbit aphid-borne yellows virus* (CABYV) genomes. **A**, Recombination hypotheses generated by more than one algorithm of the RDP3 program with a consensus score >40 and a *P* value <1%. On top there is a schematic representation of the CABYV genome with the open reading frames (ORFs) indicated as dark gray boxes. Long dashed boxes represent CABYV genomes (isolate code above the box); internal, pale light-gray segments indicate recombinant regions and short black boxes represent the minor parents. Codes of the major (first) and minor (second) parents are indicated. For example, recombinant isolate CABYV-C (recombinant region indicated by the internal gray segment) has Sq/2004/1.9 as a major parent and an unknown minor parent (short black box). **B**, Minimum evolution trees for the coding regions upstream and downstream of the putative recombination point in ORF 5 of Sq/2005/9.2. Isolates with incongruent phylogenetic relationships are marked with arrows. **C**, RNA secondary structure of the region of the putative transcription origin of the subgenomic RNA2 predicted by the RNAalifold program. The putative recombination break point is marked with an arrow and coincides with a trinucleotide insertion in the Sq/2005/9.2 genome.



potential impact on CABYV epidemiology is *E. elaterium*. As shown in this study, *E. elaterium* plants grown in the proximity of melon and squash crops are frequently infected with CABYV. *E. elaterium* persists from crop season to crop season (M. Juárez, unpublished data) and, thus, could be a virus reservoir of significant importance (48). Interestingly, diversity among *E. elaterium* populations in the Iberian Peninsula has been reported (40). On the other hand, the main aphid species caught on the green tile traps and found colonizing melon crops in the Murcia region was *A. gossypii*. This aphid species is known to be an important vector of CABYV, requiring prolonged feeding times and direct contact with vascular tissues, spreading CABYV mainly by colonizing vectors (11). By means of real-time quantitative RT-PCR, it was possible to demonstrate that *A. gossypii* was not only the most abundant aphid species in the fields but also the aphid species with the highest number of aphids carrying CABYV. Another CABYV-viruliferous aphid species found in melon crops in Murcia during this study was *Macrosiphum euphorbiae*. Although *M. euphorbiae* has been previously described as a potential CABYV vector (6), the low density of this noncolonizing aphid species suggests a very limited role on CABYV spread in this region. Taken together, this information indicates that *A. gossypii* is the main aphid species involved in the spread of CABYV in melon crops in the Murcia region. Significantly, morphological differences were observed between individuals of different *A. gossypii* colonies in Murcian melon crops (M. Juárez and M. A. Aranda, unpublished data); therefore, it is tempting to speculate that genetic differences among aphids that colonize CABYV hosts may exist and these might be associated with CABYV transmission specificities (8,29) and, thus, with CABYV diversity.

To summarize, our results support the following conclusions. First, the Spanish CABYV population is diverse and recombination might have had an important role in diversity generation and maintenance. The main evolutionary force identified was purifying selection. Second, the population appears to be genetically structured, with at least two genetic groups that coexist in the field. This structure did not seem to be associated with crop species, date, or locality of sampling and, therefore, other factors must be responsible for this observation. Among ecological factors that could potentially condition the genetic structure of CABYV populations are vector transmission and alternative host adaptation. In this regard, it would be particularly interesting to study the potential adaptation of CABYV to *A. gossypii*, its main vector in the surveyed area, and to *E. elaterium*, its main alternative host in the same area.

#### ACKNOWLEDGMENTS

This work was supported by grants 11934/PI/09 (Fundación Séneca de la Región de Murcia, Spain) and AGL2009-07552/AGR (Ministerio de Ciencia e Innovación, Spain). M. Kassem was recipient of a fellowship from the Agencia Española de Cooperación Internacional-Ministerio de Asuntos Exteriores (AECI-MAE). We thank V. Truniger for critically reading the manuscript, M. C. Montesinos and B. Gosálvez for their excellent technical assistance, and M. Fon for checking the English.

#### LITERATURE CITED

1. Ashoub, A., Rohde, W., and Prüfer, D. 1998. In plant transcription of a second subgenomic RNA increases the complexity of the subgroup 2 luteovirus genome. *Nucleic Acids Res.* 26:420-426.
2. Bananej, K., Vahdat, A., Predajna, L., and Glasa, M. 2009. Molecular characterization of geographically different *Cucurbit aphid-borne yellows virus* isolates. *Acta Virol.* 53:61-64.
3. Bencharki, B., Boissinot, S., Revollon, S., Ziegler-Graff, V., Erdinger, M., Wiss, L., Dinant, S., Renard, D., Beuve, M., Lemaître-Guillier, C., and Brault, V. 2010. Phloem protein partners of *Cucurbit aphid-borne yellows virus*: Possible involvement of phloem proteins in virus transmission by aphids. *Mol. Plant-Microbe Interact.* 23:799-810.
4. Bencharki, B., Mutterer, J., El Yamani, M., Ziegler-Graff, V., Zaoui, D., and Jonard, G. 1999. Severity of infection of Moroccan barley yellow

- dwarf virus PAV isolates correlates with variability in their coat protein sequences. *Ann. Appl. Biol.* 134:89-99.
5. Brault, V., Bergdoll, M., Mutterer, J., Prasad, V., Pfeffer, S., Erdinger, M., Richards, K. E., and Ziegler-Graff, V. 2003. Effects of point mutations in the major capsid protein of *Beet western yellows virus* on capsid formation, virus accumulation, and aphid transmission. *J. Virol.* 77:3247-3256.
6. Brault, V., Pérignon, S., Reinbold, C., Erdinger, M., Scheidecker, D., Herrbach, E., Richards, K., and Ziegler-Graff, V. 2005. The *Polerovirus* minor capsid protein determines vector specificity and intestinal tropism in the aphid. *J. Virol.* 79:9685-9693.
7. Brault, V., van den Heuvel, J. F. J. M., Verbeek, M., Ziegler-Graff, V., Reutenauer, A., Herrbach, E., Garaud, J. C., Guillely, H., Richards, K., and Jonard, G. 1995. Aphid transmission of beet western yellows luteovirus requires the minor capsid read-through protein P74. *EMBO J.* 14:650-659.
8. Burrows, M. E., Caillaud, M. C., Smith, D. M., Benson, E. C., Gildow, F. E., and Gray, S. M. 2006. The genetic regulation of polerovirus and luteovirus transmission in the aphid *Schizaphis graminum*. *Phytopathology* 96:828-837.
9. Chare, E. R., and Holmes, E. C. 2004. Selection pressures in the capsid genes of plant RNA viruses reflect mode of transmission. *J. Gen. Virol.* 85:3149-3157.
10. D'Arcy, C. J., and Domier, L. L. 2005. Luteoviridae. In: *Virus Taxonomy: VIIIth Report of the International Committee on Taxonomy of Viruses*. M. A. Mayo, J. Maniloff, U. Desselberger, L. A. Ball, and Claude M. Fauquet, eds. Academic Press, New York.
11. Fereres, A., and Moreno, A. 2009. Behavioural aspects influencing plant virus transmission by homopteran insects. *Virus Res.* 141:158-168.
12. García-Arenal, F., Fraile, A., and Malpica, J. M. 2001. Variability and genetic structure of plant virus population. *Annu. Rev. Phytopathol.* 39:157-186.
13. Gibbs, M., and Cooper, J. 1995. A recombinational event in the history of luteoviruses probably induced by base-pairing between the genomes of two distinct viruses. *Virology* 206:1129-1132.
14. Gildow, F. E., Reavy, B., Mayo, M. A., Duncan, G. H., Woodford, J. A. T., Lamb, J. W., and Hay, R. T. 2000. Aphid acquisition and cellular transport of *Potato leafroll virus*-like particles lacking P5 readthrough protein. *Phytopathology* 90:1153-1161.
15. Guillely, H., Wipf-Scheibel, C., Richards, K., Lecoq, H., and Jonard, G. 1994. Nucleotide sequence of cucurbit aphid-borne yellows luteovirus. *Virology* 202:1012-1017.
16. Guyader, S., and Ducray, D. G. 2002. Sequence analysis of *Potato leafroll virus* isolates reveals genetic stability, major evolutionary events and differential selection pressure between overlapping reading frame products. *J. Gen. Virol.* 83:1799-1807.
17. Hall, G. 2006. Selective constraint and genetic differentiation in geographically distant *Barley yellow dwarf virus* populations. *J. Gen. Virol.* 87:3067-3075.
18. Hauser, S., Stevens, M., Beuve, M., and Lemaire, O. 2002. Biological properties and molecular characterization of *Beet chlorosis virus* (BChV). *Arch. Virol.* 147:745-762.
19. Irwin, M. E. 1980. Sampling aphids in soybean fields. In: *Sampling Methods in Soybean Entomology*. M. Kogan and D. C. Herzog, eds. Springer-Verlag, New York.
20. Irwin, M. E., and Ruesink, W. G. 1986. Vector intensity: A product of propensity and activity. In: *Plant Virus Epidemics: Monitoring, Modelling and Predictions Outbreaks*, 1st ed. G. D. McLean, R. G. Garrett, and W. G. Ruesink, eds. Elsevier Science & Technology Books, Sydney, Australia.
21. Janssen, D., Velasco, L., Martín, G., Segundo, E., and Cuadrado, I. M. 2007. Low genetic diversity among *Cucurbit vein yellowing virus* isolates from Spain. *Virus Genes* 34:367-371.
22. Juárez, M., Truniger, V., and Aranda, M. A. 2004. First report of *Cucurbit aphid-borne yellows virus* in Spain. *Plant Dis.* 88:907.
23. Kassem, M. A., Sempere, R. N., Juárez, M., Aranda, M. A., and Truniger, V. 2007. *Cucurbit aphid-borne yellows virus* is prevalent in field-grown cucurbit crops of southeastern Spain. *Plant Dis.* 91:232-238.
24. Knierim, D., Deng, T. C., Tsai, W. S., Green, S. K., and Kenyon, L. 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.
25. Lai, M. M. 1992. RNA recombination in animal and plant viruses. *Microbiol. Rev.* 56:61-79.
26. Lecoq, H., Bourdin, D., Wipf-Scheibel, C., Bon, M., Lot, H., Lemaire, O., and Herrbach, E. 1992. A new yellowing disease of cucurbits caused by a luteovirus, cucurbit aphid-borne yellows virus. *Plant Pathol.* 41:749-761.
27. Lee, L., Kaplan, I. B., Ripoll, D. R., Liang, D., Palukaitis, P., and Gray, S. M. 2005. A surface loop of the *Potato leafroll virus* coat protein is involved in virion assembly, systemic movement, and aphid transmission. *J. Virol.* 79:1207-1214.

28. Lemaire, O., Gubler, W. D., Valencia, J., Lecoq, H., and Falk, B. W. 1993. First report of cucurbit aphid-borne yellows luteovirus in the United States. *Plant Dis.* 77:1169.
29. Lucio-Zavaleta, E., Smith, D. M., and Gray, S. M. 2001. Variation in transmission efficiency among *Barley yellow dwarf virus-RMV* isolates and clones of the normally inefficient aphid vector, *Rhopalosiphum padi*. *Phytopathology* 91:792-796.
30. Marco, C. F., and Aranda M. A. 2005. Genetic diversity of a natural population of *Cucurbit yellow stunting disorder virus*. *J. Gen. Virol.* 86:815-822.
31. Mastari, J., and Lapierre, H. 1999. Sequences of the 3' halves of the genome of barley yellow dwarf virus-PAV CpA isolates that vary in symptom severity. *Eur. J. Plant Pathol.* 105:801-811.
32. Mayo, M. A., Robinson, D. J., Jolly, C. A., and Hyman, L. J. 1989. Nucleotide sequence of potato leafroll luteovirus RNA. *Gen. Virol.* 70:1037-1051.
33. Mayo, M. A., and Ziegler-Graff, V. 1996. Molecular biology of luteoviruses. *Adv. Virus Res.* 46:413-460.
34. Mnari-Hattab, M., Gauthier, N., and Zouba, A. 2009. Biological and molecular characterization of the *Cucurbit aphid-borne yellows virus* affecting cucurbits in Tunisia. *Plant Dis.* 93:1065-1072.
35. Moreno, A., Garzo, E., Fernandez-Mata G., Kassem, M., Aranda, M. A., and Fereres, A. 2011. Aphids secrete watery saliva into plant tissues from the onset of stylet penetration. *Entomol. Exp. Appl.* 139:145-153.
36. Moreno, I. M., Malpica, J. M., Díaz-Pendón, J. A., Moriones, E., Fraile, A., and García-Arenal, F. 2004. Variability and genetic structure of the population of *Watermelon mosaic virus* infecting melon in Spain. *Virology* 318:451-460.
37. Moury, B., and Simon, V. 2011. dN/dS-based methods detect positive selection linked to trade-offs between different fitness traits in the coat protein of *Potato virus Y*. *Mol. Biol. Evol.* 28:2707-2717.
38. Pagán, I., and Holmes, E. C. 2010. Long-term evolution of the Luteoviridae: Time-scale and mode of virus speciation. *J. Virol.* 84:6177-6187.
39. Pazhouhandeh, M., Dieterle, M., Marrocco, K., Lechner, E., Berry, B., Brault, V., Hemmer, O., Kretsch, T., Richards, K. E., Genschik, P., and Ziegler Graff, V. 2006. F-boxlike domain in the *Polerovirus* protein P0 is required for silencing suppressor function. *Proc. Natl. Acad. Sci. USA* 103:1994-1999.
40. Perez Chiscano, J. L. 1985. Distribución geográfica de *Ecballium elaterium* (L.) Richard (Cucurbitaceae) en la Península Ibérica e Islas Baleares. *Studia Bot.* 4:57.
41. Perez, P., Collar, J. L., Avilla, C., Duque, M., and Fereres, A. 1995. Estimation of vector propensity of potato virus Y (PVY) in open-field pepper crops of central Spain. *J. Econ. Entomol.* 88:986-991.
42. Pfeffer, S., Dunoyer, P., Heim, F., Richards, K. E., Jonard, G., and Ziegler-Graff, V. 2002. P0 of *Beet western yellow virus* is a suppressor of posttranscriptional gene silencing. *J. Virol.* 76:6815-6824.
43. Prüfer, D., Kawchuk, L. M., and Rohde, W. 2006. *Polerovirus* ORF0 genes induce a host-specific response resembling viral infection. *Can. J. Plant Pathol.* 28:302-309.
44. Rao, A. L. 2006. Genome packaging by spherical plant RNA viruses. *Annu. Rev. Phytopathol.* 44:61-87.
45. Sadowy, E., Juszczuk, M., David, C., Gronenborn, B., and Hulanicka, M. D., 2001. Mutational analysis of the proteinase functions of *Potato leafroll virus*. *J. Gen. Virol.* 82:1517-1527.
46. Shang, Q., Xiang, H., Han, C., Li, D., and Yua, J. 2009. Distribution and molecular diversity of three cucurbit-infecting poleroviruses in China. *Virus Res.* 145:341-346.
47. Silva, T. F., Corrêa, R. L., Castilho, Y., Silvie, P., Bélot, J.-L., and Vaslin, M. F. S. 2008. Widespread distribution and a new recombinant species of Brazilian virus associated with cotton blue disease. *Viol. J.* 5:123-135.
48. Srinivasan, R., and Alvarez, J. M. 2008. Hairy nightshade as a potential *Potato leafroll virus* (Luteoviridae: *Polerovirus*) inoculum source in Pacific Northwest potato ecosystems. *Phytopathology* 98:985-991.
49. Torrance, L. 1992. Analysis of epitopes on potato leafroll virus capsid protein. *Virology* 191:485-489.
50. Torres, M. W., Corrêa, R. L., and Schrago, C. G. 2005. Analysis of differential selective forces acting on the coat protein (P3) of the plant virus family Luteoviridae. *Genet. Mol. Res.* 4:790-802.
51. van der Wilk, F., Verbeek, M., Dulleman, A. M., and van den Heuvel, J. F. J. M. 1997. The genome-linked protein of potato leafroll virus is located downstream of the putative protease domain of the ORF1 product. *Virology* 234:300-303.
52. Wu, B., Blanchard-Letort, A., Liu, Y., Zhou, G., Wang, X., and Elena, S. F. 2011. Dynamics of molecular evolution and phylogeography of *Barley yellow dwarf virus-PAV*. *PLoS ONE* 6:e16896
53. Xiang, H. Y., Shang, Q. X., Han, C. G., Li, D. W., and Yu, J. L. 2008. Complete sequence analysis reveals two distinct poleroviruses infecting cucurbits in China. *Arch. Virol.* 153:1155-1160.
54. Yardimci, N., and Özgönen, H. 2007. First report of *Cucurbit aphid-borne yellows virus* in Turkey. *Australas. Plant Dis. Notes* 2:59.
55. Zhang, D., Chen, J., Deng, L., Mao, Q., Zheng, J., Wu, J., Zeng, C., and Li, Y. 2010. Evolutionary selection associated with the multi-function of overlapping genes in the *Hepatitis B virus*. *Infect. Genet. Evol.* 10:84-88.
56. Ziegler-Graff, V., Brault, V., Mutterer, J. D., Simionis, M. T., Herrbach, E., Guillely, H., Richards, K. E., and Jonard, G. 1996. The coat protein of beet western yellows luteovirus is essential for systemic infection but the viral gene products P29 and P19 are dispensable for systemic infection and aphid transmission. *Mol. Plant-Microbe Interact.* 9:501-510.