Arch Virol (2010) 155:1849–1854
DOI 10.1007/s00705-010-0764-3

BRIEF REPORT

The complete genome sequence of a virus associated with cotton blue disease, cotton leafroll dwarf virus, confirms that it is a new member of the genus *Polerovirus*

Ana J. Distéfano · Ivan Bonacic Kresic · H. Esteban Hopp

Received: 15 February 2010 / Accepted: 17 July 2010 / Published online: 31 July 2010

Abstract Cotton blue disease is the most important virus disease of cotton in the southern part of America. The complete nucleotide sequence of the ssRNA genome of the cotton blue disease-associated virus was determined for the first time. It comprised 5,866 nucleotides, and the deduced genomic organization resembled that of members of the genus *Polerovirus*. Sequence homology comparison and phylogenetic analysis confirm that this virus (previous proposed name cotton leafroll dwarf virus) is a member of a new species within the genus *Polerovirus*.

The poleroviruses are an important genus of plant viruses that can infect a wide range of hosts, causing important losses in agriculture. The genera *Polerovirus*, *Luteovirus*, and *Enamovirus* constitute the family *Luteoviridae* [6]. All members of this family contain single-stranded positive-sense RNA genomes and are phloem-limited viruses that are transmitted by aphids in a persistent, circulative and non-propagative manner [13].

Cotton blue disease is an important disease present in cotton crops in America, Africa and Asia [4], transmitted by the aphid vector *Aphis gossypii* Glover [3]. Cotton plants affected by this disease show a stunting phenotype, characterized by shortening of internodes, leaf rolling, vein yellowing and intensive green color of the foliage [3]. In Argentina, the disease was first described after an outbreak affecting cotton occurred during the crop season 1982/83 in the province of Misiones [2]. Nowadays, it is still an important disease and is the object of a specific regulation for commercial transgenic cotton release in Argentina because most cultivars are susceptible. Lenardon [12] performed serological tests of infected plants from north-eastern Argentina to identify the causal agent of this disease and suggested a putative relationship to barley yellow dwarf virus serotypes RPV and PAV, two members of the family *Luteoviridae*. Recently, Corrêa et al. [5] described the sequence analysis of the capsid protein and the partial RdRp (1,405 nt) of a Brazilian isolate, revealing sequence homology to a new virus probably belonging to the genus *Polerovirus*, and the name cotton leafroll dwarf virus was proposed. We report here, for the first time, the complete sequence of the cotton blue disease-associated virus (Argentinean isolate), confirming that it should be classified in the genus *Polerovirus*, as a member of a new species.

A virus isolate was obtained from a susceptible cotton cultivar (*Gossypium hirsutum* cultivar Banda 56) showing typical blue disease symptoms, collected from a field plantation in the Province of Chaco, located in the most important cotton-producing area of Argentina. A colony of *A. gossypii* was obtained from a healthy field plantation of

Nucleotide sequence data reported here are available in the GenBank database under the accession number GU167940.

A. J. Distéfano · H. E. Hopp
Instituto de Biotecnologia, INTA Castelar, Hurlingham, Provincia de Buenos Aires, Argentina

A. J. Distéfano · H. E. Hopp
Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Pabellon 2, Argentina

I. Bonacic Kresic
EEA Saenz Peña, INTA, 3700 Presidencia Roque Saenz Peña, Chaco, Argentina

A. J. Distéfano (✉)
Instituto de Biotecnología, CICVyA, CNIA, INTA, Los Reseros y Las Cabañas s/n, B1712WAA Hurlingham, Buenos Aires, Argentina
e-mail: adistefano@cnia.inta.gov.ar

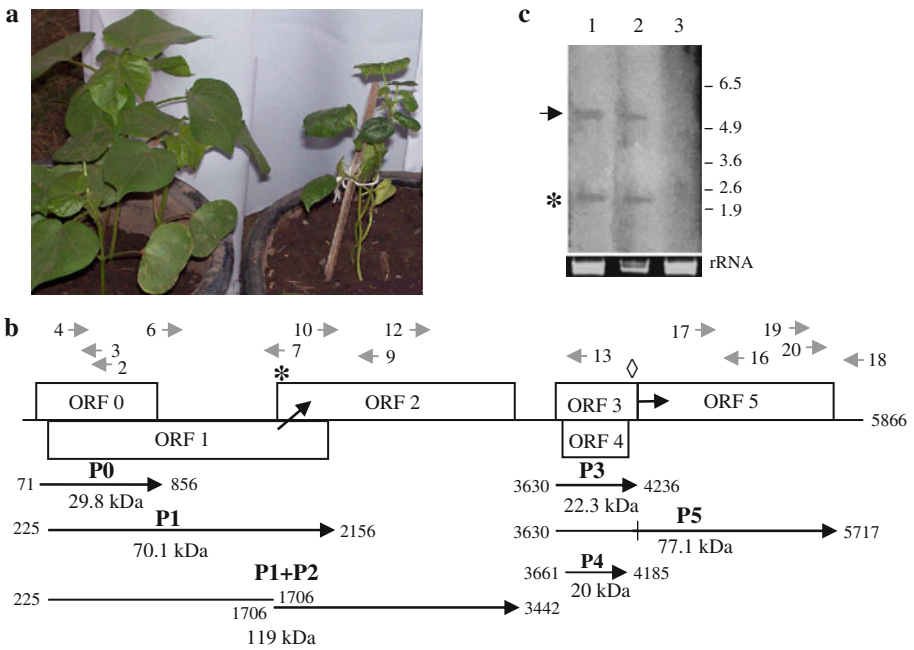

Fig. 1 **a** Healthy *G. hirsutum* cv. Banda 56 plants (uninoculated) (*left*) compared to aphid-inoculated plants of the same age with typical symptoms of blue disease (*right*) **b** CLRDV genome organization. Rectangles indicate the predicted ORFs. Protein products are represented by a *thin line* and *arrows*. Positions are marked at the extremities, and predicted sizes of the proteins are listed below the indicated protein regions. The predicted heptanucleotide frameshift is marked with an *asterisk*, and the "leaky" stop codon with a *diamond.* The gray *arrows* indicate the primers used for PCR **c** Detection of CLRDV RNA by northern blot using the capsid probe. RNA was extracted from aphid-infected *G. hirsutum* cv. Banda 56 plants (*lanes 1 and 2*) and from a healthy plant (*lane 3*). The *arrow* marks the position of genomic RNA, and the *asterisk* marks the position of subgenomic RNA. A molecular weight marker (in kbp) is shown to the right. Ribosomal RNAs were stained with ethidium bromide (*bottom*)

cotton and reared on susceptible *G. hirsutum* cv. Banda 56 in breeding chambers under controlled conditions for four generations to obtain a non-viruliferous population (healthy laboratory-reared insects). Non-viruliferous *A. gossypii* were transferred to the infected cotton plants and maintained for 2 days for virus acquisition under greenhouse conditions. Subsequently, two aphids were placed on healthy young *G. hirsutum* cv. Banda 56 plants for an inoculation period of 24 h, and symptom development was scored 20 days after inoculation. The virus was transmitted successfully, and typical blue disease symptoms were observed (stunting phenotype, leaf rolling, vein yellowing and intensive green color of the foliage) (Fig. 1a). Total RNA was isolated from symptomatic leaves using an RNeasy Mini Kit (Qiagen). Synthesis of cDNA from 3 μg of total RNA was carried out using SuperScript III reverse transcriptase and random primers (Invitrogen). The first viral sequence was amplified by PCR using the cDNA obtained by random-primed synthesis as a template and degenerate primers designed after comparing homologous conserved sequences of related viruses. The oligonucleotides used were primers 10 (5′CTCCCCMAGTTY TACYWC3′, nt 2,025–2,042) and 13 (PLR:5′TCTGA WARASWCGGCCCGAASGTGA3′, nt 3,896–3,872 [5]). The PCR reaction was carried out with a denaturation step at 94°C for 3 min followed by 40 cycles of 1 min at 94°C, 1 min at 52°C, and 2 min at 72°C, with a final extension at 72°C for 10 min. The 1,871-bp fragment amplified corresponded to parts of the polerovirus ORF2 and ORF3. Subsequent amplifications were performed with similar PCR cycling profiles but using specific primers designed based on the newly obtained viral sequences on one side and a degenerate primer on the other side. Thus, two fragments, of 2,210 bp and 1,065 bp, were amplified using primers 12 (5′ACTGATTGCTCCGGTTTCGACTGG3′ nt 2,832–2,855) and 16 (5′GTTTTRARRTCTTTTCCTT GA3′ nt 5,042–5,022), and primers 17 (5′GCTGCACG CGCAGTGGAAGTG3′ nt 4,729–4,749) and 18 (5′TGCC TATCCTTTCGGAGTCGTTCC3′ nt 5,794–5,771), respectively, completing the ORF3 and ORF5 sequences. Finally, two regions, of 1,312 bp and 1,314 bp, were obtained using primers 6 (5′CAYGCHGGYTATGCMASWTG3′ nt 934–953) and 9 (5′TTGAATTACACGCTCCCTCT3′ nt 2,246–2,227), and primers 4 (5′TRSATSTGCCCBCHC GMVGA3′ nt 412–431) and 7 (5′GGACTTGCCTTGTT TGGGCT3′ nt 1,726–1,707), respectively, which comprised part of ORF0 and ORF1 and completed the ORF2 sequence. Each base pair position in the genome was sequenced from two independent amplicons. In all cases, two to four clones from each amplicon were analyzed to

determine the consensus sequence. Amplification was performed using Platinum Pfx DNA Polymerase (Invitrogen), and the amplified products were purified using a QIAEX® II Gel Extraction Kit (Qiagen), cloned into pGEM-T Easy Vector (Life Technologies) and sequenced using an ABI 3,730 XL automated sequencer. To obtain and analyze the 5′ and 3′ termini of the CLRDV genome, Rapid Amplification of cDNA Ends (RACE) was performed [8] employing kits from Invitrogen. For 3′ RACE, the RNA was denatured and 3′ termini were polyadenylated using 5 U of polyA polymerase before performing first-strand synthesis and amplification. For 3′ RACE, the specific CLRDV oligonucleotides were primer 19 (5′AACCAACACGATCAATTTCA3′ nt 5,363–5,382) and nested primer 20 (5′AGAAGCCGTTCCAGTTGGGG3′ nt 5,536–5,555). For 5′ RACE, the specific CLRDV internal antisense oligonucleotides were primer 2 (5′CCGCGA GTGCAGAGATACTC3′ nt 519–500) and nested-specific primer 3 (5′CGACCAAAGAGCGAATAGCA3′ nt 480–461). The overall sequencing strategy is shown in Fig. 1b. The complete nucleotide sequence was assembled and analyzed using the software NTI Suite 8.0 (InforMax Inc.). For northern blot analysis, 10 μg of total RNA isolated from healthy and symptomatic leaves at 30 days postinfection was resolved by denaturing 1% agarose-gel electrophoresis, transferred onto a Hybond N^+ membrane, and finally hybridized with a radiolabelled capsid fragment to use as a probe employing the Prime-a-Gene Labeling System (Promega) in the presence of [α-^{32}P] dCTP. Phylogenetic analysis was done using PHYLIP software package version 3.5c [7].

The CLRDV genome was 5,866 nucleotides long and had six predicted major open reading frames (ORFs) that were similar to those found in other viruses belonging to the genus *Polerovirus* [13]. The features of the sequences and the sizes of the ORFs are shown in Fig. 1b. The sequence was predicted to contain a non-coding sequence of 70 nucleotides at the 5′ end, 149 nucleotides at the 3′ end, and 187 nucleotides between the two main blocks of coding sequences (5′ block and 3′ block). The CLRDV intergenic region resembles those of other members of the genus *Polerovirus*, which are around 200 nt long, contrasting with the central non-coding regions of members of the genera *Luteovirus* and *Enamovirus*, which are about 100 nt long. The 5′ terminal sequence of CLRDV has an ACAAAA sequence, which is the same as what has been reported for other poleroviruses. The 5′ block consisted of three ORFs. ORF0 encodes a putative protein (P0, 29.8 kDa) that is particularly characteristic of members of the genus *Polerovirus*, coding for the silencing suppressor protein [15]. The second ORF, ORF1, was predicted to be expressed by leaky scanning, yielding a P1 protein of 70.1 kDa. ORF2 was predicted to begin at nt 1,706 and to end at nt 3,442, and it might be expressed when ribosomes undergo a-1 frameshift from ORF1 to ORF2 to produce a fusion protein (119 kDa). By comparison with other members of the genus, the frameshift site in CLRDV was predicted to be at the heptanucleotide GGGAAAC at position 1,700–1,706. This sequence was followed six bases downstream by an inferred structured region, known as an RNA pseudoknot, which is suspected to be essential for the frameshift [9, 10]. The putative P2 polypeptide contained the motif GXXXTXXXNX$_{25–40}$GDD near the C-terminus, which is conserved in all well-known RNA-dependent RNA polymerases [13, 14]. It has been suggested for similar viruses that P1 and the fusion protein P1–P2 provide the genome-replication components [13]. The 3′ block consisted of three ORFs that, in all well-known poleroviruses, are expressed by translation of a subgenomic RNA. ORF3 encodes the putative coat protein P3 (CP, 22.3 kDa). ORF4 was embedded within the ORF3 gene segment but in a different reading frame and codes for a putative cell-to-cell movement protein (P4, 20 kDa) by a leaky scanning mechanism. ORF5 immediately follows the ORF3 stop codon and is probably translated by in-frame readthrough of this stop codon. The resulting 77.1-kDa fusion protein, P3–P5, is generally referred to as the readthrough domain (RTD) protein. At the 5′ end of ORF5, there is a C-rich domain, followed by a region of approximately 200 amino acid residues with considerable sequence conservation, which commonly exists in members of the family *Luteoviridae*. This typical proline-rich sequence encoding 7–13 proline residues and the conserved region were found in CLRDV P5 and was located just downstream of the CP stop codon. The presence of this region has been shown to be necessary for the transmission of poleroviruses by their aphid vectors and for efficient accumulation of related viruses in whole plants [1, 18].

Northern blot analysis was performed using infected and healthy plants. Two bands were observed in the RNA extracts of infected plants, corresponding to virus genomic and subgenomic RNAs (Fig. 1c).

Selected comparison between different regions of the CLRDV-ARG genome and other viruses in the family *Luteoviridae* are shown in Table 1. In all regions of the genome, CLRDV-ARG was most closely related to members of the genus *Polerovirus*. When the deduced amino acid sequence of the CP gene was analyzed, identities of 99 and 94%, were observed with CLRDV-PV1 (a Brazilian isolate) and with chickpea stunt disease associated virus (CpSDaV; an unclassified member of the family *Luteoviridae*), respectively. Similarly, when the complete CLRDV-ARG RdRp sequence (P1–P2) was analyzed, significant identities with members of the genus *Polerovirus* were also found (ranging between 45 and 52%) while identities with members of the genera *Luteovirus* and

Table 1 Amino acid sequence identity between the Argentinean isolate of CLRDV and some members of the family *Luteoviridae* (virus abbreviations and sequences used for comparison are listed in the legend of Fig. 2)

Virus	P0 (261 aa) (%)	P1 (643 aa) (%)	P1–P2 (1,073 aa) (%)	P3 (201 aa) (%)	P4 (174 aa) (%)	P5 (694 aa) (%)
CLRDV-PV1	ND	ND	98[a]	99	97	98
CpSDaV	ND	ND	ND	94	89	ND
TuYV	30	36	52	80	64	48
BWYV	27	35	48	78	64	47
BMYV	29	33	49	77	64	47
BCV-2a	27	ND	45	77	63	46
MABYV	25	33	48	75	58	56
CpCSV	NI	32	48	74	54	53
CABYV	26	35	50	64	56	62
PLRV	NI	33	48	65	49	43
BYDV-PAV	ND	31	25	64	50	39
CYDV-RPV	NI	31	46	62	47	44
CTRLV	NI	33	51	57	40	41
SCYLV	ND	36	49	45	39	35
PEMV	ND	NI	35	31	ND	41

ND No data available,
NI No identity detected

[a] Only 634 aa available

Enamovirus were much lower (Table 1). We also found a 98% identity with the partial RdRp sequence from CLRDV-PV1 (from which only a 634-amino-acid-long sequence is available). P0 and P1 were the least conserved proteins, ranging in their identity between 25 and 35% compared to other members of the genus *Polerovirus*. Phylogenetic trees were constructed by the neighbor-joining method [16]. Bootstrap analysis (*n* 100) of the data indicated that the branches were well supported. Phylogenetic analysis of the coat protein, the RdRp protein and the complete genome sequence confirmed the relationships among the viruses and placed CLRDV-ARG in the *Polerovirus* branch of the trees (Fig. 2). The topology of the CP tree and the RdRp tree showed that CLRDV-ARG was very closely related to CLRDV-PV1 (the Brazilian isolate) and to CpSDaV in the CP tree.

The coat protein is conserved in poleroviruses and recapitulates *Polerovirus* phylogeny [13]. Based on the criteria established by the International Committee on Taxonomy of Viruses [6], members of different species in the family *Luteoviridae* are expected to have CP amino acid sequence identities below 80%, and their entire genomes are expected to be less than 85% identical. The CLRDV-ARG CP sequence obtained displays 99% identity to CLRDV-PV1 (Brazilian isolate), showing that both viruses could just be geographic variants of the same species, most probably because there are no major geographic barriers between the cotton-producing regions of southeastern Brazil and northeastern Argentina. We also found 94% identity with CpSDaV, suggesting that both viruses could be strains of the same species, although their economically important hosts are not phylogenetically related (*Malvaceae* versus *Favaceae*).

There are some viruses, such as BWYV and BMYV, that share strong sequence identity in their 3′ part of the genome but have little similarity in the 5′ region. Comparative analysis of complete genome sequences showed that BMYV is a recombinant between two poleroviruses, and for this reason, it is not considered a strain of BWYV [11]. Our results showed that the CLRDV-ARG P0 and P1 proteins were the least conserved when compared to those other polerovirus, in contrast to P3, which was the most conserved of the deduced amino acid sequences. Unfortunately, the P0 and P1 protein sequences from the most closely related isolates, CLRDV-PV1 and CpSDaV, are not publically available yet; therefore, it is not possible to confirm that these three viruses are indeed strains of the same virus. In a recent study of the genetic diversity of CLRDV in Brazil, the authors found three divergent isolates in association with plants showing atypical blue disease symptoms. Upon analysis of the partial RdRp sequence (the 92 C-terminal amino acids of the viral protein), they established that the three recombinant isolates shared almost 70% sequence identity with the CLRDV-PV1 RdRp, whereas the identity for the CPs ranged between 97 and 100%. The authors propose that recombination events possibly occurred between CLRDV-PV1 and a still unidentified member of the genus *Polerovirus* [17]. In accordance with this, we found 73% sequence identity of RdRp sequence between CLRDV-ARG and these three Brazilian isolates. Thus, further studies on virus diversity are necessary to complete the characterization of cotton blue disease in Argentina.

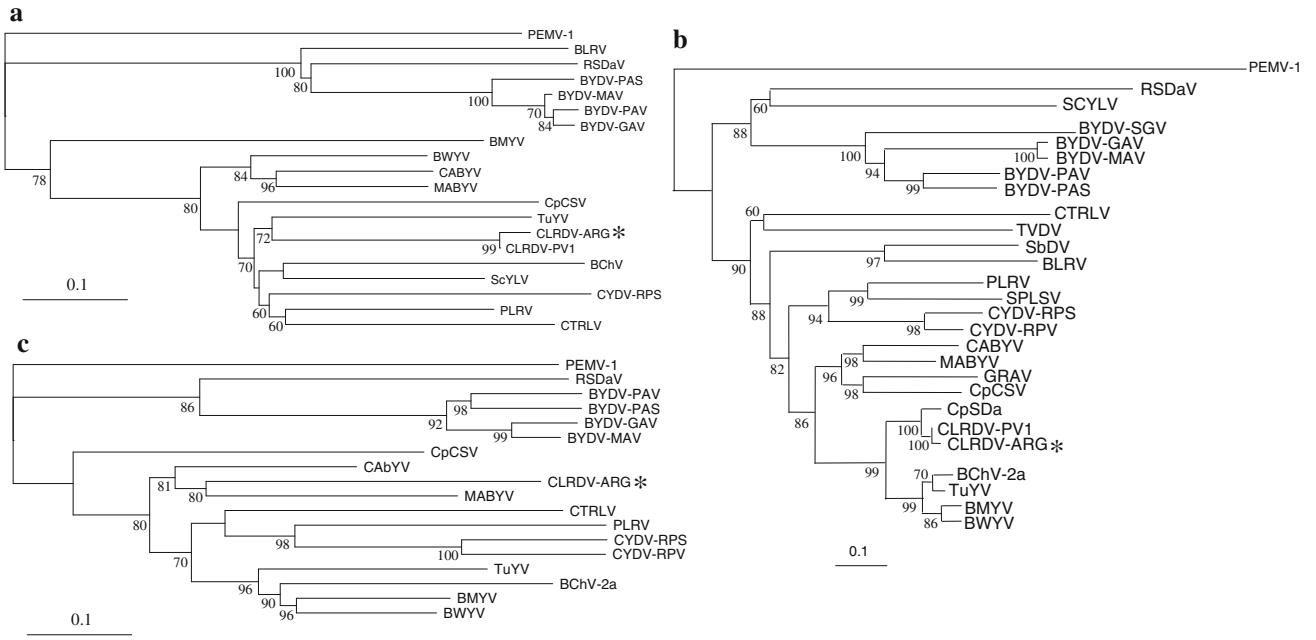

Fig. 2 Phylogenetic trees based on the codon-aligned nucleotide sequences of **a** the RNA-dependent RNA polymerase, **b** the coat protein, and **c** the complete genome sequence of members of the family *Luteoviridae*. Distance trees were constructed using the neighbor-joining method. The scale bar shows the number of nucleotide substitutions in the aligned sequence. Figures below the lines indicate the frequency of the cluster after bootstrap analysis (100 replicates; shown only when >60%). Virus abbreviations and GenBank accession numbers are as follows: barley yellow dwarf virus-PAV (BYDV-PAV, D85783), barley yellow dwarf virus-SGV (BYDV-SGV, U06865), barley yellow dwarf virus-PAS (BYDV-PAS, AF218798), barley yellow dwarf virus-GAV (BYDV-GAV, AY220739), barley yellow dwarf virus-MAV (BYDV-MAV, D11028), bean leafroll virus (BLRV, AF441393), beet chlorosis virus-2a (BChV-2a, AF352024), beet mild yellowing virus (BMYV, ×83,110), beet western yellows virus (BWYV, AF473561), carrot red leaf virus (CTRLV, AY695933), cereal yellow dwarf virus-RPS (CYDV-RPS, AF235168), cereal yellow dwarf virus-RPV (CYDV-RPV, I25299), chickpea chlorotic stunt virus (CpCSV, AY956384), chickpea stunt disease associated virus (CpSDav, Y11530), cotton leafroll dwarf virus-ARG (CLRDV-ARG, GU167940), cotton leafroll dwarf virus-PV1 (CLRDV-PV1, GQ379224), cucurbit aphid-borne yellows virus (CABYV, ×76,931), groundnut rosette assistor virus (GRAV, Z68894), melon aphid-borne yellows virus (MABYV, EU000534), pea enation mosaic virus-1 (PEMV-1, L04573), potato leafroll virus (PLRV, D00530), rose spring dwarf-associated virus (RSDaV, EU024678), soybean dwarf virus isolate (SbDV, EU419584), sugarcane yellow leaf virus (ScYLV, AY236971), sweet potato leaf speckling virus (SPLSV, DQ656700), tobacco vein distorting polerovirus (TVDV, NC010732), turnip yellows virus (TuYV, ×13,063)

In summary, we report for the first time the complete genomic sequence of the cotton blue disease-associated virus, an economically important disease of cotton in South America and other parts of the world, and clearly place it within the genus *Polerovirus*.

Acknowledgments This work was supported by PICT 060 from ANPCyT and by AEGR2412 from INTA. We thank Lic. Betiana Parodi for the reviewing the manuscript. AJD is a career member of CONICET.

References

1. Brault V, Van den Heuvel JF, Verbeek M, Ziegler-Graff V, Reutenauer A, Herrbach E, Garaud JC, Guilley H, Richards K, Jonard G (1995) Aphid transmission of beet western yellow luteovirus requieres the minor capsid read-through protein P74. EMBO J 14:650–659
2. Campagnac NA, Bonacic Kresic R, Poisson J (1986) Mal de Misiones: nueva enfermedad del algodón de probable origen virósico. VI Jornadas fitosanitarias Argentinas t II, pp 503–511
3. Cauquil J, Vaissayre M (1971) La "maladie bleue" du cotonnier en Afrique: transmission de cotonnier a cotonnier par *Aphis gossypii* glover. Cot Fib Trop 26:463–466
4. Cauquil J (1977) Etudes sur une maladie d'origine virale du cotonnier: la maladie bleue. Cot Fib Trop 32:259–278
5. Corrêa RL, Silva TF, Simoes-Araújo JL, Barroso PAV, Vidal MS, Vaslin MFS (2005) Molecular characterization of a virus from the family *Luteoviridae* associated with cotton blue disease. Arch Virol 150:1357–1367
6. D'Arcy CJ, Domier LL (2005) Family *Luteoviridae*. In: Fauquet CM, Mayo MA, Maniloff J, Desselberger U, Ball LA (eds) Virus taxonomy. Eight Report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, San Diego, pp 891–900
7. Felsenstein J (1993) PHYLIP (Phylogeny Inference Package) version 3.5c. Department of Genetics, Seattle
8. Frohman MA, Dush MK, Martin GR (1988) Rapid production of full-length cDNAs from rare transcripts: amplification using a

single gene-specific oligonucleotide primer. Proc Natl Acad Sci USA 85:8998–9002
9. Garcia A, Vanduin J, Pleij CWA (1993) Differential response to frameshift signals in eukaryotic and prokaryotic translational systems. Nucleic Acids Res 21:401–406
10. Giedroc DP, Theimer CA, Nixon PL (2000) Structure, stability and function of RNA pseudoknots involved in stimulating ribosomal frameshifting. J Mol Biol 298:167–185
11. Guilley H, Richards KE, Jonard G (1995) Nucleotide sequence of Beet mild yellowing virus RNA. Arch Virol 140:1109–1118
12. Lenardon S (1994) Caracterización del agente causal de la enfermedad azul del algodón en el área centro chaqueña. Informe del PIC 320.536.-IFFIVE (CICA) del INTA
13. Mayo MA, Ziegler-Graff V (1996) Molecular biology of luteoviruses. Adv Virus Res 46:413–460
14. Miller WA, Dineshkumar SP, Paul CP (1995) Luteovirus gene expression. CRC Crit Rev Plant Sci 14:179–211
15. Pfeffer S, Dunoyer P, Heim F, Richards KE, Jonard G, Ziegler-Graff V (2002) P0 of beet western yellows virus is a suppressor of posttranscriptional gene silencing. J Virol 76:6815–6824
16. Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic tree. Mol Biol Evol 4:406–425
17. Silva TF, Corrêa RL, Castilho YG, Silvie P, Belot JL, Vaslin MFS (2008) Widspread distribution and a new recombinant spicies of Brazilian virus associated with cotton blue disease. Virol J 5:123
18. van den Heuvel JF, Bruyére A, Hogenhout SA, Ziegler-Graff V, Brault V, Verbeek M, van der Wilk F, Richards K (1997) The N-terminal region of the luteovirus readthrough domain determines virus binding to Buchnera GroEL and is essential for virus persistence in the aphid. J Virol 71:7258–7265