Variability of the coat protein gene of G*rapevine leafroll-associated virus* 3 in Brazil

Thor V.M. Fajardo[1], Érico C. Dianese[2], Marcelo Eiras[3], Daniela M. Cerqueira[2], Daniela B. Lopes[4], Marisa A.S.V. Ferreira[2] & Cláudia R.F. Martins[2]

[1]Embrapa Uva e Vinho, Cx. Postal 130, CEP 95700-000, Bento Gonçalves, RS, Brasil, e-mail: thor@cnpuv.embrapa.br; [2]Departamento de Fitopatologia, Instituto de Ciências Biológicas, Universidade de Brasília, CEP 70910-900, Brasília, DF, Brasil; [3]Instituto Biológico, Av. Cons. Rodrigues Alves, 1252, CEP 04014-002, São Paulo, SP; [4]Embrapa Semi-Árido, Cx. Postal 23, CEP 56300-000, Petrolina, PE, Brasil

Author for correspondence: Thor V.M. Fajardo

FAJARDO, T.V.M., DIANESE, E.C., EIRAS, M., CERQUEIRA, D.M., LOPES, D.B., FERREIRA, M.A.S.V. & MARTINS, C.R.F. Variability of the coat protein gene of G*rapevine leafroll-associated virus* 3 in Brazil. Fitopatologia Brasileira 32:335-340. 2007

ABSTRACT

Leafroll is an economically important disease affecting grapevines (*Vitis* spp.). Nine serologically distinct viruses, *Grapevine leafroll-associated virus*-1 through 9, are associated with this disease. The present study describes the coat protein gene sequence of four GLRaV-3 isolates occurring in the São Francisco River basin, Northeastern Brazil. The viral RNA was extracted from GLRaV-3 ELISA-positive plants and the complete coat protein gene was amplified by RT-PCR. Sequences were generated automatically and compared to the complete coat protein sequence from North American (NY1) and Chinese (Dawanhong Nº2 and SL10) GLRaV-3 isolates. The four studied isolates, named Pet-1 through 4, showed deduced amino acid identities of 98-100% (Pet-1 through 3) and 95% (Pet-4) with North American and Chinese isolates. A total of seventeen amino acid substitutions was detected among the four characterized isolates in comparison to the NY1, Dawanhong No.2 and SL10 sequences. The results indicated the existence of natural variation among GLRaV-3 isolates from grapevines, also demonstrating a lack of correlation between sequence data and geographic origin. This variability should be considered when selecting regions of the viral genome targeted for reliable and consistent virus molecular detection.

Additional keywords: *Closteroviridae, Ampelovirus*, GLRaV-3.

RESUMO

Variabilidade do gene da proteína capsidial do *Grapevine leafroll-associated virus* 3 no Brasil

O enrolamento da folha é uma doença economicamente importante que afeta videiras (*Vitis* spp.). Nove vírus sorologicamente distintos, *Grapevine leafroll-associated virus*-1 a -9, estão associados à doença. Este estudo descreve a seqüência do gene da proteína capsidial de quatro isolados do GLRaV-3 encontrados no Vale do Rio São Francisco, Nordeste do Brasil. O RNA viral foi extraído de plantas positivas em ELISA para o GLRaV-3 e o gene da proteína capsidial completo foi amplificado por RT-PCR. As seqüências foram geradas automaticamente e comparadas a seqüências completas do gene da proteína capsidial de isolados Norte-Americano (NY1) e Chineses (Dawanhong N°2 e SL10) de GLRaV-3. Os quatro isolados estudados, denominados Pet-1 a 4, exibiram identidades de aminoácidos deduzidos de 98-100% (Pet-1 a 3) e 95% (Pet-4) com os isolados Norte-Americano e Chineses. Um total de dezessete substituições de aminoácidos foi detectado entre os quatro isolados caracterizados em comparação com as seqüências do NY1, Dawanhong No.2 e SL10. Os resultados indicaram a existência de variação natural entre os isolados de GLRaV-3 de videiras, demonstrando também a falta de correlação entre dados de sequência e origem geográfica. Esta variabilidade deve ser considerada quando se selecionam regiões do genoma viral para uma detecção molecular confiável e consistente.

Palavras-chave adicionais: *Closteroviridae, Ampelovirus*, GLRaV-3.

Leafroll occurs in all major grapevine growing regions of the world, reducing productivity and quality of both wine and table grapes. It is estimated that leafroll accounts for 62% of the world losses of grape production due to viruses (Little *et al*., 2001; Gugerli, 2003). In Brazil, grapevine leafroll is an economically important disease in the São Francisco River basin, one of the main tropical suppliers of table grapes for exportation, with 8,000 ha dedicated to grape production.

Numerous reports have shown that flexuous, filamentous closterovirus-like particles ranging from 1,400 to 2,200 nm in length and ca. 12 nm in diameter are closely associated with grapevine leafroll disease. Currently, there are nine serologically distinct viruses, collectively referred to as grapevine leafroll-associated viruses and designated GLRaV-1 through GLRaV-9. GLRaV-1 and -3 are the most widespread and economically important of these viruses (Martelli *et al*., 2002; Gugerli, 2003).

GLRaV-3 is phloem-limited and spreads naturally in the field in a semi-persistent manner, being transmitted by several species of pseudococcid and coccid insects. It is a member of the family *Closteroviridae* in the genus *Ampelovirus* and the best characterized virus associated with leafroll. Particles of GLRaV-3 are about 2,000 nm and possess a coat protein of ca. 43 kDa estimated by SDS-PAGE/Western blot which is significantly larger than the coat protein of most members of the family *Closteroviridae* (Little *et al*., 2001; Martelli *et al*., 2002; Gugerli, 2003; Ling *et al*., 2004).

Enzyme-linked immunosorbent assay (ELISA) has been widely used for GLRaV-3 detection, although molecular techniques such as molecular hybridization, standard reverse transcription polymerase chain reaction (RT-PCR) or RT-PCR coupled with immunocapture (IC-RT-PCR) have also been used for GLRaV detection. However, the natural occurrence of GLRaV-3 variants and the little sequence information available for this virus may limit the use of these techniques (Turturo *et al*., 2005).

The present study describes the coat protein (CP) gene sequences of four GLRaV-3 isolates sampled from grapevines grown in the São Francisco River basin, Northeastern Brazil. The entire coat protein gene sequences of these isolates were compared with sequences available at the GenBank.

Initially, ELISA was performed to detect GLRaV-3-infected grapevines (*Vitis* spp.) originated from commercial production areas. In 2002, nine different samples (stems) of cultivar Alicante Bouschet were collected in Lagoa Grande county and of seven other cultivars (Canner, Liberty, CG40016, 10-6, CG351, CG28467 and CG39915) in Petrolina county, both locals in the State of Pernambuco, Brazil. These samples were maintained in the Biological Experimental Station of Universidade de Brasília (Brasília, DF). Other three samples (Petite Syrah, Seybel and Franciscana cultivars) from Petrolina were maintained in the *Vitis* collection of Embrapa Uva e Vinho, Bento Gonçalves, RS, Brazil, since 1997.

Samples (petioles and midribs) were taken from mature leaves, ground and diluted 1:5 (w/v) in 0.5 M Tris-HCl, pH 8.2, 0.8% (w/v) NaCl, 2% (w/v) PVP 40000, 1% (w/v) PEG 6000 and 0.05% (v/v) Tween 20, and analyzed by DAS-ELISA using antisera and controls supplied by Sanofi Diagnostic Pasteur (France). Reaction conditions were in accordance with the manufacturer's instructions. A total of 19 samples from different plants were tested and four, showing leafroll symptoms, were positively reactive in ELISA. The four GLRaV-3 ELISA positive samples were submitted to RNA extraction and RT-PCR.

RNA extraction was performed using the method described by Mackenzie *et al*. (1997), with modifications. A total of 0.1 g of leaf tissue (including veins and petioles) was ground in liquid nitrogen. Lysis buffer [4M guanidine isothiocyanate, 0.2 M sodium acetate, pH 5.0, 25 mM EDTA, 2.5% (w/v) PVP-40, and 1% (v/v) 2-mercaptoethanol] was added. The tissue was ground again and centrifuged at 3,000 g for 5 min. The supernatant was transferred to another tube and the QIAamp RNA extraction kit protocol (Qiagen) was used. When RNeasy Plant Mini kit (Qiagen) was used for RNA extraction, the cDNA reaction was performed using the LR3-9445C primer.

The reverse transcription reaction was carried out in a final volume of 20 µL (200 U of reverse transcriptase, 150 µg of random primers and 7 µL of extraction reaction). The amplification of the 942 bp fragment of the viral coat protein gene was performed using primers LR3-8504V (nts 13269 to 13288; 5' ATGGCATTTGAACTGAAATT 3') and LR3-9445C (nts 14191 to 14210; 5' CTACTTCTTTTGCAATAGTT 3') (Ling *et al*., 2004). PCR was carried out in 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 0.2 mM each dNTP, 0.8 µM each primer, 2 U *Taq* polymerase (Invitrogen), and 5 µL cDNA solution. The samples were denatured at 95ºC for 2 min following 35 cycles of denaturation at 95ºC for 1 min, annealing at 48ºC for 1 min, elongation at 72ºC for 1 min and final elongation at 72ºC for 7 min. The amplified fragments were purified using a commercial kit (GFX PCR DNA and Gel Band Purification Kit, Amersham Biosciences), and cloned into the pGEM-T Easy vector (Promega).

At least two different PCR products from each of the infected plants were cloned and sequenced in both directions (forward and reverse), in a Megabace System sequencer (Amersham-Pharmacia). M13 forward and reverse primers (Promega) were used in sequencing reactions, generating the complete coat protein gene sequence, including its 5' and 3' ends. For each primer pair, at least two independent PCR products and sequences in both orientations were generated to exclude PCR artifacts.

The generated sequences (942 bp each) from the four different isolates were compared to three GLRaV-3 complete coat protein nucleotide sequences available at the GenBank (AF037268, DQ119574, DQ911148), using BLAST and CLUSTAL X 1.8 programs (Thompson *et al*., 1997). Multiple alignments were done for the complete deduced amino acid sequences using CLUSTAL X 1.8, visualized with the GeneDoc program (Nicholas *et al*., 1997; Thompson *et al*., 1997), and the cladogram was obtained using neighbor-joining with the MEGA 3.2 program (Kumar *et al*., 2004). The phylogenetic tree, based on the alignment of the 484 nucleotides of the 5' end of the CP of 24 different GLRaV-3 isolates available at the GenBank, was constructed by maximum likelihood using PAUP 4.0b10 (Swofford, 1999) and modelfit tests using the hierarchical likelihood ratio test implemented in MODELTEST 3.04 (Posada & Crandall, 1998). The model parameters selected were HKY+G using base frequencies f(A)=0.31, f(C)=0.18, f(G)=0.29 and f(T)=0.21, and gamma distribution shape parameter alpha=0.3630. Searches were done using neighbor-joining and bootstrap percentage values were obtained for 2000 replications. A preliminary restriction enzyme study was performed using NEBcutter V2.0 (http://tools.neb.com/NEBcutter2/index.php).

As expected, a fragment of 942 bp was generated by PCR from the cDNA of all four isolates (Pet-1 and Pet-2 from different Alicante Bouschet plants, Pet-3 from cultivar CG351and Pet-4 from cultivar Petite Syrah). BLAST analysis of the 942 bp sequence of the four isolates showed deduced amino acid identities of 98-100% (Pet-1 through 3) and 95% (Pet-4) with North American (NY1, AF037268) and Chinese isolates (Dawanhong No.2, DQ119574 and SL10, DQ911148) (Table 1). An *Eco* RI site located at coat protein nucleotides 505-510 was unique to the NY1 isolate. The six other isolates (Pet-1 through 4, Dawanhong and SL10) exhibited a substitution of G to A at position 505.

The geographic origin and the GenBank accession numbers of the nucleotide sequences of GLRaV-3 isolates used in this study are indicated in Figures 1 and 2. The amino acid sequence alignment among the four Brazilian isolates, NY1, Dawanhong Nº2 and SL10 isolates revealed a total of 17 substitutions (Figure 1A). The Pet-1 isolate showed 100% of amino acid identity with Dawanhong Nº2, and Pet-1, -2 and -3 isolates were phylogeneticaly closer to Dawanhong Nº2, SL10 and NY1 isolates than to the Pet-4 isolate (Figure 1B), demonstrating a lack of correlation between sequence identity and geographic origin. This was also observed by Turturo *et al*. (2005) for GLRaV-3 isolates. This virus has a slow-movement, low transmission efficiency vector, and is spread mainly by contaminated propagative material rather than by mealybug transmission (Gugerli, 2003).

Phylogenetic analysis was performed for the 484 nucleotides of the 5' terminus of the CP sequences of the four isolates described here, and compared with 24 different GLRaV-3 sequences from the GenBank. GLRaV-3 isolates were clustered in three distinct groups with high bootstrap values (Figure 2). The Pet-1, Pet-2 and Pet-3 isolates were clustered in "group I", while the Pet-4 isolate was placed in an individual branch in "group II", also supported by a high bootstrap value (Figure 2).

Intratype pairwise evolutionary distances (number of nucleotide variations relative to the total number of nucleotides sequenced) varied from 0.53-0.74% (Pet-1, Pet-2 and Pet-3 isolates) to 7.11% (Pet-4 isolate), indicating that isolates Pet-1 to 3 represent strains of the NY1 isolate, whose sequence has been used as a reference.

In a recent study, Turturo *et al*. (2005) observed that 15% of the collection of 45 GLRaV-3 isolates was composed of a combination of two or more variants of the coat protein gene. Estimation of genetic diversity and phylogenetic analysis disclosed the possible existence of vines with mixed infections by diverging sequence variants, showing, in some cases, possible recombination events. Furthermore, differences in the genetic diversity and constraints existing in the RNA-dependent RNA polymerase (RdRp), heat-shock protein 70 homologue (HSP-70) and coat protein (CP) genes analyzed in the above mentioned study indicated a higher variability in the CP gene (Turturo *et al*., 2005).

The variability of GLRaV-3 isolates from the São Francisco river basin may represent a natural variation associated with error prone RNA-dependent RNA polymerase of this virus within grapevines. It could also be attributed to the vegetative propagation that provides broad opportunity for sequence variation, and to the perennial cycle of the plant. The variation may also be influenced by the plant-pathogen interaction that may be responsible for originating viral variants (Little *et al*., 2001). The amino acid substitutions detected were mainly non-conservative (Figure 1A). This may have some influence on the protein product, indicating

TABLE 1 - Pairwise percent identities of deduced amino acids (above diagonal line) and nucleotides (below diagonal line) of the complete sequences of the coat protein gene among the studied isolates and other *Grapevine leafroll-associated virus 3* isolates*

Isolates	Pet-1	Pet-2	Pet-3	Pet-4	NY1	Dawanhong	SL10
Pet-1	__	99.6	99.6	95.8	99.0	100.0	99.0
Pet-2	99.8	__	99.3	95.5	98.7	99.6	98.7
Pet-3	99.7	99.6	__	95.5	98.7	99.6	98.7
Pet-4	93.4	93.3	93.2	__	94.8	95.8	95.5
NY1	99.4	99.3	99.2	92.8	__	99.0	98.0
Dawanhong	99.6	99.5	99.4	93.2	99.1	__	99.0
SL10	99.3	99.2	99.1	92.9	98.8	99.2	__

*The isolates are described as following: Origin (Isolate name = GenBank accession number). Brazil (Pet-1 = DQ680141, Pet-2 = DQ680142, Pet-3 = DQ062152 and Pet-4 = AY753208); USA (NY1 = AF037268 or NC_004667); China (Dawanhong Nº2 = DQ119574 and SL10 = DQ911148).

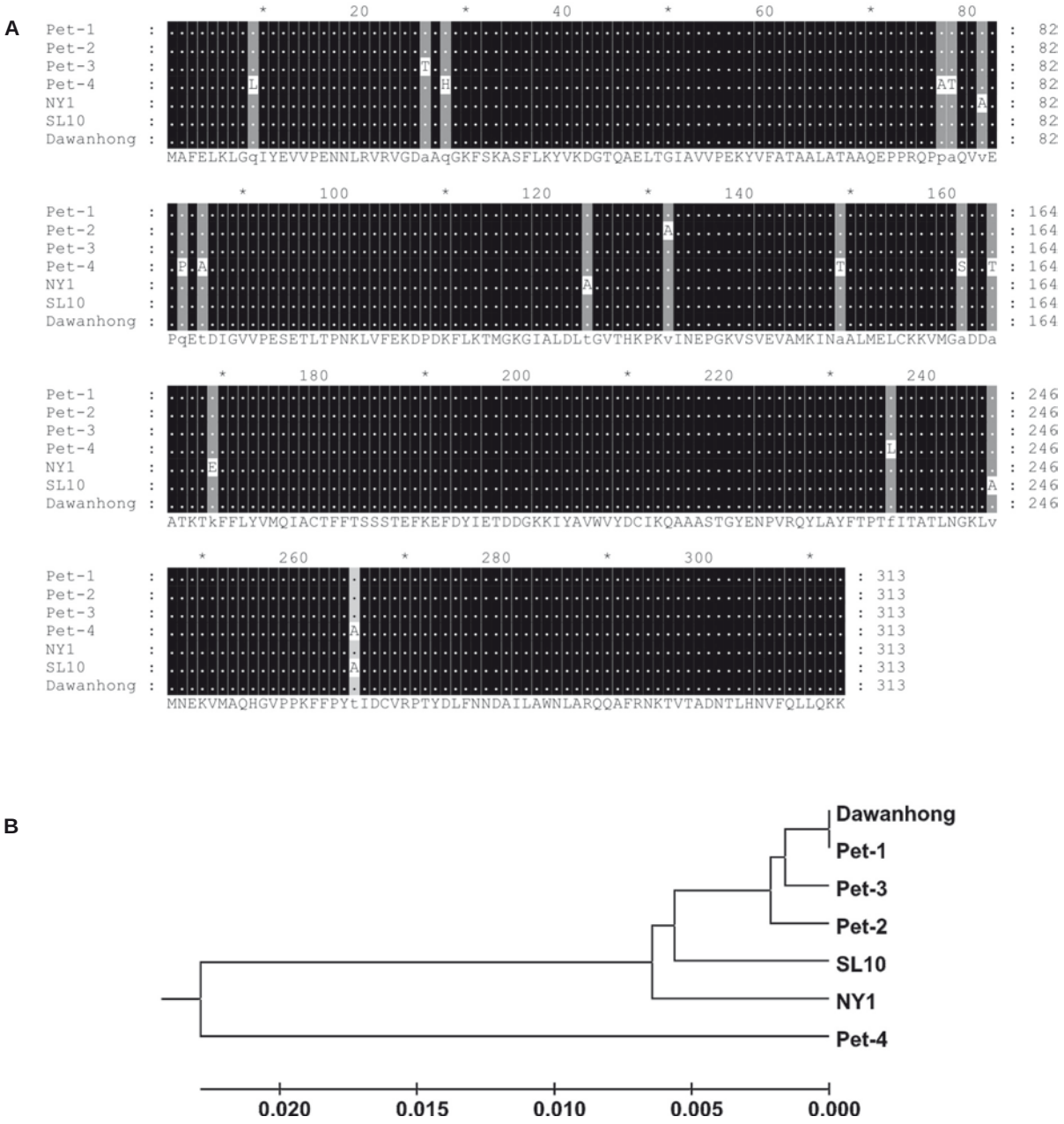

FIG. 1 - (A) Multiple alignment of the complete deduced amino acid sequences of the coat protein of GLRaV-3 isolates from Brazil, China (Dawanhong, DQ119574 and SL10, DQ911148) and USA (NY1, AF037268 or NC_004667). Pet-1 (DQ680141), Pet-2 (DQ680142), Pet-3 (DQ062152) and Pet-4 (AY753208) are isolates from Northeastern Brazil. The consensus amino acids sequence is shown at the bottom. (B) Cladogram based on the alignment of the amino acids sequences; searches were done using neighbor-joining. Bar=number of substitutions per site.

a possible alteration of physical-chemical properties of the original residues. However, the consequences of non-conservative amino acid changes on viral biological or pathogenic properties are still unknown.

Characterizing GLRaV-3 isolates is an important aspect of the development of tools for the detection of this virus. In a previous study, we were able to analyze the variability of the 3' terminal of the polymerase gene of the same GLRaV-3 isolates. We observed a lower frequency of substitutions in the deduced amino acid sequence of this gene (97.1-100% compared to NY1) (Fajardo *et al*., 2002; Dianese *et al*., 2005) when compared to the coat protein gene described in the present work. Higher conservation is probably explained by the functional activity of the polymerase gene, thus, variations could affect the fitness of the virus.

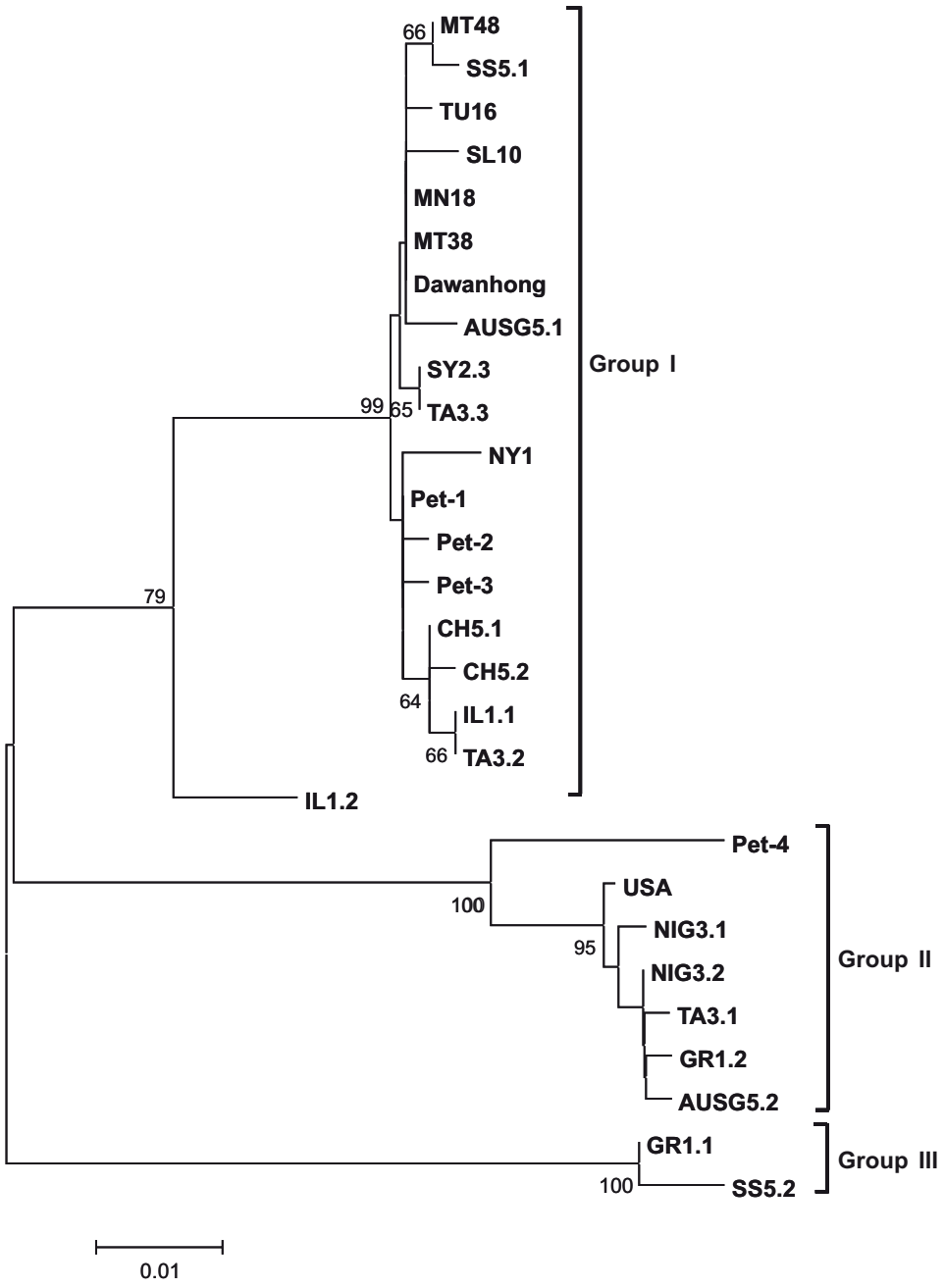

FIG. 2 - Unrooted phylogenetic tree based on the alignment of the 484 nucleotides of the 5' terminus of the CP of different GLRaV-3 isolates. Searches were done using neighbor-joining. Bootstrap percentage values are shown on the branches. Bar=number of substitutions per site. The isolates are described as following: Origin (Isolate name = GenBank accession number). Brazil (Pet-1 = DQ680141, Pet-2 = DQ680142, Pet-3 = DQ062152 and Pet-4 = AY753208); China (Dawanhong Nº2 = DQ119574, SL10 = DQ911148, CH5.1 = AJ606356, CH5.2 = AJ606357); USA (NY1 = AF037268 or NC_004667); Tromelin Island (USA = AJ606340); Nigeria (NIG3.1 = AJ606341, NIG3.2 = AJ606358); Italy (TA3.1 = AJ606345, MN18 = AJ606349, MT38 = AJ606350, MT48 = AJ606351, TA3.2 = AJ606344, TA3.3 = AJ606346, SS5.1 = AJ606348, SS5.2 = AJ606347); Greece (GR1.1 = AJ606342, GR1.2 = AJ606343); Áustria (AUSG5.1 = AJ606338, AUSG5.2 = AJ606339); Israel (IL1.1 = AJ606355, IL1.2 = AJ606354); Tunisia (TU16 = AJ606353) and Syria (SY2.3 = AJ606352).

The information about sequence diversity among variants may help to select regions of the virus genome targeted for specific virus detection. ELISA is routinely used in phytosanitary screening and certification programmes for the detection of GLRaVs. As ELISA is based on the specificity of interaction between antibodies and the virus coat protein, the occurrence of mutations in the coat protein may limit the ability of virus detection

by this method. Intraspecies sequence diversity in GLRaVs makes the selection of conserved regions for primer design rather difficult, affecting reliable detection by PCR. These results highlight the relevance of knowledge of sequence variability when designing primers for a reliable detection of all known virus variants (Little *et al*., 2001; Dovas & Katis, 2003). Moreover, small divergences observed in the coat protein gene of the four studied GLRaV-3 isolates suggest that the biological properties between variants of the this virus may differ, as mentioned by Jooste & Goszczynski (2006), based on variability detected in genes other than coat protein.

ACKNOWLEDGEMENTS

We thank Kevin P. Karol and Osmar Nickel for reviewing this manuscript. This work was supported by Embrapa, CNPq, FINATEC and Banco do Nordeste. E.C. Dianese was supported by a CAPES fellowship.

REFERENCES

DIANESE, E.C., RAMALHO, E.D., CERQUEIRA, D.M., LOPES, D.B., FAJARDO, T.V.M., FERREIRA, M.A.S.V. & MARTINS, C.R.F. Variability of the 3' terminal of the polymerase gene of *Grapevine leafroll-associated virus 3* isolates from Vale do São Francisco. Fitopatologia Brasileira 30:173-176. 2005.

DOVAS, C.I. & KATIS, N.I. A spot multiplex nested RT-PCR for the simultaneous and generic detection of viruses involved in the aetiology of grapevine leafroll and rugose wood of grapevine. Journal of Virological Methods 109:217-226. 2003.

FAJARDO, T.V.M., KUHN, G.B., EIRAS, M. & NICKEL, O. Detecção de *Closterovirus* em videira e caracterização parcial de um isolado do *Grapevine leafroll-associated virus 3*. Fitopatologia Brasileira 27:58-64. 2002.

GUGERLI, P. Grapevine leafroll and related viruses. Extended Abstracts, 14th Meeting of the International Council for the Study of Virus and Virus-like Diseases of the Grapevine, Locorotondo, Italy. 2003. pp. 25-31.

JOOSTE, A.E.C. & GOSZCZYNSKI, D.E. Differentiation between two distinct molecular variants of GLRaV-3. Extended Abstracts, 15th Meeting of the International Council for the Study of Virus and Virus-like Diseases of the Grapevine, Stellenbosch, South Africa. 2006. pp. 32-33.

KUMAR, S., TAMURA, K. & NEI, M. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Briefings in Bioinformatics 5:150-163. 2004.

LING, K.S., ZHU, H.Y. & GONSALVES, D. Complete nucleotide sequence and genome organization of *Grapevine leafroll-associated virus 3*, type member of the genus *Ampelovirus*. Journal of General Virology 85:2099-2102. 2004.

LITTLE, A., FAZELI, C.F. & REZAIAN, M.A. Hypervariable genes in *Grapevine leafroll-associated virus 1*. Virus Research 80:109-116. 2001.

MACKENZIE, D.J., MCLEAN, M.A., MUKERJI, S. & GREEN, M. Improved RNA extraction from woody plants for the detection of viral pathogens by reverse transcription-polymerase chain reaction. Plant Disease 81:222-226. 1997.

MARTELLI, G.P., AGRANOVSKY, A.A., BAR-JOSEPH, M., BOSCIA, D., CANDRESSE, T., COUTTS, R.H.A., DOLJA, V.V., FALK, B.W., GONSALVES, D., JELKMANN, W., KARASEV, A.V., MINAFRA, A., NAMBA, S., VETTEN, H.J., WISLER, G.C. & YOSHIKAWA, N. The family *Closteroviridae* revised. Archives of Virology 147:2039-2044. 2002.

NICHOLAS, K.B., NICHOLAS, H.B. & DEERFIELD, D.W.II GeneDoc: Analysis and visualization of genetics variation. EMBNEW News 4:14. 1997.

POSADA, D. & CRANDALL, K.A. MODELTEST: testing the model of DNA substitution. Bioinformatics 14:817-818. 1998.

SWOFFORD, D.L. PAUP*: Phylogenetic Analysis Using Parsimony (*and other methods), version 4.0. Sinauer Associates, Sunderland, Massachusetts, USA. 1999.

THOMPSON, J.D., GIBSON, T.J., PLEWNIAK, F., JEANMOUGIN, F. & HIGGINS, D.G. The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research 25:4876-4882. 1997.

TURTURO, C., SALDARELLI, P., YAFENG, D., DIGIARO, M., MINAFRA, A., SAVINO, V. & MARTELLI, G.P. Genetic variability and population structure of *Grapevine leafroll-associated virus 3* isolates. Journal of General Virology 86:217-224. 2005.

Received 6 Dezember 2006 - Accepted 15 June 2007 - FB 6129