Acta virologica **56**: 119 – 124, 2012 doi:10.4149/av_2012_02_119

Molecular analysis of gooseberry vein banding associated virus

K. PETRZIK, J. PŘIBYLOVÁ, J. ŠPAK

Department of Plant Virology, Institute of Plant Molecular Biology, Biology Centre of the Academy of Sciences of the Czech Republic, v.v.i., Branišovská 31, 370 05 České Budějovice, Czech Republic

Received February 23, 2012; accepted April 19, 2012

Summary. – The intraspecies variability of gooseberry vein banding associated virus (GVBaV) was analyzed by using 5 complete and 9 partial sequences and compared with other badnavirus species. GVBaV was recognized to be a monophyletic and very homogeneous species with up to 94% identities in distinct proteins. Analysis of non-synonymous and synonymous substitution ratios (d_{NS}/d_S) revealed higher values for ORF4 in comparison with other genes. This could reflect different evolutionary pressure upon this ORF. A highly variable region with possible diagnostic value has been localized in the intergenic region of the virus.

Keywords: badnavirus; complete genome; phylogeny

Badnaviruses are highly adapted viruses with a double-stranded DNA genome often restricted to plant vascular system (Geering and Hull, 2012). Their infection can lead to development of flecks, freckles, streaks, chlorosis or yellowing along veins, or there may be no visible symptoms whatsoever. In the past decade, badnaviruses have been the subject of increased attention since the discovery of their endogenous sequences in various plant genomes (Geering *et al.*, 2005, 2011; Gayral *et al.*, 2008; Kenyon *et al.*, 2008; Bousalem *et al.*, 2009; James *et al.*, 2011) and risk of their putative spread through infected germplasm and/or activation by an abiotic stress during propagation (Dallot *et al.*, 2001).

The badnavirus genome contains monopartite, open circular, double-stranded DNA with single-strand discontinuity at one site in each strand. Upon entry into the cell, the discontinuities are sealed and transcribed by host DNA-dependent RNA polymerases to yield a transcript larger than a genome. Only one strand of DNA is the coding sequence. Open reading frames (ORFs) found on the minus strand are not thought to be expressed in these viruses (Rothnie *et al.*, 1994). Three ORFs are believed to be typical for badnaviruses, although additional ORFs have been found in some species: ORF IV, V and VI in citrus yellow mosaic badnavirus (Borah *et al.*, 2009), ORF4 coding for a putative protein of 95 amino acids (aa) inside the ORF3 in dioscorea sansibarensis bacilliform virus (Seal and Muller, 2007), and four ORFs with products longer than 100 aa inside and after ORF3 in dracaena mottle virus (Su *et al.*, 2007). The function of the ORF1 product (about 23 kDa in GVBaV) is unknown. The product of ORF2 (15 kDa in GVBaV) contains two motives conserved in badnaviruses and caulimoviruses: the $^{K}/_{R}Q^{N}/_{L}N$ motif at the N-terminal region and the $^{K}/_{Q}DPK$ motif at the C-terminal region. The latter motif plays a role in binding of the protein to DNA (Borah *et al.*, 2009). The product of ORF3 is a polyprotein (216 kDa) harboring a movement protein, a coat protein, an RNA-binding site, as well as aspartyl protease and a reverse transcriptase/RNAse H (RT/RNase H), respectively. In contrast to the pol gene product of retroviruses, it has no integrase (Geering and Hull 2012). To date, GVBaV has not been found integrated into the host genome.

The use of error-prone reverse transcription in the replication of badnaviruses is assumed to be a driving force for creating variants between isolates. Mostly the RT/RNase H coding sequence is used for detecting badnaviruses and evaluating differences. Comparison of this sequences in yams (*Dioscorea*) revealed a very high degree of variability

E-mail: petrzik@umbr.cas.cz; phone: +420-38-7775549.

Abbreviations: GVBaV = gooseberry vein banding associated virus; GVCV = grapevine vein-clearing virus; SCBV = sugarcane bacilliform virus; ORF(s) = open reading frame(s); RT = reverse transcriptase, RC = red currant; BC = black currant; GB = gooseberry

between isolates and resulted in identifying many novel badnaviruses (Kenyon *et al.*, 2008; Bousalem *et al.*, 2009), as the recent species demarcation criteria in the badnavirus genus suppose more than 20% nucleotide (nt) differences in this region (Geering and Hull 2012). Seven phylogenetic groups have been characterized according to the 529 nt long RT/RNase H segment among 35 sugarcane bacilliform virus (SCBV) isolates, and three additional SCBV viruses were proposed using the 20% nt difference threshold (Muller *et al.*, 2011).

Our knowledge of sequence variability and its putative diagnostic and distinguishing potential in GVBaV is not yet sufficient. Although GVBaV infects Ribes (currants and gooseberry) species worldwide, it causes only less noticeable or transient symptoms in black or red currant and reduced vigor and yield in gooseberry cultivars (Adams 1979; Jones *et al.*, 2001). We performed a complete genome sequence of one Czech isolate from red currant and evaluated the variability of distinct genes, the RT/RNase H sequence, and variability of a noncoding region applicable to isolate diagnostics. Further isolates were obtained from shoots of the black currant 'Titania'; the red currants 'Holandský červený', 'Vitan', 'Rubigo' and one unknown cultivar, the white currant 'Blanka', and the gooseberry 'Priori' originating from a field germplasm collection and plantations during 2009–2011 in four regions within the Czech Republic.

DNA was isolated from phloem material using the NucleoSpin plant II kit (Macherey-Nagel) according to the manufacturer′s recommendations, including 15 min RNAse digestion of the plant extract and elution with 30 µl of water.

GVBaV-specific primers (Jones *et al.*, 2001; Xu *et al.*, 2011) were used to amplify overlapping segments covering the whole genome of GVBaV. The noncoding region between nt 7248 and 345 (numbered according to the sequence of the RIB9001 isolate, GenBank Acc. No. HQ852251) was amplified with primers 293: 5′-GATTCGTCATCGCTTACGCCATC and 294: 5′-GTC GGCTACTGCTCTAGATACTC. All PCR products were cloned into pJET vector (Fermentas) and further sequenced with pJET forward and reverse primers using a BigDye ver.3.1 sequencing kit (Applied Biosystems). The complete genome sequence from red currant 'Holandský červený' (RC HC) was deposited in GenBank with accession number JQ316114. Partial nucleotide sequences of the RT/RNase H region and of the noncoding region of the other isolates were deposited under AC: JQ388484-94 and JQ316115-120. The complete genome sequences of an RC isolate from red currant from the Netherlands, the isolate GB1 from gooseberry from the United Kingdom, isolate RIB9001 from red currant from the USA, a BC isolate infecting black currant from Canada (HQ852248-51) (Xu *et al.*, 2011), and the partial sequence AF298883 (Jones *et al.*, 2001) were used for phylogenetic analyses (Table 1).

The complete genome of the Czech RC HC isolate consists of 7659 nt. Three typical open reading frames of 672, 441, and 5730 nt and an additional ORF4 of 180 nt were identified. ORF1 and 2 are in the same reading frame but are separated by a termination codon. ORF2 and 3 overlap by 41 nt and ORF3 is in a –2 translational frame relative to ORF2. These

Table 1. Isolates and sequences used in this work

Isolate	Host	AC	Reference
RC HC	red currant cv. Holandský červený	JQ316114	this work
RC Roz 1	red currant unknown cv.	JQ316115	this work
RC Roz 2	red currant unknown cv.	JQ316116	this work
WC B	white currant cv. Blanka	JQ316117	this work
RC Lhe	red currant cv. Vitan	JQ316118	this work
RC VL	red currant cv. Vitan	JQ316119	this work
GB Hol 5/39	gooseberry cv. Priori	JQ316120	this work
GB1	gooseberry cv. Weisse Neckartaal	HQ852248	Xu *et al.* (2011)
RC	red currant cv. Augustus	HQ852249	Xu *et al.* (2011)
BC	black currant unknown cv.	HQ852250	Xu *et al.* (2011)
RIB9001	red currant cv. London Market Ribes sp.	HQ852251	Xu *et al.* (2011)
		AF298883	Jones *et al.* (2001)
BC 19/5	black currant cv. Titania	JQ388484	this work
RC 10/11	red currant cv. Vitan	JQ388485	this work
RC 10/10	red currant cv. Vitan	JQ388486	this work
RC 10/3	red currant cv. Rubigo	JQ388487	this work
RC 9/6	red currant cv. Holandský červený	JQ388488	this work
RC 9/5	red currant cv. Holandský červený	JQ388489	this work
RC 9/4	red currant cv. Holandský červený	JQ388490	this work
GB Pri	gooseberry cv. Priori	JQ388491	this work

Table 2. Amino acid sequence identity (%) between corresponding proteins of RC HC, BC, GB1, RC, and RIB9001 GVBaV isolates and sequence AF298883

	ORF I				ORF II				ORF III				ORF IV				
	BC	GB1	RC	RIB 9001	BC	GB	RC	RIB 9001	BC	GB	RC	RIB 9001	BC	GB	RC	RIB 9001	AF28883
RC HC	98.7	99.1	99.1	98.7	97.3	95.2	93.9	95.9	98.9	98.8	98.9	98.9	96.7	98.3	85.3	95.0	95.0
BC		99.6	99.6	99.1		95.2	93.9	95.9		99.0	99.3	99.3		95.0	80.9	91.7	91.7
GB1			100.	99.6			97.3	98.0			99.0	99.0			82.4	96.7	96.7
RC				99.6				96.6				99.1				82.4	79.4
RIB9001																	93.3

ORFs encode proteins of 224, 147, and 1910 aa (25.9 kDa, 16 kDa, and 215.8 kDa). As mentioned above, there are badnaviruses in whose genomes additional ORFs have been found (Borah *et al.*, 2009; Seal and Muller, 2007; Su *et al.*, 2007). GVBaV contained a fourth putative ORF in all isolates. The putative ORF4 is arranged in a convenient context in the genome. Its start codon overlaps the stop codon of ORF3 and is in a −1 translational frame relative to the preceding ORF3. It encodes for a protein 60 aa long (68 aa in the RC isolate), or 6.4 kDa, or it could be transcribed continuously with ORF3 and thereby produce a protein of about 222 kDa. However, no such protein has been detected in infected plants. Identity of coded proteins with corresponding proteins of BC, GB1, RC and RIB9001 isolates are in the range 93.9% to 100% (Table 2) and are in accordance with intraspecies identities of sugarcane bacilliform badnaviruses (Muller *et al.*, 2011) and strawberry vein banding pararetrovirus (Mráz *et al.*, 1998). Amino acid identity of the putative protein coded

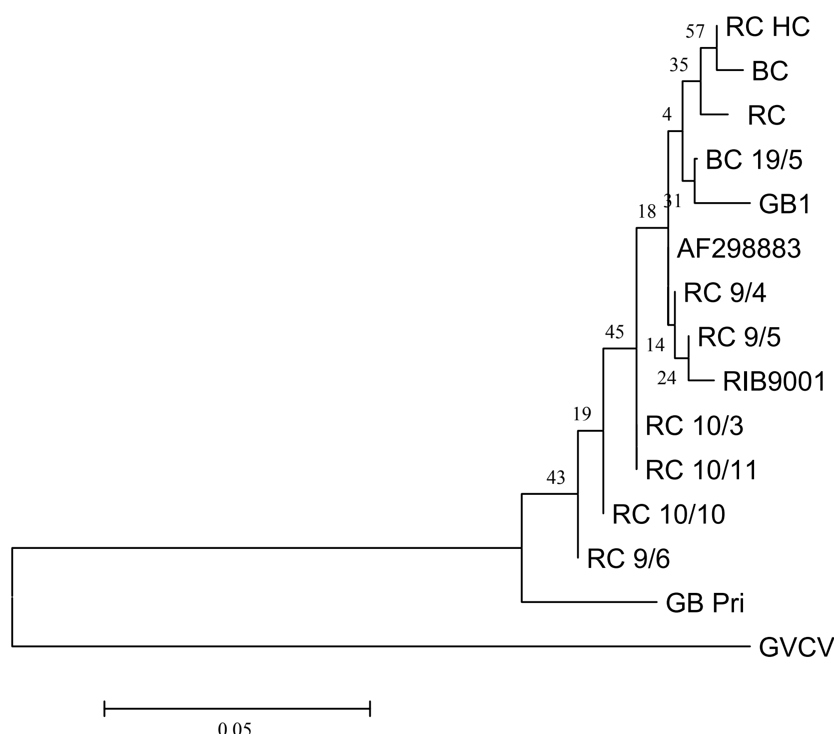

Fig. 1

Rooted neighbor-joining tree obtained from alignment of the RT/RNaseH region sequences

Bootstrap analysis was made using 1000 replicates. Grapevine vein-clearing virus (GVCV, NC_015784) was used as an out-group. Viruses and isolates were as follows: red currant 'Holandský červený' (RC HC, JQ316114), (RC 9/6, JQ388488), (RC 9/5, JQ388489), (RC 9/4, JQ388490); red currant 'Vitan' (RC 10/10, JQ388486), (RC 10/11, JQ388485); black currant 'Titania' (BC 19/5, JQ388484); gooseberry 'Priori' (GB Pri, JQ388491); gooseberry 'Wiesse Neckartaal' (GB1, HQ852248); red currant 'Augustus' (RC, HQ852249); black currant (BC, HQ852250); red currant 'London Market' (RIB9001, HQ852251); Ribes sp. (AF298883); and red currant 'Rubigo' (RC 10/3, JQ388487).

```
BC       (1) MIYLIMEADAPDDFTINDSSPLTPSALLLFYFCKCACSALSQGILSLMSFSHTFGGSGCF--------
RC HC    (1) MIYLIMEADAPDDFTINDSSSLTPSALLLLYFCKCACSALSQGILSLMSFSHTFGGSGCF--------
GB1      (1) MIYLIMEADAPDDFTINDSSSLTPSALLLLYFCKCASSALSQGILSLMSFSHTFGGSGCF--------
RC       (1) MTYPIMEADAPDDFTINDSSSITPSALLLLYFCKCACSALSQGILSLMSFSHTFGGSGCFEYKHYCKS
RIB9001  (1) MIYPIVEADAPDDFTINDSSSLTPSALLLLYFCKCASSALSQGILSLMSFSHTFGGSGCF--------
AF298883 (1) MIYLIMEADAPDDFTINDSSSLTPSALLLLYFCKCVSSALSQGILSLMSFSHTFGGGGCF--------
Alignment    *.*.*.**************..*******.*****..*******************.***........
```

Fig. 2

Amino acid alignment of putative ORF4

Non-conserved positions are highlighted.

by ORF4 that is present in all isolates sequenced to date is 79.4% to 98.3% (Table 2), but if trimming the eight aa of the RC sequence, the identity is 90% to 98.3%, comparable with identities of ORFs 1–3.

Sequence analysis of non-synonymous and synonymous substitution ratios d_{NS}/d_S of distinct GVBaV genes of five isolates (six with the trimmed ORF4) were conducted in MEGA5 (Tamura *et al.*, 2011) using the modified Nei-Gojobori model. Analysis showed large differences between the genes, as the d_{NS}/d_S of the putative ORF4 (0.622) was about 10 times higher than that of ORF1 (0.062) and ORF3 (0.061) and about 4 times greater than the d_{NS}/d_S of ORF2 (0.165). This could reflect different evolutionary pressure on this ORF.

Nucleotide sequence comparison of the 726 nt long RT/RNase H segment of 14 available GVBaV isolates (Table 1) showed a monophyletic distribution having no correlation with the host species or geographic origin of isolates (Fig. 1). Nucleotide sequence identity of the isolates was 97.5% to 100% (average 98.9%) in this region. This is in agreement with the intraspecies nucleotide sequence identity of three SCBV viruses, which was between 88% and 99% (Muller

Fig. S1

Neighbor-joining tree obtained from alignment of the variable noncoding region

Bootstrap analysis was applied using 1000 replicates. Grapevine vein-clearing virus (GVCV, NC_015784) was used as an out-group. Viruses and isolates were as follows: red currant 'Holandský červený' (RC HC, JQ316114); red currant unknown cultivar (HC Roz 1, JQ316115); (HC Roz 2, JQ316116), white currant 'Blanka' (WC B, JQ316117), red currant 'Vitan' (RC Lhe, JQ316118); (RC VL, JQ316119); gooseberry 'Priori' (GB Hol 5/39, JQ316120); gooseberry 'Wiesse Neckartaal' (GB1, HQ852248); red currant 'Augustus' (RC, HQ852249); black currant (BC, HQ852250); red currant 'London Market' (RIB9001, HQ852251); and Ribes sp. (AF298883).

et al., 2011), but not with banana streak virus, where high variability in the corresponding coding segment has been detected even in one geographic location (Jaufeerally-Fakim *et al.*, 2006). In this point, GVBaV is an exceptionally homogeneous species.

A highly variable part of the GVBaV genome was recognized within the intergenic region between the end of ORF3 and start of ORF1. This finding is surprising because this region contains promoter elements and important secondary structures. This region in GVBaV is 841 to 855 nt long, depending on the isolate. The negative-strand primer-binding site (Geering *et al.*, 2011), poly (A) signal (AATAAA), TATA box (TTATTT), and several GT-1 boxes (GGAAAA) were identified there. There is a strong (−3.34 kcal/mol), stable secondary structure (5´-AGACGUGACAGUCU, with a 4 nt stem and 6 nt loop) conserved in all GVBaV isolates before the AUG start codon of ORF1 (analyzed with the RNA fold in RNA web server on http://rna.tbi.univie.ac.at/cgi-bin/RNA fold.cgi, Zuker and Stiegler, 1981).

We analyzed sequences of 12 isolates in this region and tried to find some evolutionary consequences. A hypervariable region was located between nt 7473 and 7594 (numbered according to isolate RIB9001). Phylogenetic analysis of the intergenic region nucleotide sequences obtained by the neighbor-joining method showed no reliable clustering of isolates depending upon either region or host plant (Fig. S1, supplementary material). In any case, this region is of diagnostic value for discriminating distinct isolates.

Acknowledgements. This work was supported by project AV0Z50510513 of the Academy of Sciences of the Czech Republic and grant QH91224 of the Czech National Agency for Agricultural Research. The authors thank O. Kubešová for her excellent technical assistance.

References

Adams AN (1979): The effect of gooseberry vein banding virus on the growth and yield of gooseberry and red currant. J. Hortic. Sci. 54, 23–25.

Borah BK, Johnson AMA, Gopal DVRS, Dasgupta I (2009): Sequencing and computational analysis of complete genome sequences of Citrus yellow mosaic badnavirus from acid lime and pummelo. Virus Genes 39, 137–140. http://dx.doi.org/10.1007/s11262-009-0367-9

Bousalem M, Durand O, Scarcelli N, Lebas BSM, Kenyon L, Marchand J-L, Lefort F, Seal SE (2009): Dilemmas caused by endogenous pararetroviruses regarding the taxonomy and diagnosis of yam (Dioscorea spp.) badnaviruses: analyses to support safe germplasm movement. Arch. Virol. 154, 297–314. http://dx.doi.org/10.1007/s00705-009-0311-2

Dallot S, Acuna P, Rivera C, Ramirez P, Cote F, Lockhart BEL, Caruana ML (2001): Evidence that the proliferation stage of micropropagation procedure is determinant in the expression of Banana streak virus integrated into the genome of FHIA 21 hybrid (MusaAAAB). Arch. Virol. 146, 2179–2190. http://dx.doi.org/10.1007/s007050170028

Gayral P, Noa-Carrazana J-C, Lescot M, Lheureux F, Lockhart BEL, Matsumoto T, Piffanelli P, Iskra-Caruana M-L (2008): A single banana streak virus integration event in the banana genome as the origin of infectious endogenous pararetrovirus. J. Virol. 82, 6697–6710. http://dx.doi.org/10.1128/JVI.00212-08

Geering ADW, Hull R (2012): Caulimoviridae. In CM Fauquet, MA Mayo, J Maniloff, U Desselberger, LA Ball (Eds): Virus taxonomy, Classification and Nomenclature of Viruses, Ninth Report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, Amsterdam, pp. 429–443.

Geering ADW, Olszewski NE, Harper G, Lockhart BEL, Hull R, Thomas JE (2005): Banana contains a diverse array of endogenous badnaviruses. J. Gen. Virol. 86, 511–520. http://dx.doi.org/10.1099/vir.0.80261-0

Geering ADW, Parry JN, Thomas JE (2011): Complete genome sequence of a novel badnavirus, banana streak IM virus. Arch. Virol. 156, 733–737. http://dx.doi.org/10.1007/s00705-011-0946-7

James AP, Geijskes RJ, Dale JL, Harding RM (2011): Development of a novel rolling-circle amplification technique to detect Banana streak virus which also discriminates between integrated and episomal virus sequences. Ann. Appl. Biol. 158, 346–353. http://dx.doi.org/10.1111/j.1744-7348.2011.00466.x

Jaufeerally-Fakim Y, Khorugdharry A, Harper G (2006): Genetic variants of banana streak virus in Mauritius. Virus Res. 115, 91–98. http://dx.doi.org/10.1016/j.virusres.2005.06.015

Jones AT, McGavin WJ, Geering ADW, Lockhart BEL (2001): A new badnavirus in Ribes species, its detection by PCR, and its close association with Gooseberry vein banding disease. Plant Dis. 85, 417–422. http://dx.doi.org/10.1094/PDIS.2001.85.4.417

Kenyon L, Lebas MSB, Seal SE (2008): Yams (Dioscorea spp.) from the South Pacific islands contain many novel badnaviruses: implications for international movement of yam germplasm. Arch. Virol. 153, 877–889. http://dx.doi.org/10.1007/s00705-008-0062-5

Mráz I, Petrzik K, Šíp M, Fránová-Honetšlegrová J (1998): Variability of coat protein sequence homology among American and European sources of Strawberry vein banding virus. Plant Dis. 82, 544–546. http://dx.doi.org/10.1094/PDIS.1998.82.5.544

Muller E, Dupuy V, Blondin L, Bauffe F, Daugrois JH, Nathalie L, Iskra-Caruana M-L (2011): High molecular variability of sugarcane bacilliform viruses in Guadeloupe implying the existence of at least three new species. Virus Res. 160, 414–419. http://dx.doi.org/10.1016/j.virusres.2011.06.016

Rothnie HM, Chapdelaine Y, Hohn T (1994): Pararetroviruses and retroviruses: a comparative review of viral structure and gene expression strategies. Adv. Virus Res.

44, 1–67. http://dx.doi.org/10.1016/S0065-3527(08)60327-9

Seal S, Muller E (2007): Molecular analysis of a full-length sequence of a new yam badnavirus from Dioscorea sansibarensis. Arch. Virol. 152, 819–825. http://dx.doi.org/10.1007/s00705-006-0888-7

Su L, Gao S, Huang Y, Ji C, Wang D, Ma Y, Fang R, Chen X (2007): Complete genomic sequence of Dracaena mottle virus, a distinct badnavirus. Virus Genes 35, 423–429. http://dx.doi.org/10.1007/s11262-007-0102-3

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011): MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731–2739. http://dx.doi.org/10.1093/molbev/msr121

Xu D, Mock R, Kinard G, Li R (2011): Molecular analysis of the complete genomic sequences of four isolates of Gooseberry vein banding associated virus. Virus Genes 43, 130–137. http://dx.doi.org/10.1007/s11262-011-0614-8

Zuker M, Stiegler P (1981): Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acid Res. 9, 133–148. http://dx.doi.org/10.1093/nar/9.1.133