

Genetic variability of blueberry scorch virus isolates from highbush blueberry in New York State

Elzbieta Kalinowska¹ · Patricia Marsella-Herrick² · Marc Fuchs²

Received: 13 December 2014 / Accepted: 15 March 2015 / Published online: 27 March 2015
© Springer-Verlag Wien 2015

Abstract The genetic variability of blueberry scorch virus (BIScV) isolates from New York was determined within a portion of the RNA-dependent RNA polymerase gene and the triple gene block and coat protein (CP) genes. Phylogenetic analysis of 19 New York isolates and other isolates for which sequence information is available in GenBank revealed two distinct clades, regardless of the coding region analyzed, and limited variability within (0.029 ± 0.007) and between (0.183 ± 0.032) phylogroups. Recombination events were identified in the CP gene of three New York isolates, and codons of the five BIScV genes characterized were found to be under neutral or negative selective pressure.

Keywords *Blueberry scorch virus* · Evolution · Genetic variability · Highbush blueberry · Purifying selection · Recombination

Electronic supplementary material The online version of this article (doi:10.1007/s00705-015-2402-6) contains supplementary material, which is available to authorized users.

✉ Marc Fuchs
mf13@cornell.edu
Elzbieta Kalinowska
ela_kalinowska@op.pl; marc.fuchs@cornell.edu

¹ Department of Plant Pathology, Faculty of Horticulture, Biotechnology and Landscape Architecture, Warsaw University of Life Sciences, 159 Nowoursynowska Street, 02-776 Warsaw, Poland

² Section of Plant Pathology and Plant Microbe Biology, School of Integrative Plant Science, New York State Agriculture Experiment Station, Cornell University, Geneva, NY 14456, USA

Blueberry scorch virus (BIScV), a member of the genus *Carlavirus* in the family *Betaflexiviridae* [2], is one of the most widespread viruses of highbush blueberry (*Vaccinium corymbosum* L.) [30]. The virus has been reported in the United States [4, 5, 7, 8, 23], Canada [28], Italy [16], and Poland [21]. Expression of BIScV symptoms depends on the cultivar, viral strain, and environmental conditions [4, 28]. In severe cases of infection during early bloom, blossoms and leaves blight and dry up, resulting in a general decline. Blight symptoms can be restricted to single blossoms and shoots or be more evenly distributed throughout a diseased bush. BIScV infection can be latent for one or two years, or be symptomless in some cultivars [4].

BIScV is disseminated by vegetative propagation and grafting through the use of infected cuttings. The virus is also transmitted by aphids (*Ericaphis fimbriata* Richards and *Myzuspersicae* Sultzer) in a non-persistent mode [4]. The genome of BIScV consists of a single-stranded RNA molecule of 8,512 nucleotides (nt) in length. The viral RNA is presumed to have a 5' cap structure and a 3' polyadenylated tail. The genome has six open reading frames (ORF) that encode (from the 5' end to the 3' end) the RNA-dependent RNA polymerase (RdRp), a triple gene block of three overlapping coding regions, e.g., triple gene block 1 (TGB1), triple gene block 2 (TGB2), and triple gene block 3 (TGB3), the coat protein (CP), and ORF6 [2, 5]. Expression of the RdRp is from the genomic RNA, while the other ORFs are presumably expressed from subgenomic RNAs.

So far, genetic variability studies of BIScV have predominantly focused on partial CP gene fragments of isolates from Canada, Italy, and the United States, with nucleotide sequence identity ranging from 72 to 99 % [15, 27]. No information is available on genetic variability

outside of the CP gene with the exception of three isolates, one from New Jersey (NJ) and two from British Columbia (BC-1 and BC-2), for which the complete genome sequence is available. The aim of this study was to determine the genetic variability of BISCv isolates from New York State within the full-length CP gene, the three full-length TGB genes, and a partial RdRp gene. The genome sequences from New York isolates were compared with those of other isolates and analyzed to determine the strength and direction of selective pressure acting on BISCv populations.

Blueberry leaves and flowers were collected during the spring-summer of 2014 in plantings located in Steuben (four plantings) and Wayne (one planting) counties in New York State. Plantings in these two counties, which are 160 km apart, were selected for this study because previous survey efforts documented the presence of BISCv [7]. Selected plantings were mature (20–25 years old) and primarily consisted of the cultivars ‘Bluecrop’ and ‘Blueray’. Plant material was assayed by DAS-ELISA using commercial antibodies (Agdia, Elkhart, IN, USA). BISCv was detected in 59 out of 66 (89 %) highbush blueberry samples from Steuben County, and in 11 out of 31 (35 %) highbush blueberry samples from Wayne County. BISCv-infected bushes from Steuben County exhibited typical symptoms, such as blossom blight and twig dieback, whereas those from Wayne County were symptomless. To confirm the presence of the virus, total RNA was extracted from

blueberry leaf and/or blossom samples that reacted positively to BISCv in ELISA, using an E.Z.N.A.® Plant RNA Kit (Omega Bio-Tek Inc., USA). Specific primers (Table 1) were designed based on the three complete BISCv genome sequences available in the GenBank database (isolate NJ, accession no. NC_003499; isolate BC-1, accession no. AY941198, and isolate BC-2, accession no. AY941199) and used in immunocapture (IC) reverse transcription (RT)-polymerase chain reaction (PCR) [19] or RT-PCR using a QIAGEN One-Step RT-PCR Kit (Hilden, Germany). The genetic variability of 19 randomly selected BISCv isolates, 12 from Steuben County and seven from Wayne County, was investigated within five coding regions (Supplementary Table 1).

BLAST analysis of a 434-bp fragment from the 5' end of the RdRp gene of 19 BISCv isolates from New York revealed sequence identities of 97.9–100 % and 94.4–100 % at the nt and amino acid (aa) level, respectively. The sequences of the New York isolates had 86.4–100 % nt sequence identity and 90.3–100 % aa sequence identity to isolates BC-1, BC-2, and NJ. The RdRp fragment analyzed in this study codes for an RNA methyltransferase (MTR) domain, which contains conserved regions in aa position 66–80 (HSHPVCKTLENYILY) and 212–129 (VSSADKIR) [12]. The MTR region of the BISCv population had the lowest variability at the nt and aa level when compared to the other four ORFs characterized in this study (Table 2). Single simple repetitions (SSRs)

Table 1 Primers used for the amplification of the partial RNA-dependent RNA polymerase (RdRp) gene, full-length triple gene block (TGB) genes, and full-length coat protein (CP) gene of 19 blueberry scorch virus isolates from New York State by IC-RT-PCR or RT-PCR

Open reading frame	Primer name	Sequence 5'→3'	Position
RdRp	RDP_1 ^{ab}	ATGGCACTCACATACAGAAGTCC	59-81
	RDP_2 ^{ab}	TGCCTCTTCAATGACGATGTTC	488-510
CP	njCP1_F ^a	CTGAGTCCAGAGCATATCGTTGC	7092-7114
	njCP1_R ^a	GTA AACGGAGCTACTGACGCT	7679-7699
	njCPnj2_F ^a	ACCCTTATGGGCGGTTCTCAATTG	7506-7529
	njCPnj2_R ^a	TTTTTATTATACGATATACCCGGCAC	8349-8374
	bc2Cp1_F ^b	CTGAGTCCAGAGCATATCGTGGCA	7101-7124
	bc2Cp1_R ^b	CCGGCTTCCTTTTTCATGATTGC	7769-7791
	bc2Cp2_F ^b	ATGAGCTGTACAAGATGGATGTGCA	7539-7563
	bc2Cp2_R ^b	GGAAACACACGATAACAACGTTTC	8301-8323
TGB1	njTGB1_F ^a	AATTATGCCATTGAGGTGTCTTTTGCT	5792-5818
	njTGB1_R ^a	CACTGAAAGCTCAACGCACG	5597-6616
	bc2TGB1_F ^b	TCTTTTGCTTATCTCATGGGGGAGCG	5817-5842
	bc2TGB1_R ^b	TGCCTYGTCAAGCACTGAAAGCT	6612-6634
TGB2 and TGB3	njTGB2_F ^a	GGGGTTTGTCTTGCWGTCATTC	6486-6508
	njTGB2_R ^a	ATGTATAATCARACGCACACCTGT	6996-7019
	bc2TGB2_F ^b	GGAATTGCGTGGCCAGACTTTTG	6536-6558
	bc2TGB2_R ^b	GCTTCTTTCGGAGGCATAATCGCTT	7194-7218

^a Primers designed based on the BISCv-NJ sequence (NC_003499)

^b Primers designed based on the BISCv-BC-2 sequence (AY941199)

Table 2 Predicted evolutionary selection pressure exerted on five proteins encoded by blueberry scorch virus isolates^a

Open reading frame	Amino acid under negative selection ^b	dN/dS ^c
Partial RdRp	None	0.042
Full-length CP	None	0.042
Partial CP ^d	None	0.042
Full-length TGB1	26 , 50, 55, 63, 66, 79 , 81 , 86, 88, 94 , 96 , 110, 125, 146, 148, 149, 152, 160, 162, 165, 171, 180 , 181 , 187 , 190 , 192 , 196, 202, 209, 212	0.087
Full-length TGB2	20 , 39 , 57, 59, 60, 61, 64, 75 , 95, 98	0.092
Full-length TGB3	31 , 32 , 33 , 35	0.119

^a The number and ratio of non-synonymous (dN) and synonymous (dS) substitutions were estimated using the Datamonkey server [12] using random effects likelihood (REL), single likelihood ancestor counting (SLAC), and fixed effects likelihood (FEL) models. Only codons with evidence of selection by all three models were taken into account. No amino acid was found under positive selection

^b Amino acids located within conserved motifs are in bold

^c Ratio of nonsynonymous (dN) to synonymous (dS) substitutions

^d All CP gene sequences available in GenBank

have been described in ORF1 of carlaviruses, including B1ScV isolate BC-2, at nt positions 348-353 and 380-385 [1]. The presence of a one-nucleotide (T) SSR repetition was found in 15 (RC-18, RC-28, RC-10, RC-6, RC-5, RC-4, H-22-20, H-12-60, H-7-50, H-2-40, F-27-30, F-17-30, RC-22, B-10-60, and A-1-40) out of the 22 B1ScV isolates analyzed, including the 19 New York isolates and isolates BC-1, BC-2, and NJ. This repetition was extended to eight nucleotides in the sequence of five B1ScV isolates (A-5-50, B-1-30, F-12-40, BC-1, and NJ). A second SSR consisting of six T's was present in the sequence of four isolates (A-5-50, F-12-40, B-1-30, and BC-1). No recombination events were identified in the 5'-terminal ORF1 sequence using the RDP4 Beta 4.39 program [14], and the selective pressure exerted on the corresponding aa was neutral, as shown by estimates of the ratio of non-synonymous (dN) and synonymous (dS) substitutions using the Datamonkey server [13] (Table 2).

For the full-length TGB1 gene of 19 B1ScV isolates from New York, the sequence identity was 82.9-100 % and 90.4-100 % at the nt and aa level, respectively. When this coding region was compared with the corresponding sequence of isolates BC-1, BC-2, and NJ, identities were 73.7-100 % at the nt level and 84.6-100 % at the aa level. TGB1 aa alignments revealed five motifs characteristic of viral helicases [11]. Selective pressure analysis revealed negative selection on V/I26 in the first motif ([V/I]HSVPGAGKSTL[L/I]) in aa position 26-38. The second conserved motif (TLIDEY) at aa position 79-84 was surrounded by two (I and L) instead of four canonical hydrophobic residues [11], and negative selection was found to act on aa I81 and T79. A G/A substitution in the third conserved motif (FAVFADPLQ) was found at aa position 94-102, confirming previous observations for Kalanchoë latent virus (KLV), another carlavirus [18] with three (one V and two F) instead of four canonical hydrophobic

residues preceding the motif [11]. Purifying selection is acting on hydrophobic residues F94 and V96. A T/C substitution was found at aa position 178 of the fourth conserved motif (CIKE[V/L]RGQTF[D/E][S/C]VTFV) at a position 178-193. This substitution was also reported for several other carlaviruses, as was a Q/R substitution at position 183 [10, 15, 18, 20, 24, 30]. Negative selective pressure is acting on K180, E181, F187, V190, and F192 (Table 2). A V/Q substitution at position 208 of the fifth conserved motif (FQC[L/I]TR) at aa position 207-212 has been documented for other carlaviruses [10, 15, 18, 20, 24, 26, 30]. An A/C substitution at position 209 of the B1ScV motif was observed, but this replacement is usually neutral for protein function [3]. It is noteworthy that purifying selection is exerted on C209 but no SSR tract was identified, contradicting previous observations [1].

For the full-length TGB2 genes of 19 B1ScV isolates from New York, the sequence identity was 85-100 % and 88.6-100 % at the nt and aa level, respectively. Sequence identity of 78.2-100 % at the nt level and of 85.8-100 % at the aa level was found between New York isolates and isolates BC-1, BC-2 and NJ. The B1ScV TGB2 protein contains two hydrophobic transmembrane domains at aa positions 9-30 and 75-93 that are separated by a conserved central motif, **GD**_{x6}**GG**_x**Y****DG** [17], at aa position 38-53. Purifying selection is acting on I20 and D39, which are near the N-terminus of TGB2, and on G75, which is localized near the C-terminus of TGB2 (Table 2).

For the full-length TGB3 gene of 19 B1ScV isolates from New York, the sequence identity was 90.7-100 % and 95.3-100 % at the nt and aa level, respectively. This ORF has the second lowest genetic diversity observed within the B1ScV genome (Table 2). When the TGB3 sequence of New York isolates was compared with the corresponding sequences of isolates BC-1, BC-2 and NJ, identity values ranged between 83 and 100 % at both the nt and aa level. A

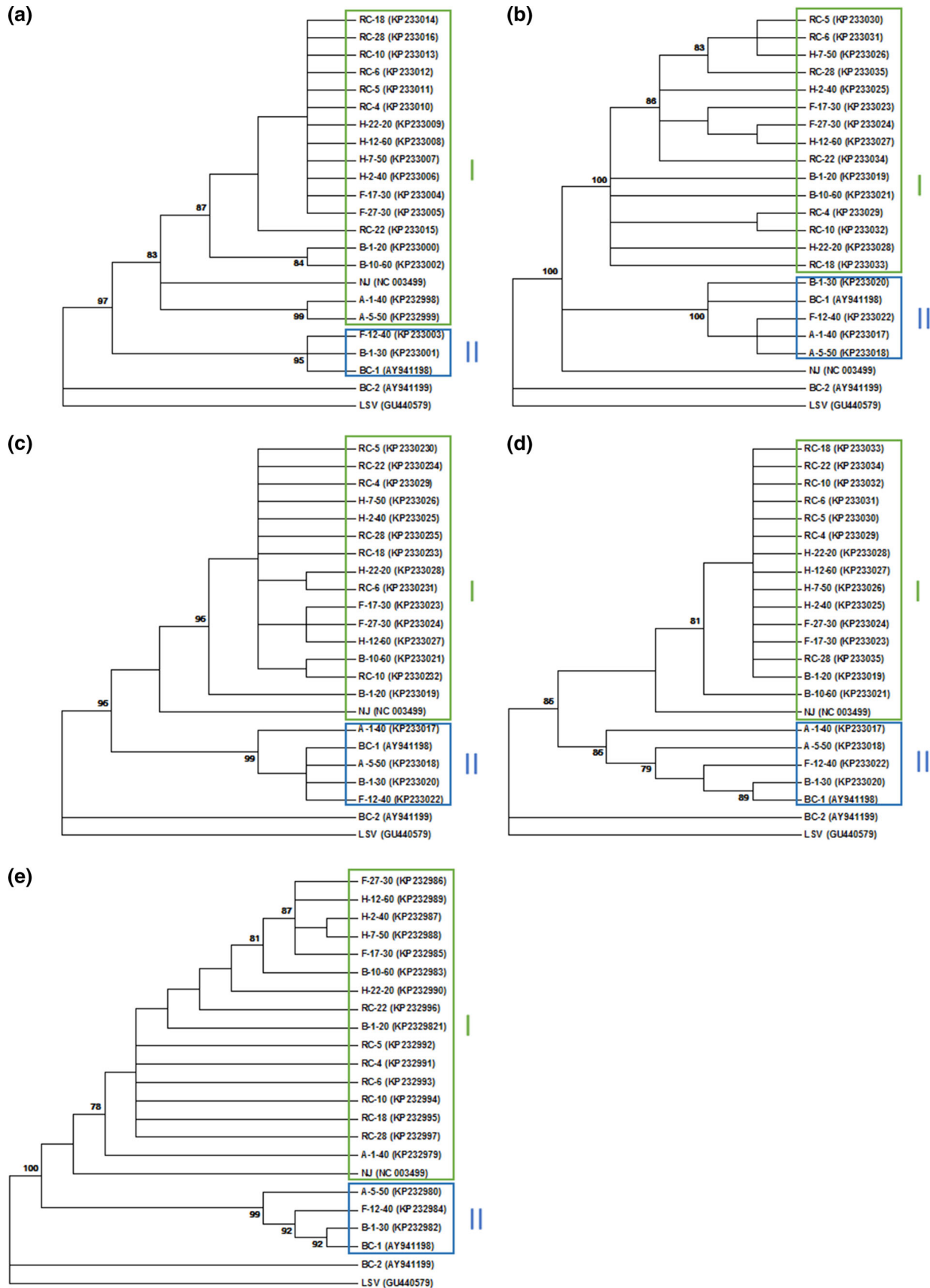


Fig. 1 Phylogenetic tree of a) a 434-nt-long RNA-dependent RNA polymerase gene fragment, b) the 690-nt full-length triple gene block 1 coding region, c) the 321-nt full-length triple gene block 2 coding region, d) the 195-nt full-length triple gene block 3 coding region, and e) the 939-nt full-length coat protein (CP) gene of 19 blueberry scorch virus isolates from New York State and isolates BC-1, BC-2, and NJ. GenBank accession numbers are in brackets. Bootstrap values are given next to branches. The sequence of lily symptomless virus (LSV) was used as an outgroup

transmembrane domain is present in aa position 4–22, followed by the carlavirus conserved motif (CX₅GX₈C) at aa position 30–44 [17]. Negative selection pressure was found to act on residues V31, I32, V33, and T35 (Table 2).

For the full-length CP gene of 19 B1ScV isolates from New York, the sequence identity was 82.3–100 % and 84–100 % at the nt and aa level, respectively. When this coding region was compared with the full-length CP gene sequence of isolates BC-1, BC-2, and NJ, the values were 78.5–100 % at the nt level and 84.6–100 % at the aa level. By analogy to the CPs of viruses containing potex-like TGBs [6], the B1ScV CP may be necessary for intercellular transport of the viral genome, and this function could explain the high frequency of synonymous changes in this part of the genome (Table 2). The three-nucleotide (CAC) SSR tract described in isolate BC-2 [1] was not confirmed in this study. Selection pressure analysis of the full-length CP gene of 22 B1ScV isolates revealed neutral selection (Table 2). Analysis of CP gene sequences using the RDP4 Beta 4.39 program [13] indicated a recombination event between nt 454–515 (beginning breakpoint) and 889 (ending breakpoint) with strong statistical support for isolates A-5-50 (p -value = 1.592×10^{-17}) and B-1-20 (p -value = 3.03×10^{-22}). For recombinant A-5-50, the putative major parent was isolate F-12-40 and the putative minor parent was isolate F-27-30. For recombinant B-1-20, the putative major and minor parents were isolates F-27-30 and B-1-30, respectively. Another recombination event between nt 40 and 293–371 was confirmed with strong statistical support (p -value = 1.126×10^{-12}) for isolate A-1-40. The putative major parent of recombinant A-1-40 was isolate RC-10, and the putative minor parent is unknown. The fact that a recombination event was suspected in only three out of 22 B1ScV isolates analyzed in this study is consistent with the rare occurrence of genetic exchange within the CP gene of carlaviruses [9, 22, 25, 29].

Phylogenetic relationships within the five ORFs characterized in this study using MEGA6 [27] revealed a grouping of B1ScV isolates into two distinct clades, regardless of the ORF analyzed (Fig. 1). Fifteen B1ScV isolates from New York clustered with isolate NJ in the predominant clade I, and four isolates from New York State, including the three recombinants, clustered with isolate BC-1 in the minor clade II. Isolate BC-2 formed a

separate branch by itself in the five phylogenetic trees. Nucleotide sequence divergence within (0.029 ± 0.007) and between (0.183 ± 0.032) phylogroups was low. In addition, there was no apparent association between sequence variability and the geographic origin of B1ScV isolates. Similarly, the phylogenetic trees did not group apart New York isolates from symptomatic and asymptomatic highbush blueberry.

In conclusion, while there is an abundance of information on the genetic variability within a portion of the CP gene [15], there is relatively little sequence information from other parts of the genome of B1ScV isolates. Five coding regions of 19 isolates from New York State were characterized by sequencing, showing that purifying selection and recombination are important evolutionary mechanisms in the genetic diversification of B1ScV. Although the five B1ScV coding regions are apparently under negative selection, as shown for other carlaviruses [9, 22, 29], codons of the TGB1 and TGB2 genes are under the strongest purifying selection; a moderate purifying selection is exerted on codons of the TGB3 coding region, and neutral selection is exerted on the RdRp and CP genes. TGB translation products are involved in plant cell-to-cell and long-distance movement [2]; therefore genetic conservation allows maintenance of this critical function. Recombination has been described in the CP genes of other carlaviruses [9, 22, 25, 29], but this is the first report of recombination in B1ScV. Together, our findings shed light on the relationship between B1ScV isolates from New York State and other regions, and on the evolution of B1ScV populations.

Acknowledgments We are grateful to the growers who provided access to their blueberry operations and acknowledge the support from USDA-NIFA through the New York State Agriculture Experiment Station and the Polish National Science Center (grant no. 2012/07/N/NZ9/01797). Thanks to Serena Minutillo and David MacUmbert for assistance with sample collection. Elzbieta Kalinowska was supported by a fellowship from Warsaw University of Life Sciences for a three-month visit to Cornell University.

References

1. Alam MC, Singh KA, Sharfuddin C, Ali S (2014) Genome-wide scan for analysis of simple and imperfect microsatellites in diverse carlaviruses. *Infect Genet Evol* 21:287–294
2. Adams MJ, Candresse T, Hammond J, Kreuze JF, Martelli GP, Namba S, Pearson MN, Ryu KH, Saldarelli P, Yoshikawa N (2012) Family Betaflexiviridae. In: King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ (eds) *Virus Taxonomy: ninth report of the international committee of taxonomy of viruses*. Elsevier Academic Press, San Diego, p 926
3. Betts MJ, Russell RB (2003) Amino acid properties and consequences of substitutions. In: Barnes MR, Gray IC (eds) *Bioinformatics for Geneticists*. Wiley, New York, pp 289–314
4. Bristow PR, Martin RR, Windom GE (2000) Transmission, orchard spread, cultivar response, and impact on yield in highbush

- blueberry infected with Blueberry scorch virus. *Phytopathology* 90:474–479
5. Cavileer TD, Halpern BT, Lawrence DM, Podleckis EV, Martin RR, Hillman BI (1994) Nucleotide-sequence of the carlavirus associated with blueberry scorch and similar diseases. *J Gen Virol* 75:711–720
 6. Chapman S, Hills G, Watts J, Baulcombe D (1992) Mutational analysis of the coat protein gene of potato virus X: effects on virion morphology and viral pathogenicity. *Virology* 191:223–230
 7. Cox K, Cox R, Fuchs M (2012) Emerging blueberry virus concerns for NY: result from blueberry virus survey efforts. *N Y Fruit Q* 20:9–11
 8. DeMarsay A, Hillman BL, Petersen FP, Oudemans PV, Schloemann S (2004) First report of Blueberry scorch virus on highbush blueberry in Connecticut and Massachusetts. *Plant Dis* 88:572
 9. Ge B, He Z, Zhang Z, Wang H, Li S (2014) Genetic variation in potato virus M isolates infecting pepino (*Solanum muricatum*) in China. *Arch Virol* 159:3197–3210
 10. Hatlestad GJ, Elam L, Gonzalez A, Lloyd AM (2011) Mirabilis jalapa mottle virus: a new carlavirus infecting four o'clocks. *Arch Virol* 156:2109–2111
 11. Kadaré G, Haenni AL (1997) Virus-encoded RNA helicases. *J Virol* 71:2583–2590
 12. Koonin EV, Dolja VV (1993) Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences. *Crit Rev Biochem Mol* 28:375–430
 13. Kosakovsky Pond SL, Frost SDW (2005) Datamonkey: rapid detection of selective pressures on individual sites of codon alignments. *Bioinformatics* 21:2531–2533
 14. Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefevre P (2010) RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* 26:2462–2463
 15. Matoušek J, Schubert J, Ptáček J, Kozlová P, Dědic P (2005) Complete nucleotide sequence and molecular probing of potato virus S genome. *Acta Virol* 49:195–205
 16. Moretti M, Ciuffo M, Gotta P, Prodorutti D, Bragagna P, Turina M (2011) Molecular characterization of two distinct strains of Blueberry scorch virus (BIScV) in Northern Italy. *Arch Virol* 7:1295–1297
 17. Morozov SY, Solovyev A (2003) Triple gene block: modular design of a multifunctional machine for plant virus movement. *J Gen Virol* 84:1351–1366
 18. Nicolaisen M, Nielsen SL (2001) Analysis of the triple gene block and coat protein sequences of two strains of Kalanchoë latent carlavirus. *Virus Genes* 22:265–270
 19. Nicolasco G, de Blas C, Torres V, Ponz F (1993) A method combining immunocapture and PCR amplification in microliter plate for the routine diagnosis of plant viruses and subviral pathogens. *J Virol Methods* 45:201–218
 20. Ochoa-Corona FM, Tang J, Lebas BS, Rubio L, Gera A (2010) Diagnosis of broad bean wilt virus 1 and Verbena latent virus in *Tropaeolum majus* in New Zealand. *Australasi Plant Path* 39:120–124
 21. Paduch-Cichal E, Kalinowska E, Chodorska M, Sala-Rejczak K, Nowak B (2011) Detection and identification of viruses of highbush blueberry and cranberry using serological ELISA test and PCR technique. *Acta Sci Pol Hortorum Cultus* 10:201–215
 22. Pramesh D, Baranwal VK (2013) Molecular characterization of coat protein gene of Garlic common latent virus isolates from India: an evidence for distinct phylogeny and recombination. *Virus Genes* 47:189–193
 23. Schilder A (2010) Shock and scorch survey update. *Michi Blueberry Newsl* 4:2
 24. Scott SW, Zimmerman MT (2008) The complete sequence of ligustrum necrotic ringspot virus, a novel carlavirus. *Arch Virol* 153:393–396
 25. Singh AK, Mahinghara BK, Hallan V, Ram R, Zaidi AA (2008) Recombination and phylogeographical analysis of Lily symptomless virus. *Virus Genes* 36:421–427
 26. Tang J, Harper SJ, Wei T, Clover GR (2010) Characterization of hydrangea chlorotic mottle virus, a new member of the genus Carlavirus. *Arch Virol* 155:7–12
 27. Tamura K, Stecher G, Peterson D, Filipiński A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30:2725–2729
 28. Wegener LA, Martin RR, Bernardy MG, MacDonald L, Punja ZK (2006) Epidemiology and identification of strains of Blueberry scorch virus on highbush blueberry in British Columbia, Canada. *Can J Plant Pathol* 28:250–262
 29. Zanardo LG, Silva FN, Lima AT, Milanese DF, Castilho-Urquiza GP, Almeida AM, Zerbini FM, Carvalho CM (2014) Molecular variability of cowpea mild mottle virus infecting soybean in Brazil. *Arch Virol* 159:727–737
 30. Zheng H-Y, Chen J, Zhao M-F, Lin L, Chen J-P, Antoniów JF, Adams MJ (2003) Occurrence and sequences of Lily mottle virus and Lily symptomless virus in plants grown from imported bulbs in Zhejiang province, China. *Arch Virol* 148:2419–2428