Nucleotide sequence of the carlavirus associated with blueberry scorch and similar diseases

Timothy D. Cavileer,1† Barbara T. Halpern,1 Diane M. Lawrence,1 Edward V. Podleckis,2 Robert R. Martin and Bradley I. Hillman1*

1 Department of Plant Pathology, Rutgers University, New Brunswick, New Jersey 08903, U.S.A. and 2 Agriculture Canada, 6660 NW Marine Drive, Vancouver, Canada V6T 1X2

We have synthesized and mapped a library of cDNA clones representing the RNA genome of a strain of blueberry scorch carlavirus (BBScV) associated with a disease known locally, in New Jersey, U.S.A., as Sheep Pen Hill disease. The nucleotide sequence of that strain was determined to be 8514 residues, excluding the poly(A) tail. In addition, cDNA clones representing the 3' terminus of another strain of the virus from the same field were synthesized, mapped and sequenced. The overall identity between sequences of these two strains was approximately 90% spanning the 1634 residue overlap, confirming their identity as distinct strains and not simply different isolates of a single strain. Finally, the coat protein gene of a distinct strain of the virus, isolated from plants with blueberry scorch disease in the Puyallup Valley in Washington State, U.S.A., was cloned from total cDNA by PCR. Sequence analysis revealed that the strain from Washington was more divergent from the two New Jersey strains than they were from each other. Comparisons of these sequences with other carlavirus sequences indicated that BBScV is more closely related to lily symptomless virus and potato virus S than to potato virus M, Helinium virus S, carnation latent virus or poplar mosaic virus. BBScV and potato virus M shared approximately 54% nucleotide sequence identity overall.

Introduction

Two similar diseases of significant economic importance have been described in blueberry-growing areas of the east and west coasts of North America. The first, known locally in New Jersey as Sheep Pen Hill disease, was first noticed in the late 1970s (Stretch, 1983). The second, named blueberry scorch disease, was first observed in 1980 near Puyallup, Wash., U.S.A. Although both are blighting diseases, specific symptoms vary in some details according to geographical location. In particular a line pattern, which may in its most dramatic form be a full oak-leaf pattern, is often observed in association with Sheep Pen Hill disease shortly before leaf drop in autumn, but this symptom has rarely been noted on the west coast. These diseases are distinct from blueberry necrotic shock disease, which is caused by an ilarivirus (MacDonald et al., 1991). Several cultivar differences have been described for the two diseases (Stretch, 1983; Martin et al., 1992), and the virus associated with blueberry scorch disease has been transmitted to blueberry plants by aphids (R. R. Martin, unpublished). The virus associated with Sheep Pen Hill disease has been transmitted to Chenopodium quinoa Willd. and C. amaranticolor Coste & Reyn., but despite repeated efforts the causal agent of blueberry scorch disease has yet to be transmitted to herbaceous hosts.

Both Sheep Pen Hill disease and blueberry scorch disease have been associated with carlaviruses (Podleckis et al., 1986; Podleckis & Davis, 1989; Martin & Bristow, 1987), but only the latter has been characterized to a significant extent (Martin & Bristow, 1988). Although polyclonal antisera have been raised against virus particles associated with both diseases, attempts to determine the degree of their serological relatedness led to somewhat ambiguous results (Martin et al., 1992). Those studies, however, provided evidence that the viruses were more closely related to each other than to the other carlaviruses tested. Until recently, little was known about the organization and expression of carlavirus genomes. Prior to this study, potato virus M (PVM) was the only carlavirus for which the complete nucleotide sequence had been determined.
(Rupasov et al., 1989; Zavriev et al., 1991). Partial sequences of several other carlaviruses have been published, including carnation latent virus (CLV; Meehan & Mills, 1991; Haylor et al., 1990). Helinemium virus S (HeVS; Foster et al., 1990c), lily symptomless virus (LSV; Memelink et al., 1990), potato virus S, ordinary and Andean strains (PVS-O; Foster & Mills, 1992b and PVS-An; MacKenzie et al., 1989; Foster et al., 1990a), chrysanthemum virus B (Levay & Zavriev, 1991) and poplar mosaic virus (PopMV; Henderson et al., 1992). Although similar in many respects to potexviruses, carlaviruses have several distinguishing features, including a larger first open reading frame (ORF1), a larger coat protein (CP) gene, and a gene encoding a cysteine-rich protein at the 3' terminus. In vitro translation studies have confirmed the differences suggested from the sequences.

We began examining the carlaviruses associated with diseases of blueberry that manifest scorch-like symptoms, including the North American east coast Sheep Pen Hill disease and the west coast blueberry scorch disease, to gain information and materials that would be useful in their detection and control. The studies here confirm that the two diseases are indeed caused by strains of a single virus, blueberry scorch virus (BBScV).

We describe the cloning and complete sequencing of one strain of BBScV, the cloning and sequencing of the 3' terminus of a closely related strain, and of the CP gene of a third, more distantly related strain. The sequence of BBScV indicates that it is organizationally similar to PVM, and evidence presented suggests that BBScV represents a biologically diverse virus species that is more closely related to PVS and LSV than to other carlaviruses.

Methods

Virus strains and isolation. Virus was isolated from infected blueberry flowers, from blueberry leaves, or from C. quinoa leaves essentially as described by Martin & Bristow (1988). Origins of the strains in this study are described in Table 1. RNA was isolated from purified virus as described by Hillman et al. (1985).

Table 1. Origins and nomenclature of blueberry scorch virus strains

<table>
<thead>
<tr>
<th>Strain</th>
<th>Origin</th>
<th>Method of cloning*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 BBScV-NJ1</td>
<td>Pemberton, N.J. 1988</td>
<td>Oligo(dT)</td>
</tr>
<tr>
<td>2 BBScV-NJ2</td>
<td>Pemberton, N.J. 1990</td>
<td>Oligo(dT), random</td>
</tr>
<tr>
<td>3 BBScV-WA1</td>
<td>Puyallup, Wash. 1990</td>
<td>Coat protein PCR</td>
</tr>
<tr>
<td>4 BBScV-WA2</td>
<td>Puyallup, Wash. 1990</td>
<td>Coat protein PCR</td>
</tr>
</tbody>
</table>

* Primers were oligo(dT)12-18 or random hexamers for strains 1 and 2, and cloning was by dC tailing and insertion to dG-tailed pUC9. Primers for strains 3 and 4 are described in Methods, and cloning of PCR products was by blunt-end ligation into the HindII site of pGEM-3ZF(+) .

Libraries of cDNA clones were synthesized essentially as described previously (Hillman et al., 1992), omitting the denaturation step. Random hexamer or oligo(dT)12-18 primers were used for these reactions, as listed in Table 1. In some cases, a size fractionation step using Size Select 400 Spin Columns (Pharmacia) was included before annealing the dC-tailed cDNA to dG-tailed pUC9. Colony selection, plasmid preparations, restriction analysis and Southern blots for mapping of clones were performed as described (Hillman et al., 1992).

DNA sequencing reactions were performed on denatured plasmid templates with [35S]dATP and Sequenase (U.S. Biochemicals) as described previously (Hillman et al., 1992). Dideoxynucleotide chain termination RNA sequencing reactions were performed on viral RNA essentially as described previously (Hillman et al., 1985). Nucleotide sequence analysis was performed with the Microgenie programs (Queue & Korn, 1984) and the Wisconsin Genetics Computer Group (GCG) programs (Devereux et al., 1984). Sequences were aligned with the aid of the Microgenie Align program, the CLUSTAL V alignment programs (Higgins & Sharp, 1989; Higgins et al., 1992), and by inspection. Cluster dendograms were produced and displayed using the PROTPARS and DRAWGRAM programs from the PHYLIP version 3.4 program package (Felsenstein, 1989).

Cloning and expression of the CP gene of strain BBScV-NJ1 fused to
Cloning and expression of the CP gene of strain BBScV-NJ1 fused to the carboxy terminus of the β-galactosidase gene were performed in the vector pUR290 (Rüthner & Müller-Hill, 1983). The CP gene was amplified by PCR from clone pBBScV-NJ1 with primers complementary to the termini of the coding region, containing flanking SalI and EcoRI sites at the 5' and SalI and XbaI sites at the 3' ends. The amplified CP gene was digested with SalI and inserted into pUR290 at the SalI site. Expression and analysis of fusion proteins were performed as described by Zudema et al. (1987).

Results

Although we successfully isolated BBScV from blueberry leaves and flowers, flower tissue was used when available because virus yield was found to be several times higher than in leaves. After C. quinoa and C. amaranticolor had been successfully infected with the BBScV strains associated with Sheep Pen Hill disease, they became the propagation hosts of choice. Yields of CsCl density gradient-purified virus from these herbaceous plants were typically 2 mg per 100 g tissue.

The first cDNA library was synthesized from RNA isolated from a virus preparation obtained from a blueberry field in Pemberton, N.J., U.S.A. (BBScV-NJ1). This library, initiated with oligo(dT) primers, contained many poly(A)+ clones that were proved to be coterminal by restriction and nucleotide sequence analyses. Since most of these clones were shorter than 2 kbp in length, a second library, primed with random hexamers, was synthesized from RNA isolated from a virus preparation obtained from the same field in Pemberton, N.J. 2 years later (BBScV-NJ2). A number of clones from this second library were substantially larger, many over 3 kbp in length. Mapping of the second library by restriction and Southern blot analyses indicated that the entire genome was represented many times, with the exception of the 5' terminus. The proximity of clones to the 5' terminus of the viral genome was determined by comparisons of their terminal nucleotide sequences with sequences at the 5' termini of PVM (Zavriev et al., 1991) and PVS-An (Monis & de Zoeten, 1990). An oligonucleotide complementary to the sequence 86 to 102 residues from the 5' end was then used to prime the RNA sequencing reactions in order to determine the sequence at the 5' end (Fig. 1), and the 5'-terminal sequence of BBScV-NJ2 was found to be similar to the PVM and PVS-An sequences. The 5' end appeared to be capped, based on the finding of two bands at the end of the template after shorter exposures of the autoradiogram shown in Fig. 1. Nucleotide sequence analysis of clones representing the 3' termini of the two BBScV isolates from the same field revealed that two distinct strains were represented. The entire nucleotide sequences of the second strain isolated (BBScV-NJ2) and the 3'-terminal region of BBScV-NJ1 are provided in Fig. 2.

Fig. 3 provides a map of open reading frames deduced from the sequence of BBScV-NJ2 and a matrix comparison of its nucleotide sequence with the PVM sequence. Overall identity between these two sequences is approximately 54%, and their organizations are similar. The 5' leader of BBScV-NJ2 has a length of 58 residues, slightly shorter than the 75-residue leader of PVM. BBScV-NJ2 ORF 1 has the potential to encode a protein of 223K, approximately the same size as the PVM ORF 1 product. Direct comparison of the amino acid sequences deduced from BBScV and PVM ORF1 regions is presented in Fig. 4. Not surprisingly, the three regions of highest conservation are those for which putative functions have been assigned: the presumed methyltransferase at the N terminus, the first of two presumed NTP-binding/helicase domains, and the core polymerase domain near the C terminus.

Following ORF1 is the first of two intergenic regions. A triple block of overlapping coding regions reminiscent of those in other carlaviruses, potexviruses, hordeiviruses and furoviruses has the coding capacity for a helicase homologue (ORF2, p25) and two basic proteins (ORF3, p12 and ORF4, p7). Alignments of the deduced amino acid sequences of these three proteins with their homologues from several other carlaviruses are provided in Fig. 5. The CP gene (ORF5, p33) follows the second intergenic region. We demonstrated that this gene encoded the CP by expression as a β-galactosidase fusion in Escherichia coli and subsequent immunoblot analysis using antiserum generated against the native CP (results not shown). The 3'-terminal gene, with the potential to encode a cysteine-rich protein (p16), substantially overlaps the CP gene.

Since previous serological tests to determine the relatedness of BBScV strains from the east and west coasts of the U.S.A. had been equivocal, we proceeded to analyse virus isolated from a field in Puyallup, Wash., U.S.A. in more detail (Table 1). Total cDNA was synthesized from randomly primed RNA samples, amplified with primers specific to the CP gene of BBScV-NJ1, and the amplification products were cloned into the plasmid vector pGEM-3Zf(+) (1987). Three clones were sequenced, one in its entirety (BBScV-WA2) and two partially. The two clones that were partially sequenced, represented by BBScV-WA1, were identical to one another and differed by only one residue from the sequence of BBScV-NJ1 over the 350 residues sequenced. The different residue, a C to T transition, was observed in both BBScV-WA1 clones. The sequence of the third clone, designated BBScV-WA2 in Fig. 2, was distinct from any other BBScV clone sequenced. Its overall nucleotide sequence identity with the other three isolates represented in Table 1 was only approximately 70%. There was no evidence for progressive evolution of the
Fig. 2. For legend see opposite.
We have provided evidence that diseases termed blue-}

isolates in Table 1. The most closely related strains were
BBsCV-NJ1 and -NJ2, which shared 90% amino acid
identity through the CP genes. BBsCV-WA2 shared 74% 
amino acid identity with BBsCV-NJ1 and 76% amino
acid identity with BBsCV-NJ2. Of 66 residues that were
shared by only two of the three different BBsCV CP
amino acid sequences, 56 were common to only strains
NJ1 and NJ2, three were common to strains NJ1 and
WA2, and seven were common to strains NJ2 and WA2.
Thus, strain NJ2 appears to be slightly more closely
related to strain WA2 than is strain NJ1.

Discussion
We have provided evidence that diseases termed blue-

berry scorch on the west coast and Sheep Pen Hill on the
east coast of the U.S.A. are associated with several
strains of BBsCV. The complete nucleotide sequence of
BBsCV indicates that it is similar in overall organization
to the only other carlavirus for which the complete
sequence has been determined, PVM, and similar to the
other carlaviruses for which partial sequences have been
published. The deduced product of the first ORF of
BBsCV, p223, shared 57% similarity and 44% identity
with its PVM homologue. Like PVM p223, BBsCV p223
contained the conserved domains typical of proteins
involved in RNA virus replication (Morozov et al.,
1990b) and is considerably larger than the ORF1
products of potexviruses, which typically have sizes of
approximately 160K.
Overall comparison of the 3' terminal sequences of BBScV revealed closer similarities with PVS and LSV than with other carlaviruses. To examine more closely the relationships among carlaviruses sequenced to date, alignments of the three 'triple block' proteins and cluster analysis of deduced amino acid sequences of the CP and cysteine-rich proteins were performed. As with other carlaviruses (MacKenzie et al., 1989; Rupasov et al., 1990; Memelink et al., 1990; Henderson et al., 1992), p25 of BBScV contains a putative NTP-binding motif, the highly conserved GKS core of which is shown boxed in Fig. 5. The potential translation products of the other two triple block genes, ORF3 (p12) and ORF4 (p7), appear to be hydrophobic, possibly membrane-associated proteins as described by Morozov et al. (1990a, 1991). Both p12 and p25 were highly conserved, having more than 42% identical residues in pairwise comparisons with their homologues in the other four viruses, with a maximum value of 64% for p12 of BBScV/PVS. BBScV p7 was considerably less conserved, with values ranging from 31 to 45% identity in pairwise comparisons. Similar to results reported by Morozov et al. (1991), however, there was considerable amino acid conservation between the two rigidly conserved cysteine residues toward the C terminus.

In CP sequence comparisons, the three BBScV strains, the two PVS strains and LSV formed a cluster that was distinct from the other carlaviruses. PVM and HelVS formed a separate cluster (Fig. 6). The most distant viruses were CLV and PopMV.

Cell-free translations of fractionated CLV (Tavantzis, 1991) or PVS-An (Foster & Mills, 1992a) RNAs have implicated the CP in processing of the ORF1 product in trans. The region of the CP that has been suggested as the possible catalytic site of the putative protease involved in this ORF1 protein processing contains the threonine-glycine–glycine triplet near the C terminus, with histidine and aspartic acid residues upstream. Like other carlavirus CPs sequenced to date, all three BBScV strains contain this putative catalytic cluster.

The 3' terminal ORFs of BBScV strains NJ1 and NJ2 contain four cysteine residues in a pattern typical of the conserved motif that characterizes the putative zinc-binding domain (Klug & Rhodes, 1987) found in other carlavirus RNAs (Gramstat et al., 1990; Haylor et al., 1990; Zavriev et al., 1991; Koonin et al., 1991; Henderson et al., 1992). Comparisons of the cysteine-rich proteins revealed affinities among the viruses examined that differed slightly from results of the CP comparisons (Fig. 7). Although the cysteine-rich protein...
Fig. 4. For legend see opposite.
of BBScV was most closely related to that of LSV, the cysteine-rich protein of Hel/NS was more closely related than PVS-An, PVM, CLV or PopMV. Pairwise comparisons of the BBScV and Hel/NS cysteine-rich proteins reveal particularly close similarities at the C termini, in which 29 of 33 residues are identical.

Details concerning the expression of the carnavirus

![Alignment and cluster dendrogram of amino acids deduced from carnavirus CP gene sequences. Alignments (a) were performed using the program CLUSTAL V, as described in the legend to Fig. 5. Trees (b) were generated with the Phylip version 3.5c programs PROTDIST and NEIGHBOR, with no outgroup selected. Branch lengths shown are proportional. M values are shown on the right. In addition to sequences listed in Fig. 5, Helicanium virus S (Hel/NS; Foster et al., 1990c) was included.](image-url)
Alignments
CLUSTAL V and Phylip version 3.5c, as described in the legends to Fig. 6. Sequences were from the same sources as in Fig. 6, but sequences were from Haylor et al., 1990). It should be noted that the low level of replication of BBScV renders it a poor candidate for studies of gene expression. Unlike luteoviruses, in which the AUG context of the second ORF of the large subgenomic RNA tends to be in a more favourable context than the first, and is thus translated efficiently despite its internal position (Dinesh-Kumar et al., 1992), the first AUG of the BBScV triple block is the most favourable for translation (Kozak, 1989). Indeed, evidence for low-concentration subgenomic RNA species representing internal triple block genes in other systems is accumulating (e.g. Gilmer et al., 1992). Among the other motifs that have been attributed to carlaviruses, the presence of sequences similar to chloroplast ribosome binding sites immediately upstream from the 25K protein gene, the first gene of the triple block, and from the CP gene (Foster & Mills, 1991). Examination of this region of the BBScV genome reveals similar motifs in the same area of the BBScV genome. As in other carlaviruses, the relevance of these motifs to BBScV gene expression is unknown.

BBScV is a heterogeneous collection of strains, as demonstrated by the identification of different strains from a single field in New Jersey and an even more distantly related strain from Washington. The 90% sequence identity (or 10% sequence divergence) between BBScV-NJ1 and -NJ2 distinguishes these viruses as distinct strains, and not merely isolates of a single strain. This heterogeneity has implications for BBScV detection by antibody or DNA-based methods, particularly using monoclonal antibodies or PCR. We currently have no information as to whether the strain differences evident from the sequences is reflected in different biological properties, such as transmission or symptomatology of BBScV.

The editorial comments of Dr P. H. Berger are gratefully acknowledged. This research was supported by USDA Specific Cooperative Agreement No. 58-3615-9-066 and New Jersey Agricultural Experiment Station Project No. 11192.

References


Foster, G. D. & Mills, P. R. (1992b). The 3'-nucleotide sequence of an ordinary strain of potato virus S. Virus Genes 6, 213–220.


Foster, G. D., Millar, A. W., Meehan, B. M. & Mills, P. R.

Fig. 7. Alignments and cluster dendrogram of amino acids deduced from carlaviruses 3'-terminal genes, encoding 'cysteine-rich' proteins. Alignments (a) and trees (b) were constructed using the program CLUSTAL V and Phylip version 3.5c, as described in the legends to Fig. 3 and 6. Sequences were from the same sources as in Fig. 6, but only one BBScV strain (BBScV-NJ2) was compared because of close similarity of BBScV-NJ1 and -NJ2 in these regions, and CLV sequences were from Haylor et al. (1990).


(Received 4 August 1993; Accepted 10 November 1993)