Complete nucleotide sequence of Kashmir bee virus and comparison with acute bee paralysis virus

J. R. de Miranda,1 M. Drebot,2 S. Tyler,2 M. Shen,1 C. E. Cameron,3 D. B. Stoltz4 and S. M. Camazine1

Correspondence
J. R. de Miranda
joachimdemiranda@yahoo.com

1Department of Entomology, Penn State University, University Park, PA 16803, USA
2National Microbiology Laboratory, Health Canada, Winnipeg, Manitoba, Canada R3E 3R2
3Department of Biochemistry, Penn State University, University Park, PA 16803, USA
4Department of Microbiology and Immunology, Dalhousie University, Halifax, Nova Scotia, Canada B3H 4R2

Received 21 January 2004
Accepted 26 April 2004

The complete nucleotide sequence of a novel virus is presented here together with serological evidence that it belongs to Kashmir bee virus (KBV). Analysis reveals that KBV is a cricket paralysis-like virus (family Dicistroviridae: genus Cripavirus), with a non-structural polyprotein open reading frame in the 5’ portion of the genome separated by an intergenic region from a structural polyprotein open reading frame in the 3’ part of the genome. The genome also has a polyadenylated tail at the 3’ terminus. KBV is one of several related viruses that also includes acute bee paralysis virus (ABPV). Although KBV and ABPV are about 70 % identical over the entire genome, there are considerable differences between them in significant areas of the genome, such as the 5’ non-translated region (42 % nucleotide identity), between the helicase and 3C-protease domains of the non-structural polyprotein (57 % amino acid identity) and in a 90 aa stretch of the structural polyprotein (33 % amino acid identity). Phylogenetic analyses show that KBV and ABPV isolates fall into clearly separated clades with moderate evolutionary distance between them. Whether these genomic and evolutionary differences are sufficient to classify KBV and ABPV as separate species remains to be determined.

INTRODUCTION

Kashmir bee virus (KBV) is a potentially lethal virus of honeybees that has recently come to prominence as one of several viruses closely associated with colony collapse because of infestation with varroa mites (Ball & Bailey, 1997). Like most honeybee viruses, KBV is thought to persist as an inapparent infection within the bee community, until stress or an alternative vector (such as varroa) causes it to become epidemic and lethal. The geographical and host origins of KBV are obscure. It was discovered in 1974 as a contaminant in preparations of Apis iridescent virus from the Asian hive bee (Apis cerana) that multiplied to high titres when injected or fed to adult Apis mellifera bees (Bailey & Woods, 1977). Although it was suspected that KBV originated in A. cerana and SE Asia, the detection of KBV, or its serological relatives, in natural populations of A. mellifera from around the world (Ball & Bailey, 1997; Allen & Ball, 1995), as well as A. cerana from India (Bailey & Woods, 1977; Bailey et al., 1979), bumblebees (Bombus spp.) from New Zealand and European wasps (Vespula germanica) from Australia (Anderson, 1991) has made this difficult to prove.

KBV is serologically and biologically closely related to acute bee paralysis virus (ABPV; Allen & Ball, 1995; Anderson, 1991). Like KBV it was discovered as a contaminant, during transmission studies of chronic bee paralysis virus (Bailey et al., 1963) and is extremely lethal to adults and larvae, both by injection and in larger doses by feeding (Bailey et al., 1963; Nordstrom, 2000). It is common in seemingly normal, healthy colonies and has been heavily implicated in varroa-induced colony losses, primarily in Europe in the 1980s (Ball, 1985; Allen et al., 1986; Ball & Allen, 1988; Bailey & Ball, 1991). Varroa can transfer ABPV among adults and pupae with 50–80 % efficiency, depending on the sensitivity of the detection method used (Wiegers, 1988; Ball, 1989). This efficiency drops with successive transfers and there is no noticeable latent period between acquisition and transmission, which suggests that there is no virus replication in the mite (Wiegers, 1988). The experimental host range for ABPV includes Apis and Bombus species, but not several non-hymenopteran insects (Bailey & Gibbs, 1964).
Although KBV and ABPV are closely related they are not identical. ABPV is often associated with paralysis in adult bees whereas KBV generally is not, the VP4 proteins of the ABPV and KBV particles are serologically distinct (Stoltz et al., 1995), and the viruses can be readily distinguished by RT-PCR (Stoltz et al., 1995; Evans, 2001). KBV is also more variable than ABPV, as determined by capsid protein profiles and serology (Bailey et al., 1979; Allen & Ball, 1995). This may be an innate property of the virus or the result of an adaptation process if A. mellifera was only recently acquired as a new host of KBV (Bailey & Ball, 1991). Both viruses can coexist within the same colony (Hung et al., 1996) and even within the same bee (Evans, 2001), although the implication of this for their classification is unclear. The ABPV genomic sequence has been determined (Govan et al., 2000), as well as partial sequences of a large number of European and American ABPV isolates (Bakonyi et al., 2002). Here, we present the complete genome sequence of KBV and compare this to that of ABPV.

**METHODS**

**Purification.** Purification of the Canadian KBV isolate was described by Stoltz et al. (1995). Infections of the Pennsylvania KBV isolate were established in 50 white-eyed honeybee pupae by injecting 4 μl of a crude extract of a single adult honeybee infected with KBV between the integuments of the abdomen, as determined by direct ELISA using KBV-specific antisera (Stoltz et al., 1995), and incubating the pupae at 32°C for 3–4 days (Anderson & Gibbs, 1988). Successfully infected pupae were identified by ELISA. Virus particles were purified from these by banding clarified extracts of infected pupae on 10–40% (w/v) sucrose gradients (Bailey & Ball, 1991).

**Cloning and sequencing.** Intact viral genomic RNA was isolated from the particles and used to prepare cDNAs, employing both random and sequence-specific primers. cDNA fragments were cloned either directly or after PCR amplification with specific primers into one of several plasmid cloning vectors (pBlueScript-II, Stratagene; pGEM-T Easy, Promega; pCR2.1, Invitrogen) and were subsequently sequenced. The 3′ terminus of the KBV genome was obtained by cloning oligo(dT) primed cDNA (cDNA kit; Clontech), since the virus group to which KBV belongs naturally has a polyadenylated tail at the 3′ end. The 5′ end of the KBV genome was obtained by three independent rounds of nested 5′ RACE, using the SMART-Oligo protocol (Clontech) and three different cDNA primers. A total of 15 5′ RACE clones from these reactions were sequenced.

**Protein expression.** Ten fragments of the KBV genome, covering the two major open reading frames (ORF), were cloned into bacterial expression plasmids (pQE30-32; Qiagen) behind an N-terminal 6× Histidine ‘tag’. The polypeptides thus produced were purified by their histidine tags on Ni-NTA columns. Production and purification of the expressed proteins was confirmed by ELISA, using a horseradish peroxidase (HRP) conjugated anti-(6×His) monoclonal antibody (A7058; Sigma).

**ELISA.** Direct ELISA was performed using 1–5 ng purified virus or expressed virus protein, 1:1000 dilution of primary antibody, 40 ng HRP conjugated protein-A ml⁻¹ (P8651; Sigma) as secondary antibody and 33/55′ tetramethyl benzidine as the colorimetric substrate (Harlow & Lane, 1988). The reaction was terminated with 3 M H₂SO₄ and read at 405 nm.

**Antisera.** Bee virus antisera used in the ELISA assays were obtained from various sources. The KBV-VP4 and ABPV-VP4 antisera were produced in rabbits using the SDS-PAGE purified VP4 proteins of KBV and ABPV as antigens and are described in Stoltz et al. (1995). The DWV-VP1, BQCV-VP3 and SPV-VP1/2 antisera were similarly produced using SDS-PAGE purified VP1 of deformed wing virus (DWV), VP3 of black queen cell virus (BQCV) and VP1 and VP2 of slow paralysis virus (SPV) as antigens. The DWV-VP1 antiserum is described by de Miranda et al. (2002). The KBV-virus, BQCV-virus, CBPV-virus and SBV-virus antisera are a gift of D. Anderson (CSIRO, Canberra, Australia) and were produced in rabbits against purified virus preparations of KBV, BQCV, chronic bee paralysis virus (CBPV) and sacbrood virus (SBV), respectively. The ABPV-IgG antiserum was a gift of I. Fries (Swedish Agricultural University, Uppsala, Sweden) and comprises the IgG fraction of an antisera prepared in rabbits against purified ABPV.

**Sequence analysis.** Each nucleotide position of the KBV sequence reported here is a consensus of at least three independent clones. The consensus sequences of the Canadian and Pennsylvania isolates of KBV have been deposited at GenBank under accession numbers AY452696 (KBV-ca) and AY275710 (KBV-pa).

**Phylogenetic analysis and taxonomy.** Two multiple sequence alignments, of 44 and 45 homologous, partial ABPV-KBV sequences, were used for phylogenetic analyses. The first alignment is located in the RNA polymerase region and covers nt ABPV(5306–5719)/KBV(5406–5819), while the second alignment, located in the coat protein region, covers nt ABPV(8121–8519)/KBV(8164–8562). The alignments were assembled using CLUSTAL W (Thompson et al., 1994) and the crick paralysis virus (CrPV) sequence (Wilson et al., 2000) as an outgroup. Phylogenetic inference and bootstrapping followed maximum-parsimony criteria and heuristic search methods as implemented by PAUP (4.0b10; Swofford, 1998). Poorly supported branches were collapsed.

**RESULTS**

KBV coexists with many other honeybee viruses of similar size and shape, both in natural populations and when propagating the virus in bee pupae (Anderson & Gibbs, 1988), and virus preparations are therefore rarely free of contaminating viruses. To prove that the nucleotide sequence presented here indeed belongs to KBV, segments of the ORFs were subcloned into bacterial expression vectors, behind a 6× His tag used for purifying the expressed proteins (Fig. 1). The expressed and 6× His tag purified proteins were then matched in ELISA with antisera specific for a range of similar shaped honeybee viruses. All expressed protein segments reacted positively with a commercial anti-(6×His) monoclonal antibody, proving that the proteins were expressed correctly and abundantly. None of the expressed proteins reacted to antisera against DWV, BQCV, CBPV, SBV or SPV and only the protein expressed from clone 8 reacted with antisera against KBV and ABPV. The observation that both these antisera reacted against this protein was not unexpected, since ABPV and KBV are known to be serologically related (Allen & Ball 1995). This suggests that the major antigenic epitopes of ABPV and KBV are located on clone 8, which corresponds to the N-terminal portion of the structural polyprotein of the virus (Fig. 1). To determine whether the virus sequence belonged to KBV or to a strain of ABPV, two more antisera were tested against the expressed proteins.
These are monospecific antisera raised against the SDS-PAGE purified VP4 of KBV (5-6 kDa) and ABPV (9-4 kDa), respectively (Stoltz et al., 1995), and are therefore free from contaminating antibodies against either bee proteins or other honeybee viruses, which can be present in antisera against whole virus particle preparations. Most importantly, these antisera are specific to their viruses in ELISA and Western blots and do not cross-react (Stoltz et al., 1995). The strong reaction of the polypeptide of clone 8 with the KBV-VP4 antiserum and the absence of such a reaction with the ABPV-VP4 antiserum, is evidence that the nucleotide sequences presented here indeed belong to KBV. These reactions also indicate that the VP4 of KBV is located on clone 8 of the structural polyprotein. The KBV-VP4 and ABPV-VP4 antisera also only react with their homologous purified virus preparations (strongly in the case of the KBV-VP4 antiserum, 10 × background; less so in the case of the ABPV-VP4 antiserum, 7 × background), confirming both their ability to distinguish between these viruses and the absence of significant amounts of ABPV in the KBV preparation and vice versa.

Analysis of the sequence reveals that KBV is a cricket paralysis-like virus (family Dicistroviridae: genus Cripavirus), whose genome consists of a single positive-strand RNA containing two ORFs, separated by an intergenic region and flanked by untranslated regions. KBV has roughly 70 % nt and aa identity to ABPV across the genome. The 5' non-translated region (5' NTR; nt 1–608) is the part of the genome that is least conserved between ABPV and KBV, with only 42 % nt identity. Although the KBV and ABPV 5' NTR are of similar length, there are many gaps in the alignment, and the KBV sequence extends about 116 nt beyond the 5' terminus of ABPV (Fig. 2a). The alignment of this region is anchored by a short, highly conserved stretch of 44 nt (shaded in Fig. 2a). The intergenic region (Fig. 2b; nt 6411–6764) and the 3' NTR (Fig. 2c; nt 9297–9506) are both highly conserved between ABPV and KBV at 79 and 76 % nt identity, respectively. The KBV intergenic region is considerably longer than that of ABPV (shaded area in Fig. 2b). However, it may be shorter than indicated, depending on whether KBV structural polyprotein is initiated at the first available AUG of the ORF (as shown in Fig. 2b) or at a prior non-AUG initiation site, as is the case for many other cripviruses (Johnson & Christian, 1998; Sasaki & Nakashima, 2000; Wilson et al., 2000; Czibener et al., 2000; van Munster et al., 2002; Nishiyama et al., 2003; Domier & McCoppin, 2003). The larger ORF is located in the 5' half of the genome and contains the three helicase domains, the 3C-protease domain and eight RNA polymerase domains (underlined in Fig. 2d) identified by Koonin & Dolja (1993). The helicase domains include the putative NTP-binding residues 566GxxGxGKS567 and 617D618 in domains A and B, respectively (Gorbalenya & Koonin, 1989). The 3C-protease domain conserves the cysteine protease motif 1335GxCG1338 and the putative substrate-binding residues 1354GxHxG1359. C1337 is the third residue of the protease catalytic triad that also involves a histidine residue and either an aspartate or glutamate residue (Koonin & Dolja, 1993), possibly H1245 and E1246, by analogy with ABPV (Govan et al., 2000). The RNA polymerase region contains the universal polymerase motif, 1783YGDD1786, in the pol-VI domain. The shorter ORF is located towards the 3' end of the genome and has two picornavirus capsid protein domains (underlined in Fig. 2e).

Within the recognized protein domains of both polyproteins there is a high degree of conservation between KBV and ABPV. Outside these domains there is far less conservation and there are several gaps throughout the amino acid alignments (Fig. 2d, e). Particularly divergent are the regions between the helicase and 3C-protease domains of the non-structural polyprotein (Fig. 2d; 57 % aa identity) and between aa 210 and 280 of the KBV structural polyprotein (Fig. 2e; 33 % aa identity). This latter region should be interesting in light of the distinct antigenic character of the VP4 proteins of ABPV and KBV.

There were 93 unique nucleotide variations among the KBV-pa sequence clones, appearing uniformly across the genome at a frequency of 1.7 × 10⁻³ per nucleotide sequenced. The vast majority (75 %) of these changes were transitions (C–U and A–G changes). High transition frequencies are common for RNA viruses and can be explained in part by the relative stability of the G–U pairing in RNA molecules during replication. A high percentage (50 %) of the nucleotide variations within the coding regions result in amino acid changes.

The results of the phylogenetic analyses of the relationships among a range of KBV and ABPV isolates are shown in Fig. 3, for both a section of the polymerase region and the coat protein region. These analyses are similar to those
of Evans (2001) and Bakonyi et al. (2002), but with the inclusion of more isolates. The isolates are classified by their original designation in GenBank (KBV or ABPV) and by their geographical origin, with the total number of isolates falling within each geographical group indicated in parentheses. The main observation is that the ABPV and KBV designated isolates are cleanly separated into their respective clades, with the exception of two solitary isolates from Hungary (one for the polymerase region and one for the coat protein region) that fall outside the 'centres' of the ABPV and KBV clades, as defined by the majority of the isolates. In the polymerase dataset, this isolate was classified as KBV (GenBank accession no. AF468967) while in the coat protein dataset the wayward isolate was classified as ABPV (AF346301).

Within the KBV and ABPV clades, a broad geographical distinction can be made between the isolates (America, Europe, etc.) with 85–90 % nt identity between isolates from different continents compared with 90–95 % nt identity between isolates from the same continent, but beyond that there is little geographical identity to the isolates.
isolates. For example, for the polymerase region there is a group of eight clearly demarked isolates from California within the KBV clade. However, in the 'USA' group there are also several isolates from California, as well as isolates from Maryland and Pennsylvania, and the 'Canada/USA' group contains series of independent isolates from Maryland, Maine and Canada that are all polytomous to each other and to the other two groups. This means that, despite their subsequent divergence, these (groups of) sequences were indistinguishable at one point in their evolution. Similar observations can be made when analysing the coat protein region of ABPV clade. Here, groups of isolates from UK, Hungary, Poland and E. Europe are polytomous at their base. The italicized numbers represent the bootstrap support for the terminal groups of isolates, which is generally strong enough to reflect genuine divergence from the other isolates.

**Fig. 2.** KBV-ABPV nucleotide and amino acid alignments of 5' NTR (a); intergenic region (b); 3' NTR (c); non-structural polyprotein (d); and structural polyprotein (e). Full stops (.) indicate identical nucleotide or amino acid residues, dashes (-) indicate a gap in the alignment. Upper-case letters in Fig. 2(a, b and c) indicate coding region. The shaded area in Fig. 2(a) marks a conserved area near the 5' end. The shaded area in Fig. 2(b) marks a section of KBV intergenic region that might be part of the structural polyprotein ORF. The shaded areas in Fig. 2(d) mark the conserved amino acids of the underlined domains (helicase domains A, B and C; the 3C-protease domain and RNA polymerase domains I–VIII) as identified by Koonin & Dolja (1993). The underlined areas in Fig. 2(e) mark two conserved picorna-like virus capsid protein domains and the shaded, lower-case letters indicate the potential amino acid residues prior to first methionine of the KBV structural polyprotein.

**DISCUSSION**

The nucleotide sequence presented here confirms that KBV and ABPV are closely related (Allen & Ball, 1995). Evidence that the nucleotide sequence indeed belongs to KBV has been presented, using the only diagnostic means available to distinguish between these viruses. Despite the close overall relationship between KBV and ABPV there are significant differences between the viruses in critical areas of the genome. The 5' NTR in particular is highly divergent. In positive-strand RNA viruses this region contains primary and secondary RNA structures that are critical for the binding of the viral RNA polymerase for genome replication (Gromeier et al., 1999). The differences between KBV and ABPV in their 5' NTRs may therefore reflect unique binding requirements of their respective RNA polymerases, and hence a means for the rival polymerases to distinguish
between homologous and non-homologous viral RNAs during mixed infections (Evans, 2001).

The comparison of closely related viruses such as KBV and ABPV is particularly useful for identifying regions of greater or lesser variability within the genome. For example, although there is great overall conservation within the RNA polymerase region there is a short section between pol-V and pol-VI where variation is either allowed to exist or perhaps even selected for (Fig. 2d). The helicase domains, embedded in a very conserved stretch of sequence, are followed by a highly variable protein sequence and there is also a highly variable region following the second capsid protein domain of the structural polyprotein. The biological functions of these variable proteins is of great interest since they may indicate unique molecular and biological properties of the viruses.

The phylogenetic analyses revealed that KBV and ABPV are distinct viruses and are found in several continents (Allen & Ball, 1996). Both KBV and ABPV isolates have only a limited geographical identity. Isolates of each virus can be broadly separated by their continent of origin, but it is more difficult to identify regional trends within each continent. The most obvious explanation for this is that the KBV and ABPV isolates lose their regional identity through the long distance transport of live bees (packages and migratory beekeeping) within each continent. An alternative explanation is that these virus isolates are united by biological rather than geographical criteria, the most obvious of which is the genetic background of the bees. Despite the considerable international trade in queen bees, regional preferences for certain strains of bees persist, which could account for the continental distinctions between the virus isolates. Both geography and host genetics were used to explain the relationships between sacbrood virus isolates collected from around the world, and from both A. mellifera and A. cerana (Grabenstein et al., 2001).

What remains to be determined is whether ABPV, KBV and other variants within this group are autonomous viruses meriting species status or strains of single species (van Regenmortel et al., 2000). Since there is no clear geographical, temporal or ecological separation between KBV and ABPV, species designation will depend heavily on the unique biological and molecular characteristics of
these viruses, such as the specificity of virus replication and encapsidation, to support their clear phylogenetic separation.

ACKNOWLEDGEMENTS

Many thanks to Dennis Anderson for the BQCV, SBV, KBV and CBPV antisera, and to Ingemar Fries for the ABPV-IgG antiserum. This work was supported by grants from the Pennsylvania Department of Agriculture (ME 400492 and ME41833) and the Bintational Agricultural Research and Development fund (US-3205-01).

REFERENCES


