Analysis and phylogenetic comparisons of full-length VP2 genes of the 24 bluetongue virus serotypes


The outer capsid protein VP2 of Bluetongue virus (BTV) is a target for the protective immune response generated by the mammalian host. VP2 contains the majority of epitopes that are recognized by neutralizing antibodies and is therefore also the primary determinant of BTV serotype. Full-length cDNA copies of genome segment 2 (Seg-2, which encodes VP2) from the reference strains of each of the 24 BTV serotypes were synthesized, cloned and sequenced. This represents the first complete set of full-length BTV VP2 genes (from the 24 serotypes) that has been analysed. Each Seg-2 has a single open reading frame, with short inverted repeats adjacent to conserved terminal hexanucleotide sequences. These data demonstrated overall inter-serotype variations in Seg-2 of 29 % (BTV-8 and BTV-18) to 59 % (BTV-16 and BTV-22), while the deduced amino acid sequence of VP2 varied from 22.4 % (BTV-4 and BTV-20) to 73 % (BTV-6 and BTV-22). Ten distinct Seg-2 lineages (nucleotypes) were detected, with greatest sequence similarities between those serotypes that had previously been reported as serologically ‘related’. Fewer similarities were observed between different serotypes in regions of VP2 that have been reported as antigenically important, suggesting that they may play a role in the neutralizing antibody response. The data presented form an initial basis for BTV serotype identification by sequence analyses and comparison of Seg-2, and for development of molecular diagnostic assays for individual BTV serotypes (by RT-PCR).

INTRODUCTION

Bluetongue virus (BTV) has a genome composed of ten linear segments of double-stranded RNA (dsRNA) and is classified as the type species of the genus Orbivirus within the family Reoviridae (Mertens et al., 2004). BTV is transmitted between its mammalian hosts by certain species of biting midges (Culicoides spp. – Diptera: Ceratopogonidae) and can infect all ruminant species, although it usually causes severe disease (bluetongue – BT) only in sheep and some species of deer. BTV occurs between latitudes 40°S and 53°N in the Americas, Africa, Asia and Australia, causing worldwide losses that were estimated in 1996 at US$3 billion per year (Tabachnick, 1996). BT was therefore given list ‘A’ status by the Office International des Epizooties (OIE) (Alexander et al., 1996; OIE, 2000) and represents a major barrier to international trade in animals and some animal products.

Historically there have been occasional but relatively short-lived incursions of BT into southern Europe, involving a single virus serotype on each occasion, although their impact on the susceptible and immunologically naïve local populations of sheep could be severe (Mertens & Mellor, 2003). However, since 1998, eight BTV strains from five different serotypes (BTV-1, 2, 4, 8, 9 and 16) have invaded Europe, collectively representing the largest outbreak of the disease ever recorded, with the deaths of over 1.8 million animals. During this period, BTV has gradually spread further west- and northwards into Europe, reflecting recent increases in the distribution of Culicoides imicola (the major vector species for BTV in the region), which has been influenced by climate change. However, there is also evidence for the involvement of other novel vector species in the region (Culicoides obsolitus and Culicoides pulicaris groups) (Purse et al., 2005).

The ten dsRNA segments of the BTV genome are packaged within a three-layered icosahedral protein capsid (approx. 90 nm in diameter) (Huismans & Erasmus, 1981; Huismans et al., 1983; Mertens et al., 2004). The outer capsid layer is composed of two proteins, VP2 and VP5, encoded by genome segments 2 (Seg-2) and 6 (Seg-6) respectively,
which can elicit a neutralizing antibody response in infected mammalian hosts (Huismans & Erasmus, 1981; Huismans et al., 1983; Roy et al., 1994). VP2 is the most variable of the BTV proteins, containing most of the epitopes that interact with neutralizing antibodies, and consequently is the main determinant of virus serotype (Huismans & Erasmus, 1981; Huismans et al., 1987; Cowley & Gorman, 1987, 1989; Mertens et al., 1989; Gould & Eaton, 1990; Mertens, 1999). VP2 is also the viral haemagglutinin and is responsible for cell attachment during the initiation of infection (Appleton & Letchworth, 1983; Eaton & Crameri, 1989; Huismans et al., 1983, 1987; Kahlon et al., 1983; Mertens et al., 1987a; http://www.iah.bbsrc.ac.uk/dsRNA_virus_proteins/btv-serotype-distribution.htm).

Full-length sequence data for Seg-2 were previously available for only nine of the 24 BTV serotypes (BTV-1, 2, 3, 10, 11, 13, 16, 17 and 23) (Bernard et al., 1997; Bonneau et al., 1999; Fukushima et al., 1987; Ghiasi et al., 1987; Gould & Pritchard, 1990; Inumaru & Roy, 1987; Pritchard & Gould, 1995; Yamaguchi et al., 1988a, b; Yamakawa et al., 1994), with partial sequence data from a further four types. We report full-length sequence data for genome segment 2 from one or more isolates of each of the 24 BTV serotypes. This is the first time that full-length sequence data have been available for a single gene from all 24 BTV types. These data were used to construct phylogenetic trees, establishing the levels of genetic relatedness between each of the different BTV serotypes in Seg-2.

METHODS

Viruses. The reference strains of the 24 BTV serotypes were obtained from the dsRNA virus collection at IAH Pirbright (see http://www.iah.bbsrc.ac.uk/dsRNA_virus_proteins/ReoID/BTV-isolates.htm). These strains were originally provided by the OIE Reference Laboratory at the Onderstepoort Veterinary Institute, South Africa.

dsRNA extraction. A 175 cm² flask (Falcon) of confluent BHK cells was infected with 1 ml (containing approx. 4 log₁₀ TCID₅₀) of each BTV serotype, and incubated at 37 °C under 5 % CO₂ until CPE reached 80 % (approx. 48–72 h). Infected cells were harvested and pelleted by centrifugation (2000 g for 10 min at 4 °C). Cells infected with each of the 24 BTV serotypes were pelleted and total RNA was extracted using Trizol (Life Technologies) according to the manufacturer’s protocol. Single-stranded RNA was removed from total RNA by precipitation with 2 M LiCl (Attoui et al., 2000). Finally, the viral dsRNA was precipitated by addition of 0.25 vols 7.5 M ammonium acetate and 3 vols 100 % ethanol (final ammonium acetate concentration of 0.44 M), washed twice with 70 % ethano- nol and resuspended in RNAse-free water. The quality of the dsRNA was assessed by 1 % agarose gel electrophoresis (AGE) (7 V cm⁻¹ for 1 h) in TAE buffer containing 0.5 µg ethidium bromide ml⁻¹.

cDNA synthesis and amplification by PCR. The concentration of dsRNA was determined by measurement of absorbance at a wave- length of 260 nm (using an absorbance coefficient of 20 mg⁻¹ ml⁻¹). Using the method described by Shapiro et al. (2005), which is based on that of Lambden et al. (1992), approximately 500 ng dsRNAs was ligated to a single-stranded ‘anchor primer’, GAGGGATCC GATTCTAAACG/Sp9/TCCAGTTAGAATCC-OH (Sp9 is a C9 spacer) that forms a self-priming hairpin structure. These reactions were carried out in a final volume of 10 µl, containing 1 µl RNA ligase buffer and 10 U T4 RNA ligase (New England Biolabs). The reactions were incubated at 4 °C for 12 h, the products were then precipitated by adding 0.1 vols 3 M sodium acetate and an equal volume of 2-propanol. After the addition of 11 µl water to the pellet, the tube was boiled for 4 min to denature the dsRNA and the first strand of cDNA was synthesized by using avian myeloblastosis virus reverse transcriptase (Promega) at 37 °C for 1 h (using the manufacturer’s recommended conditions). PCR was subsequently performed with the Triple Master PCR system (Eppendorf) with a single primer, 5’-GAGGGATCCGATTCTAAACG-3’ (primer 5-15-1) as described by Shapiro et al. (2005), which is partially complementary (underlined region) to the anchor primer oligonucleotide sequence and contains a BanHI site (in bold) under the following conditions for 30 cycles: denaturing at 95 °C for 30 s and annealing and extension at 68 °C for 3 min.

Cloning of full-length Seg-2 cDNAs. Full-length PCR amplicons were generated for all 10 genome segments from the reference strains of the 24 BTV serotypes. The amplicons from the different segments of each virus strain were separated on a 1 % TAE agarose gel and the Seg-2 cDNA (approx. 2.9 kb) was excised and purified using a GFX PCR DNA and gel band purification kit (Amersham Pharmacia Biotech) as per the manufacturer’s instructions. The PCR products generated using the Triple Master PCR system were purified and A-tailed at 70 °C for 30 min by using 0.25 U Taq DNA polymerase and 0.2 mM dATP (final concentration). The A-tailed full-length Seg-2 amplicons were purified by phenol/chloroform extraction, then ligated to pGEM-T Easy vector (Promega). The recombinant plasmids were used to transform Escherichia coli (ElectroMAX DH10B) cells. The cloned insert from Seg-2 was amplified by PCR with anchor-specific primer 5-15-1; its identity was confirmed by the length of the PCR product (approx. 2.9 kb) and restriction mapping using BanHI.

Sequencing of full-length Seg-2. Plasmid DNA was prepared from positive clones with inserts using a Genelute Plasmid miniprep kit (Sigma). The termini of the Seg-2 cDNA inserts were sequenced with universal M13/T-7 forward and SP-6 reverse primers (Promega) using a Cycle Sequencing Ready Reaction kit (CEQ DTCS Beckman Coulter) on a Beckman capillary sequencer. Internal forward and reverse primers were designed from the resulting terminal sequence data, making it possible to generate full-length sequence of Seg-2 in both directions. At least three clones from different cDNAs were completely sequenced for each segment and the sequence of at least one of the clones was determined from both directions. The sequences obtained were used to design gene-specific primers for direct sequencing of the RT-PCR products from Seg-2 (without cloning), which confirmed the original sequence data for all 24 serotypes. No differences were detected between the sequences for the PCR products obtained in this way (which represent the whole RNA population in each case) and the consensus sequence for the cloned cDNAs.

Nucleotide sequence data assembly and analysis. The Seg-2 sequence data from computerized databases were analysed and compared using the BLAST program (Altschul et al., 1997). Sequence assembly and analysis of Seg-2, as well as multiple alignments of the deduced VP2 amino acid sequences and phylogenetic comparisons, were performed using MEGA2 (Kumar et al., 2001), Open Reading Frame (ORF) Finder (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=genbank) and CLUSTAL X v. 1.81 (Thompson et al., 1997). Phylogenetic trees (neighbour-joining) created using CLUSTAL X were viewed using TreeView (v. 1.6.6) (http://taxonomy.zoology.gla.ac.uk/rod/ treeview.html).
RESULTS

Cloning and sequencing of VP2 genes

The methods previously used to clone full-length cDNA copies of large genes are technically demanding (Vreede et al., 1998). However, more rapid and reliable sequence-independent cloning and sequencing methods were recently optimized for use with dsRNAs (Shapiro et al., 2005). Seg-2 cDNAs were therefore synthesized for representative isolates of the 24 BTV serotypes, including the 24 reference strains, using these methods. The Seg-2 cDNAs were subsequently purified and separated from the copies of other BTV genome segments by AGE. Full-length Seg-2 cDNAs were cloned into pGEM-T Easy vector and positive clones identified by insert size, restriction fragment pattern (after treatment with BamHI) and by sequencing of near-terminal regions of the cDNA inserts, to identify the conserved terminal hexanucleotides that are characteristic of BTV genome segments (Mertens & Sangar, 1985; Mertens et al., 2004; see below).

Sequence analysis of Seg-2

The nucleotide sequences of Seg-2, and the deduced amino acid sequences of VP2, from each of the BTV strains analysed, have been submitted to GenBank (see GenBank accession numbers in Table 1). BLAST (NCBI) analysis of each of the sequences confirmed that they were all cognate with VP2/Seg-2 sequences that have previously been published for other BTV strains (BTV-1, 2, 3, 10, 11, 13, 16, 17 and 23 – GenBank accession nos: X06464, AF155218, L29026, M17437, D00153, AF530067, AF017278 and U04200, respectively). However, no data were previously available for the reference stains of any of the BTV serotypes. As previously reported for the genome segments of other orbiviruses, Seg-2 from each of the BTV serotypes has conserved terminal hexanucleotides, monocistronic coding, short untranslated regions at the upstream and downstream termini, with 3′-terminal nucleotides ---AC-3′ on both strands (Table 1) (Mertens & Sangar, 1985; Mertens et al., 2004).

Near-terminal non-coding regions (NCRs)

Comparison of the near-terminal NCRs of Seg-2 from the 24 BTV reference strains revealed differences in their length: 16–21 nt at the upstream terminus and 34–37 nt at the downstream terminus (Table 1). The NCRs showed a higher overall level of conservation between serotypes than the coding region of Seg-2 (Fig. 1), with 2.9–64 % variation in the 3′ NCR and 5.9–62 % variation in the 5′ NCR. Comparisons between orbivirus species also showed higher conservation in these terminal non-coding regions than in the ORF (Fig. 1).

Almost all of the terminal hexanucleotides of Seg-2 from the 24 BTV reference strains are in agreement with the conserved sequences previously determined for BTV and Epizootic hemorrhagic disease virus (EHDV) genome segments (positive strand 5′-GUUAAA------ACUUAC-3′) (Mertens & Sangar, 1985; Mertens et al., 2004) (Table 1). The only exception was BTV-24, which has G instead of A at the sixth position from the 3′ end of the positive strand. This sequence change (in comparison to the other BTV strains) was observed by sequence analysis of both cloned cDNAs and by direct sequence analysis of PCR products from BTV-24 and therefore does not appear to be a cloning artefact. The sequence contexts of the initiation codon on the positive strand are in agreement with the consensus sequence for strong eukaryotic translation initiation sites (Kozak, 1987) with a purine (A or G) at position −3 (3 nt upstream of the AUG codon) and a guanosine at position +4 (1 nt downstream of the AUG codon). This arrangement was found in each Seg-2 analysed, except for BTV-24, which has a ‘C’ at position −3.

Comparison of full-length Seg-2 nucleotide sequences

The full-length Seg-2 sequences from the 24 BTV reference strains were aligned using CLUSTAL X (with gaps). In order to allow for the gaps included in the alignment, nucleotide positions are referred to in the text with reference to their relative position in the reference strain of BTV-1 (RSArrrr/01). Although the overall base composition of each Seg-2 is similar, the nucleotide sequence varied between reference strains of different BTV serotypes by 29 % (BTV-8 and BTV-18) to 59 % (BTV-16 and BTV-22) (Supplementary Table S1, available in JGV Online). Comparisons of Seg-2 sequences showed that the terminal hexanucleotides are the longest regions that are fully conserved across the reference strains of all 24 types. The only other region of more than 3 nt that is fully conserved is bases 18–22 ----AUGGA--- (positive strand in RSArrrr/01). This corresponds to the translation initiation site, which clearly has an essential functional role that is reflected by its conservation. Another sequence of 3 nt (GAC), which is fully conserved in Seg-2 of the 24 reference strains, was detected near to the downstream terminus (positive position 2911–2913 in RSArrrr/01).

The Seg-2 nucleotide alignment showed evidence of several insertions and deletions between BTV types (Table 2; Supplementary Fig. S1), some of which were unique to individual serotypes. For example, there was an insertion of 12 bases at nt position 122–134 in the aligned sequence of BTV-8 (position 112–113 with reference to RSArrrr/01) and an insertion of 6 nt at position 830–835 in the aligned sequence of BTV-19 (position 773–774 with reference to RSArrrr/01). Other variations were detected that were present in strains of several different BTV types. For example, Seg-2 of types 12, 15 and 22 contained several longer insertions in the aligned sequences (e.g. at nt 360–369), as well as deletions (between positions 1888 and 1926), compared to other Seg-2s (Table 2). This reflects the relatively close genetic relationship that was detected between types 15 and both 12 and 22, even though they are classified within separate nucleotype groups (nucleotype
Table 1. Characteristics of dsRNA segment 2 (cDNA copy) and VP2 protein of the reference strain of each of the 24 BTV serotypes

Letters in bold are the 5’- and 3’-terminal conserved sequences in genome segment 2 of BTV. Start codons and termination codons are underlined. There is a substitution of G for A at the sixth position from the 3’ end in the genome segment of BTV RSArrrr/24 (shaded).

<table>
<thead>
<tr>
<th>BTV type</th>
<th>IAH dsRNA virus reference collection no.</th>
<th>GenBank accession no.</th>
<th>Segment length (bp)</th>
<th>Open reading frame (nt: including stop codon)</th>
<th>G+C content (mol%)</th>
<th>Size of protein (aa)</th>
<th>Predicted protein molecular mass (kDa)</th>
<th>5’-Terminal sequences of the positive strand</th>
<th>3’-Terminal sequences of the positive strand</th>
</tr>
</thead>
<tbody>
<tr>
<td>RSArrrr/01</td>
<td>AJ58512</td>
<td>2940</td>
<td>18–2903</td>
<td>42.76</td>
<td>961</td>
<td>112</td>
<td>GTTAAATGAGCCGCGAGT</td>
<td>TAGCTCTGCTGACGCCGAGATCCGCGGAGCATTAGACTTAC-3’</td>
<td></td>
</tr>
<tr>
<td>RSArrrr/02</td>
<td>AJ58513</td>
<td>2943</td>
<td>18–2906</td>
<td>41.11</td>
<td>962</td>
<td>112</td>
<td>GTTAAAAACAGGATGCGAGT</td>
<td>TAGCTCTGCTGACGCCGAGATCCGCGGAGCATTAGACTTAC-3’</td>
<td></td>
</tr>
<tr>
<td>RSArrrr/03</td>
<td>AJ58514</td>
<td>2935</td>
<td>22–2901</td>
<td>42.18</td>
<td>959</td>
<td>112</td>
<td>GTTAAAAAACGCTGGCGAGGATG</td>
<td>TAAACCCGCTGACAGGGGTTGGGCGGACATTACCTTAC-3’</td>
<td></td>
</tr>
<tr>
<td>RSArrrr/04</td>
<td>AJ58515</td>
<td>2926</td>
<td>20–2890</td>
<td>42.89</td>
<td>956</td>
<td>111</td>
<td>GTTAAAAAGAGGTGCCCAATTAG</td>
<td>TAGCTCTGCTGACGCCGAGATCCGCGGAGCATTAGACTTAC-3’</td>
<td></td>
</tr>
<tr>
<td>RSArrrr/05</td>
<td>AJ58516</td>
<td>2921</td>
<td>18–2885</td>
<td>41.83</td>
<td>955</td>
<td>110</td>
<td>GTTAAAAAGCTTCTGAGATG</td>
<td>TAGCTCTGCTGACGCCGAGATCCGCGGAGCATTAGACTTAC-3’</td>
<td></td>
</tr>
<tr>
<td>RSArrrr/06</td>
<td>AJ58517</td>
<td>2922</td>
<td>18–2885</td>
<td>42.78</td>
<td>955</td>
<td>110</td>
<td>GTTAAAAATGAGCTGCGAGATG</td>
<td>TAGCTCTGCTGACGCCGAGATCCGCGGAGCATTAGACTTAC-3’</td>
<td></td>
</tr>
<tr>
<td>RSArrrr/07</td>
<td>AJ58518</td>
<td>2936</td>
<td>22–2901</td>
<td>43.19</td>
<td>959</td>
<td>111</td>
<td>GTAATTAAAAAGACCTGGCGAGATG</td>
<td>TAAACCCGCTGACAGGGGTTGGGCGGACATTACCTTAC-3’</td>
<td></td>
</tr>
<tr>
<td>RSArrrr/08</td>
<td>AJ58519</td>
<td>2939</td>
<td>18–2903</td>
<td>42.4</td>
<td>961</td>
<td>111</td>
<td>GTAATTAAAAAGACCTGGCGAGATG</td>
<td>TAAACCCGCTGACAGGGGTTGGGCGGACATTACCTTAC-3’</td>
<td></td>
</tr>
<tr>
<td>RSArrrr/09</td>
<td>AJ58520</td>
<td>2921</td>
<td>18–2885</td>
<td>43.65</td>
<td>955</td>
<td>110</td>
<td>GTAATTAAAAAGACCTGGCGAGATG</td>
<td>TAAACCCGCTGACAGGGGTTGGGCGGACATTACCTTAC-3’</td>
<td></td>
</tr>
<tr>
<td>RSArrrr/10</td>
<td>AJ58521</td>
<td>2926</td>
<td>20–2890</td>
<td>41.87</td>
<td>956</td>
<td>111</td>
<td>GTAATTAAAAAGACCTGGCGAGATG</td>
<td>TAAACCCGCTGACAGGGGTTGGGCGGACATTACCTTAC-3’</td>
<td></td>
</tr>
<tr>
<td>RSArrrr/11</td>
<td>AJ58522</td>
<td>2926</td>
<td>20–2890</td>
<td>42.96</td>
<td>956</td>
<td>111</td>
<td>GTAATTAAAAAGACCTGGCGAGATG</td>
<td>TAAACCCGCTGACAGGGGTTGGGCGGACATTACCTTAC-3’</td>
<td></td>
</tr>
<tr>
<td>RSArrrr/12</td>
<td>AJ58523</td>
<td>2904</td>
<td>17–2869</td>
<td>43.53</td>
<td>950</td>
<td>110</td>
<td>GTAATTAAAAAGACCTGGCGAGATG</td>
<td>TAAACCCGCTGACAGGGGTTGGGCGGACATTACCTTAC-3’</td>
<td></td>
</tr>
<tr>
<td>RSArrrr/13</td>
<td>AJ58524</td>
<td>2935</td>
<td>22–2901</td>
<td>42.11</td>
<td>959</td>
<td>112</td>
<td>GTAATTAAAAAGACCTGGCGAGATG</td>
<td>TAAACCCGCTGACAGGGGTTGGGCGGACATTACCTTAC-3’</td>
<td></td>
</tr>
<tr>
<td>RSArrrr/14</td>
<td>AJ58525</td>
<td>2922</td>
<td>18–2885</td>
<td>42.47</td>
<td>955</td>
<td>110</td>
<td>GTAATTAAAAAGACCTGGCGAGATG</td>
<td>TAAACCCGCTGACAGGGGTTGGGCGGACATTACCTTAC-3’</td>
<td></td>
</tr>
<tr>
<td>RSArrrr/15</td>
<td>AJ58526</td>
<td>2909</td>
<td>17–2875</td>
<td>43.14</td>
<td>952</td>
<td>111</td>
<td>GTAATTAAAAAGACCTGGCGAGATG</td>
<td>TAAACCCGCTGACAGGGGTTGGGCGGACATTACCTTAC-3’</td>
<td></td>
</tr>
<tr>
<td>RSArrrr/16</td>
<td>AJ58527</td>
<td>2935</td>
<td>22–2901</td>
<td>42.25</td>
<td>959</td>
<td>112</td>
<td>GTAATTAAAAAGACCTGGCGAGATG</td>
<td>TAAACCCGCTGACAGGGGTTGGGCGGACATTACCTTAC-3’</td>
<td></td>
</tr>
<tr>
<td>RSArrrr/17</td>
<td>AJ58528</td>
<td>2923</td>
<td>20–2887</td>
<td>41.77</td>
<td>955</td>
<td>111</td>
<td>GTAATTAAAAAGACCTGGCGAGATG</td>
<td>TAAACCCGCTGACAGGGGTTGGGCGGACATTACCTTAC-3’</td>
<td></td>
</tr>
<tr>
<td>RSArrrr/18</td>
<td>AJ58529</td>
<td>2927</td>
<td>18–2891</td>
<td>43.76</td>
<td>957</td>
<td>110</td>
<td>GTAATTAAAAAGACCTGGCGAGATG</td>
<td>TAAACCCGCTGACAGGGGTTGGGCGGACATTACCTTAC-3’</td>
<td></td>
</tr>
<tr>
<td>RSArrrr/19</td>
<td>AJ58530</td>
<td>2922</td>
<td>22–2910</td>
<td>43.5</td>
<td>962</td>
<td>111</td>
<td>GTAATTAAAAAGACCTGGCGAGATG</td>
<td>TAAACCCGCTGACAGGGGTTGGGCGGACATTACCTTAC-3’</td>
<td></td>
</tr>
<tr>
<td>RSArrrr/20</td>
<td>AJ58531</td>
<td>2925</td>
<td>20–2890</td>
<td>42.7</td>
<td>956</td>
<td>111</td>
<td>GTAATTAAAAAGACCTGGCGAGATG</td>
<td>TAAACCCGCTGACAGGGGTTGGGCGGACATTACCTTAC-3’</td>
<td></td>
</tr>
<tr>
<td>RSArrrr/21</td>
<td>AJ58532</td>
<td>2922</td>
<td>18–2885</td>
<td>42.44</td>
<td>955</td>
<td>111</td>
<td>GTAATTAAAAAGACCTGGCGAGATG</td>
<td>TAAACCCGCTGACAGGGGTTGGGCGGACATTACCTTAC-3’</td>
<td></td>
</tr>
<tr>
<td>RSArrrr/22</td>
<td>AJ58533</td>
<td>2907</td>
<td>17–2872</td>
<td>43.07</td>
<td>951</td>
<td>111</td>
<td>GTAATTAAAAAGACCTGGCGAGATG</td>
<td>TAAACCCGCTGACAGGGGTTGGGCGGACATTACCTTAC-3’</td>
<td></td>
</tr>
<tr>
<td>RSArrrr/23</td>
<td>AJ58534</td>
<td>2922</td>
<td>18–2885</td>
<td>42.57</td>
<td>953</td>
<td>111</td>
<td>GTAATTAAAAAGACCTGGCGAGATG</td>
<td>TAAACCCGCTGACAGGGGTTGGGCGGACATTACCTTAC-3’</td>
<td></td>
</tr>
<tr>
<td>RSArrrr/24</td>
<td>AJ58535</td>
<td>2924</td>
<td>21–2888</td>
<td>41.96</td>
<td>955</td>
<td>111</td>
<td>GTAATTAAAAAGACCTGGCGAGATG</td>
<td>TAAACCCGCTGACAGGGGTTGGGCGGACATTACCTTAC-3’</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Amino acid deletions and insertions in aligned VP2 sequences of different BTV serotypes

See Supplementary Fig. S2.

<table>
<thead>
<tr>
<th>Serotype</th>
<th>Positions (aa) in the aligned sequences (positions with respect to RSArrrrr/01)</th>
<th>Deletions</th>
<th>Insertions unique to a serotype</th>
<th>Aligned regions not showing insertions or deletions</th>
</tr>
</thead>
<tbody>
<tr>
<td>BTV-5 and 9</td>
<td>414–419 (395–396)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BTV-7 and 19</td>
<td>749–752 (725–726)</td>
<td>259 (242–243); 266–267 (248–249)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BTV-19</td>
<td>750 (725–726)</td>
<td>172 (158–159)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BTV-3, 13, 16</td>
<td>724–228 (209–210)</td>
<td>36–38 (32–33)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BTV-8</td>
<td>513 (491–492); 630–633 (608–609)</td>
<td>292–410 (273–391); 418–454 (398–434); 452–471</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BTV-4, 10, 11, 17, 20, 24</td>
<td>288–292 (269–270); 669–671 (646–647)</td>
<td>13 (11–12); 103–108; 241 (225–226); 794 (769–770); 942 (916–917)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BTV-12, 15, 22</td>
<td>171–175 (158–159); 216 (201–202); 241 (225–226); 794 (769–770); 942 (916–917)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BTV-15</td>
<td>621–638 (600–601)</td>
<td>101–108 (95–96)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BTV-12</td>
<td>621–636 (600–601)</td>
<td>103–108 (95–96)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BTV-1</td>
<td>472 (450–452)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
J and G, respectively) (Fig. 2; Table 2). There were also several areas of Seg-2 from the 24 serotypes where no insertions or deletions were evident (e.g. positions 370–424, 900–1250, 1580–1850, etc., in the aligned sequences; see Supplementary Fig. S1).

**VP2 amino acid sequence variation**

Seg-2 encodes a protein that ranges from 950 aa (2852 bp ORF) in BTV-12, to 962 aa (2888 bp ORF) in BTV-19 (Table 1). The predicted amino acid sequences of VP2 from the 24 reference strains were aligned using the program CLUSTAL X (Supplementary Fig. S2, available in JGV Online). VP2 showed between 22.7 % (BTV-4 and BTV-20) and 72.9 % (BTV-6 and BTV-22) amino acid variation between BTV types (Fig. 1; Supplementary Table S1). Only the carboxy terminus (aa 946–961) and residues at positions 338–379 appeared to be relatively conserved. A maximum of only five consecutive residues were conserved across all 24 serotypes (PCLRG at position 362–366), with three conserved residues at two other positions (NFL at position 412–414 and EKD at position 500–502). The amino terminus of VP2 appears to be one of the least conserved regions (Supplementary Fig. S2).

VP2 contains a number of conserved cysteine residues [as previously reported for VP2 of the African horse sickness virus (AHSV) by Williams et al. (1998)]. Five cysteine residues (at positions 162, 281, 363, 617 and 851) were conserved in all of the reference strains from 24 of the BTV serotypes. Four cysteine residues were present in most of the serotypes, at positions 446 and 928 (with the exception of BTV-5 and 9), at position 870 (with the exception of BTV-7, 10, 12, 15, 19 and 22) and at position 937 (with the exception of BTV-5, 7, 9, 12, 15, 19 and 22). However, cysteine residues at positions 149, 392, 479, 699 and 842 in the aligned sequences (positions 136, 373, 458, 675 and 817 with reference to RSArrrr/01) were limited to a few serotypes (Supplementary Fig. S2). Many regions of low amino acid sequence identity were detected between VP2 proteins from different BTV types and several ‘gaps’ (deletions) were required to achieve a best-fit alignment, reflecting the very variable nature of the protein (Table 2; Supplementary Fig. S2). For example, an insertion of eight aa was detected at positions 101–108 of the aligned sequences in BTV-15 (positions 95–96 with reference to RSArrrr/01), as well as an insertion of 6 aa at the same site (position 103–108) in BTV-12 and BTV-22. Several other insertions were recorded in the aligned amino acid sequences of the 24 types (as shown in Table 2 and Supplementary Fig. S2). Overall (as might be expected), greater similarity in VP2 amino acid sequences was detected between those BTV types that have previously been reported to cross-react serologically (Erasmus, 1990; Figs 1, 3 and Supplementary Fig. S2).

**Phylogenetic analyses**

Phylogenetic packages including CLUSTAL X and MEGA2 were used for sequence assembly and phylogenetic comparisons of Seg-2 and the deduced VP2 amino acid sequences. Unrooted neighbour-joining (NJ) trees were constructed to visualize the relationships between the 24 BTV reference strains. The trees constructed using neighbour-joining methods with p distance, Jukes–Cantor or Kimura two-parameter algorithms all gave identical phylogenetic groupings. Minimum evolution or UPGMA methods also generated identical trees, confirming the validity of the phylogenetic groupings (which is also indicated by high bootstrap values) and the relationships between the various serotypes. These showed that the different BTV ‘types’ are distinct, with > 28.6 % and > 23 % nucleotide and amino acid divergence, respectively. The amino acid and nucleotide trees both showed nine major branching points, which correlate with the 10 ‘nucleotypes’ suggested by Gould & Pritchard (1990) and reviewed by Mertens (1999). These include: BTV-4, 10, 11, 17, 20 and 24 (identified here as nucleotype A); BTV-3, 13 and 16 (nucleotype B); BTV-6, 14 and 21 (nucleotype C); BTV-8, 18 and 23 (nucleotype D); BTV-5 and 9 (nucleotype E); BTV-7 and 19 (nucleotype F); BTV-12 and 22 (nucleotype G); BTV-1 (nucleotype H); BTV-2 (nucleotype I); and BTV-15 (nucleotype J).

Previously, comparisons of the reference strains of BTV using serological methods (Erasmus, 1990) identified antigenic relationships between the different serotypes. In most cases, these similarities (particularly the stronger cross-reactions) were reflected by grouping of the reference strains
As previously reported, there were some similarities between VP2 of BTV and that of the other Culicoides-borne orbiviruses EHDV-1 (Iwata et al., 1991, 1992a) and AHSV-4 (Iwata et al., 1992b). The VP2 sequence from EHDV was most closely related to that of BTV, with 78.3% amino acid variation (for EHDV-1, GenBank accession no. D10767, and BTV-12) up to 81.5% (for EHDV-1 and BTV-23). This is higher than the maximum variation detected between different BTV types (72.9% for BTV-6 and BTV-22). The level of variability detected in VP2 increased still further between BTV and the other more distantly related orbivirus species: to a minimum of 89.1%, between BTV-24 and Chuzan virus (Palyam species, GenBank accession no. AB014725); 91% between BTV-22 and AHSV-5 (accession no. AY163331); 91.4% between BTV-11 and Broadhaven virus 2 (BRDV-2, accession no. M87875); and 94.8% between BTV-13 and St Croix river virus (SCRV, accession no. AF133432) (Supplementary Fig. S3, available in JGV Online). In each case the BTV types cited are those that showed greatest similarity to these other orbiviruses. It is apparent from the phylogenetic tree (Supplementary Figure S3) that VP2 of different orbivirus species are maintained as distinct monophyletic clusters, with the BTV branch more closely related to EHDV than to AHSV.

**DISCUSSION**

This paper describes the first comparison of full-length sequence data for genome segment 2 from representative (reference) isolates of all 24 BTV serotypes.

Conserved terminal sequences are a common feature of the genome segments of members of the Reoviridae (Mertens & Sangar, 1985; Mertens et al., 2004) and they may provide important recognition signals for initiation of transcription and/or RNA packaging (Rao et al., 1983). Conserved 5′ terminal hexanucleotides (5′-GUUAAA---) were identified on the positive strand of Seg-2 from each of the 24 BTV reference strains. A conserved hexanucleotide is also present at the downstream 3′ termini of most of the BTV Seg-2s analysed, with the exception of BTV-24. In this case the sixth nucleotide from the terminus was G instead of A (---GCUUAC-3′ instead of ---ACUUAC-3′). It is not clear what effect, if any, this change has on the virus or if it might be a modification caused by virus growth in cell culture. Small inverted repeats were observed adjacent to the conserved terminal hexanucleotides, which may also form part of recognition signals for sorting and packaging of the RNA segments during replication, as suggested for other members of the genus Orbivirus (Iwata et al., 1991, 1992a; Moss et al., 1992; Roy et al., 1991), Wound tumor virus (Anzola et al., 1987) and influenza virus (Muramoto et al., 2006; Noda et al., 2006). It is also possible that these motifs are involved in transcription or replication by the virus-specific RNA-dependent RNA polymerase, or in translation (Yamakawa et al., 1999). Two other conserved sequences were observed in the sequences of Seg-2; these include the translation initiation site and a 5 nt sequence of unknown function in the downstream NCR (nt 2911–2913).
Previous studies of reassortant viruses, created using parental strains from two different BTV serotypes, have demonstrated that both Seg-2 and Seg-6 (encoding the two outer capsid proteins VP2 and VP5, respectively) can influence virus serotype, although the influence of Seg-2 is more significant (Cowley & Gorman, 1989; Mertens et al., 1989). Phylogenetic comparisons of VP2 from the 24 reference strains, as well as from over 200 other BTV isolates from around the world (Maan, 2004), show a perfect correlation between sequence variation in Seg-2/VP2 and BTV serotype. This confirms the importance of VP2 as the major determinant of BTV serotype. The low level of Seg-2 sequence identity detected between different BTV serotypes (Fig. 2) agrees with previous sequencing studies (Gould & Pritchard, 1990; de Mattos et al., 1994; Pritchard & Gould, 1995; Bonneau et al., 1999). It also agrees with the results of cross-hybridization studies, which show that partial or full-length Seg-2 probes hybridize in a serotype-specific manner to the denatured dsRNAs of different BTV isolates (Mertens et al., 1987b).

The Seg-2/VP2 sequences of the reference strains of different BTV serotypes clustered as ten distinct evolutionary lineages, identified as nucleotypes A–J (Figs 1, 2). Within each nucleotype, the Seg-2 ORF and NCRs were uniform in length. The grouping of different BTV types within these nucleotypes also reflects the serological relationships between different BTV serotypes (Erasmus, 1990; Zhang et al., 1999; Gould, 1988; Gould & Pritchard, 1990) (Fig. 3).

A multiple alignment of VP2 amino acid sequences from all of the 24 BTV serotypes showed from 22.7 % to 72.9 % variation. Despite the overall sequence variability, some features of VP2 appeared to be conserved across serotypes, including the hydrophobicity profile (data not shown), charge distribution (data not shown) and the position of certain cysteine residues. This suggests that the basic structure of VP2 remains similar and, despite differences in neutralizing epitopes, there are significant constraints on protein variability. The amino acid sequence alignment showed that, although VP2 is very variable, several regions are relatively more conserved between serotypes, which may reflect structural or functional constraints on the protein (Supplementary Fig. S2, available in JGV Online). This supports earlier conclusions from limited datasets (Roy, 1989; Yamakawa et al., 1994). Conserved sequences are most evident in the carboxy-terminal region (aa 946–961) and the central region of the VP2 molecule (aa 357–398) (Supplementary Fig. S2). The longest conserved region of BTV VP2, which contains five consecutive amino acids (---PCLRG, residues 362–366), may play a significant role in the structure or function of the protein. The conserved and variable regions of VP2 identified here are similar to those previously identified by comparison of the four US serotypes (BTV-2, 11, 13 and 17) (Yamaguchi et al., 1988b). It has been suggested that conservation of the carboxy terminus of BTV and AHSV VP2 may reflect functional constraints caused by interactions with VP5 in the outer capsid (Roy, 1989; Williams et al., 1998).

Multiple alignment of the full-length VP2 aa sequences from the 24 BTV types showed lower levels of identity in the regions of VP2 sequence that have previously been indicated as antigenically important. These regions are mostly hydrophilic, suggesting that they could be located on the virion surface and may function as or contribute to neutralization epitopes (DeMaula et al., 1993, 2000; Gould, 1988; Pritchard & Gould, 1995). The amino terminus represents one of the more variable regions of VP2, suggesting that it could play a significant role in determination of virus serotype (Yamaguchi et al., 1988b). The region spanning aa 330–338 of VP2 of BTV-1 has previously been identified as a potential determinant of virus serotype (Gould, 1988). Studies using neutralizing monoclonal antibody escape mutants (DeMaula et al., 1993, 2000) have also linked aa regions 199–213 and 321–346 of VP2 with the determination of virus serotype in BTV-10. Pierce et al. (1995) reported that a neutralizing epitope common to BTV-10 and BTV-17 had a similar location in VP2 (330–338). Less definitive studies using peptide blocking or sequence analysis of VP2 from different BTV strains also identified regions that might contribute to interactions with neutralizing antibodies (aa 199–213 and aa 642–651) (Bernard et al., 1997; Hwang & Li, 1993).

Comparisons of VP2s from the members of different orbivirus species (AHSV, EHDV, BRDV, SCRV and Chuzan virus) showed only low levels of amino acid sequence identity (data not shown). Phylogenetic analyses confirmed that Seg-2s/VP2s of these orbiviruses exist as distinct monophyletic groups, despite their overlapping global distributions (Supplementary Fig. S3). This provides further evidence that distinct orbivirus species do not exchange genetic information by reassortment. It also confirms previous serological and sequencing studies, which indicate that BTVs are more closely related to EHDVs than to AHSVs or Palyam viruses (Supplementary Fig. S3).

The initial database of Seg-2 sequences for the 24 reference strains of different BTV serotypes (http://www.iah.bbsrc.ac.uk/dsRNA_virus_proteins/btv_sequences.htm) generated by the studies described here, represents a valuable resource for the comparison and identification of new BTV isolates. It also provides an initial basis for molecular epidemiology studies of BTV, which will provide more detailed information concerning the origin and movements of BTV strains (Maan et al., 2004a; Potgieter et al., 2005). This database will be strengthened by the generation and addition of sequence data for additional BTV strains.

A multiple alignment of full-length Seg-2 sequences is currently being used to identify potential oligonucleotide primer sets, to specifically amplify selected regions of Seg-2 from individual BTV serotypes, for development of RT-PCR-based typing assays. Serotype-specific RT-PCR-based
assays have already been developed for all 24 BTV serotypes, including the six BTV serotypes currently circulating in Europe (BTV-1, 2, 4, 8, 9 and 16) (Maan et al., 2004b). Sequence data for the amplified cDNAs can then be used to confirm BTV serotype within 30 h of sample receipt. BTV typing by these methods is considerably faster than by conventional serological techniques (which can take several weeks), representing a significant improvement for the design and application of an appropriate vaccination strategy.

ACKNOWLEDGEMENTS

These studies were supported by The Commonwealth Scholarship Commission, UK, in the form of a PhD studentship to S.M. The work was also supported by EU contract ‘ReoID’ (QLK2-CT-2000-00143) by DEFRA and by BBSRC.

REFERENCES


