Sequence analysis of the equid herpesvirus 2 chemokine receptor homologues E1, ORF74 and E6 demonstrates high sequence divergence between field isolates

Emma L. Sharp,1 Helen E. Farrell,1,2 Kerstin Borchers,3 Edward C. Holmes4 and Nicholas J. Davis-Poynter1,2

Correspondence
Emma L. Sharp
emma.sharp@aht.org.uk

1Department of Infectious Diseases, Animal Health Trust, Kentford, Newmarket CB8 7UU, UK
2Sir Albert Sakzewski Virus Research Centre, University of Queensland, Herston, QLD 4029, Australia
3Institute for Virology, FU Berlin, Königin-Luise-Str. 49, 14195 Berlin, Germany
4Department of Biology, The Pennsylvania State University, PA 16802, USA

Equid herpesvirus 2 (EHV-2), in common with other members of the subfamily Gammaherpesvirinae, encodes homologues of cellular seven-transmembrane receptors (7TMR), namely open reading frames (ORFs) E1, 74 and E6, which each show some similarity to cellular chemokine receptors. Whereas ORF74 and E6 are members of gammaherpesvirus-conserved 7TMR gene families, E1 is currently unique to EHV-2. To investigate their genetic variability, EHV-2 7TMRs from a panel of equine gammaherpesvirus isolates were sequenced. A region of gB was sequenced to provide comparative sequence data. Phylogenetic analysis revealed six ‘genogroups’ for E1 and four for ORF74, which exhibited approximately 10–38 and 11–27 % amino acid difference between groups, respectively. In contrast, E6 was highly conserved, with two genogroups identified. The greatest variation was observed within the N-terminal domains and other extracellular regions. Nevertheless, analysis of the number of non-synonymous (dN) and synonymous (dS) substitutions per site generally supported the hypothesis that the 7TMRs are under negative selective pressure to retain functionally important residues, although some site-specific positive selection (dN > dS) was also observed. Collectively, these data are consistent with transmembrane and cytoplasmic domains being less tolerant of mutations with adverse effects upon function. Finally, there was no evidence for genetic linkage between the different gB, E1, ORF74 and E6 genotypes, suggesting frequent intergenic recombination between different EHV-2 strains.

INTRODUCTION

Equid herpesvirus 2 (EHV-2) is a lymphotropic herpesvirus with a high prevalence in horse populations worldwide (Borchers et al., 1997). Although the clinical significance of EHV-2 infection is uncertain, EHV-2 has been implicated in conjunctivitis, immunosuppression in foals, pneumonia, respiratory disease and poor racing performance (Collinson et al., 1994; Kershaw et al., 2001; Murray et al., 1996; Nordengrahn et al., 1996). It may also serve as a trans-activating factor to trigger or upregulate equid herpesvirus 1 (EHV-1) and equid herpesvirus 4 (EHV-4) reactivation from latency (Purewal et al., 1992; Welch et al., 1992). EHV-2 establishes latency in B lymphocytes (Drummer et al., 1996) and has been detected in peripheral blood mononuclear cells, the respiratory tract and draining lymph nodes. The virus has also been detected in both peripheral and central nervous systems (Rizvi et al., 1997). EHV-2 and the closely related equid herpesvirus 5 (EHV-5) are members of the genus Rhadinovirus of the subfamily Gammaherpesvirinae.

EHV-2 encodes three homologues of cellular seven-transmembrane receptors (7TMRs), also known as G-protein-coupled receptors (GPCRs), encoded by the genes E1, E6 and ORF74 (Telford et al., 1995). Homologues of 7TMRs are characteristic of the beta- and gammaherpesvirus subfamilies and generally show similarity to cellular chemokine receptors. Chemokine receptors belong to the rhodopsin-like family of 7TMRs (Murphy et al., 2000). Their binding of chemokines, released by cells at
sites of infection and inflammation, induces leukocyte migration and activation through the modulation of intracellular signalling cascades. Therefore, it has been suggested that the immunomodulatory potential of these proteins may be exploited by herpesviruses that encode viral chemokine receptor ‘mimics’, by subversion of host immunity in the vicinity of infected cells or otherwise enhancing virus replication and/or dissemination (Ahuja et al., 1994; Davis-Poynter & Farrell, 1996). Indeed, several in vivo studies have demonstrated the importance of these receptors for virus replication and pathogenesis in the natural host (Beisser et al., 1998, 1999; Davis-Poynter et al., 1997; Oliveira & Shenk, 2001).

E1, which has no known counterparts in other herpesviruses, is present within the 18 kb direct repeat at either end of the genome and hence is EHV-2’s only duplicated gene (Telford et al., 1995). E1 has a high degree of sequence similarity (47% amino acid identity) to human CCR3 and is a functional receptor for the chemokine eotaxin (Camarda et al., 1999).

EHV-2 ORF74 is collinear with homologues encoded by a number of rhadinoviruses, including herpesvirus saimiri (HVS), human herpesvirus 8 (HHV-8), rhesus rhadinovirus (RRV) and mouse herpesvirus strain 68 (MHV-68), all of which share the greatest sequence similarity with the cellular CXC chemokine receptor, CXCR2. Most ORF74 receptors exhibit constitutive signalling activity and promiscuous affinity for CXC chemokines (Ahuja & Murphy, 1993; Arvanitakis et al., 1997; Couty et al., 2001; Rosenkilde et al., 2000, 2004), and have been associated with triggering of cell proliferation and transformation (Bais et al., 1998; Estep et al., 2003; Wakeling et al., 2001; Yang et al., 2000). A highly conserved feature of rhodopsin-like 7TMRs is a ‘DRY’ motif on the intracellular face of the third transmembrane region, the Arg of which is one of the most conserved residues of 7TMRs and has generally been considered to be essential for G-protein coupling. The ORF74 family members encode variants of the ‘DRY’ motif, namely VRY (HHV-8), LRC (HVS), HRC (MHV-68) and IRC (RRV). EHV-2 is the only ORF74 family member encoding a variant motif (DTW) that lacks the central Arg. A recent study of the molecular pharmacology of EHV-2 ORF74 highlighted the novelty of this receptor, which was functional and exhibited both constitutive and ligand-induced G-protein-mediated signalling (Rosenkilde et al., 2005). Furthermore, EHV2-ORF74, in contrast to other ORF74 receptors, displayed a comparatively narrow ligand-binding spectrum, restricted to CXCL6, and a limited rather than promiscuous activation of G proteins.

EHV-2 E6 is positioned between two conserved gene blocks and is poorly conserved in comparisons with cellular chemokine receptors. It is homologous and collinear with members of the Epstein–Barr virus (EBV) BILFI 7TMR gene family (Beisser et al., 2005; Davis-Poynter & Farrell, 1996; Paulsen et al., 2005), which include ORF A5 of porcine lymphotropic herpesvirus (PLHV) types 1, 2 and 3, and ORF A5 of alcelaphine herpesvirus 1 (AIHV-1). BILFI is a highly constitutively active receptor, signalling through G proteins of the Gi/o class (Beisser et al., 2005; Paulsen et al., 2005).

Considerable biological, antigenic and genomic variability has been observed among EHV-2 field isolates (Browning & Studdert, 1987, 1989; Plummer et al., 1973), but investigations of sequence variability are limited. We therefore conducted a comprehensive sequence analysis of the three EHV-2 7TMRs from a panel of field isolates, to (i) determine whether the extent and structure of the genetic variability were equivalent among the three 7TMR ORFs, (ii) investigate the likely importance of specific residues and domains and (iii) determine phylogenetic relationships and assess the frequency of recombination between different genotypes.

EHV-5 is related closely to EHV-2 and, although genomic sequence data are limited, the complete sequence of EHV-5 gB has been determined (Holloway et al., 1999; Telford et al., 1993). It is not known whether EHV-5 encodes 7TMRs homologous to those of EHV-2. If present, the PCR primers used for the amplification of EHV-2 7TMR may potentially amplify homologous regions of EHV-5. A hypervariable region of gB was sequenced to verify whether isolates were EHV-2 or EHV-5 and to provide comparative data for a variable genetic locus unrelated to the 7TMR.

METHODS

**Virus isolates.** Viral DNA extracts were obtained from 33 tissue-culture-grown field isolates originating from Australia (*n*=9), England (*n*=13), Germany (*n*=8), Hungary (*n*=1) and Sweden (*n*=2). Isolate Aus-1 originally designated EHV2.141) has been described previously (Holloway et al., 2000). In addition, two laboratory strains, LK4 (Plummer & Waterson, 1963) and T400 (Thein & Hartl, 1976), were included (Table 1). Viral DNA prepared from virus-infected primary equine kidney cells (equine embryonic lung or equine embryonic kidney) was either supplied by the contributing laboratories or prepared in house, using a High Pure PCR template preparation kit (Roche). The virus isolates and details of PCR results for the various genes tested are shown in Table 1.

**PCR and sequencing.** Primers used for gene amplification and sequencing are listed in Table 2. For E1, E6 and ORF74, PCR amplification primers were initially designed to bind to the extreme ends of the prototypic ORFs of the sequenced EHV-2 strain 86/67 (Telford et al., 1995). Variation within the primer-binding sites of the E6 and ORF74 sequences of some isolates necessitated the use of alternative ORF-flanking primers. PCR was performed by using a proof-reading ‘Expand Long Template’ PCR kit (Roche) as follows: 3 min at 94 °C, 35 cycles of 30 s at 94 °C, 15 s at 58 °C (50 °C used for alternative forward primers) and 1 min at 68 °C, and 7 min at 68 °C. PCR products were gel-extracted by using a QIAquick gel extraction kit (Qiagen) and then A-tailed (by incubation with Tag DNA polymerase and dATP at 72 °C for 30 min) and cloned into the pGEM-T vector (Promega) for sequencing with M13 vector-specific primers (Promega), or sequenced directly by using the gene-amplification primers and internal gene primers designed to bind the prototypic gene sequence. Gene-variant-specific primers were...
required to complete the sequencing of some isolate E1 and ORF74 sequences. For each cloned PCR product, between two and five clones were sequenced.

For some isolates, there were large discrepancies between clone sequences, indicating the presence of strain mixtures at a particular gene locus. These sequences exhibited $\geq 10$ amino acid changes and were entered separately into the phylogenetic analysis, distinguishable by an alphabetical extension added to the isolate designation (for example, Eng-1a and Eng-1b). For all other, seemingly ‘pure’ isolates, cloned sequences were either identical or exhibited a low number of nucleotide differences. Where such differences were observed, the isolate consensus sequence was taken as the sequence represented by the highest number of clones, or in the case of an equal number of clones for each sequence, as the sequence exhibiting least divergence from other similar sequences. Sequencing was performed by using Big Dye Terminator chemistry version 3, and an ABI 3100 automated sequencer (both from Applied Biosystems).

<table>
<thead>
<tr>
<th>Isolate designation</th>
<th>Origin</th>
<th>Date*</th>
<th>DNA source†</th>
<th>PCR product sequencing‡</th>
<th>gB§</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eng-1</td>
<td>England</td>
<td>1998</td>
<td>TCV</td>
<td>CL</td>
<td>DS</td>
</tr>
<tr>
<td>Eng-2</td>
<td>England</td>
<td>1998</td>
<td>TCV</td>
<td>CL</td>
<td>DS</td>
</tr>
<tr>
<td>Eng-3</td>
<td>England</td>
<td>1998</td>
<td>TCV</td>
<td>CL</td>
<td>DS</td>
</tr>
<tr>
<td>Eng-4</td>
<td>England</td>
<td>1998</td>
<td>TCV</td>
<td>DS</td>
<td></td>
</tr>
<tr>
<td>Eng-5</td>
<td>England</td>
<td>1998</td>
<td>TCV</td>
<td>DS</td>
<td></td>
</tr>
<tr>
<td>Eng-6</td>
<td>England</td>
<td>1998</td>
<td>TCV</td>
<td>DS</td>
<td></td>
</tr>
<tr>
<td>Eng-7</td>
<td>England</td>
<td>1998</td>
<td>TCV</td>
<td>CL</td>
<td>DS</td>
</tr>
<tr>
<td>Eng-8</td>
<td>England</td>
<td>1998</td>
<td>TCV</td>
<td>DS</td>
<td></td>
</tr>
<tr>
<td>Eng-9</td>
<td>England</td>
<td>2004</td>
<td>TCV</td>
<td>CL M</td>
<td>DS</td>
</tr>
<tr>
<td>Eng-10</td>
<td>England</td>
<td>2004</td>
<td>TCV</td>
<td>CL</td>
<td>DS</td>
</tr>
<tr>
<td>Eng-11</td>
<td>England</td>
<td>2004</td>
<td>TCV</td>
<td>CL</td>
<td>DS</td>
</tr>
<tr>
<td>Eng-12</td>
<td>England</td>
<td>2004</td>
<td>TCV</td>
<td>CL</td>
<td>DS</td>
</tr>
<tr>
<td>Eng-13</td>
<td>England</td>
<td>2004</td>
<td>TCV</td>
<td>CL</td>
<td>DS</td>
</tr>
<tr>
<td>T400</td>
<td>England</td>
<td>1998</td>
<td>DNA</td>
<td>CL M</td>
<td>DS</td>
</tr>
<tr>
<td>LK4</td>
<td>England</td>
<td>1998</td>
<td>DNA</td>
<td>CL</td>
<td>DS</td>
</tr>
<tr>
<td>Ger-1</td>
<td>Germany</td>
<td>1999</td>
<td>DNA</td>
<td>CL</td>
<td>DS</td>
</tr>
<tr>
<td>Ger-2</td>
<td>Germany</td>
<td>1999</td>
<td>DNA</td>
<td>CL</td>
<td>DS</td>
</tr>
<tr>
<td>Ger-3</td>
<td>Germany</td>
<td>1999</td>
<td>DNA</td>
<td>CL</td>
<td>DS</td>
</tr>
<tr>
<td>Ger-4</td>
<td>Germany</td>
<td>1999</td>
<td>DNA</td>
<td>CL</td>
<td>DS</td>
</tr>
<tr>
<td>Ger-5</td>
<td>Germany</td>
<td>1999</td>
<td>DNA</td>
<td>CL</td>
<td>DS</td>
</tr>
<tr>
<td>Ger-6</td>
<td>Germany</td>
<td>1999</td>
<td>DNA</td>
<td>CL</td>
<td>DS</td>
</tr>
<tr>
<td>Ger-7</td>
<td>Germany</td>
<td>1999</td>
<td>DNA</td>
<td>CL</td>
<td>DS</td>
</tr>
<tr>
<td>Ger-8</td>
<td>Germany</td>
<td>1999</td>
<td>DNA</td>
<td>CL</td>
<td>DS</td>
</tr>
<tr>
<td>Swe-1</td>
<td>Sweden</td>
<td>1991</td>
<td>TCV</td>
<td>CL</td>
<td>DS</td>
</tr>
<tr>
<td>Swe-2</td>
<td>Sweden</td>
<td>2002</td>
<td>TCV</td>
<td>CL</td>
<td>DS</td>
</tr>
<tr>
<td>Hun-1</td>
<td>Hungary</td>
<td>2002</td>
<td>TCV</td>
<td>CL</td>
<td>DS</td>
</tr>
<tr>
<td>Aus-1 [EHV2.14]</td>
<td>Australia</td>
<td>2000</td>
<td>TCV</td>
<td>CL M</td>
<td>DS</td>
</tr>
<tr>
<td>Aus-2</td>
<td>Australia</td>
<td>1998</td>
<td>DNA</td>
<td>DS M</td>
<td>CL</td>
</tr>
<tr>
<td>Aus-3</td>
<td>Australia</td>
<td>1998</td>
<td>DNA</td>
<td>DS</td>
<td>CL</td>
</tr>
<tr>
<td>Aus-4</td>
<td>Australia</td>
<td>1998</td>
<td>DNA</td>
<td>DS</td>
<td>CL</td>
</tr>
<tr>
<td>Aus-5</td>
<td>Australia</td>
<td>1998</td>
<td>DNA</td>
<td>DS M</td>
<td>CL</td>
</tr>
<tr>
<td>Aus-6</td>
<td>Australia</td>
<td>1998</td>
<td>DNA</td>
<td>DS M</td>
<td>CL</td>
</tr>
<tr>
<td>Aus-7</td>
<td>Australia</td>
<td>1998</td>
<td>DNA</td>
<td>DS M</td>
<td>CL</td>
</tr>
<tr>
<td>Aus-8</td>
<td>Australia</td>
<td>1998</td>
<td>DNA</td>
<td>DS M</td>
<td>CL</td>
</tr>
<tr>
<td>Aus-9</td>
<td>Australia</td>
<td>1998</td>
<td>DNA</td>
<td>DS M</td>
<td>CL</td>
</tr>
</tbody>
</table>

*Date of isolation, or reference cited for previously reported isolates.
†DNA provided by contributing laboratory (DNA) or prepared from tissue-culture-grown virus (TCV).
‡PCR products were either sequenced directly (DS) or cloned for sequencing (CL). ND, Not done; M, presence of strain mixtures within isolate. PCR products obtained by the use of alternative ORF-flanking primers are indicated by italics.
§Variable region of glycoprotein B corresponding to aa 235–631 of EHV-2. Isolates found to be EHV-5 from gB sequencing are indicated in bold.
and sequenced for 10 isolates belonging to different E1 'genogroups', using E1 group-specific, internal reverse primers and a forward primer that bound 1226 bp upstream of the E1 start codon. PCR products were sequenced directly. Eng-3 exhibited a mixture of upstream sequences, so cloned PCR products were sequenced for this isolate.

Sequence analysis. Sequence files were assembled by using Seq-ManII (DNASTAR), and CLUSTAL_W multiple alignments of consensus sequences were constructed by using BioEdit v6.0.7 (Hall, 1999). Manual rearrangements of gapped data were applied to achieve optimal results. Phylogenetic trees were inferred by using the maximum-likelihood method available in the PAUP* package (version 4; Swofford, 2003). In all cases, we used the general time-reversible GTR+I+C model of nucleotide substitution, employing tree bisection–reconstruction (TBR) branch-swapping. To assess the reliability of each node on the tree, we conducted a bootstrap-resampling analysis, employing 1000 neighbour-joining trees estimated by using the GTR+I+C substitution model.

### Table 2. Primers used for amplification and sequencing of E1, E6, ORF74 and gB

<table>
<thead>
<tr>
<th>Gene</th>
<th>Reference</th>
<th>Type*</th>
<th>Primer sequence (5’→3’)†</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1</td>
<td>E1Ef</td>
<td>S</td>
<td>ttgaaattcacagtaaaATGCGAACCAC</td>
</tr>
<tr>
<td></td>
<td>E1Er†</td>
<td>AS</td>
<td>ttgaaattCACAGCGGGTTGGGGCCCCCT</td>
</tr>
<tr>
<td></td>
<td>E1seq1</td>
<td>S</td>
<td>GTCAATAAGATACGATGAGATC</td>
</tr>
<tr>
<td></td>
<td>E1seq2</td>
<td>S</td>
<td>TTTCTGAGGAGGTTGTTG</td>
</tr>
<tr>
<td></td>
<td>E1seq3</td>
<td>S</td>
<td>CTCTCTGCTGTCATTCCCT</td>
</tr>
<tr>
<td></td>
<td>E1seq4</td>
<td>AS</td>
<td>TGGCTCAGAAAAGTCGAAAG</td>
</tr>
<tr>
<td></td>
<td>E1seq5</td>
<td>AS</td>
<td>GATGCAACAGTGTTGTTGAGG</td>
</tr>
<tr>
<td></td>
<td>E1group6F</td>
<td>S</td>
<td>CAGTATGCTTGTCATTCCCT</td>
</tr>
<tr>
<td></td>
<td>E1upstreamF</td>
<td>S</td>
<td>gcttgacagagagagac</td>
</tr>
<tr>
<td></td>
<td>E1upstream2F</td>
<td>S</td>
<td>ctcaactatcaccaaatc</td>
</tr>
<tr>
<td></td>
<td>E1upstreamR</td>
<td>AS</td>
<td>gggcaagagagatataac</td>
</tr>
<tr>
<td></td>
<td>E1group1/2R</td>
<td>AS</td>
<td>AGTGAACAGCGAGTAAAG</td>
</tr>
<tr>
<td></td>
<td>E1group3R</td>
<td>AS</td>
<td>TGTTAACAGTGTTGAGG</td>
</tr>
<tr>
<td></td>
<td>E1group4R</td>
<td>AS</td>
<td>GGTGAAGGCTCATGTTGAGG</td>
</tr>
<tr>
<td></td>
<td>E1group5R</td>
<td>AS</td>
<td>GGGTTAAGCAGTTGTTGAGG</td>
</tr>
<tr>
<td>E6</td>
<td>E6Ef</td>
<td>S</td>
<td>ttgaaattcagATGGAGGAGATGAGACAC</td>
</tr>
<tr>
<td></td>
<td>E6Er†</td>
<td>AS</td>
<td>ttgaaattcct TGAAGAGCGCCGCG</td>
</tr>
<tr>
<td></td>
<td>E6flank4F</td>
<td>S</td>
<td>ttaactctgctgaaacaca</td>
</tr>
<tr>
<td></td>
<td>E6seq1</td>
<td>S</td>
<td>CGTCAAGTGCTAGTCGA</td>
</tr>
<tr>
<td></td>
<td>E6seq2</td>
<td>S</td>
<td>AAGTTCTCATGTCATGCG</td>
</tr>
<tr>
<td></td>
<td>E6seq3</td>
<td>AS</td>
<td>CACATGCACAGGCTCACAG</td>
</tr>
<tr>
<td>ORF 74</td>
<td>74Ef</td>
<td>S</td>
<td>ttgaaattcagATGGAGGAGATGAGG</td>
</tr>
<tr>
<td></td>
<td>74Er†</td>
<td>AS</td>
<td>ttgaaattcct TTAAGGATGCGGCG</td>
</tr>
<tr>
<td></td>
<td>74flank3F</td>
<td>S</td>
<td>ccaataacccaccccaaac</td>
</tr>
<tr>
<td></td>
<td>74flank4R</td>
<td>AS</td>
<td>caataacccaccccaaac</td>
</tr>
<tr>
<td></td>
<td>74seq1</td>
<td>S</td>
<td>ATGACCTTTAGACATAGGGA</td>
</tr>
<tr>
<td></td>
<td>74seq2</td>
<td>S</td>
<td>GAGACACATAGCCTTGGGA</td>
</tr>
<tr>
<td></td>
<td>74seq3</td>
<td>AS</td>
<td>CACAGATATAGGAAGACATATC</td>
</tr>
<tr>
<td></td>
<td>VarORF74seq1.f</td>
<td>S</td>
<td>GGTGTCATGGCTAGTATGTTATATTACTG</td>
</tr>
<tr>
<td></td>
<td>VarORF74seq2.f</td>
<td>S</td>
<td>GATAGTGGCTCGATGATGCTGTGCATG</td>
</tr>
<tr>
<td></td>
<td>VarORF74seq3.r</td>
<td>AS</td>
<td>AACAAGACGCAAGATGCCA</td>
</tr>
<tr>
<td>gB</td>
<td>gBconserved.f</td>
<td>S</td>
<td>AGAACAACACAGTCACTGC</td>
</tr>
<tr>
<td></td>
<td>gBconserved.r†</td>
<td>AS</td>
<td>GTAGTGTGACCTCGCTGCCA</td>
</tr>
<tr>
<td></td>
<td>gBint.1f</td>
<td>S</td>
<td>CACACACAAGGCAAGCTGCC</td>
</tr>
<tr>
<td></td>
<td>gBint.1R</td>
<td>AS</td>
<td>GTCATAGCGAAGCTGCC</td>
</tr>
<tr>
<td></td>
<td>gBint.2f</td>
<td>S</td>
<td>CACAGGTGTCTTCTTGCAG</td>
</tr>
<tr>
<td></td>
<td>gBint.2r</td>
<td>S</td>
<td>CACAGGTGTCTTCTTCACAG</td>
</tr>
<tr>
<td></td>
<td>5gB.1f</td>
<td>S</td>
<td>TATACCAACGCGACTGAGGC</td>
</tr>
<tr>
<td></td>
<td>5gB.2f</td>
<td>S</td>
<td>GACCACGCGAGGAAACCTAC</td>
</tr>
<tr>
<td></td>
<td>5gB.1r</td>
<td>AS</td>
<td>TAGGCAATGCTGCTTATCG</td>
</tr>
<tr>
<td></td>
<td>5gB.2r</td>
<td>AS</td>
<td>CACAGGTGTCTCAGCTTCTGCG</td>
</tr>
</tbody>
</table>

* S, Sense; AS, antisense.
† Upper-case letters denote coding regions. Underlined letters mark EcoRI restriction sites.
‡ Primers used for PCR amplification.
minimum and maximum percentage amino acid differences among isolate pairs between arbitrarily defined phylogenetic groups were calculated by using MEGA version 2.1 (Kumar et al., 2001). Gene- and site-specific selection pressures for each gene were measured as the ratio of non-synonymous ($d_\alpha$) to synonymous ($d_\theta$) substitutions per site, estimated by using the fixed effects likelihood (FEL) method available at the Datamonkey facility (http://www.datamonkey.org/; Pond & Frost, 2005). This analysis also incorporated the GTR model of nucleotide substitution, with input phylogenetic trees inferred by using the neighbour-joining method.

RESULTS AND DISCUSSION

Phylogenetic analysis of gB sequences from EHV-2/EHV-5 isolates

The panel of presumptive EHV-2 field isolates was initially screened, by sequencing a hypervariable region of gB, to verify whether they comprised EHV-2 or the closely related EHV-5 (or a mixture of both). Characterization of EHV-5 gB (Holloway et al., 1999) revealed 66% amino acid identity with the EHV-2 gB sequence, with the region immediately N-terminal to the predicted endoproteolytic cleavage site exhibiting the highest degree of variability between EHV-2 and EHV-5. PCR products spanning the endoproteolytic-cleavage site, corresponding to EHV-2 gB aa 235–609 and EHV-5 gB aa 234–605, were sequenced.

Of the 35 isolates, 31 were identified as EHV-2 and the remaining four as EHV-5. PCR products from 17 isolates consisted of mixtures of gB sequences (Table 1). Each comprised a mixture of EHV-2 gB sequences, apart from a single isolate that comprised a mixture of EHV-5 gB sequences. The 18 isolates with unambiguous gB sequences were analysed and gave rise to similar phylogenetic trees for both nucleotide (Fig. 1a, GenBank accession numbers EF182695–EF182712) and amino acid sequences (data not shown). Based on the 360 amino acid sites used for the analysis, the EHV-2- and EHV-5-lineage sequences differed from each other at the amino acid level by approximately 30–35%. Eleven of the 15 EHV-2 isolate sequences clustered with the reference sequence (86/67) in group 1. Based on the 367 amino acid sites used for the analysis, sequences in different lineages differed by 31–38% at the amino acid level, with a minimum of 10% variation between isolates from different groups. The extent of amino acid variation within each group was generally low (<2%), with the highest variability seen in group 2 (up to 8%).

Despite this high level of variation, all isolate sequences were predicted to conserve the general 7TMR structure (Fig. 2b) comprising an extracellular N terminus (N-term), an intracellular C terminus (C-term) and seven transmembrane domains (TM1–TM7), linked sequentially by three cytoplasmic loops (IC1–IC3) and three extracellular loops (EC1–EC3). Structural ‘hallmarks’ associated with rhodopsin-like 7TMRs were also conserved, including ‘fingerprint’ residues that are the most highly conserved within transmembrane domains of rhodopsin-like receptors (Gether, 2000; Mirzadegan et al., 2003), N-linked glycosylation sites within the extracellular N terminus (two conserved in all groups plus a third specific to lineage A sequences), a ‘DRY’ motif at the cytoplasmic side of TM3 and two cysteine residues involved in disulphide-bond formation between the extracellular face of TM3 and the second extracellular loop (EC2). Additional conserved features particularly associated with chemokine receptors (Chen et al., 2004) included two cysteine residues likely to form a disulphide bond between the N terminus and third extracellular loop (EC3), an acidic N-terminal region and a short, basic third intracellular loop (IC3). Furthermore, E1 sequences retained the conserved TxF motif in TM2, through which TM2 may interact strongly with TM3 (Govaerts et al., 2001), the conserved NPxxY(x)5,6F motif in TM7, which may be important for receptor activation (Fritze et al., 2003), and a cluster of aromatic residues between TM2 and TM3, which may be involved in chemokine-induced activation (Govaerts et al., 2003).

None of the EHV-5 isolates yielded PCR products for any of the 7TMR homologues, whereas these could be amplified from the majority of EHV-2 isolates (see below). It is likely that EHV-5 encodes counterparts of one or more of the EHV-2 7TMRs, but presumably sequence mismatches with the PCR primers prevented their detection.

Phylogenetic analysis of E1 sequences

E1 PCR products were amplified successfully from each of the 31 EHV-2 isolates. E1 amplicons were cloned and sequenced for 23 isolates, of which five comprised a mixture of two different sequence types (Table 1). A further eight isolates were characterized by direct sequencing of the uncloned PCR products; four of these had a mixture of E1 sequence types and were not analysed further. The nucleotide sequences (Fig. 1b, GenBank accession numbers EF182622–EF182653) and amino acid sequences (data not shown) clustered into six distinct phylogenetic groups arising from two main lineages, designated A and B. The prototypic sequence (strain 86/67) was assigned to group 1. Based on the 367 amino acid sites used for the analysis, sequences in different lineages differed by 31–38% at the amino acid level, with a minimum of 10% variation between isolates from different groups. The extent of amino acid variation within each group was generally low (<2%), with the highest variability seen in group 2 (up to 8%).
Fig. 1. Maximum-likelihood phylogenetic trees for partial gB (a) and complete E1 (b), ORF74 (c) and E6 (d) nucleotide sequences for a panel of EHV-2 field isolates. All genogroup designations are shown. All bootstrap values >90% are depicted and horizontal branch lengths are scaled according to the number of nucleotide substitutions per site. All trees are mid-point-rooted for the purposes of clarity only.
Fig. 2. Amino acid alignments of sequences representative of the main phylogenetic groups (defined in Fig. 1, indicated in parentheses). For each alignment, dots mark identities to the group 1 reference sequence of EHV-2 strain 86/67, and dashes mark gaps in the sequence inserted to maximize alignment. (a) Amino acid alignment of EHV-2/EHV-5 gB sequences corresponding to aa 235–609 of the reference EHV-2 gB sequence. Top and bottom sequences correspond to the reference EHV-2 and EHV-5 gB sequences, respectively. The positions of conserved N-glycosylation sequons (boxed, *), the EHV-2/EHV-5 endoproteolytic cleavage site (boxed, 1), and conserved SPF (boxed, 2) and GQLG (boxed, 3) motifs are shown. (b–d) Amino acid alignments of EHV-2 7TMR sequences. Putative transmembrane regions (horizontal bars) with flanking extracellular N termini (N-term), extracellular loops (EC1–EC3), intracellular loops (IC1–IC3) and intracellular C termini (C-term) are shown.
in addition to a number of transmembrane ‘fingerprint’ residues widely conserved among rhodopsin-like 7TMRs (unmarked closed boxes) and conserved cysteine residues likely to be involved in disulphide bond formation (boxed, *). Arrows denote sites that may be positively selected \((P < 0.05)\). (b) Amino acid alignment of E1 sequences. Additional features include two \(N\)-glycosylation sequons (dashed boxes), with a third present in lineage A sequences (groups 1–3), and a DRY motif (*). (c) Amino acid alignment of ORF74 sequences. Additional features include three \(N\)-glycosylation sequons (dashed boxes) and a DRC variant of the DRY motif (*). Two conserved group 2 amino acid sequence types differ by a single conserved amino acid change (boxed, *).

### (c)

<table>
<thead>
<tr>
<th>N-term</th>
<th>TM1</th>
<th>IC1</th>
<th>TM2</th>
<th>EC1</th>
<th>TM3</th>
<th>IC2</th>
<th>TM4</th>
<th>IC3</th>
<th>TM5</th>
<th>IC4</th>
<th>TM6</th>
<th>EC3</th>
</tr>
</thead>
<tbody>
<tr>
<td>86/67</td>
<td>NEKEGBGQD</td>
<td>L254</td>
<td>FFYVFV</td>
<td>6</td>
<td>DUEA</td>
<td>DQ</td>
<td>V</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Eng</td>
<td>3</td>
<td>203</td>
<td>205</td>
<td>207</td>
<td>209</td>
<td>211</td>
<td>213</td>
<td>215</td>
<td>217</td>
<td>219</td>
<td>221</td>
<td>223</td>
</tr>
<tr>
<td>Aus</td>
<td>3</td>
<td>203</td>
<td>205</td>
<td>207</td>
<td>209</td>
<td>211</td>
<td>213</td>
<td>215</td>
<td>217</td>
<td>219</td>
<td>221</td>
<td>223</td>
</tr>
</tbody>
</table>

### (d)

<table>
<thead>
<tr>
<th>N-term</th>
<th>TM1</th>
<th>IC1</th>
<th>TM2</th>
<th>EC1</th>
<th>TM3</th>
<th>IC2</th>
<th>TM4</th>
<th>IC3</th>
<th>TM5</th>
<th>IC4</th>
<th>TM6</th>
<th>EC3</th>
</tr>
</thead>
<tbody>
<tr>
<td>86/67</td>
<td>NEKEGBGQD</td>
<td>L254</td>
<td>FFYVFV</td>
<td>6</td>
<td>DUEA</td>
<td>DQ</td>
<td>V</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Eng</td>
<td>3</td>
<td>203</td>
<td>205</td>
<td>207</td>
<td>209</td>
<td>211</td>
<td>213</td>
<td>215</td>
<td>217</td>
<td>219</td>
<td>221</td>
<td>223</td>
</tr>
<tr>
<td>Aus</td>
<td>3</td>
<td>203</td>
<td>205</td>
<td>207</td>
<td>209</td>
<td>211</td>
<td>213</td>
<td>215</td>
<td>217</td>
<td>219</td>
<td>221</td>
<td>223</td>
</tr>
</tbody>
</table>
The N-terminal domain was the most highly divergent between isolates, whereas the C terminus and several of the transmembrane regions were more highly conserved.

An analysis of the selection pressures acting on E1 indicated an overall occurrence of negative (purifying) selection (mean $d_S : d_S = 0.35$); that is, a tendency for natural selection against amino acid changes that disrupt function and hence reduce fitness. Using a significance cut-off of $P<0.05$, 55 sites were identified as being under negative selection. However, two codon sites (positions 42 and 332 in Fig. 2b) were found to be positively selected. Further sequence analysis of a subset of isolates (two from each of groups 1–5) indicated that, in contrast to the E1 coding sequence, the upstream region is strongly conserved. Thus, an analysis of approximately 1.15 kb upstream of the E1 start codon revealed a maximum of approximately 3% nucleotide site difference across 1110 sites (data not shown, GenBank accession numbers EF182654–EF182666). As there are no known ORFs in the region upstream of E1, this suggests that, whilst substantial alterations to the E1 amino acid sequence have been tolerated, the 5’ elements (presumably controlling expression of E1) have been relatively well conserved. Taken together, these data suggest that E1 diversification has occurred under selective pressure to maintain expression and key functional properties.

### Phylogenetic analysis of ORF74 sequences

PCR amplification of ORF74 with primers binding close to the ORF start and stop codons was successful for 19 of the EHV-2 isolates. PCR products for 11 of the remaining 12 isolates were obtained with the use of an alternative pair of ORF-flanking primers, which bound a short distance upstream and downstream of the original forward and reverse primers respectively. ORF74 amplicons were cloned and sequenced for 20 of the isolates, one of which was found to have a mixture of two distinct sequence types. A further 10 isolates were characterized by direct sequencing of uncloned PCR products, of which two were found to have a mixture of ORF74 sequence types and were not characterized further. Phylogenetic analysis generated similar trees for both nucleotide (Fig. 1c, GenBank accession numbers EF182593–621) and amino acid (data not shown) sequences, with four distinct groups. The reference sequence of strain 86/67 was assigned to group 1. The minimum amino acid divergence between groups was 11%, based on the 315 sites used for the analysis. Group 1 amino acid sequences were all identical to the prototypic strain 86/67 ORF74 sequence, whereas sequences varied from one another by up to 9% in groups 2 and 4.

Variability of ORF74 amino acid sequences was lower than that observed for E1, with variable sites largely concentrated within the N-terminal region and the other three extracellular domains (Fig. 3). Previous analysis of the prototypic ORF74 sequence (strain 86/67) identified several sequence hallmarks of rhodopsin-like 7TMRs, such as the four cysteine residues involved in disulphide-bond formation, and we noted certain differences from the majority of the rhodopsin-like 7TMRs that tend to be observed among other ORF74 7TMRs (Rosenkilde et al., 2005). Despite the high level of variation among EHV-2 ORF74 sequences, the similarities and differences with rhodopsin-like 7TMRs described previously were conserved, albeit with one exception. In common with some other ORF74 homologues, the prototypic EHV-2 ORF74 lacks the Asn and Pro of the NPxxY(x)$_{3,6}$F motif (TM7/C-term), but retains the Tyr and Phe (Rosenkilde et al., 2005). The Tyr and Phe residues of the NPxxY(x)$_{3,6}$F motif are believed to be the most important for signalling (Fritze et al., 2003). However, the ORF74 group 2 sequences encoded Leu rather than Phe (Fig. 2c). As noted above, EHV2 ORF74 is the first example of a naturally occurring 7TMR to exhibit both constitutive and ligand-induced signalling in the absence of the Arg residue in a variant of the ‘DRY’ motif. The novel DTW motif was shown to be functionally important, as reconstitution to ‘DRY’ resulted in a decrease in constitutive signalling (Rosenkilde et al., 2005). The DTW motif was conserved in all isolates analysed in this study, with only synonymous nucleotide substitutions at the third position of the Thr codon for all group 2, 3 and 4 isolates. An analysis of selection pressures again revealed a dominance of negative selection (mean $d_S : d_S = 0.30$), with 30 amino acid sites identified as being under negative selection ($P<0.05$), indicating that ORF74 is under selective pressure to retain functionally important residues. However, against, we found evidence for positive selection at two codon sites (positions 23 and 44 in Fig. 2c).

### Phylogenetic analysis of E6 sequences

PCR amplification of E6 using the original primers was successful for 21 of the EHV-2 isolates. PCR amplicons were cloned and sequenced for 10 of these isolates and sequenced directly from uncloned PCR products for the remaining 11 isolates. PCR products for the remaining 10 EHV-2 isolates were obtained using an alternative forward primer that bound upstream of the E6 ORF, and the uncloned PCR products were sequenced directly. Three of the isolates had mixed prototypic and variant E6 sequence types and were not analysed further. Phylogenetic analysis indicated two distinct groups, comprising 21 sequences related closely (>99% nucleotide identity) to the prototypic sequence (group 1) and seven sequences related closely to each other (group 2) (Fig. 1d; GenBank accession numbers EF182667–EF182694). The group 2 sequences differed by 11–19 amino acid site differences (<6% amino acid divergence overall) from the prototypic sequence, with substitutions concentrated within the N-terminal extracellular region (Fig. 2d).

E6 shows relatively poor sequence conservation with cellular 7TMR, with few of the ‘hallmark’ residues characteristic of rhodopsin-like 7TMR (Gether, 2000; Mirzadegan et al., 2003) (Fig. 2d). A DRC motif at the cytosolic side of TM-3, which is also present in other members of the BILF1
7TMR family, namely PLHV-1, -2 and -3 (Paulsen et al., 2005), is likely to be a functional variant of the 'DRY' motif. The group 2 sequences had conserved features at the N terminus that may be functionally important, namely three potential N-glycosylation sites and a cysteine residue. An analysis of the selection pressures acting on E6 revealed a preponderance of negative selection (mean $d_{NS}$: $d_S = 0.21$), with 11 negatively selected sites and no positively selected sites. These data support the hypothesis that ORF E6 is under selective pressure to retain functionally important residues.

**Comparison of EHV-2 7TMR sequence variability**

E1 had the highest level of amino acid variability (six genogroups, up to 38% difference) and ORF74 sequences were similarly diverse (four genogroups, up to 27% difference). In contrast, E6 sequences were well conserved (two genogroups, <6% difference) and amino acid changes were concentrated in the N terminus. For both E1 and ORF74, variable amino acid positions were more prevalent in some domains than others, as shown in Fig. 3. Sequence variability was highest in the N terminus and other extracellular regions (approx. 50–60% amino acid divergence), whereas the predicted transmembrane and intracellular regions were less variable (approx. 30% amino acid divergence for E1 and 10–20% for ORF74). These differences in variability may have arisen because the transmembrane and intracellular regions of the 7TMR, which are critical for controlling the activation state of the receptor, coupling to G proteins and receptor trafficking, are less able to tolerate mutations than the extracellular regions, without having adverse effects upon function. Alternatively, the extracellular regions may be more variable due to positive selection as a result of antibody responses or interaction with extracellular ligands. However, our analysis of the relative numbers of synonymous and non-synonymous substitutions found little evidence for positive selection in all three EHV-2 7TMRs and did not highlight the extracellular regions as sites of positive selection, suggesting that constraint of tolerated substitutions in the transmembrane and cytoplasmic regions is a more likely explanation. However, it is important to note that comparisons of $d_{NS}$: $d_S$ are an inherently conservative means of detecting positive selection. Furthermore, although the alignments are robust for most of their length, due to the periodic presence of conserved blocks, such as the signature motifs, and very few gaps, there is a high degree of nucleotide divergence and a requirement for insertion of several gaps for the N-terminal regions. Consequently, the assessment of positive versus negative selection for the N termini may not be reliable. The extracellular regions of the 7TMR influence signalling via interaction with ligands (e.g. chemokines). The prototypic sequence of E1 has been shown to be a functional receptor for the chemokine eotaxin (Camarda et al., 1999), and the prototypic EHV-2 ORF74 has been shown to exhibit both constitutive and ligand-modulated signalling activity (Rosenkilde et al., 2005). It will now be of interest to evaluate the structural and functional properties of the 7TMR variants identified in this study, with a view to identifying variability of their cellular expression, constitutive and ligand-induced activity, ligand specificity and ligand-binding affinities.

There appeared to be an inverse relationship between the sequence diversity and the order that the EHV-2 7TMR were first acquired from the host. E6 is believed to be the most anciently acquired 7TMR of EHV-2 and has the lowest sequence similarity to cellular 7TMRs. It is a member of the BILF1 gene family, which includes receptors encoded by rhadinoviruses and lymphocryptoviruses (Beisser et al., 2005; Davis-Poynter & Farrell, 1996; Paulsen et al., 2005), and presumably evolved from a 7TMR homologue that was incorporated into the genome of a progenitor gammaherpesvirus prior to the division into the genera *Rhadinovirus* and *Lymphocryptovirus*. ORF74 is shared with many members of the genus *Rhadinovirus*, but absent from members of the genus *Lymphocryptovirus*, and was presumably acquired after the division of these two genera. E1 has the highest reported similarity to cellular chemokine receptors of any viral 7TMR (Camarda et al., 1999) and
lacks sequence homologues in other gammaherpesviruses, indicating that it was acquired relatively recently.

EHV-2 7TMR gene variability is considerably greater than that identified for clinical isolates of human cytomegalovirus (CMV). US28 (Rasmussen et al., 2003) and UL78 (Michel et al., 2005) were found to be highly conserved, with <2 % amino acid sequence divergence compared with the prototypic (AD169) strain. Similarly, analysis of M33 sequence variability via heteroduplex mobility assay for a panel of field isolates of mouse CMV demonstrated very little sequence variation (Smith et al., 2003). However, considerable 7TMR variation, as a result of gene duplication accompanied by sequence divergence, has been reported for viral 7TMRs encoded by primate CMVs. Sequencing of rhesus and simian CMV has identified a tandemly repeated gene family comprising five genes homologous to US28 of human CMV (Penfold et al., 2004). High levels of amino acid diversity were noted for both gene clusters, with amino acid site difference of 63–85 % for rhesus CMV and 48–79 % for simian CMV. The presence of two copies of E1 in the EHV-2 terminal repeats might enable a single virus strain to encode two different variants of the gene. This possibility was not tested in the current study, but previous analysis of restriction fragment-length polymorphisms for EHV-2 isolates has indicated that the terminal repeats are perfect copies (Browning & Studdert, 1989).

In considering the aetiology of diversification of the EHV-2 E1 and ORF74 7TMR, at least two scenarios may be postulated. First, independent evolution of the 7TMR may have occurred in separated populations of EHV-2, for example in populations of horses or other equids in discrete geographical locations. In recent times, mixing of equid (and hence EHV-2) populations may have occurred so that multiple genotypes of the 7TMR are now present simultaneously at diverse geographical locations. Alternatively, 7TMR may have diversified in the context of a mixture of co-circulating EHV-2 strains. Potentially, the observed tendency of EHV-2 to exist in mixed populations in primary isolates may have favoured sequence diversification via trans-complementation of deleterious mutations, allowing suboptimal gene sequences to persist for longer in the population and hence increasing the time available for acquisition of additional, function-restoring mutations. This latter scenario is analogous to the gene duplication accompanied by diversification seen for the CMV US28 homologues.

Lack of association of genotypes among the EHV-2 genes

Eleven of the 31 EHV-2 isolates were free of mixtures at each of the four gene loci. The distribution of genotypes for these isolates is shown in Table 3. No association between different phylogenetic groups is apparent, with only two sets of isolates (Ger-1/Ger-5 and Ger-6/Ger-7) having the same combinations of genotypes, the phylogenetic trees showing widespread incongruence (Fig. 1). For example, the three isolates with group 2 sequences for E6 differed in their ORF74 (group 1, group 2, group 4), E1 (group 1, group 4, group 6) and gB (group 1, group 2) sequences. Similarly, the four isolates with group 4 sequences for ORF74 differed in their E1 (group 2, group 4, group 6) and E6 (group 1, group 2) sequences. This suggests that different permutations of the various genotypes at the four loci have arisen via recombination, which is expected, given the observed co-infection of horses with different strains. Previously, multiple concurrent infection of individual horses with different EHV-2 strains has been reported (Browning & Studdert, 1987), and in this study, 20 of the 31 EHV-2 isolates appeared to comprise more than one strain on the basis of detection of mixed sequence types for one or more of the four genes analysed. Approximately half of these 20 isolates had mixed genotypes at just one of the four loci investigated. It is possible that the mixed isolates consisted of relatively similar strains that differed at just a few loci. Alternatively, the choice of primers may have excluded additional genogroups present at some loci from the analysis.

Recombination is thought to be an important source of genetic variation in herpesviruses, allowing the combination of genetic features shared by different strains. It has been demonstrated in vitro and/or in vivo between distinguishable strains of several alphaherpesviruses (collectively reviewed by Schyns et al., 2003), has resulted in naturally occurring interspecies recombinants between EHV-1 and -4 (Pagamjav et al., 2005) and has been noted as a major contributor to genetic diversity of clinical isolates of human herpesviruses (Kakoola et al., 2001; Norberg et al., 2004; Poole et al., 1999; Rasmussen et al., 2002; Zong et al., 2002). Recombination of EHV-2 is likely

### Table 3. Comparison of the gene-defined phylogenetic groups of 11 EHV-2 isolates, for which the presence of strain mixtures was not detected

<table>
<thead>
<tr>
<th>Isolate</th>
<th>Phenylogenetic group*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>gB</td>
</tr>
<tr>
<td>Aus-9</td>
<td>Group 1</td>
</tr>
<tr>
<td>Eng-2</td>
<td>Group 1</td>
</tr>
<tr>
<td>Ger-1</td>
<td>Group 1</td>
</tr>
<tr>
<td>Ger-5</td>
<td>Group 1</td>
</tr>
<tr>
<td>Ger-6</td>
<td>Group 1</td>
</tr>
<tr>
<td>Ger-7</td>
<td>Group 1</td>
</tr>
<tr>
<td>Eng-13</td>
<td>Group 1</td>
</tr>
<tr>
<td>Ger-8</td>
<td>Group 1</td>
</tr>
<tr>
<td>Eng-3</td>
<td>Group 1</td>
</tr>
<tr>
<td>Ger-3</td>
<td>Group 2</td>
</tr>
<tr>
<td>Swe-2</td>
<td>Group 3</td>
</tr>
</tbody>
</table>

*Phylogenetic groups assigned from phylogenetic analysis (Fig. 1). Phylogenetic groups containing the prototypic gene sequence (strain 86/67) are indicated in bold.
to have contributed to the considerable biological, antigenic and genomic variability that has been observed among isolates (Browning & Studdert, 1987, 1989; Plummer et al., 1973) and, through frequent reinfection of equids by antigenically different EHV-2 strains, may have pathogenic implications.

ACKNOWLEDGEMENTS

This work was supported by funding from the Horserace Betting Levy Board and the British Council. The authors gratefully acknowledge the provision of virus isolates or purified DNA from Dr S. Raidal (Murdoch University, Western Australia) and Professor S. Belak (National Veterinary Institute, Sweden).

REFERENCES


