Genetic characterization by composite sequence analysis of a new pathogenic field strain of equine infectious anemia virus from the 2006 outbreak in Ireland

Michelle Quinlivan,1 Frank Cook,2 Rachel Kenna,1 John J. Callinan3 and Ann Cullinane1

1Virology Unit, Irish Equine Centre, Johnstown, Naas, Co. Kildare, Ireland
2Gluck Equine Research Centre, Department of Veterinary Science, University of Kentucky, Lexington, KY 40545, USA
3Veterinary Science Centre, University College Dublin, Belfield, Dublin 4, Ireland

Equine infectious anemia virus (EIAV), the causative agent of equine infectious anaemia (EIA), possesses the least-complex genomic organization of any known extant lentivirus. Despite this relative genetic simplicity, all of the complete genomic sequences published to date are derived from just two viruses, namely the North American EIAV_WYOMING (EIAV_WY) and Chinese EIAV_LIAONING (EIAV_LIA) strains. In 2006, an outbreak of EIA occurred in Ireland, apparently as a result of the importation of contaminated horse plasma from Italy and subsequent iatrogenic transmission to foals. This EIA outbreak was characterized by cases of severe, sometimes fatal, disease. To begin to understand the molecular mechanisms underlying this pathogenic phenotype, complete proviral genomic sequences in the form of 12 overlapping PCR-generated fragments were obtained from four of the EIAV-infected animals, including two of the index cases. Sequence analysis of multiple molecular clones produced from each fragment demonstrated the extent of diversity within individual viral genes and permitted construction of consensus whole-genome sequences for each of the four viral isolates. In addition, complete env gene sequences were obtained from 11 animals with differing clinical profiles, despite exposure to a common EIAV source. Although the overall genomic organization of the Irish EIAV isolates was typical of that seen in all other strains, the European viruses possessed $\leq 80\%$ nucleotide sequence identity with either EIAV_WY or EIAV_LIA. Furthermore, phylogenetic analysis suggested that the Irish EIAV isolates developed independently of the North American and Chinese viruses and that they constitute a separate monophyletic group.

INTRODUCTION

Equine infectious anemia virus (EIAV) is a member of the genus Lentivirus in the family Retroviridae and is genetically related to human immunodeficiency virus type 1 (Stephens et al., 1986). However, its genome organization is the least complex of any known extant lentivirus and it is the only member of the genus identified to date that lacks the equivalent of a functional vif gene. EIAV causes a persistent, sometimes fatal, infection in equids and, whilst clinical signs can vary even in horses infected with the same strain of virus (Hammond et al., 2000; Leroux et al., 2001), the disease often progresses through three distinct phases that were first described in 1904 (Valleé & Carré, 1904). The first or acute phase, typically characterized by pyrexia, depression and thrombocytopenia, is followed by a chronic stage lasting 12–24 months where recurring febrile episodes are accompanied by anaemia, weight loss, lethargy and oedema (Issel & Coggins, 1979). If the animal survives, all overt disease signs gradually cease and the animal enters a prolonged third or inapparent stage where it may serve as a source of virus to cohorts (Issel et al., 1982). Immune suppression of inapparent carriers causes an increase in virus replication and sometimes recrudescence of disease (Craigo et al., 2007a; Kono et al., 1976). Because infected animals can control replication of the virus but cannot eliminate it, EIAV serves as a unique model for the study of lentivirus replication under immune control.

The GenBank/EMBL/DDBJ accession numbers for the sequences of EIAV described in this paper are JX480631–JX480841.

Six supplementary figures and a table are available with the online version of this paper.
Natural transmission of EIAV commonly occurs via the mechanical transfer of infected blood by haematophagous insect vectors such as horseflies. However, it can also be spread efficiently via iatrogenic sources such as contaminated veterinary instruments (Hawkins et al., 1973; More et al., 2008; Williams et al., 1981). In 2006, there was an outbreak of equine infectious anaemia (EIA) in Ireland, the source of which appeared to be contaminated plasma imported from Italy (Cappelli et al., 2011; Quinlivan et al., 2007). During this outbreak, most of the horses infected showed disease symptoms (Cullinane et al., 2007; More et al., 2008), demonstrating that this virus, designated EIAV_{IRE}, was pathogenic. The outbreak in Ireland resulted in the death or euthanasia of 35 confirmed cases and movement restrictions for over 1500 horses, with the performance of >90,000 serological tests required for disease investigation, surveillance and business continuity (Brangan et al., 2008). Initial, albeit limited, studies based on the gag gene suggested that EIAV_{IRE} comprised a separate phylogenetic lineage to North American viruses (Capomaccio et al., 2012; Quinlivan et al., 2007).

Unfortunately, with the possible exception of the aforementioned gag sequences, very little is known about genetic diversity in EIAV. Although several full-length genomic sequences have been published, they are all based on either the North American Wyoming isolate (EIAV_{WY}) plus some of its fibroblast cell-adapted variants (EIAV_{PV}, EIAV_{UK} and EIAV_{WSUS}) (Cook et al., 1998; Gradinaru et al., 1981; Hussain et al., 1987; Leroux et al., 1997, 2001; Maury et al., 2003; McGuire et al., 2000; Payne et al., 1994; Perry et al., 1992; Sponseller et al., 2007), or the virulent Chinese Liaoning strain (EIAV_{LIA}) along with various derivative attenuated vaccine strains (Liang et al., 2006; Shen et al., 2006; Tu et al., 2007; Wei et al., 2009). The V70 and V26 viruses described as Japanese EIAV isolates (Zheng et al., 2000) are in fact originally derived from EIAV_{WY} and were imported from the USA for infection experiments in Japanese horses (Kobayashi & Kono, 1967; Kono et al., 1970).

Determining the extent of genetic variation in viral populations of EIAV is essential not only for gaining insight into viral pathogenesis but also for the development of diagnostic reagents and vaccines. The importance of these studies is further highlighted by the fact that EIA currently poses a serious threat to the equine industry in the European Union. In the current study, we report the first known, complete genomic sequences for European EIAV isolates derived from the tissues of four animals infected during the 2006 Irish outbreak. These sequences were based on consensus information generated by the analysis of multiple molecular clones of each PCR-generated fragment and, as such, provide additional data concerning the predominant viral genotypes present during clinical episodes. Genomic analysis demonstrated that, as expected, viruses from each of the four cases were closely related but shared \( \leq 80\% \) nucleotide sequence identity with either EIAV_{WY} or EIAV_{LIA}. Furthermore, nucleotide variation between EIAV_{IRE} and EIAV_{WY} or EIAV_{LIA} was distributed throughout the entire genome, suggesting that the European isolate was not derived from either the North American or the Chinese viruses via recombination events. Indeed, phylogenetic studies suggested that EIAV_{IRE} forms a separate monophyletic group equidistant from EIAV_{WY} and EIAV_{LIA} and as such constitutes a novel field isolate. This study highlights the amount of genetic variation that exists between different strains of EIAV, the extent of which has been grossly underestimated in the past.

### RESULTS AND DISCUSSION

#### Overview of the 2006 EIA outbreak in Ireland and sample history

The sample history is summarized in Table 1. The 2006 EIA outbreak appeared to originate from a contaminated horse plasma product administered intravenously to thoroughbred foals. Representative genomic sequences of the causal virus were determined from assembled portions of EIAV genomes from four infected horses. These four horses, comprising three foals [foals 2–4 (F2–F4)] and one mare [horse 3 (H3)] were in the care of the same veterinary practice. Foal 2 (Table 1) was one of the original recipients of the product, and iatrogenic transmission was considered the most likely route of infection for foals 3 and 4 (More et al., 2008). Horse 3 was the dam of foal 4. The mode of infection of horse 3 was never established, but natural haematophagous insect-mediated transmission was not suspected. The three foals died after a protracted illness, whereas the mare was euthanized the day after first presentation of clinical signs of disease (Cullinane et al., 2007; More et al., 2008). Complete proviral sequences were amplified by PCR in the form of 12 overlapping fragments, R1–R12 (Fig. 1) from all four animals. These were molecularly cloned and the nucleotide sequences of individual clones determined.

The complete env gene was PCR amplified and sequenced from 11 horses that shared a barn in a veterinary hospital. The probable source was an acutely affected mare (horse 1, the dam of foal 2; Table 1) that suffered a severe haemorrhage in the barn involving the loss of several litres of blood. The route of exposure at the hospital remains under investigation. Eight of the horses were euthanized within days of exhibiting clinical signs consistent with EIA and the remaining three were described as subclinical.

#### Full-length proviral genome of EIAV_{IRE}

Consensus full-length proviral sequences based on the predominant clones present were constructed for EIAV isolates from H3 along with F2–F4 (GenBank accession nos JX480631–JX480634). These varied in size, being 8266, 8254, 8275 and 8272 bp, respectively, although all EIAV_{IRE} clones were larger than the proviral genomes of EIAV_{WY}.
(8228 bp) and EIAV\textsubscript{LIA} (8235 bp). Nucleotide sequence identity between the four EIAV\textsubscript{IRE} consensus proviral sequences was $\geq 99.0\%$ (Table 2), strongly supporting the viewpoint that the 2006 EIA outbreak in Ireland originated from a single source. However, nucleotide sequence identity between the four EIAV\textsubscript{IRE} consensus proviral genomes and either EIAV\textsubscript{WY} or EIAV\textsubscript{LIA} was $\leq 80\%$ (Table 2). Interestingly, this level of identity was also very similar to that between EIAV\textsubscript{WY} and EIAV\textsubscript{LIA} (78.9\%), suggesting that similar evolutionary distances exist between each of these complete genomic sequences. Furthermore, a similarity plot performed using EIAV\textsubscript{LIA} as the query sequence demonstrated that none of the four consensus EIAV\textsubscript{IRE} sequences contained extensive regions of genomic homology with the previously characterized virus strains EIAV\textsubscript{LIA} or EIAV\textsubscript{UK} (data not shown). Similar results were obtained using EIAV\textsubscript{WY} as the query sequence, and studies also revealed that there were no extensive regions of genetic identity between EIAV\textsubscript{WY} and EIAV\textsubscript{LIA}. Therefore, it appears unlikely that EIAV\textsubscript{WY} was derived by recombination with EIAV\textsubscript{LIA} (or vice versa) and equally unlikely that recombination involving either EIAV\textsubscript{WY} or EIAV\textsubscript{LIA} was recently involved in the generation of EIAV\textsubscript{IRE}. Results (data not shown) supporting these conclusions were obtained using the DualBrothers recombination detection algorithm (Minin et al., 2005; Suchard et al., 2002, 2003).

Table 1. Sample information
Adapted from Cullinane et al. (2007) and More et al. (2008). Mares with foals that were also sampled are indicated with foal (F) number in parentheses beside their name.

<table>
<thead>
<tr>
<th>Name</th>
<th>Date sampled</th>
<th>Sample</th>
<th>Probable source and date</th>
<th>Proposed incubation or seroconversion period (days)</th>
<th>Clinical signs</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Foal 2</td>
<td>10/05/06</td>
<td>Liver/Spleen</td>
<td>Iatrogenic, 01/03/06</td>
<td>Unknown</td>
<td>Yes</td>
<td>Genome</td>
</tr>
<tr>
<td>Foal 3</td>
<td>26/05/06</td>
<td>Liver/Spleen</td>
<td>Iatrogenic, Unknown</td>
<td>Unknown</td>
<td>Yes</td>
<td>Genome</td>
</tr>
<tr>
<td>Foal 4</td>
<td>12/06/06</td>
<td>Liver/Spleen</td>
<td>Iatrogenic, 03/05/06</td>
<td>Unknown</td>
<td>Yes</td>
<td>Genome</td>
</tr>
<tr>
<td>Horse 3 (F4)</td>
<td>14/07/06</td>
<td>Liver/Spleen</td>
<td>Foal contact, Unknown</td>
<td>Unknown</td>
<td>Yes</td>
<td>Genome</td>
</tr>
<tr>
<td>Foal 5</td>
<td>21/07/06</td>
<td>Plasma</td>
<td>Horse 1(F2), 12/06/06</td>
<td>36</td>
<td>Yes</td>
<td>Env</td>
</tr>
<tr>
<td>Horse 4</td>
<td>21/07/06</td>
<td>Plasma</td>
<td>Horse 1, 12/06/06</td>
<td>38</td>
<td>Yes</td>
<td>Env</td>
</tr>
<tr>
<td>Horse 6</td>
<td>21/07/06</td>
<td>Plasma</td>
<td>Horse 1, 12/06/06</td>
<td>37</td>
<td>Yes</td>
<td>Env</td>
</tr>
<tr>
<td>Horse 10 (F5)</td>
<td>21/07/06</td>
<td>Plasma</td>
<td>Horse 1, 12/06/06</td>
<td>41</td>
<td>Yes</td>
<td>Env</td>
</tr>
<tr>
<td>Horse 11</td>
<td>24/07/06</td>
<td>Plasma</td>
<td>Horse 1, 12/06/06</td>
<td>44</td>
<td>Yes</td>
<td>Env</td>
</tr>
<tr>
<td>Horse 12 (F8)</td>
<td>27/07/06</td>
<td>Plasma</td>
<td>Horse 1, 12/06/06</td>
<td>45</td>
<td>Yes</td>
<td>Env</td>
</tr>
<tr>
<td>Foal 8</td>
<td>01/09/06</td>
<td>Plasma</td>
<td>Dam contact, Unknown</td>
<td>38</td>
<td>Yes</td>
<td>Env</td>
</tr>
<tr>
<td>Horse 15</td>
<td>03/09/06</td>
<td>Plasma</td>
<td>Horse 1, 12/06/06</td>
<td>78</td>
<td>No</td>
<td>Env</td>
</tr>
<tr>
<td>Horse 16</td>
<td>03/09/06</td>
<td>Plasma</td>
<td>Horse 1, 12/06/06</td>
<td>78</td>
<td>No</td>
<td>Env</td>
</tr>
<tr>
<td>Horse 20</td>
<td>23/09/06</td>
<td>Plasma</td>
<td>Horse 1, 12/06/06</td>
<td>100</td>
<td>Yes</td>
<td>Env</td>
</tr>
<tr>
<td>Horse 21</td>
<td>17/11/06</td>
<td>Plasma</td>
<td>Horse 1, 12/06/06</td>
<td>157</td>
<td>No</td>
<td>Env</td>
</tr>
</tbody>
</table>

Table 2. Percentage genome identity between strains of EIAV
Percentage identity was calculated at the nucleotide level. WY, Wyoming strain; DLV, Chinese vaccine strain; LIA Chinese virulent strain; UK North American Wyoming derivative.

<table>
<thead>
<tr>
<th>Strain</th>
<th>WY</th>
<th>DLV</th>
<th>LIA</th>
<th>UK</th>
<th>F2</th>
<th>F3</th>
<th>F4</th>
<th>H3</th>
</tr>
</thead>
<tbody>
<tr>
<td>WY</td>
<td>78.6</td>
<td>78.9</td>
<td>98.6</td>
<td>79.0</td>
<td>78.8</td>
<td>78.8</td>
<td>78.8</td>
<td>78.8</td>
</tr>
<tr>
<td>DLV</td>
<td>78.6</td>
<td>96.8</td>
<td>78.2</td>
<td>79.2</td>
<td>78.9</td>
<td>79.0</td>
<td>79.0</td>
<td>79.0</td>
</tr>
<tr>
<td>LIA</td>
<td>78.9</td>
<td>96.8</td>
<td>78.4</td>
<td>80.0</td>
<td>79.7</td>
<td>79.7</td>
<td>79.7</td>
<td>79.7</td>
</tr>
<tr>
<td>UK</td>
<td>98.6</td>
<td>78.2</td>
<td>78.4</td>
<td>78.4</td>
<td>78.4</td>
<td>78.4</td>
<td>78.4</td>
<td>78.3</td>
</tr>
<tr>
<td>F2</td>
<td>79.0</td>
<td>79.2</td>
<td>80.0</td>
<td>78.6</td>
<td>99.4</td>
<td>99.1</td>
<td>99.0</td>
<td>99.0</td>
</tr>
<tr>
<td>F3</td>
<td>78.8</td>
<td>78.9</td>
<td>79.7</td>
<td>78.4</td>
<td>99.4</td>
<td>99.3</td>
<td>99.1</td>
<td>99.1</td>
</tr>
<tr>
<td>F4</td>
<td>78.8</td>
<td>79.0</td>
<td>79.7</td>
<td>78.4</td>
<td>99.1</td>
<td>99.3</td>
<td>99.2</td>
<td>99.2</td>
</tr>
<tr>
<td>H3</td>
<td>78.8</td>
<td>79.0</td>
<td>79.7</td>
<td>78.3</td>
<td>99.0</td>
<td>99.1</td>
<td>99.2</td>
<td>99.2</td>
</tr>
</tbody>
</table>

Fig. 1. Map of cloned regions R1–R12 in relation to the EIAV genome. The location of the cloned and sequenced regions R1–R12 is shown with reference to a schematic diagram of the EIAV genome. Details of the primers used for generation of R1–R12 are given in Table S1, available in JGV Online. LTR, Long terminal repeat.
Phylogenetic analysis of complete EIAV proviral sequences

The complete proviral sequences of EIAVWY and EIAVLIA along with the genomes of selected derivative strains were aligned using CLUSTAL W against the four consensus EIAVIRE proviruses and phylogenetic analysis conducted by the neighbour-joining method with bootstrap values determined over 1000 iterations. Evolutionary distances were estimated by maximum composite likelihood. The resultant unrooted phylogenetic tree (Fig. 2) comprised three monophyletic groups in which the EIAVwY-like, EIAVLIA-like and four EIAVIRE consensus sequences were separated by similar evolutionary distances. This, along with the SimPlot analysis, strongly suggested that EIAVWY, EIAVLIA and EIAVIRE have evolved independently of one another since diverging from a common ancestor. Consequently, EIAVIRE represents a novel field virus strain that can be regarded as pathogenic in that it possesses the ability to cause disease in a majority (>70%) of recipients (Cullinane et al., 2007; More et al., 2008). The phylogenetic analysis reported here extends previously published results based on just 313 bp of the viral gag gene sequence, which also found the strain responsible for the 2006 EIA outbreak to be associated with a different monophyletic group from either EIAVwY or EIAVLIA (Capomaccio et al., 2012).

Sequence variation within molecular clones

The majority of molecular clones generated from proviral DNA isolated from F2–F4 along with H3 possessed intact ORFs and only these were incorporated into subsequent alignment, phylogenetic and percentage identity analyses. However, some clones were identified as containing substitutions, insertions or deletions resulting in the introduction of stop codons or frameshift mutations. The incidence of predicted premature stop codons in each major structural gene was: 13.3% in gag (one clone of F3, two clones of F2 and one clone of H3), 23.5% in pol (one clone of F3 and three clones of H3) and 12% in env (three clones of F4). Furthermore, as might be expected, some differences in predicted amino acid sequences were observed between the molecular clones. Unfortunately, without more information about the replicative capabilities of the proviruses from which they were derived, it was not possible even to begin to evaluate the role that these infrequent differences have within the viral quasispecies.

Given that the F2–F4 had prolonged illness compared with H3 who was euthanized the day after exhibiting clinical signs, the diversity of EIAV populations within each horse was calculated to establish whether there was a relationship with the duration of illness. The identity among clones for the Gag and Pol polyproteins or precursors was quite similar for each of the animals (98–100 and 97.7–100%, respectively), although the results may have been biased by the relatively low number of clones sequenced for each animal coupled with the possibility that the PCR conditions were not optimal for amplification of minor populations. However, the foals had much greater diversity in their env genes compared with the mare. H3 had 99.5 and 99.4% identity between clones for gp90 and gp45

---

Fig. 2. Phylogenetic analysis of the EIAV genomes. A phylogenetic tree of aligned complete proviral sequences was constructed by the neighbour-joining method with bootstrap values determined over 1000 iterations. Branch lengths are proportional to the distance existing between the sequences. Bar, Nucleotide substitutions per site. The Genbank accession numbers of the strains are: LIA, AF327877; DLV, AF327878; DLV18-8, HM141923; DV117, HM141912; DV35-20, HM141911; UK, AF016316; WY, AF033820; WSU5, AF247394; V70, AB008196; V26, AB008197.
respectively, whereas the foals had 92.5–95.5 and 94.9–97.3% identity, respectively, for the glycoproteins. As mutations in gp90 are responsible primarily for the viral escape mechanism of EIAV during febrile episodes of disease (Lichtenstein et al., 1996), these differences among the animals would suggest that the viral population in H3 did not have sufficient time to diversify.

The data were analysed further to generate a consensus genome for each of the four animals, and EIAV protein sequences were then compared. The percentage identities of these consensus sequences among the four animals for Gag, Pol, Tat, S2 and Rev were 99.2–100, 99, 100, 98.6–100 and 97–98.5%, respectively. However, greater variation was observed in the predicted Env proteins, with sequence identities ranging from 97.8% between F2 and F3 to 95.3% between F2 and H3.

**Characterization of the EIAV<sub>IRE</sub> LTR and structural genes**

The proviral EIAV<sub>IRE</sub> genome was identical to all known strains of this equine lentivirus, containing three major structural genes (gag, pol and env) and three ancillary genes (tat, s2 and rev) bounded by LTRs.

**LTR.** The LTR consensus sequence for EIAV<sub>IRE</sub> proviruses isolated from H3 and F2–F4 was 313 bp, which is shorter than that of EIAV<sub>WY</sub> (323 bp), EIAVLIA (316 bp), an attenuated vaccine strain (DLV) derived from EIAVLIA (334 bp) and a fibroblast cell culture-adapted variant (19-2) of EIAV<sub>WY</sub> (322 bp) (Malmquist et al., 1973; Payne et al., 1994). In addition to size differences, the consensus EIAV<sub>IRE</sub> LTR sequences shared just 77.3–77.6 and 78.1–78.4% nucleotide sequence identity with the virulent EIAV<sub>WY</sub> and EIAVLIA strains, respectively. However, the majority of this variation occurred in the U3 region upstream from previously identified transcription factor-binding motifs and in U5 (Fig. 3). In fact, there was extensive conservation of the MDBP motif, TATA box, stem–loop comprising the TAR element (Carvalho & Derse, 1991; Hoffman & White, 1995) and poly(A) addition site (Fig. 3). Furthermore, with the exception of 19-2, which contains PEA-2 sites believed to be associated with fibroblast adaptation, all other viruses, including EIAV<sub>IRE</sub>, possessed three PU.1 (GTTCCE)-binding motifs consistent with equine macrophage tropism (Maury et al., 2000).

**Gag.** There was good general agreement between the EIAV<sub>IRE</sub> gag gene sequence published in 2007 (Quinlivan et al., 2007) and data from the EIAV<sub>IRE</sub> clones (Fig. S1). Furthermore, there was ≥98% nucleotide sequence identity between the consensus EIAV<sub>IRE</sub> gag genes and those obtained from Italian horses (designated ITA1, ITA2 and ITA3), strongly supporting the belief that the 2006 EIA outbreaks in Italy and Ireland were derived from the same contaminated horse plasma product (Cappelli et al., 2011; Cullinane et al., 2007; Quinlivan et al., 2007).

Although there were numerous differences in predicted amino acid sequence between the gag genes of EIAV<sub>IRE</sub> EIAV<sub>WY</sub> and EIAVLIA, comparative analysis demonstrated extensive conservation of all previously identified structural and functional motifs in p15, p26, p11 and p9. These motifs included a potential leucine-rich-type nuclear export signal (NES) in p15, specific residues within the major homology region (MHR) in p26 that are believed to play an essential role in virus particle assembly (Grund

---

Fig. 3. Transcription factor-binding motifs within the LTR. WY, Wyoming strain; 19-2, attenuated cell culture-adapted variant of WY; LIA, Chinese strain; DLV, attenuated vaccine strain of LIA; F2, foal 2. U3, R and US are indicated by arrows, and the +1 start G and poly(A) addition sites are underlined. The transactivation response (TAR) element stem–loop is 31 nt downstream of the +1 start. The U3 hypervariable enhancer is indicated in consensus with bold text; the methylated DNA-binding protein site (MDBP), PU.1 GTTCCE (x3) and TATA box by white boxes, the CAAT virulence marker (WY x2, 19-2, F2) by grey shaded boxes, and the ACCGCA PEA-2 (19-2 only) by grey shading.
et al., 1994), the two CX2CX4HX4C zinc-binding domains that in all EIAV strains analysed to date are separated by a characteristically short five-residue basic linker (RAPKV) (Amodeo et al., 2006) and the tyrosine-proline-aspartic acid-leucine (YPDL) late assembly domain in p9 (Chen et al., 2005; Li et al., 2002; Puffer et al., 1997, 1998).

**Pol.** Alignment of pol gene sequences based on the predicted five Pol proteins [protease (PR), reverse transcriptase (RT), RNase H, dUTPase (DU) and integrase (IN)] demonstrated that the nucleotide sequence identity between EIAVIRE-, EIAVWY- and EIAVLIA-derived viruses (Fig. S2a–e) was 89.4–94.2 % in PR, IN and RT, 85.2–87.8 % in RNaseH and 85.1–85.8 % in DU.

Whilst sequence identity among different retroviral PRs is low, the residues forming the substrate-binding sites are relatively conserved (Eizert et al., 2008; Weber et al., 1993). Therefore, it was perhaps not surprising to find that these residues were conserved among all EIAV strains examined, with the exception of a generally highly conservative isoleucine-to-valine substitution at aa 54 (I54V) seen in EIAVWY.

RNase H contains amino acid residues that are highly conserved among all retroviruses including four acidic residues, D443, E478, D498 and D549, essential for enzyme catalysis and metal ion binding (Davies et al., 1991; Doolittle et al., 1989). These residues, along with a glycine-rich motif (aa 534–545), were all present in the EIAVIRE, EIAVWY and EIAVLIA virus strains.

Previously identified conserved residues and motifs in the dUTPase were preserved among all virus strains except for some individual EIAVIRE clones. These included the five highly conserved motifs present in all known DU proteins (N67, G69, Y75, Q80 and I82; McGeoch, 1990) important in binding the uracil ring (Dauter et al., 1999), along with D72 involved in deoxyribose interactions (Steagall et al., 1995).

The zinc-binding HHCC motif (H12, H16, C40, C43) of IN, as well as residues that form part of the hydrophobic core (I5, A8, L22, A33 and I36), are strictly conserved among immunodeficiency viruses (Eijkelboom et al., 1997). In addition, there are three central catalytic core (D64, D116 and E152) so-called invariant residues (Engelman & Craigie, 1992), all of which were present in all the viruses analysed in this study (Fig. S2e).

**Env.** The env gene encodes the surface unit (gp90) and transmembrane (gp45) glycoproteins. These proteins are highly variable in infected animals, with mutations in gp90 being responsible for the generation of neutralizing-antibody escape mutants (Leroux et al., 2001; Lichtenstein et al., 1996; Zheng et al., 1997). Therefore, as might be expected, there was significant variation between EIAVIRE and the other strains analysed, with nucleotide sequence identities of 63–68 % in gp90 and 65–67 % in gp45 (Fig. S3).

In longitudinal studies on horses or ponies infected with a single EIAV strain (EIAVpv), variation over time within gp90 was restricted mostly to eight 'hypervariable regions' (Leroux et al., 1997, 2001; Zheng et al., 1997). Interestingly, amino acid variation in gp90 sequences amplified from different animals infected during the 2006 outbreak in Ireland was also restricted mainly to these same regions (Fig. 4, Fig. S3a). In contrast, amino acid variation in gp90 between geographically distinct EIAV isolates was not confined to these locations and instead appeared to be distributed throughout the entire coding sequence (Fig. 4, Fig. S3a). This supports a previous observation (Craigo et al., 2009) concerning geographically distinct isolates and has significant implications for the design of potential immunogens, as in the EIAV system gp90 is a primary determinant of vaccine efficacy (Craigo et al., 2007b).

Despite considerable variation in gp90 sequences, it was interesting to note that cysteine residues were extensively conserved among EIAV strains and, although the positions of potential N-linked glycosylation sites varied, the number of sites was similar. In predicted gp45 sequences, variation among virus strains was particularly evident in the cytoplasmic domain, with EIAVIRE having just 58.7 and 57.4 % amino acid identity with EIAVLIA and EIAVWY, respectively. In comparison, EIAVIRE differed from EIAVLIA and EIAVWY by 78.5 and 79.1 % in predicted amino acid identity of the extracellular domain of gp45, and all four potential N-linked glycosylation sites were conserved.

**S1 Tat.** The EIAVIRE Tat consensus sequence had 78.4 % amino acid identity with EIAVLIA and 74.5 % with EIAVWY. The 15 aa core domain responsible for Tat function was completely conserved among all strains examined. This is not surprising, given its homology with tat genes of other lentiviruses including simian, bovine and human immunodeficiency viruses (Carroll et al., 1991). The majority of Tat variation among different strains was localized to the N-terminal 10 aa and C-terminal 14 aa (Fig. S4). The N terminus has been shown to be functionally dispensable for Tat; however, the C terminus contains the basic and C-terminal amino acids responsible for TAR binding (Darse & Newbold, 1993). Importantly, the C-terminal leucine is conserved among all strains and clones.

**S2.** S2 was not extensively conserved among EIAV strains, as exemplified by the fact that EIAVIRE had only 58.6 and 50.7 % sequence identity with EIAVLIA and EIAVWY, respectively. The S2 protein is an important virulence determinant believed to interact with cellular proteins and enhance pro-inflammatory cytokine responses, although the exact mode of action is unknown (Covaleda et al., 2010a, b; Li et al., 2000). Previously, potential functional motifs including a nucleoporin motif, an SH3-binding motif, a nuclear localization signal, myristylation signals, an N-linked glycosylation site and a number of phosphorylation...
sites have been defined for EIAV (Yoon et al., 2000). The nucleoceptor motif, one myristylation signal and some casein kinase II and protein kinase C sites were conserved in S2 sequences between the different EIAV isolates (Fig. 5), suggesting functional significance. Although an SH3 domain was present in all EIAVWY and EIAV_{LIA} isolates, a P28C substitution and a P25L substitution eliminated it from EIAV_{IRE} and EIAV_{PA}, respectively. In addition, the putative C-terminal nuclear localization signal (RRKQETKK; Li et al., 2000) was modified to EETQDTKK and EETQDTQK in EIAV_{IRE} and EIAV_{PA}, respectively, suggesting functional significance. Although an SH3 domain was present in all EIAVWY and EIAV_{LIA} isolates, a P28C substitution and a P25L substitution eliminated it from EIAV_{IRE} and EIAV_{PA}, respectively. In addition, the putative C-terminal nuclear localization signal (RRKQETKK; Li et al., 2000) was modified to EETQDTKK and EETQDTQK in EIAV_{IRE} and EIAV_{PA}, respectively, suggesting functional significance.

The second exon of EIAV_{IRE} Rev possessed significant variation compared with the other strains, with sequence identities of 72.6 and 68.2 % with EIAV_{LIA} and EIAV_{WY}, respectively. Although the sequence of the N-terminal polar effector domain (Mancuso et al., 1994) varies among strains and includes some apparently non-conservative substitutions, there was complete conservation of the hydrophobic residues (L36, L45, L49 and W41) that mutagenesis studies have shown are absolutely required for the nuclear export function of Rev (Belshan et al., 1998; Harris et al., 1998).

In Rev derived from EIAV_{WY}, RNA-binding is believed to be dependent on the 20 C-terminal residues along with the RRDRW (aa 76–80) and an ERLE (aa 93–96) motif (Harris et al., 1998; Lee et al., 2006). However, equivalent positions in most EIAV_{IRE} clones comprised KERW and EQLE, respectively, and in EIAV_{LIA}-derived viruses, aa 76–80 were occupied by RRDSW. This suggests that aa 76 and 77 should be positively charged, whilst a negative charge is required at aa 78 and hydrophobic W is mandated at aa 80. Similarly, the presence of negatively charged E at aa 93 and 96 plus hydrophobic L at aa 95 is important, whilst some variability is permissible at aa 94. These conclusions are supported

Fig. 4. Pattern of amino acid variation within EIAV gp90. Amino acid variation in the SU envelope glycoprotein between subjects infected during the 2006 outbreak in Ireland [represented by foals 2–5 (F2–F5) and horses 3, 6 and 12 (H3, H6 and H12)] was generally restricted to the hypervariable (V) regions (V3–V7 are indicated by boxes) identified by analysis of gp90 sequences obtained during sequential febrile episodes (II–V) from a pony (567) infected with the EIAV_{WY} (PV) strain (Leroux et al. 2001). In contrast, amino acid variation in gp90 between geographically distinct isolates including those from China (LIA) and Pennsylvania, USA (PA) was not confined to the hypervariable regions.
experimentally by the observation that, whilst substitution of L95 reduced Rev interactions with viral RNA, there was
near wild-type binding where R94 had been replaced by A (Lee et al., 2006).

Nuclear localization of EIAV<sub>WY</sub> Rev–green fluorescent protein fusion proteins is abolished by the introduction of substitutions within the C-terminal KRRRK motif (Harris et al., 1998), and work by Lee et al. (2006) has confirmed that this cluster of basic residues functions as a nuclear localization signal. The importance of this motif was further highlighted by the fact that it was conserved in all but one of the EIAV<sub>IRE</sub> clones.

In summary, complete genomic sequence information was obtained for the pathogenic EIAV strain EIAV<sub>IRE</sub> responsible for the 2006 EIA outbreak in Ireland. This European isolate possessed &lt;80 % nucleotide sequence identity with the previously described North American EIAV<sub>WY</sub> and Asian EIAV<sub>LIA</sub> strains and in phylogenetic terms probably constitutes a separate monophyletic group. However, despite significant genetic variation, almost all of the structural and functional motifs that have been identified in proteins encoded by EIAV<sub>WY</sub>-derived viruses were either conserved or contained highly conservative amino acid substitutions. This study adds important new information, particularly in terms of EIAV pol and env gene sequences, as these have not been characterized extensively among geographically distinct viral isolates. It is interesting to note that, whilst genetic variation in env gene sequences encoding gp90 during the course of infection in horses was mainly restricted to eight hypervariable regions, variation among different EIAV strains was distributed throughout the molecule, an observation that may further complicate the design of vaccines against this lentivirus.

METHODS

Samples. Post-mortem tissues (liver or spleen) from three foals (F2–4) and one mare (H3) infected with EIAV<sub>IRE</sub> in 2006 (Table 1) were used to generate EIAV<sub>IRE</sub> full-length genome sequence data. Approximately 0.1 mg tissue was homogenized in 5 ml minimum essential medium supplemented with 2 % FCS, 100 U penicillin ml<sup>−1</sup>
and 100 µg streptomycin ml<sup>−1</sup> (all from Gibco, Invitrogen). Plasma samples were collected from 11 EIAV<sub>IRE</sub>-infected animals showing clinical signs of disease or just before they were euthanized. DNA extracted from these samples was subjected to nested PCR and directly sequenced for env gene analysis. DNA was extracted from 200 µl homogenate supernatant or plasma using a QIAamp DNA Mini kit (Qiagen) according to the manufacturer’s instructions.

Primers. Oligonucleotide primers (Table S1) were used to amplify the EIAV<sub>IRE</sub> genome in the form of 12 overlapping fragments (R1–R12). Similarly, for sequencing of the env gene, overlapping regions EI1–EI8 of the gene were chosen. Outer PCRs for five regions were initially employed to provide product for primers encompassing these env regions as follows: E01 (EI1), R9 (EI2–EI4), R10 (E5 and E6), R11 (EI7) and EO2 (EI8). Primers were designed and analysed based on EIAV genome sequences deposited in GenBank using the online application Primer3 (Rozem & Skeatsky, 2000).

PCR, cloning and sequencing. A 50 µl PCR contained 0.125 U AmpliTaq DNA polymerase µl<sup>−1</sup>, 1 × Buffer and 200 µM each dNTP (all from Applied Biosystems), 0.5 µM each primer and 5 µl DNA. Initial denaturation was carried out at 95 °C for 5 min, followed by amplification with 30 cycles of 95 °C for 15 s, 45–65 °C (primer dependent) for 30 s and 72 °C for 1 min per kb of ampiclon, and a final elongation at 72 °C for 5 min (G-Storm; Gene Technologies). For nested PCRs, 1 µl outer product was used as PCR template and reactions were purified using a QIAquick PCR Purification kit or a QIAquick Gel Extraction kit (Qiagen). Amplicons were then sequenced with the primers used to generate them. Amplicons generated for tissue samples were similarly purified and cloned into vector pCR-4-TOPO (Invitrogen) according to the manufacturer’s recommendations. The inserts of positive TOPO clones were sequenced with M13 primers. At least three clones per animal per genome region were sequenced giving a minimum of 12 sequences per region. All sequencing was performed by Qiagen Sequencing Services (Germany).

Genomic analysis. The deduced amino acid sequences of EIAV<sub>IRE</sub> proteins from the nucleotide sequences were translated using ExpASY software (Gasteiger et al., 2003), and sequence alignments along with percentage identities were produced using CLUSTAL W (Larkin et al., 2007). EIAV<sub>IRE</sub> sequence data were compared with global strains of the virus. In alignments, the first digit describing the sequence source represents the region cloned (1–12) and this was followed by F2–F4 or H3 depending on the tissue source. Clones of a particular region from one animal were then further differentiated by the letters a–f depending on the number of clones, giving a unique sequence. For example, 7F2c was the third unique sequence for a region 7 clone from F2. Phylogenetic analysis was conducted by the neighbour-joining method with bootstrap values determined over 1000 iterations (Drummond et al., 2012) and evolutionary distances estimated by maximum composite likelihood (Tamura & Nei, 1993; Tamura et al., 2004).

ACKNOWLEDGEMENTS

M. Q. was funded by an IRCSET EMPOWER postdoctoral fellowship and experimental work was additionally funded by the Department of Agriculture under the National Development Plan and the Clay Fellowship Fund, Gluck Equine Research Center.

REFERENCES


