The integration of a macrophage-adapted live vaccine strain of equine infectious anaemia virus (EIAV) in the horse genome

Qiang Liu,1 Xue-Feng Wang,1 Cheng Du,1 Yue-Zhi Lin,1 Jian Ma,1 Yu-Hong Wang,2 Jian-Hua Zhou1 and Xiaojun Wang1,*

Abstract
Integration is an important feature of retroviruses and retrovirus-based therapeutic transfection vectors. The non-primate lentivirus equine infectious anaemia virus (EIAV) primarily targets macrophages/monocytes in vivo. Investigation of the integration features of EIAVDLV121 strains, which are adapted to donkey monocyte-derived macrophages (MDMs), is of great interest. In this study, we analysed the integration features of EIAVDLV121 in equine MDMs during in vitro infection. Our previously published integration sites (IS) for EIAVFDDV13 in fetal equine dermal (FED) cells were also analysed in parallel as references. Sequencing of the host genomic regions flanking the viral IS showed that reference sequence (RefSeq) genes were preferentially targeted for integration by EIAVDLV121. Introns, AT-rich regions, long interspersed nuclear elements (LINEs) and DNA transposons were also predominantly biased toward viral insertion, which is consistent with EIAVFDDV13 integration into the horse genome in FED cells. In addition, the most significantly enriched Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways, specifically gag junctions for EIAVDLV121 and tight junctions for EIAVFDDV13, are regulators of metabolic function, which is consistent with the common bioprocesses, specifically cell cycle and chromosome/DNA organization, identified by gene ontology (GO) analysis. Our results demonstrate that EIAV integration occurs in regions that harbour structural and topological features of local chromatin in both macrophages and fibroblasts. Our data on EIAV will facilitate further understanding of lentivirus infection and the development of safer and more effective gene therapy vectors.

INTRODUCTION
The integration site (IS) distribution of retroviruses in the host genome is a crucial factor for viral transcription and therefore heavily influences viral replication efficacy, which determines viral pathogenicity levels and host immune responses. Viral genome insertion into the host chromosome is not random, and each retrovirus genus has its own pattern of IS distribution. For example, lentiviruses prefer to integrate into transcription units [1], while gamma-retroviruses and foamy viruses prefer active enhancers and CpG islands, respectively [2–4]. The insertion of retroviral DNA copies depends on numerous factors related to the virus and target cells. Several mechanisms have been proposed to explain the favourable or unfavourable distribution of IS, and the tethering mechanism has been experimentally confirmed, particularly for lentiviruses and gamma-retroviruses. Studies of human immunodeficiency virus-1 (HIV-1) revealed that viral integration is mediated by cellular factors, such as lens epithelium-derived growth factor (LEDGF/p75), and viral proteins, such as integrase (IN) [5, 6]. In the gamma-retrovirus model, a cellular factor, the bromo- and extra-terminal domain (BET) protein, covalently binds mouse leukaemia virus (MLV) IN and mediates viral integration near transcription start sites (TSS) [7, 8].

Equine infectious anaemia virus (EIAV), family Retroviridae, genus Lentivirus, is an important pathogen of equids and is characterized by persistent infection and high genomic variability, which strongly impact the prevention and treatment of lentivirus infection [9]. EIAV primarily targets macrophages/monocytes [10] but also replicates in endothelial cells in vivo [11] and in certain types of fibroblasts in vitro [12]. EIAV-infected horses serve as models of acquired
immunodeficiency syndrome (AIDS) that conform to a variety of experimental parameters, including many of the hallmarks of HIV-1 infection and pathogenesis [13]. Recent publications have highlighted the role of tissue macrophages as critical contributors to HIV-1 persistence during long-term suppressive antiretroviral therapy (ART) [14]. Shen et al. previously attenuated a Chinese pathogenic EIAV strain by successively adapting the virus in cultivated primary donkey monocyte-derived macrophages (MDMs) and fetal donkey dermal (FDD) cells [12, 15]. The specific tropisms of different EIAV strains may lead to differences in other virology characteristics.

The characteristics of EIAV integration were first investigated by Hacker et al. in 2006 by infecting the human embryonic kidney cell line HEK293T with a VSV-G-envelope pseudovirus [16]. In our previous study, the integration of an FDD-adapted strain EIAV_FDDV13 in fetal equine dermal (FED) cells was not randomly distributed but rather selected certain patterns of genomic structures in the horse genome [17]. These results reveal similarities in EIAV IS selection between the horse and human genomes. However, these studies focused on the EIAV IS in non-immune cells during acute infection using either a replication-incompetent EIAV pseudovirus or a live EIAV strain. EIAV replicates differently in macrophages versus other target cells such as endothelial cells and fibroblasts, resulting in different host cell responses [9, 18]. Integration of the provirus into the host genome represents a central event in the replication of all lentiviruses, including EIAV. Determining the diversity of host genes containing integrants in EIAV-infected cells may yield useful information regarding pathways by which the host affects viral replication.

Thus, it is of great interest to determine the characteristics of EIAV ISs in the horse genome, particularly in its primary target cells, macrophages, and to determine whether there are any differences compared to ISs in fibroblasts. In the study presented here, we characterized IS selection of the macrophage-tropic EIAV_DLV121 in the equine MDM genome and analysed insertion preferences and the functions of annotated genes harbouring an IS.

**RESULTS**

The sequences surrounding EIAV IS are AT-rich and weakly palindromic

The horse genomic sequences flanking the IS were amplified from a gDNA library of MDMs infected with EIAV_DLV121 for 24 h and digested with the restriction enzyme Dra I or Ssp I. The 5’ LTR and 3’ LTR are the two termini of the lentivirus genome. To determine whether there was a base preference in regions surrounding the IS, 40-nt sequences flanking the 5’ LTR and 3’ LTR of EIAV were obtained and referred to as position −40 to −1 and position +1 to +40 on the X-axis, respectively. Viral integration occurred at positions between −1 and 1 on the X-axis. A total of 525 sequences were aligned (Table 1), and the frequencies of each base at each position within positions −40 to +40 of the flanking sequences were tabulated using the WebLogo program. The length of each base at each position represents the proportion of A/T/G/C nucleotides at each of the 80 positions in the flanking sequences, based on total IS numbers. The same method was used to determine the flanking sequences of EIAV IS in the genome of equine fibroblasts (EIAV_FDDV13) [17]. This analysis identified the specific nucleotide composition, which differed from that in the MRC group (Fig. 1c), within 40 bases on each side of an IS. Based on these results, EIAV strains favoured integration in A/T-rich consensus sequences (Fig. 1a, b), similar to HIV-1 in the human genome [19].

Genome-wide analyses of host–virus junctions have revealed that retroviruses have a weak preference for specific primary DNA sequences in host genomes, which are unique for each retrovirus genus and commonly have a palindromic structure. For example, HIV-1 and MLV favour the consensus sequences 5’-GTWAC-3’ and 5’-VTAB-3’, respectively, when labelled using Standard International Union of Biochemistry Codes (B=C, G or T; V=A, C or G; W=A or T) [20–22]. By referring to studies examining HIV-1 integration, the 5’ and 3’ genomic sequences surrounding IS were measured. The consensus sequence in EIAV_DLV121 at these positions was 5’-VTWAC-3’, which was highly similar to that of the fibroblast-tropic EIAV_FDDV13 (Fig. 1d). In addition, a palindromic sequence was found near the IS. The palindromic sequence was centred on the offset −3 and was

<table>
<thead>
<tr>
<th>Group</th>
<th>Virus or vector</th>
<th>Cell type</th>
<th>Number of integration sites</th>
<th>Accession numbers (GenBank)</th>
<th>Source of reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>EIAV_DLV121†</td>
<td>MDM†</td>
<td>525§</td>
<td>KS297078–KS297602§</td>
<td>This study</td>
</tr>
<tr>
<td>B</td>
<td>EIAV_FDDV13ǁ</td>
<td>FED¶</td>
<td>477§</td>
<td>KO454223–KO454699§</td>
<td>[15]</td>
</tr>
</tbody>
</table>

*EIAV donkey leukocyte-attenuated vaccine strain.
†Equine monocyte-derived macrophage.
‡Number of sequences flanking the integration sites in each group.
§Accession number of sequences in GenBank.
ǁEIAV fetal equine dermal cell-adapted vaccine strain.
¶Fetal equine dermal cell.
observed in the EIAV<sub>DLV121</sub> and EIAV<sub>FDDV13</sub>-integrated horse genome when the most frequently appearing nucleotides at each site were lined up based on the base frequency at each site surrounding the IS (Fig. 1d). These results indicated a weak palindromic structure near the IS in the horse genome.

**Fig. 1.** Base preferences within 40 bp regions flanking IS in the horse genome. The consensus sequences 40 bp upstream (−40) and 40 bp downstream (+40) (relative to the 5’ LTR end) of each IS were plotted using the WebLogo program (http://weblogo.berkeley.edu/). Consensus sequences were determined for the EIAV<sub>DLV121</sub> dataset (a), EIAV<sub>FDDV13</sub> dataset (b) and MRC dataset (c). The Y-axis represents the percentages of each nucleotide at the indicated sites (1 indicates 100 %), and the X-axis shows the sites upstream (−40) and downstream (+40) of each IS. (d) A weak palindromic sequence centred on base −3 was identified for EIAV<sub>DLV121</sub> and EIAV<sub>FDDV13</sub>.

<table>
<thead>
<tr>
<th>Virus</th>
<th>Most frequent base at integration site</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>−6</td>
</tr>
<tr>
<td>EIAV&lt;sub&gt;DLV121&lt;/sub&gt;</td>
<td>A</td>
</tr>
<tr>
<td>EIAV&lt;sub&gt;FDDV13&lt;/sub&gt;</td>
<td>A</td>
</tr>
</tbody>
</table>
RefSeq genes and introns are favoured targets for EIAV integration

Next, the frequency of EIAV<sub>DLV121</sub> integration in coding genes was determined based on horse reference sequence (RefSeq) genes as a criterion (release 2007) and compared to the frequency of random IS within RefSeq genes. The results demonstrated that 275 of 525 (52.4%) integration events were within defined RefSeq genes. In comparison, 257 of 477 (53.9%) EIAV<sub>FDDV13</sub> integration events were located within RefSeq genes, which was not statistically significant (P>0.05). As the control, 34.8% (3481/10 000) of matched random controls were identified within RefSeq genes (P<0.01) (Table 2). A total of 94.5% (260/275) of genes contained one EIAV<sub>DLV121</sub> IS, and the remainder contained two or three independent IS per gene, e.g. two IS in the centrosomal protein 290 (CEP290) gene and three IS in the tetratricopeptide repeat protein 8 (TTC8) gene (Table 3).

Furthermore, the orientations of EIAV integration relative to the direction of the hosted gene were examined. The percentage of sense-direction EIAV<sub>DLV121</sub> insertions in RefSeq genes was 48.7% (134/275), which was not significantly different from that of the antisense-direction insertions (P>0.05). This orientation-independent insertion was also detected for EIAV<sub>FDDV13</sub>, with 55.3% sense-direction insertions (142/257) (Table 2). In addition, the frequency of integration in RefSeq gene introns was analysed. EIAV<sub>DLV121</sub> appeared to preferentially integrate within introns, with a frequency as high as 97.5% (268/275). Similarly, 97.7% (251/257) of EIAV<sub>FDDV13</sub> insertions were within introns (Table 2). The intron-favouring nature of EIAV is consistent with HIV-1 integration into the human genome [1]. Based on these results, EIAV integration into both cultivated primary equine macrophages and fibroblasts demonstrated a similar strong preference for RefSeq genes and introns.

EIAV integration prefers LINE and DNA transposons

Integrated proviral DNA affects the expression of surrounding genes, either by reducing gene expression via disruptive insertion or enhancing gene expression via a promoter effect, and the distance from IS to TSS in a host gene mediates the effects of insertion [23]. Like HIV-1, EIAV displays a preferential bias for insertion into genes. In this study, distances from the viral IS to upstream or downstream of the TSS closest to the IS of EIAV<sub>DLV121</sub> were detected and divided into three intervals: ±0–5, ±5–10 and ±10–25 kb. The frequency of each interval was calculated as a fold increase compared to the corresponding MRC data. A gradually increasing preference for viral integration (1.49-, 2.07- and 2.13-fold, respectively) was observed as the distance from the TSS to the IS increased. The distance from the TSS to the EIAV<sub>FDDV13</sub> IS also appeared to be similar but had a relatively weaker correlation with integration events. In contrast, intergenic regions were less likely to harbour integrations than the MRC group for both EIAV<sub>DLV121</sub> and EIAV<sub>FDDV13</sub> (0.69- and 0.70-fold increase, respectively) (Fig. 2).

Next, the potential preference of EIAV<sub>DLV121</sub> for integration into repetitive elements in the horse genome was assessed. Several repetitive elements, including LINEs, SINEs, DNA transposons and LTR retrotransposons, were identified as regions targeted for the integration of both EIAV<sub>DLV121</sub> and EIAV<sub>FDDV13</sub>. Specifically, compared to the proportion of LINEs in MRC, which were generated as described by Mitchell et al. and Berry et al. [21, 24], EIAV<sub>DLV121</sub> strongly favoured integration within LINE, with a fold enrichment of 1.80 (Fig. 2). Notably, although DNA transposons represent a low proportion (approximately 2%) of the horse genome, EIAV<sub>DLV121</sub> integration displayed a strong preference for this element with a fold increase of 2.19. In contrast, SINEs and LTR retrotransposons appeared to be disfavoured targets for EIAV<sub>DLV121</sub> integration (0.56- and 0.60-fold compared, respectively, to MRC) (Fig. 2). The other class of repetitive elements, including satellites and low-complexity repeats, accounted for a lower proportion within the horse genome, and no differences were observed in terms of the numbers of integration events within these repetitive elements. Again, EIAV<sub>FDDV13</sub> had a similar preference to that of EIAV<sub>DLV121</sub> for these repetitive elements. In addition, we also provide Tables S1 and S2 (available in the online Supplementary Material) to show actual insertion numbers.

Based on these results, both EIAV<sub>DLV121</sub> and EIAV<sub>FDDV13</sub> possessed a decreasing preference for integration in regions close to the TSS, as well as a preference for LINEs and DNA transposons.

Functional characterization of genes harbouring an IS

The aforementioned sequence analysis results revealed that both EIAV<sub>DLV121</sub> and EIAV<sub>FDDV13</sub> preferred to target

<table>
<thead>
<tr>
<th>Table 2. Comparison between groups for integration sites within genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>RefSeq genes represent reference sequence genes. P1: comparison to the random group using a chi-squared test. P2: comparison to the EIAV&lt;sub&gt;DLV121&lt;/sub&gt; group using a chi-squared test. NA, not applicable; ND, indicates not done; NS, not significant.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Random No. (%)</th>
<th>EIAV&lt;sub&gt;DLV121&lt;/sub&gt; No. (%)</th>
<th>EIAV&lt;sub&gt;FDDV13&lt;/sub&gt; No. (%)</th>
<th>P1</th>
<th>P2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>10 000 (100.0)</td>
<td>525</td>
<td>477</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>In RefSeq</td>
<td>3481 (34.8)</td>
<td>275 (52.4)</td>
<td>257 (53.9)</td>
<td>&lt;0.01</td>
<td>NS</td>
</tr>
<tr>
<td>Sense</td>
<td>ND</td>
<td>134 (48.7)</td>
<td>142 (55.3)</td>
<td>ND</td>
<td>NS</td>
</tr>
<tr>
<td>Introns</td>
<td>ND</td>
<td>268 (97.5)</td>
<td>251 (97.7)</td>
<td>ND</td>
<td>NS</td>
</tr>
</tbody>
</table>
the most frequently enriched KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways were related to gap junctions, which harboured more enriched KEGG pathways with weaker significances were identified, among which the most enriched pathway was associated with tight junctions ($P=5.2 \log_2$), followed by ubiquitin-mediated proteolysis, adherens junctions and purine metabolism ($P$ values ranged from 3.2 to 2.9 $\log_2$).

GO analysis showed that in terms of biological processes, mitosis and chromosome segregation were the top two functional areas of genes for EIAV$_{DLV121}$ IS and had much more significant $P$ values than the biological processes listed in Fig. 4a. In contrast, genes inserted by EIAV$_{FDDV13}$ involved biological processes such as chromosome organization, DNA metabolism, the cell cycle and positive regulation of the macromolecule process, of which the second and third processes ranked fourth and third, respectively, among the functions of EIAV$_{FDDV13}$-integrated genes (Fig. 4b). Based on these results, the two most significantly enriched KEGG pathways, specifically gap junctions and tight junctions, which harboured insertions by both EIAV$_{DLV121}$ and EIAV$_{FDDV13}$, are regulators of metabolic function, which is consistent with the common bioprocesses such as the cell cycle and chromosome DNA organization identified by GO analysis.

**DISCUSSION**

Combined with previously published data, our studies demonstrate that two EIAV strains, the macrophage-tropic EIAV$_{DLV121}$ and fibroblast-tropic EIAV$_{FDDV13}$, share the same preferences for integration in the host genomes of favoured cells. In addition, similar enriched KEGG pathways and bioprocesses related to the genes hosting EIAV integrations were observed for these two viruses. Based on these results, EIAV integration occurs in genomic regions that harbour structural and topological features of local chromatin in both macrophages (terminally differentiated immune cells) and fibroblasts (dividing, non-immune cells). The above data facilitate further understanding of lentivirus infection.

*In vitro* experimental controls revealed that the use of restriction endonucleases to digest host gDNA during viral IS cloning did not introduce detectable bias [1]. More importantly, various cloning and analytic strategies for examining viral IS in the host genome, including the fragmentation of gDNA using restriction enzymes or physical approaches [25], amplification with inverse PCR (I-PCR) [26] or LM-PCR [27] and high-throughput sequencing [28], resulted in similar conclusions regarding integration characteristics in different types of host cell [29, 30]. We used two different endonucleases, Ssp I and Dra I, to fragment gDNA in a previous study investigating EIAV integration in equine fibroblasts, and observed similar results following the analysis of IS data from each group of clones prepared using these endonucleases. In addition, this strategy did not reveal statistically significant differences when controls were treated using the MRC method [15]. For consistency, Ssp I and Dra I were used to fragment MDM gDNA in this study.

---

**Table 3. Host genes with 2 or 3 independent insertions of EIAV**

<table>
<thead>
<tr>
<th>Number group</th>
<th>EIAV$_{DLV121}$</th>
<th>EIAV$_{FDDV13}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2*</td>
<td>CEP290†</td>
<td>BBS9</td>
</tr>
<tr>
<td></td>
<td>CEP152</td>
<td>BRCA1</td>
</tr>
<tr>
<td></td>
<td>DYM</td>
<td>CAMSAP2</td>
</tr>
<tr>
<td></td>
<td>GPD2</td>
<td>RC3H1</td>
</tr>
<tr>
<td></td>
<td>HSPBAP1</td>
<td>SPIDR</td>
</tr>
<tr>
<td></td>
<td>LRBA</td>
<td>SGCD</td>
</tr>
<tr>
<td></td>
<td>MOGAT1</td>
<td>THADA</td>
</tr>
<tr>
<td></td>
<td>NTSDC1</td>
<td>UBN2</td>
</tr>
<tr>
<td></td>
<td>SRFBP1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>POLQ</td>
<td></td>
</tr>
<tr>
<td></td>
<td>UTRN</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>TTC8</td>
<td>HMCN1</td>
</tr>
<tr>
<td></td>
<td>PDS5B</td>
<td>FAM227B</td>
</tr>
</tbody>
</table>

*Numbers of integrants.
†Genes are indicated by their gene symbol.

RefSeq genes. Thus, the specific pathway involvement and ontology of the genes harbouring viral integration events were analysed using gene annotation enrichment analysis and gene ontology (GO) annotation tools from DAVID, respectively. Gene annotation enrichment analysis revealed that the most frequently enriched KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways were related to gap junctions ($P=9.2 \log_2$), followed by progesterone-mediated oocyte maturation, insulin signalling, GnRH signalling and melanogenesis pathways ($P$ values ranged from 6.7 to 6.0 $\log_2$) for the IS of EIAV$_{DLV121}$ (Fig. 3a). For EIAV$_{FDDV13}$, fewer KEGG pathways with weaker significances were identified, among which the most enriched pathway was associated with tight junctions ($P=5.2 \log_2$), followed by ubiquitin-mediated proteolysis, adherens junctions and purine metabolism ($P$ values ranged from 3.2 to 2.9 $\log_2$) (Fig. 3b).
Fig. 3. KEGG pathway analysis. KEGG pathways enriched among the coding genes hosting an IS using the DAVID program (https://david.ncifcrf.gov/) for the EIAV<sub>DLV121</sub> dataset (a) and the EIAV<sub>FDDV13</sub> dataset (b) are indicated. GnRH, gonadotropin-releasing hormone; MAPK, mitogen-activated protein kinase; mTOR, mechanistic or mammalian target of rapamycin; VEGF, vascular endothelial growth factor.
EIAV integration events are non-random in the horse genome. More than 97% were preferentially integrated into introns in host genes. Similar integration patterns were reported for HIV-1 [31]. A possible reason for this phenomenon is the high percentage of sequence length represented by introns. For instance, in the horse and human genomes,
AT-rich regions are uniformly favoured for EIAV and HIV-1 integration. These regions are gene-poor and have small exons, which encode an average of only 50 codons, and many sprawling genes are separated by long introns, some of which exceed 10 kb [32]. More interestingly, two EIAV\textsubscript{DLV121} proviral DNA copies were integrated into different introns of the CEP290 gene, and three IS were in the TTC8 gene. There are two possible explanations for multiple insertions in a single gene: 1) CEP290 and TTC8, which respectively contain 93 202 and 56 925 nt, are longer than the median and mean values (14 000 and 27 000 nt, respectively) of human protein-coding genes [32]. 2) Lentiviral integration is favoured in active transcription units, which allows efficient viral gene expression after integration [27, 28, 33]. Both CEP290 and TTC8 are centrosomal proteins, which are involved in cell cycle regulation and are essential for the growth of many cell types [34, 35]. We speculate that these genes are relatively more active. Indeed, our unpublished data demonstrated that CEP290 and TTC8 underwent high levels of transcription in EIAV\textsubscript{DLV121}\textsuperscript{-}infected MDMs.

As shown in several studies, the physical location of viral IS in the host genome tends to significantly influence viral transcription. For instance, specific characteristics of genomic regions, e.g. intergenic regions, gene deserts, heterochromatin and centromeres, are able to downregulate the expression of integrated viral genomes and therefore favour latency [27, 36]. The tendency of EIAV\textsubscript{DLV121} to insert within LINEs is consistent with the tendency of EIAV\textsubscript{FDDV13} [17] but differs from HIV-1, which preferentially integrates within SINEs in the human genome [1]. This difference occurs because although LINE content in the horse genome is similar to that in the human genome (22.6 vs 19.7 %), SINEs comprise only 7.3 % of the former, which is approximately half of the SINE content in the latter (13.4 %) [37]. Genomic regions close to LINEs are usually expressed at low levels or exist in a resting state [38, 39]. After the acute stage of clinical equine infectious anemia, most horses enter a chronic stage characterized by a progressive decrease in the frequency and intensity of symptom onset over approximately one year, after which the animals enter an inapparent phase with no overt clinical symptoms [9]. However, whether or to what extent this behaviour in LINE content is associated with virological behaviour requires further investigation.

According to our results, the most significant enrichment of KEGG was observed in gap junctions for EIAV\textsubscript{DLV121} and tight junctions for EIAV\textsubscript{FDDV13}. Gap junctions contain intercellular channels through which small molecules, including ions, nucleotides and amino acids, directly transfer between the cytosolic compartments of adjacent cells [40], whereas tight junctions, in the form of multifunctional complexes, constitute a permeability barrier in the paracellular space and mediate functions involved in epithelial cell polarity, cell proliferation and gene transcription [41]. Notably, these two significantly enriched KEGG pathways are regulators of metabolic function, which is consistent with the common bioprocesses, specifically the cell cycle and chromosome/DNA organization, identified by GO analysis. Therefore, EIAV-integrated genes in either MDMs or FEDs are uniformly categorized into specific functions, such as cell cycle, although the differential and cell type-specific expression of genes between MDM and FED cells should be considered. Host genes with integrants in virus-infected cells may be involved in viral replication. A recent study reported that IS detected in certain clonal expansion CD4 + T cells of HIV-1-infected patients after prolonged combination antiretroviral therapy (cART) were enriched in genes controlling cell cycle and survival [25, 42]. These data suggest that HIV-1 integration into specific genes promotes the proliferation of HIV-1-infected CD4 + T cells and slows viral clearance during cART, supporting a role for host genes with integrants contributing to persistent viral infection. To strengthen this hypothesis, further proof of IS in this region is needed, which requires the generation and analysis of additional IS. In addition, comparative studies of gene expression in these cells would facilitate assessment of the role of these integrations.

While MLV infects only dividing cells and has a strong preference for TSS, particularly in enhancer regions, lentiviruses such as HIV-1 and EIAV also preferentially differentiate cells like macrophages and preferentially target RefSeq genes by transferring the pre-integration complex (PIC) through nuclear pores. Since transcriptional active sites in chromatin are associated with nuclear transport machinery, PIC might integrate into transcriptionally active sites when passing through nuclear pores, where the structure of chromatin is usually open [43]. Differences in IS selection between lentiviruses and MLV are attributable to interactions between viral IN and host cellular factors. The strong bias in MLV integration is influenced by the interaction of MLV IN with BET proteins [7, 8]. In contrast, lentivirus-encoded IN tightly binds to certain cellular proteins, the best characterized of which is LEDGF/p75. This host protein is important for HIV-1 infectivity and integration targeting [5]. In addition, HIV-1 IN residue Ser119 was found to affect viral integration [44], and HIV-1 capsid (CA) protein also affects IS selection [45, 46]. Our previously published data indicated that IN and CA in EIAV\textsubscript{DLV121} and EIAV\textsubscript{FDDV13} shared highly conserved structural features [47]. These viral determinants of target-site selection provide reasonable evidence for the similar integration preferences of different EIAV strains in the horse genome.

Lentivirus-based vectors, such as HIV-1 and EIAV vectors, are widely used to correct target gene function in genetic diseases. The non-primate lentivirus EIAV-derived delivery vector has demonstrated safety and efficacy for gene therapy, particularly for ocular diseases, e.g. adult NHP photoreceptors [48] and neonatal rodents [49]. The integration characteristics demonstrated in this study and our previous study [17], such as the preference for coding genes and LINEs in the genome of equine MDMs (terminally
differentiated immune cells) and FEDs (dividing, non-immune cells), facilitate a better understanding of EIAV-based vectors, which are promising alternative viral vectors in gene therapy.

**METHODS**

**Viral strains**

The macrophage-adapted EIAV strain EIAV\textsubscript{DLV121} was developed by successively passaging an EIAV virulent strain, EIAV\textsubscript{DLV12}, in primary donkey MDMs. EIAV\textsubscript{DLV17} was derived from the pathogenic strain EIAV\textsubscript{LN40} via 116 passages in donkeys, and was highly pathogenic in both horses and donkeys [15]. This strain was provided by the Division of Equine Infectious Diseases and Lentivirus Research of Harbin Veterinary Research Institute (HVRI). EIAV\textsubscript{FDDV13} is a fibroblast-tropic strain derived from 13 passages of EIAV\textsubscript{DLV21} in FDD. Both EIAV\textsubscript{DLV121} and EIAV\textsubscript{FDDV13} are avirulent in horses and donkeys [12].

**Cell culture and viral infection**

Equine MDM preparation and cultivation were performed as previously described [50]. Briefly, 200 ml of fresh heparinized whole blood was carefully placed in a 250 ml sterile culture flask and maintained at room temperature for 30–60 min. The plasma supernatant was collected in 50 ml sterile centrifuge tubes (Costar, Corning, USA) and centrifuged at 1000 r.p.m. at room temperature for 10 min. After discarding the supernatant, the cells remaining in the plasma were resuspended and washed twice with Ca\textsuperscript{2+}- and Mg\textsuperscript{2+}-free PBS, cultivated at a density of 1 \times 10^5 cells/ml in RPMI 1640 medium (Gibco, Carlsbad, CA, USA) containing 30 % heat-inactivated calf serum, 20 % heat-inactivated horse serum (Gibco), 100 IU penicillin and 100 µg ml\textsuperscript{-1} streptomycin in 25 cm\textsuperscript{2} culture flasks (Corning) and incubated overnight at 37 °C in 5 % CO\textsubscript{2}. After incubation for 24 h, non-adherent cells were removed via two PBS washes. The remaining adherent cells were detached with 0.9 % NaCl and resuspended at a density of 5 \times 10^6 cells/ml in 20 ml of RPMI 1640, placed into 75 cm\textsuperscript{2} flasks and then incubated overnight at 37 °C in 5 % CO\textsubscript{2}. The adherent cells differentiated into macrophages after 24 h. EIAV\textsubscript{DLV121} was inoculated at a multiplicity of infection (MOI) of 10 for 1 h to facilitate viral adsorption and infection, and 20 ml of culture medium was added to the flask for further incubation. As shown in a previous study, lentiviral cDNA enters the nucleus within 24 h post-infection, and viral particles are produced and released within 48 h post-infection [51, 52]. To maximize provirus formation but minimize repeat infections after integration, MDMs were harvested 24 h post-infection and stored at −80 °C until further use.

**Cloning of host–virus junction sequences**

The genomic DNA (gDNA) of infected cells was extracted using a QIAamp DNA Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer’s instructions and electrophoresed on a 0.6 % agarose gel at 4 °C to eliminate the effects of unintegrated forms. Gel slices containing DNA bands larger than 8.3 kb were collected and extracted from the gel using a QIAquick PCR Purification Kit (Qiagen). Linker-mediated polymerase chain reaction (LM-PCR) was performed to clone the horse sequences flanking IS, as previously described [53]. Specifically, gDNA was digested with both the blunt-end restriction endonuclease Dra I and Sp I (NEB, USA). Fragments from each endonuclease-digested DNA were ligated with a specific adaptor, which was formed by annealing two complementary single-stranded links of Linker 1 (5’-CTAATACGACTATAGGCGTCGAGCGCCCGCCGGAGGT–3’) and Linker 2 (5’-[Phosp]-ACCTCCCC-[Amino]–3’). Fragments containing the EIAV LTR region and the flanking host sequence were amplified from gDNA libraries digested with Dra I or Sp I using nested PCR. Primary PCR was performed using primers matching the viral LTR (EIAV514 : 5’-CTTTTGAGACCCCTTGTACCGTACCTTT-3’) and the adaptor (APF1 : 5’-GGATCCTAATACGAGTCACTATAGCGC-3’), and secondary PCR was performed using the viral LTR primers (EIAV359 : 5’-GTTTCAGCAGGCCAGGTTCT-3’) and the adaptor primers (APF2 : 5’-CTATAAGCGCTCAGCGC-3’). The product sizes derived from secondary PCR ranged from 0.1 to 5.0 kb and were examined by performing agarose gel electrophoresis. DNA fragments 0.1–2.5 kb and 2.5–5.0 kb in size were separately extracted from gel slices, purified using a QIAquick Kit and ligated with the pMD18-T vector (Takara, Shiga, Japan). The ligated products were transformed into DH5α competent cells (Takara). Positive clones were selected and sequenced.

**Sequence analysis**

The BLAT program (http://genome.ucsc.edu/) was applied to map sequence reads to the horse genome (UCSC, assembled in Sep 2007). Other programs for analysis were used as previously described [17]. An IS was determined to be authentic if it met the following criteria: 1) more than 95 % identity with a horse genome sequence and no more than a single horse genetic locus; 2) contained both the upstream adaptor sequence and the downstream viral LTR; 3) ‘TGTGAG,’ which is the 5’-terminal sequence of viral LTR, was detected as the initial sequence; and 4) had a length of at least 20 bp, which was the lower limit of the BLAT program.

A horse genome dataset was downloaded from the genome browser website of the University of California Santa Cruz (UCSC). Integration was considered to occur in a gene only if it was located in the RefSeq mRNA sequences on the chromosome as determined using the BioMart program (http://asia.ensembl.org). Furthermore, the insertion direction was determined based on its orientation in the RefSeq mRNA sequences. Additional information, including the distance relative to transcription initiation or termination sites and the frequency of exons or introns hosting an IS, was obtained from BioMart when an IS occurred in RefSeq mRNA sequences. The nucleotide sequences around the IS were obtained from UCSC. The base frequency of each site within 40 bp of the IS was analysed using the WebLogo.
program (http://weblog.berkeley.edu/). Integration was determined to occur in a repetitive element only if it was located within the boundaries of one of the repetitive elements, including long interspersed nuclear elements (LINEs), short interspersed nuclear elements (SINEs), DNA transposons and LTR retrotransposons. These repetitive elements were determined by the RepeatMasker program (http://www.repeatmasker.org/cgi-bin/AnnotationRequest). In addition, functional annotation and classification of the genes harbouring an IS were analysed using Annotation, Visualization and Integrated Discovery (DAVID) tools (http://david.abcc.ncifcrf.gov).

Statistical analysis

To control for possible sampling and analytical biases, in addition to using two different restriction endonucleases to digest the gDNA, an approach called Matched Random Control (MRC) was applied to generate a set of 10,000 random insertion sites [21, 24].

First, a restriction map of the horse genome was constructed for the restriction enzyme used for library construction, specifically Dra I or Ssp I in this study. The corresponding sequences of this set of deduced fragments (approximately 1–1000 bp) were collected as Dataset A. Dataset B was based on the distance of each actual IS (525 for EIAV_PDLV121 in this study and 477 for EIAV_FDDV13 in a previous study) from the restriction enzyme site (Dra I or Ssp I). Each of the fragments in Dataset B was matched to the ten fragments with the closest lengths in Dataset A. A Chi-square test was performed to compare the integration patterns of Dataset A and Dataset B. Differences of P<0.05 and P≤0.01 were considered statistically significant or highly significant, respectively. Statistical analysis was carried out using the Statistics Analysis System (SAS) 9.2, and all images were drawn using GraphPad Prism 5 (GraphPad Software, Inc., CA, USA).

Nucleotide sequence accession numbers

Genomic nucleotide sequences flanking the EIAV IS were deposited in GenBank, and the accession numbers for the Genomic nucleotide sequences flanking the EIAV IS were Nucleotide sequence accession numbers.

GraphPad Prism 5 (GraphPad Software, Inc., CA, USA).

References


Funding information

This study was supported by grants from the National Natural Science Foundation of China (31070809 to JHZ, 31302066 to XFW).

Conflicts of interest

The authors declare that there are no conflicts of interest.

Downloaded from www.microbiologyresearch.org by
IP: 54.70.40.11
On: Wed, 12 Dec 2018 14:59:17


