Article

A Unique Evolution of the *S2* Gene of Equine Infectious Anemia Virus in Hosts Correlated with Particular Infection Statuses

Xue-Feng Wang [1], Shuai Wang [1], Qiang Liu [1], Yue-Zhi Lin [1], Cheng Du [1], Yan-Dong Tang [1], Lei Na [1], Xiaojun Wang [1,*] and Jian-Hua Zhou [1,2,*]

[1] State Key Laboratory of Veterinary Biotechnology, Harbin Veterinary Research Institute of Chinese Academy of Agricultural Sciences, Harbin 150001, China;
E-Mails: xuefengwang1982@126.com (X.-F.W.); 8111781252@163.com (S.W.); liuqiangtriumph@163.com (Q.L.); sndhr@163.com (Y.-Z.L.); chengdu1981@163.com (C.D.); tangyandong2008@163.com (Y.-D.T.); nl2zy@163.com (L.N.)

[2] Harbin Pharmaceutical Group Biovaccine Company, Harbin 150069, China

* Authors to whom correspondence should be addressed; E-Mails: xjw@hvri.ac.cn (X.W.); jianhua_uc@126.com (J.-H.Z.); Tel.: +86-189-4606-6285 (X.W.); +86-137-9608-5512 (J.-H.Z.); Fax: +86-0451-5199-7166 (X.W.).

Abstract: Equine infectious anemia virus (EIAV) is a member of the *Lentivirus* genus in the *Retroviridae* family that exhibits a genomic structure similar to that of HIV-1. The S2 accessory proteins play important roles in viral replication *in vivo* and in viral pathogenicity; however, studies on *S2* evolution *in vivo* are limited. This study analyzed the evolutionary characteristics of the *S2* gene of a pathogenic EIAV strain, $EIAV_{LN40}$, in four experimentally infected horses. The results demonstrated that 14.7% (10 of 68 residues) of the stable amino acid mutations occurred longitudinally in S2 during a 150-day infection period. Further analysis revealed that six of the ten mutated residues were positively selected during the infection. Alignment and phylogenetic analyses showed that the *S2* gene sequences of viruses isolated from the infected horses at the early stage of $EIAV_{LN40}$ infection were highly homologous and similar to the vaccine-specific sequence. The *S2* gene variants isolated from the febrile episodes and late phase of infection became homologous to the *S2* gene sequence of the inoculating $EIAV_{LN40}$ strain. Our results

indicate that the *S2* gene evolves in diversity and divergence *in vivo* in different stages of EIAV infection and that this evolution correlates with the pathogenicity of the virus.

Keywords: equine infectious anemia virus; EIAV; *S2* gene; evolution; diversity

1. Introduction

Equine infectious anemia virus (EIAV) belongs to the *Lentivirus* genus in the *Retroviridae* family. This virus predominantly infects *Equidae*, and the onset of disease typically occurs within three to four weeks post-infection. Viral replication could be controlled in most infected horses after approximately one year of infection, and these horses become asymptomatic carriers for life. Carriers can transmit the virus through blood, and the clinical symptoms of equine infectious anemia (EIA) might recur in carriers when the immune system is naturally or experimentally suppressed [1]. These phenomena indicate that infected horses can control EIAV replication without eliminating the virus. Therefore, EIAV has been recognized as a unique model for the study of *lentivirus* replication under immune control [2].

The EIAV genome is the least complex among lentiviruses. In addition to the structural proteins encoded by the *gag*, *pol*, and *env* genes, EIAV encodes three small proteins, which include the Tat and Rev regulatory proteins and the S2 accessory protein [3]. The *S2* gene encodes a protein of 65 or 68 amino acid residues [4,5]. The S2 protein is unique to EIAV, and its function is unclear. The S2 protein potentially interacts with Gag, without incorporation into EIAV particles. It has been reported that the S2 protein interacts with the cellular protein OS-9 [6]. The S2 protein has no sequence homology with other *lentivirus* proteins. There are a few common predicted motifs that are shared in the amino acid sequences of EIAV S2 and the primate *lentivirus* Nef, including a myristoylation site, an SH3 domain binding motif and a casein kinase 2 phosphorylation site. Therefore, EIAV S2 might have a function similar to Nef of HIV-1 or SIV [7]. The HIV-1 Nef protein is an important factor associated with the virulence and infectious efficacy of the virus. Studies have demonstrated that the primate lentivirus *Nef* gene has numerous polymorphisms and presents different variants in different disease states *in vivo* [8,9].

The EIAV *S2* gene showed remarkable variation after successive passages in donkey monocyte-derived macrophages (dMDM), an attenuating process that greatly reduced the pathogenicity of the precursor EIAV$_{LN40}$ [5]. Deletion of *S2* in EIAV does not affect viral replication *in vitro* [10]. Additionally, the information regarding *S2* evolution in infected horses is limited. Two studies that used the EIAV$_{PV}$ infectious clone-derived virus to infect ponies found that the EIAV *S2* gene was highly conserved in infected animals *in vivo* [11]. The deletion of *S2* significantly reduced viral replication *in vivo* and attenuated the virulence of the virus [11]. Our previous studies on the Chinese EIAV attenuated vaccine revealed that the *S2* gene was one of most varied regions compared with its parental virulent strain [5]. The reversal of four major stable mutations in the *S2* gene of an infectious clone of the vaccine EIAV strain (EIAV$_{FDDV3-8}$) in the sequence of the virulent strain resulted in a modest but statistically significant increase in the plasma viral load and clinical signs (increased body temperature and decreased platelet number) [12]. This result suggests that the S2 protein is crucial; however, it is not the only factor

affecting EIAV pathogenicity *in vivo*. In this study, we longitudinally analyzed the sequence of the *S2* gene during the first five months of $EIAV_{LN40}$ infection in horses. The results demonstrated that the *S2* gene is highly diverse *in vivo* and significantly evolves under positive selection pressure.

2. Materials and Methods

2.1. EIAV Strains

$EIAV_{LN40}$ is a virulent strain derived from a field strain isolated in Liaoning Province in China by 16 successive passages in horses. $EIAV_{LN40}$ is highly lethal in horses; however, it causes no apparent clinical EIA symptoms in donkeys. An attenuated, live EIAV vaccine, $EIAV_{DLV121}$, was developed by passaging $EIAV_{LN40}$ in donkey monocyte-derived macrophages (dMDMs). A fibroblast-adapted derivate, $EIAV_{FDDV13}$, was subsequently developed by 13 passages of $EIAV_{DLV121}$ in fetal donkey dermal (FDD) cells. $EIAV_{DLV121}$ and $EIAV_{FDDV13}$ proliferate in horses and donkeys without causing EIA symptoms and induce protective immunity against experimental and natural infections with pathogenic EIAV strains. These three EIAV strains were stocked at the Harbin Veterinary Research Institute of the Chinese Academy of Agricultural Sciences [5].

2.2. Horses Experimentally Infected with EIAV

Four horses were infected with EIAV in a previous study [13]. Briefly, two male 4-year-old horses (#25 and #26) were subcutaneously injected in the neck with 1×10^6 $TCID_{50}$ of $EIAV_{LN40}$; these animals died from typical EIA at day 28 and day 30 days post infection (dpi), respectively. Another two male 4-year-old horses (#4 and #10) were subcutaneously injected in the neck with 10-fold less $EIAV_{LN40}$ (1×10^5 $TCID_{50}$). No EIA clinical symptoms were observed for 5 months post infection in both horses. Samples of 100 ml of peripheral blood were taken at the time points indicated in Figure S1.

The use of horses and the related experimental protocols in this study were approved by the Institutional Animal Care and Use Committee (IACUC) of the Harbin Veterinary Research Institute (HVRI), Chinese Academy of Agricultural Sciences. At the end of the experiment, or when severe disease-associated symptoms resulting in distress appeared, the horses infected with pathogenic EIAV strains were euthanized by an intravenous injection of pelltobarbitalum natricum (100 mg/kg body weight, dissolved in saline) in the jugular vein by veterinarians according to protocols approved by the IACUC of HVRI.

2.3. Analysis of S2 Gene Variation

Virions were collected from the plasma of the blood samples from the horses infected with $EIAV_{LN40}$ or culture supernatants of equine monocyte-derived macrophages (eMDMs) infected with $EIAV_{DLV121}$ and fetal donkey dermal (FDD) cells infected with $EIVA_{FDDV13}$ by centrifugation. The viral genomic RNA was extracted from the pellets using a QIAamp Viral RNA Mini Kit (QIAGEN, Hilden, Germany). The full-length *S2* gene cDNA fragments were amplified by three independent nested reverse transcription (RT)-PCR experiments, as previously described [13]. The PCR products were excised from 0.8% agarose gels and ligated into the pMD18-T vector (TaKaRa, Dalian, China). Eight to 25

positive recombinant clones of each sample were sequenced. The alignment and phylogenetic analysis of the nucleotide sequences were performed with the SeqMan II tool of the Lasergene DNAStar program (version 6.0, DNAStar Inc., Madison, WI, USA, 2001) and the Molecular Evolutionary Genetics Analysis (MEGA) program (version 5.0, Center for Evolutionary Functional Genomics Biodesign Institute, Arizona State University , Tempe, AZ, USA, 2011). The phylogenetic tree was constructed with nucleotide sequences using the bootstrap neighbor-joining method. The bootstrap values were calculated from 1000 replicates of the alignment. Statistical analysis of sequence variations in S2 sequences was performed by SAS (Statistical Analysis System) (version 9.2, SAS Institute Inc, SAS Campus Drive, Cary, NC, USA, 2008).

2.4. Detection of Selection Pressures

The detection of the selection pressures of the evolutionary processes of the EIAV *S2* gene was performed using the codeml method of the PAML software package (version 4.3, Department of Biology, University College London, London, UK, 2009) [14]. The ratio ω of non-synonymous to synonymous substitutions (d_N/d_S) is an important indicator of selection pressure at the codon level, and $\omega = 1$, <1 and >1 reflect neutral, purifying and positive selection, respectively. The Model M0 (ratio of one), M1a (nearly neutral), M2b (positive selection), M3 (discrete), M7 (β) and M8 (β and ω) of the codeml program are typically applied in the detection of codon-specific positive selection in a virus gene codon [14–16].

3. Results

3.1. The S2 Gene Highly Varied among the Isolates of Experimentally Infected Horses and in Vitro Attenuated Strains

Four horses were experimentally infected with a pathogenic EIAV strain, EIAV$_{LN40}$, in a previous study of the evolution of EIAV *in vivo* [13]. Two of these horses (#25 and #26) were inoculated with 1×10^6 TCID$_{50}$ EIAV$_{LN40}$ and died of acute EIA at 28 and 30 days post infection (dpi), respectively. Another two horses (#4 and #10) were infected with 10-fold less of the virus, which resulted in sub-clinical infections (Figure S1).

To investigate the evolution of the *S2* gene in long-term infection in the host, the *S2* gene fragment was amplified by RT-PCR from 18 plasma samples of four horses infected with EIAV$_{LN40}$. The PCR product was not obtained from Sample 4-1, which indicates Time Point 1 of horse #4. In addition, the *S2* genes of two attenuated EIAV strains, *i.e.*, eMDM-adapted EIAV$_{DLV121}$ and FDD-adapted EIAV$_{FDDV13}$, were amplified from virions collected from the culture supernatants of the infected cells. Eight to 26 *S2* nucleotide sequences were determined from randomly selected PCR clones of each sample, and a total of 387 nucleotide sequences (HQ008940-HQ009261, HQ223289-HQ223335) were obtained. The S2 nucleotide and deduced amino acid sequences of EIAV$_{LN40}$ isolated from experimentally infected horses were compared with the S2 sequence of EIAV$_{LN40}$ before inoculation, as well as those of two attenuated strains, EIAV$_{DLV121}$ and EIAV$_{FDDV13}$.

Table 1. Comparison of nucleotide and amino acid genetic distances of the *S2* genes of viruses isolated from different samples infected with EIAV$_{LN40}$ and the vaccine strains EIAV$_{DLV121}$ and EIAV$_{FDDV13}$.

Sample		EIAV$_{LN40}$		EIAV$_{DLV121}$		EIAV$_{FDDV13}$	
name	dpi	nt	aa	nt	aa	nt	aa
DLV121	V	2.32(0.97–4.53)	6.84(2.99–14.20)				
FDDV13	V	2.31(1.47–4.00)	6.65(4.51–10.86)	1.17(0–3.49)	3.21(0–7.64)		
25-1	14	2.75(2.47–4.00)	7.02(6.06–10.86)	1.81(0.49–3.50)	5.22(1.48–10.86)	1.78(0.98–3.50)	4.96(2.99–7.64)
25-2 *	18	0.44(0–1.96)	1.03(0–4.51)	2.32(0.97–4.51)	6.78(2.99–12.52)	2.31(1.47–3.99)	6.60(4.51–9.24)
25-3 *	28	0.32(0–1.47)	0.74(0–4.51)	2.21(0.97–4.00)	6.50(2.99–12.52)	2.20(1.47–3.49)	6.30(4.51–9.24)
26-1	14	2.89(2.47–4.50)	8.57(7.64–14.20)	1.84(0.49–3.47)	5.28(1.48–10.86)	1.79(0.98–3.47)	5.04(2.99–9.24)
26-2 *	18	1.74(0–4.00)	5.17(0–10.86)	1.37(0–3.49)	3.97(0–10.86)	0.73(0–2.97)	1.88(0–7.64)
26-3 *	30	0.35(0–1.47)	0.87(0–4.51)	2.17(0.49–4.00)	6.43(1.48–12.52)	2.15(0.98–3.49)	6.23(2.99–9.2)
4-2	28	2.05(0.98–3.49)	6.2(2.98–10.86)	2.17(0.49–5.04)	6.51(1.48–14.20)	1.81(0.98–3.49)	5.27(2.99–9.24)
4-3	42	1.69(1.47–2.47)	5.03(4.51–7.64)	1.42(0.49–2.97)	4.18(1.48–7.64)	0.64(0.49–1.47)	1.66(1.48–2.9)
4-5	70	2.04(1.47–4.00)	5.72(4.51–10.86)	2.41(0.49–5.59)	6.86(1.48–14.20)	1.94(0.98–4.02)	5.27(2.99–9.24)
4-7	112	2.36(1.97–3.50)	3.92(2.99–9.24)	3.67(1.47–6.11)	7.96(2.99–14.20)	3.2(1.47–4.53)	6.26(2.99–9.2)
4-8	162	1.88(0.98–3.49)	5.47(2.99–10.86)	1.31(0.49–3.47)	3.68(1.48–9.24)	0.90(0-2.48)	2.30(0–6.06)
10-1	14	2.8(2.47–4)	8.44(7.64–12.52)	1.75(0.49–2.99)	5.08(1.48–9.24)	1.70(0.98–2.99)	4.83(2.99–7.64)
10-2	28	2.72(2.47–4)	8.27(7.64–12.52)	1.67(0.49–2.99)	4.90(1.48–9.24)	1.62(0.98–2.99)	4.67(2.99–7.64)
10-3	42	2.34(0.49–3.49)	6.92(1.48–10.86)	2.26(0.49–5.57)	6.59(1.48–14.20)	2.12(0.98–4.00)	6.06(2.99–10.81)
10-5 *	70	0.87(0.49–2.48)	2.07(1.48–4.51)	2.77(1.47–5.06)	7.90(4,51–12.52)	2.76(1.96–4.53)	7.70(6.06–9.24)
10-7	112	0.86(0.49–2.48)	2.38(1.48–6.06)	2.77(1.47–5.06)	8.25(4.51–14.20)	2.76(1.96–4.54)	8.05(6.06–10.81)
10-8	162	1.34(0.97–2.47)	3.85(2.99–7.6)	1.83(0.98–4.00)	5.37(2.99–10.87)	1.28(0.98–2.48)	3.53(2.99–6.06)

EIAV: Equine Infectious Anemia Virus; * Indicates typical febrile episodes; dpi: days post infection; nt: nucleotide; aa: amino acid; V: vaccine.
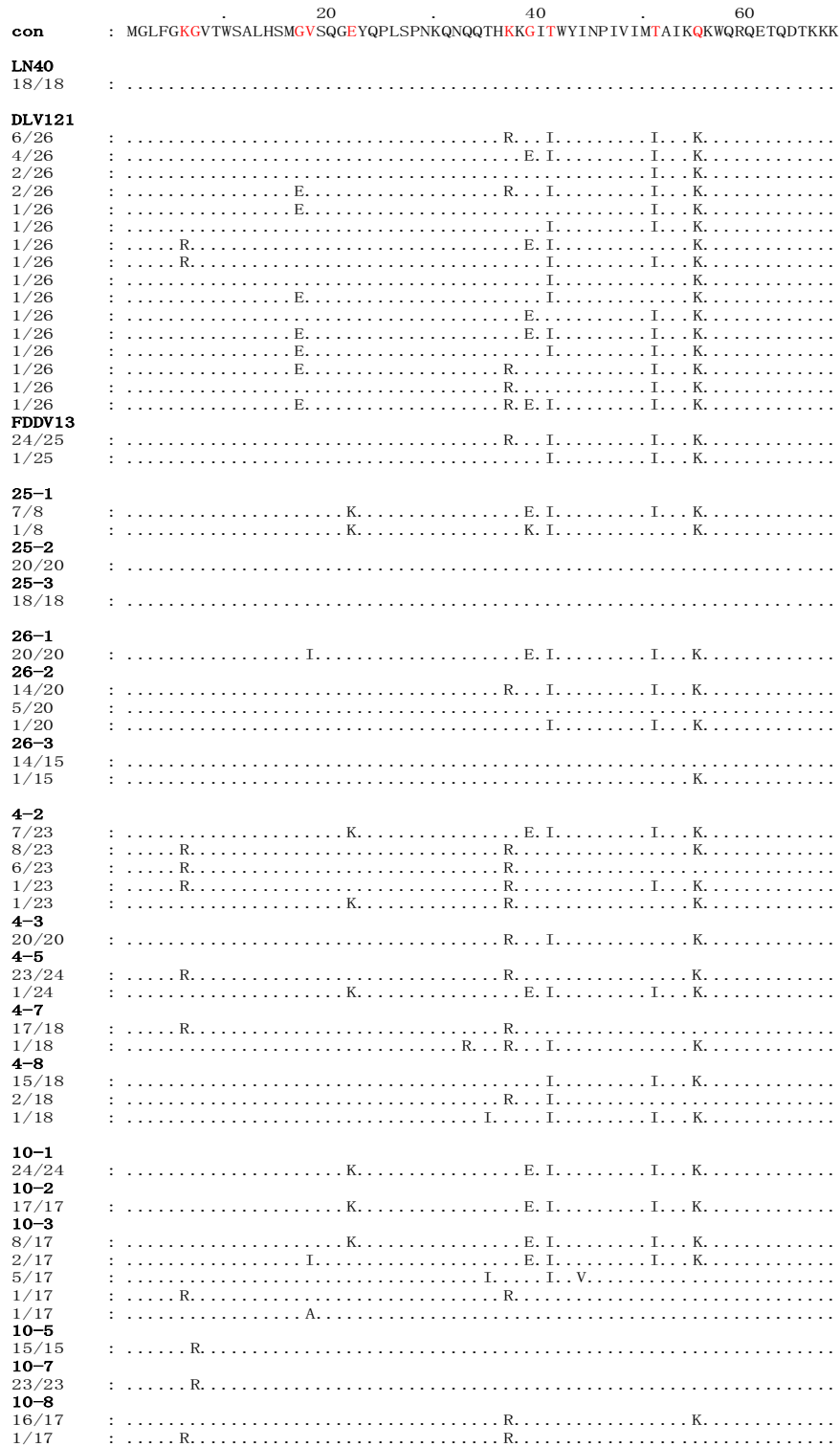
```
                    .         20         .         40         .         60
con      : MGLFGKGVTWSALHSMGVSQGEYQPLSPNKQNQQTHKKGITWYINPIVIMTAIKQKWQRQETQDTKKK

LN40
18/18    : ................................................................

DLV121
6/26     : .........................................R...I.........I...K.............
4/26     : .........................................E.I.........I...K.............
2/26     : .....................................................I...K.............
2/26     : .....................E...................R...I.........I...K.............
1/26     : .....................E...............................I...K.............
1/26     : .....................................................I.I...K.............
1/26     : ......R..............................E.I...........K.............
1/26     : ......R..............................I.........I...K.............
1/26     : .....................................I.........I...K.............
1/26     : .....................E...............I.........I...K.............
1/26     : .....................................E.I.........I...K.............
1/26     : .....................E...............E.I.........I...K.............
1/26     : .....................E...............I.........I...K.............
1/26     : .....................E...............R.............I...K.............
1/26     : .....................................R.............I...K.............
1/26     : .....................E...............R.E.I.........I...K.............

FDDV13
24/25    : .........................................R...I.........I...K.............
1/25     : .....................................I.........I...K.............

25-1
7/8      : ....................K...............E.I.........I...K.............
1/8      : ....................K...............K.I.............K.............

25-2
20/20    : ................................................................

25-3
18/18    : ................................................................

26-1
20/20    : .................I...................E.I.........I...K.............

26-2
14/20    : .........................................R...I.........I...K.............
5/20     : ................................................................
1/20     : .....................................I.........I..K.............

26-3
14/15    : ................................................................
1/15     : .............................................K.............

4-2
7/23     : ....................K...............E.I.........I...K.............
8/23     : ......R..............................R.............K.............
6/23     : ......R..............................R.............K.............
1/23     : ......R..............................R.............I...K.............
1/23     : ....................K...............R.............K.............

4-3
20/20    : .........................................R...I.............K.............

4-5
23/24    : ......R..............................R.............K.............
1/24     : ....................K...............E.I.........I...K.............

4-7
17/18    : ......R..............................R.............................
1/18     : .....................................R...R...I.............K.............

4-8
15/18    : .............................................I.........I...K.............
2/18     : .....................................R...I.............................
1/18     : ...............................I.....I.........I...K.............

10-1
24/24    : ....................K...............E.I.........I...K.............

10-2
17/17    : ....................K...............E.I.........I...K.............

10-3
8/17     : ....................K...............E.I.........I...K.............
2/17     : .................I...................E.I.........I...K.............
5/17     : .............................I.....I..V.........................
1/17     : ......R..............................R.............................
1/17     : ................A...............................................

10-5
15/15    : ......R..........................................................

10-7
23/23    : ......R..........................................................

10-8
16/17    : .....................................R.............K.............
1/17     : ......R..............................R.............................
```

Figure 1. The deduced multiple the deduced sequence alignment of amino acids encoded by the *S2* genes of viruses isolated from four EIAV_{LN40}-infected horses. The sequences of all the viral clones were aligned to the consensus sequence (con) of EIAV_{LN40}, shown at the top. Only the amino acid residues that are different from the consensus sequence of EIAV_{LN40} are indicated. The dots indicate residues identical to the consensus sequence. The numbers in the left column in each sequence indicate the number of a specific sequence in the total detected clones of each sample. The red letters denote the stable substitutions.

As summarized in Table 1, the divergence of the *S2* nucleotide sequence between EIAV isolated from the four horses after inoculation for 150 days and EIAV$_{LN40}$, EIAV$_{DLV121}$ and EIAV$_{FDDV13}$ were 0%–4.50%, 0%–6.11%, and 0%–4.54%, respectively. The percentages of divergence of the corresponding amino acid sequences were 0%–14.20%, 0–14.20, and 0%–10.81%, respectively. An alignment analysis of the deduced S2 amino acids sequences between the EIAV$_{LN40}$ consensus reference and the isolates from experimentally infected horses was performed (Figure 1). The variable amino acid positions were scored by the existence of at least three clones containing an amino acid residue that differed from the EIAV$_{LN40}$ consensus sequence, which was summarized from 18 clones. The results revealed that ten stable amino acid variations in S2, including 6K/R, 7G/R, 17G/I, 18V/I, 22E/K, 37K/R, 39G/E, 41T/I, 51T/I, and 55Q/K (Figure 1). These mutated residues accounted for 14.7% (10/68) of the total deduced S2 amino acid residues.

3.2. Mutations in the S2 Gene Were the Result of Positive Selection

To investigate whether the mutations in *S2* occurred randomly or as a result of evolutionary pressure, which implicates the involvement of the target gene in an altered phenotype, the *S2* sequences were analyzed for the evidence of positive selection using a maximum likelihood (ML) method implemented in the PAML4.3 software package. This method is based on the ratio of non-synonymous (dN) to synonymous (dS) substitutions. Three model pairs (M0/M3, M1a/M2a, and M7/M8) were employed to evaluate the likelihood of positive selection for the mutations in *S2*. The expected ratio (ω) of the dN to dS substitutions in a gene is one (ω = 1). The former models in these pairs are neutral models that do not permit positive selection (ω < 1), and the latter ones are alternative models that permit positive selection (ω > 1). The analytic data shown in Table 2 indicate that the comparisons of M0 to M3, M1a to M2a, and M7 to M8 were significant and that the models that permit positive selection are a better fit to these data, *i.e.*, the *S2* gene was under positive selection pressure. The amino acids that are most likely responsible for the non-neutral pattern were detected. Six codons (17, 18, 22, 37, 39, and 41, see Table 2) appeared to be under positive selection pressure (the posterior probabilities were over 95%).

Table 2. Summary of the parameter estimates of different codon evolution models for the EIAV *S2* gene. dN: non-synonymous; dS: synonymous.

Model	dN/dS	Parameters [b,c]	2ΔlnL	Positively Selected Codons [d]
M0 (ratio of one) 1 [a]	2.2197	ω = 2.2197		None
M3 (discrete) 5 [a]	2.4099	$P_0 = 0.54824$ $p_1 = 0.34821$ $p_2 = 0.10355$ $\omega_0 = 0.79215$ $\omega_1 = 3.59132$ $\omega_2 = 7.00158$	34.71 $p < 0.01$	17, 18, 22, 37, 39, 41
M1a (neutral) 1 [a]	0.8434	$p_0 = 0.17661$ $p_1 = 0.82339$ $\omega1 = 0.11319$ $\omega_1 = 1.00000$		Not permitted
M2a (selection) 3 [a]	2.4296	$p_0 = 0.06756$ $p_1 = 0.56448$ $p_2 = 0.36796$ $\omega_0 = 0.32811$ $\omega_1 = 1.00000$ $\omega_2 = 5.00850$	47.544 $p < 0.01$	17, 18, 22, 37, 39, 41

Table 2. *Cont.*

Model	dN/dS	Parameters [b,c]	2ΔlnL	Positively Selected Codons [d]
M7 (β) 2 [a]	1.0000	P = 97.53765 q = 0.00500		Not permitted
M8 (β & ω) 4 [a]	2.4337	$p_0 = 0.63252$ $p_1 = 0.36748$ p = 0.19920 q = 0.01258 ω = 5.02072	51.384 $p < 0.01$	17, 18, 22, 37, 39, 41

[a] Number of degrees of freedom used; [b] Values in parentheses are not free parameters; [c] Parameters p and q are the shape parameters of the beta distribution that underlies M7 and M8; [d] Codons inferred to be under selection at a level of 85%. Numbers in bold refer to codons with posterior probabilities over 95%.

3.3. The Vaccine- and Pathogenic-Specific S2 Sequences Were Identified in Inoculated Horses, Depending on the Inoculation Dose and Infection Status

To examine whether the *S2* gene evolved at different infection stages to best fit the environmental and immunity pressures of the host, the mutations that presented as changes in the deduced amino acid residues in the S2 protein were longitudinally analyzed by comparing the *in vivo* isolated sequences with the consensus S2 sequence of EIAV$_{LN40}$.

The phylogenetic analyses of the *S2* gene sequences of the EIAV$_{LN40}$ clones isolated from the infected horses at different sampling times were performed by comparing the sequences EIAV$_{LN40}$ before inoculation with the sequences of two attenuated vaccines strains, which consisted of 26 clones of EIAV$_{DLV121}$ and 25 clones of EIAV$_{FDDV13}$. The phylogenetic analysis showed that the *S2* sequences were obviously split into three branches, Branches A, B, and C, on the phylogenetic tree (Figure 2). The sequences of EIAV$_{LN40}$ isolated from the horses at different infection stages were identified in different branches that correlated with the period of infection and the clinical symptoms of the host, which were largely determined by the doses of inoculated virus.

Branch A contained all of the clones of samples 4-8, 10-1, 10-2, 25-1, and 26-1 and some of the clones of samples 4-2, 10-3, and 26-2, as well as all of the clones of the EIAV$_{DLV121}$ and EIAV$_{FDDV13}$ vaccine strains (Figure 2). Except for sample 4-8, all of the clones clustered in this branch were isolated from the horses in the early stages of EIAV$_{LN40}$ infection (14 to 28 dpi). In particular, the clones of samples 10-1, 10-2, 10-3, and 25-1 and some of the clones of sample 4-2 were highly homologous, and the following mutations were identified: 22E/K, 39G/E, 41T/I, 51T/I and 55Q/K (Figures 1 and S1). Additionally, the same sub-branch of Branch A included all of the clones of 26-1, which contained an extra 18V/I mutation and did not contain 22E/K. Twelve clones of sample 26-2 had an *S2* gene sequence identical to that of the EIAV vaccine strains (EIAV$_{DLV121}$ and EIAV$_{FDDV13}$, which were attenuated *in vitro*) and were co-located at the same sub-branch of Branch A. The mutations identified were 37K/R, 41T/I, 51T/I and 55Q/K (Figure 1). In addition, the *S2* sequence of a few clones of the vaccine strains clustered with clones of sample 4-8 at the same sub-branch of Branch A. The following mutations were found: 41T/I, 51T/I and 55Q/K (Figure 1). Further analysis revealed that with the exception of 6K/R and 7G/R, all consensus mutations were predominantly observed in sequences clustered in Branch A (Table 3).

Table 3. Differences between S2 sequences derived from Branch A and Branch C.

S2 Variation [a]	No. (%) of Occurrences in:		*p* Value [b]
	A (*n* = 192)	C (*n* = 144)	
6K/R	25(13)	22(15)	> 0.05
7G/R	0(0)	38(26)	< 0.01
17G/E	8(4)	0(0)	< 0.05
18V/I	22(11)	0(0)	< 0.01
22E/K	64(33)	0(0)	< 0.01
37K/R	72(38)	24(17)	< 0.01
39G/E	93(48)	0(0)	< 0.01
41T/I	162(84)	7(5)	< 0.01
51T/I	164(85)	0(0)	< 0.01
55Q/K	190(99)	1(1)	< 0.01

[a] Positions are those indicated by letters denote in Figure 1; [b] Determined by the chi-square test of the SAS 9.2.

In contrast, Branch C contained all of the clones of the initial EIAV$_{LN40}$ strain; samples 25-2, 25-3, 26-3, 10-5, and 10-7; 5/20 clones of sample 26-2; 17/18 clones of sample 4-7; 4/23 clones of 4-2; and a few clones of samples 10-3 and 4-8 (Figure 2). These *in vivo* isolated clones included most of the viruses isolated at the time points of clinical EIA presentation (body temperature >39 °C and platelet count <100,000 unit/µL) of the horses inoculated with the higher EIAV dose (1×10^6 TCID$_{50}$), including samples 25-2, 25-3, 26-2, and 26-3. The *S2* gene sequences of these viruses were identical to that of the pathogenic EIAV$_{LN40}$ strain (Figure 1). In addition, the samples from horses inoculated with a lower dose of EIAV (1×10^5 TCID$_{50}$) and isolated after the fever peak (sample 10-5) and sample 10-7 were co-located at the same sub-branch with EIAV$_{LN40}$ and had only one mutated site (7G/R) compared with the initial EIAV$_{LN40}$ sequence (Figure 1). Other mutations in clones, including most of 4-7 and a few of 4-2 and 4-8, formed another sub-branch in Branch C; the mutations were predominantly 6K/R and 37K/R (Figure 1). A comparison of the deduced amino acid sequences of the isolated *S2* genes demonstrated that among the ten consensus mutations presented in Table 3, only 7G/R was specifically generated in sequences clustered in the Branch C, and 6K/R showed no distribution trend in Branches A and C. No other consensus mutations were predominantly detected in sequences clustered in Branch C.

In addition to Branches A and C, which contained the *S2* sequences of the attenuated vaccine strains and the pathogenic strain, respectively, some clones clustered and formed a third branch, Branch B. These clones primarily contained sequences from some samples taken from horses inoculated with a low dose of EIAV, *i.e.*, samples 4-3, 4-5 and 10-8, as well as some of the clones of sample 4-2 and one clone of sample 4-7 (Figure 2). The mutation sites of these clones were varied and included 22E/K, 39G/E, and 41T/I; 37K/R, 41T/I, and 55Q/K; or 37K/R and 55Q/K (Figures 1 and S1).
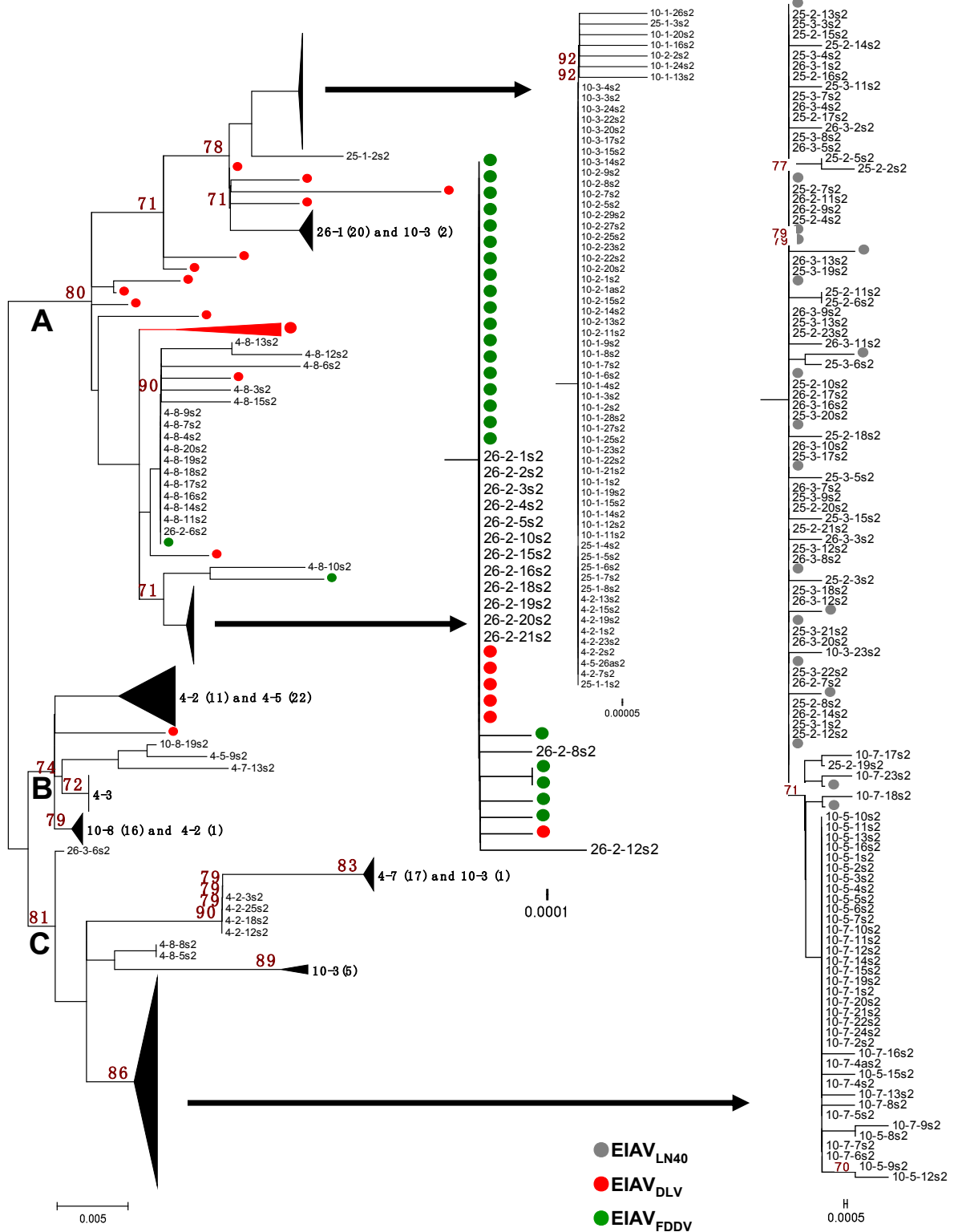
Figure 2. Phylogenetic analysis of the *S2* gene. Sequences include the EIAV_LN40 stock virus, plasma viral RNA obtained from four EIAV_LN40-infected horses, the donkey monocyte-derived macrophage (MDM)-adapted vaccine strain EIAV_DLV121, and the donkey dermal (FDD) cell-attenuated vaccine strain EIAV_FDDV13. The phylogenetic trees were constructed by the neighbor-joining method, calculated with the Kimura 2 parameter in MEGA software (version 5.0, Center for Evolutionary Functional Genomics Biodesign Institute, Arizona State University, Tempe, AZ, USA, 2011).

To investigate whether the aforementioned amino acid substitutions in S2 were associated with different stages and statuses of infection, correlations between S2 consensus mutations and the pathogenesis, as well as the phase of infection were analyzed. Sequences of samples 25-2, 25-3, 26-2, 26-3 and 10-5 were grouped as symptomatic, and sequences from the other time points were grouped as asymptomatic. As examined by the Pearson's chi-squared (X^2) test, the results demonstrated that with the exceptions of 7G/R and 17G/E, eight of the ten primary mutations were significantly correlated with the asymptomatic time points (Table S1). Further, the aforementioned mutation pattern was also significantly correlated with the samples taken during the early phase (earlier than 15 dpi) and the late phase (later than 15 dpi) of infection, when examined by the X^2 test (Table S2). Combined with the analytic data of the location of mutated sequences in the branches of the phylogenetic tree shown in Figure 2, statistical analyses of the ten consensus mutated residues in S2 revealed that most of these mutations were generated in a specific group of EIAV species and were clustered in Branch A, which emerged in only a particular phase of infection. The quasispecies of pathogenic EIAV$_{LN40}$ with or without these mutations exhibited different capabilities to cause active EIA (Table 3).

4. Discussion

In this study, we found that the *S2* genes of pathogenic EIAV$_{LN40}$ strains were highly diverse at different infectious stages, which correlated with the doses of inoculated virus. The higher and lower doses mimicked the acute and chronic EIA, respectively. The divergence of the S2 protein sequence between EIAV isolated *in vivo*, as well as that of EIAV$_{LN40}$ before inoculation, ranged from 0%–14.20%. As many as 14.7% (10 of 68) of the residues were found to be mutated in at least three of the 387 total sequenced clones isolated during 150 days of an *in vivo* inoculation course in four horses. Phylogenetic analysis revealed that these *in vivo* isolated *S2* sequences could be separated into three groups, attenuated-specific Branch A, pathogenic-specific Branch C, and Branch B, which was located between Branches A and C. Interestingly, in all four experimentally infected horses, the original *S2* sequences of the inoculated EIAV$_{LN40}$ were invariably replaced, either completely or partially (partially in horse #4, in which sequences from the first time point were not detectable), by the vaccine-specific sequence (shifted from Branch C to Branch A of the phylogenetic tree). A similar and important observation is that the pathogenic-specific S2, which is identical or very similar to the initial inoculating virus in amino acid sequence, re-emerged after a particular period, either the short period of 15 dpi in horses infected with a high dose of EIAV (1×10^6 TCID$_{50}$) or the long period of approximately 40 dpi to 100 dpi (with the intermediate phase in Branch B) in horses infected with a low dose of the virus (1×10^5 TCID$_{50}$). Statistical analysis revealed that ten primary mutations were significantly different in viruses isolated from time points of active disease or subclinical infection as well as in viruses isolated from the early or late stages of infection (Tables S1 and S2), suggesting an evolutionary tendency of the *S2* gene *in vivo*. Most of these mutations were determined to have resulted from positive selection pressure. These results revealed that *in vivo* selection pressure promotes longitudinal sequence changes of *S2* to allow for adaptation to the host, which in turn alters EIAV pathogenicity, possibly via concurrent effects of mutations in other viral genes. The selective forces that drive these sequence variations and the phenotypic appearance of these viruses isolated from infected hosts will need to be investigated in future studies.

The results in this study showed that the clones from samples 25-2, 25-3, 26-3, 10-5, and 10-7 and 5/20 clones from sample 26-2 are located in the same sub-branch on Branch C, along with the initial EIAV$_{LN40}$ strain. Particularly, the S2 sequences of clones isolated just after febrile episodes, *i.e.*, samples 25-2, 25-3, 26-3 and 26-2 (5/20), were identical to that of the initial EIAV$_{LN40}$ strain, which was previously obtained from the plasma of an EIAV-infected horse during a febrile episode. Sample 26-2, which was obtained from the plasma at the beginning of the first febrile episode, contained both pathogenic-specific and vaccine-specific clones. The pathogenic-specific clones became dominant when detected from sample 26-3 at the time just before death; most of the sequences were identical to those of the initial EIAV$_{LN40}$ strain. These lines of evidence indicate the re-emergence of the pathogenic-specific *S2* sequence, which is correlated with active EIA (Table S1). In addition, although clones from samples 10-5 and 10-7 clustered in Branch C, their sequences differed by one residue (7G/R) from EIAV$_{LN40}$. Although this horse (#10) did not present active EIA, several waves of slight fever (rectal temperature of approximately 39 °C) and minor decreases in the platelet count (below 100,000/μL) occurred before sample 10-5 was collected (Figure 1). These results suggest that the aforementioned longitudinal S2 sequence alteration is associated with disease development (Table 3, Tables S1 and S2). Notably, the virus in horses #4 and #10 did not re-emerge with the EIAV$_{LN40}$ sequence, as was observed in horses #25 and #26, but eventually did evolve into Branch B or mixed A/C and B/C (Figure S1). We presume that certain levels of immunity were induced after the initial infection. If the horses do not die of acute EIA, then the immune suppression drives EIAV to further evolve into a less pathogenic and/or latent status. This hypothesis is supported by previous reports indicating that horses that clinically recovered from chronic EIA acquired resistance to subsequent EIAV infection [17].

At the early phase of infection, we observed the appearance of clones containing *S2* sequences clustered with the sequence of vaccine strains, which implies the selection of vaccine-like viral species in the quasispecies pool of EIAV, as well as the re-emergence of clones containing *S2* sequences clustered with the sequence of EIAV$_{LN40}$ strain at the late infection phase; this pattern of sequence clustering is correlated with typical EIA clinical symptoms. This tendency for *in vivo* evolution was observed in the *S2* gene of all four horses examined in this study as well as in the previously examined *gp90* gene [13]. This feature of EIAV *in vivo* evolution is consistent with the trophic change of HIV-1 in different infection phases and the re-boost of latent strains in clinically cured patients [18]. A general pattern might exist for the adaption of pathogenic EIAV$_{LN40}$ in the host. Our earlier studies on cultivated primary target cells of EIAV revealed that the pathogenic and vaccine strains induced different panels of cytokines at different stages of infection [19,20]. A particular phenotype pattern, including the viral structure and function controlled by *S2* and *gp90,* is considered beneficial in a specific infection phase to successfully establish infection. Studies have shown that EIAV with wild-type *S2* induced significantly higher levels of inflammatory cytokines and chemokines than the *S2*-deficient control in eMDMs. It has been proposed that the enhanced cytokines created a suitable intracellular environment for viral infection and replication and affected the disease development [21]. The clinical signs (febrile responses, thrombocytopenia and anemia) associated with acute EIA are mediated by pro-inflammatory cytokines, such as TNFα, IL-6, and TGFβ [22]. A number of studies on *lentiviruses*, including HIV-1, SIV and feline immunodeficiency virus (FIV), have confirmed that the expression of a variety of cytokines could affect viral replication and disease progression.

The emergence of quasispecies containing the mutated *S2* gene could be considered the result of the generation of new mutations in predominant viral species, the selection of existing underrepresented species, or both. The editing activities of APOBECs and ADAR 1 are major known selection pressures that stimulate G to A and A to G mutations in retroviruses and double-stranded RNA viruses [23,24]. An altered environment could boost the replication of certain existing underrepresented viral species. In a previous study, using the single-genome amplification (SGA) method, we identified 41 unique *gp90* V3–V5 sequences from 73 clones of the EIAV vaccine strain EIAV$_{FDDV13}$, among which contained nine sequences with a 3-nucleotide insert that was previously identified only in the sequences of pathogenic strains [25]. Studies on HIV-1 found that a single transmitted founder (TF) virus, which possesses specific phenotypic properties compared with chronic control viruses, is sufficient to establish a new infection [26]. Clones containing the vaccine-specific *S2* gene isolated from EIAV$_{LN40}$-infected horses could be selected from existing viral species. If this finding was true, an extensive sequence analysis of specific quasispecies of a *lentivirus* strain would facilitate antigen design for vaccine development.

Author Contributions

J.-H.Z., X.W, and X.-F.W. designed the experiments and wrote the paper. X.-F.W, S.W., Q.L., Y.-Z.L., C.D., Y.-D.T., and L.N. performed the experiments and analyzed the data. All authors read and approved the final manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1. Issel, C.J.; Adams, W.V., Jr.; Meek, L.; Ochoa, R. Transmission of equine infectious anemia virus from horses without clinical signs of disease. *J. Am. Vet. Med. Assoc.* **1982**, *180*, 272–275.
2. Craigo, J.K.; Ezzelarab, C.; Cook, S.J.; Chong, L.; Horohov, D.; Issel, C.J.; Montelaro, R.C. Envelope determinants of equine lentiviral vaccine protection. *PLoS One* **2013**, *8*, e66093.
3. Leroux, C.; Cadore, J.L.; Montelaro, R.C. Equine Infectious Anemia Virus (EIAV): What has HIV's country cousin got to tell us? *Vet. Res.* **2004**, *35*, 485–512.
4. Fagerness, A.J.; Flaherty, M.T.; Perry, S.T.; Jia, B.; Payne, S.L.; Fuller, F.J. The S2 accessory gene of equine infectious anemia virus is essential for expression of disease in ponies. *Virology* **2006**, *349*, 22–30.
5. Wang, X.; Wang, S.; Lin, Y.; Jiang, C.; Ma, J.; Zhao, L.; Lv, X.; Wang, F.; Shen, R.; Kong, X.; Zhou, J. Genomic comparison between attenuated Chinese equine infectious anemia virus vaccine strains and their parental virulent strains. *Arch. Virol.* **2010**, *156*, 353–357.

6. Covaleda, L.; Gno, B.T.; Fuller, F.J.; Payne, S.L. Identification of cellular proteins interacting with equine infectious anemia virus S2 protein. *Virus Res.* **2010**, *151*, 235–239.

7. Payne, S.L.; Fuller, F.J. Virulence determinants of equine infectious anemia virus. *Curr. HIV Res.* **2010**, *8*, 66–72.

8. Kirchhoff, F.; Easterbrook, P.J.; Douglas, N.; Troop, M.; Greenough, T.C.; Weber, J.; Carl, S.; Sullivan, J.L.; Daniels, R.S. Sequence variations in human immunodeficiency virus type 1 Nef are associated with different stages of disease. *J. Virol.* **1999**, *73*, 5497–5508.

9. Walker, P.R.; Ketunuti, M.; Choge, I.A.; Meyers, T.; Gray, G.; Holmes, E.C.; Morris, L. Polymorphisms in Nef associated with different clinical outcomes in HIV type 1 subtype C-infected children. *AIDS Res. Hum. Retrovir.* **2007**, *23*, 204–215.

10. Li, F.; Puffer, B.A.; Montelaro, R.C. The S2 gene of equine infectious anemia virus is dispensable for viral replication *in vitro*. *J. Virol.* **1998**, *72*, 8344–8348.

11. Li, F.; Leroux, C.; Craigo, J.K.; Cook, S.J.; Issel, C.J.; Montelaro, R.C. The S2 gene of equine infectious anemia virus is a highly conserved determinant of viral replication and virulence properties in experimentally infected ponies. *J. Virol.* **2000**, *74*, 573–579.

12. Gao, X.; Jiang, C.G.; Wang, X.F.; Lin, Y.Z.; Ma, J.; Han, X.E.; Zhao, L.P.; Shen, R.X.; Xiang, W.H.; Zhou, J.H. Reverse mutation of the virulence-associated S2 gene does not cause an attenuated equine infectious anemia virus strain to revert to pathogenicity. *Virology* **2013**, *443*, 321–328.

13. Wang, X.; Wang, S.; Lin, Y.; Jiang, C.; Ma, J.; Zhao, L.; Lv, X.; Wang, F.; Shen, R.; Zhou, J. Unique evolution characteristics of the envelope protein of EIAVLN40, a virulent strain of equine infectious anemia virus. *Virus Genes* **2011**, *42*, 220–228.

14. Yang, Z.; Wong, W.S.; Nielsen, R. Bayes empirical bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* **2005**, *22*, 1107–1118.

15. Wong, W.S.; Yang, Z.; Goldman, N.; Nielsen, R. Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. *Genetics* **2004**, *168*, 1041–1051.

16. Soares, A.E.; Soares, M.A.; Schrago, C.G. Positive selection on HIV accessory proteins and the analysis of molecular adaptation after interspecies transmission. *J. Mol. Evol.* **2008**, *66*, 598–604.

17. Craigo, J.K.; Montelaro, R.C. EIAV envelope diversity: Shaping viral persistence and encumbering vaccine efficacy. *Curr. HIV Res.* **2010**, *8*, 81–86.

18. Zhu, T.; Mo, H.; Wang, N.; Nam, D.S.; Cao, Y.; Koup, R.A.; Ho, D.D. Genotypic and phenotypic characterization of HIV-1 patients with primary infection. *Science* **1993**, *261*, 1179–1181.

19. Lin, Y.Z.; Shen, R.X.; Zhu, Z.Y.; Deng, X.L.; Cao, X.Z.; Wang, X.F.; Ma, J.; Jiang, C.G.; Zhao, L.P.; Lv, X.L.; *et al.* An attenuated EIAV vaccine strain induces significantly different immune responses from its pathogenic parental strain although with similar *in vivo* replication pattern. *Antivir. Res.* **2011**, *92*, 292–304.

20. Ma, J.; Wang, S.S.; Lin, Y.Z.; Liu, H.F.; Wei, H.M.; Du, C.; Wang, X.F.; Zhou, J.H. An attenuated EIAV strain and its molecular clone strain differentially induce the expression of Toll-like receptors and type-I interferons in equine monocyte-derived macrophages. *Vet. Microbiol.* **2013**, *166*, 263–269.

21. Covaleda, L.; Fuller, F.J.; Payne, S.L. EIAV S2 enhances pro-inflammatory cytokine and chemokine response in infected macrophages. *Virology* **2010**, *397*, 217–223.

22. Cook, R.F.; Leroux, C.; Issel, C.J. Equine infectious anemia and equine infectious anemia virus in 2013: A review. *Vet. Microbiol.* **2013**, *167*, 181–204.

23. Harris, R.S.; Liddament, M.T., Retroviral restriction by APOBEC proteins. *Nat. Rev. Immunol.* **2004**, *4*, 868–877.

24. Ota, H.; Sakurai, M.; Gupta, R.; Valente, L.; Wulff, B.E.; Ariyoshi, K.; Iizasa, H.; Davuluri, R.V.; Nishikura, K. ADAR1 forms a complex with Dicer to promote microRNA processing and RNA-induced gene silencing. *Cell* **2013**, *153*, 575–589.

25. Wei, H.M.; Wang, X.F.; Wang, S.S.; Du, C.; Liu, H.F.; Liu, Q.; Zhou, J.H. The application of single-genome amplification and sequencing in genomic analysis of an attenuated EIAV vaccine. *Bing Du Xue Bao* **2012**, *28*, 431–438 (In Chinese).

26. Keele, B.F.; Giorgi, E.E.; Salazar-Gonzalez, J.F.; Decker, J.M.; Pham, K.T.; Salazar, M.G.; Sun, C.; Grayson, T.; Wang, S.; Li, H.; *et al.* Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 7552–7557.