Research in Veterinary Science 112 (2017) 185–191

Contents lists available at ScienceDirect

Research in Veterinary Science

journal homepage: www.elsevier.com/locate/rvsc

Nonsynonymous changes of equine lentivirus receptor-1 (ELR1) gene in amino acids involved in the interaction with equine infectious anemia virus (EIAV)

C.M. Corbi-Botto[a,b], S.A. Sadaba[a,b], M.E. Zappa[a], P. Peral-García[a], S. Díaz[a,*]

[a] IGEVET - Instituto de Genética Veterinaria "Ing. Fernando Noel Dulout", (UNLP-CONICET LA PLATA), Facultad de Ciencias Veterinarias, Universidad Nacional de La Plata, CC 296. CP B1900AVW, La Plata, Argentina

[b] Research fellows from Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina

ARTICLE INFO

Keywords:
Equine lentivirus receptor
EIAV
Single nucleotide polymorphism
Equids

ABSTRACT

Equine lentivirus receptor-1 (ELR1) has been characterized as the specific functional receptor that mediates equine infectious anemia virus (EIAV) entrance to horse macrophages. This receptor is tumor necrosis factor receptor superfamily member 14 (*TNFRSF14*). The aim of this study was to investigate the occurrence of allelic variants in the coding sequence of equine *TNFRSF14* gene by screening for single-nucleotide polymorphisms (SNPs) in different equine populations. Forty seven horse samples were randomly selected from a reservoir of EIAV-seropositive and seronegative samples collected from different outbreaks and regions of Argentina. DNA samples were scanned via PCR and direct sequencing of exon 3 and exon 5 of *TNFRSF14* gene. A total of 21 SNPs were identified, of which 11 were located in coding sequences. Within exon 5, four SNPs caused nonsynonymous substitutions, while two other SNPs caused synonymous substitutions in crucial residues (Ser112 and Thr114) implicated in the interaction with EIAV. Despite some of exon 5 variants occurred exclusively in EIAV-positive or EIAV-negative horses, critical residues for the function of the mature protein were conserved, accounting for selective pressures in favor of preserving the specific function of TNFRSF members and the host immune response. To our knowledge, this is the first report of the existence of allelic variations involving some crucial amino acid residues in horse ELR1. Further, it could be an initial step to test the possible functional relevance and relationship of these variants with EIAV infection and disease progression as well as to develop preventive strategies.

1. Introduction

Equine infectious anemia (EIA) is a chronic disease that affects members of the *Equidae* family, such as horses, donkeys and mules (Cook et al., 1996; Spyrou et al., 2003). The etiological agent of EIA is equine infectious anemia virus (EIAV), a lentivirus which belongs to the *Retroviridae* family and causes swamp fever disease, with characteristic episodes of fever, anemia, depression, weight loss and edema (Sellon, 1993; Cook et al., 1996). The cellular receptor that mediates EIAV entry into target cells is known as equine lentivirus receptor-1 (ELR1) and is encoded by tumor necrosis factor receptor superfamily member 14 (*TNFRSF14*) gene (Zhang et al., 2005), a type I transmembrane protein expressed on the surface of equine macrophages. The protein ectodomain contains four cysteine-rich domains (CRD1–4) (Zhang et al., 2008) defined by intrachain disulfide bonds generated by highly conserved cysteine residues within the receptor chains (Smith et al., 1994). The activities of TNFRSF members encompass a wide range of biological functions, such as cell proliferation, survival, differentiation and apoptosis of responding cells, hence regulating the normal physiology of the immune system. In humans also plays a role in signal transduction pathways that activate inflammatory and inhibitory T-cell immune response (Locksley et al., 2001).

Ligands of the TNFRSF have pivotal roles in the organization and function of the immune system. TNF ligands share a common structural motif, the TNF homology domain (THD), which binds to CRDs of TNF receptors. CRDs are composed of structural modules, whose variation in number and type confers heterogeneity upon the family (Bodmer et al., 2002). Protein folds reminiscent of the THD and CRD are also found in other protein families, raising the possibility that the mode of interaction between TNF and TNF receptors might be conserved in other contexts. In this regard, despite a lack of amino acid sequence homology, it has been demonstrated that lentivirus envelope proteins

* Corresponding author at: IGEVET (UNLP-CONICET LA PLATA), Facultad de Ciencias Veterinarias, UNLP, 60 y 118 S/N CC 296, 1900 La Plata, Argentina.
E-mail address: sdiaz@igevet.gob.ar (S. Díaz).

http://dx.doi.org/10.1016/j.rvsc.2017.05.001
Received 26 October 2016; Received in revised form 17 March 2017; Accepted 1 May 2017

share similar structural features (Gallaher et al., 1995). Moreover, a number of DNA and RNA viruses have evolved a convergent mechanism to invade cells by targeting the CRD1 of several TNFRSF members, and taking advantage of receptor-mediated endocytosis. This mechanism allowed some viruses to exploit and manipulate the signaling pathways transduced by TNFRSF members, regulating cell death and survival of the infected cells, acting as a strong selective pressure in the evolution of host defenses (Cheung et al., 2005; Benedict et al., 2003; Kinkade and Ware, 2006).

In contrast to most other lentiviruses, such as human immunodeficiency virus (HIV-1), simian immunodeficiency virus (SIV) and feline immunodeficiency virus (FIV), which require co-receptors for successful infection, EIAV appears to depend only on a functional ELR1 for the invasion of target cells (Zhang et al., 2005). Furthermore, comparative sequence analysis of horse ELR1 with herpesvirus entry mediator (HVEM) and FIV receptor CD134 indicated that Tyr61, Leu70 and Gly72 in CRD1 of ELR1 are important residues for binding to the surface glycoprotein (gp90) of EIAV, while Ser112 and Thr114 in CRD2 of ELR1 are crucial for functional binding (Qian et al., 2015). Also, amino acids within CRD1 were described as the predominant determinant of functional gp90 binding to ELR1 (Zhang et al., 2008). In addition, several studies indicate that gp90 of the envelope protein is the predominant site of EIAV antigenic variation (Rushlow et al., 1986; Payne et al., 1987; Leroux et al., 2001).

TNFRSF14 gene is located on Chromosome 2: 47,305,618 (ENSECAG00000008751) and spans 6605 bp, with nine exons of which eight are coding. Multiple alternative splicing variants have been reported previously, including a soluble isoform of ELR1 (Lin et al., 2013). Until now, only protein structural information and a few exonic single-nucleotide polymorphisms (SNPs) of horse ELR1 have been reported (www.ensembl.org; Zhang et al., 2005, 2008; Sun et al., 2008). The characterization of *TNFRSF14* polymorphisms could shed light on the molecular and functional mechanisms underlying ELR1 and EIAV interaction that may affect virus entrance and subsequent disease progression.

In order to investigate the existence of allelic variants in the coding sequence of the equine *TNFRSF14* gene, we screened for new SNPs in different equine populations; the SNPs reported here could provide a prerequisite for studies of its functional importance in the horse.

2. Materials and methods

2.1. Horse samples

The study samples included 47 unrelated mix-breed horses with known EIA serological status. DNA samples were already available at our laboratory; they belonged to retrospective studies on EIA prevalence in different locations of Chaco and Buenos Aires provinces. Sixteen seropositive and 31 seronegative horses, as well as a free EIA donkey, were included. Genomic DNA was extracted from whole blood by using DNAzol® (Invitrogen, Carlsbad, CA, USA), and quantified in a NanoVue™ (GE Healthcare) spectrophotometer. The serological diagnosis was performed by the agar gel immunodiffusion test (AGID) prescribed by the Office International des Épizooties (OIE) for international trade of horses (Coggins' test) in laboratories approved by the National Service of Agri-Food Health and Quality (SENASA, for its acronym in Spanish) sanitary authorities.

2.2. PCR amplification and sequencing of TNFRSRF14 exon 3 and exon 5

TNFRSRF14 publicly available genomic sequence (ENSECAG00000008751) was used as reference in order to design specific primers with Primer3 software (http://bioinfo.ut.ee/primer3-0.4.0/). PCR fragments of coding sequences from the *TNFRSF14* gene were obtained by designing different sets of primers for each independent exon. However, due to the characteristics of the exon flanking sequences, only exon 3 and exon 5 set of primers met the quality requirements for PCR amplification. Target regions included the exon and the nucleotide sequence corresponding to the 3′ end of the adjacent intron (Table S1).

Amplicons of ELR1 exon 3 and exon 5 were obtained separately by PCR amplification using 48 equine DNA samples. Overall, 31 and 38 samples were successfully amplified for exon 3 and 5, respectively. PCR was carried out in a volume of 50 µl containing genomic DNA (50–100 ng), PCR-primers (0.25 µM ELR1e3F, ELR1e3R; 0.2 µM ELR1e5F, ELR1e5R), dNTPs (0.08 mM), Mg^{++} (2.0 mM), *Taq* polymerase (0.5 U), 1 × PCR buffer and 0.66 × enhancer solution provided by the supplier (Inbio Highway, Tandil, Argentina). Amplification conditions for both exons were the following: initial denaturation cycle at 95 °C for 5 min, 35 amplification cycles of 95 °C for 30 s, 58 °C for 45 s, and 72 °C for 1 min. A final extension step at 72 °C for 10 min was applied. PCR products were purified with AccuPrep® PCR Purification Kit (BIONEER, Korea), quantified by spectrophotometry (Nanovue – GE Healthcare) and sequenced by Macrogen, Inc. (Korea) with the same forward primers.

2.3. Sequence and statistical analysis

The raw sequencing data were edited to remove poor quality sequences or segments, either manually or using Chromas Lite version 2.1.1 software (http://technelysium.com.au). The 3′ end of the adjacent intron was also analyzed. The exonic sequences of exon 3 (109 bp) and exon 5 (156 bp) were analyzed independently. Nucleotide alignments were performed by ClustalW version 1.83 (http://www.genome.jp/tools/clustalw/). The DNA sequences were then translated into their predicted amino acid sequences using the Translate tool from the ExPASy Molecular Biology server (http://www.expasy.org/) and aligned to the reference sequence (ENSECAG00000008751) using ClustalW version 1.83 (http://www.genome.jp/tools/clustalw/) and BioEdit multiple alignment software version 7.2.5 (Hall, 1999).

In order to estimate haplotypes within each exon, SNPs linkage phases were analyzed using PHASE v2.1 software (http://stephenslab.uchicago.edu/software.html). When available, pedigree reconstruction was considered to verify the allele identity by segregation analysis in related horses. With the purpose of corroborating haplotypes of rare occurrence, linkage disequilibrium (LD) parameters were estimated by using HAPLOVIEW ver.4.2 software (Barrett et al., 2005); this package was also employed to analyze haplotype frequencies and single locus and multi-marker haplotype association tests. Population genetic analyses were performed for EIAV-positive and EIAV-negative individuals in both exons. For this purpose, unbiased expected heterozygosity (He), observed heterozygosity (Ho), F-statistics (FIS), allele number (Na) and Hardy-Weinberg equilibrium (HWE) were estimated using GENEPOP software package ver.4.2 (Rousset, 2008). For HWE, p-values < 0.05 were considered significant. Predictive analysis of the potential effect of the amino acid changes were performed by Protein Variation Effect Analyzer (PROVEAN) Software (Choi and Chan, 2012) to score for the neutral effect on protein biological function using default score ranges (cut-off score = − 2.5).

2.4. Evolutionary analysis of TNFRSF14 sequences

The equine *TNFRSF14* exon 3 and exon 5 sequences were subjected to phylogenetic analysis by using MEGA version 7.0.14 (Tamura et al., 2013) with the Neighbour Joining method (boostrap = 1000) (Sentsui and Kono, 1987). Horse *TNFRSF14* exon 5 sequences (ENSECAE00000042908), *ELR1e5_001.1*, *ELR1e5_002.1*, *ELR1e5_002.2* and *ELR1e5_003.1* were compared with orthologous sequences of the following species: Human (HUMANENSE00003526946), Gorilla (ENSGGOE00000216900), Chimpanzee (ENSPTRE00000000466), Gibbon (ENSNLEE00000116734), Macaque (ENSMMUE00000158204), Olive Baboon (ENSPANE00000124104), Mouse (ENSMUSE00012253020000), Rat (ENSRNOE0000013178300000), Guinea Pig (ENSCPOE00000122113) and *Equus asinus* (NW_014638278), in

order to evaluate whether the gene has evolved under positive selection or which individual codons had experienced selective pressures. For each codon, the numbers of inferred synonymous and nonsynonymous substitutions were estimated with the joint Maximum Likelihood (ML) reconstructions of ancestral states under a Muse-Gaut model (Muse and Gaut, 1994) of codon substitution and Felsenstein 1981 model (Felsenstein, 1981) of nucleotide substitution. The statistical test dN-dS was used to detect codons that had undergone positive selection, where dS is the number of synonymous substitutions per site (s/S) and dN is the number of nonsynonymous substitutions per site (n/N). A positive value indicated an overabundance of nonsynonymous substitutions, and the probability of rejecting the null hypothesis of neutral evolution (*p*-value).

3. Results

3.1. TNFRSRF14 SNPs

To identify any new SNPs within the *TNFRSF14* gene, sequences for exon 3 (ENSECAE00000042690) and exon 5 (ENSECAE00000042908) were obtained from 29 EIA-seropositive and 38 EIA-seronegative unrelated horses, respectively. Target regions from exon 3 consisted of 151 bp (42 bp from intron 2 and 109 bp from exon 3); the exon 5 region comprised 303 bp (159 bp corresponding to intron 4 and 144 bp from exon 5) (Figs. 1 and 2).

A total of 21 SNPs were identified, ten in the intronic sequences (one SNP every 19 nucleotides), eight in exon 5 (one SNP every 19 nucleotides), and three in exon 3 (one SNP every 36 nucleotides) of *TNFRSF14* gene. Analysis of the nucleotide sequence of exon 3 showed the two previously reported SNPs (www.ensembl.org): rs395939045 and rs397088947 and a new SNP C > T was identified at position 47309385 of chromosome 2 (Table 1). This SNP was validated in the horse population by its presence in both homozygous and heterozygous unrelated individuals. All three SNPs were transitions. On the other hand, screening of exon 5 nucleotide sequences showed eight SNPs: the two previously reported rs68535248 and rs68535250, and six new SNPs; among them five transitions: 47311282 G > A, 47311284 C > T, 47311323 G > A, 47311295 A > G, 47311301 A > G and one transversion: 47311358 T > G (Table 1). SNP at position 47309360 within intron 2 C > T* was only detected in a donkey. Only SNPs in exon coding sequences were further analyzed. All newly identified SNPs were submitted to the NCBI dbSNP Short Genetic Variations database (www.ncbi.nlm.nih.gov/project/SNP) and assigned a submitted SNP number or #ss (Table 1).

The three SNPs analyzed on the exon 3 region showed minor allele frequency (MAF) values ranging from 0.172 to 0.414 (Table 1). The C > T transition newly described at position 47309385 of exon 3 region presented allele T as the most common, with a 78% frequency (MAF = 0.224). Analysis of the predicted amino acid sequences showed that the three SNPs caused synonymous substitutions.

On the other hand, in *TNFRSF14* exon 5, MAF values ranged from 0.013 to 0.224. SNP at nucleotide position 47311301 showed allele A as the most common (78% frequency). Among the eight SNPs analyzed in exon 5, SNP 47311310 C > T (rs68535248) was monomorphic, only showing allele C in all analyzed horses. For LD analysis, this marker was excluded. Five nonsynonymous substitutions were detected on *TNFRSF14* exon 5: the previously reported transition 47311347 A > G (rs68535250) which is predicted to cause a change from glutamic acid to lysine (Table 1), and four new SNPs. Among the newly described SNPs, transition 47311282 A > G is predicted to cause a change from arginine to lysine; transition 47311284 C > T allele is predicted to cause a change from arginine to tryptophan; transition 47311323 A > G allele is predicted to cause a change from aspartic acid to asparagine; and transversion 47311358 T > G is predicted to cause a change from aspartic acid to glutamic acid.

The analysis of the effect of the amino acid substitutions showed that variants Arg108Lys, Asp122Asn, Glu130Lys, Asp133Glu had a PROVEAN score (cut-off = − 2.5), indicative of neutral effect on protein biological function (score range − 1.388–0.565). In contrast, the analysis of substitution Arg109Trp showed that it may have a deleterious effect on protein function (PROVEAN score = − 4.357, cut-off = − 2.5).

3.2. TNFRSF14 allelic variants

LD and haplotype analysis allowed the identification of six different polymorphism combinations within *TNFRSF14* exon 3. They represent six allelic variants (Fig. 1) named as follows: *ELR1e3-002*, *ELR1e3-003*, *ELR1e3-004*, *ELR1e3-005*, *ELR1e3-006* and *ELR1e3-007* (Acc. Nos.:

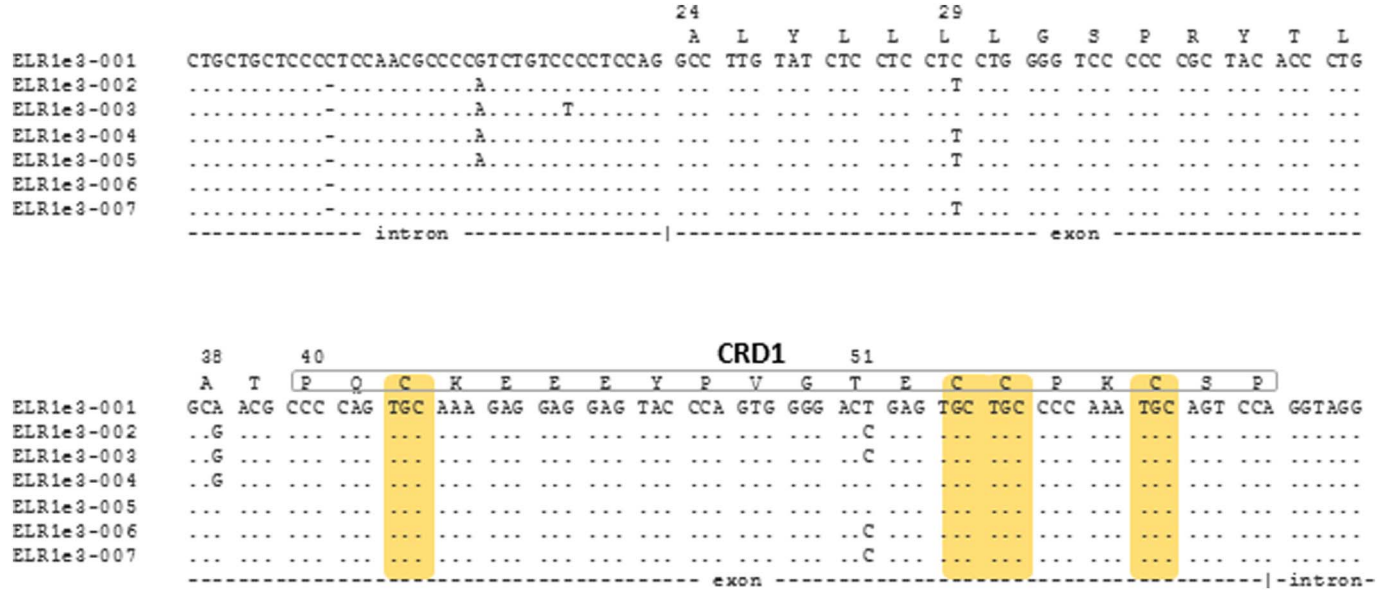

Fig. 1. ELR1 exon 3 allelic variants. Nucleotide alignment of the 109 bp corresponding to the complete ELR1 exon 3 sequence variants. The different detected nucleotide sequences were sequentially numbered as ELR1e3-002 (KT956427), ELR1e3-003 (KT956428) ELR1e3-004 (KT956429), ELR1e3-005 (KT956430), ELR1e3-006 (KT956431) and ELR1e3-007 (KT956432). Note: Accession numbers are given between brackets; predicted amino acid sequences are shown above the alignment; numbers indicate amino acid positions; the conserved cysteines within CRD1 are highlighted in yellow; exon-intron boundaries are indicated below in the alignment. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

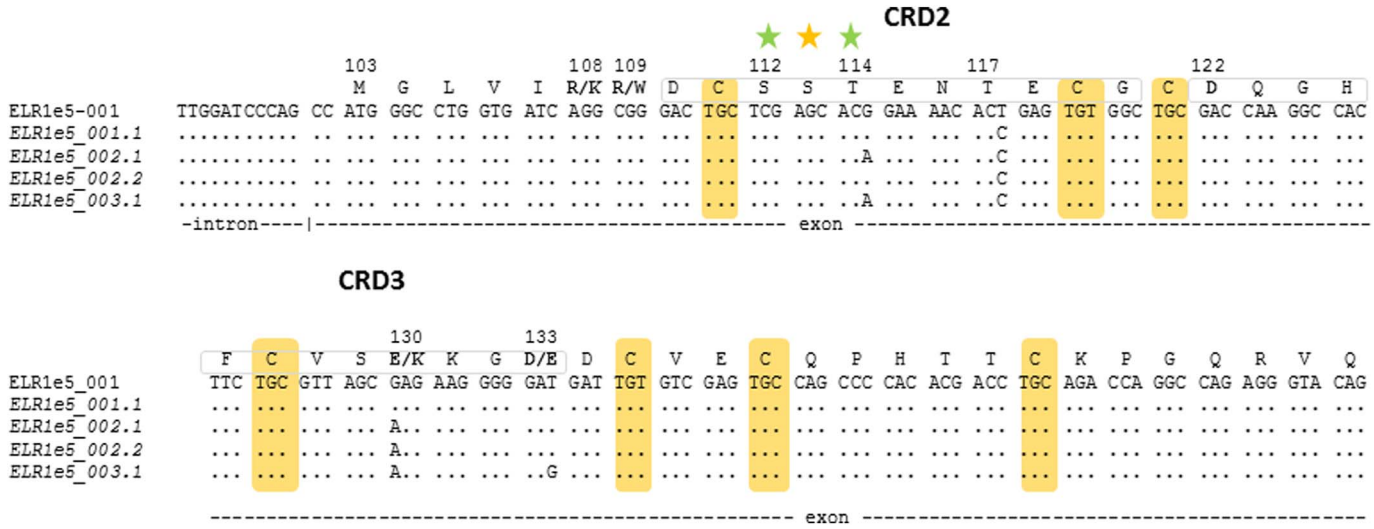

Fig. 2. ELR1 exon 5 allelic variants. Nucleotide alignment of the complete ELR1 exon 5 sequence variants. The different detected nucleotide sequences were sequentially numbered as *ELR1e5_001.1* (KT932625), *ELR1e5_002.1* (KT932626), *ELR1e5_002.2* (KT932627) and *ELR1e5_003.1* (KT932628). Note: Accession numbers are given between brackets; predicted amino acid sequences are shown above the alignment; numbers indicate amino acid positions; the conserved cysteines within CRD2 and CRD3 are highlighted in yellow; exon-intron boundaries are indicated below in the alignment. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

KT956427-32). *TNFRSF14* exon 3 allelic variants *ELR1e3-002* and *ELR1e3-007* were the most commonly detected, at frequencies of 46 and 16%, respectively (Table 2; Table S2).

The occurrence of the different allelic variants evaluated in the EIAV-positive and EIAV-negative horses showed that variants *ELR1e3-002*, *ELR1e3-006* and *ELR1e3-007* were observed in both the EIAV-positive and EIAV-negative horses, with frequencies ranging from 1 to 50%, being *ELR1e3-002* variant the most frequently found (45–50%). In contrast, variants *ELR1e3-001* and *ELR1e3-005* only occurred in EIAV-positive horses, while *ELR1e3-003* and *ELR1e3-004* were only observed in the EIAV- negative horses (Table 2).

The analysis of the eight SNPs detected in exon 5 showed eight different haplotype combinations (Table 2; Fig. S2; Table S3). However, only four of them were validated by their occurrence in several heterozygous and homozygous individuals and in different independent PCR products. These validated haplotype combinations conformed four allelic variants: *ELR1e5_001.1*, *ELR1e5_002.1*, *ELR1e5_002.2* and *ELR1e5_003.1* (Accession Numbers: KT932625-28, Table 2) (Fig. 2; Table S3). The frequency of *TNFRSF14* exon 5 allelic variant *ELR1e5_002.2* was 63% and that of the remaining variants was < 15% in the studied horses. The haplotype corresponding to the reference sequence (ENSECAE00000042908), named as *ELR1e5_001*, was not detected in the analyzed horses. Among EIAV-negative horses, all variants were detected, while in the EIAV-positive group, only *ELR1e5_002.1*, *ELR1e5_002.2* and *ELR1e5_003.1* were present. Four additional variants were estimated: *ELR1e5_008*, *ELR1e5_009* and *ELR1e5_010*, which occurred only in EIAV-negative horses at frequencies under 5%, and *ELR1e5_004*, which was only detected in EIAV-positive horses, at a frequency close to 3% (Table 2).

HWE tests showed highly significant values in both EIAV-negative

Table 1
TNFRSF14 gene SNPs detected by sequencing *TNFRSF14* exon 3, exon 5 and flanking introns. The location of each polymorphism refers to the position on ENSEMBL accession ENSECAG00000008751. * *This variant was only observed in donkeys*. For coding regions, MAF, HWE *p*-value and amino acid change are indicated. NA: Not Analyzed; **bold:** Minor Allele; rs: reference SNP on SNPdb (RefSNP); ss: Submitted SNP web report.

Region	Position	Change	Type	MAF	*p*-Value	AA change	ID
Intron 2	47309353	A/G	Noncoding	NA			ss1998444782
Intron 2	47309360	C/T*	Noncoding	NA			This study
Exon 3	47309385	T/**C**	Synonymous	0.224	0.839		ss1998444784
Exon 3	47309412	G/**A**	Synonymous	0.414	$8.09E^{-05}$		rs395939045
Exon 3	47309451	C/**T**	Synonymous	0.172	0.046		rs397088947
Intron 4	47311106	A/G	Noncoding	NA			ss1998444787
Intron 4	47311118	C/G	Noncoding	NA			ss1998444788
Intron 4	47311124	C/T	Noncoding	NA			rs397236404
intron 4	47311135	A/C	Noncoding	NA			ss1998444790
Intron 4	47311155	C/T	Noncoding	NA			ss1998444791
Intron 4	47311162	A/G	Noncoding	NA			rs395177919
Intron 4	47311175	C/T	Noncoding	NA			ss1998444791
Intron 4	47311193	A/C	Noncoding	NA			ss1998444792
Exon 5	47311282	G/**A**	Nonsynonymous	0.039	0.080	Arg108Lys	ss1998444793
Exon 5	47311284	C/**T**	Nonsynonymous	0.026	1.000	Arg109Trp	ss1998444794
Exon 5	47311295	G/**A**	Synonymous	0.013	1.000		ss1998444795
Exon 5	47311301	G/**A**	Synonymous	0.224	0.001		ss1998444796
Exon 5	47311310	C/T	Synonymous	0.000	1.000		rs68535248
Exon 5	47311323	G/**A**	Nonsynonymous	0.013	1.000	Asp122Asn	ss1998444798
Exon 5	47311347	A/**G**	Nonsynonymous	0.079	0.772	Glu130Lys	rs68535250
Exon 5	47311358	T/**G**	Nonsynonymous	0.079	0.016	Asp133Glu	ss1998444800

Table 2
Distribution of exon 3 and exon 5 allelic variant frequencies of the Equine lentivirus receptor-1 (ELR1) in EIAV negative and EIAV positive horses. The proposed allele variant name is shown for each exon, and corresponding Accession numbers are indicated, as well as the number of horses (N). Genetic and genotypic diversity was also estimated. HWE = Hardy-Weinberg equilibrium. *p-value 0.05.

ELR1 exon 3	Accession number	EIAV-negative (N = 18)	EIAV-positive (N = 11)
ELR1e3-001	ENSECAE0000042690	0.000	0.136
ELR1e3-002	KT956427	0.500	0.454
ELR1e3-003	KT956428	0.111	0.000
ELR1e3-004	KT956429	0.056	0.000
ELR1e3-005	KT956430	0.000	0.227
ELR1e3-006	KT956431	0.111	0.091
ELR1e3-007	KT956432	0.222	0.091
	He	0.740	0.692
	Ho	0.273	0.333
	*HWE**	0.000	0.000
	Gene differentiation*	0.003	
	Genotypic differentiation*	0.039	

ELR1 exon 5	Accession number	EIAV-negative (N = 25)	EIAV-positive (N = 16)
ELR1e5_001	ENSECAE0000042908	0.000	0.000
ELR1-e5_001.-1	KT932625	0.071	0.000
ELR1-e5_002.-1	KT932626	0.190	0.094
ELR1-e5_002.-2	KT932627	0.595	0.781
ELR1-e5_003.-1	KT932628	0.048	0.094
ELR1e5_004		0.000	0.031
ELR1e5_008		0.024	0.000
ELR1e5_009		0.024	0.000
ELR1e5_010		0.048	0.000
	He	0.656	0.333
	Ho	0.320	0.125
	*HWE**	0.000	0.030
	Gene differentiation	0.021	
	Genotypic differentiation	0.063	

and EIAV-positive groups for both exons. Ho and He ranged from 0.125 to 0.333 and 0.333 to 0.740, respectively. Genetic diversity was lower than expected in EIAV-positive and negative horses, in both exon 3 and exon 5, whereas gene differentiation was highly significant between groups (Table 2).

Each SNP was evaluated with single marker association analysis test. Only the previously reported SNP (rs68535250) was statistically significant ($p = 0.026$). This A > G transition is predicted to cause a change from glutamic acid to lysine (Table 1), and occurred in variants *ELR1e5_002.1*, *ELR1e5_002.2* and *ELR1e5_003.1* (Table 2, Fig. 2; Table S3).

3.3. Multispecies TNFRSF14 exon 3 and exon 5 sequences comparative analysis

Neighbor-Joining trees from nucleotide and predicted amino acid sequences of exon 3 and exon 5 of *TNFRSF14* from horses and other species showed that horse allelic variants grouped together into a unique cluster (bootstrap = 1000), including the *Equus asinus* sequence, and separated from the other species. Since the nucleotide changes detected on exon 3 were synonymous, and hence did not produce amino acid changes among the predicted protein products, the phylogenetic analysis was performed only on exon 5.

The deduced amino acid sequence of horse *TNFRSF14* exon 5 variants was compared against its ortholog in several species. The analysis involved 17 amino acid sequences and the final dataset consisted of 47 amino acid positions. Amino acid identity ranged from 94 to 98% among equine ELR1 variants (including *E. asinus*); amino acid sequence identities were higher when compared to dog, cat and guinea pig orthologs (64–79%), and lower than 55% when compared with other mammalian orthologs (data not shown). The predicted amino acid sequence comparison of *TNFRSF14* exon 5 and its orthologs from different phylogenetic origin is shown in Fig. S1 and Table S4. Results of the maximum likelihood analysis of natural codon-by-codon selection of the predicted amino acid sequences were not significant and showed dN-dS values lower than 1 ($p = 0.05$).

4. Discussion

The use of TNFRSF as virus receptors has been extensively reported (Qian et al., 2015; Hehlgans and Pfeffer, 2005), indicating that several residues are critical for the interaction and/or affinity to the surface glycoprotein of the viral particle. Under this hypothesis, the functional capabilities would be mediated by two structural regions on TNFR: CRD1 mediates the functional union by direct contact, and CRD2 influences the surface protein binding affinity. Because ELR1 acts as both the receptor of EIAV and a member of the TNFR family, the finding of ELR1 isoforms and the evidence that the soluble form of ELR1 can compete with its membrane-associated form implies the involvement of this protein in both viral infectivity and host response (Lin et al., 2013). Altogether, these observations suggest that the existence of different allelic variants could be relevant in terms of host immune response and viral infection efficiency.

In this study, we identified and evaluated *TNFRSRF14* gene polymorphisms on the exons that code for CRD1, CRD2 and part of the CRD3 domains in a group of 47 horses from different genetic backgrounds, including EIAV-positive and EIAV-negative horses. When considering *TNFRSF14* exon 3 genomic variants, genetic differentiation was significant ($p < 0.05$) when the population was splitted into EIAV-positive and EIAV-negative horses (Table 2). Moreover, none of the three SNPs on exon 3 caused amino acid changes, either in the signal peptide or the proximal sequence (~ 20 amino acids) of CRD1 (Table 1; Fig. 1). Therefore, the seven exon 3 allelic variants lead to the same polypeptide product. These results were somehow expected, taking into account the putative function of the domains encoded by exon 3, i.e., part of the signal peptide and CRD1, with its crucial role in terms of ELR1 binding capacity to EIAV gp90 (Zhang et al., 2008).

TNFRSF14 exon 5 codes for CRD2 and part of CRD3 domains. As it has been pointed out by Zhang et al. (2008), even though CRD2 is not crucial in terms of binding to EIAV gp90, the amino acid composition could have an impact on the receptor conformation and interactions with gp90 protein, hence influencing the protein binding affinity. Bearing this in mind, the occurrence of several nonsynonymous changes could be reflecting the effects of antigenic variation, according to the notion that selection pressures could be exerted by the antigen, thus leading to variations in the receptor sequence. In order to further explore whether selective pressures were acting on amino acids encoded by *TNFRSF14* exon 5, the ratio of nonsynonymous and synonymous substitutions was estimated (McDonald and Kreitman, 1991). The estimated ratio dN/dS was negative and equal to 1 in residues Ser112 and Thr114 (CRD2); thus, only synonymous changes were present in these amino acid positions. On the other hand, although substitutions Glu130Lys and Asp133Glu showed dN/dS positive values, they did not statistically exceed 1 (Table S5). It has been postulated that at genetic level, the pressure to select for changes that alter the host-virus interaction leads to a much greater likelihood that a nonsynonymous substitution will be selected than will a synonymous one (Doherty and Malik, 2012).

Analysis of the probable effects of the predicted amino acid changes

detected on exon 5 of *TNFRSF14* evidenced two kinds of consequences for nonsynonymous substitutions: i) a neutral effect on protein biological function, and ii) a probable deleterious effect on protein function. Among the latter, SNP #ss1998444794, detected in the coding sequence corresponding to CRD2, is predicted to cause the nonsynonymous change Arg109Trp, with a PROVEAN score = −4.357, which implies a high confidence that the substitution affects protein function or structure. This could be related with the different properties of the R group of these two amino acids. The change of arginine, a simple basic amino acid, to tryptophan, a non-polar and voluminous aromatic residue, and the probable consequent change in secondary structure may interfere with the domain functions. The role of this amino acid change in receptor integrity and/or function should be further investigated in relation to the functional interaction and binding to EIAV gp90, since functional changes in proteins interacting with a certain virus could result in rapid evolution and could account for differences in the response, and even the progression, to disease (Barreiro and Quintana-Murci, 2010).

This observation arises the question as to whether these variants could be related to disease progression in terms of receptor integrity and/or modification of signaling pathways; yet, additional studies in other EIAV affected and unaffected groups of horses are necessary to validate this hypothesis.

In the comparative analysis of *TNFRSF14* amino acid sequences among different mammalian species, important residues remained conserved all along the allelic variants detected in this study. For instance, CRDs were conserved among the species analyzed, with CRD2 and CRD3 containing three intradomain disulfide bonds expected for TNFRSF members. Furthermore, within the horse exon 5 variants, SNPs ss1998444795 and ss1998444796 caused synonymous substitutions in two crucial residues: Ser112 and Thr114 (Qian et al., 2015), which are proposed to form a hydrophilic interaction with gp90 by comparison to other lentivirus receptor complexes (Qian et al., 2015). In addition, several studies indicate that gp90 of the envelope protein is the predominant site of EIAV antigenic variation (Rushlow et al., 1986; Payne et al., 1987; Leroux et al., 2001). Evidence that the ELR1-binding domains of EIAV gp90 are located in the C-terminal two-thirds of EIAV gp90 is provided by the fact that whereas nucleotide substitutions or deletions in the N-terminal third of gp90 retained their receptor-binding activity, substitutions or deletions in the C-terminal two-thirds of gp90 eliminated such receptor-binding activity (Sun et al., 2008). Thus, it could be hypothesized that the structural constraints that determine the EIAV gp90 binding capacity could be reinforced or otherwise eliminated by the amino acid changes in receptor positions that are critical for their successful interaction. In this regard, even though variants *ELR1e5_002.1*, *002.2* and *003.1* were present both in seropositive and seronegative horses, genetic and genotypic differentiation were significantly different (Table 2; Table S3). Moreover, these were the only variants exhibited by EIAV-seropositive horses. The causative nucleotide nonsynonymous change G/A was present in variants only occurring in EIAV-positive horses.

As pointed out by several authors (Cheung et al., 2005; Benedict et al., 2003: Kinkade and Ware, 2006), the mechanisms that allowed some viruses to exploit and manipulate the signaling pathways transduced by TNFRSF members, regulating cell death and survival of the infected cells, acted as a strong selective pressure in the evolution of host defenses. The pattern of interspersed variable and conserved regions observed in both exons may suggest selection for variation in certain regions but not others, reflecting the division of the protein into regions of very different function. Taken together, these observations suggest that despite the occurrence of allelic variants, critical residues for the function of the mature protein are conserved, accounting for selective pressures in favor of preserving the specific function of TNFRSF members and the host immune response.

Despite the alteration of structural or expression levels of the protein itself, an interesting question to be explored is whether nonsynonymous variations in CRDs may have a role in ELR1 signaling pathways, or elucidate whether ELR1 allelic variants are related to the three major splicing variants of ELR1 mRNA or the additional isoforms identified with indels at different sites of the coding regions (Lin et al., 2013).

We noted that there are some limitations in our study. We cannot claim to have fully tagged the *TNFRSF14* gene region, and other SNPs in *TNFRSF14* may show a stronger association that may be identified by finer mapping of this gene. Moreover, our data apply only to individuals of a heterogeneous horse population. Differences in the strength or even the presence of genetic associations may or may not be found in other breeds or groups of horses exposed to EIAV.

Common genetic variants do sometimes modify the biological function of gene products or the expression levels and can be associated with susceptibility to diseases. Although equine populations sort vast genetic diversity throughout the genome, most variations may not affect the function of genes. Similarly, allelic variation in ELR1 may lead to diverse and not necessarily mutually exclusive situations: first, increase overall genomic diversity within a given horse population; secondly, decrease the variability repertoire by the presence of individuals lacking key alleles; and finally, act as a putative barrier for the interaction and binding affinity throughout viral infections.

The productive infection of target cells by animal viruses requires access to highly specific entry pathways that introduce the virion components into the cell cytoplasm for subsequent replication processes. Nonspecific entry or entry by an incorrect pathway typically results in inappropriate processing and degradation of the virion, leading to a nonproductive infection (Jin et al., 2005). The ELR1 protein appears to be sufficient for mediating productive EIAV infection in the absence of any co-receptor (Jin et al., 2005). Our results demonstrate the existence of SNPs that produce amino acid changes in residues of the protein that are postulated to have a role in the binding affinity and/or functional binding of the EIAV particle. Moreover, we detected that the amino acid variable sites are present in particular combinations (allelic variants), giving rise to the question whether individual genetic diversity at this particular receptor can contribute to the various factors that determine the success in the entry pathway utilized by EIAV in natural susceptible individuals. To our knowledge, this is the first report of the existence of allelic variations on the horse ELR1, and could be an initial step to test the possible functional relevance of EIAV entry in target cells.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.rvsc.2017.05.001.

Conflict of interest

The authors declare no conflict of interest in connection with the submitted article.

Acknowledgments

This work was supported by grants from the Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT-PICT2012-2610, -PID2007), Programa Nacional de Incentivos Docentes (Grant V206) from the Universidad Nacional de La Plata. The authors wish to thank M.V. E. Durante and J. Dayraut from SENASA. Thanks are also due to A. Di Maggio for manuscript correction.

References

Barreiro, L.B., Quintana-Murci, L., 2010. From evolutionary genetics to human immunology: how selection shapes host defense genes. Nat. Rev. Genet. 11, 17–30.
Barrett, J.C., Fry, B., Maller, J., Daly, M.J., 2005. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21, 263–265.
Benedict, C.A., Banks, T.A., Ware, C.F., 2003. Death and survival: viral regulation of TNF signaling pathways. Curr. Opin. Immunol. 15, 59–65.
Bodmer, J.L., Schneider, P., Tschopp, J., 2002. The molecular architecture of the TNF

superfamily. Trends Biochem. Sci. 27, 19–26.
Cheung, T.C., Humphreys, I.R., Potter, K.G., 2005. Evolutionarily divergent herpesviruses modulate T cell activation by targeting the herpesvirus entry mediator cosignaling pathway. Proc. Natl. Acad. Sci. U. S. A. 102, 13218–13223.
Choi, Y., Chan, A.P., 2012. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics btv195.
Cook, R.F., Montelaro, R.C., Issel, C.J., 1996. Equine infectious anemia. In: Studdert, M.J. (Ed.), Virus Infections of Equines. Elsevier, Amsterdam, pp. 297–323.
Felsenstein, J., 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17 (6), 368–376.
Gallaher, W.R., Ball, J.M., Garry, R.F., Martin-Amedee, A.M., Montelaro, R.C., 1995. A general model for the surface glycoproteins of HIV and other retroviruses. AIDS Res. Hum. Retrovir. 11, 191–202.
Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 41, 95–98.
Hehlgans, T., Pfeffer, K., 2005. The intriguing biology of the tumour necrosis factor/tumour necrosis factor receptor superfamily: players, rules and the games. Immunology 115, 1–20.
Jin, S., Zhang, B., Weisz, O.A., Montelaro, R.C., 2005. Receptor-mediated entry by equine infectious anemia virus utilizes a pH-dependent endocytic pathway. J. Virol. 79, 14489–14497.
Kimura, M., 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16 (2), 111–120.
Kinkade, A., Ware, C.F., 2006. The DARC conspiracy–virus invasion tactics. Trends Immunol. 27, 362–367.
Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol. Biol. Evol msw054.
Leroux, C., Craigo, J.K., Issel, C.J., Montelaro, R.C., 2001. Equine infectious anemia virus genomic evolution in progressor and nonprogressor ponies. J. Virol. 75, 4570–4583.
Lin, Y.Z., Yang, F., Zhang, S.Q., Sun, L.K., Wang, X.F., Du, C., Zhou, J.H., 2013. The soluble form of the EIAV receptor encoded by an alternative splicing variant inhibits EIAV infection of target cells. PLoS One 8, e79299.
Locksley, R.M., Killeen, N., Lenardo, M.J., 2001. The TNF and TNF receptor superfamilies: integrating mammalian biology. Cell 104 (4), 487–501.
McDonald, J.H., Kreitman, M., 1991. Adaptive protein evolution at the Adh locus in Drosophila. Nature 351, 652–654.
Muse, S.V., Gaut, B.S., 1994. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. Mol. Biol. Evol. 11 (5), 715–724.
Payne, S.L., Fang, F.D., Liu, C.P., Dhruva, B.R., Rwambo, P., Issel, C.J., Montelaro, R.C., 1987. Antigenic variation and lentivirus persistence: variations in envelope gene sequences during EIAV infection resemble changes reported for sequential isolates of HIV. Virology 161, 321–331.
Qian, L., Han, X., Liu, X., 2015. Structural insight into equine lentivirus receptor 1. Protein Sci. 24, 633–642.
Rousset, F., 2008. Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. Mol. Ecol. Resour. 8, 103–106.
Rushlow, K., Olsen, K., Stiegler, G., Payne, S.L., Montelaro, R.C., Issel, C.J., 1986. Lentivirus genomic organization: the complete nucleotide sequence of the env gene region of equine infectious anemia virus. Virology 155, 309–321.
Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4 (4), 406–425.
Sellon, D.C., 1993. Equine infectious anemia. Vet. Clin. North Am. Equine Pract. 9, 321–335.
Sentsui, H., Kono, Y., 1987. Complement-mediated hemolysis of horse erythrocytes treated with equine infectious anemia virus. Arch. Virol. 95, 53–66.
Smith, C.A., Farrah, T., Goodwin, R.G., 1994. The TNF receptor superfamily of cellular and viral proteins. Activation, costimulation, and death. Cell 76, 959–962.
Spyrou, V., Papanastassopoulou, M., Psychas, V., Billinis, Ch., Koumbati, M., Vlemmas, J., Koptopoulos, G., 2003. Equine infectious anemia in mules: virus isolation and pathogenicity studies. Vet. Microbiol. 95, 49–59.
Sun, C., Zhang, B., Jin, J., Montelaro, R.C., 2008. Binding of equine infectious anemia virus to the equine lentivirus receptor-1 is mediated by complex discontinuous sequences in the viral envelope gp90 protein. J. Gen. Virol. 89 (8), 2011–2019.
Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30, 2725–2729.
Zhang, B., Jin, S., Jin, J., Li, F., Montelaro, R.C., 2005. A tumor necrosis factor receptor family protein serves as a cellular receptor for the macrophage-tropic equine lentivirus. Proc. Natl. Acad. Sci. U. S. A. 102, 9918–9923.
Zhang, B., Sun, C., Jin, S., Cascio, M., Montelaro, R.C., 2008. Mapping of equine lentivirus receptor 1 residues critical for equine infectious anemia virus envelope binding. J. Virol. 82, 1204–1213.