

# Genomic Studies and genetic characterization of equid herpesvirus type 1 (EHV-1) strains: estimating the global similarity by distance methods.

Giselle P. Martín Ocampos<sup>1, 2,3</sup>, Claudio G. Barbeito<sup>2,3</sup>, Cecilia M. Galosi<sup>1,4</sup>

<sup>1</sup>Department of Virology, <sup>2</sup> Histology and Embryology. School of Veterinary Sciences, 60 & 118, P.O.BOX 296,1900, La Plata, Buenos Aires, Argentina, <sup>3</sup> National Research Council (CONICET), Rivadavia 1917, Buenos Aires, Argentina, <sup>4</sup> Scientific Research Commission (CIC-PBA), 526 & 11, La Plata, Buenos Aires, Argentina.

**Equid Herpesvirus type 1 (EHV-1)** is a member of the genus *Varicellovirus*, which belongs to the subfamily *Alphaherpesvirinae*, family *Herpesviridae*. EHV-1 contains five non-coding regions of variable sizes. One of them, called intergenic region (IR), is located between ORF 62 and ORF 63. This region was amplified by polymerase chain reaction and its products were sequenced. We analyzed 25 EHV-1 strains isolated from different geographical areas. Genomic analysis showed a high percentage of identity of the strains of the intergenic region. In 13 Argentinean strains, 18-bp insertions were found located in the 109168 bp position, whereas, in the AR6 strain, a 9-bp insert was found in the same position. The strains studied were classified using global similarity techniques and the analysis of genetic distances revealed that seven strains are grouped with foreign strains. Within this group a subgroup comprising an Argentinean and a British strain was identified. No relationship between the isolation years and the geographical origins of EHV-1 strains was found.

**Keywords:** Genomic characterization, global similarity, EHV- 1

## INTRODUCTION

*Equid Herpesvirus type 1* (EHV-1) is a member of the genus *Varicellovirus*, which belongs to the subfamily *Alphaherpesvirinae*, family *Herpesviridae*. Within this genus, 17 species are distinguished, among them: *Equid Herpesvirus type 1*, *Equid Herpesvirus type 4*, *Bovine Herpesvirus*, and *Human Herpesvirus 3*.

The EHV-1 genome measures approximately 150 kbp, contains 80 genes, 72 of which correspond to genes of unique copy and 4 to duplicate genes, and has 56% G-C content. Its molecular weight is 92x10<sup>6</sup> Da, and its buoyant density in sucrose and potassium tartate is

1.18s g/ cm<sup>3</sup>.

The double chain of linear DNA is composed of two regions covalently bound, one called L ("long"), and another one called S ("short"). Region S is flanked by two inverted repeat regions: an internal repeat sequence (IRS) and a terminal repeat sequence (TRS), which allow the generation of two equimolar isomers, whereas region L consists of a unique sequence flanked by small inverted repeats (IRS and TRS)

Five non-coding regions of variable size have been identified in the EHV-1 genome: a) the first region is approximately 1kbp and is located on the right side at the end of the viral genome, b) the second region of 1kbp is located between the open reading frame (ORF) 39 and ORF 40, c) the third region is approximately 1.5 kbp and is located between ORF 62 and ORF 63, d) the fourth region of 2 kbp is located between ORF 63 and ORF 64

in a 1.5 kbp union U<sub>L</sub>/IR<sub>s</sub>, and e) the fifth region of 2.5 kbp is located between ORF 64 and ORF 65. (Ibrahim et al. 2004).

The intergenic region (IR) located between ORF 62 and ORF 63 has two variable and two conserved domains. The conserved domains are found between positions 108803 and 108945 (domain 1) and 109551 and 110286 (domain 3); whereas, the variable domains are found between positions 108946 and 109550 (domain 2) and 110287 and 110385 (domain 4). The variable domain located between positions 108946 and 109550 has insertions and deletions in Japanese strains, as compared with the British strains Ab4. The insertions include a repetitive sequence of 18 bp. The changes between ORF 62 and ORF 63 would have a fundamental role in virulence and viral growth (Ibrahim et al. 2004).

Early epidemiological molecular studies on EHV-1 were based on the analysis of DNA fragments from different strains (Allen, 1983). With this methodology, more than 16 electropherotypes were identified as comprising two genotypes: EHV-1P and EHV-1B (Allen et al. 1985). EHV-1P differs from EHV-1B in the number of identified sites of the restriction enzyme *Bam*HI in the IRS and TR regions; EHV-1P has two sectioning sites in the IRS region and other two in the TR region, whereas EHV-1B has only one site *Bam*HI, in each of these regions.

Analysis of 69 isolated strains from the year 1979 showed that only two of them correspond to electropherotypes 1B. From these results, we hypothesize that there has been genomic heterogeneity among Argentinean isolates since 1996 (Martinez et al. 2006).

In Argentina, EHV-1 was first isolated from an aborted fetus in 1979 (Etcheverrigaray et al. 1982). In 1985, the virus was isolated again from plasma rich in leukocytes drawn from an animal with respiratory signs (Galosi et al. 1989). Subsequently, numerous viral strains have been isolated. However, differences related to their abortive potential have not yet been studied.

Whole genome sequencing of strains that cause neurological disease show a point mutation in the 2254 site of the genomic region that encodes the DNA polymerase (ORF 30), resulting in Guanine (G<sub>2254</sub>) in neuropathogenic strains compared to Adenine (A<sub>2254</sub>) in non-neuropathogenic ("wild type") strains (Nugent et al. 2006).

The latest studies of Argentinean strains have also shown that the neuropathogenic biotype (G<sub>2254</sub>) has been present in the country since at least 1996 and that it is significantly associated with the clinical manifestation of the neurological disease (Vissani et al. 2008).

In this work, we aimed to study the intergenic region located between ORF 62 and ORF 63 in Argentinean, Japanese, American and British EHV-1 strains and to establish global similarity that will allow us to genetically

characterize them by the use of genetic distance methods, specifically the neighbor-joining algorithm.

## MATERIALS AND METHODS

### Viral isolation

The strains were originally isolated in the cell line RK13 (from rabbit kidney) or in primary cultures of equine fetal kidney. To this end, we used between two and five passages of each of the strains in RK13 cells developed in minimal essential medium (MEM) supplemented with 10% fetal bovine serum (FBS) as growth medium (MEM-G) or with 2% of FBS as maintenance medium (MEM-M).

#### 2.2 DNA extractions.

Extraction of total genomic DNA from the viral-infected RK13 cells was as described by Galosi et al. (1998) with each of the isolated strains.

Cells were infected with different strains of the virus until a multiplicity of infection (MOI) of 1 was obtained, and then left in adsorption for 1 h at 37°C in a CO<sub>2</sub> atmosphere and incubated in the same conditions with MEM-M.

At about 18-24 h postinoculation pi and when the cytopathogenic effect was ~ 80%, cells were washed twice with phosphate buffer NaCl 0.13 M, KCl 0.002 M, Na<sub>2</sub>HPO<sub>4</sub> 0.008 M, KH<sub>2</sub>PO<sub>4</sub> 0.0014 M – PBS-). Later, they were digested with Proteinase K in a final concentration of 0.2 mg/ml in PK buffer [100 mM of Tris-HCl pH 7.5, 12.5 mM of ethylenediaminetetracetic acid (EDTA) pH 8.0, 150 mM NaCl and 1% sodium dodecyl sulfate (SDS)]. We then carried out the first extraction with saturated phenol in TE buffer (10 mM Tris-HCl pH 8.0 and 1 mM EDTA pH 8.0) and a second extraction with a phenol-chloroform-isoamyl alcohol mixture (25:24:1). DNA was precipitated by adding two volumes of ethanol, washed with 70% ice-cold ethanol, and resuspended in sterile distilled water.

### PCR amplification of the genomic locus under study.

The genomic locus corresponding to the intergenic region (IR) located between ORF 62 and ORF 63 (108486-110681 positions; GenBank accession number AY 665713) was amplified using four pairs of primers:

SF1: 5' CCG GTC GTT CGG TTG AGC AAG TTT TTG ATG 3';  
SR1: 5' CCT CCA GTC CAC AGA TAT GAC ATC CAA AGG 3'.  
SF2: 5' ACC GGA AGC TTG TCA TAT TTG TGA GCC TGG 3';  
SR2: 5' TGT GAA CAT CAC CAC CAA TAC CAA GCA CGG 3'.  
SF3: 5' CC AAT TAG CCC CCA ATT GGC ACA TGG TAA 3';  
SR3: 5' TTA CAA AAA CCT ATG CAG GGG TGT GGG TGG 3'.  
SF4: 5' TTC CCC CGG GCC TTA TAT CTT GCA GCT TTA 3';  
SR4: 5' TTG TTT TAG TCG ACC GAA GCT CTG AGG GAG 3'.

Final PCR amplification volume of 25 µl consisted of 2 µl DNA, 1.5 µl of MgCl<sub>2</sub> (25 mM), 2.5 µl PCR 10X buffer, 1 µl of DNA *Taq* Polymerase, 2 µl of dNTP mixture (0.2 mM of each), 1 µl of each primer (20 pM of each) and 12 µl of sterile distilled water. DNA was amplified with the following program: a) initial denaturation of 94°C for 4 min, b) 30 cycles of 30 sec at 94°C, 20 sec at 60°C and 60 sec at 72°C, and c) final extension of 4 min at 72°C.

The amplification products obtained were analyzed by electrophoresis on 1.5% agarose gel stained in ethidium bromide solution. The sizes of the amplified fragments were compared with a molecular weight marker. PCR products were stored at -20°C. Purified products were analyzed and quantified by electrophoresis

in agarose gel using a molecular weight marker as size standard.

### Analysis of amplified sequences.

Sequences corresponding to 25 viral strains from different geographical regions were analyzed. Twenty one of the strains were isolated from different places of Argentina. The strains JA and US were isolated in Japan and the USA respectively. Sequences corresponding to the Ab4 and V592 strains were obtained from GenBank. Ab4 and V592, herein called UK 1 and UK 2 respectively, were isolated in the United Kingdom (Table 1).

The sequence of strain NS80567, belonging to the species *Equid Herpesvirus 4* (**Accession number to Genbank: NC 001844**), was used as 'outgroup'.

The sequences were manually edited using text processors and formatted using Bio-Edit (Hall, 1999), Proseq 3.0 (Filatov, 2002), and Genedoc (Nicholas et al. 1997) software.

Diversity in the genetic blueprint and the percentage of identity among EHV-1 strains were calculated using Swaap version 1.0.3 (Pride, 2000).

Sequence alignment was performed using Clustal X software (Thompson et al. 1997), with the following parameters: elimination of gaps before isolation (*reset all gaps before alignment*), GOP (*gap opening penalty*) = 15 and GEP (*gap extension penalty*) = 6.66. The two latter parameters established the penalty for opening the gap and for gap extension respectively.

Global similarity was analyzed by constructing distance matrices, and a dendrogram built using the neighbor-joining algorithm (Saitou and Nei, 1987).

## RESULTS

The genomic analysis of the intergenic region located between ORF62 and ORF 63 showed a high percentage of identity among strains. In thirteen of the Argentinean strains (AR8, AR7, AR4, AR2, AR3, AR10, AR9, AR18, AR20, AR15, AR19, AR21 and AR22), 18-bp insertions located in the 109168 bp position were found, whereas, in strain AR6, a 9-bp insertion was found in the same position (Figure 1).

The other sectors of this region analyzed presented a high percentage of identity among strains. The genomic sectors located between positions 108803 and 110385 presented a high percentage of identity (between 99% and 100%).

Data from these genomic loci were used in phylogenetic analysis.

The dataset presented 1510 sites, of which only 13 were variable. The nucleotide composition of this locus was T= 27.4 %, A= 26.1 %, C= 21.3 %, and G= 25.2 %.

The percent identity (percentage of identical nucleotides along the studied sector) among Argentinean isolates, between Argentinean and foreign isolates and among foreign isolates varied between 99% and 100%.

The analysis of genetic distances revealed that the Argentinean strains AR1, AR14, AR11, AR16, AR17, AR6 and AR12 are grouped with the foreign strains JA,

UK1 and UK2. The analysis showed that a subgroup made of the Argentinean strain AR6 and the British strain UK2 can be identified within this group (Figure 2). The isolates included in this subgroup were isolated from animals with different signs.

Global similarity among Argentinean isolates and global similarity among foreign isolates ranged between 99% and 100%.

## DISCUSSION

The genetic characterization of the strains and the association of the region analyzed with changes in virulence of the strains analyzed in the experimental model allowed us to infer that the intergenic region (IR) located between ORF 62 and ORF 63 has an identity higher than 99% among strains.

Strains AR8, AR7, AR4, AR2, AR3, AR10, AR9, AR18, AR20, AR15, AR19, AR21, and AR22 showed 18-bp insertions, whereas AR6 showed 9-bp insertions.

Ibrahim and his collaborators (2004) described four domains of the IR located between ORF 62 and ORF 63. In their work, they describe the first and third domains as conserved, and domains 2 and 4 as variable. In the strains analyzed by these authors, the second domain had 18-bp insertions, whereas, in our case, the second domain had 18-bp insertions in 13 strains and a 9-bp insertion in one strain. The remaining nine strains showed no genetic mutations in this locus. As regards the fourth domain, the strains analyzed in this work presented a high percentage of conservation. According to our analysis, Argentinean and British strains, together with the US and JA strains, did not show variable domains. These strains had four highly conserved domains.

The classification of the strains studied using global similarity techniques and the analysis of genetic distances revealed that seven strains are grouped with foreign strains. Within this group, a subgroup composed of an Argentinean and a British strain (both non-producers of neurological signs) was identified. We observed no relationship between the isolation years and the geographical origin of the strains. It is important to point out that the only existing background on genetic characterization of EHV-1 was carried out with a small number of strains and with the aim to reconstitute the evolutionary history of the group. Another important aspect is that the use of distance methods, as is the case of the neighbor-joining algorithm, is controversial since the conversion of discrete data (as are the DNA sequences in matrices of genetic distances) is not recommended for a phylogenetic reconstruction, since they do not usually represent the evolutionary history of the group (Farris, 1981). However, the methods of genetic distances, specifically the neighbor-joining

**Table 1.** EHV-1 isolations used in the analysis.

| <b>Isolation</b> | <b>Origin</b>        | <b>Biotype</b> | <b>Date and place of isolation</b>                  | <b>AN<br/>(Accession<br/>Number )</b> |
|------------------|----------------------|----------------|-----------------------------------------------------|---------------------------------------|
| <b>AR 1</b>      | Aborted Fetus        | NNP            | Galosi (1979). La Plata. Argentina                  | <b>EU 366292</b>                      |
| <b>AR 2</b>      | Rhinopneumonitis     | NNP            | Galosi (1979). La Plata. Argentina                  | <b>EU 366293</b>                      |
| <b>AR 3</b>      | Aborted Fetus        | NNP            | Galosi (1990). 25 de Mayo. Argentina                | <b>EU 366294</b>                      |
| <b>AR 4</b>      | Aborted Fetus        | NNP            | Galosi (1990).Buenos Aires. Argentina               | <b>EU 366295</b>                      |
| <b>AR 6</b>      | Aborted Fetus        | NNP            | Galosi (1990).Tucumán. Argentina                    | <b>EU366296</b>                       |
| <b>AR 7</b>      | Aborted Fetus        | NNP            | Galosi (1990).Capitán Sarmiento. Argentina          | <b>EU 366297</b>                      |
| <b>AR 8</b>      | Aborted Fetus        | NNP            | Galosi (1996). Magdalena. Argentina                 | <b>EU 366298</b>                      |
| <b>AR 9</b>      | Aborted Fetus        | NNP            | Galosi (2000) La Pampa. Argentina                   | <b>EU 366299</b>                      |
| <b>AR 10</b>     | Aborted Fetus        | NNP            | Galosi (2002). San Antonio de Areco. Argentina.     | <b>EU 366300</b>                      |
| <b>AR 11</b>     | Aborted Fetus        | NNP            | Barrandeguy (2004). San Antonio de Areco. Argentina | <b>EU 366301</b>                      |
| <b>AR 12</b>     | Aborted Fetus        | NNP            | Galosi (2001). Trenque Lauquén. Argentina           | <b>EU 366302</b>                      |
| <b>AR 13</b>     | Aborted Fetus        | NNP            | Barrandeguy (1997). General Villegas. Argentina     | <b>EU 366303</b>                      |
| <b>AR 14</b>     | Neurological Disease | NNP            | Barrandeguy (1998). Pilar. Argentina.               | <b>EU 366304</b>                      |
| <b>AR 15</b>     | Neonatal Disease     | NNP            | Barrandeguy (1999). San Antonio de Areco. Argentina | <b>EU 366305</b>                      |
| <b>AR 16</b>     | Aborted Fetus        | NNP            | Galosi (1997). Entre Ríos. Argentina                | <b>EU 366306</b>                      |
| <b>AR 17</b>     | Neonatal Disease     | NNP            | Galosi (1999). Cañuelas. Argentina.                 | <b>EU 366307</b>                      |
| <b>AR 18</b>     | Aborted Fetus        | NNP            | Barrandeguy (1999). General Pueyrredón. Argentina   | <b>EU 366308</b>                      |
| <b>AR 19</b>     | Aborted Fetus        | NNP            | Barrandeguy (2000). General Pueyrredón. Argentina.  | <b>EU 366309</b>                      |
| <b>AR 20</b>     | Aborted Fetus        | NNP            | Barrandeguy (1999). San Antonio de Areco. Argentina | <b>EU 366310</b>                      |
| <b>AR 21</b>     | Aborted Fetus        | NNP            | Galosi (2001). Córdoba. Argentina                   | <b>EU 366311</b>                      |
| <b>AR 22</b>     | Aborted Fetus        | NNP            | Galosi (1998). Trenque Lauquen. Argentina           | <b>EU 366312</b>                      |
| <b>JA</b>        | Aborted Fetus        | NNP            | Kawakami (1970).Japan                               | <b>EU 366313</b>                      |
| <b>US</b>        | Aborted Fetus        | NNP            | Doll (1954). USA                                    | <b>EU 366314</b>                      |
| <b>UK 1</b>      | Neurological Disease | NP             | Crowhurst( 1981) United Kingdom                     | <b>AY 665713</b>                      |
| <b>UK 2</b>      | Abortion Storms      | NNP            | Mumford( 1987) United Kingdom.                      | <b>AY 464052</b>                      |

algorithm, are of great use to establish global similarity among strains, because they calculate the global similarity by establishing a disparity index among strains. In summary, this study focuses on the genetic characterization of the IR between ORF 62 and ORF 63 which would be related to virulence. The strains analyzed in this study showed the absence of four well-defined domains, as all have a high degree of conservation and

variability observed corresponds only to the presence of insertions and deletions in the strains. The overall similarity analysis showed a high percentage of identity in the group, which confirms the high degree of conservation of the locus analyzed. This is the first study that uses the neighbor-joining algorithm to establish global similarities between strains of EHV-1.

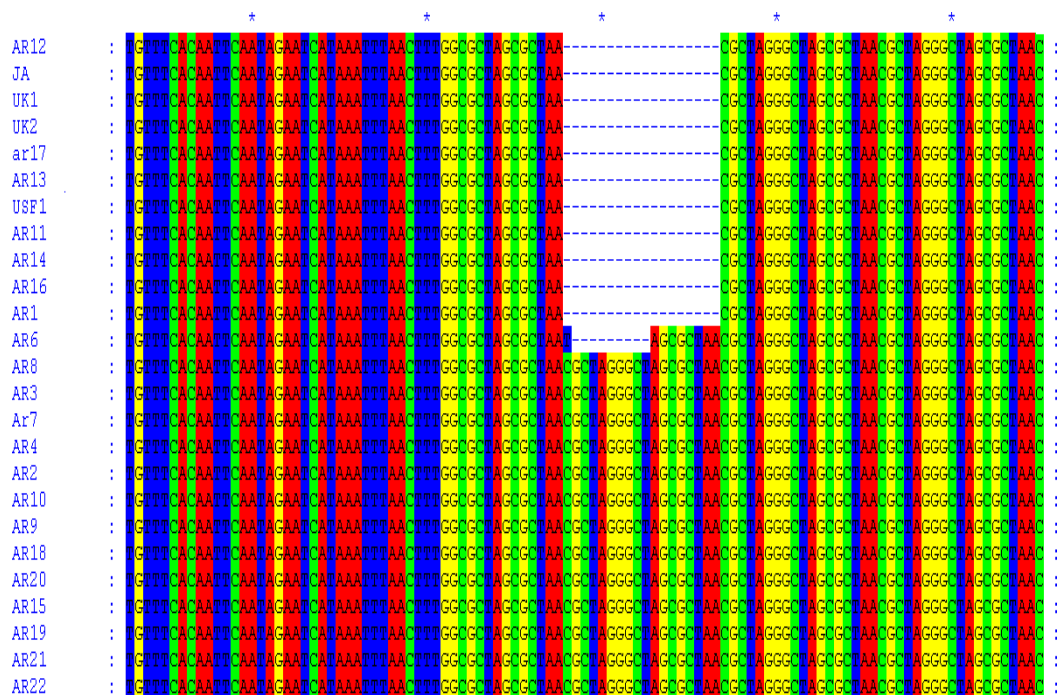


**Figure 1.** Alignment of 25 EHV-1 sequences corresponding to the genomic sector flanked between positions 108803 and 109168. Deletions appear with the mark (-).

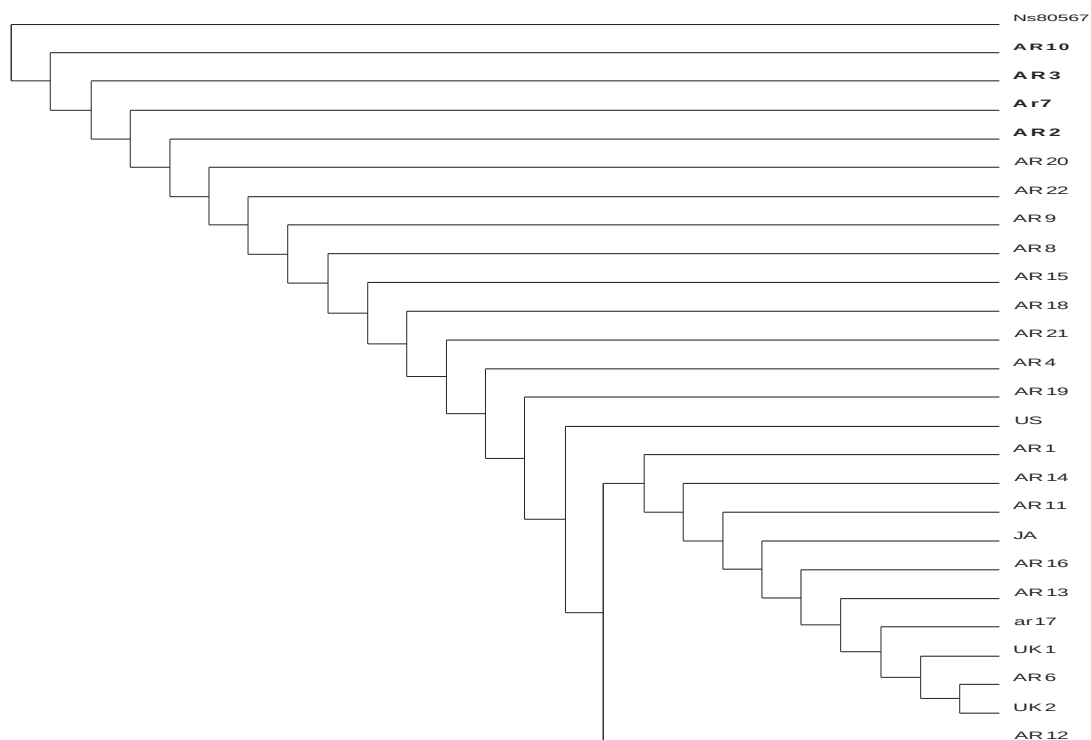


**Figure 2.** Dendrogram obtained by the distance calculus of Hasegawa, Kishino and Yano (HKY 85) using the Neighbor-Joining algorithm (Saitou and Nei, 1987).

## ACKNOWLEDGEMENTS

This research was supported by grants from National University of La Plata and CIC-PBA. GPMO, NAF, MLE are fellows from CONICET (National Research Council). CGB and CMG are Career researchers of CONICET and CIC-PBA, respectively.

## REFERENCES

- Etcheverrigaray ME, Oliva, GA, Gonzalez ET, Nosetto EO, Martin AA (1982). Comportamiento de una cepa de HVE-1 aislada de un feto abortado. *Rev. Mil Vet* 30: 138-139.
- Farris JS (1981). Distance data in phylogenetic analysis. In Funk VD & Brooks DR (eds). *Advances in Cladistics. Proceedings of the First Meeting of the Willi Hennig Society*, New York, Botanical Garden ; New York.
- Filatov DS (2002) Proseq: A software for preparation and evolutionary analysis of DNA sequence data set. *Molec. Ecology Notes* 2: 621-624.
- Galosi CM, Nosetto EO, Gimeno EJ, Gomez Dunn C, Etcheverrigaray ME, Ando Y (1989). Equine Herpesvirus 1 (EHV-1): Characterisation of a viral strain isolated from e quine plasma in Argentina. *Rev Sci Tech Int Epizz Vol* 8 (1): 117-122.
- Galosi CM, Echeverria MG, Vila roza MV, Cid de la Paz V, Oliva GA, Etcheverrigaray ME (1998) Virus herpes equino tipo 1 (EHV-1): patrones de restricción de ADN, perfiles proteicos y estudios de patogenicidad en ratones. *Analecta Veterinaria* 18 ½: 35-40
- Hall TA (1999). Bio-Edit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Series*. 41 95-98.
- Ibrahim el SM, Pagmajav O, Yamaguchi T, Matsumura T, Fuskushi H (2004) Growth and virulence alterations of equine herpesvirus 1 by insertion of a green fluorescent protein gene in the intergenic region between ORF 62 and ORF 63. *Microbial Immunol* 48 (11): 831-842.
- Martinez JP, Martin Ocampos GO, Fernandez LC, Fuentealba NA, Cid de la Paz V, Barrandeguy M, Galosi CM (2006). First detection of equine herpesvirus genome 1B in Argentina. *Rev Sci Tech Off Int Epizz* 25(3): 1075-1079.
- Nicholas KB, Nicholas HB, Deerfield DW( 1997) Genedoc :Analysis of Genetic variation EMBNEW NEW 4:14.
- Nugent J, Birch-Machin I, Smith KC, Mumford JA, Swann Z, Newton JR, Bowden RJ, Allen GP,, Davis-Poynter N (2006). Analysis of equid herpesvirus strain variation reveals a point mutation of the DNA polymerase strongly associated with neuropathogenic versus non-neuropathogenic disease outbreak. *J Virol* 80: 4047-4060.
- Pagamjav O, Sakata T, Matsumura T, Yamaguchi T, Fukushi H (2005) Natural recombinant between Equine Herpesvirus 1 and 4 in the ICP 4 gene. *Microbiol Immunol* 49(2): 167-179.
- Pride DT (2000) SWAAP: a tool for analyzing substitutions and similarity in multiple alignments. Distributed by the author.
- Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic tree. *Mol Biol Evol* 4: 406-425.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The Clustal X: windows interface flexible strategies for multiple sequences alignment aided by quality analysis and tools. *Nucleic Acids Research* 25: 4876-4882.
- Vissani Ms, Miño S, Becerra L, Tordoya MS, Olgion Prigliore C, Barrandeguy M (2008) Identificación del biotipo neuropatogénico de Herpes virus equino en fetos abortados en la República Argentina. *Rev Arg Microbiol* 40(1):98.