

# Isolation, sequencing and phylogenetic analysis of the hemagglutinin, neuraminidase and nucleoprotein genes of the Chilean equine influenza virus subtypes H7N7 and H3N8

<sup>1</sup>ILSE MÜLLER, <sup>4</sup>BELTRÁN JAUREGUIBERRY and <sup>1,2,3</sup>PABLO D.T. VALENZUELA

<sup>1</sup> Bios Chile IGSA, <sup>2</sup> Fundación Ciencia para la Vida and <sup>3</sup> Instituto Milenio MIFAB. Av. Zañartu 1482, Santiago, Chile <sup>4</sup> Veterquímica S.A. Av. Pedro Aguirre Cerda 5461, Santiago, Chile

## ABSTRACT

We report here on the isolation and sequencing of the hemagglutinin, neuraminidase and nucleoprotein genes of the Chilean equine influenza virus subtypes H7N7 (A/equi-1/Santiago/77, Sa77) and H3N8 (A/equi-2/Santiago/85, Sa85). The sequences obtained allowed a variability analysis, which indicated significant differences when compared with other isolates. We found that Chilean isolates are more similar to the North American variety than to European isolates. Isolate Sa77 is a good candidate for inclusion in a vaccine as it is the latest isolate of the subtype H7N7 and is probably better-adapted to the equine host. Isolate Sa85, of subtype H3N8, also appears to be a good candidate since it has no significant differences in the main antigenic sites with recent isolates.

**Key terms:** equine influenza, hemagglutinin, neuraminidase, nucleoprotein genes

## INTRODUCTION

Viral equine influenza is one of the major diseases to affect horses worldwide [6, 16, 35, 36] and requires a great deal of veterinary attention [31]. The disease is caused by the equine influenza type A virus, an RNA virus belonging to the *Orthomyxoviridae* family. The viral RNA consists of eight independent segments that allow easy recombination, which gives rise to new viral subtypes [12, 27]. In horses there are two subtypes of influenza A virus. H3N8 (A/equine-2, first recognized in Miami in 1963), which is the most commonly isolated, and subtype H7N7 (A/equine-1, first described in Prague in 1956), which is more rarely

isolated [13, 18, 25]. In Chile, both subtypes, H7N7 (A/equi-1/Santiago/77, [4]) and H3N8 (A/equi-2/Santiago/85, [3]) have been isolated.

Equine influenza has severe consequences; sick animals remain incapacitated for up to four weeks. In addition they require a training period to return to competitive levels. Pregnant mares may abort or suffer fetal reabsorption during the first months of pregnancy.

In Chile, the disease has emerged in both summer and winter months at irregular intervals of four to eight years. As a consequence of the lack of preventive measures or of poorly-planned vaccination programs, assorted outbreaks have caused

heavy damage to livestock throughout the country.

One of the evasion mechanisms of the viruses is the antigenic drift, which allows evasion from the host's immune system. As a consequence, each subtype carries significant antigenic changes that allow the virus to persist for several years.

Because of the high morbidity and economic losses associated with outbreaks of equine influenza, intense vaccination programs are employed in an effort to control infection. However, the vaccines currently available are based on inactivated whole virus and only offer a limited short-term protection. Outbreaks of the disease have occurred in vaccinated horses, although they have presented milder symptoms as compared with non-vaccinated animals. Vaccine failure is primarily attributed to viral antigenic changes [34], and new approaches to vaccination are therefore needed.

As a first step toward the development of vaccines for the Chilean and South American markets we are isolating viral genes from the Chilean subtypes. This report describes the isolation and sequence features of the hemagglutinin (HA), neuraminidase (NA) and nucleoprotein (NP) coding regions of both subtypes. The purpose of this study is to understand the bases of the antigenic changes of the virus from this region in order to better design an effective vaccine.

## MATERIALS AND METHODS

### 1. Virus culture and isolation of viral RNA

Viruses from isolates of A/equi-1/Santiago/77 (Sa77) and A/equi-2/Santiago/85 (Sa85)

were cultured in embryonated 10 or 11-day-old chicken eggs as described elsewhere [14]. The isolates we used were maintained through a process involving multiple viral transfers. This is particularly relevant in the case of isolate Sa85, whose RNA exhibited sequence differences compared with the original sequence reported by Kawaoka et al. [11]. The isolate from which the HA gene was sequenced in this paper is named Santiago 85b (Sa85b). Viral RNA was extracted from the allantoic fluid using Trizol LS (Invitrogen) according to Chomczynski [5] as described previously [30]. This RNA was used to amplify and isolate relevant genes by RT-PCR using random hexanucleotides as primers according to Sambrook and Russell [23].

### 2. Amplification and cloning

Genes were amplified by PCR using primers designed after the conserved regions of the desired sequences (Table I) and cloned into vector pCR2.1 (Invitrogen). Recombinant plasmids were selected after analysis by PCR and digestion with endonucleases EcoRI, Sall and XbaI. Plasmid DNA was prepared as described previously [15].

### 3. Sequencing

Relevant genes were sequenced according to Sanger et al. [24] using the BigDye Terminator v1.1 Cycle Sequencing kit (Applied Biosystems), processed by capillary electrophoresis and visualized in an ABI Prism 310 Genetic Analyzer. The nucleoprotein sequences obtained are registered in the Genbank database under

TABLE I.

Oligonucleotides used as primers to amplify the nucleoprotein, neuraminidase and hemagglutinin genes by RT-PCR.

Gene	Oligo Forward (5'→3')	Oligo Reverse (5'→3')
NP	ATGGCGTCTCAAGGCACCAAA	TTAACTGTCAAACCTCCTCAGC
NA7	AAGCAGGGTAATTTTGAATG	TTACGAAAAGTATTTAATTTG
NA8	ATGAATCCAAATCAAAAGATA	CGTAAATTACATCTTATCGAT
HA7	ATGAACACTCAAATTTCTAATATTAGCCAC	TCATATACAAATAGTGCACCCGCATGTTTC
HA3	ATGAAGACAACCATTATTTTGATACTACTG	TCAAATGCAAATGTTGCATCTGATGTTGCC

accession numbers AY383752 and AY383753 (Sa77 and Sa85, respectively). Neuraminidase sequences have accession numbers AY383757 (Sa77) and AY383754 (Sa85); hemagglutinin sequences have accession numbers AY383756 (Sa77) and AY383755 (Sa85b).

#### 4. Bioinformatic analysis

Sequence analyses were performed using the Blast N and Clustal X1.8 programs [29]. Phylogenetic trees were constructed using Saitou and Nei's Neighbor-joining method [22] and displayed employing the Treeview drawing program [19]. The statistical validity of this analysis is supported by the bootstrap analysis performed using 1,000 bootstrapped data sets.

## RESULTS

### *Sequence analysis of genes coding for nucleoprotein. Alignment and comparison with other isolates of equine influenza virus*

Nucleoprotein is the primary protein of the nucleocapsid and is coded by the fifth viral RNA segment. It is a type-specific antigen used to classify the equine influenza virus into groups A, B, and C. As an internal protein, it is not affected by selection pressures from the host immune system.

The nucleoprotein open reading frame consists of 1,497 base pairs coding for a protein of 498 amino acids.

In this study we compare nucleotide and amino acid sequences of the two isolates we studied with six sequences obtained from the NCBI database. The results are shown in Table II. Despite comparing sequences from two different subtypes (H7N7 and H3N8), a high degree of correspondence was found between them with the exception of Prague 1956 (Pr56), which is unrelated to other subtype H7N7 isolates. Some authors have suggested that subtype H7N7 might have undergone a genomic rearrangement that resulted in a nucleoprotein that differs from that of Pr56 [2]. The fact that this nucleoprotein has not been found again indicates that it may be an extinct lineage.

A high degree of correlation between amino acid changes and the year of the isolate may be appreciated with the exception of Jilin 1989 (Ji89). Although it is similar to the avian isolates [7], it should also be remembered that it is separated from Pr56 by thirty years. Isolates from the same decade are quite comparable as it can be seen from the data for 1986 (North American) and those from 1985 and 1977 (Chilean).

The resulting phylogenetic tree is shown in Figure 1. The tree has four branches; one corresponds to Pr56, the second to Ji89, the

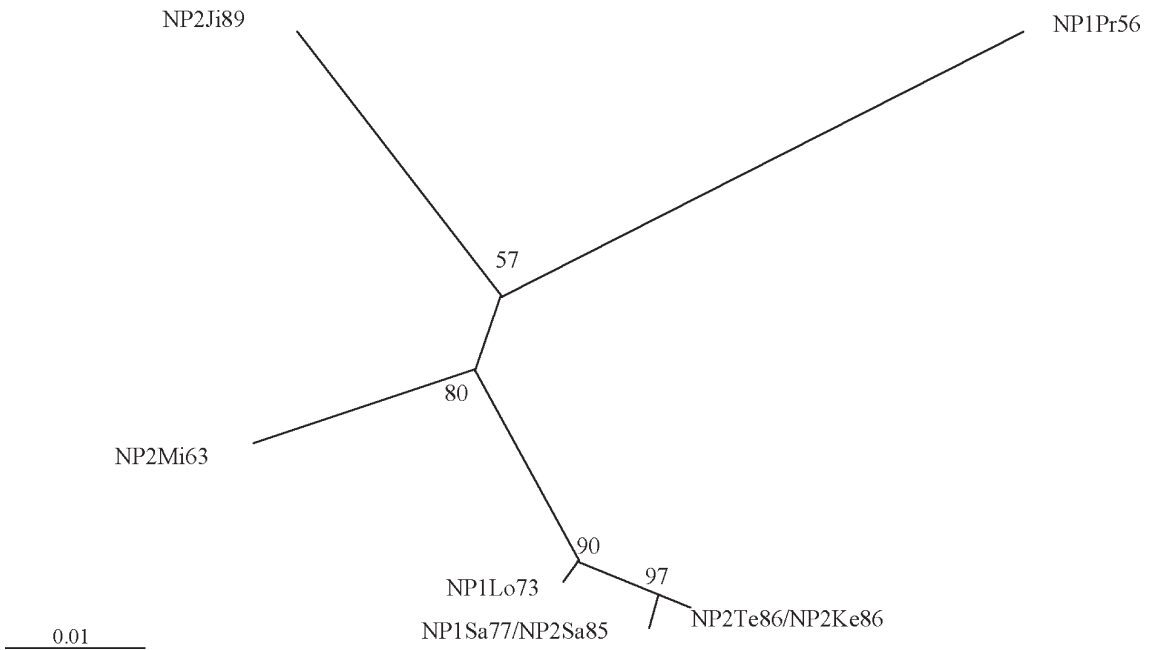
TABLE II

Comparative matrices of nucleotide and amino acid changes of nucleoproteins in different isolates of equine influenza virus of subtypes H7N7 (\*) and H3N8 (‡).  
Te86 corresponds to isolate Tennessee 1986.

	Pr56	Mi63	Lo73	Sa77	Sa85	Ke86	Te86	Ji89
Pr56*		270	264	266	263	268	268	237
Mi63‡	33		62	70	76	80	80	191
Lo73*	33	18		28	34	36	36	209
Sa77*	37	20	6		16	26	26	203
Sa85‡	37	20	6	3		12	12	203
Ke86‡	38	21	5	3	3		0	210
Te86‡	38	21	5	3	3	0		210
Ji89‡	33	20	24	28	28	29	29	

Amino acid variations

Nucleotide variations



**Figure 1.** Phylogenetic tree of all known amino acid sequences of nucleoproteins of the equine influenza virus, subtypes H7N7 (NP1) and H3N8 (NP2).

Bootstrap values obtained after 1,000 re-samplings. Figures represent percentage values. Bar scale represents number of amino acid substitutions per site.

third to Miami 1963 (Mi63) and the fourth to the most recent isolates. The nucleoprotein from the isolate Pr56, which is the most primitive and closest to the avian virus ancestor [7], shows the fewest similarities. Isolate Ji89, although more recent, has been classified as similar to the avian [8]. The remaining isolates show a high degree of homology. From this data we estimate a mutation rate of 0.33 events per year.

*Sequence analysis of genes coding for neuraminidase. Alignments and comparison with other isolates of the equine influenza virus.*

Coded by the sixth segment of viral RNA, this enzyme splits the  $\alpha$ -keto bond that joins a terminal sialic acid and the next sugar residue, thereby allowing the release of the viral progeny from infected cells. There are 9 subtypes of this protein identified in nature. They all have two structural regions, a stalk, and a head. Subtypes N7 and N8 have been found in viruses affecting horses. All N8 proteins

studied have 470 amino acids in which the first 8 are highly conserved, followed by a region rich in hydrophobic amino acids considered to be the transmembrane domain of the enzyme. The stalk is made up of the following 51 amino acids, and the head region begins with Cys90. The catalytic site of the enzyme is found in this last region.

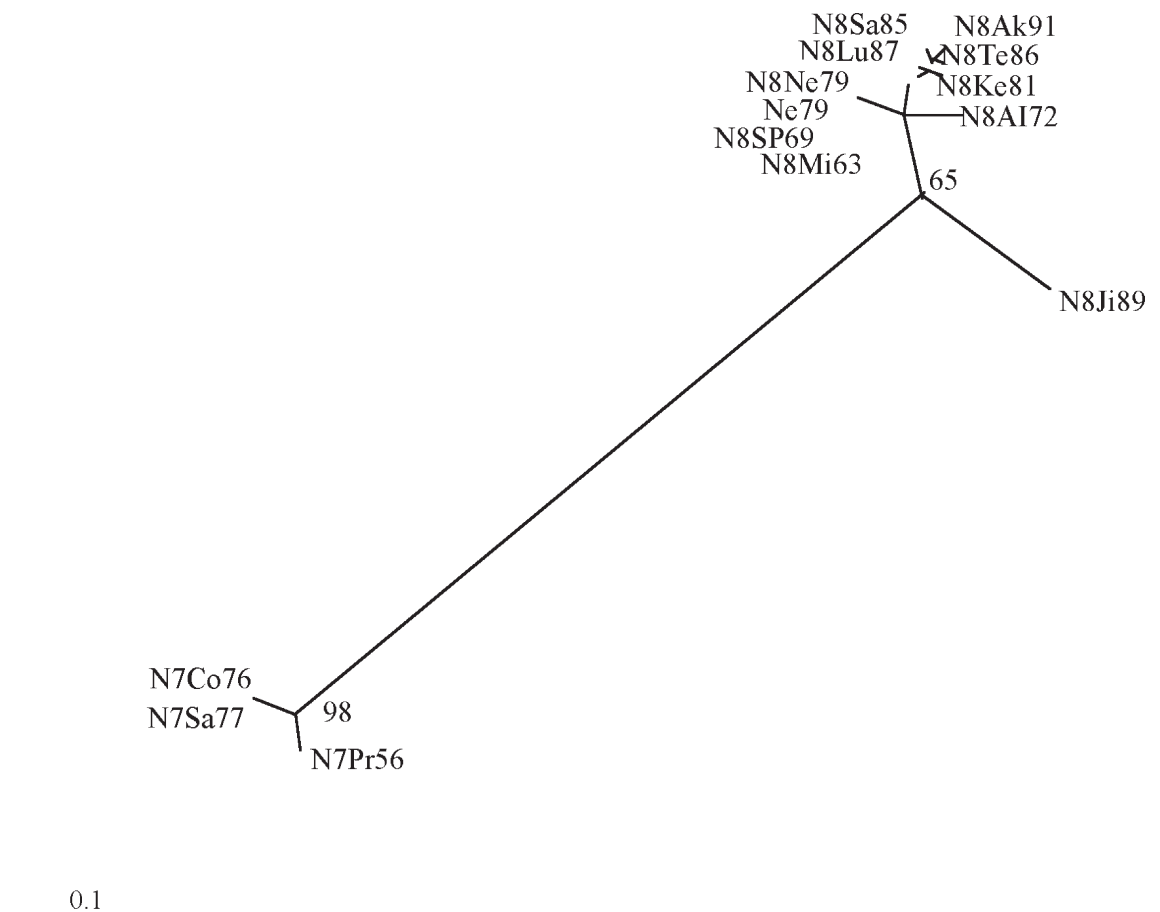
We compared the N7 neuraminidase coding region of isolate Sa77 with two sequences available in the database, all having 1,410 base pairs. Additionally, the neuraminidase nucleotide sequence from the Chilean isolate Sa85 of subtype N8, obtained in this work was compared with nine published sequences of different lengths, but we only used the codogenic region corresponding to 1,416 base pairs. The homology between them ranges from 89-99% (results not shown).

We also compared the cysteine residues and glycosylation sites of NA of subtype N8 with the nine published sequences. All of the N8 proteins analyzed have 17 highly-conserved cysteine residues; one is in the stalk region and the others in the head section. Cysteine residues from the head

section are highly conserved in other neuraminidase subtypes such as N1, N2, N5, N7 and N9 [21]. All sequences analyzed have 7 putative N-glycosylation sites, with the exception of isolate Mi63, Sao Paulo 1969 (SP69), and Ji89, which have only 6, and Alaska 1991 (Ak91), which has 8. Conserved sites in all isolates are located in residues 46, 54, 144, 293, and 400. The stalk region has only one cysteine residue and at least 3 glycosylation sites, while Sa85 has 4

reputed sites. Homology ranges from 92-99% in amino acid sequence, in which the stalk region (amino acids 39 to 89) is the most divergent (88-98%).

Figure 2 shows the phylogenetic tree drawn from the amino acid sequences deduced from the nucleotide sequences of all isolates. The data suggests that two subtypes of neuraminidase can be singled out. For N7, 21 of the 470 amino acids in the sequence of isolate Pr56 are different



**Figure 2.** Phylogenetic tree of all known amino acid sequences of the neuraminidases of the equine influenza virus, subtypes H7N7 and H3N8.

Bootstrap values obtained after 1,000 re-samplings are shown only at major nodes. Figures represent percentage values. Bar scale represents number of amino acid substitutions per site. Lu87 corresponds to isolate Ludhiana 1987.

with respect to the isolates of the 1970's (including Sa77), while in the Sa77 and Cornell 1976 (Co76) proteins there are only 2 different amino acids, resulting in a 99% homology. In N8, the phylogenetic tree shows that the Ji89 isolate is quite distant and that the Algiers 1972 (Al72) isolate is a little closer. Two groups or lineages can be noted: isolates from the 1980's (including 1979 and 1991), where Sa85 is located, and isolates from the 1960's. The mutation rate was estimated by plotting the year of the isolate against the length of the branch from a reference node in the phylogenetic tree resulting in 0.82 substitutions per year.

*Sequence analysis of genes coding for hemagglutinin. Alignments and comparison with other isolates of equine influenza virus*

This membrane glycoprotein is responsible for the adsorption of the virus into the host cell. It is the main antigen to which neutralizing antibodies are directed and its antigenic variation is the major cause of influenza epidemics. The hemagglutinin is coded by the fourth viral RNA segment. The protein has a signal peptide of 16 amino acids and two polypeptides (HA1 & HA2) joined by disulfide bridges where HA1 includes the amino terminal end, and HA2 has the carboxyl end [6]. This last polypeptide includes a hydrophobic region that allows the anchoring of the hemagglutinin to the viral membrane.

Different subtypes of hemagglutinin have been described, although only subtypes H3 and H7 are found in infected horses.

*Characterization of hemagglutinin subtype H7.*

Subtype H7 was the first isolated in horses in 1956 in Prague, and its latest report was in 1978 [32]. The nucleotide sequence from isolate Sa77 showed that the hemagglutinin coding region has 1,713 base pairs which code a polypeptide of 570 amino acids. In this study, 10 sequences obtained from Genbank database are compared with isolate Sa77. Results are shown in Table III. The 11 isolates show a remarkable homology in their nucleotide sequences (96%-99%), corresponding to a conserved amino acid sequence (95%-99%).

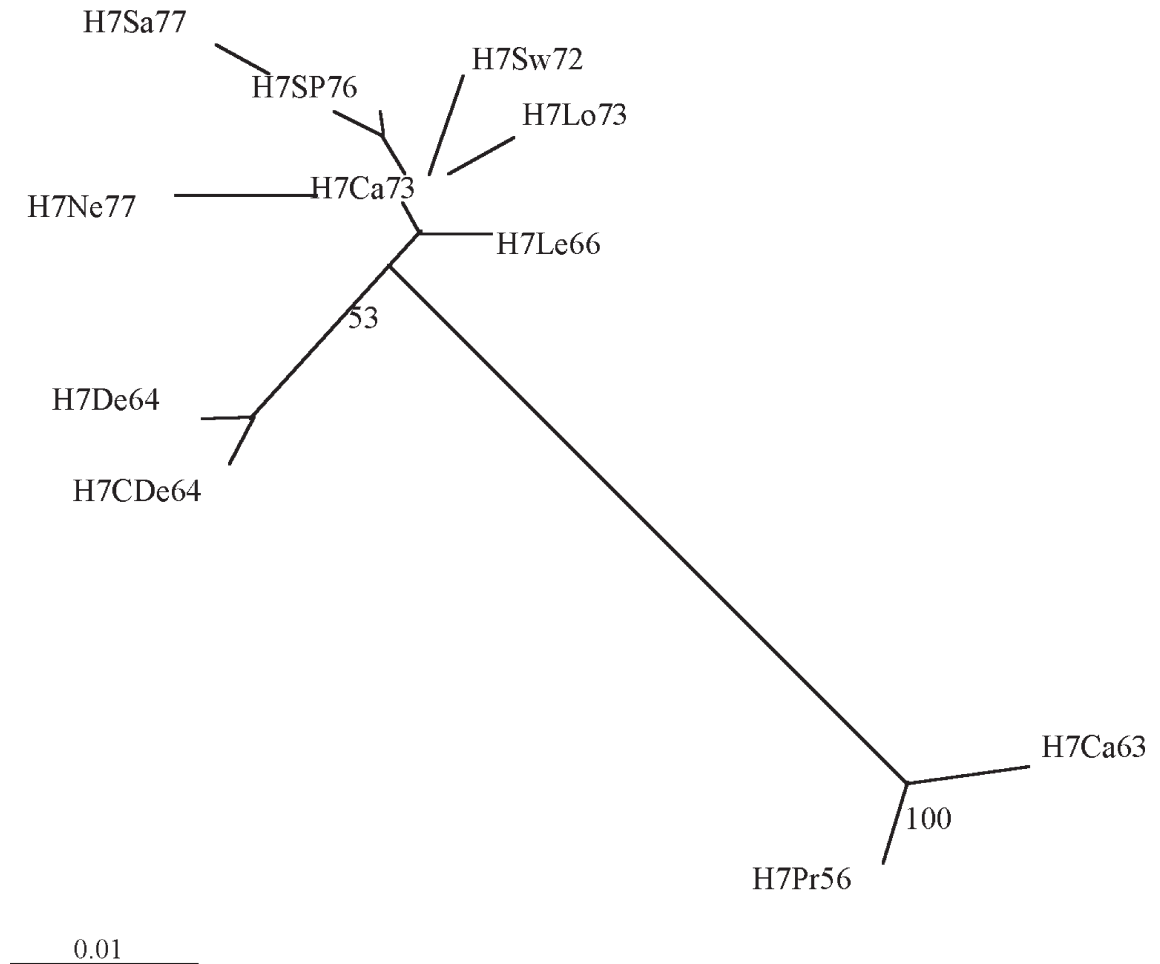
Figure 3 shows the phylogenetic tree obtained from the amino acid sequences of H7 from different isolates. The results show two well-defined groups, the earlier European isolates Pr56 and Cambridge 1963 (Ca63) and the others, reflecting a distant relation between them. Isolates from 1964 on show a high homology (~98%). We can also distinguish 3 groups. One is the North American lineage group constituted by the two Detroit 1964 isolates (De64, CDe64), the second is the European lineage, and a third is the South American lineage which comprises isolates Sao Paulo 1976 (SP76) and Sa77.

TABLE III  
Comparative matrices of amino acid and nucleotide changes of hemagglutinin H7.

	Nucleotide variations										
	Pr56	Ca63	De64	CDe64	Le66	Sw72	Lo73	Ca73	SP76	Ne77	Sa77
Pr56		<b>17</b>	<b>85</b>	<b>86</b>	<b>81</b>	<b>98</b>	<b>96</b>	<b>97</b>	<b>100</b>	<b>102</b>	<b>105</b>
Ca63	5		<b>90</b>	<b>91</b>	<b>86</b>	<b>101</b>	<b>101</b>	<b>100</b>	<b>103</b>	<b>106</b>	<b>108</b>
De64	26	26		<b>20</b>	<b>27</b>	<b>46</b>	<b>46</b>	<b>45</b>	<b>52</b>	<b>52</b>	<b>56</b>
CDe64	26	26	3		<b>30</b>	<b>47</b>	<b>47</b>	<b>46</b>	<b>53</b>	<b>53</b>	<b>57</b>
Le66	22	22	9	9		<b>42</b>	<b>42</b>	<b>41</b>	<b>48</b>	<b>52</b>	<b>54</b>
Sw72	26	26	13	13	6		<b>27</b>	<b>26</b>	<b>33</b>	<b>37</b>	<b>39</b>
Lo73	23	25	12	12	5	5		<b>24</b>	<b>33</b>	<b>37</b>	<b>41</b>
Ca73	23	23	10	10	3	3	2		<b>32</b>	<b>36</b>	<b>40</b>
SP76	25	25	12	12	5	5	4	2		<b>43</b>	<b>31</b>
Ne77	25	27	12	12	9	9	8	6	8		<b>49</b>
Sa77	28	28	13	13	8	8	9	7	5	11	

**Amino acid variations**

CDe64 corresponds to isolate Detroit Championship



**Figure 3.** Phylogenetic tree of all known amino acid sequences of the hemagglutinin isolates of the equine influenza virus, subtype H7N7. Bootstrap values obtained after 1,000 re-samplings are shown only at major nodes. Figures represent percentage values. Bar scale represents number of amino acid substitutions per site.

The amino acid sequence of this hemagglutinin shows a high degree of conservation in the first 9 amino acids; the tenth amino acid of the isolate Sa77 is invariable with respect to isolates from the 1950's and 1960's and isolates from the 1970's. There are 5 antigenic sites defined for H3 from human isolates [37] that can be extrapolated to H7, where there are amino acid substitutions in positions 164, 168,

169, 198, 202 and 208 of the antigenic site B (located in the most exposed region of the protein) (Table IV). Mutations present in residues 164, 169 and 198 correspond to differences among earlier isolates (1958-1963), in contrast to the rest of the isolates. In position 208, a mutation is present in the last isolates of this subtype, Sa77 and Switzerland 1972 (Sw72). In antigenic site D there are mutations in residues 182, 183,



Table IV  
Amino acid variations in relevant antigenic sites of hemagglutinin H7

Isolate	Site B						Site D						Site E		Site C	
	164	168	169	198	202	208	182	183	211	220	223	253	67	96	54	277
<b>Pr56</b>	S	G	T	T	R	N	K	V	I	Q	V	T	V	V	T	Q
<b>Ca63</b>	S	G	I	T	R	N	K	V	I	Q	V	T	V	V	T	Q
<b>De64</b>	N	G	V	A	R	N	R	E	I	Q	A	N	I	I	K	Q
<b>CDe64</b>	N	E	V	A	R	N	R	E	I	Q	A	N	I	I	K	Q
<b>Le66</b>	N	G	V	A	R	N	R	E	I	Q	A	T	I	I	K	Q
<b>Sw72</b>	N	G	V	A	K	D	R	E	I	Q	A	T	I	I	K	K
<b>Lo73</b>	N	G	V	A	R	N	R	E	I	Q	A	T	I	I	K	Q
<b>Ca73</b>	N	G	V	A	R	N	R	E	I	Q	A	T	I	I	K	Q
<b>SP76</b>	N	G	V	A	R	N	R	E	V	R	A	T	I	I	K	Q
<b>Ne77</b>	N	G	V	A	R	N	K	E	I	Q	A	N	I	I	K	Q
<b>Sa77</b>	N	G	V	A	R	D	R	E	V	R	A	N	I	I	K	Q

211, 220, 223 and 253; isolate Sa77 has three of them. In antigenic site E there are mutations in positions 67 and 96 only in the earlier isolates (Pr56, Ca63).

In this study we also analyzed the glycosylation sites of hemagglutinin subtype H7, which are involved in the immune evasion mechanism of the virus. In polypeptide HA1 we have two highly-conserved putative sites in positions 46 and 249. In polypeptide HA2, four conserved putative sites are located in positions 431, 484, 495, and 503. Differences found in HA1 include isolates Ca63 and London 1973 (Lo73), which have 3 N-glycosylation sites, while isolates Cambridge 1973 (Ca73), Sw72, SP76, Lexington 1966 (Le66) and Pr56 have four putative sites and isolates Sa77, Newmarket 1977 (Ne77), De64 and CDe64 have 5 glycosylation sites.

#### *Characterization of the hemagglutinin subtype H3.*

Sequencing of the hemagglutinin-coding region of subtype H3 shows an open reading frame of 1,698 base pairs, which codes a polypeptide of 565 amino acids. H3 is one of the most studied subtypes of this protein, with 64 sequences published between 1963 and 1999. Since the Chilean isolate is from 1985, we compared Sa85b with 26 published sequences reports from 1963 to 1989. All have 565 amino acids, except Ji89, which has 566 amino acids. Only the first 4 amino acids are conserved (excluding Ji89) at the N-terminal end.

There are 5 highly conserved putative N-glycosylation sites located in residues 37, 53, 78, 180, and 300 of HA1. The 26 sequences from the literature show 6 possible glycosylation sites in HA1 with the exception of the Uruguay 1963 isolate (Ur63), which has 5. In HA2 isolates there is a single highly-conserved site in position 498, which is the only N-glycosylation site in this polypeptide.

In HA1, there are 9 cysteine residues with the exception of Mi63 that has 8; they form 4 intra-chain disulfide bridges and one of them links HA1 to HA2. In HA2 there are 8 highly-conserved cysteine residues.

The amino acid variation in relevant antigenic sites of hemagglutinin is shown in Table V. In antigenic site A (residues of the amino acids 147-161) there are changes in residue 152 that distinguish isolates from years 1963, 1971, and 1972 from the rest with the exception of Ji89. There are also changes in residue 155 characterizing isolates from 1985 onward. In antigenic site B (amino acids 202-214), there is a mutation in residue 202 that characterizes all isolates between 1985 and 1989 (with the exception of Ji89) and in residue 204, which characterizes isolates of 1989 (with the exception of Ji89). Within site C (amino acids 67-70 and 288-293), proline is substituted for serine in residue 70, characterizing all isolates from 1985 onward (except Ji89, Sa85 and Sa85b). In the antigenic site D (amino acids 186-189, 222-230 and 256-261), in residue 187 there is a mutation from aspartic acid, found in



TABLE V  
Amino acid variations in relevant antigenic sites of hemagglutinin H3

	Site A			Site B			Site C			Site D			Site E			
Residues	147-161			202-214			67-70	288-293	186-189	222-230	256-261	59-63	94-99			
<b>Isolates</b>																
<b>Mi63</b>	QNGGSSACRRGSADS TNNEQTKLYVQAS CNNP PIDTCV NDNF KRSQQTIIP DVLMIN QSTST FXYENW															
<b>Ur63</b>	R															G
<b>Tk71</b>			S				S				L					
<b>A172</b>	R		S								L	V				
<b>Fo76</b>	R	G				E L			N		I			I		
<b>Fr76</b>		G				E L		W						I	Q	
<b>Ne76</b>	E	G		S	D	E			N		I			I	K	
<b>Ne79</b>	R	G				E			N		I			I	Q	
<b>So79</b>	S	G				E L			N		I			I	Q	
<b>Ro80</b>		G				E L			N		I			I		
<b>Ke81</b>		G			I	E			N		I			I	Q	
<b>Ab84</b>	R	G		S		E L			N		I			I	Q	
<b>Te85</b>	R	G	K		S	I	E	S			I			I	I	
<b>Sa85</b>	R	G	K		S	I	E		N		I	V		I	I	
<b>Sa85b</b>	R	G	K		S	I	E		N	R	I			I	I	
<b>Ke86</b>	R	G	K		S	I	E	S			I	I		I	I	
<b>Ke87</b>	R	G	K		S	I	E	S			I			I	I	
<b>Sk88</b>	R	G	K		S	I	E	S			I			I	I	
<b>El89</b>	R	G	K		S	K	I	E	S	G	K	E	V	I	I	
<b>Li89</b>	R	G	K		S	K	I	E	S		K	E	V	I	I	
<b>Su89</b>	R	G	K		S	K	I	E	S	L	G	K	E	V	I	
<b>Ne89</b>	R	G	K		S	K	I	E	S		K	E	V	I	I	
<b>Be89/3</b>	R	G	K		S	K	I	E	S		K	E	V	I	I	
<b>Be89/4</b>	R	G	K		S	K	I	E	S		K	E	V	I	I	
<b>Sf89</b>	R	G	K		S	K	I	E	S		K	E	V	I	I	
<b>Yv89</b>	R	G	K		S	K	I	E	S		K	E	V	I	I	
<b>Ji89</b>		K	T	S		Q	E		S	I	R	K	V		V	
														S	Q	
															T	

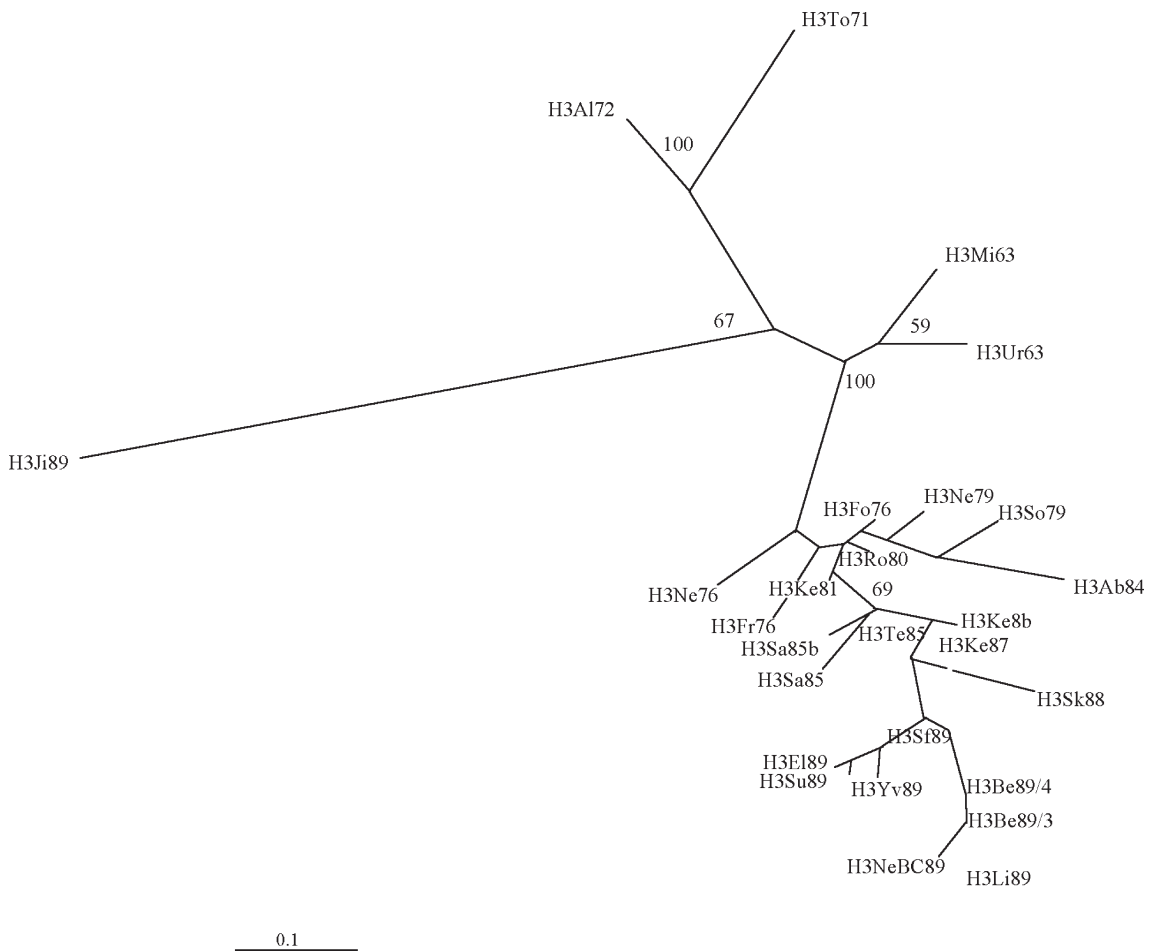
The following abbreviations are not shown in text: Ab84: Aby 1984; Be89/3: Berlin 1989; Be89/4: Berlin 1989; El89: Ella 1989; Fo76: Fontainebleau 1976; Fr76: France 1976; Ke87: Kentucky 1987; Li89: Lichfield 1989; Ne76: Newmarket 1976; Ne79: Newmarket 1979; Ne89: Newmarket 1989; Ro80: Romania 1980; Sk88: Skara 1988; Sf89: Suffolk 1989; So79: Solvalla 1979; Su 89: Sussex 1989; Tk71: Tokio 1971; Yv89: Yvelines 1989

old isolates, into arginine in isolates from 1976 to 1988, or into lysine in isolates of 1989 (with the exception of Ji89). In the antigenic site E (amino acids 59-63 and 94-99), residue 63 shows a mutation in which isoleucine is substituted for threonine, characterizing isolates from 1976 on.

Comparing the amino acid sequences of HA from isolate Sa85b (sequenced in this work) to isolate Sa85 [11], 7 differences are found. Two of them correspond to antigenic sites E and D (residues 95 and 260 respectively). Five are unique; they are not found in 25 of the 26 isolates listed in the literature. Of these five, two (residues 108 and 238) appear in both, proving that it is

unique for isolates Sa85 and Sa85b. Another, a change in residues 88 and 237 (Sa85) reported by Kawaoka [11], was not confirmed in our sequence. A third difference, found in residue 225 between isolates Sa85 and Sa85b, was observed in other isolates after repeated cultures in embryonated eggs [9].

Analyzing the phylogenetic tree of the 27 sequences (Fig. 4) we can describe two large groups, former isolates in addition to Ji89 and isolates from 1976 onward. Within the last group we can visualize three subgroups: one consisting of European isolates from years 1976 to 1984 (with the exception of Kentucky 1981, Ke81), North



**Figure 4.** Phylogenetic tree of all known amino acid sequences of the hemagglutinins of the equine influenza virus, subtype H3N8 of the years 1963 to 1989.

Bootstrap values obtained after 1,000 re-samplings are shown at selected nodes. Figures represent percent values. Bar scale represents number of amino acid substitutions per site. Ne79 corresponds to isolate Newmarket 1979 and NeBC89 corresponds to isolate Newmarket Bob Championship 1989.

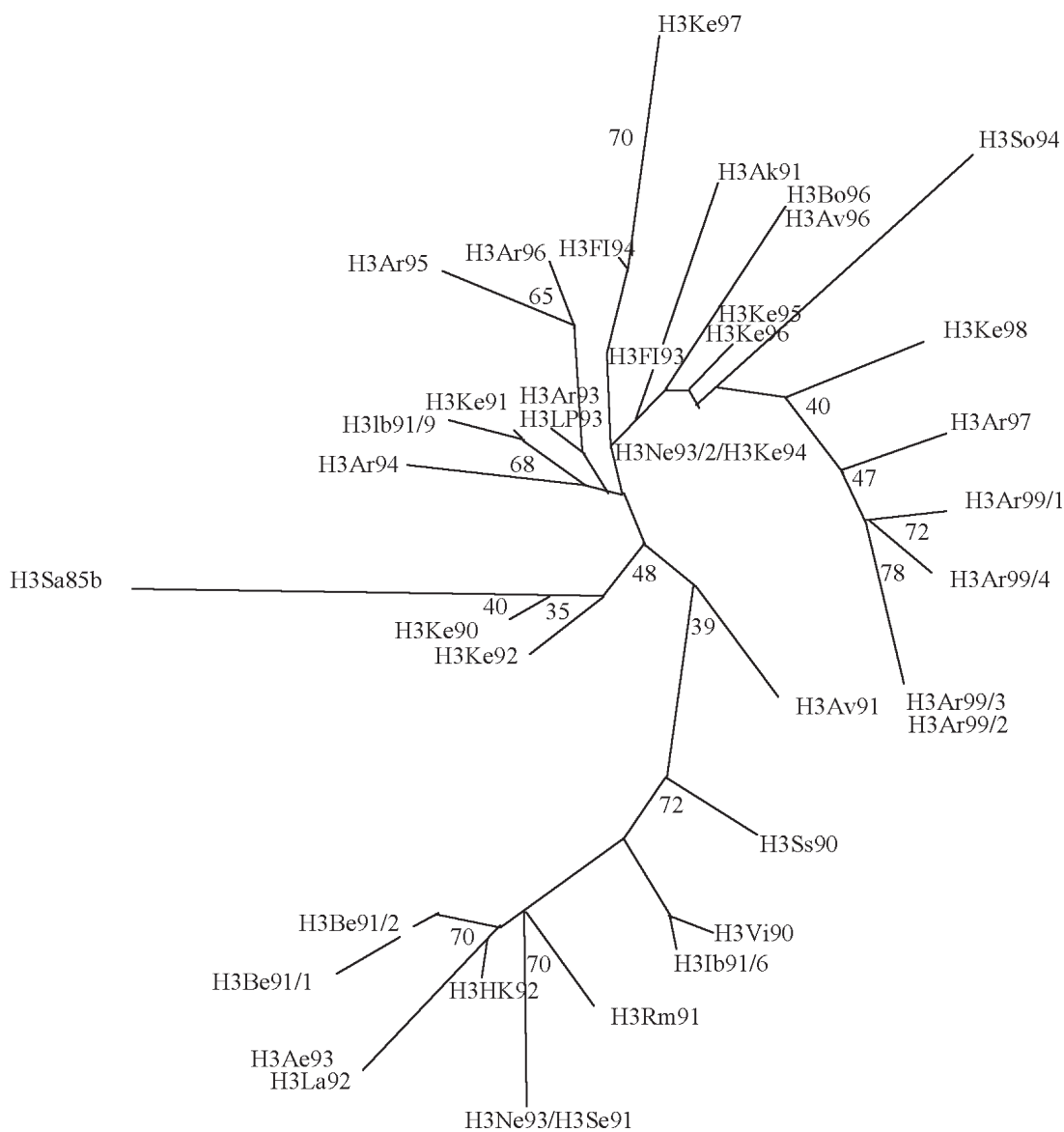
American isolates from years 1985 to 1988 and a third sub-group with the isolates of 1989, also of European lineage. The branch of the phylogenetic tree where Sa85 and Sa85b are found includes isolates Kentucky 86 (Ke86), Tennessee 1985 (Te85).

#### DISCUSSION

In agreement with previous reports [1, 28], the sequence analysis of the two equine influenza strains isolated in Chile demonstrates that the nucleoprotein is a good antigen to be considered in a vaccine formulation as it is highly-conserved in

influenza type A viruses and exclusive of viruses that infect horses. This gene undergoes mutations at a very low rate and generates cellular and humoral immunity in DNA vaccines that are not affected by maternal antibodies [20].

The phylogenetic trees show that the first virus isolates (Pr56 and Mi63) have a great similitude with avian isolates. This fact would indicate, as Gorman et al. have suggested [7], that the equine virus originated from the avian virus. Furthermore, isolate Ji89 was classified as "avian-like" and is very close to the first isolates, as we noted in all of the proteins analyzed here.



0.01

**Figure 5.** Phylogenetic tree of all known amino acid sequences of the hemagglutinins of the equine influenza virus, subtype H3N8 of isolates of the period 1990-1999.

Bootstrap values obtained after 1,000 re-samplings are shown at selected nodes. Figures represent percent values. Bar scale represents the number of amino acid substitutions per site.

The abbreviations correspond to the following isolates; Av96: Alvdalen 1996; Ar95: Argentina 1995; Ar97: Argentina 1997; Ar99/1: Argentina 1999 isolate 1; Ar99/2: Argentina 1999 isolate 2; Ar99/3: Argentina 1999 isolate 3; Ar99/4: Argentina 1999 isolate 4; Au91: Arundel 1991; Ae93: Avesta 1993; Be91/1: Berlin 1991 isolate 1; Be91/2: Berlin 1991 isolate 2; Bo96: Bollnas 1996; FI93: Florida 1993; FI94: Florida 1994; HK92: Hong Kong 1992; Ib91/6: Ibadan 1991 isolate 6; Ib91/9: Ibadan 91 isolate 9; Ke94: Kentucky 1994; Ke95: Kentucky 1995; Ke96: Kentucky 1996; Ke97: Kentucky 1997; Ke98: Kentucky 1998; La92: Lambourn 1992; Ne93/1: Newmarket 1993 isolate 1; Ne93/2: Newmarket 1993 isolate 2; Rm91: Rome 1991; Se91: Sweden 1991; Ss90: Saskatoon 1990; So94: Soderala 1994; Vi90: Visingsö 1990

Mutation rates of neuraminidase and hemagglutinin are similar, suggesting that both membrane proteins are under comparable selective pressures from the host.

Although strain H7N7 has not been isolated for more than 20 years, circulating antibodies against it have been reported [16]. It was decided that H7N7 would be maintained as a vaccine-strain, although it might be advisable to incorporate a strain derived from the post 1964 group, which could be more effective than the currently-used Pr56 isolate. Carbohydrate chains may modulate the biological activity of glycoproteins influencing the virulence of the isolate; it can be used as a blocking mechanism to antibody linking. Isolate Sa77 has the highest number of glycosylation sites.

With respect to the HA of subtype H3, in agreement with results in the literature [26] most amino acid substitutions in Sa85 are found in the protein's globular region. This globular region, whose three-dimensional structure is known [36], contains all the known antigenic sites.

Comparing the entire sequence of HA of isolate Sa85b with all sequences from the 1990's available in the Genbank (Fig. 5), we found that it belongs to the same branch of Kentucky 1990 (Ke90) and Kentucky 1992 (Ke92) isolates and is less similar to the South American isolates Argentina 1993 (Ar93), Argentina 1994 (Ar94), La Plata 1993 (LP93) and Kentucky 91 (Ke91). They have a lower mutation rate compared with the rest of the strains, of 0.27 amino acids per year or  $0.8 \times 10^{-3}$  amino acids per site per year [10]. Selecting polypeptide HA1, where relevant antigenic sites are found, site C in residues 51-54, site D in residues 240-245 and E, in residues 43-47 have no differences compared with American or European isolates. Antigenic site A of isolate Sa85b has one amino acid that differs from the American isolates and one from the European ones. In site B there are 4 differences in the amino acid sequence as compared with American isolates. Site C has 3 differences with the European isolates but none with the American ones. Site D has only one difference with isolates of the 1990's and

one with isolates Ar95 and Ar96. We therefore conclude that isolates Sa77 and Sa85 are good vaccine candidates, as they appear to preserve the most relevant antigenic determinants.

#### ACKNOWLEDGEMENTS

We thank the manuscript editing assistance of Dr. Carlos Doggenweiler and the technical assistance of Rodolfo Soto. This research has been partially funded by the Iniciativa Científica Milenio from the government of Chile.

#### REFERENCES

1. ANDREW ME, COUPAR BE, BOYLE DB, ADA GL (1987) The roles of influenza virus haemagglutinin and nucleoprotein in protection: analysis using vaccinia virus recombinants. *Scand J Immunol* 25: 21-28
2. BEAN WJ (1984) Correlation of influenza A virus nucleoprotein genes with host species. *Virology* 133: 438-442
3. BERRÍOS P, CELEDÓN M, SEPÚLVEDA C (1986) Influenza equina. Aislamiento del virus A/equi/2 (H3N8). *Arch Ciencias Vet* 1: 65-66
4. CASANOVA A, MARTÍNEZ I, ROMÁN M (1977) Aislamiento y tipificación del virus de influenza equina en Chile. *Arch Med Vet* 9: 91-93
5. CHOMCZYNSKI P (1993) A reagent for the single-step simultaneous isolation of RNA, DNA and proteins from cell and tissue samples. *Biotech* 15: 532-534, 536-537
6. CULLINANE A (1997) Viral respiratory disease In: ROBINSON N (ed) *Current therapy in equine medicine* 4<sup>th</sup> ed. Philadelphia: WB Saunders Company
7. GORMAN OT, BEAN WJ, KAWAOKA Y, WEBSTER RG (1990) Evolution of the nucleoprotein gene of influenza A virus. *J Virol* 64: 1487-1497
8. GUO Y, WANG M, KAWAOKA Y, GORMAN O, ITO T, SAITO T, WEBSTER RG (1992) Characterization of a new avian-like influenza A virus from horses in China. *Virology* 188: 245-255
9. ILOBI CP, NICOLSON C, TAYLOR J, MUMFORD JA, WOOD JM, ROBERTSON JS (1998) Direct sequencing of the HA gene of clinical equine H3N8 influenza virus and comparison with laboratory derived viruses. *Arch Virol* 143: 891-901
10. LAI AC, CHAMBERS TM, HOLLAND RE JR, MORLEY PS, HAINES DM, TOWNSEND HG, BARRANDEGUY M (2001) Diverged evolution of recent equine-2 influenza (H3N8) viruses in the Western Hemisphere. *Arch Virol* 146: 1063-1074
11. KAWAOKA Y, BEAN WJ, WEBSTER RG (1989) Evolution of the hemagglutinin of equine H3 influenza viruses. *Virology* 169: 283-292
12. LAMB RA, CHOPPIN PW (1983) The gene structure and replication of influenza virus. *Ann Rev Biochem* 52: 467-506
13. MADIC J, MARTINOVIC S, NAGLIC T, HAJSIG D, CVETNIC S (1996) Serological evidence for the

- presence of A/equine-1 influenza virus in unvaccinated horses in Croatia. *Vet Rec* 138: 68
14. MAHY B, KANGRO H (1996) *Virological Methods Manual*. London: Academic Press
  15. MÜLLER I, MEDINA-SELBY A, PALACIOS JL, MARTÍNEZ P, OPAZO P, BRUCE E, MANCILLA M, VALENZUELA PDT, YUDELEVICH A, VENEGAS A (2002) Cloning and comparison of ten gene sequences of a Chilean *H. pylori* strain with other *H. pylori* strains revealed higher variability for *VacA* and *CagA* virulence factors. *Biol Res* 35: 67-84
  16. MUMFORD J (1992) Progress in the control of equine influenza. Equine infectious diseases, VI Proceedings of the Six international Conference, R&W Publications 207-217
  17. MUMFORD J, WOOD J (1993) WHO/OIE meeting: consultation on newly emerging strains of equine influenza. 18-19 May 1992, Animal Health Trust, Newmarket, Suffolk, UK. *Vaccine* 11: 1172-1175
  18. MURPHY B, WEBSTER R (eds) (1996) *Orthomyxoviruses*. *Field's Virology*. Philadelphia: Lippincot-Raven
  19. PAGE RD (1996) TreeView: an application to display phylogenetic trees on personal computers. *Comp Appl Biosci* 12: 357-358
  20. PERTMER TM, ORAN AE, MOSER JM, MADORIN CA, ROBINSON HL (2000) DNA vaccines for influenza virus: differential effects of maternal antibody on immune responses to hemagglutinin and nucleoprotein. *J Virol* 74: 7787-7793
  21. SAITO T, KAWAOKA Y, WEBSTER RG (1993) Phylogenetic analysis of the N8 neuraminidase gene of influenza A viruses. *Virology* 193: 868-876
  22. SAITOU N, NEI M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4: 406-425
  23. SAMBROOK J, RUSSELL D (2001) *Molecular cloning. A laboratory manual*. New York: Cold Spring Harbor Laboratory Press CSH: 846-853
  24. SANGER F, NICKLEN S, COULSON AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 74: 5463-5467
  25. SINGH G (1995) A note on the concurrent isolation, from horses and ponies, of influenza A/EQ-1 and A/EQ-2 viruses from an epidemic of equine influenza in India. *Comp Immunol Microbiol Infect Dis* 18: 73-74
  26. SKEHEL JJ, WILEY DC (2000) Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. *Ann Rev Biochem* 69: 531-569
  27. STUDDERT M (1996) *Orthomyxoviridae*. In: STUDDERT M (ed) *Virus infections of equines*. pp: 281-284
  28. TAMURA S, MIYATA K, MATSUO K, ASANUMA H, TAKAHASHI H, NAKAJIMA K, SUZUKI Y, AIZAWA C, KURATA T (1996) Acceleration of influenza virus clearance by Th1 cells in the nasal site of mice immunized intranasally with adjuvant-combined recombinant nucleoprotein. *J Immunol* 156: 3892-3900
  29. THOMPSON JD, GIBSON TJ, PLEWNIAK F, JEANMOUGIN F, HIGGINS DG (1997) The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25: 4876-4882
  30. TISHCLER N, FERNANDEZ J, MÜLLER I, MARTÍNEZ R, GALENO H, VILLAGRA E, MORA J, RAMÍREZ E, ROSEMBLATT M, VALENZUELA PDT (2003). Complete sequence of the genome of the human isolate of Andes virus CH-7913: comparative sequence and protein structure analyses. *Biol Res* 36: 201-210
  31. TRAUB-DARGATZ JL, SALMAN MD, VOSS JL (1991) Medical problems of adult horses, as ranked by equine practitioners. *J Am Vet Med Assoc* 198: 1745-1747
  32. TUMOVA B (1980) Equine influenza: A segment in influenza virus ecology. *Comp Immunol Microbiol Infect Dis* 3: 45-59
  33. UPPAL PK, YADAV MP (1987) Outbreak of equine influenza in India. *Vet Rec* 121: 569-570
  34. VAN MAANEN C, CULLINANE A (2002) Equine influenza virus infections: an update. *Vet Q* 24: 79-94
  35. WEBSTER RG, GUO YJ (1991) New influenza virus in horses. *Nature* 351: 527
  36. WILEY DC, SKEHEL JJ (1987) The structure and function of the hemagglutinin membrane glycoprotein of influenza virus. *Ann Rev Biochem* 56: 365-394
  37. WILEY DC, WILSON IA, SKEHEL JJ (1981) Structural identification of the antibody-binding sites of Hong Kong influenza haemagglutinin and their involvement in antigenic variation. *Nature* 289: 373-378