Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

ELSEVIER

Virus Research 42 (1996) 41–52

Virus Research

Genetic variation in open reading frame 2 of field isolates and laboratory strains of equine arteritis virus

Jodi F. Hedges[a], Udeni B.R. Balasuriya[a], Peter J. Timoney[b], William H. McCollum[b], N. James MacLachlan[a,*]

[a]*Department of Pathology, Microbiology and Immunology, School of Veterinary Medicine, University of California, Davis CA 95616, USA*

[b]*Department of Veterinary Science, Gluck Equine Research Center, University of Kentucky, Lexington KY 40546, USA*

Received 19 December 1995; revised 9 February 1996; accepted 9 February 1996

Abstract

The open reading frame 2 (ORF2) of three laboratory strains, the live attenuated vaccine virus, and 18 field isolates of equine arteritis virus (EAV) from Europe and North America was sequenced. The ORF2 of EAV encodes the G_S protein that is abundantly expressed in infected cells but constitutes less than 2% of the virion protein mass. Variation of ORF2 among the isolates facilitated phylogenetic analysis that largely confirmed results of an earlier study based on sequence divergence of ORF5 of the same isolates of EAV, despite exposure of the proteins encoded by ORF2 (G_S) and ORF5 (G_L) to potentially different selective pressures in vivo. The data indicate that the G_S protein is highly conserved between isolates, considerably more so than the G_L protein, consistent with an important role of the G_S protein in virus replication.

Keywords: Equine arteritis virus; Phylogenetic analysis; G_S protein

1. Introduction

Equine arteritis virus (EAV) is a positive stranded RNA virus and is the cause of equine viral arteritis (Doll et al., 1957). EAV infection can cause abortion, persistent infection of stallions, interstitial pneumonia in foals, as well as clinical disease of variable severity in adult horses (Timoney and McCollum, 1993). The EAV virion is of similar size and icosahedral architecture as those of members of the Togaviridae family (Westaway et al., 1985). The genome organization, gene expression and replicase function of EAV, however, resemble those of coronaviruses and toroviruses, with formation of multiple 3′ coterminal subgenomic mRNAs (den Boon et al., 1991; Snijder and Horzinek, 1993). Simian hemorrhagic fever virus, lactate dehydrogenase-elevating virus (LDV), and porcine reproductive and respiratory syndrome virus (PRRSV) are closely re-

* Corresponding author.

PII S0168-1702(96)01294-4

lated to EAV (Conzelmann et al., 1993; Meulenberg et al., 1993; Wensvoort et al., 1993). These viruses are currently defined as members of the arterivirus genus, however it has been proposed that they share sufficient characteristics to warrant the formation of a new family, the Arteriviridae (Plagemann and Moennig, 1992; Meulenberg et al., 1994).

The EAV genome includes eight open reading frames (ORFs: 1a, 1b, 2–7), and is polyadenylated at the 3′-end (den Boon et al., 1991). ORFs 1a and 1b occupy 9.5 kb located at the 5′-end of the genome and encode the viral replicase (den Boon et al., 1991; de Vries et al., 1992). ORFs 2–7 are nested within 2.9 kb at the 3′-end of the genome. ORF5 encodes the glycosylated 29 kDa G_L protein that contains the neutralization epitopes of the virus (Balasuriya et al., 1993, 1995b; Deregt et al., 1994; Chirnside et al., 1995; Glaser et al., 1995). ORFs 6 and 7 respectively encode an unglycosylated envelope protein, M, and a phosphorylated nucleocapsid protein, N (de Vries et al., 1992). ORF2 encodes a 25kDa Class I ($N_{exo}C_{cyto}$) membrane protein (G_S) with a single potential N-glycosylation site (de Vries et al., 1992). The G_S protein has three N-terminal cysteines presumed to be important for homodimerization and protein function. The G_S protein is abundant in infected cells but occupies less than 2% of the virion protein mass, suggesting an important pre-assembly function (de Vries et al., 1992).

Many RNA viruses are found as a heterogeneous mixture of related genomes, or quasispecies (Domingo et al., 1985; Domingo and Holland, 1988). Genetic analyses of ORFs 5–7 suggest that the same is true of EAV (Sugita et al., 1994; Chirnside et al., 1994; Balasuriya et al., 1995a). Genetic differences within a virus population can be responsible for differences in tissue tropism, antigenicity, and virulence (Wege et al., 1981; Faragher et al., 1988; Weaver et al., 1992). Similarly, differences in severity of disease associated with infection by different strains of EAV and LDV have been described (Bryans et al., 1957; Doll et al., 1957; Stroop and Brinton, 1983; Brinton et al., 1986).

The goal of this study was to determine the genetic variation in ORF2 of three laboratory strains, the live attenuated virus vaccine (ARVAC®), and 18 field isolates of EAV. Because of presumed differences in the selective pressures exerted on the G_S and the G_L proteins in vivo, it was of interest to compare our data with the results of a similar genetic analysis of ORF5 of the same isolates (Balasuriya et al., 1995a). Specifically, ORF5 encodes the G_L protein that expresses the neutralization determinants of EAV, thus it would be expected to be exposed to immune selection, whereas the G_S protein constitutes only a very minor part of the virion. The data show that the G_S protein is highly conserved between isolates, considerably more so than the G_L protein, indicating an important role for the G_S protein in EAV replication.

2. Materials and methods

2.1. Cells and viruses

Viruses were propagated in rabbit kidney (RK-13) cells, as previously described (Balasuriya et al., 1995b). The cells were maintained in Dulbecco's modified Eagles medium supplemented with 10% heat-inactivated calf serum and 100 IU of penicillin and streptomycin, 100 μg/ml.

The nucleotide sequences of ORF2 of three different laboratory strains derived from the original prototype Bucyrus virus, the modified live vaccine (ARVAC®, Fort Dodge Laboratories, IA), and 18 field isolates of EAV were compared (Table 1). The virulent Bucyrus strain (VBS53) had been passaged only in horses and was harvested as pleural fluid, whereas all other isolates had previously been passaged in cell culture. The California Veterinary Diagnostic Laboratory System (CVDLS) strain of EAV (EAVCVDLS, generously provided by Dr. A. Castro, CVDLS, Davis, CA) was originally obtained from the National Veterinary Services Laboratory, Ames, IA. Our prototype strain (EAVUCD) was derived from the EAVCVDLS strain, as previously described (Balasuriya et al., 1995a,b). Another laboratory strain of the Bucyrus

Table 1
Origin and passage history of EAV isolates

Virus isolate	Year of isolation	Location/country of origin	Breed	Passage history
VBS53	1953	Bucyrus, Ohio, USA	STB	Eq15
KY63	1963	Kentucky, USA	ASB	HK 3, ED 1, RK1
SWZ64	1964	Bibuna, Switzerland[a]	Mixed	HK 5, RK 3, ED 1, RK 1
AUT68	1968	Vienna, Austria[a]	Warmblood	HK 5, RK 2
PA76	1976	Pennsylvania, USA	STB	RK 2
PLD76	1976/77	Wroclaw, Poland[b]	TB	RK 2
KY77	1977	Kentucky, USA	STB	RK 2
KY84	1984	Kentucky, USA	TB	RK 2
CAN86	1986	Alberta, Canada	TB	RK 3
AZ87	1987	Arizona, USA	AB	RK 2
NE88	1988	Nebraska, USA	AB	RK 2
MT89	1989	Montana, USA	AB	RK 2
NE89	1989	Nebraska, USA	AB	RK 3
ITA92	1992	Italy[c]	Maremmano	RK 2
IL93TB	1993	Illinois, USA	TB	RK 2
KY93	1993	Kentucky, USA	TB	RK 3
MI93	1993	Michigan, USA	STB	RK 2
IL93AB	1993	Illinois, USA	AB	RK 3
ARVAC	N/A	Fort Dodge Laboratories	STB	HK 131, RK 111, ED 24
EAVATCC	N/A	ATCC, USA	STB	Eq 15, LLC-MK 1, RK 1, BHK 1
EAVCVDLS	N/A	California Vet. Diag. Lab. Sys., USA	N/A	RK high passage
EAVUCD	N/A	University of California, Davis, USA	N/A	Vero 3, RK 2

Cells: HK, primary horse kidney; ED, equine dermis; RK, rabbit kidney; LLC-MK, Rhesus monkey kidney cell line 2; BHK, baby hamster kidney and Vero-African green monkey kidney cells.
Passage history: Each cell type is followed by the number of passages on that particular cell type (e.g. RK 3: passaged three times in rabbit kidney 13 cells; Eq 15: 15 serial passages in horses).
Horse breed: TB, thoroughbred; STB, standardbred; AB, Arabian; ASB, American saddlebred.
[a] Dr. Franz Burki, Virology Institute, Veterinary Medical School, Vienna, Austria.
[b] Dr. Berndt Klingeborn, The National Veterinary Institute, Uppsala, Sweden.
[c] Dr. G.L. Autorino, Istituto Zooprofillattico Sperimentale delle Regioni Lazio eTuscana, Rome.

virus (EAVATCC) was obtained from the American Type Culture Collection (ATCC; Rockville, MD) and subsequently passaged once each in RK-13 and BHK-21 cells. In addition, ARVAC®, a live attenuated vaccine strain of EAV, also derived from the Bucyrus strain of EAV, was included (Doll et al., 1968; McCollum, 1969; Harry and McCollum, 1981). The field isolates were collected over a 40-year period, and were isolated from horses of different breeds in the United States, western Canada, Poland, Italy, Switzerland and Austria (Table 1). The published sequence of another laboratory strain of the Bucyrus virus was used for comparison (EAVUtr; den Boon et al., 1991).

2.2. Amplification and sequencing

Total RNA was extracted from virus infected RK-13 cells using guanidinium thiocyanate and phenol (RNAzol™B; Tel-Test, Inc., TX), according to the manufacturers' instructions and as previously described (Balasuriya et al., 1995b). The ORF2 was amplified by reverse transcription-polymerase chain reaction (RT-PCR) using the same RNA preparations that were extracted from

each virus for genetic analysis of ORF5 (Balasuriya et al., 1995a). The published sequence of a laboratory strain of EAV (EAVUtr) was used to design synthetic oligonucleotide 20-mer or 21-mer primers (National Biosciences, MN) for amplification and sequencing as previously described (Balasuriya et al., 1995b) (Positive primers: 9728–9747, 9746–9765, 9984–10 004, 10 125–10 144, 10 261–10 281; Negative primers: 10 111–10 131, 10 353–10 373, 10 571–10 590). The GeneAmp RNA PCR kit (Perkin Elmer, CA) was used for RT-PCR and reactions were carried out as previously described (Balasuriya et al., 1995a). The 10 571–10 590 primer was used for reverse transcription of viral RNA into cDNA. PCR amplification was then carried out using the 9728–9747 and 10571–10590 primers. The resulting product was the entire ORF2 with approximately 75 nucleotides flanking each end. Fourteen different PCR reactions (100 μl/reaction) were carried out with each virus strain to control for variation in amplification. These reaction products were pooled, concentrated (Centricon-30, Amicon, MA), and purified using a commercial kit (Gene-Clean, BIO101, CA).

Purified cDNA from each amplicon was sequenced using the PRISM™ Ready Reaction DyeDeoxy™ Terminator (Applied Biosystems, CA) cycle sequencing kit. Sense and nonsense strands were each sequenced with the respective PCR primers as well as one or two internal primers, as necessary and for additional confirmation of mutations. Approximately 100 ng cDNA and 10 pmol primer were used in each reaction. Fluorescently-labeled dideoxynucleotides were directly incorporated into extension products during 25 cycles consisting of 30 s denaturation at 96°C, 15 s annealing at 50°C and 4 min extension at 60°C. Sequence data were collected and edited with an ABI 377 Automatic Sequencer (Applied Biosystems, CA) according to the manufacturers' instructions.

2.3. Sequence and phylogenetic analysis

Computer analyses were performed using a Power Macintosh 8100/100AV. Nucleotide sequences from ORF2 were translated into amino acids, and predictions of hydrophobicity and secondary structure were made using HIBIO MacDNASIS pro version 3.5 software (Hitachi). The PileUp program of the Wisconsin package (GCG version 8.0 software; Genetics Computer Group, 1994) was used for multiple sequence alignment. Phylogenetic analysis was done with the PHYLIP (Phylogeny Inference Package) version 3.5c for the Macintosh Power PC (Felsenstein, 1993). Distance matrices for nucleotide sequence data were calculated for aligned sequences using the DNADist program based on the Kimura 2 parameter model (Kimura, 1980) with a transition/transversion ratio of 2.0. The distance matrices were then used in the FITCH program (least-square method; Fitch and Margoliash, 1967) to generate phylogenetic trees. The FITCH program was carried out with global rearrangements and randomized input order. The resulting phylogenetic tree is unrooted as rooting the tree with ORF2 of LDV or PRRSV results in a branching pattern that is an unlikely representation of the nucleotide changes observed. ORF2s of LDV and PRRSV have very low homology to ORF2 of EAV (Plagemann, 1996).

Several treeing methods were applied to the data in an attempt to produce the dendrogram that best reflects the nucleotide changes observed. Among these were the distance matrix based on maximum likelihood (Felsenstein, 1981a; Felsenstein, 1981b; Felsenstein, 1981c), Jin and Nei analysis, (Jin and Nei, 1990), as well as the Kimura two parameter model (Kimura, 1980). Distance matrices were then used in the FITCH program (Fitch and Margoliash, 1967) or the Neighbor-Joining/UPGMA (Saitou and Nei, 1987) program of the PHYLIP package (Felsenstein, 1993). Trees were also developed using the parsimony method. These methods gave similar groupings but not identical trees to that produced by the Kimura two parameter method.

3. Results

3.1. Variation in ORF2 and the G_s protein

The ORF2 of 18 field isolates, three laboratory

strains and the modified live vaccine virus (ARVAC®) of EAV was sequenced (GenBank accession numbers given below)[1]. Each ORF2 is the same length, 684 nucleotides, and each is predicted to encode a protein of 227 amino acids. The sequences were compared to ORF2 of the published sequence, EAVUtr (den Boon et al., 1991). Most of the nucleotide changes are transitional, and many (63.5%) are silent. The field isolates differ from each other by between 1 and 66 nucleotides (99.9–90.4% homology). ORF2 of the modified live vaccine (ARVAC®) is very similar to the laboratory strains, VBS53 and PLD76 (Table 2). ORF2 of the virulent Bucyrus strain (VBS53) and EAVCVDLS have the same nucleotide sequence as laboratory strains derived from them by cell culture passage, respectively, EAVATCC and EAVUCD (Table 1). This suggests that virus strains that experience low passage in culture, without selection, can sustain a relatively constant quasispecies equilibrium, as has been previously demonstrated with other viruses (Holland et al., 1982; Domingo and Holland, 1988; Steinhauer et al., 1989; Nichol et al., 1993). It is appreciated, however, that mutations may arise during adaptation of field strains to cell culture.

Nucleotide substitutions are scattered throughout ORF2. A large number of silent nucleotide changes are clustered in the region of nucleotides 10166–10409. The regions from the start codon at 9807 to 10 165 and from 10 410 to the stop codon at 10 490 contain fewer changes although a greater proportion of these are substitutions that result in amino acid changes. There are 67 unique nucleotide changes, many of which occur in the isolate KY63 and the European isolates. Some nucleotide substitutions are present in early isolates from a certain geographical area but are not present in later isolates (e.g. 9849, 10 316, 10 352 in KY77; 10 034, 10 296, in NE88; and numerous changes in KY63). Similarly, nucleotide changes occur in later isolates that are not present in earlier isolates from the same geographical region (10 247, 10 277 in KY84; 10 334 and 10 245 in all 1993 isolates; and numerous positions in ITA92). There are nucleotide changes that are restricted to one continent (9823, 9890 and 10 265 in North American isolates; 9845 and 10 187 in the European isolates). Only one nucleotide residue (10 198) is unique to the published sequence.

The alignment of the deduced amino acid sequences of each G_S protein from the various isolates of EAV is shown in Fig. 1. The isolates differ from the published sequence (EAVUtr; den Boon et al., 1991) by between 2 and 15 amino acids (99.1–93.4% homology; Table 2). The greatest difference between field isolates is 19 amino acids (91.6% homology) between ITA92 and NE88. The three laboratory strains (EAVATCC, EAVCVDLS, and EAVUCD) and VBS53 have identical amino acid sequences, each having two amino acid substitutions as compared to the published sequence.

Four variable regions in the G_S protein of the isolates are apparent: amino acids (aa) 1–33, aa 91–92, aa 123–131, and aa 161–227 (Fig. 1). The first 24 amino acids of the G_S protein are a proposed signal sequence (de Vries et al., 1992). The region from aa 123–131 includes the non-conservative glutamine to leucine substitution (position 131) that is common to all isolates evaluated. Many unique changes occur in the region from aa 161–227, which includes the transmembrane region and C terminus of the G_S protein (de Vries, 1994). The changes from positions aa 223–227 alter the proposed endoplasmic reticulum (ER) retention motif, RRKIL. This motif has been suggested to cause retention of the G_S protein in the ER when the protein is incorrectly dimerized (de Vries, 1994). Interestingly, the nucleotide sequences that encode aa 132–160 (nucleotide positions 10 200–10 286) vary among the isolates, despite the conservation of these amino acids.

The proposed N-glycosylation site on the G_S protein is at amino acid position 156 (de Vries, 1994). There are no amino acid changes that alter

[1] The nucleotide sequence data reported in this paper have been submitted to the GenBank nucleotide sequence data base and have been assigned the accession numbers: U42305 (EAVATCC), U42306 (EAVCVDLS), U42307 (EAVUCD), U42308 (ARVAC), U42309 (VBS53), U42310 (PA76), U42311 (KY63), U42312 (KY77), U42313 (KY84), U42314 (KY93), U42315 (AZ87), U42316 (NE88), U42317 (NE89), U42318 (MT89), U42319 (MI93), U42320 (IL93TB), U42321 (IL93AB), U42322 (CAN86), U42323 (SWZ64), U42324 (AUT68), U42325 (PLD76), U42326 (ITA92).

Table 2
Nucleotide and amino acid comparison

	EAVUtr	VBS53	PA76	AZ87	NE88	NE89	MT89	KY63	KY77	KY84	KY93	IL93AB	IL93TB	MI93	CAN86	SWZ64	AUT68	PLD76	ITA92	ARVAC	ATCC	CVDLS[a]	EAVUCD
EAVUtr		4(99.4)	33(95.2)	25(96.3)	37(94.6)	33(95.2)	28(95.9)	54(92.1)	26(96.2)	33(95.2)	33(95.2)	31(95.5)	32(95.3)	25(96.3)	29(95.8)	47(93.1)	47(93.1)	7(99.0)	53(92.3)	7(99.0)	4(99.4)	3(99.6)	3(99.6)
VBS53	2(99.1)		31(95.5)	23(96.6)	37(94.6)	33(95.2)	28(95.9)	54(92.1)	26(96.2)	33(95.2)	33(95.2)	31(95.5)	32(95.3)	25(96.3)	29(95.8)	47(93.1)	47(93.1)	7(99.0)	53(92.3)	7(99.0)	0(100.0)	1(99.9)	1(99.9)
PA76	11(95.2)	9(96.0)		40(94.2)	42(93.9)	36(94.7)	34(95.0)	56(91.8)	31(95.5)	34(95.0)	36(94.7)	40(94.2)	35(94.9)	36(94.7)	39(94.3)	54(92.1)	58(91.5)	36(94.7)	62(90.9)	34(95.0)	31(95.5)	30(95.6)	30(95.6)
AZ87	8(96.5)	6(97.4)	13(94.3)		44(93.5)	48(93.0)	39(94.3)	52(92.4)	37(94.6)	42(93.9)	38(94.4)	38(94.4)	37(94.6)	36(94.7)	38(94.4)	59(91.4)	42(93.9)	27(96.1)	39(94.3)	28(95.9)	23(96.6)	22(96.9)	22(96.9)
NE88	14(93.8)	14(93.8)	12(94.7)	16(93.0)		25(96.3)	16(97.7)	66(90.4)	14(98.0)	25(96.3)	26(96.2)	30(95.6)	27(96.1)	30(95.6)	30(95.6)	57(91.7)	51(92.5)	41(94.0)	58(91.5)	42(93.9)	37(94.6)	36(94.7)	36(94.7)
NE89	13(94.3)	13(94.3)	11(95.2)	14(93.8)	7(96.9)		19(97.2)	60(91.2)	11(98.9)	20(97.1)	27(96.1)	31(95.5)	26(96.2)	31(95.5)	27(96.1)	51(92.5)	53(92.3)	37(94.6)	61(91.1)	36(94.7)	33(95.2)	32(95.3)	32(95.3)
MT89	10(95.6)	10(95.6)	10(95.6)	12(94.7)	4(98.2)	5(97.8)		62(90.9)	10(98.5)	10(98.5)	18(97.4)	20(97.1)	17(92.5)	20(97.1)	26(96.2)	52(92.4)	52(92.4)	32(95.3)	60(91.2)	33(95.2)	28(95.9)	27(96.1)	27(96.1)
KY63	15(93.4)	15(93.4)	15(93.4)	17(92.5)	18(92.1)	14(93.8)	15(93.4)		57(91.7)	59(91.4)	56(91.8)	62(90.9)	57(91.7)	59(91.4)	60(91.2)	57(91.7)	54(92.1)	58(91.5)	59(91.4)	57(91.7)	54(92.1)	53(92.3)	53(92.3)
KY77	11(95.2)	11(95.2)	9(96.0)	13(94.3)	5(97.8)	2(99.1)	3(98.7)	13(94.3)		15(97.8)	16(97.7)	20(97.1)	15(97.8)	20(97.1)	20(97.1)	50(92.7)	46(93.3)	30(95.6)	56(91.8)	31(95.5)	26(96.2)	25(96.3)	25(96.3)
KY84	13(94.3)	13(94.3)	8(96.5)	15(93.4)	6(97.4)	5(97.8)	4(98.2)	12(94.7)	3(98.7)		18(97.4)	22(96.9)	17(92.5)	24(96.5)	27(96.1)	52(92.4)	52(92.4)	37(94.6)	59(91.4)	38(94.4)	33(95.2)	32(95.3)	32(95.3)
KY93	9(96.0)	9(96.0)	10(95.6)	11(95.2)	8(96.5)	5(97.8)	4(98.2)	11(95.2)	3(98.7)	4(98.2)		6(99.1)	1(99.9)	10(98.5)	26(96.2)	50(92.7)	48(92.9)	37(94.6)	52(92.4)	38(94.4)	33(95.2)	32(95.3)	32(95.3)
IL93AB	8(96.5)	8(96.5)	15(93.4)	12(94.7)	11(95.2)	10(95.6)	7(96.9)	16(93.0)	8(96.5)	9(96.0)	5(97.8)		5(99.3)	8(98.8)	28(95.9)	52(92.4)	52(92.4)	35(94.9)	57(91.7)	36(94.7)	31(95.5)	30(95.6)	30(95.6)
IL93TB	9(96.0)	9(96.0)	10(95.6)	11(95.2)	8(96.5)	5(97.8)	4(98.2)	11(95.2)	3(98.7)	4(98.2)	0(100.0)	5(97.8)		9(98.7)	25(96.3)	49(92.8)	47(93.1)	36(94.7)	52(92.4)	37(94.6)	32(95.3)	31(95.5)	31(95.5)
MI93	6(97.4)	6(97.4)	13(94.3)	10(95.6)	11(95.2)	10(95.6)	7(96.9)	15(93.4)	8(96.5)	9(96.0)	5(97.8)	4(98.2)	5(97.8)		26(96.2)	52(92.4)	52(92.4)	29(95.8)	56(91.8)	30(95.6)	25(96.3)	24(96.5)	24(96.5)
CAN86	11(95.2)	11(95.2)	11(95.2)	11(95.2)	9(96.0)	6(97.4)	7(96.9)	12(94.7)	5(97.8)	5(97.8)	4(98.2)	7(96.9)	4(98.2)	7(96.9)		48(92.9)	52(92.4)	31(95.5)	56(91.8)	34(95.0)	29(95.8)	28(95.9)	28(95.9)
SWZ64	9(96.0)	9(96.0)	12(94.7)	11(95.2)	18(92.1)	14(93.8)	14(93.8)	12(94.7)	13(94.3)	14(93.8)	10(95.6)	13(94.3)	10(95.6)	11(95.2)	10(95.6)		50(92.7)	50(92.7)	56(91.8)	50(92.7)	47(93.1)	46(93.3)	46(93.3)
AUT68	12(94.7)	12(94.7)	13(94.3)	16(93.0)	17(92.5)	14(93.8)	13(94.3)	11(95.2)	12(94.7)	13(94.3)	9(96.0)	14(93.8)	9(96.0)	14(93.8)	13(94.3)	13(94.3)		52(92.4)	37(94.6)	52(92.4)	47(93.1)	46(93.3)	46(93.3)
PLD76	6(97.4)	4(98.2)	12(94.7)	8(96.5)	16(93.0)	15(93.4)	12(94.7)	18(92.1)	13(94.3)	15(93.4)	11(95.2)	11(95.2)	11(95.2)	9(96.0)	13(94.3)	12(94.7)	15(93.4)		55(92.0)	8(98.8)	7(99.0)	6(99.1)	6(99.1)
ITA92	14(93.8)	14(93.8)	15(93.4)	10(95.6)	19(91.6)	13(94.3)	15(93.4)	13(94.3)	14(93.8)	15(93.4)	11(95.2)	16(93.0)	11(95.2)	14(93.8)	13(94.3)	13(94.3)	12(94.7)	16(93.0)		58(91.5)	53(92.3)	52(92.4)	52(92.4)
ARVAC	5(97.8)	3(98.7)	11(95.2)	8(96.5)	16(93.0)	15(93.4)	12(94.7)	17(92.5)	13(94.3)	15(93.4)	11(95.2)	11(95.2)	11(95.2)	9(96.0)	13(94.3)	11(95.2)	14(93.8)	5(97.8)	16(93.0)		7(99.0)	6(99.1)	6(99.1)
ATCC	2(99.1)	0(100.0)	9(96.0)	6(97.4)	14(93.8)	13(94.3)	10(95.6)	15(93.4)	11(95.2)	13(94.3)	9(96.0)	8(96.5)	9(96.0)	6(97.4)	11(95.2)	9(96.0)	12(94.7)	4(98.2)	14(93.8)	3(98.7)		1(99.9)	1(99.9)
CVDLS[a]	2(99.1)	0(100.0)	9(96.0)	6(97.4)	14(93.8)	13(94.3)	10(95.6)	15(93.4)	11(95.2)	13(94.3)	9(96.0)	8(96.5)	9(96.0)	6(97.4)	11(95.2)	9(96.0)	12(94.7)	4(98.2)	14(93.8)	3(98.7)	0(100.0)		0(100.0)
EAVUCD	2(99.1)	0(100.0)	9(96.0)	6(97.4)	14(93.8)	13(94.3)	10(95.6)	15(93.4)	11(95.2)	13(94.3)	9(96.0)	8(96.5)	9(96.0)	6(97.4)	11(95.2)	9(96.0)	12(94.7)	4(98.2)	14(93.8)	3(98.7)	0(100.0)	0(100.0)	

Upper section: number of nucleotide changes (percentage of nucleotide sequence homology).
Lower section: number of amino acid changes (percentage of amino acid homology).
[a] EAVCVDLS

FIGURE 1 Aligned Amino Acid Sequences of the G_5 Protein of Isolates of Equine Arteritis Virus

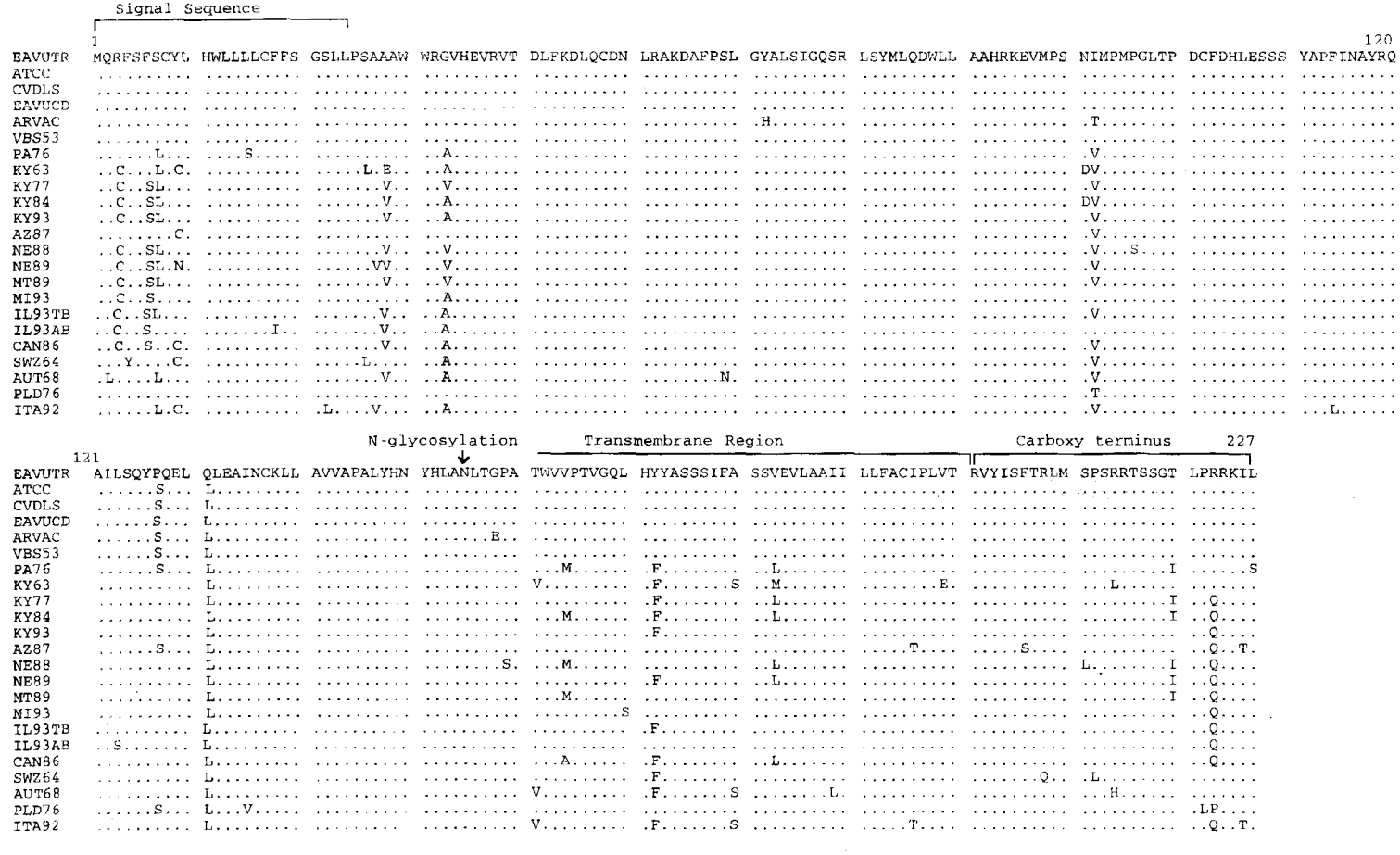


Fig. 1. Aligned amino acid sequences of the G_S protein of the published sequence of EAV (EAVUtr; den Boon et al., 1991), three laboratory strains, the ARVAC® vaccine virus, and 18 field isolates of EAV. Dots indicate homology with the published sequence at that position.

this site nor are there any that create new potential N-glycosylation sites on the G_S protein. In addition, there are no changes to any of the three luminal cysteine residues.

There are no apparent differences in the predicted hydrophobicity of the G_S protein between isolates (Kyte and Doolittle, 1982). Very minor changes were noted in the predicted secondary structures of the G_S protein of some isolates as determined by the Chou et al. (1978) method (data not shown).

3.2. Phylogenetic analysis

Nucleotide distance matrices of the ORF2 sequences were used to construct a phylogenetic tree (Fig. 2A). The viruses are grouped into three lineages, the first includes the North American viruses isolated after 1976, the second includes the European strains plus KY63, and the third is an irregular cluster that contains VBS53, the laboratory strains, the vaccine virus and PLD76. This phylogenetic tree was then compared to that obtained after sequence analysis of ORF5 of the same viruses (Fig. 2B; Balasuriya et al., 1995a).

The viruses in the North American lineage were isolated over a period of 17 years in Canada and the United States, but differ from each other at only 1–48 nucleotides in ORF2 (93.0–99.9% homology; Fig. 2A). This lineage branches to separate AZ87 from the other North American isolates. This was expected as the AZ87 strain was isolated during an outbreak of viral arteritis that resulted from the importation of a Polish stallion in 1986. AZ87 is placed in the European clade in the phylogenetic analysis based on the sequence of ORF5 of the same isolates (Fig. 2B). Although ORF2 of AZ87 is very similar to the published sequence, many of the nucleotide changes are distinctly European (nucleotide positions 9832, 9869, 9950, 10 181, 10 197, 10 295 and 10 393; and amino acid positions 9, 196 and 226; Fig. 1). The location of AZ87 in the tree is strongly dependent on the treeing method applied, for instance, it is a member of the laboratory strain cluster in a rooted tree analysis, whereas a tree made with the parsimony method places it in the European clade (data not shown).

Further branches of the North American lineage separate CAN86 and PA76 producing two groups of closely related viruses (Fig. 2A). It was expected that all 1993 isolates would group together as these strains are all derived from an outbreak of viral arteritis that originated in a Chicago clinic in the summer of 1993 and which subsequently spread to neighboring states.

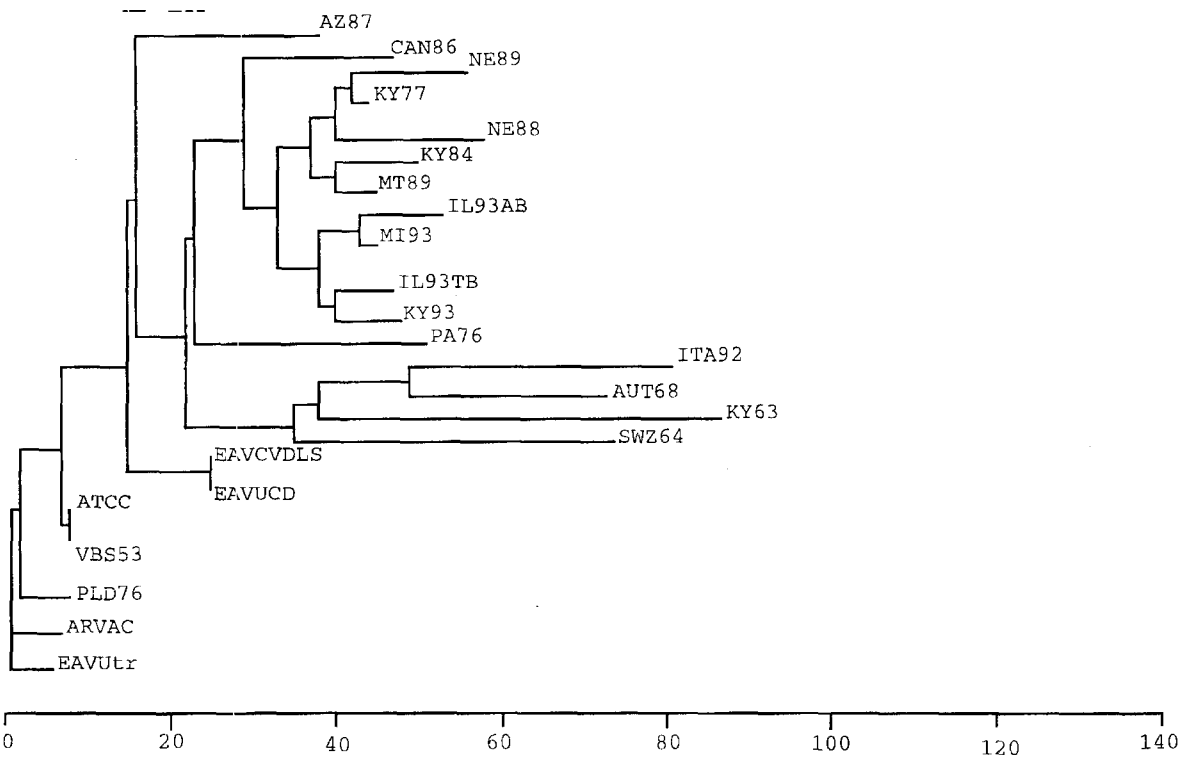
KY63 and all the European isolates, except PLD76, are closely related on the tree, but differ from each other by 37–59 nucleotides. The isolates KY63, ITA92 and AUT68 cluster together and KY63 forms a distinct lineage within the group. KY63 differs from all other isolates by 53–66 nucleotides, thus it is distantly related to both North American and European isolates.

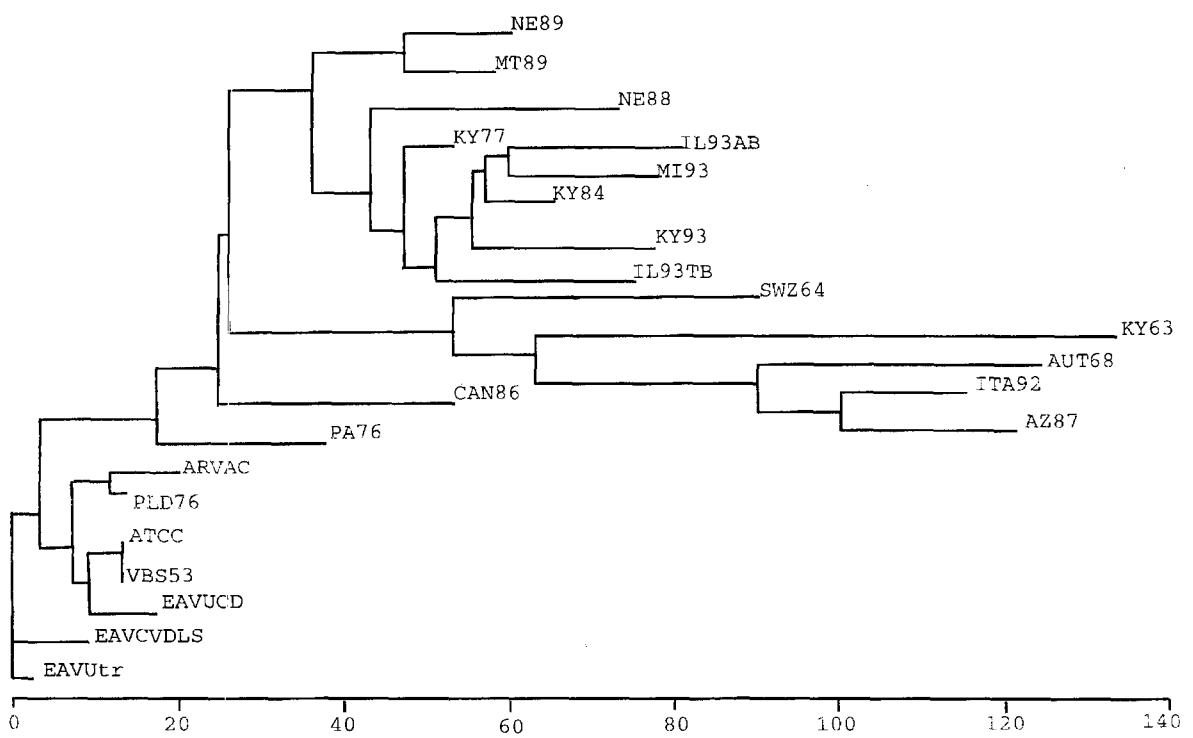
Phylogenetic trees based on the sequences of ORF2 and ORF5 of these viruses are very similar (Fig. 2A, B). The North American isolates, excluding KY63, cluster together in both cases, and CAN86 and PA76 are external to that group, more so in the case of ORF5. Similarly, the placement of the European isolates and KY63 is consistent. ORF5 (85.7–99.7% homology; Balasuriya et al., 1995a) exhibits considerably more sequence divergence than does ORF2 (90.4–99.9% homology).

4. Discussion

The primary goals of this study were to compare the ORF2 sequence of 18 field isolates, three laboratory strains, and the modified live vaccine virus strain of EAV, and to compare the results of this phylogenetic study with that done using ORF5 of the same viruses. Since the G_S protein is a minor envelope protein of EAV, whereas the G_L protein contains the neutralization determinants of the virus, it is likely that immune selective pressures on the two proteins differ. The G_S protein is abundant in infected cells, thus selective pressures on the G_S

Fig. 2. Phylogenetic tree based on sequence analysis of ORF2 of three laboratory strains, the ARVAC® vaccine virus, and 18 field isolates of EAV from North America and Europe. 2B. Phylogenetic tree based on sequence analysis of ORF5 of the same isolates as in 2A. Sequence data from Balasuriya et al. (1995a) were used to create this unrooted tree, using the same method as that in 2A. Vertical lengths are not significant. Horizontal lengths reflect genetic distances and are proportional to the scale bar.


AZ87
CAN86
NE89
KY77
NE88
KY84
MT89
IL93AB
MI93
IL93TB
KY93
PA76
ITA92
AUT68
KY63
SWZ64
EAVCVDLS
EAVUCD
ATCC
VBS53
PLD76
ARVAC
EAVUtr
0
20
40
60
80
100
120
140


NE89
MT89
NE88
KY77
IL93AB
MI93
KY84
KY93
IL93TB
SWZ64
KY63
AUT68
ITA92
AZ87
CAN86
PA76
ARVAC
PLD76
ATCC
VBS53
EAVUCD
EAVCVDLS
EAVUtr
0
20
40
60
80
100
120
140

protein may involve interactions with cellular or other viral proteins.

Data from this study indicate that the G_S protein is under a more stringent evolutionary constraint than the G_L protein. In contrast to our data with ORF2 of EAV, comparison of sequences of ORF2 from two isolates of LDV revealed only 75.3% amino acid homology (Palmer et al., 1995). Similarly, 65% amino acid homology was found between ORF2 of a US and a European isolate of PRRSV (Morozov et al., 1995). EAV is distantly related to LDV and PRRSV, whereas it has been suggested that LDV and PRRSV may have evolved recently from a common ancestor (Plagemann, 1996). The G_S protein of EAV has approximately 20% amino acid homology to the proteins encoded by ORFs2 of LDV and PRRSV (Plagemann, 1996). We are uncertain as to why so little variation is tolerated in the G_S protein of EAV when putative equivalent proteins from isolates of LDV and PRRSV are, apparently, highly variable. However, it is clear that significant variation also occurs among the other proteins of LDV and PRRSV (Palmer et al., 1995; Morozov et al., 1995). The three luminal cysteines of proteins encoded by ORF2 of EAV, LDV and PRRSV are conserved, consistent with a critical role in dimerization that is necessary for release of the protein from the endoplasmic reticulum (de Vries, 1994).

Despite the greater genetic variability of ORF5 of the field isolates of EAV, phylogenetic trees based on ORF2 and ORF5 are similar with only a few notable differences. It previously was suggested that the different trees that resulted from phylogenetic analyses of two different ORFs of EAV might be the result of recombination (Chirnside et al., 1994). However, the differences between the ORF2 and ORF5 phylogenetic analyses are most likely the result of differences in the extent of conservation of the two proteins and we found no evidence of recombination. Interestingly, both phylogenetic trees indicate a close relationship between the highly virulent strain VBS53 and the avirulent ARVAC® vaccine virus. With a few exceptions, such as KY63, viruses from North America and Europe segregate into distinct clades based on sequence analysis of either ORF2 or ORF5.

Sequence analysis of ORF2 of field isolates of EAV indicates that although very closely related, the ORF2 of each isolate is unique. Nucleotide changes common to all isolates imply an ancestor-dependent relationship between the isolates. However, changes that are present in earlier, but not later, isolates from a region, and unique changes that arise, but are not maintained, are not consistent with this hypothesis. These data suggest that nucleotide changes are not necessarily cumulative, there is a reversion of changes to the original nucleotide, and unique changes are constantly arising. Therefore, nucleotide substitutions in EAV do not necessarily follow a linear pattern of accumulation over time and the virus does not evolve via an ancestor-dependent pathway but probably behaves as a quasispecies.

In summary, we have shown the G_S protein to be highly conserved among field isolates of EAV, more so than the G_L protein of EAV and the equivalent proteins of other arteriviruses (LDV and PRRSV). Phylogenetic analyses of ORF2 and ORF5 of the same EAV isolates gave similar but not identical results, which is expected given the different selective pressures likely to be acting on the two proteins. The conserved nature of the G_S protein is consistent with a critical role in EAV replication.

Acknowledgements

The authors gratefully acknowledge Dr. John Gillespie, Department of Genetics, University of California, Davis, for providing helpful information regarding the analysis of the sequences, and John F. Patton and Dustin W. Lee for excellent computer support. These studies were supported by the Grayson Jockey Club Research Foundation, Inc., Lexington, Kentucky, USDA-NRI Competitive Grant 95-3702-2138, the California Center for Equine Health and Performance, University of California, Davis, with funds provided by the Oak Tree Racing Association, the State of California satellite wagering fund, and contributions by private donors.

References

Balasuriya, U.B.R., MacLachlan, N.J., de Vries, A.A.F., Rossitto, P.V. and Rottier, P.J.M. (1995b). Identification of a neutralization site in the major envelope glycoprotein (G_L) of equine arteritis virus. Virology 207, 518–527.

Balasuriya, U.B.R., Rossitto, P.V., DeMaula, C.D. and MacLachlan, N.J. (1993). A 29K envelope glycoprotein of equine arteritis virus expresses neutralization determinants recognized by murine monoclonal antibodies. J. Gen. Virol. 74, 2525–2529.

Balasuriya, U.B.R., Timoney, P.J., McCollum, W. H. and MacLachlan, N.J. (1995a) Phylogenetic analysis of open reading frame 5 of field isolates of equine arteritis virus and identification of conserved and nonconserved regions in the G_L envelope glycoprotein. Virology 214, 690–697.

Brinton, M.A., Gavin, E.I. and Fernandez, A.V. (1986). Genotypic variation among six isolates of lactate dehydrogenase-elevating virus. J. Gen. Virol. 67, 2673–2684.

Bryans, J.T., Doll, E.R., Crowe, M.E.W. and McCollum, W.H. (1957). The blood picture and thermal reaction in experimental viral arteritis of horses. Cornell Vet. 47, 43–52.

Chirnside, E.D., de Vries, A.A.F., Mumford, J.A. and Rottier, J.A. (1995) Equine arteritis virus-neutralizing antibody in the horse is induced by a determinant on the large envelope glycoprotein G_L. J. Gen. Virol. 76, 1989–1998.

Chirnside, E.D., Wearing, C.M., Binns, M.M. and Mumford, J.A. (1994). Comparison of the M and N gene sequences distinguishes variation amongst equine arteritis virus isolates. J. Gen. Virol. 75, 1491–1497.

Chou, P.Y., Fasman, G.D. and Rose, G.D. (1978). Prediction of the secondary structure of proteins from their amino acid sequence. Adv. Enzymol. 47, 45–148.

Conzelmann, K.K., Visser, N., Woensel, P.V. and Thiel, H.J. (1993). Molecular characterization of porcine reproductive and respiratory syndrome virus, a member of the arterivirus group. Virology 193, 329–339.

den Boon, J.A., Snijder, D.J., Chirnside, E.D., de Vries, A.A.F., Horzinek, M.C. and Spaan, W.J.M. (1991). Equine arteritis virus is not a togavirus but belongs to the coronavirus like superfamily. J. Virol. 65, 2910–2920.

Deregt, D., de Vries, A.A.F., Raamsman, M.J.B., Elmgren, L.D. and Rottier, P.J.M. (1994). Monoclonal antibodies to equine arteritis virus proteins identify G_L protein as a target for virus neutralization. J. Gen. Virol. 75, 2439–2444.

de Vries, A.A.F. (1994). The molecular biology of equine arteritis virus. Ph.D. Thesis (Utrecht University, Utrecht, the Netherlands), Addix, Wijk bij Duurstede, The Netherlands.

de Vries, A.A.F., Chirnside, E.D., Horzinek, M.C. and Rottier, P.J.M. (1992). Structural proteins of equine arteritis virus. J. Virol. 66, 6294–6303.

Doll, E.R., Bryans, J.T., McCollum, W.H. and Crowe, M.E.W. (1957). Isolation of a filterable agent causing arteritis of horses and abortion by mares. Its differentiation from the equine abortion (influenza) virus. Cornell Vet. 47, 3–41.

Doll, E.R., Bryans, J.T., Wilson, J.C. and McCollum, W.H. (1968). Immunization against equine viral arteritis using modified live virus propagated in cell cultures of rabbit kidney. Cornell Vet. 58, 497–524.

Domingo, E. and Holland, J.J. (1988). High error rates, population equilibrium, and evolution of RNA replication systems. In: E. Domingo, J.J. Holland and P. Ahlquist (Eds.), RNA Genetics, Vol. III. Variability of RNA genomes, 3–36. CRC Press, Boca Raton, FL.

Domingo, E., Martinez-Salas, E., Sobrino, F., de la Torre, J.C., Portela, A., Ortin, J., Lopes-Glands, C., Perrs-Bern, P., Villanueva, N., Najera, R., Vandepol, S., Steinhauer, D., Depolo, N. and Holland, J. (1985). The quasispecies (extremely heterogeneous) nature of viral RNA genome populations: biological relevance—a review. Gene 40, 1–8.

Faragher, S.G., Meek, A.D.J., Rice, C.M. and Dalgarno, L. (1988). Genome sequences of a mouse-avirulent and a mouse-virulent strain of Ross River virus. Virology 163, 509–526.

Felsenstein, J. (1981a). A likelihood approach to character weighting and what it tells us about parsimony and compatibility. Biol. J. Linn. Soc. 16, 183–196.

Felsenstein, J. (1981b). Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17, 368–376.

Felsenstein, J. (1981c) Evolutionary trees from gene frequencies and quantitative characters: finding maximum likelihood estimates. Evolution 35, 1229–1242.

Felsenstein, J. (1993). PHYLIP (Phylogeny Inference Package) 3.5c manual. University of Washington, Department of Genetics SK-50, Seattle, WA 98195, USA.

Fitch, W.M. and Margoliash, E. (1967). Construction of phylogenetic trees. Science 155, 279–384.

Genetics Computer Group (1994). Program manual for the GCG package, Version 8, 575 Science Drive, Madison, WI 53711.

Glaser, A.L., de Vries, A.A.F., Dubovi, E.J. (1995). Comparison of equine arteritis virus isolates using neutralizing monoclonal antibodies and identification of sequence changes in G_L associated with neutralization resistance. J. Gen. Virol. 76, 2223–2233.

Harry, T.O. and McCollum, W.H. (1981). Stability of viability and immunizing potency of lyophilized, modified equine arteritis live virus vaccine. Am. J. Vet. Res. 42, 1501–1505.

Holland, J., Spindler, K., Horodyski, F., Grabau, E., Nichol, S. and Vandepol, S. (1982) Rapid evolution of RNA genomes. Science 215, 1577–1585.

Jin, L. and Nei, M. (1990) Limitations of the evolutionary parsimony method of phylogenetic analysis. Mol. Biol. Evol. 7 (1), 82–102.

Kimura, M. (1980) A simple model for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16, 111–120.

Kyte, J. and Doolittle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157, 105–132.

McCollum, W.H. (1969). Development of a modified virus strain and vaccine for equine viral arteritis. J. Am. Vet. Med. Assoc. 155 (2), 318–322.

Meulenberg, J.J.M., Hulst, M.M., de Meijer, E.J., Moonen, P.L.J.M., den Besten, A., de Kluyver, E.P., Wensvoort, G. and Moormann, R.J.M. (1993). Lelystad virus, the causative agent of porcine epidemic abortion and respiratory syndrome (PEARS) is related to LDV and EAV. Virology 192, 62–72.

Meulenberg, J.J.M., Hulst, M.M., de Meijer, E.J., Moonen, P.L.J.M., den Besten, A., de Kluyver, E.P., Wensvoort, G. and Moormann, R.J.M. (1994). Lelystad virus belongs to a new virus family, comprising lactate dehydrogenase-elevating virus, equine arteritis virus, and simian hemorrhagic fever virus. Arch. Virol. (Suppl.) 9, 441–448.

Morozov, I, Meng, X.J. and Paul, P.S. (1995) Sequence analysis of open reading frames (ORFs) 2 to 4 of a U.S. isolate of porcine reproductive and respiratory syndrome virus. Arch. Virol. 140, 1313–1319.

Nichol, S.T., Rowe, J.E. and Fitch, W.M. (1993). Punctuated equilibrium and positive Darwinian evolution in vesicular stomatitis virus. Proc. Natl. Acad. Sci. USA 90, 10424–10428.

Palmer, G.A., Kuo, L., Chen, Z., Faaberg, K.S. and Plagemann, P.G.W. (1995). Sequence of the genome of lactate dehydrogenase-elevating virus: heterogenicity between strains P and C. Virology 209, 637–642.

Plagemann, P.G.W. (1996) Lactate dehydrogenase-elevating virus and related viruses. In: B.N. Fields, D.M. Knipe and P.M. Howley (Eds.), Fields Virology, 3rd edn. Lippincott-Raven, New York, pp. 1105–1120.

Plagemann, P.G.W. and Moennig, V. (1992) Lactate dehydrogenase-elevating virus, equine arteritis virus and simian hemorrhagic fever virus: A new group of positive stranded RNA viruses. Adv. Virus Res. 41, 99–192.

Saitou, N. and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.

Snijder, E.J. and Horzinek, M.C. (1993). Toroviruses: replication, evolution and comparison with other members of the coronavirus-like superfamily. J. Gen. Virol. 74, 2305–2316.

Steinhauer, D.A., de la Torre, J.C., Meier, E. and Holland, J.J. (1989). Extreme heterogeneity in populations of vesicular stomatitis virus. J. Virol. 63 (5), 2072–2080.

Stroop, W.G. and Brinton, M.A. (1983). Mouse strain-specific central nervous system lesions associated with lactate dehydrogenase-elevating virus infection. Lab. Invest. 49, 334–345.

Sugita, S., Kondo, T., Sekiguchi, K., Yamaguchi, S., Kamada, M., Nerome, K. and Fukunaga, Y. (1994). Molecular evolution of the M gene of equine arteritis virus. Eq. Infect. Dis. 7, 39–43.

Timoney, P.J. and McCollum, W.H., (1993) Equine viral arteritis. Vet. Clin. North Am. 9 (2), 295–308.

Weaver, S.C., Bellew, L.A. and Rico-Hesse, R. (1992). Phylogenetic analysis of alphaviruses in the Venezuelan equine encephalitic complex and identification of the source of epizootic viruses. Virology 191, 282–290.

Wege, H., Stephenson, J.R., Wege, H. and Meulen, V. (1981). Genetic variation of neurotropic and non-neurotropic murine coronaviruses. J. Gen. Virol. 54, 67–74.

Wensvoort, G., Meulenberg, J.J.M., Murtaugh, M., Benfield, D., Nelson, E.A., Conzelmann, K., Thiel, H.J., Albina, E. and Drew, T.W. (1993). The porcine reproductive and respiratory syndrome characteristics and diagnosis of the causative virus. Vet. Biotechnol. Newslett. 3, 113–120.

Westaway, E.G., Brinton, M.A., Gaidamovich, S.Y., Horzinek, M.C., Igarashi, A., Kaariainen, L., Lvov, D.K., Porterfield, J.S., Russell, P.K. and Trent, D.W. (1985). Togaviridae. Intervirology 24, 125–139.