Genetic Variation of ORFs 3 and 4 of Equine Arteritis Virus

JODI F. HEDGES[1], UDENI B.R. BALASURIYA[1], JB TOPOL[1], DUSTIN W. LEE[2], AND N. JAMES MACLACHLAN[1]

[1]The Bernard and Gloria Salick Equine Viral Diseases Laboratory, Department of Pathology, Microbiology and Immunology, School of Veterinary Medicine, and [2]Department of Computer Science, University of California, Davis, CA 95616, USA

1. INTRODUCTION

The open reading frames (ORFs) 3 and 4 of equine arteritis virus (EAV) encode the GP3 and GP4 proteins, respectively. The GP3 protein of EAV is an extensively glycosylated membrane protein that is likely anchored by the uncleaved signal sequence (Hedges et al., 1999a). ORF 4 is predicted to encode a membrane glycoprotein. The goal of this study was to determine the variation in ORFs 3 and 4 and their encoded GP3 and GP4 proteins amongst a large number of EAV strains, including those amplified directly from the semen of carrier stallions.

2. GENETIC AND PROTEOMIC VARIATION

ORFs 3 and 4 of 70 field isolates and laboratory strains of EAV were compared. The description and passage history of 18 field isolates and one laboratory strain (EAVATCC) of EAV used in these studies, as well as methods have been previously described (Hedges et al., 1996). Also included in the analyses were 29 strains of EAV amplified directly from the semen of 10 different carrier stallions and 14 viruses present in the tissues of foals affected in recent outbreaks of EVA (Hedges et al., 1999b; Patton et

The Nidoviruses (Coronaviruses and Arteriviruses).
Edited by Ehud Lavi *et al.*, Kluwer Academic/Plenum Publishers, 2001.

al., 1998; Balasuriya et al., 1999). The published sequences of ORFs 3 and 4 from eight strains of EAV were also included (Archambault et al., 1997). The number of variable sites and type of substitution in the ORFs 3 and 4 of the 70 strains of EAV were determined by comparison to the prototype EAV ATCC sequence using a program developed in our laboratory (Table 1). The variability of ORFs encoding the four known structural proteins of most of the same strains of EAV was also determined. The ORF 5 was more variable than the other ORFs and had a high proportion of non-synonymous changes. In every ORF, specific synonymous changes occurred more frequently in individual rather than multiple strains, indicating that these changes occur randomly. In contrast, in ORFs 3, 4, 5 and 6 most (77.9%, 63%, 52.2% and 55% respectively) non-synonymous nucleotide changes occurred in more than one virus strain suggesting that many coding nucleotide changes that occur during the evolution of EAV are not random.

Table 1. Percentages of variable sites, non-synonymous and synonymous changes in EAV ORFs as compared to the prototype sequence (EAVATCC)

ORF	Variable Nucleotide Sites (%)	Non-synonymous Changes (%)	Synonymous Changes (%)
2b	26.6	34.6	65.4
3	40	60.3	39.7
4	37.4	37.9	62.1
5	67.8	59.7	40.3
6	24.9	30.5	69.9
7	14.7	31.5	68.5

Hypervariable regions of the GP3 protein were amino acids 1-41 and 88-131. Only two of the six prototypical asparagine (N)-linked glycosylation sites are conserved amongst all 70 strains of EAV; at position 96, where position 98 wobbles between serine (S) and threonine (T), thus preserving the glycosylation site requirements, and at position 106. There is a new glycosylation site at position 39 of the GP3 protein in many strains, and at position 115 where the N substitution occurs only where there is a T or an S in position 117 in most cases, suggesting that the N substitution is advantageous when it is also a glycosylation site. We have determined that the lack of glycosylation of N^{28} and N^{29} had no obvious effect on either the signal sequence cleavage or membrane association (Hedges et al., 1999a). The glycosylation site at asparagine 96, the cysteine residues at positions 13, 33 and 109 and, the tryptophan residue at position 146 are conserved in the GP3 proteins of all 70 strains of EAV and of the other three prototype *Arteriviruses*. Some EAV strains had altered hydrophobicity profiles in the signal sequence region.

a.

MGRAYSGPVA LLCFFLYFCF ICGSVGSNNT TICMHTTSDT SVHLFYAANV TFPSHFQRHF
AAAQDFVVHT GYEYAGVTML VHLFANLVLT FPSLVNCSRP VNVFANASCV QVVCSHTNST
TGLGQLSFSF VDEDLRLHIR PTLICWFALL LVHFLPMPRC RGSQFYLH

b.

MKIYGCILGL LLFVGLPCCW CTFYPCHAAE ARNFTYISHG LGHVHGHEGC RNFINVTHSA
FLYLNPTTLT APAITHCLLL VLAAKMEHPN ATIWLQLQPF GYHVAGDVIV NLEENKRHPY
FKLLRAPALP LGFVAIVYVL LRLVRWAQQC YL

Figure 1. Amino acid sequences of the GP3 protein (a) and the GP4 protein (b) of the prototype EAVATCC. Bold letters indicate amino acids that varied amongst 70 strains of EAV, italicises indicate conservation amongst all *Arteriviruses*, glycosylation sites are underlined, strikethrough letters indicate amino acids found on the GP3 protein of the modified live virus vaccine only.

The GP4 protein was more conserved than was the GP3 protein. The GP4 protein of the EAV ATCC strain has three potential glycosylation sites, but only the site at amino acid 90 was conserved amongst all isolates. The majority of the cysteines in the GP4 protein were conserved and the cysteine at position 19 is conserved in those of the prototype strains of the other *Arteriviruses* as well.

3. CONCLUSION

We have determined the genetic and proteomic diversity of the uncharacterised ORFs 3 and 4 and encoded GP3 and GP4 proteins of EAV. Variation in the predicted signal sequence region amongst the GP3 proteins of different EAV strains may influence membrane interaction, as is perhaps the case with divergent strains of the closely related porcine *Arterivirus* (van Nieuwstadt et al., 1996; Mardassi et al., 1998). There were a large number of non-synonymous substitutions in both ORF 3 and ORF 5 of EAV strains. Amino acid changes in the ORF5 encoded G_L protein alter the neutralisation phenotype of EAV strains (Balasuriya et al., 1997; Hedges et al., 1999b). The GP3 protein is under strong selective pressure during long term persistent infection of stallions, and carrier stallions and horses vaccinated with the live attenuated vaccine virus develop antibodies specific for the GP3 protein (Hedges et al., 1999a,b). The large number of non-synonymous substitutions in ORF 3 amongst strains of EAV is consistent with a strong selective pressure exerted on the GP3 protein. Clearly, individual EAV proteins are under unique selective pressure and structural/functional constraint and the corresponding ORFs evolve independently. Regions of the GP3 protein that vary may be important for the maintenance of persistent infection of carrier stallions, whereas conserved regions are likely important

for protein function in virus replication. The significance of nucleotide and amino acid substitutions identified in this study can now be determined using the infectious cDNA clone of EAV.

ACKNOWLEDGMENTS

These studies were supported by the Center for Equine Health at the University of California-Davis, the Bernard and Gloria Salick Endowment and USDA National Research Initiative Competitive Grant 97-35204-4736.

REFERENCES

Archambault, D., Laganiere, G., Carman, S., and St-Laurent, G. (1997). Comparison of nucleic and amino acid sequences and phylogenetic analysis of open reading frames 3 and 4 of various equine arteritis virus isolates. Vet. Res. 28, 505-516.

Balasuriya, U.B.R., Hedges, J.F., Timoney, P.J., McCollum, W.H., and MacLachlan, N.J. (1999). Genetic stability of equine arteritis virus during horizontal and vertical transmission in an outbreak of equine viral arteritis. J. Gen. Virol. 80, 1949-1958.

Balasuriya, U.B.R., Patton, J.F., Rossitto, P.V., Timoney, P.J., McCollum, W.H., and MacLachlan, N.J. (1997). Neutralization determinants of laboratory strains and field isolates of equine arteritis virus: Identification of four neutralization sites in the amino-terminal ectodomain of the G_L envelope glycoprotein. Virology 232, 114-128.

Hedges, J.F., Balasuriya, U.B.R., Timoney, P.J., McCollum, W.H., and MacLachlan, N.J. (1996). Genetic variation in open reading frame 2 of field isolates and laboratory strains of equine arteritis virus. Virus Res. 42, 41-52.

Hedges, J.F., Balasuriya, U.B.R., and MacLachlan, N.J. (1999a). The open reading frame 3 of equine arteritis virus encodes an immunogenic glycosylated, integral membrane protein. Virology 264, 92-98.

Hedges, J.F., Balasuriya, U.B.R., Timoney, P.J., McCollum, W.H., and MacLachlan, N.J. (1999b). Genetic divergence with emergence of phenotypic variants of equine arteritis virus during persistent infection of stallions. J. Virol. 73, 3672-3681.

Mardassi, H., Gonin, P., Gagnon, C.A., Massie, B., and Dea, S. (1998). A subset of porcine reproductive and respiratory syndrome virus GP3 glycoprotein is released into the culture medium of cells as a non-virion-associated and membrane-free (soluble) form. J. Virol. 72, 6298-6306.

Patton, J.F., Balasuriya, U.B.R., Hedges, J.F., Schweidler, T.M., Hullinger, P.J., and MacLachlan, N.J. (1998). Phylogenetic characterization of a highly attenuated strain of equine arteritis virus from the semen of a persistently infected standardbred stallion. Arch. Virol. 144, 817-827.

van Nieuwstadt, A.P., Meulenberg, J.J.M., van Essen-Zandbergen, A., Peterson-den Besten, A., Bende, R.J., Moormann, R.J., and Wensvoort, G. (1996). Proteins encoded by open reading frames 3 and 4 of the genome of lelystad virus (Arteriviridae) are structural proteins of the virion. J. Virol. 70, 4746-4772.