106

GENETIC VARIATION AND PHYLOGENETIC ANALYSIS OF OPEN READING FRAMES 3 AND 4 OF VARIOUS EQUINE ARTERITIS VIRUS ISOLATES

D. Archambault, G. Laganière, and G. St-Laurent

Université du Québec à Montréal
Département des Sciences Biologiques
C.P. 8888, Succursale Centre-Ville
Montréal, Québec, Canada, H3C 3P8

1. ABSTRACT

The genetic variation in equine arteritis virus (EAV) nonstructural (NS) protein-encoding open reading frames (ORF) 3 and 4 genes was investigated. Nucleotide and deduced amino acid sequences from seven different EAV isolates (one European, one American and five Canadian isolates) and the Arvac vaccine strain were compared with those of the Bucyrus reference strain. ORF 3 nucleotide and amino acid sequence identities amongst these isolates (including the Arvac vaccine strain) and the Bucyrus reference strain ranged from 85.6 to 98.8%, and 85.3 to 98.2%, respectively, whereas ORF 4 nucleotide and amino acid sequence identities ranged from 90.4 to 98.3%, and 90.8 to 97.4%, respectively. Phylogenetic tree analysis based on the ORF 3 nucleotide sequences showed that the European Vienna isolate could be classified into a genetically divergent group from all other isolates and the Arvac vaccine strain. In contrast, a phylogenetic relationship among all EAV isolates and the Arvac vaccine strain based on the ORF 4 nucleotide sequences was observed.

2. INTRODUCTION

Equine arteritis virus (EAV) is the causative agent of equine viral arteritis, a debilitating respiratory disease which may cause abortion in pregnant mares (Timoney and McCollum, 1993). However, the clinical outcome following EAV exposure varies widely

Coronaviruses and Arteriviruses, edited by Enjuanes *et al.*
Plenum Press, New York, 1998

from subclinical infection to systemic EAV disease, indicating that virulence varies among EAV isolates (see Timoney and McCollum, 1993). The EAV genome is a polyadenylated positive single-stranded RNA of 12.7 kb in length, and eight open reading frames (ORFs) have been identified in EAV (Den Boon et al., 1991). ORF 1a/1b encoded for the viral polymerase, and ORFs 2, 5, 6, and 7 coded for EAV structural proteins (De Vries et al., 1992; Den Boon et al., 1991). These include a 14 kDa nucleocapsid protein (N, ORF 7), a 16 kDa unglycosylated membrane protein (M, ORF 6), a heterogeneously glycosylated 30 to 42 kDa large membrane protein (G_L, ORF 5), and a glycosylated 25-kDa small membrane protein (G_s, ORF 2). The ORF 3- and 4-encoded products are believed to be glycosylated nonstructural (NS) proteins (Den Boon et al., 1991; De Vries et al., 1992).

There is only one serotype identified in EAV (see Timoney and McCollum, 1993). However, RNA oligonucleotide fingerprint analysis has indicated genetic variation among various EAV isolates (Murphy et al., 1992). Comparison of M, N, G_L and G_S nucleotide sequences, whose products induce an antibody response (De Vries et al., 1992; Chirnside et al., 1995), has also demonstrated variation among EAV isolates (Chirnside et al., 1994; Balasuriya et al., 1995; Hedges et al., 1996; Lepage et al., 1996; St-Laurent et al., 1997). However, no sequence data are yet available for the EAV NS protein-encoding genes. Because the NS proteins are less likely to be exposed to immunological pressure, the ORF 3 and 4 genes might be relevant indicators of EAV random evolution. In this study, we report the characterization of the genetic variation and evolutionary relationships of various EAV isolates, including the Bucyrus reference (Den Boon et al., 1991) and the Arvac vaccine (Timoney and McCollum, 1993) strains, by comparison of the nucleotide sequences of ORFs 3 and 4. The deduced amino acid sequences of ORF 3 and 4-encoded proteins were also compared.

3. MATERIALS AND METHODS

The EAV isolates originated from Canada (n = 5), the United States (n = 1) and from Europe (n = 1) (Table 1). The Arvac vaccine EAV strain which is the attenuated form of the virulent Bucyrus strain was also included in this study. Virus propagation in cell cultures and EAV genomic RNA extraction of each isolate/strain were performed essentially as described (Lepage et al., 1996; St-Laurent et al., 1997).

Table 1. Characteristics of the equine arteritis virus (EAV) isolates used in this study

Isolate	Origin (year of isolation)	Source	Passage history*
T1329	Ontario, Canada (1988)	Neonatal lung**	RK, P5
11958	Ontario, Canada (1990)	Semen	RK, P5
15492	Ontario, Canada (1991)	Semen	RK, P5
19933	Ontario, Canada (1992)	Semen	RK, P5
Vienna	Vienna, Austria (1968)	Nasal swab	ED, P1/RK, P2
84KY-A1	Kentucky, U.S.A. (1984)	Nasal swab	RK, P5
86AB-A1	Alberta, Canada (1986)	Fetus	RK, P3/V, P1
Arvac***	Fort Dodge Laboratories	–	HK, p131/RK, p111/ED, p24/V, P2

*Cells: RK; rabbit kidney-13, ED; equine dermis, HK; primary horse kidney, V; Vero, P; refers to passage number.
**EAV was isolated from 5 day old standard bred foal.
***Vaccine strain of EAV.

The sense and antisense primer pairs [PEV-20 and PEV-21; (9807–9824, 10503–10486); PEV-30 and PEV-31 (10289–10305, 10783–10766); PEV-40 and PEV-41 (10683–10700, 11165–11148)] were used for reverse transcription-PCR (RT-PCR) amplification in order to obtain overlapping cDNA fragments which represent the entire nucleotide sequence of ORFs 3 and 4 (with an expected length of 492 and 459 bp, respectively). These primers were selected according to the nucleotide sequence of the EAV Bucyrus reference strain genome (Den Boon et al., 1991). RT-PCR assays were carried out as previously described (St-Laurent et al., 1994). The resulting amplified cDNA fragments of each EAV isolate were cloned, and sequenced using the Sanger dideoxynucleotide chain termination method (Sanger et al, 1977). Two or more independent cDNA clones were sequenced from RT-PCR products for each EAV isolate/strain.

Comparison and multiple alignments of nucleotide and deduced amino acid sequences were carried out with the University of Wisconsin Genetics Computer Groups software package (GCG, version 8.0). Phylogenetic analyses based on the analysis of ORF 3 and 4 nucleotide sequences were performed using the DNADIST and FITCH programs of the Phylogenic Inference Package (PHYLIP, version 3.5c) (Felsenstein, 1993). The ORFs 3 and 4 of lactate dehydrogenase-elevating virus (LDV) (Godeny et al., 1993) were used to generate outgroup-rooted trees. Bootstrap analysis was carried out on 1000 replicate datasets to assess the confidence limits of the branch pattern.

4. RESULTS

4.1. Nucleotide and Amino Acid Sequence Analysis of ORF 3

A total of 311 nucleotide mutations were observed within the ORF 3 sequence of all the EAV isolates the analyzed. Of the nucleotide substitutions observed, 45.3% lead to amino acid changes. The expected 492 bp ORF 3 gene was observed for all EAV isolates, except for the 11958 isolate and the Arvac vaccine strain whose ORF 3 length was determined to be 483 and 507 bp, respectively. This was due to nucleotide substitutions at position 481 (11958 isolate) which generates a stop codon, and at position 490 (Arvac vaccine strain) which abolishes the stop codon observed for the Bucyrus strain and all other EAV isolates, rendering the amino acid sequence in frame with a downstream stop codon located at nucleotide positions 505 to 507 (data not shown).

The levels of identity for ORF 3 nucleotide sequences between each EAV isolate/strain and the Bucyrus reference strain ranged from 85.6 (Vienna isolate) to 98.8% (Arvac vaccine strain) (Table 2). Deduced amino acid identities for the ORF 3-encoded product between each EAV isolate/strain and the Bucyrus strain ranged from 85.3 (Vienna isolate) to 98.2% (Arvac vaccine strain) (data not shown). When the EAV isolates were compared to each other, the Vienna and 19933 isolates showed the lowest levels of nucleotide sequence and amino acid identities at 82.7 and 76.1%, respectively. When the Canadian EAV isolates were compared to each other, the T1329 and 19933 isolates showed the lowest level of nucleotide sequence identity (87.8%), whereas the lowest level of amino acid identity was shown between the 19933 and 15492 isolates at 81.0%. The 11958 and 86AB-A1 Canadian isolates were found more closely related to each other with nucleic and amino acid identities of 96.7 and 94.4%, respectively. When the Arvac vaccine strain was compared to all EAV isolates, the lowest levels of nucleotide sequence and amino acid identities were shown with the Vienna isolate at 85.0 and 83.4%, respectively, whereas the highest were observed with the American 84KY-A1 isolate at 93.1 and 91.4%.

Table 2. Nucleotide sequence identities (%) of ORF 3 and 4 genes of equine arteritis virus (EAV) isolates*

Isolate	Bucyrus	T1329	11958	15492	19933	Vienna	84KY-A1	86AB-A1	Arvac**
Bucyrus		93.3	90.1	93.3	89.2	85.6	94.3	92.1	98.8
T1329	98.3		88.4	93.7	87.8	86.2	95.3	90.6	92.1
11958	93.2	94.1		90.1	92.1	83.4	90.3	96.7	89.0
14592	93.2	94.5	92.2		88.6	86.0	94.1	91.3	92.1
19933	93.9	94.3	96.5	93.5		82.7	88.0	93.9	88.0
Vienna	90.4	91.7	89.5	90.2	89.8		86.0	84.3	85.0
84KY-A1	94.8	96.1	94.5	95.4	95.9	91.5		91.3	93.1
86AB-A1	94.3	94.8	97.2	93.2	98.0	90.0	96.1		90.8
Arvac	98.0	97.2	92.6	92.2	93.2	90.0	94.1	93.7	

Upper section: ORF 3
Lower section: ORF 4
*The ORF 3 and 4 nucleotide sequence data reported for these isolates have been deposited in the Genbank Database under accession numbers: AF001079 to AF001094.
**Vaccine strain of EAV.

4.2. Nucleotide and Deduced Amino Acid Sequence Analysis of ORF 4

No insertions or deletions were found in the ORF 4 sequence for any of the EAV isolates including the Arvac vaccine strain. A total of 201 nucleotide changes were observed. Only 34.3% of all substitutions observed lead to amino acid changes. Nucleotide sequence identities with the Bucyrus reference strain ranged from 90.4 (Vienna isolate) to 98.3% (T1329 isolate) (Table 2). Deduced amino acid identities for the ORF 4-encoded product between each EAV isolate/strain and the Bucyrus strain ranged from 90.8 (Vienna isolate) to 97.4% (T1329 isolate) (data not shown). When the EAV isolates and Arvac vaccine strain were compared to each other, the Vienna and 11958 isolates were found to be the most divergent at the nucleotide level at 89.5% of identity, whereas the 19933 and 86AB-A1 Canadian isolates were found to be the most closely related with an identity of 98.0%. The Vienna isolate and Arvac vaccine strain showed the lowest level of amino acid identity at 91.4%, whereas the highest identity level (97.4%) was observed between the 11958 and 86AB-A1 Canadian isolates, and the 15492 and 84KY-A1 isolates. When the Canadian EAV isolates were compared to each other, nucleotide sequence identities ranged from 92.2 (11958 and 15492 isolates) to 98.0% (19933 and 86AB-A1 isolates), whereas amino acid identities between these isolates ranged from 94.7 (15492 and 11958 isolates, and 86AB-A1 and 15492 isolates) to 97.4% (11958 and 86AB-A1 isolates). The Arvac vaccine strain ORF 4 sequence showed nucleotide and amino acid sequence identities ranging from 90.0 (Vienna isolate) to 97.2% (T1329 isolate), and from 91.4 (Vienna isolate) to 95.4% (T1329, 19933 and 86AB-A1 isolates), respectively.

4.3. Phylogenetic Analysis of EAV Isolates Based on Their ORFs 3 and 4 Nucleotide Sequences

The phylogenetic analysis based on nucleotide sequence of ORF 3 (Figure 1) showed that the European Vienna isolate evolved independently and formed a distinct phylogenetic group (group E) from all other isolates and the Arvac vaccine strain. The North American isolates and the Arvac vaccine strain were clustered into a large clade

(group A) which could be further subdivided into two distinct monophylogenic lineages. However, the confidence value obtained by bootstrap analysis of the lineage that includes the Bucyrus reference and Arvac vaccine strains and three other isolates (84KY-A1, T1329 and 15492) was low (37.3%). This indicates that the order of descent of this lineage was not fully resolved by this analysis, and suggests that the North American isolates and the Arvac vaccine strain are likely to represent a large phylogenetically-related group on the basis of the ORF 3 nucleotide sequences.

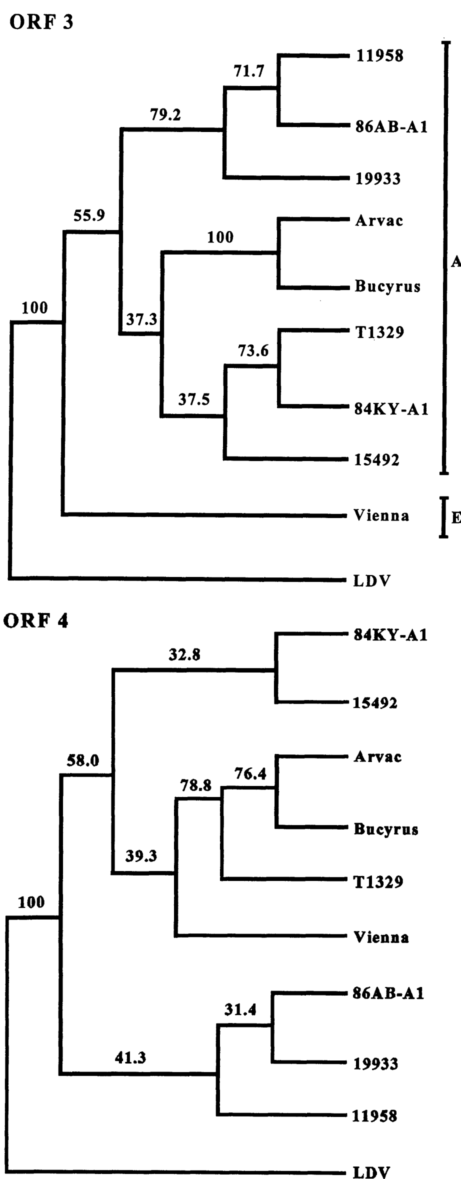

Figure 1. Phylogenetic relationship of nine equine arteritis virus (EAV) isolates including the Bucyrus reference and Arvac vaccine strains based on their ORF 3 or ORF 4 nucleotide sequences. The ORFs 3 and 4 of lactate dehydrogenase-elevating virus (LDV) were used as outgroups. The numbers indicate the branching node confidence values (%) obtained by bootstrap analyses.

In contrast, the phylogenetic tree based on the nucleotide sequence of ORF 4 (Figure 1) showed that the European Vienna isolate could not be clustered in a distinct group from the North American isolates. Although three Canadian isolates (11958, 19933 and 86AB-A1) might be classified as a distinct monophylogenic group from all other isolates, the order of descent of this lineage was not fully resolved by this analysis (confidence value of 41.3%). Indeed, the high levels of nucleotide and amino acid sequence identities described above indicate a close relationship between the ORF 4 sequences in all analyzed EAV isolates.

5. DISCUSSION

The results presented in this study demonstrate a high degree of nucleotide and deduced amino acid sequence conservation in the EAV ORF 4 among the analyzed EAV isolates. In contrast, the EAV ORF 3 gene was less conserved and was found to be genetically divergent such that the European Vienna isolate and the North American isolates could be classified into distinct phylogenetic groups on the basis of nucleotide sequences.

The results also showed that the EAV ORF 3 appears to be more variable genetically, based on percentages of identity of the nucleotide and amino acid sequence, than the ORF 2- ,5- , 6- and 7-encoded EAV structural proteins (Chirnside et al., 1994; Balasuriya et al., 1995; Hedges et al., 1996; Lepage et al., 1996; St-Laurent et al., 1997), the elicitors of antibody responses in EAV-infected horses (De Vries et al., 1992; Chirnside et al., 1995). This is particularly surprising for the ORF 5-encoded G_L protein which expresses the EAV neutralizing epitopes (Deregt et al., 1994; Balasurya et al., 1995; Chirnside et al., 1995; Glaser et al., 1995), and which, therefore, is likely to be the primary target of selective pressure by the host immune system.

The significance of the genetic variation observed in EAV ORF 3 has yet to be determined. The ORF 3-encoded product is believed to be an NS protein (Den Boon et al., 1991). However, the structural nature of the ORF 3-encoded product can not be ruled out. Genetic variation has also been reported with NS proteins of other RNA viruses, the conserved regions of which are believed to be important in the regulation of virus replication (Rao et al., 1995; Oberste et al., 1996). Analysis of the deduced amino acid sequence alignment of the ORF 3-gene product of all EAV isolates with the reference Bucyrus strain revealed the presence of two variable regions that encompassed residue positions 3 to 41 in the N-terminal region, and positions 98 to 121 in the central portion of the predicted 163 amino acid ORF 3-encoded product (data not shown). The remaining amino acid sequence of the protein was relatively well conserved among EAV isolates. Thus, it is likely that these conserved regions within the ORF-3 encoded protein might also have an important role in EAV pathogenesis. More in-depth studies are needed to resolve the molecular basis of EAV ORF 3 and 4 genetic variation and to determine the biological function of ORFs 3 and 4 in EAV biogenesis.

ACKNOWLEDGMENTS

This work was supported by an operating grant from the National Sciences and Engineering Research Council of Canada to D. Archambault. D. Archambault is the holder of a research scholarship from the Fonds de la Recherche en Santé du Québec (FRSQ). We also thank Peter Timoney and William McCollum (Gluck Equine Research Center, Col-

lege of Agriculture, University of Kentucky, Lexington, KY, USA) for providing the Canadian 86AB-A1, American and European EAV isolates, and the Arvac vaccine strain, and Susy Carman (Veterinary Laboratory Services, Guelph, Ontario, Canada) for providing the Canadian EAV field isolates.

REFERENCES

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

Balasuriya U.B.R., Timoney P.J., McCollum W.H. and MacLachlan N.J., 1995, 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.

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

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

Den Boon J.A., Snijder E.J., Chirnside E.D., De Vries A.A.F., Horzinek M.C. and Spann W.J.M., 1991, Equine arteritis virus is not a togavirus but belongs to the coronaviruslike superfamily, *J. Virol.* **65:** 2910–2920.

Deregt D., De Vries A.A.F., Raamsman M.J.B. and Elmgren L.D., 1994, Monoclonal antibodies to equine arteritis virus protein identify the G_L protein as a target for virus neutralization. *J. Gen. Virol,* **75:** 2439–2444.

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.

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

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.

Godeny E.K., Chen L., Kumar S.N., Methven S.L., Koonin E.V. and Binton M.A., 1993, Complete genomic sequence and phylogenetic analysis of the lactate dehydrogenase-elevating virus (LDV), *Virology* **194:** 585–596.

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 isolate and laboratory strains of equine arteritis virus, *Vir. Res.* **42:** 41–52.

Lepage N., St-Laurent G., Carman S. and Archambault D., 1996, Comparison of nucleic and amino acid sequences and phylogenic analysis of the Gs protein of various equine arteritis virus isolates, *Virus Genes* **13:** 87–91.

Murphy T.W., McCollum W.H., Timoney P.J., Klingeborn B.W., Hyllseth B., Golnik W. and Erasmus B., 1992, Genomic variability among globally distributed isolates of equine arteritis virus, *Vet. Microbiol.* **32:** 101–115.

Oberste M.S., Parker M.D. and Smith J.F., 1996, Complete sequence of Venezuelan equine encephalitis virus subtype IE reveals conserved and hypervariable domains within the C terminus of nsP3, *Virology* **219:** 314–320.

Rao C.D., Das M., Ilango P., Lalwani R., Bhargavi S.R. and Gowda K., 1995, Comparative nucleotide and amino acid sequence analysis of the sequence-specific RNA-binding rotavirus nonstructural protein NSP 3, *Virology* **207:** 327–333.

Sanger F., Nicklen S. and Coulson A.R., 1977, DNA sequencing with chain-terminating inhibitors, *Proc. Natl. Acad. Sci.* USA **74:** 5463–5467.

St-Laurent G., Morin G. and Archambault D., 1994, Detection of equine arteritis virus following amplification of structural and nonstructural viral genes by reverse transcription-PCR, *J. Clin. Microbiol.* **32:** 658–665.

St-Laurent G., Lepage N., Carman S. and Archambault D., 1997, Genetic and amino acid analysis of the G_L protein of Canadian, American and European equine arteritis virus isolates, *Can. J. Vet. Res.* **61:** 72–76.

Timoney P.J. and McCollum W.H., 1993, Equine viral arteritis, *Vet. Clinics North Amer.: Equine practice* **9:** 295–309.