Comparison of nucleic and amino acid sequences and phylogenetic analysis of open reading frames 3 and 4 of various equine arteritis virus isolates
D Archambault, G Laganière, S Carman, G St-Laurent

To cite this version:
D Archambault, G Laganière, S Carman, G St-Laurent. Comparison of nucleic and amino acid sequences and phylogenetic analysis of open reading frames 3 and 4 of various equine arteritis virus isolates. Veterinary Research, BioMed Central, 1997, 28 (6), pp.505-516. <hal-00902498>
Comparison of nucleic and amino acid sequences and phylogenetic analysis of open reading frames 3 and 4 of various equine arteritis virus isolates

D Archambault¹*, G Laganière¹, S Carman², G St-Laurent¹

¹ Département des Sciences Biologiques, Université du Québec à Montréal, C P 8888, Succ Centre-Ville, Montreal, PQ H3C 3P8; ² Animal Health Laboratory, Laboratory Services Division, University of Guelph, ON N1H 6R8, Canada

(Received 27 March 1997; accepted 28 July 1997)

Summary – The genetic variation in equine arteritis virus (EAV) protein-encoding open reading frames (ORFs) 3 and 4 genes was investigated. Nucleic 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 between 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 analysis and estimation of genetic distances based on the ORF 3 nucleic acid 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, the nucleic acid sequences of ORF 4 were found to be less variable, with a closer phylogenetic relationship evident among the EAV isolates and the Arvac vaccine strain.

equine arteritis virus / open reading frames 3 and 4 / genetic variation

Résumé – Comparaison des séquences d’acides nucléiques et d’acides aminés et analyse phylogénétique des cadres de lecture ouverts 3 et 4 de divers isolats du virus de l’artérite équine. Dans cette étude, les séquences d’acides nucléiques et d’acides aminés prédites des protéines associées aux cadres de lecture ouverts (CLO) 3 et 4 de sept isolats du virus de l’artérite équine (VAE) et de la souche vaccinale Arvac ont été comparées à celles de la souche de référence Bucyrus. Les niveaux d’identité des acides nucléiques et des acides aminés associés au CLO 3 ont varié de 85,6 à 98,8% et de 85,3 à 98,2% respectivement, alors que ceux associés au CLO 4 ont varié de 90,4 à 98,3%.

* Correspondence and reprints
Tel: (1) 514 987 3000, ext 4622; fax: (1) 514 987 4647; e-mail: archambault.denis@uqam.ca
et de 90,8 à 97,4 % respectivement. L'analyse phylogénétique basée sur les séquences d'acides nucléiques du gène CLO 3 a indiqué que l'isolat européen Vienne pouvait être classé en un groupe distinct de tous les autres isolats de VAE et de la souche vaccinale Arvac. En revanche, tous les isolats de VAE et la souche vaccinale Arvac ont montré une plus étroite relation phylogénétique sur la base des séquences d'acides nucléiques du gène CLO 4.

**Virus de l'artérite équine / cadres de lecture ouverts 3 et 4 / variation génétique**

**INTRODUCTION**

Equine arteritis virus (EAV) is an enveloped positive-stranded RNA virus of 70 nm in diameter that was first isolated in Bucyrus, Ohio (Doll et al, 1957). EAV, present in horses throughout the world, is the causative agent of equine viral arteritis, a respiratory disease with the most severe form resulting in abortion in pregnant mares (Timoney and McCollum, 1993). Several EAV strains showing different levels of pathogenicity have been reported (Burki, 1970; McCollum and Swerczek, 1978).

EAV belongs to the arterivirus group, which includes lactate dehydrogenase-elevating virus (LDV), porcine reproductive and respiratory syndrome virus (PRRSV) and simian haemorragic fever virus (SHFV) (Cavanagh et al, 1994). The EAV genome is a polyadenylated single-stranded RNA of 12.7 kb in length. Eight open reading frames (ORFs) have been identified in EAV (Den Boon et al, 1991). ORFs 1a/1b encode the viral polymerase (Snijder et al, 1994). One minor and three major structural EAV proteins have been identified and assigned to ORFs 2, 5, 6 and 7, respectively. ORFs 6 and 7 encode an unglycosylated membrane (M) protein of 16 kDa and a 14 kDa nucleocapsid (N) protein, respectively (Den Boon et al, 1991; De Vries et al, 1992). ORF 5 codes for the heterogeneously glycosylated 30 to 42 kDa large (G_

small (G_

small membrane protein, which only represents 1 to 2% of the total virion proteins (De Vries et al, 1992). ORF 3- and 4-encoded products are believed to be glycosylated nonstructural proteins (Den Boon et al, 1991; De Vries et al, 1992).

Although it is believed that there is only one EAV serotype, minor antigenic differences among various EAV isolates have been reported (Fukunaga and McCollum, 1977; Fukunaga et al, 1994; Glaser et al, 1995). RNA oligonucleotide fingerprint analysis has indicated genetic variation between EAV isolates (Murphy et al, 1992). Comparison of M, N, G_

and G_

nucleic acid sequences, whose products have been shown to induce variable antibody responses (De Vries et al, 1992; Chirnside et al, 1995; Kheyar et al, 1997), has also demonstrated variation among EAV isolates (Chirnside et al, 1994; Balasuriya et al, 1995b; Hedges et al, 1996; Lepage et al, 1996; St-Laurent et al, 1997). However, these reported variations in the genetic or phylogenetic relationships among EAV isolates have not yet been supported by sequence information for ORF 3- and 4-encoded products. Because these are believed to be nonstructural proteins (Den Boon et al, 1991; De Vries et al, 1992), the ORF 3- and 4-encoded proteins are less likely to be exposed to immunological pressure. Therefore, ORFs 3 and 4 might be better indicators of random evolution than M, N, G_

or G_

nucleic acid sequences. The ORF 3 and 4 proteins and their genes should also be studied because they may have important roles in EAV biogenesis.
The goal of this study was to determine the genetic variation and evolutionary relationships between 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 nucleic acid sequences of ORFs 3 and 4, as determined by reverse transcriptase-polymerase chain reaction (RT-PCR) sequencing. The deduced amino acid sequences of ORF 3- and 4-encoded proteins of the EAV isolates/strains were also compared.

**MATERIALS AND METHODS**

**Viruses and cells**

The characteristics of the EAV isolates used in the present study are found in table I. These isolates originated from Canada ($n = 5$), the United States ($n = 1$) and from Europe ($n = 1$). The Arvac vaccine EAV strain, which is the attenuated form through cell culture passages of the virulent Bucyrus strain, was also included in this study. Viruses were propagated in rabbit kidney (RK-13) or Vero cells and purified by ultracentrifugation as described previously (St-Laurent et al, 1994).

**Oligonucleotide primers**

The sense and antisense primer pairs [PEV-20 and PEV-21: nucleotides (nt) 9807-9824, nt 10503-10486; PEV-30 and PEV-31: nt 10289-10305, nt 10783-10766; PEV-40 and PEV-41: nt 10683-10700, nt 11165-11148] were selected according to the EAV Bucyrus reference strain genome sequence published by Den Boon et al (1991) (Genbank Database accession number X53459). These primers were used for RT-PCR amplification in order to obtain overlapping cDNA fragments, which represented the entire nucleic acid sequence of ORFs 3 and 4 (with an expected length of 492 and 459 bp, respectively).

**RT-PCR, cloning and sequencing**

EAV genomic RNA of each isolate/strain was extracted from virus particles by the guanidium isothiocyanate method (Chomczynski and Sacchi, 1987) using the Trizol reagent (Gibco/BRL, Mississauga, Ontario, Canada). The RT-PCR assays were carried out as described (St-Laurent et al, 1994). Briefly, genomic viral RNA was reverse transcribed at 42 °C for 90 min by the avian myeloblastosis virus (AMV) reverse transcriptase (Pharmacia, Uppsala, Sweden) with the antisense primer to generate first-strand cDNA. This first-strand cDNA template was then copied into double-stranded cDNA by using the Taq DNA

---

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

<table>
<thead>
<tr>
<th>Isolate</th>
<th>Origin (year of isolation)</th>
<th>Source</th>
<th>Passage history</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1329</td>
<td>Ontario, Canada (1988)</td>
<td>Neonatal lung$^a$</td>
<td>RK, P5$^b$</td>
</tr>
<tr>
<td>11958</td>
<td>Ontario, Canada (1990)</td>
<td>Semen</td>
<td>RK, P5</td>
</tr>
<tr>
<td>15492</td>
<td>Ontario, Canada (1991)</td>
<td>Semen</td>
<td>RK, P5</td>
</tr>
<tr>
<td>19933</td>
<td>Ontario, Canada (1992)</td>
<td>Semen</td>
<td>RK, P5</td>
</tr>
<tr>
<td>Vienna</td>
<td>Vienna, Austria (1968)</td>
<td>Nasal swab</td>
<td>ED, P1/RK, P2</td>
</tr>
<tr>
<td>86AB-A1</td>
<td>Alberta, Canada (1986)</td>
<td>Fetus</td>
<td>RK, P3/V, P1</td>
</tr>
<tr>
<td>Arvac$^c$</td>
<td>Fort Dodge Laboratories</td>
<td>—</td>
<td>HK, P131/RK, P111/ED, P24/V, P2</td>
</tr>
</tbody>
</table>

$^a$EAV was isolated from 5-day-old standardbred foal; $^b$cells: RK, rabbit kidney-13; ED, equine dermis; HK, primary horse kidney; V, Vero; P refers to passage number; $^c$valvaccine strain of EAV.
polymerase (Promega Corporation, Madison, Wisconsin, USA) using both sense and antisense primers (100 pmol each). DNA amplification consisted of 30 cycles of 1 min at 95 °C, 1 min at 50 °C and 2 min at 72 °C and a final extension step at 72 °C for 10 min in a programmable thermal cycler (Temp Tronic; Barnstead/Thermolyne, Dubuque, Iowa, USA). The resulting amplified cDNA fragments of each EAV isolate were cloned into PCR'M II TA cloning (Invitrogen, San Diego, California, USA) or pBluescript KS(+) (Stratagene, Palo Alto, California, USA) vectors, 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.

Computer sequence analysis

Comparison and multiple alignments of nucleic acid and deduced amino acid sequences were carried out with the Gap, Pileup and Distance programs from the software provided by the University of Wisconsin Genetics Computer Groups (GCG, version 8.0) (Devereux et al, 1984), and the PSORT program (Nakai and Kanehisa, 1992). Phylogenetic analyses based on ORF 3 and 4 nucleic acid sequences were performed using the DNADIST and FITCH programs of the Phylogenetic Inference Package (PHYLIP, version 3.5c) (Felsenstein, 1993). The ORFs 3 and 4 of 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.

RESULTS

Nucleic and deduced amino acid sequence analysis of ORF 3

Complete nucleic and deduced amino acid sequences of ORF 3 (and ORF 4, see below) were determined for seven EAV isolates and the Arvac vaccine strain. A total of 311 nucleic acid mutations was observed within the ORF 3 sequence of all analysed EAV isolates, including both transitions (70.7%) and transversions (29.3%). Of the nucleic acid substitutions observed, 45.3% led to amino acid changes. The remaining nucleic acid substitutions observed were silent, a feature not uncommon for RNA viruses (Strauss and Strauss, 1988) consistent with the quasispecies model as proposed by Eigen (1993). No deletions or insertions of nucleic acids were found in the ORF 3 gene for six analysed EAV isolates.

ORF 3 nucleic and amino acid identities with the Bucyrus reference strain ranged from 85.6 (Vienna isolate) to 98.8% (Arvac vaccine strain), and 85.3 (Vienna isolate) to 98.2% (Arvac vaccine strain), respectively (table II). When the EAV isolates were compared to each other, the Vienna and 19933 isolates showed the lowest levels of nucleic 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 nucleic acid 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 to be 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 nucleic 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%.

Analysis of the deduced amino acid sequence alignment (fig 1) of ORF 3 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 ORF 3 product. Cysteine residue substitutions were observed at position 19 (Vienna and 84KY-A1 isolates), position 22 (all EAV isolates except the Arvac vaccine strain), and position 160 (Arvac vac-
cine strain). Only the Vienna isolate amino acid sequence showed a new cysteine residue at position 3. The two putative N-linked glycosylation sites observed for the Bucyrus reference strain at residue positions 28 and 118 were absent for some of the EAV isolates (i.e., the 11958, Vienna and 84KY-A1 isolates), and all EAV isolates except the Arvac vaccine strain, respectively. However, new putative N-linked glycosylation sites in EAV ORF 3 sequences were found at amino acid residue positions 39 (T1329, Vienna and 84KY-A1 isolates), 115 (T1329, 15492 and 84KY-A1 isolates) or 117 (11958 and 86AB-A1 isolates).

Finally, the ORF 3 gene product of the 11958 isolate and the Arvac vaccine strain were three amino acids shorter and five amino acids longer, respectively, than the predicted 163 amino acid ORF-3 gene product of the Bucyrus reference strain and other EAV isolates analysed in this study. This was due to nucleic acid substitutions at position 481 (11958 isolate), which generated a stop codon, and at position 490 (Arvac vaccine strain), which abolished 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 nucleic positions 505 to 507. The biological significance of these base substitutions has yet to be determined.

### Nucleic and deduced amino acid sequence analysis of ORF 4

No insertions or deletions of nucleic acids were found in the ORF 4 sequence for any of the EAV isolates and the Arvac vaccine strain analysed. A total of 201 nucleic acid changes was observed and found to be transitions (82.1%) or transversions (17.9%). Of the nucleic acid substitutions observed, 34.3% led to amino acid changes.

Nucleic and deduced amino acid identities with the Bucyrus reference strain ranged from 90.4 (Vienna isolate) to 98.3% (T1329 isolate), and 90.8 (Vienna isolate) to 97.4% (T1329 isolate), respectively (table III). 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 nucleic acid level at 89.5% of identity, whereas the 19933 and 86AB-A1 Canadian isolates were

<table>
<thead>
<tr>
<th>Isolate</th>
<th>Bucyrus</th>
<th>T1329</th>
<th>11958</th>
<th>15492</th>
<th>19933 Vienna</th>
<th>84KY-A1</th>
<th>86AB-A1</th>
<th>Arvac²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bucyrus</td>
<td>93.3</td>
<td>90.1</td>
<td>93.3</td>
<td>89.2</td>
<td>85.6</td>
<td>94.3</td>
<td>92.1</td>
<td>98.8</td>
</tr>
<tr>
<td>T1329</td>
<td>92.0</td>
<td>88.4</td>
<td>91.3</td>
<td>91.7</td>
<td>86.2</td>
<td>95.3</td>
<td>90.6</td>
<td>92.1</td>
</tr>
<tr>
<td>11958</td>
<td>88.7</td>
<td>83.7</td>
<td>90.1</td>
<td>92.1</td>
<td>83.4</td>
<td>90.3</td>
<td>96.7</td>
<td>89.0</td>
</tr>
<tr>
<td>15492</td>
<td>91.6</td>
<td>91.1</td>
<td>84.4</td>
<td>88.6</td>
<td>86.0</td>
<td>94.1</td>
<td>91.3</td>
<td>92.1</td>
</tr>
<tr>
<td>19933</td>
<td>85.9</td>
<td>82.2</td>
<td>86.9</td>
<td>81.0</td>
<td>82.7</td>
<td>88.0</td>
<td>93.9</td>
<td>88.0</td>
</tr>
<tr>
<td>Vienna</td>
<td>85.3</td>
<td>84.0</td>
<td>78.7</td>
<td>84.7</td>
<td>76.1</td>
<td>86.0</td>
<td>84.3</td>
<td>85.0</td>
</tr>
<tr>
<td>84KY-A1</td>
<td>93.2</td>
<td>91.4</td>
<td>85.6</td>
<td>89.6</td>
<td>81.6</td>
<td>83.4</td>
<td>91.3</td>
<td>93.1</td>
</tr>
<tr>
<td>84AB-A1</td>
<td>91.4</td>
<td>87.1</td>
<td>94.4</td>
<td>87.1</td>
<td>88.3</td>
<td>81.0</td>
<td>86.5</td>
<td>90.8</td>
</tr>
<tr>
<td>Arvac²</td>
<td>98.2</td>
<td>90.2</td>
<td>86.9</td>
<td>90.8</td>
<td>84.0</td>
<td>83.4</td>
<td>91.4</td>
<td>89.6</td>
</tr>
</tbody>
</table>

Upper section: percentage of nucleic acid sequence identity; lower section: percentage of amino acid sequence identity; *the ORF 3 nucleotide sequence data reported for these isolates have been deposited with the Genbank Database under accession numbers AF001079 to AF001086; **vaccine strain of EAV.

Table II. Nucleic and deduced amino acid sequence identities (%) of ORF 3 of equine arteritis virus (EAV) isolates².
**Fig 1.** Alignment of deduced amino acid sequences of the ORF 3-encoded protein of seven equine arteritis virus (EAV) isolates and the Arvac vaccine strain with the Bucyrus reference strain.
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, nucleic acid 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 nucleic 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.

Analysis of the deduced amino acid sequence alignment (fig 2) of ORF 4 of all EAV isolates (except the Vienna isolate, see bel w) including the Arvac vaccine strain with the Bucyrus reference strain showed no major variation in either the number and position of cysteine residues or putative N-linked glycosylation sites. Only the Vienna isolate amino acid sequence showed an additional cysteine residue at position 4. Amino acid substitutions were scattered throughout the entire sequence. However, all EAV isolates/strain, when compared to the Bucyrus reference strain, exhibited an amino acid change at residue position 149 (R → Q) in the ORF 4-encoded product.

Phylogenetic relationship of EAV isolates based on their ORF 3 and 4 nucleic acid sequences

In order to illustrate phylogenetic relationship among EAV isolates, phenograms were constructed based on ORF 3 or ORF 4 nucleic acid sequences (fig 3). The phylogenetic analysis based on nucleic acids of ORF 3 (fig 3A) showed that the European Vienna isolate evolved independently and formed a distinct phylogenetic group (group I) 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 II), which could be further subdivided into two distinct monophyletic lineages. However, the confidence value obtained by bootstrap analysis of the lineage that included the Bucyrus reference and Arvac vaccine strains and three other isolates (84KY-A1, T1329 and 15492) was low (37.3%), thereby indicating that the order of descent of this lineage was not fully resolved by this analysis. This suggested 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 nucleic acid sequences.

In contrast, the phylogenetic tree based on the nucleotide sequence of ORF 4 (fig 3B) 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 in a distinct monophyletic group from all other isolates, the order of descent of this lineage was not fully resolved by this analysis (confidence value of 41.3%). Thus, the genetic variation observed in EAV ORF 4 does not appear to be consistent with the different geographic origins. Indeed, the high levels of nucleic and amino acid identities (table III) indicated a close relationship between the ORF 4 sequences in all analysed EAV isolates.

DISCUSSION

In previous studies (Burki, 1970; McCol-lum and Swerczek, 1978), epidemiological surveys revealed a great degree of variation in the severity of clinical symptoms during
Fig 2. Alignment of deduced amino acid sequences of the ORF 4-encoded protein of seven equine arteritis virus (EAV) isolates and the Arvac vaccine strain with the Bucyrus reference strain.
Fig 3. Phylogenetic relationship of nine equine arteritis virus (EAV) isolates/strains including the Bucyrus reference and Arvac vaccine strains based on their ORF 3 (A) or ORF 4 (B) nucleic acid 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.

Table III. Nucleic and deduced amino acid sequence identities (%) of ORF 4 of equine arteritis virus (EAV) isolatesa.

<table>
<thead>
<tr>
<th>Isolate</th>
<th>Bucyrus</th>
<th>T1329</th>
<th>11958</th>
<th>15492</th>
<th>19933</th>
<th>Vienna</th>
<th>84KY-A1</th>
<th>86AB-A1</th>
<th>Arvac</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bucyrus</td>
<td>98.3</td>
<td>93.2</td>
<td>93.2</td>
<td>93.9</td>
<td>90.4</td>
<td>94.8</td>
<td>94.3</td>
<td>98.0</td>
<td></td>
</tr>
<tr>
<td>T1329</td>
<td>97.4</td>
<td>94.1</td>
<td>94.5</td>
<td>94.3</td>
<td>91.7</td>
<td>95.6</td>
<td>94.8</td>
<td>97.7</td>
<td></td>
</tr>
<tr>
<td>11958</td>
<td>94.1</td>
<td>95.4</td>
<td>92.2</td>
<td>96.5</td>
<td>89.8</td>
<td>94.5</td>
<td>94.1</td>
<td>92.6</td>
<td></td>
</tr>
<tr>
<td>15492</td>
<td>94.1</td>
<td>96.7</td>
<td>94.7</td>
<td>93.5</td>
<td>90.2</td>
<td>95.4</td>
<td>95.4</td>
<td>93.2</td>
<td></td>
</tr>
<tr>
<td>19933</td>
<td>96.0</td>
<td>96.0</td>
<td>96.7</td>
<td>95.4</td>
<td>89.8</td>
<td>95.9</td>
<td>98.0</td>
<td>93.2</td>
<td></td>
</tr>
<tr>
<td>Vienna</td>
<td>90.8</td>
<td>92.8</td>
<td>92.1</td>
<td>93.4</td>
<td>93.4</td>
<td>91.5</td>
<td>90.0</td>
<td>90.0</td>
<td></td>
</tr>
<tr>
<td>84KY-A1</td>
<td>94.1</td>
<td>96.7</td>
<td>96.0</td>
<td>97.4</td>
<td>96.7</td>
<td>94.7</td>
<td>96.1</td>
<td>94.1</td>
<td></td>
</tr>
<tr>
<td>86AB-A1</td>
<td>95.4</td>
<td>95.4</td>
<td>97.4</td>
<td>94.7</td>
<td>96.7</td>
<td>92.8</td>
<td>96.0</td>
<td>93.7</td>
<td></td>
</tr>
<tr>
<td>Arvacb</td>
<td>96.7</td>
<td>95.4</td>
<td>93.4</td>
<td>92.1</td>
<td>95.4</td>
<td>91.4</td>
<td>93.4</td>
<td>95.4</td>
<td></td>
</tr>
</tbody>
</table>

Upper section: percentage of nucleic acid sequence identity; lower section: percentage of amino acid sequence identity. aThe ORF 4 nucleotide sequence data reported for these isolates have been deposited with the Genbank Database under accession numbers AF001087 to AF001094. bVaccine strain of EAV.
several EAV outbreaks. These observations were thought to be an indication of genetic variation among EAV isolates. A few years later, RNA oligonucleotide fingerprint analysis indeed detected the existence of genomic variation between EAV isolates (Murphy et al, 1992). This genetic variability was then confirmed, at variable degrees, by comparison of M, N, Gs and G_L structural protein-encoding genes of various EAV isolates (Chirnside et al, 1994; Balasuriya et al, 1995b; Hedges et al, 1996; Lepage et al, 1996; St-Laurent et al, 1997), which are the elicitors of antibody responses in EAV-infected horses (De Vries et al, 1992; Chirnside et al, 1995; Kheyar et al, 1997). This is particularly surprising since the ORF 5-encoded G_L protein expresses EAV-neutralizing determinants (Deregts et al, 1994; Balasuriya et al, 1995a; Chirnside et al, 1995; Glaser et al, 1995), and would therefore be expected to be a good candidate for being the primary target of a selective pressure by the host immune system. However, this does not preclude the possibility that other EAV gene products (including those of ORFs 3 and 4) may also contain neutralizing epitopes, able to induce neutralizing antibodies. Therefore genes other than ORF 5 could also be subjected to pressure by the host immune response. Further studies are needed to clarify this point.

The results presented here also showed that the EAV ORF 3 appeared to be more variable genetically, based on the percentages of nucleic and amino acid identity, than ORFs 2, 5, 6 and 7 (Chirnside et al, 1994; Balasuriya et al, 1995b; Hedges et al, 1996; Lepage et al, 1996; St-Laurent et al, 1997), which are the elicitors of antibody responses in EAV-infected horses (De Vries et al, 1992; Chirnside et al, 1995; Kheyar et al, 1997). This result is comparable to other studies (Lepage et al, 1996; St-Laurent et al, 1997) which have analysed nearly the same EAV isolates. In these prior studies, the Vienna isolate could also be subjected to selective pressure by the host immune system. However, this does not preclude the possibility that other EAV gene products (including those of ORFs 3 and 4) may also contain neutralizing epitopes, able to induce neutralizing antibodies. Therefore genes other than ORF 5 could also be subjected to pressure by the host immune response. Further studies are needed to clarify this point.

The significance of the genetic variation observed in EAV ORF 3 has yet to be determined. Using the PSORT program, ORF 3-encoded EAV protein is predicted to be a type IIIa protein (whereas the ORF 4-encoded product is predicted to be a membrane-associated 1a-like protein). Genetic variation has also been reported with non-structural proteins of other RNA viruses, the conserved regions of which are believed to be important in the regulation of virus replication (Durga Rao et al, 1995; Oberste et al, 1996). Here, in spite of the divergence of the deduced amino acid sequence in two distinct regions of the EAV ORF 3 protein, 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 protein activity and/or in EAV replication.
Further studies are needed to resolve the molecular basis of EAV ORF 3 and 4 genetic variation and to determine the biological function and intracellular expression of ORFs 3 and 4 during EAV cell infection. The recent determination of the extreme 5' end of the EAV genome (Kheyar et al., 1996), and, thereafter, the development of an EAV cDNA infectious clone (Van Dinten et al., 1997) should provide powerful molecular means to investigate the precise role of ORF 3- and 4-encoded proteins 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 are grateful to C Villeneuve for secretarial work. We also thank P Timoney and W McCollum (Gluck Equine Research Center, College of Agriculture, University of Kentucky, Lexington, KY, USA) for providing the American and European EAV isolates, and the Arvac vaccine strain.

REFERENCES

Balasuriya UBR, MacLachlan NJ, De Vries AAF, Rossito PV, Rottier PJM (1995a) Identification of a neutralization site in the major envelope glycoprotein (G1) of equine arteritis virus. Virology 207, 518-527


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


Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74, 5463-5467


