Genetic variation in porcine reproductive and respiratory syndrome virus isolates in the midwestern United States

V. Kapur, M. R. Elam, T. M. Pawlovich and M. P. Murtaugh*

Department of Veterinary Pathobiology, University of Minnesota, 1971 Commonwealth Avenue, St Paul, MN 55108, USA

The nucleotide sequence of a 3266 bp region encompassing open reading frames (ORFs) 2 through 7 of the porcine reproductive and respiratory syndrome virus (PRRSV) was determined for 10 isolates recovered from the midwestern United States. Pairwise comparisons showed that genetic distances between isolates ranged from 2.5% to 7.9% (mean 5.8% ± 0.2%) whereas the Lelystad strain from Europe was, on average, 34.8% divergent from US clones. Thus, US and European PRRSV isolates represent genetically distinct clusters of the same virus. ORF 5, which encodes the envelope glycoprotein, was the most polymorphic [total nucleotide diversity (π) = 0.097 ± 0.007] and ORF 6, encoding the viral M protein, was the most conserved (π = 0.038 ± 0.003). The substantial differences in nucleotide diversity among ORFs suggests that the virus is evolving by processes other than simple accumulation of random neutral mutations. In support of this hypothesis, statistical analyses of the nucleotide sequence provided strong evidence for intragenic recombination or gene conversion in ORFs 2, 3, 4, 5 and 7, but not in ORF 6. An excess of synonymous (silent) substitutions was observed in all six ORFs, indicating an evolutionary pressure to conserve amino acid sequences. Taken together, the data indicate that despite intragenic recombination among extant PRRSV isolates, purifying selection has acted to maintain the primary structure of individual ORFs.

Introduction

A new viral disease of pigs was detected in North America in 1987 (Hill, 1990) and in Europe in 1990 (Paton et al., 1991). The disease, now known as porcine reproductive and respiratory syndrome (PRRS), is characterized by late-term abortions and stillbirths in sows and respiratory difficulties in nursery pigs (Pol et al., 1991; Wensvoort et al., 1991; Collins et al., 1992). The virus is recovered primarily from alveolar macrophages of infected swine and is a small, enveloped positive-stranded RNA virus (Benfield et al., 1992; Wensvoort et al., 1992a). It is provisionally classified in the arterivirus group based on morphology, genome organization, transcriptional regulation and macrophage specificity (Plagemann & Moennig, 1992). The Arteriviridae include the PRRS virus, equine arteritis virus (EAV; den Boon et al., 1991), lactate dehydrogenase-elevating virus (LDV; Plagemann & Moennig, 1992) and simian haemorrhagic fever virus (SHFV; Godeny et al., 1993).

The PRRS virus and other arteriviruses share the genome organization and expression strategy of the coronaviruses (Spaan et al., 1988; den Boon et al., 1991; Plagemann & Moennig, 1992; Meulenberg et al., 1993). Seven proteins are expressed from a nested set of RNA transcripts with overlapping 3′ ends. The first two open reading frames (ORFs 1a and 1b) encode the viral RNA polymerase. ORFs 2, 3 and 4 encode proteins of unknown function, ORF 5 encodes an N-glycosylated envelope protein, ORF 6 encodes a non-glycosylated integral membrane protein and ORF 7 encodes the nucleocapsid protein (Meulenberg et al., 1995). ORF 2 may encode an envelope glycoprotein based on homology to EAV G, (de Vries et al., 1992).

The American and European isolates, VR2332 and Lelystad virus (LV), respectively, show a high level of molecular and antigenic variation. In a comparison of 24 field sera and seven viral isolates from Europe and North America, Wensvoort et al. (1992b) found that the...
European and American isolates were immunologically distinct and had no common antigens. The complete 15 kb nucleotide sequence of a European isolate (Lebstad strain) of the PRRS virus was determined by Meulenberg et al. (1993), and the 3' hydroxy-terminal region was recently characterized from a US isolate (VR-2332) by Murtaugh et al. (1995). These studies showed that while European and US isolates of the PRRS virus evolved from a common ancestor, there was substantial variation at the nucleotide and amino acid levels. For instance, there was only 59% identity in the amino acid sequence encoded by ORF 5 between the US and European isolates (Murtaugh et al., 1995).

In order to determine the amount of genetic variation among isolates recovered from infected swine in the US and to understand the molecular mechanisms that govern the evolution of the PRRS virus, nucleotide sequences from ORFs 2 through 7 of ten US isolates were characterized. The results reveal substantial nucleotide diversity among the individual isolates and suggest that the virus is evolving by processes other than the simple accumulation of random neutral mutations. Moreover, the data indicate that, while intragenic recombination has played an important role in the genetic diversification of PRRS virus isolates, evolutionary pressure has acted to maintain the primary amino acid structure of individual ORFs.

**Methods**

*Virus and cells.* The ATCC VR-2332 strain of the PRRSV was a fourth cell culture passage grown on CL2621 cells (Collins et al., 1992). Other virus strains were midwestern US field isolates cloned at the University of Minnesota Veterinary Diagnostic Laboratory by limiting dilution. All isolates were obtained between 1989 and 1992 from lung tissue cultured on CL2621 cells. Four of the ten isolates did not grow on porcine alveolar macrophages. A description of the isolates is available on request from the authors. Viral stocks were grown on CL2621 cells as described (Bautista et al., 1993). Cells were cultured in MEM supplemented with 4–10% fetal calf serum in a humidified 5% CO₂ atmosphere at 37°C.

*Isolation of RNA, RT-PCR and DNA sequencing.* Total RNA from infected cell supernatants was isolated by acid guanidine phenol extraction as described by Chomczynski & Sacchi (1987). Reverse transcription was performed using random primers as directed by the manufacturer (Perkin-Elmer Cetus). For PCR reactions, primers were synthesized based on VR-2332 sequence and designed to amplify regions encompassing each ORF. The primer sequences are available on request from the authors. Nucleic acids were denatured at 94°C, followed by 30 cycles at 55°C for 30 s, 72°C for 45 s and 93°C for 45 s. Amplified fragments were polished at 72°C for 10 min and purified using SpinBind columns (FMC). Automated DNA sequencing reactions were performed according to the directions provided with the Taq DyeDeoxy Terminator Cycle Sequencing Kit (Applied Biosystems) using a PE 480 Thermocycler (Perkin Elmer). The nucleic acid sequences encoding ORFs 2 through 7 of the isolates are deposited in GenBank with the following accession numbers L39361 (IA-1), L39362 (IA-6), L39368 (NE-1), L39367 (MO-1), L39365 (KY-1), L39364 (KS-1), L39363 (IL-1), L39366 (MN-1), L39369 (SG-1) and U00153 (VR-2332).

**Analysis of nucleotide sequences.** Sequence data were assembled and analyzed with the computer programs SEQMAN and MEGALIGN (DNASTAR). Phylogenetic analysis of the compiled sequences was performed with distance-matrix, maximum-likelihood and parsimony methods, with PHYLIP 3.4 (J. Felsenstein, Department of Genetics, University of Washington, Seattle, Wash., USA) the programs NJTREE and NJBOOT (T. S. Whittam, Institute of Molecular Evolutionary Genetics, Pennsylvania State University, Penn., USA) and PAUP 3.1.1 (Swofford, 1991). The robustness of the phylogenetic analysis and significance of the branch order were determined by bootstrap analysis (Felsenstein, 1985) with 2000 replications. Statistical analysis for clustering of polymorphic sites was performed by Sawyer's method (Sawyer, 1989) and individual ORFs were analysed for pairwise comparisons of the number of synonymous substitutions per synonymous site (dₛ) and coding substitutions per non-synonymous site (dₓ) by the method of Nei & Gojobori (1986).

**Results and Discussion**

**Properties of PRRS virus gene sequences**

Complete nucleotide sequences (range 3256–3260 bp) of ORFs 2 through 7 were determined for the 10 PRRSV isolates. Alignment of the full-length nucleotide sequences from the 10 isolates by the program CLUSTAL V resulted in a consensus sequence of 3266 bp with a total of 611 polymorphic nucleotide sites. The total nucleotide diversity (π) of individual ORFs ranged from 0.038 ± 0.003 for ORF 6 to 0.097 ± 0.007 for ORF 5, with a mean nucleotide diversity for all six ORFs of 0.060 ± 0.008 (Table 1). These substantial differences in the levels of total nucleotide diversity among the individual ORFs suggest that the virus is evolving by mechanisms other than the simple accumulation of random neutral mutations.

**Patterns of nucleotide substitution**

The ratio of the number of nucleotide substitutions that result in amino acid replacement (non-synonymous substitutions) to those that do not (synonymous substitutions) was determined for each ORF to evaluate the relative evolutionary pressure on viral genes either to conserve or change amino acid sequence. Since it is difficult to distinguish between synonymous and non-synonymous substitutions by visual examination alone, the numbers of synonymous changes per synonymous site (dₛ) and coding changes per non-synonymous site (dₓ) were calculated by the unweighted pathway method of Nei & Gojobori (1986). The results indicate a three- to five-fold excess of synonymous over non-synonymous substitutions in the six ORFs (Table 1). For instance, in ORF 5, the dₛ values were 0.213 ± 0.022 while the dₓ values were 0.063 ± 0.006, a dₛ/dₓ ratio of 3.38:1. Similarly, for ORF 6, the dₛ/dₓ ratio was 5.6:1. These
Table 1. Properties of nucleotide sequences of 6 open reading frames characterized from 10 PRRSV isolates.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Sequence length</th>
<th>Number of polymorphic sites</th>
<th>$d_s^*$</th>
<th>$d_v^+$</th>
<th>Total diversity‡</th>
<th>SSCF§ (P)</th>
<th>SSUF∥ (P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORF 2</td>
<td>771</td>
<td>134</td>
<td>0.119 ± 0.014</td>
<td>0.030 ± 0.004</td>
<td>0.050 ± 0.004</td>
<td>0.0110</td>
<td>0.0750</td>
</tr>
<tr>
<td>ORF 3</td>
<td>765</td>
<td>145</td>
<td>0.143 ± 0.016</td>
<td>0.042 ± 0.004</td>
<td>0.065 ± 0.005</td>
<td>0.0100</td>
<td>0.0070</td>
</tr>
<tr>
<td>ORF 4</td>
<td>537</td>
<td>98</td>
<td>0.141 ± 0.020</td>
<td>0.034 ± 0.005</td>
<td>0.058 ± 0.006</td>
<td>0.0100</td>
<td>0.0460</td>
</tr>
<tr>
<td>ORF 5</td>
<td>603</td>
<td>157</td>
<td>0.213 ± 0.022</td>
<td>0.063 ± 0.006</td>
<td>0.097 ± 0.007</td>
<td>0.0000</td>
<td>0.0050</td>
</tr>
<tr>
<td>ORF 6</td>
<td>525</td>
<td>60</td>
<td>0.101 ± 0.015</td>
<td>0.018 ± 0.004</td>
<td>0.038 ± 0.004</td>
<td>0.4080</td>
<td>0.1520</td>
</tr>
<tr>
<td>ORF 7</td>
<td>372</td>
<td>64</td>
<td>0.135 ± 0.023</td>
<td>0.027 ± 0.005</td>
<td>0.059 ± 0.006</td>
<td>0.0000</td>
<td>0.3000</td>
</tr>
</tbody>
</table>

* $d_s$, Number of synonymous substitutions per synonymous site (Nei & Gojobori, 1986).
† $d_v$, Number of non-synonymous substitutions per non-synonymous site (Nei & Gojobori, 1986).
‡ Total nucleotide diversity ($\pi$) (Nei & Gojobori, 1986).
§ SSCF, sum of squares of condensed fragment lengths (Sawyer, 1989).
∥ SSUF, sum of squares of uncondensed fragment lengths (Sawyer, 1989).

Table 2. Pairwise distances among midwestern US isolates of the PRRS virus for the 3266 bp region encoding ORF 2 through ORF 7

Mean and absolute distances are presented above and below the diagonal, respectively. Mean distance is expressed as the percent polymorphic sites. Absolute distance is the total number of polymorphic sites.

<table>
<thead>
<tr>
<th></th>
<th>IA1</th>
<th>IA6</th>
<th>IL1</th>
<th>KS1</th>
<th>KY1</th>
<th>MN1</th>
<th>MO1</th>
<th>NE1</th>
<th>SG1</th>
<th>VR2332</th>
</tr>
</thead>
<tbody>
<tr>
<td>IA1</td>
<td>-</td>
<td>0.045</td>
<td>0.055</td>
<td>0.053</td>
<td>0.071</td>
<td>0.056</td>
<td>0.073</td>
<td>0.059</td>
<td>0.058</td>
<td>0.047</td>
</tr>
<tr>
<td>IA6</td>
<td>1.46</td>
<td>-</td>
<td>0.048</td>
<td>0.059</td>
<td>0.072</td>
<td>0.060</td>
<td>0.077</td>
<td>0.055</td>
<td>0.063</td>
<td>0.045</td>
</tr>
<tr>
<td>IL1</td>
<td>180</td>
<td>157</td>
<td>-</td>
<td>0.052</td>
<td>0.074</td>
<td>0.055</td>
<td>0.079</td>
<td>0.059</td>
<td>0.065</td>
<td>0.049</td>
</tr>
<tr>
<td>KS1</td>
<td>172</td>
<td>193</td>
<td>168</td>
<td>-</td>
<td>0.072</td>
<td>0.033</td>
<td>0.074</td>
<td>0.050</td>
<td>0.064</td>
<td>0.037</td>
</tr>
<tr>
<td>KY1</td>
<td>232</td>
<td>235</td>
<td>242</td>
<td>236</td>
<td>-</td>
<td>0.067</td>
<td>0.047</td>
<td>0.066</td>
<td>0.064</td>
<td>0.054</td>
</tr>
<tr>
<td>MN1</td>
<td>181</td>
<td>194</td>
<td>180</td>
<td>108</td>
<td>218</td>
<td>-</td>
<td>0.073</td>
<td>0.047</td>
<td>0.065</td>
<td>0.032</td>
</tr>
<tr>
<td>MO1</td>
<td>239</td>
<td>252</td>
<td>256</td>
<td>241</td>
<td>153</td>
<td>238</td>
<td>-</td>
<td>0.073</td>
<td>0.065</td>
<td>0.061</td>
</tr>
<tr>
<td>NE1</td>
<td>192</td>
<td>180</td>
<td>193</td>
<td>162</td>
<td>216</td>
<td>152</td>
<td>237</td>
<td>-</td>
<td>0.058</td>
<td>0.025</td>
</tr>
<tr>
<td>SG1</td>
<td>188</td>
<td>204</td>
<td>211</td>
<td>208</td>
<td>208</td>
<td>211</td>
<td>212</td>
<td>190</td>
<td>-</td>
<td>0.030</td>
</tr>
<tr>
<td>VR2332</td>
<td>154</td>
<td>145</td>
<td>159</td>
<td>122</td>
<td>175</td>
<td>103</td>
<td>200</td>
<td>83</td>
<td>162</td>
<td>-</td>
</tr>
</tbody>
</table>

Data suggest an evolutionary pressure to conserve the primary structure of these viral proteins.

Phylogenetic relationships among PRRSV isolates

Pairwise comparisons of the 10 aligned nucleotide sequences revealed mean nucleotide divergence ranging from a low of 2.5% for the isolates NE1 and VR2332 to a high of 7.9% for the isolates IL1 and MO1 (Table 2). An unrooted bootstrap phylogenetic tree constructed with the program PAUP (Swofford, 1991) indicates that the ten US isolates grouped together into three clusters, consisting of isolates IA1, IA6 and IL1; KS1, MN1, NE1 and VR2332; and SG1, KY1 and MO1 (Fig. 1a). Phylogenetic trees constructed with the neighbour-joining and maximum-likelihood methods gave similar results. When the nucleotide sequence from the corresponding region of Lelystad virus (Muelenberg et al., 1993) was included in the analysis, it always clustered as an outgroup with a branch length of, on average, more than an order of magnitude longer than the distance between any two of the US virus isolates.

Intragenic recombination in PRRS virus genes

Prompted by the finding of heterogeneity in nucleotide substitution rates among the PRRSV gene sequences, we examined the pattern of distribution of polymorphic sites in individual ORFs for evidence of intragenic recombination or gene conversion. An analysis of the 611 polymorphic nucleotide sites in the aligned sequences by the method of Sawyer (1989) revealed strong statistical evidence for intragenic recombination. The sum of squares of condensed fragments (SSCF) and uncondensed fragments (SSUF) and the maximum length of condensed fragments (MCF) and uncondensed fragments (MUF) were calculated for the observed sequences and compared with those from an artificial data set obtained from 10000 random permutations (Sawyer, 1989). In a pairwise comparison of the 10 viral strains, the observed scores for SSCF (144987), MCF (68), SSUF (6689164) and MUF (411) were all highly significant ($P \sim 0.005$). In addition, a plot of the distribution of observed gap lengths and the probability of occurrence of the same gap length in 10000 randomly permuted sequences revealed the presence of several.
fragments between 9 and 68 nucleotides in length that may represent intragenic recombination events.

In order to identify specific genes of the PRRS virus genome in which intragenic recombinational events may have occurred, the six individual ORFs were analysed by Sawyer's method. The data provide statistical evidence for intragenic recombination in ORFs 2, 3, 4, 5 and 7, but not ORF 6 (Table 2). The apparent lack of recombination in ORF 6 may be due to selection against recombinant forms, as described in coronaviruses (Ban-
ner & Lai, 1991). Possible intragenic recombinational events in ORF 3 from isolates MN1, IL1 and KS1 were visualized in a haplotype plot (Fig. 1b). These results indicate that individual regions of the PRRS virus genome may have different evolutionary histories, or that the PRRSV genome may represent a mosaic of RNA fragments from several distinct genetic back-
grounds.

How old is the PRRS virus?

Although sequence similarities between North American and European isolates of PRRSV show they arose from a common ancestor, it is apparent that these groups of
isolates diverged before the recent appearance of disease in the last decade (Hill, 1990; Paton et al., 1991). However, within the midwestern US it is of interest to determine whether the outbreak of disease is consistent with the appearance of a single viral type which then radiated, or whether numerous genetically distinct types were present, one or more of which were responsible for the new disease. An estimate of the time of divergence can be obtained from the rate of synonymous nucleotide substitutions in genes and genomes (Nei, 1987). For the influenza virus, the rates of synonymous substitution sites per year of the haemagglutinin, neuraminidase and non-structural protein genes were estimated to be 0.014, 0.011 and 0.009, respectively (Hayashida et al., 1985). The rate of accumulation of synonymous to non-synonymous substitutions in ORF 6 (5:6:1) is about the same as in mammalian and influenza virus genes (Li et al., 1985, Hayashida et al., 1985). Moreover, ORF 6 does not show evidence of recombination. If the rate of synonymous substitutions in ORF 6 is similar to that in structural and non-structural protein genes of influenza virus, then the total amount of nucleotide variation in midwestern US PRRSV isolates can be accounted for by 6 to 14 years of virus evolution (mean ~ 9 years). These results are consistent with the epidemiological observations that suggest that the virus emerged as a swine pathogen approximately a decade ago.

In conclusion, our results indicate that PRRS virus isolates recovered from swine in the midwestern US are evolving by both the accumulation of random neutral mutations and by intragenic recombination. Further, the data indicate that there is substantial nucleotide variation between US and European PRRS virus isolates.

We thank Elida Bautista, Sagar Goyal and Thomas Molitor for providing cloned virus stocks. Financial support for this project was provided by Boehringer Ingelheim Animal Health and the Minnesota Agricultural Experiment Station.

References


(Received 18 September 1995; Accepted 5 January 1996)