In the late 1980s and early 1990s, swine farmers in North America and Europe noticed the emergence of a new disease, characterized by abortions, stillbirths, delivery of mummified foetuses and birth of weak piglets (Keffaber, 1989). After the first waves of abortion, the affected farms suffered severe respiratory problems in nursery and fattening units. By its unique nature, and since its cause was unknown at that time, the disease was called 'mystery swine disease'. Early research showed this syndrome to be caused by an RNA virus that resembled equine arteritis virus (EAV) and lactate dehydrogenase-elevating virus (LDV) (Wensvoort et al., 1991). Thus, the 'mystery disease' agent was classified as a member of the genus Arterivirus and named porcine reproductive and respiratory virus (PRRSV).

Further studies demonstrated that the PRRSV genome is about 15 000 bases in length and comprises eight ORFs. ORFs 1a and 1b encode the viral polymerase and account for most of the genome, ORFs 2–4 encode virion-associated proteins, ORF5 encodes the envelope glycoprotein protein GP5, ORF6 encodes the membrane protein and ORF7 encodes the nucleocapsid protein (Dea et al., 2000; Meulenberg et al., 1997; Snijder & Meulenberg, 1998). It is thought that GP5 could be one of the leading targets for vaccine design, since antibodies directed against it can neutralize virus infectivity (Pirzadeh & Dea, 1997, 1998; Gonin et al., 1999). In addition, since this protein is exposed in the outer part of the virion, it is submitted to a selective pressure exerted by antibodies of infected or vaccinated animals. As a result, GP5 has been shown to be polymorphic (Kapur et al., 1996).

Two viral genotypes have been described for PRRSV: American and European (Mardassi et al., 1994; Meng et al., 1995a). The reference strain for the American genotype is ATCC VR-2332, while Lelystad virus (LV) is the reference strain for European genotypes. Globally, genetic similarity between these two types is about 65 % and some diversity within a given genotype can exist. Thus, some reports have shown different clusters of European strains that can coexist in the same herd (Indik et al., 2000). These facts are of major relevance for the design of vaccines, as these variations could potentially affect their efficacy under field conditions.

In this paper, we present the results of a phylogenetic analysis with European-type PRRSV strains detected in farms of Spain.

A total of 21 PRRSV strains was selected randomly from PRRSV-positive serum samples submitted between 2000 and 2001 directly to our laboratory, or referred by another laboratory (Laboratorios Hipra, Amer, Spain), for diagnosis by PCR. Two modified live vaccines (v1 and v2) currently used in Spain and made of European strains were also included. Total RNA was extracted using Nucleospin RNA (Macherey–Nagel), according to the manufacturer’s
recommendations. This procedure yielded 50 μl total RNA from 150 μl serum. Reverse transcription was done using 1× RT buffer, 300 ng random hexamers, 0·5 μM dNTPs, 5 mM DTT, 1 μl RNase inhibitor, 100 U RT Superscript II (Invitrogen) and 2 μl total RNA in a total volume of 20 μl. For the detection of PRRSV, a PCR aimed specifically to amplify the whole ORF5 sequence was designed. Primers L1F (forward, 5′-TGAGGGGCTACACACCTT-3′) and L1R (reverse, 5′-AGGCTAGCGACGCTTGGT-3′), which are based on the cloned sequence of LV (accession no. M92962), were used. The expected PCR product was calculated to be 702 bp and included the complete ORF5 and flanking regions of ORF4 and ORF6. PCR was done in a 50 μl volume using 1× PCR buffer, 1·5 mM MgCl2, 0·2 mM each dATP, dTTP, dCTP and dGTP, 25 pmol of each primer, 2 U Taq DNA polymerase (all products from E. coli), and 2·5 μl cDNA. The detection limit of the PCR, carried out for 35 cycles at 94 °C for 45 s, 55 °C for 45 s and 72 °C for 45 s, was about 50 virus particles ml⁻¹ serum of the European-type strains.

Selected PCR products were purified and then sequenced using an ABI Prism sequencer, ABI 373 A Stretch (Applied Biosystems). To simplify the identification of data, each sequence was coded with a three letter name (uab) followed by a serial number (1–21). These sequences were deposited in GenBank (accession nos AF495499–AF495521), together with ORF5 sequences of several European strains deposited by other researchers: LV (accession no. M92962); the first sequenced Spanish strain, Olot-91 (accession no. M92962); the first Spanish strain, Olot-91 (accession no. X92942); 15 other Spanish strains isolated before 1996 (accession nos PRU40687–PRU40702); seven Czech strains (accession nos AF253531–AF253537).

Sequence alignment was done using the CLUSTAL W software (Thompson et al., 1994). Variability for a given position in the sequence was calculated by means of an entropy plot using the BIOEDIT software (Hall, 1999). Imported alignments were analysed further using the neighbour-joining method of Saitou & Nei (1987), with correction for multiple substitutions using 1000 bootstrap trials. Evolutionary parsimony tree and evolutionary distance matrices were calculated according to Felsenstein (1993).

Predicted amino acid sequences were analysed in a similar manner to nucleotide sequences to obtain an unrooted parsimony tree and an evolutionary matrix. In addition, transmembrane segments were evaluated using TMAP (Persson & Argos, 1994) and hydrophobicity profiles were calculated according to Kyte & Doolittle (1982). Potential signal peptides were predicted by means of the IPSIGNAL program, version 2.0 (Nielsen et al., 1997). Potential sites of N-linked glycosylation were determined using the EXPASY proteomics tools available at http://us.expasy.org/tools.

For our strains, nucleotide similarities in ORF5 compared to the reference strain, LV, ranged from 83·1 (uab12) to 93·4 % (uab4), with only 7 of 21 strains showing similarities above 90 % (uab2, uab3, uab4, uab5, uab7 and uab21). In contrast, ORF5 sequences of European strains retrieved from GenBank were similar to LV, with similarities ranging from 99·8 to 87·6 %, with only two strains (Czech isolates AF253531 and AF253537) below 90 %. Two Spanish strains, uab15 and uab16, were closely related ( > 99 % similarity) to variant Czech strains AF253531 and AF253537, respectively. v1 shared 93·6 % similarity with LV and v2 shared 88·5 % similarity with LV.

The evolutionary tree calculated by the maximum-parsimony method showed three main branches deriving from LV. The first one would have given rise to the Czech strains AF253532–AF253536, the second one would have originated three groups of Spanish strains isolated before 1996 (PRU40687–PRU40702) and the third one grouped the variant Czech strains AF253531 and AF253537 described in 2000, together with all of our field strains and Spanish vaccines as well as the first Spanish isolate, Olot-91 (accession no. X92942). None of the isolates seemed to be derived from the vaccines, even in farms where vaccination is currently used (Fig. 1).

For most of the strains, the predicted proteins of ORF5 were 201 aa long; however, isolates uab2, uab5, uab11 and uab20 yielded proteins truncated at the C-terminus. Comparison of the predicted protein sequences indicated that similarity with LV ranged from 99·5 to 31 %. Major differences were found in three recent Spanish strains and one vaccine (uab2, uab6, uab8 and v1) which were below 80 % similarity and two others (uab12 and uab17) which were below 40 % (data not shown).

Analysis of predicted proteins showed that cleavage of the signal peptide was most likely to occur between residues 33 and 34 in most sequences, but not in LV, where this site was calculated to be between residues 32 and 33. A strongly conserved region between residues 37 and 55 was detected for which the consensus sequence (80 %) corresponded to Asn–Ser–Ser–Thr–Tyr–Gln–Tyr–Ile–Tyr–Asn–Leu–Thr–Ile–Cys–Glu–Leu–Asn–Gly–Thr. However, in LV and X92942 (Olot-91), the first Asn residue was not present. Thus, most of our strains and Czech variant strains, but not LV, had three potential N-linked glycosylation sites (Asn–X–Ser/Thr). In contrast, some of them (uab2, uab11 and uab13) only had one. Transmembrane segments were predicted to be between aa 10 and 31, 60 and 82, and 104 and 132. Regarding mean hydrophobicity profiles, it was seen that some strains, particularly uab11, uab12 and v2, had GP5 segments with markedly lower levels of hydrophobicity than LV (Fig. 2).
Evolutionary analysis of the ORF5 protein sequences rooted on LV showed two main branches for the GP5 of European-type PRRSV strains. The first one was composed of five of the Czech isolates (AF253532–AF253536) and the second one, which comprised all Spanish strains and Czech variants, divided in two evolutionary ways, one with Spanish isolates retrieved from GenBank (PRU40687–PRU40702) and the other with our field strains, vaccines and the first Spanish PRRSV isolate (X92942) (Fig. 3).

Genetic diversity in RNA viruses is a well-known fact. This variability is driven by several factors, including random mutation, recombination competition and selection (Domingo et al., 1996). As a result, some viral variants acquired advantageous features, such as masking of critical epitopes, an enhanced ability to establish persistent infections or a given tissue tropism (Faaberg et al., 1995; Chen et al., 1998). This variability frequently affects those genes encoding for the more exposed parts of the virion, particularly the envelope proteins (Kapur et al., 1996), probably because of the pressure exerted by the immune system through neutralization mechanisms among other factors (Meng et al., 1995b; Pirzadeh & Dea, 1997, 1998; Gonin et al., 1999).

Many reports have been published about the genetic diversity of American-type PRRSV isolates but much less is known about European-type field isolates. The similarity between sequences of these two genotypes is relatively low (about 65%). However, both types share a similar identity to LDV (47%). This fact leads to the notion that the European and American PRRSV genotypes diverged from a common ancestor (Murtaugh et al., 1995).

Among a given genotype, variability of ORF5 has been calculated to be as much as 12% for American- (Key et al., 2001; Meng, 2000) or European-type strains (Suárez et al., 1996; Indik et al., 2000), with the exception of some strains from Eastern Europe, in which divergence can be higher than 30% (Stadejek et al., 2002). Most of this variability has been found in a so-called hypervariable region, corresponding to the N-terminal segment of GP5. In our study, we found that percentages of sequence identity were between 83.1 and 93.4%. As in other studies, maximum variation was found in that hypervariable region. Interestingly, some of our sequences clustered very closely with variant Czech strains described recently (Indik et al., 2000). Some of the Spanish farms from where these Czech-like strains were detected had very serious problems of mortality in weaners and fatteners that were attributed to a combination of PRRSV and porcine circovirus type 2 infections.

The evolutionary tree constructed by the maximum-parsimony method showed that all of our field strains, the two examined vaccines and the variant Czech strains might belong to a distinct evolutionary branch with respect to the other older European-type ORF5 sequences retrieved from GenBank. Since our sequences were randomly selected from positive samples coming from regions all over Spain, this could indicate the current existence of a well-established variant of PRRSV in our country. Whether this type can in fact be prevalent is unknown to us and requires further investigation.

PRRSV GP5 forms the envelope of the virus. This protein is composed by some 200 aa, with a molecular mass of about 22–25 kDa. As known from studies with LV, GP5 has a signal sequence of 32 aa, a transmembrane region between residues 108 and 131 and present two glycosylation sites in the short ectodomain close to the N-terminal region (Meulenberg et al., 1997). The predicted protein sequences of our isolates showed considerably lower percentages of similarity to LV than the nucleotide sequences. In the case of...
the two isolates that shared less than 40% identity with LV, this was due to an insertion of a codon and a later deletion of three nucleotide residues that shifted the frame in the N-terminal segment of the protein. This change did not affect the very conserved region between aa 37 and 55. In LDV and EAV, it has been shown that the neutralization epitope is located in the ectodomain of the ORF5 glycoprotein (Li et al., 1998; Snijder & Meulenberg, 1998; Balasuriya et al., 1997; Chirnside et al., 1995). Viable neutralization escape mutants of EAV can arise by means of substantial deletions in that segment. By analogy, it is possible that insertions could also contribute to this escape of neutralization.

The analysis of the predicted amino acid composition of our sequences showed that most of them introduced a third potential glycosylation site in the short ectodomain of GP5. This site arises from a change of an aspartic acid residue at position 36 (Asp\(^{36}\)) to an asparagine (Asn\(^{36}\)). A similar fact has been observed by Rowland et al. (1999), who described the occurrence of Asn\(^{34}\) PRRSV mutants during the course of an experimental infection initiated in utero with the American VR-2332 strain. These mutants were not detected before 1 week after birth but were detected subsequently as piglets developed neutralizing antibodies. The neutralizing activity of antibodies directed to Asn\(^{34}\) mutants was lower than that of the antibodies directed to Asp\(^{34}\) viruses. In addition, these Asn mutants were always detected in the lymph nodes or tonsils of infected pigs. Rowland et al. (1999) suggested that these mutant variants might represent the result of a favourable selection of viruses with the ability to persist in lymphoid tissues. Accordingly, our field strains

**Fig. 2.** Mean hydrophobicity plot (Kyte & Doolittle, 1982) of the GP5-deduced protein sequences of LV compared with v2. Boxes show predicted transmembrane segments.

**Fig. 3.** Evolutionary tree (maximum-parsimony method) based on predicted protein sequences of PRRSV ORF5. The tree was rooted using LV (M98262) as an outgroup. The suffix ‘sp’ indicates strains of Spanish origin obtained from GenBank, whereas the suffix ‘cr’ denotes sequences from European-type strains of Czech origin deposited in GenBank. Spanish-type vaccines are marked as v1 and v2. All other sequences correspond to Spanish PRRSV strains detected by PCR in our laboratory.
may represent variants adapted to the pressure exerted by the immune system after infection or vaccination. Regarding this latter point, it is interesting to note that none of the two vaccines examined had this additional N-glycosylation site. Thus, it would be expected that they would induce good neutralizing antibody responses against similar strains.

Comparisons of the mean hydrophobic profiles of our sequences showed that some isolates, particularly the vaccines examined, were more hydrophilic than LV. This fact, together with the lack of extra glycosylation sites, could explain the immunogenicity of these vaccines, which, according to our experience when sera of vaccinated pigs were examined by means of indirect ELISA, induce the production of high antibody levels compared to inactivated or American-type vaccines currently in use in Spain.

Taken together, our results suggest that variant PRRSV strains might have developed with time, and at the present time are common in pigs in Spain. Such strains would have some features (extra glycosylation sites, changes in the ectodomain composition, etc.) that, on a theoretical basis, can give them an advantage to escape the immune system or to establish persistent infections. These data should be taken into account in the design of new vaccines.

ACKNOWLEDGEMENTS

Our grateful thanks to Laboratorios Hipra (Amer, Spain) for its collaboration in this study and for providing us with 11 samples. We would also like to thank Dr Anna Barceló of the Servei de Sequenciació de DNA at the Universitat Autònoma de Barcelona for her help in sequencing.

REFERENCES


