Classical swine fever virus diversity and evolution

Paul Lowings,* Georgina Ibata, Jennifer Needham and David Paton

Department of Virology, Central Veterinary Laboratory (Weybridge), Addlestone, Surrey KT15 3NB, UK

By analysing the nucleotide sequence data generated from both the E2 (gp55) and the NS5B genes of classical swine fever virus (CSFV), in addition to previously published data from the 5'NCR, we were able to divide 115 CSFV isolates into two major groups, five subgroups and two disparate isolates. Further discrimination was possible by analysis of sequence data from the E2 region. The three sequencing based methods were compared to monoclonal antibody (MAb) typing and to limited restriction enzyme (RE) mapping. Although both MAb and RE methods confirmed the previous classification the resolution was inferior. We estimated an approximate evolution rate for CSFV from an analysis of the virus variation observed in a single geographical area over a 6 year period. Applying this proposed rate to each of our deduced CSFV subgroups enabled us to calculate the approximate dates of divergence for each subgroup.

Introduction

Classical swine fever virus (CSFV; also known as hog cholera virus) is a member of the genus Pestivirus within the family Flaviviridae. Classical swine fever (CSF), the disease caused by this agent, is of world-wide economic importance, and is generally subject to statutory control, involving slaughter of affected pig herds and restrictions on movements of pigs and pig products from affected regions. The genome of the virus is positive polarity RNA of about 12.3 kb in length which encodes a single polyprotein. This polyprotein is both co- and post-translationally processed to yield mature viral proteins. In common with all flaviviruses, the CSFV genome encodes structural proteins, including the major glycoprotein E2 (gp55), in the 5' half of the genome, and non-structural proteins, including the putative polymerase gene (NS5B), at the 3' end. Non-coding regions (NCRs) exist at both the 5' and 3' ends of the genome. The function of the NCRs is not fully understood but they contain highly conserved sequences and may be involved in the initiation of translation and replication. Antibodies raised against the E2 glycoprotein are often virus neutralizing and are directed at epitopes located towards the N terminus of the protein (Paton et al., 1992; van Rijn et al., 1994). This region of the protein is encoded by one of the most variable parts of the pestivirus genome. Typing with monoclonal antibodies (MAbs) (Edwards & Sands, 1990) and some limited molecular data (Harding et al., 1994; Hofmann et al., 1994; Lowings et al., 1994) have suggested that CSFV, although probably the least variable member of the pestivirus genus, can be subdivided into distinct groups, the most obvious division being between old and recent isolates. The identification of virus subtypes should improve our understanding of CSFV evolution and epidemiology and might provide markers for biological differences, such as in virulence. In this study, we have assembled much of the available information regarding CSFV differentiation and have included our own data from a 190 bp region of E2 (112 viruses) and a 210 bp region of NS5B (50 viruses). We have analysed all the information using phylogenetic techniques. This has allowed us for the first time to reproducibly subgroup these viruses and to draw some conclusions as to their possible evolutionary history.

Methods

Viruses. The viruses used in this study have been isolated over a period of 49 years (1945–1994) from 14 countries. In many cases the full isolation details cannot be traced. One-hundred and twelve viruses were cultured in PK15 cells including 12 from the United States (US) (1946–1968), 10 from Malaysia (a Malaysian ALD strain and isolates from 1967–1986), four from Brazil (one vaccine strain and three isolates from 1987), 11 from the UK (1956–1986), five from Japan (up to 1974), 15 from Italy (1945–1992), nine from Austria (1990–1992), 15 from Germany (the Riems Vaccine strain and isolates from 1986–1994), six from Poland (1991–1993), two from Holland (1977), one from Russia, 13 from Belgium (1986–1994), five from France (1966–1993), one from Mexico and three of unknown origin. None of the isolates used in this study was biologically cloned. In addition, the sequence of
the E2 region was obtained from the published data for three additional viruses: Alfort (Meyers et al., 1989), Brescia (Moormann et al., 1990) and Weybridge (Yu et al., 1993). Sequence data for the NS5B region were obtained from the former two sources.

**PCR and sequencing.** The molecular cloning and sequencing of some of the Italian isolates have been described (Lowings et al., 1994). The same PCR primers were used to amplify the remaining virus gene segments which were sequenced directly without molecular cloning. A 210 bp region of the NS5B product was sequenced as previously described using two internal primers. Likewise, a 190 bp region of the E2 product was partially sequenced using two further primers TC(GA) (AT)CCA ACC AA(TC) GAG ATA GGG (corresponding to Alfort nucleotides 2467–2487) and CAC AG(TT) CC(TG) AA(TC) CC(TG) AA(GG) TCA TC (corresponding to the complement of Alfort nucleotides 2738–2716). The sequences generated with the primers were complementary. In total, 112 E2 and 50 NS5B segments were sequenced.

Viruses were selected for NS5B analysis using the E2 data and were judged to contain candidate members from the major groups described by the E2 analysis. In addition to our own data, the CSFV sequence data from Hofmann et al. (1994) were analysed. This data included partial 5'NCR sequence from 48 CSFVs. Using two primers, CTTAG TAC AGG ACA GTC TGC TGC ATG TAC (corresponding to the complement of Alfort nucleotides 157–178) and TCA ACT CCA TGT GCC ATG TAC (corresponding to the complement of Alfort nucleotides 373–353; Wirz et al., 1993), in combination with direct sequencing, we analysed a 93–94 bp region of the 5'NCR sequence from seven of our own viruses and included the sequences obtained with the Hofmann data analysis. To allow the inclusion of our sequences Hofmann’s dataset was truncated by 23–24 bp at the 3' end. However, this truncation did not affect phylogenetic discrimination between isolates.

**Restriction site data.** The restriction site data of Harding et al. (1994) were used to produce a crude restriction map of the p45/p75 region corresponding to Alfort nucleotide sequence 5067–5574 (data not shown). These data were analysed as described below. For comparison, cDNAs from Italian viruses c1W, c3D and s1D were amplified and digested as described (Harding et al., 1994), with the exception that the isoschizomer Sau3AI was substituted for the enzyme MboI. In addition, to confirm the restriction endonuclease (RE) result, the region amplified was sequenced directly using Harding’s primers. Restriction map data from the two reference strains Alfort (Meyers et al., 1989) and Brescia (Moormann et al., 1990) were also included in the analysis.

**Monoclonal antibody typing.** The MAbs and the methods used for typing the viruses have been described (Paton et al., 1995a). A panel of 10 MAbs was selected on the basis of previously demonstrated ability to discriminate between CSFV isolates.

**Computer analysis.** The PHYLIP (Felsenstein, 1989) and GCG (Devereux et al., 1984) software used for the analysis of molecular sequences and MAb data were as described (Lowings et al., 1994). In addition, the PHYLIP program RestML was used to analyse RE derived data (Harding et al., 1994). To search for conserved peptide motifs present in all pestiviruses, the nucleotide data obtained in this study were translated and aligned with the translation of all available pestivirus sequences including (for the E2 region): (a) 21 bovine viral diarrhoea virus (BVDV) isolates – NADL (Collett et al., 1988), SD1 (Deng & Brock, 1992), Singer (Yu et al., 1994), Oregon C24V (Paton et al., 1992), R2727 (Becher et al., 1994), FB-1 and FP-1 (Paton et al., 1995a), Odssos (Renard et al., 1987) and 13 Swedish field isolates (Paton et al., 1995b); (b) three border disease virus (BDV) isolates – Morehead (Lowings & Paton, 1993), L83/84 and Clover lane (Becher et al., 1994); and (c) one ‘atypical’ pestivirus – BD-78 (Sullivan et al., 1994). For the NS5B region only NADL, Odssos and SD1 data were available. The GCG program FINDPATTERNS was used to search the SwissProt database for conserved motifs.

**Estimation of virus divergence times.** The proposed rate of E2 evolution was calculated from three sources: (a) the rate of divergence of the central Italian viruses c1W and c4D (Lowings et al., 1994) which were strongly suspected of being from a closed virus population; (b) by averaging the rates obtained from all suspected terminal branch progenitor/progeny relationships produced by E2 phylogenetic analysis; and (c) by assuming that all of the current isolates originated from viruses similar to our oldest isolates (1940s).

The apparent divergence times were calculated by selecting paired viruses from each of the proposed subgroups (each with a similar isolation date) and applying the equation \( T = K/(2r) \) (Li & Graur, 1991) where \( T \) is the divergence time between the subgroups, \( K \) is the number of substitutions per site between the representative viruses and \( r \) is the rate of E2 evolution derived from the Italian figure in substitutions per site per year. The selected viruses were EVI 192 (1987), VR4167 (1986), VR1277 (1986), c1W (1985) and V744 (1986).

**Results**

**Nucleotide sequence of the E2 fragment**

A 190 bp region corresponding to the variable 5' part of the E2 gene of 112 viruses was sequenced, aligned with published data and used to construct phylogenetic trees (Fig. 1). Using this data, CSFVs could be subdivided into two major groups with a further two disparate isolates (Congenital Tremor and Kanagawa). Group 1 consisted of all of the viruses predating 1964, in addition to more recent isolates mainly from the Far East and Brazil. It also included all the vaccine strains examined (Fig. 1). The second group consisted only of viruses isolated after 1970, the vast majority of which came from Europe and were isolated after 1985.

Most of the isolates in group 1 were of US or UK origin and date from 1945 (Brescia) to 1987 (e.g. EVI 100). Two subgroups were identified: subgroup 1.1 included the reference strain Alfort 187 from France and subgroup 1.2 included the Italian reference strain Brescia. Subgroup 1.1 could be further subdivided into isolates mainly from the 1960s, e.g. Alfort 187 (France, 1966) and isolates mainly from the US and UK from the 1940s and 1950s, e.g. ALD (US, 1946). In addition to these two clusters, subgroup 1.1 also included the PS-Porco and PS-Cells Porton isolates which lay equidistant from the Alfort 187 and ALD groups. The Fort Dodge (US, 1954) and the Brazilian EVI isolates (1987) formed two distinct lineages of which the latter fell between subgroups 1.1 and 1.2. Subgroup 1.2, the Brescia group, was more divergent than subgroup 1.1 and contained Brescia, a lone Malaysian isolate (VRI 201, 1967), Baker A (US) and HCV-37 Porcivac. Even less related to strain Brescia was a group of Malaysian viruses isolated in 1986.

Group 2 consisted mostly of viruses isolated since 1985, with only three viruses being older: Zoelen and Bergen (Netherlands, 1971) and Osaka (Japan, 1971). The group could be further divided into three subgroups. Group 2.1 consisted of a cluster of Malaysian viruses,
isolated in the same year (1986) as the 1.2 Malaysian group. On the periphery of this cluster was \( \text{ss}7\text{D}_2 \), the Italian isolate of unknown origin described previously (Lowings et al., 1994). Group 2.2 consisted of the Italian central group viruses (1985–1991) described in the same paper and a series of Austrian viruses (1990–1992).
Fig. 2. For legend see facing page.
Within this subgroup but quite distinct were the Dutch Zoelen and Bergen viruses (1977). The final subgroup, 2.3, contained only one non-European isolate (Osaka, Japan) which was actually isolated during quarantine from a pig imported from the UK (Kamijo et al., 1977). Osaka, also the oldest virus in the group (1971), was present on its periphery but was closely related to a French isolate from 1986 (Bas Rhin). Several further divisions of this subgroup 2.3 could be made: e.g. the single Italian D isolate (D4878), Osterode and the remaining Italian D virus group and the Oldenberg group. With the exception of the Rostock, 4165 P4, LK and South group Italian viruses (s7D and s8D), the remaining viruses formed a final division that showed limited variation.

Bootstrapping the complete dataset produced poor confidence levels for the deduced groups. However, re-analysis with ALD as an outgroup and including only viruses post-dating 1985 produced identical groups and subgroups, but with confidence levels at or greater than 94% (not shown). Calculated distance values gave results analogous to those of bootstrapping with inter-subgroup distances much greater than intra-subgroup values when the post-1985 viruses were analysed.

Nucleotide sequence of the NS5B fragment

Comparison of the sequences of 210 nucleotide long fragments of NS5B genes from 52 viruses selected to represent each of the E2 subgroups showed similar virus relationships when analysed by phylogenetics (Fig. 2A). Both major groups were present and similar subgroups, although less pronounced, were identifiable. A major difference was seen in group 1 where Eystrup grouped closer to Brescia (1.2) than it did to the Alfort 187 group (1.1). Bootstrapping the data gave similar confidence levels to the groupings derived from the E2 region (not shown).

Nucleotide sequence of the 5'NCR

Using the CSFV data of Hofmann et al. (1994), supplemented with seven of our sequences for the candidate subgroups 1.1 (EVI 192), 1.2 (VRI 4167), 2.1 (VRI 2277), 2.2 (c4D) and 2.3 (s7D) and the two disparate viruses (Kanagawa and Congenital Tremor), we have generated a phylogenetic tree (Fig. 2B) which showed, as they did, that this region could be used to differentiate between CSFV strains. In this case, even with the limited number of viruses, all the subgroups could be resolved. However, this method produced a slightly different grouping of our subgroup 1.1 (EVI 192) and 1.2 (VRI 4167) candidate viruses.

Comparison of sequenced regions

The viruses s7D, c4D, Alfort, Brescia, Kaernten, Rostock, EVI 192, VRI 4167, VRI 2277, Kanagawa and Congenital Tremor were common to all three datasets so could be used to compare the suitability of the three regions of the genome for phylogenetic analysis. In the E2 region, the viruses varied from 3·2% (Alfort and Rostock) to 25·0% (s7D with Kanagawa). The mean variation for all 11 viruses was 15·9% (standard deviation was 4·8). Using the NS5B region, variation was from zero (Alfort with s7D) to 19·0% (Alfort with Kanagawa). The mean was 13·6% with a standard deviation of 5·3. The 5'NCR segment produced distance values of between zero (s7D and Alfort) and 11·8% (Kanagawa with c4D). The mean variation for this region was 6·5% with a standard deviation of 2·8.

RE cleavage within the p45/75 gene junction

Phylogenetic analysis was carried out using RE cleavage data presented by Harding et al. (1994) supplemented with the restriction cleavage and sequence data from viruses c1W, c4D, and s7D and the predicted cleavage of the published Alfort (Meyers et al., 1989) and Brescia (Moormann et al., 1990) viruses. Two groups were distinguishable (Fig. 2C), one of which was analogous to the group 1 proposed from sequence data. The other group was analogous to group 2 and contained the more recent isolates.

MAb reactivity

When reactivity data from 10 MAbs (WB212, WH216, WH217, WH220, WH174, WH179, WH180, WH181, WH304, WH308 and CA72) that show differentiation...
between CSFV isolates were used to generate the most parsimonious tree of virus relationships, the isolates were subdivided into five major clusters (Fig. 2D) and a series of monophyletic branches. Three of the clusters and these single species branches contained the historical isolates equivalent to group 1 of the nucleotide analysis. The remaining two clusters contained most of the viruses previously classified as group 2; however, there were some discrepancies (Fig. 2D), particularly with those isolates shown to be more divergent by E2 sequencing.

Rates of evolution and divergence times

Using the central Italian data the rate of E2 evolution (between c1W and c4D) was calculated as $2.7 \times 10^{-3}$ substitutions per base per year. A similar value was obtained when the average rate of evolution was calculated from all probable terminal branch progenitor/progeny relationships demonstrated by the E2 tree. Assuming a virus similar to ALD was the progenitor of all the current isolates, evolution rates vary from $1.9 \times 10^{-3}$ (EVI 192) to $4.8 \times 10^{-3}$ (VRI 2277) with an average of $3.3 \times 10^{-3}$ substitutions per base per year (Fig. 3). If Kanagawa and Congenital Tremor evolved from an ALD type virus, the rates of evolution would have to be $8 \times 10^{-3}$ and $9.5 \times 10^{-3}$ substitutions per base per year respectively. Similar rate values were also obtained when using Brescia as the progenitor. Suggested evolution rates and divergence times (using the Italian E2 rate) for the viruses from each subgroup are given in Figs 3 and 4 respectively. Data from seven viruses which formed divergent branches within the defined subgroups are also plotted on Fig. 3.

Conserved peptide motifs

When a consensus sequence of the amino acid residues from the E2 region from all the available pestivirus data was generated a number of conserved motifs were

---

**Fig. 3.** Proposed rates of evolution of candidate viruses from each subgroup assuming that (A) ALD (US) or (B) Brescia was similar to its progenitor. The estimated evolution rate of the central Italian viruses is shown in the bottom right of each graph whilst the rate required for the two disparate isolates and Alfort to have evolved from the progenitor virus is shown with continuous dotted lines. The proposed subgroups are labelled in bold and the required evolution rate ($\times 10^{-3}$ nucleotide substitutions per base per year) are given in parentheses. The viruses used were EVI 192 (1.1), VRI 4167 (1.2), VRI 2277 (2.1), c4D (2.2) and Rostock II (2.3). A series of divergent members of each subgroup was also plotted on both of the graphs. These isolates included (1) VRI 201, (2) Zoelen, (3) DS021, (4) D4878, (5) Bas Rhin, (6) Fort Dodge and (7) Colorado.

**Fig. 4.** Proposed evolutionary history of the isolates examined. The estimated divergence dates of the groups and subgroups are given on the right-hand side. The viruses used were EVI 192 (1.1), VRI 4167 (1.2), VRI 2277 (2.1), c1W (2.2) and V744 (2.3).
Discussion

Nucleotide sequence comparisons of two different regions of CSFV genomes or analysis of the binding patterns of MAbs to various isolates allowed us to define two major subgroups of CSFV. The analysis of RE data provided by Harding et al. (1994), supplemented with that from additional viruses, also reveals the same two subgroups. These results are consistent with earlier investigations using MAbs (Edwards & Sands, 1990) and sequence analysis (Lowings et al., 1994) and with the 5'NCR sequence comparisons of Hofmann et al. (1994). Comparing the nucleic acid datasets, the restriction map information and the MAb typing revealed that the sequencing methods produced greatest discrimination. Of the three regions analysed by sequencing, the E2 region produced the best discrimination and was able to distinguish between many of the viruses that were identical by analysis of the other two regions. The NS5B region produced superior resolution when compared with the 5'NCR. Considering the limitations of each method a high degree of concordance was obtained. MAAb discrimination (Fig. 2D), although broadly consistent with the other methods, produced several groupings that were not seen with the other techniques.

The RE data of Harding et al. (1994) were analysed by generating a crude restriction map of the p45/p75 region with three enzymes [NcoI, MboI (Sau3AI) and AvaII]. Additional viruses representative of the subgroups defined by sequence data included Alfort and s7D (subgroup 2.3), Brescia (subgroup 1.2) and two examples of subgroup 2.2 viruses (c4D and c1W). The analysis (Fig. 2C) successfully resolved the two main groups and showed some ability to separate subgroup 2.2 from 2.3 and 1.2 from subgroup 1.1. No known examples of subgroup 2.1 were included.

All of the 112 viruses that were investigated could be amplified using the described primers in the E2 region and of these the 50 viruses investigated further could be amplified with the NS5B primers. As an epidemiological tool, the sequencing and phylogenetic analysis of the E2 and NS5B regions has already proved useful (Lowings et al., 1994). To compare the discriminatory abilities of both these regions with that of the 5'NCR we examined a subset of the 11 viruses common to all three datasets. The greatest variation between isolates was seen in the E2 set where no isolates were identical and the mean variation was the greatest (see results). This discriminatory power was confirmed by comparing the E2 and NS5B phylogenetic trees in the group 2.3 region. Of the nine viruses that were identical in the NS5B region, six groups were resolved in the E2 region. Likewise, Glentorf, Baker A and Alfort 187 were clearly discriminated using the E2 fragment but remained identical using 5'NCR sequence. Thus, as one would have predicted on the basis of known sequence variation along the pestivirus genome, the E2 region provides the greatest discriminatory ability of the regions studied. Nevertheless, there was only one major contradiction in the subgrouping predicted by any of the three methods (seen using the 5'NCR). When the complete E2 dataset was bootstrapped the consensus tree was similar to the single run data; however, values obtained for many of the nodes were low. This was due to near equal similarities between various of the older isolates and two or more current groups. We re-analysed the data rooting the tree at ALD (US), and with this exception used only viruses isolated after 1985. The bootstrapping values for each proposed group or subgroup were then equal to or greater than 94%. All of these values are probably significant as the normal confidence level of 95% is thought to be conservative when applied to bootstrap analysis (Brown, 1994). Bootstrap values of nodes describing the relationship between the subgroups within the two groups remained low in some cases due to the similar distances of these subgroups to the root (ALD); however, the Brazilian subgroup 1.1 and the Malaysian subgroup 1.2 remained the closest to ALD in 95% of trees and a confidence level of 94% was obtained for the position of the group 2.1 and 2.2 viruses (which separate the group 1 viruses from subgroup 2.3). The relationship between individual isolates within the subgroups was, however, sometimes uncertain. Analysis of both NS5B and the 5'NCR showed that all the proposed subgroups could be resolved (although in the latter case there was some difficulty in resolving subgroup 1.1 from 1.2). By including of some of our 5'NCR sequences with Hofmann's data, we were able to show that subgroup 2.1 was analogous to the 5'NCR Hannover 633/90 group. This has important implications for European CSFV epidemiology as it suggests that all three subgroups of group 2 are currently present in Europe.

The analysis of any group of viruses can be distorted by sampling bias. We are aware that even in the E2
region, where 115 viruses have been analysed, not all virus lineages may be represented. Our earliest data are from isolates from the UK and US whilst our later data are biased towards continental Europe. The limited data that we have from South American and Far East viruses suggests that different subgroups may be present in these regions. Factors other than sample bias are likely to distort the tree morphology; for instance: (1) vaccination escape could account for the appearance of recent isolates with sequences similar to older vaccine strains. This could explain the presence of the Brazilian viruses (subgroup 1.1) and one of the Malaysian subgroups (1.2) in the mid-1980s. However, it is worth noting that none of the recent European isolates seem to be derived directly from these old vaccines. (2) Likewise, the re-introduction of virus from infected frozen materials could also give the impression of zero evolution of the released strain. (3) Differing rates of evolution of certain lineages may occur but are difficult to establish with our existing data. The two isolates Congenital Tremor and Kanagawa could be candidates for rapid evolution (but could also be artefacts of sample bias). (4) Virus isolate cross-contamination may have occurred during the long passage histories of some of these viruses (see Alfort and Alfort 187 below). It is not clear to what extent sample bias and these other factors have distorted the phylogenetic tree. Another factor which could confuse the interpretation of data without distorting tree structure is animal trade. This can account for apparent jumps in virus group type present within a country. For instance Osaka, which was known to have been imported from the UK, could have introduced subgroup 2.3 virus into Japan. Other relationships such as Glentorf and 331 or Liphook and Jen Sal could also be explained by trade. In the case of the Malaysian isolates, viruses from the two major groups were circulating simultaneously (although were separated by local geography which again suggests a trade introduction or, as mentioned above, a vaccine escape.

When phylogenetic analysis studies are performed it is often said that regions of hyper-variability or areas under strong positive or negative selective pressure should be avoided as (a) if variation is too great evolutionary change can be destroyed by reversion mutations, or (b) selective pressures can appear to speed up or slow down evolution depending on the nature of this variable external force. We have found that analysis of one the most variable regions of the CSFV genome is suitable for determining the relationships between viruses. In addition to increasing our confidence in the discriminatory ability of this technique, the large number of isolates now sequenced provides an indication of the variability of this group which will hopefully be useful in future epidemiological studies. Although all of the discriminatory methods evaluated could discriminate between group 1 and 2 viruses, this alone may be of limited epidemiological value, particularly in Europe where most current isolates fall within group 2. The additional discrimination possible with E2 sequencing is therefore likely to be of great value.

It is clear from the data that in some situations CSFV can be remarkably stable: for instance, Italian isolates c1W (1985) and c3D (1991) were identical over the 190 bp analysed in the E2 region. It seems unlikely that this could be explained as a cross-contamination of virus or RNA stocks as differences are observed between these viruses within the NS5B region (Fig. 2A). Even more dramatic was the complete conservation between Moore (UK, 1964) and Old Lederle (US, 1946). This latter case seems so extreme as to suggest an escape or a cross-contamination event. There are other examples of good conservation: for example, the Polish viruses (1991–1993), the Belgian viruses (1990–1994) and the Austrian SP viruses (four of which are completely conserved in both the E2 and NS5B regions).

The sequenced E2 region is known to encode epitopes crucial to antibody mediated virus neutralization (van Rijn et al., 1994) and so may be under immunological selective pressure. The analogous region of hepatitis C virus may be unchangeable during acute infections but may change rapidly during persistent infections that are accompanied by a humoral immune response (van Doorn et al., 1995). As with hepatitis C, CSFV can occur in a variety of forms. Infection with virulent CSFV is an acute disease, characterized by a high mortality, with death or recovery within 10–30 days. Furthermore, affected pigs may be killed sooner due to the imposition of slaughter policies. It is therefore possible that the duration of the disease is usually insufficient to allow the selection of variable mutants, for instance by antibody maturation. This situation would be analogous to acute infections with hepatitis C virus which result in little virus variation. It might, however, be expected that greater variation would be evident in viruses of lesser virulence that produce chronic disease, sometimes accompanied by fluctuating antibody responses. We have found little evidence for rapid changes in the deduced amino acid sequences of the viruses that we have analysed and are not able to discern a clear relationship between variability and virulence. Some of the most variable viruses (Congenital Tremor, Kanagawa, Bergen and Zoelen) are of low virulence, but others such as Brescia are not. Also, our data suggest proportionally similar variations between viruses when analysing different genomic regions (including the 5′NCR) and using either nucleotide or derived amino acid comparisons, which is not consistent with immunological selective pressure exerted at a particular point (e.g. E2). An exception is
virus s7D2, which for E2 shows little divergence from other group 2 viruses at the nucleic acid level, but which is the second most distinctive virus of all 115 analysed at the deduced amino acid level. The origins of this virus are unknown but it was a component of a virus mixture (see Lowings et al., 1994) which has been classified as being of low virulence.

The rate of CSFV evolution can be estimated from nucleotide sequence variability by comparing viruses with a common source, but with widely separated isolation dates. This creates a paradox as no system in the field is completely closed and hence there is always a risk (which increases with time) of the introduction of an unrelated isolate. The best model for a rate of evolution that we have is the central Italian group (c1W to c4D). There are a number of reasons for its unique suitability. (1) The viruses form a distinct cluster allowing them to be readily distinguished from new introductions. (2) The viruses are strongly suspected of originating from a single source at a known time (swill feeding from an International camp site in 1985). (3) The area of origin is relatively geographically isolated, with respect to movement of wild boar, and trade in pigs. (4) The 6 year time span between the first and last isolate reduces the risk of obtaining a distorted value due to chance. We have already stated that c1W (1985) is identical to c3D (1991), suggesting that the rate can, in some instances, be near to zero. However, comparison of c1W with c4D (1991) results in an estimated rate of $2.7 \times 10^{-3}$ substitutions per base per year in the E2 region. In addition, a similar value was obtained when all probable terminal branch progenitor/progeny relationships shown on the E2 tree were analysed and averaged. The Italian rate was used to test a series of hypotheses concerning the evolutionary history of the other CSFVs. The first hypothesis was that viruses closely related to our oldest viruses, e.g. ALD (1946) or Brescia (1945), are the progenitors for all the current groups. Plotting the year of isolation against the percentage nucleotide substitutions for the ALD isolate compared to members of the current subgroups (Fig. 3A) reveals that the required evolution rates would be comparable with the Italian rate. A second hypothesis would be that the two subdivisions of group 1.1 have arisen by divergent virus evolution. Again, the proposed evolution rate required for this ($2.3 \times 10^{-3}$) is compatible with our estimation from the Italian viruses. Similar rates (Fig. 3B) were obtained when Brescia was used as the progenitor of the current subgroups with all predicted rates falling within those of the ALD set with the exception of Alfort 187, which would require more rapid evolution. A proposed evolution from ALD or Brescia to Kanagawa and Congenital Tremor would require rates of evolution up to five times those of the Italian viruses, making this suggestion unlikely without postulating some strong selective pressure. Interestingly, despite the highly divergent nature of these two isolates they remain much more closely related to CSFV than to any other known pestivirus group. Including data from divergent members of each subgroup in Fig. 3 showed that generally the required rates of evolution would have to be slightly higher than those of the more typical members of the subgroup. A proposed evolution from Brescia to Fort Dodge or Colorado would require extremely rapid evolution, suggesting that these two viruses evolved from a progenitor more like ALD. Despite the agreement of evolution rates and the apparent logic of these hypotheses, it should be noted that the unrooted phylogenetic tree presented in Fig. 1 does not fully support the theory of evolution of recent isolates from a progenitor similar to ALD or Brescia. Reasons for this discrepancy could be a genuinely uneven rate of evolution for different lineages or one or more of the other distorting factors described above, such as the reintroduction of the same virus at different times.

Using the equation given in Methods in combination with our proposed rate of evolution and the genetic distance between two viruses of the same age, it was possible to calculate the approximate times of divergence of the virus subtypes. Fig. 4 gives a pictorial representation of the proposed evolution of the CSFVs. We suggest that during the early 1950s a progenitor virus split into two lineages, one of which split again during the early 1960s to eventually form the 1986 EVI (subgroup 1.1) and Malaysian group (subgroup 1.2) viruses. The other lineage formed by the 1950s’ split divided into three lineages during the mid-1960s. These three groups eventually formed the 2.1, 2.2 and 2.3 subgroups. We do not have any examples of viruses from the 1960s that appear to be prototypes for these groups; however, we have one virus, Osaka, a Japanese isolate imported from the UK in 1971, which appears to fit this hypothesis reasonably well, being near to the progenitor of subgroup 2.3 and separating it from 2.2 and 2.1. If this evolutionary scenario is correct, then the type of virus at present affecting Europe (mostly subgroup 2.3 viruses) is genetically quite distinct form those responsible for the outbreaks in the 1940s and 1950s. This model assumes that evolution rates are constant between all branches. Clearly this may not be the case; however, the branch point dates are remarkably consistent, with all subgroups predicting the same evolution pattern to within 3 years. The Alfort and Alfort 187 viruses purport to be derived from a single isolation in 1966 (Aynaud, 1968), but this now seems unlikely. Alfort 187 (group 1.1) appears similar to viruses isolated in the 1950s and 1960s in Europe and the US. This fits well with its derivation by end point biological cloning in pigs using the original Alfort (Dahle et al., 1987) but throws into question the
nature of the virus sequenced by Meyers et al. (1989), which appears to be a typical group 2.3 virus similar to many viruses isolated in Europe in the mid 1980s. For a transition to have occurred from one type to the other, post-isolation, would have required an exceptionally high evolution rate (Fig. 3) during passage in culture (Kassimi et al., 1995). In addition, the pattern of evolution would have had to mimic the direction of the proposed natural evolution, i.e. from type 1.1 to type 2.3.

When the E2 regions of all the available pestivirus sequences were translated and aligned, a consensus sequence revealed a number of short conserved motifs. Of the three motifs analysed, only RYL(AV)(X)(LV)H was sufficiently rare to warrant further investigation. Curiously, the motif that most closely matched this pattern (RYL(AV)H) was found exclusively on a series of cellular receptors. This may be coincidental or may indicate pestivirus mimicry. To speculate, pestiviruses could, for instance, use such a motif to dimerize with the receptor or an associated cellular factor as a prelude to membrane fusion and/or some form of cell stimulation. Whatever the function of the motif, it is present in all 115 CSFVs, 21 BVDVs, three BVDs and one atypical pestivirus and in a reduced form (RYL) in the NS1 gene of dengue-2 virus (although apparently not in hepatitis C virus). The motif adjoins an immunological epitope found in BVDV (Paton et al., 1992) but not CSFV (van Rijn et al., 1994). The cysteine at Alfort amino acid residue 737 is also conserved throughout the studied pestiviruses and may therefore be a critical determinant of the secondary structure of the E2 protein.

We have analysed multiple strains of CSFV using sequence data from three regions of the genome, restriction sites from one region and MAb reactivity. All methods support the division of CSFV into two groups. The sequencing based methods allow further classification to a subgroup level. Of the three regions analysed, the E2 segment gives the greatest resolution with the greatest statistical confidence. At a low resolution, the isolates have been shown to be divided on the basis of their age (Edwards & Sands, 1990; Harding et al., 1994; Hofmann et al., 1994). It is possible that the old viruses have evolved into the more recent ones. Approximate rates of evolution obtained independently supported this hypothesis. Using the evolution rate obtained, we are able to suggest divergence dates for the major subgroups. Two highly variable isolates, Kanagawa and Congenital Tremor, have either diverged earlier than the main progenitor (ALD or Brescia type) or have experienced strong selective pressure (possibly as a result of their low virulence). Comparison between the translated region of E2 from all pestivirus types produced a conserved consensus sequence which had homology to a series of cellular receptors; the significance of this remains to be demonstrated.

We thank all of those who supplied us with isolates of CSFV, notably, Dr R. Ahl, Dr F. Koenen, Dr R. Krassung, Dr Z. Pejsak, Dr P. Roche and Dr D. Rutili. This work was supported by MAFF, UK and benefited from the use of the SEQuNET facility.

References


Diversity of CSFV


(Received 6 October 1995; Accepted 23 January 1996)