Molecular epidemiology of current classical swine fever virus isolates of wild boar in Germany

Immanuel Leifer,1 Bernd Hoffmann,1 Dirk Höper,1 Thomas Bruun Rasmussen,2 Sandra Blome,1 Günter Strebelow,1 Detlef Höreth-Böntgen,3 Christoph Staubach3 and Martin Beer1

Correspondence
Martin Beer
martin.beer@fli.bund.de

1Institute of Diagnostic Virology, Friedrich-Loeffler-Institut, Südufer 10, D-17493 Greifswald-Insel Riems, Germany
2DTU National Veterinary Institute, Technical University of Denmark, Lindholm, DK-4771 Kalvehave, Denmark
3Institute of Epidemiology, Friedrich-Loeffler-Institut, Seestrasse 55, D-16868 Wusterhausen, Germany

Classical swine fever (CSF) has caused significant economic losses in industrialized pig production, and is still present in some European countries. Recent CSF outbreaks in Europe were mainly associated with strains of genogroup 2 (subgroup 2.3). Although there are extensive datasets regarding 2.3 strains, there is very little information available on longer fragments or whole classical swine fever virus (CSFV) genomes. Furthermore, there are no detailed analyses of the molecular epidemiology of CSFV wild boar isolates available. Nevertheless, complete genome sequences are supportive in phylogenetic analyses, especially in affected wild boar populations. Here, German CSFV strains of subgroup 2.3 were fully sequenced using two different approaches: (i) a universal panel of CSFV primers that were developed to amplify the complete genome in overlapping fragments for chain-terminator sequencing; and (ii) generation of a single full-length amplicon of the CSFV genome obtained by long-range RT-PCR for deep sequencing with next-generation sequencing technology. In total, five different strains of CSFV subgroup 2.3 were completely sequenced using these newly developed protocols. The approach was used to study virus spread and evolutionary history in German wild boar. For the first time, the results of our study clearly argue for the possibility of a long-term persistence of genotype 2.3 CSFV strains in affected regions at an almost undetectable level, even after long-term oral vaccination campaigns with intensive monitoring. Hence, regional persistence in wild boar populations has to be taken into account as an important factor in the continual outbreaks in affected areas.

INTRODUCTION

Classical swine fever (CSF) is one of the most important diseases of pigs and is caused by classical swine fever virus (CSFV). Symptoms of acute infection are pyrexia, respiratory and gastrointestinal symptoms, haemorrhages and ataxia with moderate to high mortality depending on the strain (Kaden et al., 2000) and the age class of animals involved (Moennig et al., 2003). During the last decades, the disease has been present in several European countries as well as in many other areas of the world (Depner et al., 2006; Edwards et al., 2000; Pol et al., 2008).

CSFV, Border disease virus and Bovine viral diarrhea virus (BVDV) are members of the genus Pestivirus in the family Flaviviridae (Fauquet & Fargette, 2005). Pestiviruses are positive-sense, ssRNA viruses with a genome of approximately 12 300 nt (Meyers & Thiel, 1996). The genome of CSFV comprises one ORF encoding a polyprotein of about 4000 aa. The ORF is flanked by terminal 5′- and 3′-NTRs. The polyprotein is processed co- and post-translationally by viral and cellular proteases (Meyers & Thiel, 1996) to obtain four structural (C, ENS, E1 and E2) and nine non-structural (NPSO, p7, NS2, NS3, NS2-3, NS4A, NS4B, NS5A and NS5B) proteins (Rümkenaf et al., 1993; Tautz et al., 1993, 1997; Thiel et al., 1989; Wegelt et al., 2009).

CSFV strains can be divided into three different genogroups (1–3). Each genogroup consists of three or four subgroups (Greiser-Wilke et al., 1998). Recent CSF outbreaks in Europe were mainly associated with moderately virulent isolates of
subgroup 2.3 (Kaden et al., 2004; Pol et al., 2008). The present method for classification of CSFV strains uses two short genome fragments: a 150 nt region of the highly conserved 5′-NTR and a 190 nt region of the more variable E2 protein-encoding region (Greiser-Wilke et al., 2006). Mainly based on these fragments, a CSFV database is maintained at the Community Reference Laboratory (CRL) for CSF in Hannover, Germany (Greiser-Wilke et al., 2000, 2006). When this database and genetic typing started, mainly German isolates were available. Using the 5′-NTR sequences, it was found that the isolates seemed to be characteristic for certain regions where outbreaks occurred and, to distinguish between isolates, they were given names of places where they were first found, e.g. 2.3* Uelzen, 2.3* Rostock, 2.3* Güstrow (Fritzemeier et al., 2000). Based on these sequences, outbreak history could be followed in Germany, and isolate clusters also seemed to be characteristic for CSF scenarios in affected wild boar populations (e.g. 2.3* Rostock vs 2.3* Uelzen).

In 2009, CSF in wild boar was detected in the German federal state North Rhine Westphalia in an area close to an old vaccination zone against CSF in wild boar, but separated from this region by the large river Rhine. While the old restriction zones lay on the west side surrounding the city of Euskirchen, the new cases were found on the east side. On the west side, CSF had been detected with some time gaps between 2002 and mid-2007. The virus type found there was so far attributed to 2.3* Rostock. The first 2009 isolate was quite similar, but showed distinct differences to the traditional 2.3* Rostock cluster. Data obtained by partial sequencing carried out as part of routine diagnosis did not allow for detailed molecular epidemiology and showed problems for subgroup nomenclature. This was mainly due to the high identity among different isolates. For this reason, a high-resolution approach such as full-genome sequencing seemed more adequate.

This project was aimed at designing novel RT-PCR systems for different sequencing approaches and analysis of their applicability for full-genome sequencing. In addition, the obtained sequence data, together with sequence information available in GenBank (http://www.ncbi.nlm.nih.gov/GenBank/), were used for phylogenetic analysis as a basis for molecular epidemiology of recent German CSFV isolates from wild boar. In total, five different German CSFV isolates were fully sequenced in the framework of this project and fed into GenBank. These results were compared to analyses done after partial sequencing.

RESULTS

Partial sequencing of E2 and 5′-NTR fragments

During the 2009 outbreak among German wild boars, two closely related but distinct isolates were found in the affected region. The restriction zones and the isolate classification made after partial sequencing are depicted in Fig. 1. Interestingly, it was observed that, especially with the isolate found near the city of Roesrath, the historical nomenclature of CSFV isolates described above was misleading. Due to the fact that this isolate bore characteristics of both clusters 2.3* Uelzen and 2.3* Rostock, the close relationship to virus strains isolated in the neighbouring Euskirchen area remained hidden. In detail, the CSFV isolated in Roesrath showed at position 207 (based on the sequence of Alfort187) a tyrosine that is characteristic for the cluster 2.3* Uelzen (cytosine for 2.3* Rostock), and also a tyrosine at position 138 that is characteristic for the 2.3* Rostock cluster (guanidine for 2.3* Uelzen). Only after the second isolate was detected did the full extent of the relationship become clear. This isolate that was detected near the city of Hennef turned out to be identical to former isolates from the previous vaccination area of Euskirchen, to the west of the river Rhine where the last virus-positive case had been detected in July 2007. Comparing the isolates ‘Roesrath’ and ‘Hennef’, a unique nucleotide exchange was observed at position 64. While both clusters 2.3* Uelzen and 2.3* Rostock showed a guanine at this position, both new isolates as well as the old ‘Euskirchen’ isolate showed an adenine. In addition, unique patterns were found in the glycoprotein E2-encoding region, also justifying a detailed analysis by full-length sequencing.

Complete genome sequences

Five CSFV isolates of genotype 2.3 and one isolate of genotype 1.1 were completely sequenced, and sequencing data were released into GenBank with the following accession numbers: GU233731 (CSFV/2.3/dp/CSF857/2006/Borken), GU233732 (CSFV/2.3/wb/XXX0608/2005/Euskirchen), GU233733 (CSFV/2.3/wb/CSF1046/2009/Hennef), GU233734 (CSFV/2.3/wb/CSF1045/2009/Roesrath), GU324242 (CSFV/2.3/wb/XXX0609/2004/Uelzen) and HM237795 (CSFV/1.1/dp/CSF0382/XXX/Koslov). In order to confirm the final genome sequences, they were translated into protein sequences, and a 3898 aa polyprotein was verified. With the exception of CSFV strain ‘Borken’, all CSFV strains of genotype 2.3 possessed a 373 nt 5′-NTR and a 227 nt 3′-NTR. In contrast, CSFV strain ‘Borken’ had a 5′-NTR of 374 nt and a 3′-NTR of 226 nt.

Phylogenetic analysis, molecular epidemiology and quasispecies analysis

Comparing nucleotide differences in the individual genome regions, interestingly, the p7 protein-encoding region showed the highest identity (<1% nucleotide differences), followed by the NS3-encoding region (approx. 1% nucleotide differences) and the 5′-NTR (approx. 1.2% nucleotide differences). Slight higher variability was observed in the E1 (approx. 1.25% nucleotide differences), ENS (approx. 1.3% nucleotide differences) and NS5 (approx. 1.7% nucleotide differences)-encoding regions. The most variable regions were the 3′-NTR (approx. 2.3% nucleotide differences) and the E2-, NPP- and NS4-encoding regions, with about 2.1% nucleotide differences (Table 1).
Looking at amino acid differences in the respective protein-encoding regions, the overall situation mirrors the nucleotide level. The more conserved regions at the nucleotide level stay the same and this is also the case for the more conserved regions at the amino acid level. However, there are many silent mutations and, on average, one can find one amino acid exchange per 5.7 nucleotide exchanges. Remarkably, this is not the case for the ERNS and the NS3 proteins. The NS3 protein is the most conserved protein with 0.06 % amino acid exchanges, followed by the E1 (0.22 %) and ERNS (0.25 %) proteins (Table 1). In the NS3 protein the ratio between amino acid exchanges and nucleotide exchanges is about 1 : 40.

Based on the full-length sequences, isolates CSFV/2.3/wb/CSF1046/2009/Hennef and CSFV/2.3/wb/CSF1045/2009/Roesrath showed the highest genome identity, with only 13 nucleotide differences. Compared with isolate CSFV/2.3/wb/XXX0608/2005/Euskirchen, they exhibited 44 (‘Hennef’) and 53 (‘Roesrath’) nucleotide differences, respectively. These three isolates are highly similar to the former 2.3* Rostock cluster. If the isolate ‘Euskirchen’ is defined as the common ancestor of ‘Hennef’ and ‘Roesrath’, the nucleotide-exchange rate would be about 10–15 nt year$^{-1}$ for the full-length genome.

In contrast, the CSFV strain isolated in Borken showed 156 nucleotide differences from isolate ‘Euskirchen’. This isolate could be placed into the former 2.3* Güstrow cluster. Taking into account the small differences observed between isolates ‘Euskirchen’, ‘Hennef’ and ‘Roesrath’, isolate ‘Borken’ does not seem to have a direct correlation with the latter. Compared again with isolate ‘Euskirchen’, the ‘Uelzen’ isolate was most divergent as it displayed 396 nucleotide exchanges. It could be clearly placed into the old cluster 2.3* Uelzen. The Spanish and the Alfort/Tübingen isolates in the database could not be directly placed into one of the German clusters of subgroup 2.3 (Fig. 2).

Neighbour-joining (NJ) trees of all seven CSFV isolates discussed here were constructed. Fig. 2 shows the NJ trees based on the complete genome (a), the 5’-NTR (b), Npro protein (c)- or E2 protein (d)-encoding sequences. Considering complete genome sequences (Fig. 2a), CSFV/2.3/wb/CSF1046/2009/Hennef and CSFV/2.3/wb/CSF1045/2009/Roesrath, together with the next most closely related isolate CSFV/2.3/wb/XXX0608/2005/Euskirchen, could be grouped into one subgroup ‘Euskirchen’. These three isolates also represent the former 2.3* Rostock cluster. The CSFV/2.3/dp/CSF857/2006/Borken isolate is the most
closely related to this subgroup ‘Euskirchen’, whereas isolates CSFV/2.3/wb/XXX0609/2004/Uelzen and Alfort/Tuebingen are the most divergent. The same allocation is found if the NJ trees of the 5′-NTR and the E2 protein sequences are considered. Looking at the N pro NJ tree, the ‘Euskirchen’ subgroup stays at the same position, but the most divergent isolate is the Spanish one.

The allocation displayed when using the complete genome sequence is generally confirmed, independent of the region used for NJ tree construction. The ‘Euskirchen’ subgroup is always present, and the most divergent isolate is either the Spanish genogroup 2.3 isolate or CSFV/2.3/wb/XXX0609/2004/Uelzen.

In order to address the phenomenon of quasispecies evolution among the closely related strains ‘Euskirchen’ and ‘Hennef’, the complete datasets obtained after pyrosequencing were compared with the most likely consensus sequences of these strains. Eleven nucleotide positions showed characteristics of quasispecies among the three isolates ‘Euskirchen’, ‘Hennef’ and ‘Roesrath’. The variable nucleotide positions are shown in Table 2, with both the corresponding nucleotides and the percentage of variability. Here, the nucleotide differences of the isolate ‘Hennef’ are already found as a quasispecies proportion of 30–50% in the presumable parental isolate ‘Euskirchen’. On the other hand, the nucleotide differences from isolate ‘Euskirchen’ are still maintained in the ‘Hennef’ isolate within the range 10–20% (Table 2).

**DISCUSSION**

Sequencing and genetic typing are widely used to support epidemiologists in their efforts to trace the origin and spread of the virus causing CSF outbreaks. It also gives an indication regarding genetic variation and quasispecies evolution (Li et al., 2006). Moreover, it can be useful for the differentiation of infected from vaccinated animals in scenarios involving (emergency) vaccination (Leifer et al., 2009).

Different genome regions have been recommended for differentiation and genotyping of CSFV isolates, such as the 5′-NTR (Greiser-Wilke et al., 1998, 2006; Paton et al.,...
Fig. 2. Phylogeny and classification of seven CSFV strains by NJ using the bootstrap method with 500 bootstrapping replicates. The NJ trees are based on complete genome (a), 5’-NTR (b), Npro protein-encoding (c) or E2 protein-encoding (d) sequences. (e) Complete genome sequences of the genogroup 2.3 isolates compared with two isolates of genogroup 1 (CSFV/1.1/XX/CSF0902/1968/Alfort187 and CSFV/1.1/dp/CSF0382/XXX/Koslov). For reconstruction of NJ trees, the MEGA 4 software (Center for Evolutionary Functional Genomics) was used. Bars, relative phylogenetic distance between respective isolates. Bootstrap values (%) are indicated next to branches.
Table 2. Comparison of complete datasets obtained after pyrosequencing with the most likely consensus sequences of the individual strains

Variable nucleotide positions with the corresponding nucleotides in the genomes and the percentage of variability of the nucleotides found in the related strains at the same position are shown. Variability data from pyrosequencing were available for the strains CSFV 'Euskirchen' and CSFV 'Hennef'.

<table>
<thead>
<tr>
<th>Nucleotide position</th>
<th>CSFV 'Euskirchen'</th>
<th>CSFV 'Hennef'</th>
<th>CSFV 'Roesrath'</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nucleotide in</td>
<td>Variability</td>
<td>Nucleotide in</td>
</tr>
<tr>
<td></td>
<td>consensus sequence</td>
<td></td>
<td>consensus sequence</td>
</tr>
<tr>
<td>523</td>
<td>A</td>
<td>37% C</td>
<td>C</td>
</tr>
<tr>
<td>1423</td>
<td>C</td>
<td>46% T</td>
<td>T</td>
</tr>
<tr>
<td>4312</td>
<td>A</td>
<td>46% G</td>
<td>G</td>
</tr>
<tr>
<td>4111</td>
<td>C</td>
<td>41% A</td>
<td>A</td>
</tr>
<tr>
<td>7324</td>
<td>A</td>
<td>50% G</td>
<td>G</td>
</tr>
<tr>
<td>7492</td>
<td>A</td>
<td>37% G</td>
<td>G</td>
</tr>
<tr>
<td>7505</td>
<td>A</td>
<td>37% G</td>
<td>G</td>
</tr>
<tr>
<td>7540</td>
<td>A</td>
<td>38% G</td>
<td>G</td>
</tr>
<tr>
<td>8161</td>
<td>T</td>
<td>49% C</td>
<td>C</td>
</tr>
<tr>
<td>8911</td>
<td>A</td>
<td>41% G</td>
<td>G</td>
</tr>
<tr>
<td>10980</td>
<td>A</td>
<td>33% G</td>
<td>A</td>
</tr>
</tbody>
</table>

2000), the Npro-encoding region (Becher et al., 1997), the E2-encoding region (Lowings et al., 1999) and the NS5B-encoding region (Bjorklund et al., 1999). For a fast diagnostic approach, short 5'S-NTR (150 nt)- and E2 (190 nt)-encoding sequence fragments are usually used for sequencing and subsequent typing. Based on these fragments, several studies have been published with CSFV isolates from all over the world (Bartak & Greiser-Wilke, 2000; Biagetti et al., 2001; Blacksell et al., 2004; Cha et al., 2007; Jemersic et al., 2003; Li et al., 2006; Pereda et al., 2005; Sabogal et al., 2006; Stadejek et al., 1997; Tu et al., 2001; Vilcek et al., 1997; Vlasova et al., 2003). Regarding these fragments, extensive datasets are available at the CRL (Greiser-Wilke et al., 2006). Using this approach, German CSFV isolates of subgroup 2.3 are grouped into further clusters. These are, among other clusters, 2.3* Güstrow, 2.3* Rostock, 2.3* Spreda, 2.3* Uelzen and 2.3* Warnow (Fritzemeier et al., 2000; Kaden et al., 2004).

While this approach is suitable for rapid characterization in outbreak situations, it has its limitations when highly similar CSFV strains are compared. Therefore, full-length genome sequencing can be a valuable tool for detailed molecular epidemiology, and must also be considered in future outbreaks of CSFV. Unfortunately, the availability of complete genome sequences of subgroup 2.3 isolates in databases is very limited and there is so far no agreement on how to define new groups, subgroups or clusters. The most recent sequence data are from a 2001 Spanish isolate (GenBank accession no. FJ265020.1) and a German sample (J04358.2/HCVCGSA), where time and place of the first occurrence are unknown.

Here, two approaches for full-genome sequencing were used to obtain reliable sequence data. Newly developed (fragmented amplification) or adjusted (long-range) RT-PCR systems were used to generate stock material for Sanger sequencing and next-generation sequencing, respectively. Complete sequences of five current German isolates of subgroup 2.3 are reported, evaluated and used for detailed molecular epidemiology. In general, both methods produced reliable sequencing results in the first round of sequencing. However, a few single nucleotide differences and gaps were only resolved after further sequencing and a close inspection of the individual results. Most discrepancies probably result from sequencing errors. These can result from algorithm errors in the analysing software or emerge from RT or PCR steps. Moreover, selection of enzymes used for amplification can be crucial for accuracy (Eckert & Kunkel, 1990; Koehavong & Thilly, 1989; Tindall & Kunkel, 1988). Especially with the Sanger method, sequencing errors could be attributed to weak signals in individual sequencing read-outs in the 3130 Genetic Analyzer. For the fragmented approach, reliability of sequencing also depends on the overlap and the length of the sequenced fragments. However, the main disadvantage of Sanger sequencing is the low sequence depth. In contrast, using GS FLX pyrosequencing, 100–1000-fold coverage of the complete target is possible depending on the starting material, amount and size of template, providing very reliable sequencing results (Rozera et al., 2009). GS FLX pyrosequencing has therefore some benefits over Sanger sequencing, such as: (i) one primer pair is sufficient for genome amplification; (ii) it is more economical and faster; (iii) statistical analysis of a quasispecies is possible; (iv) higher genome coverage gives more reliable sequencing results; and (v) several CSFV genomes can be sequenced in parallel in a single run.

Phylogenetic analysis of the obtained genome sequences showed that isolates ‘Euskirchen’, ‘Hennef’ and ‘Roesrath’ are closely related. Using the traditional approach, isolates
‘Euskirchen’ and ‘Hennef’ are identical and can be grouped into the former German cluster 2.3* Rostock. This fits with observations of a geographical distribution pattern of CSFV isolates in affected wild boar populations (von Rüden et al., 2008). Using this method with very short sequence fragments, the ‘Roesrath’ isolate showed one nucleotide exchange in both the 5’-NTR and the E2 fragments. This slight but distinct difference led to the problem that the ‘Roesrath’ isolate could not be easily placed into this cluster, as it bore both characteristics of 2.3* Rostock and 2.3* Uelzen. Further investigations of those short fragments showed that these three isolates have a significant pattern in the E2 fragment that differentiates them from other strains in both clusters 2.3* Uelzen and 2.3* Rostock. Using complete genome sequences, this study showed 13 nucleotide differences between the ‘Hennef’ and ‘Roesrath’ isolates, leading to the assumption that both isolates may originate from one parental CSFV isolate. Taking into account that the ‘Hennef’ isolate shows 100% identity with the ‘Euskirchen’ isolate based on the traditional fragments, the ‘Euskirchen’ isolate from 2005 could be the parental virus (Fig. 2). Compared with the 2005 CSFV strain ‘Euskirchen’, 44 nucleotide differences were found in the complete genome of the 2009 strain ‘Hennef’ and 53 nucleotide differences in the 2009 ‘Roesrath’ isolate (Table 1).

Working with the reasonable hypothesis that the ‘Euskirchen’ strain could be the ancestor of the recent CSFV isolates ‘Hennef’ and ‘Roesrath’, regional evolution of quasispecies can be addressed. The first cases were detected in the region of Euskirchen in 2002. Unfortunately, the strains involved were not available for full-length sequencing. The first available strain dated from 2005. Based on this strain, it can be speculated that, after 2005, CSFV ‘Euskirchen’ appeared in the new region. Assuming an exchange rate of about 10–15 nt year⁻¹, the 13 nucleotide differences between the 2009 CSFV isolates ‘Hennef’ and ‘Roesrath’ would suggest a separation of these strains in 2008, and both strains are likely to have diverged from the ‘Euskirchen’ strain. In conclusion, this hypothesis could also suggest the presence of 2.3 CSFV strains in German wild boar populations in 2008 that were not detected by intensive surveillance during this period (e.g. 1760 km² in the Euskirchen area; n=8712 samples without any positive cases in 2008; 95% confidence interval for the virological prevalence, 0.000–0.0344%). Should this be the case, very low virus prevalence could have led to non-detection. It is known that virological prevalence is always low, even in outbreak situations without vaccination, due to the natural development of population immunity (Anonymous, 1999; Thulke et al., 2009). As one can expect that any introduction and spread of CSF in a naïve population would trigger a significant increase of seroprevalence, these findings are surprising, especially when taking into account that the sample size taken was statistically high enough to detect an introduction during the study period (with 95% probability to detect at least 1.74% seroprevalence each year). Thus, in the light of the aforementioned facts, namely the relationship of virus isolates over time and the persistence of CSFV in the affected region, it seems likely that CSFV persisted in the affected region despite extensive use of vaccination and surveillance measures. This could have serious implications for future campaigns. So far, vaccination campaigns were carried out to eradicate the disease from wildlife in order to protect domestic pig herds. However, it seems that domestic pigs were sufficiently protected, but the disease has not been eradicated. Although tremendous efforts were necessary to test a statistically sufficient sample set, the virus seems to have escaped testing from mid-2007 to the end of 2008. Future options need to be discussed based on these findings. In addition, it must be discussed how persistence was possible. One hypothesis is, for example, that single chronically infected animals remain undetected in vaccinated, and thus almost immune, sounders, shedding the virus for up to several months. This way, contact with only a small proportion of unvaccinated and unprotected wild boar could have led to the spread and flare-up of CSF. Chronically diseased animals are known to survive for months and they constantly shed high amounts of virus (Moennig et al., 2003).

Quasispecies evolution was analysed using the complete datasets obtained after pyrosequencing. It could be shown that 11 variable nucleotide positions showed quasispecies characteristics with different proportions of the respective quasispecies among the isolates in question. The results substantiate the assumption that CSFV strains ‘Euskirchen’, ‘Roesrath’ and ‘Hennef’ have a common ancestor and evolved over time.

In summary, this study provides new insights into the phylogeny of CSFV strains of the relevant subgroup 2.3, namely the regional emergence of viral quasispecies as described for the ‘Euskirchen’, ‘Hennef’ and ‘Roesrath’ isolates. Furthermore, the methods described here enable easy and economical sequencing of complete CSFV genomes. In order to test the applicability of both methods for other CSFV genogroups, the complete genome of the highly virulent CSFV strain ‘Koslov’ (CSFV/1.1/dp/CSF0382/XXXX/Koslov) was also sequenced and compared (data not shown).

Besides the phylogenetic approach, full-genome sequencing also facilitates research into virulence markers and genetic modification of CSFV for DIVA (Differentiating Infected from Vaccinated Animals) vaccine development. Finally, this kind of molecular epidemiology study of CSFV will also support control measures in future CSF outbreaks.

**METHODS**

**Cells and viruses.** Cells and viruses were grown in Dulbecco’s modified Eagle’s medium supplemented with 10% FBS (both from Biochrom AG) free of BVDV at 37 °C in a humidified atmosphere containing 5% CO₂. The porcine kidney cell line 15 (PK15; ATCC)
was obtained from the Collection of Cell Lines in Veterinary Medicine, Friedrich-Loeffler-Institut, Insel Riems, Germany.

Virus strains were chosen to represent the recent outbreak among German wild boars and prototype viruses of clusters 2.3\textsuperscript{a} Rostock and 2.3\textsuperscript{a} Uelzen according to the traditional nomenclature. In addition, CSFV ‘Koslov’ represented genogroup 1.1 strains.

To streamline the nomenclature of CSFV isolates, a new approach that was discussed with the CRL for CSF is introduced here. The notation contains the most relevant information about the virus. These are: (i) the species CSFV; (ii) the genogroup including the subgroup; (iii) the source of isolation (dp, domestic pig; wb, wild boar); (iv) the CRL CSVF database entry number consisting of CSF and a four-digit number for virus strains available at the CRL; and XXX and a four-digit number for strains where only sequences are available at the CRL; (v) the year of isolation (where not available, XXXX was put as a placeholder); and (vi) the place of first occurrence.

All virus isolates (CSFV/1.1/dp/CSF0382/XXXX/Koslov; CSFV/2.3/dp/CSF0857/2006/Borken, CSFV/2.3/wb/XXX0609/2004/Uelzen, CSFV/2.3/wb/XXX0608/2005/Euskirchen, CSFV/2.3/wb/CSF1046/2009/Hennef and CSFV/2.3/wb/CSF1045/2009/Roesrath) were obtained from the German National Reference Laboratory for CSF. CSFV strains were cultured by passing them three times on PK15 cells, according to standard protocols (Technical Annex to Commission Decision 2002/106/EC). CSFV isolate CSVF/2.3/wb/XXX0609/2004/Uelzen was passaged five times on PK15 cells to obtain an adequate virus titre for RT-PCR amplification and subsequent sequencing.

**RNA isolation.** For RNA isolation from blood or cell-culture supernatant, an RNeasy Mini kit (Qiagen GmbH) was used according to the manufacturer’s instructions. For full-length RT-PCR and RACE RT-PCR, the following modifications were made: instead of RLT buffer (part of the RNeasy Mini kit), 750 µl TRIzol LS (Invitrogen) was added to 250 µl sample and incubated at room temperature for 5 min and, subsequently, 200 µl chloroform was added. After 3 min incubation, the liquid phase was separated by centrifugation for 15 min at 5300 g and 4 °C and used for further purification according to the standard protocol. DNA was eluted from the columns in three serial elution steps using 30 µl water for each step. Second and third eluates were used for long RT-PCR (Rasmussen et al., 2008, 2010).

**Primer and probe selection.** A new panel of CSFV-specific primers for amplification of the complete genome in approximately 1000 bp-long overlapping fragments was chosen after alignment of 28 different complete CSFV genome sequences available from the NCBI nucleotide database. Primers included in the panel were selected to fit all known CSFV sequences for universal amplification of any isolate independent from its genogroup. ClustalW2 (EMBL-EBI; http://www.ebi.ac.uk/) (Larkin et al., 2007) and BioEdit (IBIS Biosciences) (Hall, 1999) were used for alignment and Beacon Designer 5.0 (Premier Biosoft International) was used for primer selection. Oligonucleotides were provided by Eurogentec (Eurogentec GmbH) (see Supplementary Table S1, available in JGV Online). Primers for long-range and RACE RT-PCR were selected after alignment of 5’ and 3’ ends sequence data of all CSFV isolates available in the NCBI nucleotide database. The long-range approach was designed in accordance with the system applied by Rasmussen et al. (2008, 2010). Modifications concerned the use of different primers (see Supplementary Table S1).

**PCR and RT-PCR.** Commercial kits for RACE RT-PCR (5’ RACE System and 3’ RACE System; Invitrogen) were used according to the manufacturer’s recommendations. For 3’ RACE, viral RNA was first polyadenylated using an A-Plus Poly(A) Polymerase Tailing kit (Epicentec Biotechnologies).

For reverse transcription of complete genome viral RNA, the SuperScript III First-Strand Synthesis System for RT-PCR (Invitrogen) was used. Reverse transcription was carried out at 50 °C for 90 min. For PCR amplification of complete genome cDNA, an AccuPrime Taq DNA Polymerase High Fidelity kit (Invitrogen) was used. Complete CSFV cDNA amplification was carried out over 40 cycles of 15 s at 94 °C (DNA denaturation), 30 s at 65 °C (annealing) and 12 min 30 s at 68 °C (elongation), with one initial 15 s denaturation step at 94 °C and one final elongation step for 12 min 30 s at 68 °C.

The long-range amplification technique was developed according to a method recently described for other pestiviruses and CSFV strains ‘Paderborn’ and C-strain ‘Riems’ (Rasmussen et al., 2008, 2010). In the framework of this study, the system was modified and optimized to amplify the known CSFV groups and subgroups.

RT-PCR for fragmented amplification of the viral genome was performed with a SuperScript III One-Step RT-PCR with Platinum Taq kit (Invitrogen) using the following RT-PCR program: 30 min reverse transcription at 50 °C followed by 2 min denaturation at 94 °C and 45 cycles of 30 s at 94 °C (denaturation), 30 s at 55 °C (annealing) and 2 min at 68 °C (elongation).

All RT-PCR and PCR protocols were performed with a Mastercycler gradient S (Eppendorf AG).

RT-PCR for the amplification of traditional E2 and 5’-NTR fragments was carried out as described elsewhere (Depner et al., 2006).

**Sequencing of PCR products.** While RT-PCR products obtained with the overlapping fragment approach and the traditional protocol were sequenced by the Sanger method (Sanger & Coulson, 1975), amplicons obtained by long-range RT-PCR were utilized for pyrosequencing with a Genome Sequencer (GS) FLX machine (Roche). For sequencing of the extreme 5’ and 3’ ends, commercial kits for RACE (Invitrogen) were used.

**Sanger sequencing.** DNA fragments obtained by standard RT-PCR for Sanger sequencing were isolated from agarose gels with a QIAquick Gel Extraction kit (Qiagen). Sequencing was carried out using a BigDye Terminator v1.1. Cycle Sequencing kit (Applied Biosystems). Nucleotide sequences were obtained with a 3130 Genetic Analyzer (Applied Biosystems). For construction of complete genome sequences, BioEdit (Ibis Biosciences) and Genetics Computer Group version 11.0 (Accelrys Inc.) software was used.

**Pyrosequencing with the genome sequencer FLX.** Full-genome DNA fragments (complete CSFV genomes) were isolated after agarose gel electrophoresis using a Zymoclean Gel DNA Recovery kit (Zymo Research Corporation) and were analysed using the GS FLX. To this end, DNA libraries were prepared according to the protocol of Wiley et al. (2009) followed by binding to library capture beads and recovery of the ssDNA template library. These libraries were used for sequencing in 1/16 of a pico titre plate according to the manufacturer’s instructions. The obtained raw sequence data were assembled using the GS assembler software NEWBIE v. (2.0.00.22; Roche).

**Generation of complete CSFV genome sequences.** Individual fragments obtained by Sanger sequencing were aligned and assembled to complete sequences of five German isolates (CSFV/2.3/dp/CSF857/2006/Borken, CSFV/2.3/wb/XXX0609/2004/Uelzen, CSFV/2.3/wb/XXX0608/2005/Euskirchen, CSFV/2.3/wb/CSF1046/2009/Hennef and CSFV/2.3/wb/CSF1045/2009/Roesrath). Three consensus sequences were obtained from long-range RT-PCR products by GS FLX pyrosequencing (CSFV/2.3/dp/CSF857/2006/Borken, CSFV/2.3/wb/XXX0608/2005/Euskirchen and CSFV/2.3/wb/CSF1046/2009/Hennef). For determination of the 3’ and 5’ ends, sequences obtained after RACE protocols were added.
For results obtained by Sanger sequencing, sequence depth is low. In contrast, results from GS FLX pyrosequencing are based on higher sequence depth. Here, complete datasets cover the genome up to 1000 times. For consensus sequence generation, about 10% of the complete GS FLX data were used. Fig. 3 shows an example of the depth of the GS FLX sequences using the reduced data. As can be seen, even in the reduced dataset, nucleotides were sequenced up to 70 times, thereby leading to the generation of very reliable consensus sequences. For comparison, data characterizing the sequence depths of all three isolates sequenced with the GS FLX are also depicted in Fig. 3.

Sequences from both methods were analysed using the online alignment software BLASTN (http://blast.ncbi.nlm.nih.gov/blast.cgi) (Altschul & Erickson, 1985; Altschul et al., 1990). After assembly and revision, matchable sequences were compared. To follow up minor differences, sequencing reactions and analyses were repeated until one final sequence of 12,297 bp for each isolate could be consolidated.

Phylogenetic analysis. For phylogenetic classification of CSFV strains (subgroup 2.3), neighbour-joining (NJ) trees were constructed using the bootstrap method with 500 bootstrapping replicates. To this end, the MEGA 4 software (Center for Evolutionary Functional Genomics, Tempe, AZ, USA) (Kumar et al., 2008) was used. The NJ trees are based on the complete genome, 5’-NTR, Npro protein- or E2 protein-encoding sequences.

The different strains were comparatively analysed, and variable as well as conserved regions within the CSFV strains of subgroup 2.3 were defined. For this purpose, the complete genomes were split into the regions encoding the different viral proteins as well as the 3’- and 5’-NTRs. Full-length sequences of CSFV subgroup 2.3 isolates available in GenBank were included in the comparison. In order to get an insight into the relationship between the strains, CSFV strain CSFV/2.3/wb/XXX0608/2005/Euskirchen, the possible ancestor of the German 2009 strains, was used as a reference. All full-length sequences were compared with this strain in terms of fragment identity. Moreover, the mean number of nucleotide exchanges and the percentage of exchanges per fragment were calculated.

To address the question of quasispecies, the appropriate reference sequences (‘Euskirchen’ and ‘Hennef’) were mapped using the GS FLX reference mapper software (version 2.3; Roche).

ACKNOWLEDGEMENTS

We thank Moctezuma Reimann and Kersten Biebl for excellent technical assistance and Michael Eschbaumer for critically reading this manuscript. Further thanks go to Andreas Fröhlich for statistical advice. This study was supported by the FP7 EU project ‘CSFV_goDIVA’ (no. 227003) and the German Federal Ministry of Food, Agriculture and Consumer Protection.

REFERENCES


