101

FURTHER ANALYSIS OF THE GENOME OF PORCINE EPIDEMIC DIARRHOEA VIRUS

A. Bridgen,[1,2] R. Kocherhans,[1] K. Tobler,[1] A. Carvajal,[1] and M. Ackermann[1]

[1]Institute of Virology
University of Zurich
Winterthurerstrasse 266a
CH 8057 Switzerland
[2]Institute of Virology
Church Street
Glasgow G11 5JR, Scotland

1. ABSTRACT

We report here the continued determination and analysis of the nucleotide sequence of both wild type (wt) and cell culture adapted (ca) porcine epidemic diarrhoea coronavirus (PEDV). These studies were undertaken with two objectives in mind: the identification of common and divergent features in the genomic sequences of wt and ca PEDV which can explain the differences in virulence of these isolates and the further exploration of the relationship of PEDV to other coronaviruses.

2. INTRODUCTION

PEDV is, as its name implies, responsible for causing diarrhoea in pigs, particularly in neonates. We have previously completed the sequencing of the M, sM, ORF 3 and N genes of the CV777 strain of PEDV. The S, M, sM, ORF 3 and N genes of a British strain of PEDV have also been sequenced (Duarte *et al.*, 1994; Duarte and Laude, 1994 and Bridgen *et al.*, 1993). These results indicated that PEDV occupies a position intermediate between the two well characterised members of the group 1 coronaviruses, transmissible gastroenteritis virus (TGEV) and human coronavirus (HCV) 229E.

Our original cloning approach involved designing primers based on conserved regions of the coronavirus M and N genes to amplify the equivalent as then unknown PEDV sequence (Bridgen *et al.*, 1994). In this study, we were interested in knowing whether this

Coronaviruses and Arteriviruses, edited by Enjuanes *et al.*
Plenum Press, New York, 1998

technique could be used to clone the ORF 1 of PEDV. Such a method is useful for viruses which do not grow to high titre, avoids lengthy screening of clones and could potentially be applied to the cloning of any group 1 coronavirus. However, the large size of the ORF (more than 20,000 nucleotides) and the paucity of sequence data from other coronaviruses make this an ambitious objective. Only two group 1 coronaviruses have been sequenced completely, HCV229E and TGEV (Herold *et al.*, 1993 and Eleouet *et al.*, 1995). In addition, two strains of MHV, JHM and A59 and also IBV have also been completely sequenced. A number of conserved functional domains have been identified in the predicted ORF 1 products, but these domains are mainly located in the ORF 1b region and leave large regions of the 1a product with no known function and only a low level of sequence conservation between different coronavirus genomes.

Wild type (wt) and cell culture adapted (ca) PEDV exhibit remarkably different phenotypes with regard to virulence in piglets (Bernasconi et al., 1995). It was therefore of interest to compare the sequences of the two virus types at the nucleotide and predicted amino acid levels.

3. MATERIALS AND METHODS

3.1. Growth of PEDV and Preparation of Viral DNA

Growth of ca PEDV was performed essentially as has been described previously (Bridgen *et al.*, 1993), except that virus-infected cells were harvested at approximately 18 h pi. Cells were freeze-thawed three times and cell debris removed by low speed centrifugation. Virus was pelleted by centrifugation for 2 h at 22,000 rpm and 4°C in a SW28 rotor. Virus pellets were resuspended in 0.5 ml Trizol™, and RNA was prepared as recommended by the manufacturer. Pre-existing cDNA clones (Tobler and Ackermann, 1995) derived from RNA of gut tissue of pigs infected with either Belgian (Bernasconi et al., 1995) or Spanish (Carvajal et al., 1995) strains of PEDV were used for the determination wt PEDV sequences.

3.2. cDNA Synthesis and PCR Amplification of Viral Sequences

RNA prepared from two 175 cm^2 flasks of virus-infected cells was denatured for 10 min at 65°C and first strand cDNA was prepared in a 50 µl total reaction volume for 70 min at 42°C using Superscript™ (Gibco-BRL), with conditions as recommended by the manufacturer. Each 50 µl reaction contained up to six anti-genome sense primers (Schmidheini AG) based on PEDV sequences or conserved regions of the HCV229E and TGEV genomes (see Table 1). Samples were treated with RNaseH for 30 min at 37°C prior to PCR amplification in a Perkin Elmer DNA Thermal Cycler 480. Amplifications were performed using pfu polymerase (Stratagene) in a 30 µl reaction volume containing 0.75 µM each primer. Reaction conditions varied with the primer pair but were typically one cycle of 95°C for 5 min, 50°C for 5 min, 74°C for 5 min followed by thirty five cycles of 95°C for 50 sec, 50°C for 50 sec, 74°C for 2 min.

3.3. Cloning of PCR Fragments

PCR fragments were phenol/chloroform extracted and ethanol precipitated prior to phosphorylation for 30 min at 37°C using T4 polynucleotide kinase in the presence of

Table 1. Primers used for RT-PCR amplification of the ORF 1a (1st column) and ORF 1b regions (2nd column) of the PEDV genome

Primer[a]	Sequence[b]	Primer[c]	Sequence[d]
o199	ggaacatttgttattgacatg	o203	tatgatgattctttctgatga
o200	ccatatggttgtgccatttt	o204	tgctgtgaacaaaattcatg
o205	gacaagttttacaagcgcatt	o207	tctcgtaatggcaacgttaa
o216	aggtttkccatcrgcrccaca	o208	tttgtcttgtgggattatgaa
o217	taaaaccggtccaatcwccaac	o209	tcttcttccattgtgctgcga
o238	cattwgwsagcatytttacgcagtt	o210	tgtcccamccgccataaaactt
o246	gagccacttgctactatcta	o211	catactattcttatatgatgcct
o258	gmgttwatccarcarttrttrtc	o226	caccaaatatatcactcttaac
o261	gatkttattgtwaatgctgcwaa	o239	gttgtmgtkgatgaggtstctatg
o263	ggtaagttytgyaaraarca	o240	ttrttttggaaytgtaatgtrga
o266	ctgcctgaataatacttgtact	o251	aaccaattcattcctgtgcac
o267	agctgtccwattgtwgtwgg	o284	tgacctacagcgagtatcaaa
o269	tgagttacaattcaactctaca		
o272	gcagctatgtayaargargc	o214[e]	gcaaccagaagtaaattaaagac
o286	ccaaacgtctcaacataccaaa	o117[e]	gttagctctttttctagacc
o287	aaataaacacctgctgggatacc	o241[e]	ggatagttagctctttttctag

[a]Designation and [b]nucleotide sequence of primers used for the amplification of the ORF 1a region.
[c,d]Primers designated for the amplification of the ORF 1b region.
[e]Primer o214 is in the S gene and primers o117 and o241 in the leader sequence.
Abbreviations are standard, ie w=a/t; s=c/g; r=a/g; y=c/t; m=a/c ; k=g/t.

500 μM ATP. Blunt-ended vector was prepared by digestion of pBS KS+ with *Eco* RV, dephosphorylation and phenol extraction to remove the alkaline phosphatase. Both PCR fragment and vector were electrophoresed through a 1% agarose gel, excised and purified by gene clean.

4. RESULTS

4.1. RT-PCR Cloning of the PEDV ORF 1

In order to clone and determine the sequences for the PEDV ORF 1, the predicted amino acid sequences of the HCV229E and TGEV polymerase open reading frames were aligned and homologous regions identified. The HCV229E and TGEV ORFs were sufficiently closely related to allow complete alignment of the predicted expression products. In contrast, the MHV and IBV sequences were much more divergent, and could only be aligned with the group 1 sequences in some of the conserved regions. Conservation between the HCV229E and TGEV sequences was much higher in the pol 1b than the pol 1a ORF, in which considerable regions of the predicted polypeptides showed under 30% homology.

A mixture of degenerate and non-degenerate primers were designed from amino acids conserved between the predicted products of the HCV229E and TGEV and, where possible, the MHV and IBV ORF 1. These primers were used both to prime reverse transcription and for the PCR amplification. Initially, 3 primer pairs yielded products. Two of them mapped to the pol 1b region and one to pol 1a. Sequencing of the amplification products allowed the design of PEDV-specific primers and confirmed the expected homology between the PEDV and the HCV229E and TGEV sequences. Continued use of this strategy using the primers described in Table 1 eventually led to RT-PCR amplification and cloning of almost the entire ORF 1 of PEDV (Figure 1).

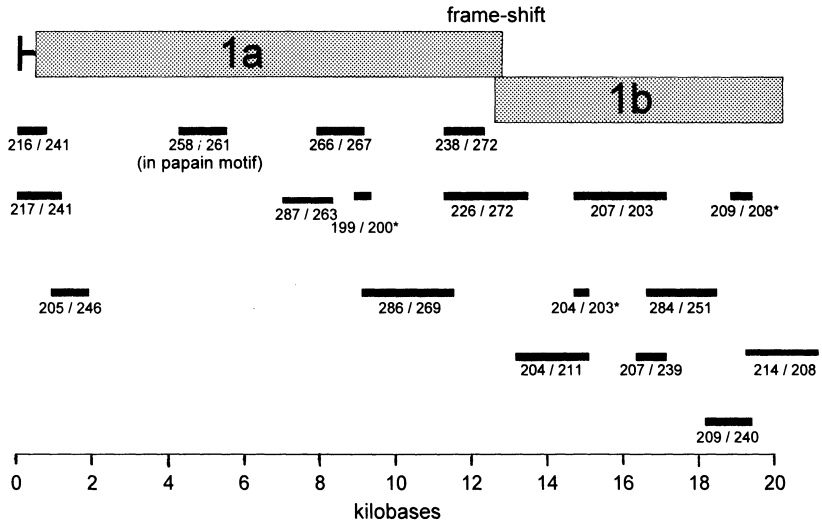

Figure 1. RT-PCR amplification of the PEDV ORF 1 using primers designed from conserved coronavirus sequences. The figure is drawn to scale, with the exception of the frame shift region and 5' leader sequence. PCR products are numbered according to the primers used for the amplification. Stars indicate the first products obtained by RT-PCR.

Analysis of the PEDV sequences continued to confirm the close relationship of PEDV with both TGEV and HCV229E. Similar results were obtained when the PEDV sequence was compared with both HCV229E and TGEV: a homology of approximately 70% and 25–30% at the amino acid level was determined for the pol 1b and pol 1a products, respectively. However the PEDV polymerase product showed a consistent slightly greater homology to the corresponding HCV229E than the TGEV product.

4.2. Sequence Comparisons of Wild Type and Cell Culture Adapted PEDV

Comparison of the two wt and CV777 ca PEDV genomes remarkably revealed no differences in either the coding or the non-coding sequences between the 3' end of the ORF 3 gene and the 3' end of the genome. Differences, however, between wt and ca strains were found within ORF 3 and within the S gene.

The sequence variations and truncations which were observed in ORF 3 of both a ca Belgian and a ca British strain (Duarte *et al.*, 1994) were not detected in wt isolates of either Belgian or Spanish origin. In contrast, the maximum possible length of the ORF 3, which had never been detected in a ca isolate, was the only form of ORF 3 which could be detected in the wt isolates. This observation suggested that an intact ORF 3 was maintained in wt viruses, whereas it seemed to be incompatible with replication of PEDV in cell culture.

The sequence of the S gene was determined only from one strain, CV777, the wt parent virus of our Belgium ca PEDV. The sequence comparison was made to the se-

Table 2. Five nucleotide changes within the S gene cause five differences between the predicted amino acid sequences of wild type and cell culture adapted PEDV

Nucleotide	From[a]	To[b]	Amino acid	From[a]	To[b]	Remarks
347	T	C	116	I (aliphatic)	T (hydrophilic)	creates new potential glycosylation site in ca PEDV
1168	C	T	390	P (imino acid)	S (hydrophilic)	
1264	A	T	422	N (hydrophilic)	Y (aromatic)	destroys potential glycosylation site in ca PEDV
3224	C	T	1075	S (hydrophilic)	L (hydrophobic)	S at this position is typical of group 1 coronaviruses
4123	C	G	1375	Q (neutral)	E (acidic)	

[a]Sequence in wt PEDV
[b]Sequence in ca PEDV

quence of the British ca PEDV isolate, which had been published previously (Duarte and Laude, 1995). Most interestingly only 5 nucleotide changes were observed within the coding sequences for S. All of those changes, however, were meaningful and led to changes in the predicted amino acid sequence. The results are summarised in Table 2. The overall number of 29 potential glycosylation sites remained unchanged. One of the changes in the amino acid sequence destroyed such a site, another change created a new one at a different location. A potential cleavage site for S1-S2 cleavage, which has been described for other coronaviruses, was not detected in either the wt or the ca sequences.

5. DISCUSSION

The technique of RT-PCR cloning the PEDV ORF 1 using primers based on conserved coronavirus sequences has proved successful, despite the lack of conserved sequences in the ORF 1a, and we would therefore recommend this technique as a means of cloning other group 1 ORF 1 sequences.

The sequence results obtained for the ORF 1 of PEDV revealed the presence of conserved domains typical of all the coronaviruses sequenced to date, such as the ribosomal frame shift site and polymerase and helicase domains. The 5' end of the genome was successfully amplified using a primer based on the leader sequence of PEDV (Tobler and Ackermann, 1995), thus confirming the relationship between these two sequences. In addition, the sequence data confirmed the homology of PEDV to HCV229E and TGEV.

The comparison of the ca and wt PEDV sequences has identified changes in both ORF 3 and the S gene between these isolates. However, in the absence of a reverse genetics system to study coronavirus genomes, completion of the entire sequence of both the wt and ca isolates will be necessary before we know which changes are responsible for the different biological properties of the isolates.

ACKNOWLEDGMENTS

These studies were supported by the Swiss National Science Foundation, grant #31–43503.95.

REFERENCES

Bernasconi C., Guscetti F., Utiger A., Van Reeth K., Ackermann M., and Pospischil A., 1995, Experimental infection of gnotobiotic piglets with a cell culture adapted porcine epidemic virus: clinical, histopathological and immunohistochemical findings, in: *Immunobiology of viral infections* (M. Schwyzer, M. Ackermann, G. Bertoni, R. Kocherhans, K. McCullough, M. Engels, R. Wittek, and R. Zanoni, eds.), Fondation Marcel Merieux, Lyon, France, pp 542–546.

Bridgen A., Duarte M., Tobler K., Laude H. and Ackermann M., 1993, Sequence determination of the nucleocapsid protein gene of the porcine epidemic diarrhoea virus confirms that this virus is a coronavirus related to human coronavirus 229E and porcine transmissible gastroenteritis virus, *J. Gen. Virol.* **74**: 1795–1804.

Bridgen A., Tobler K. and Ackermann, M., 1994, Identification of coronaviral conserved sequences and application to viral genome amplification. In: Coronaviruses: Molecular Biology and Virus-Host Interactions (H. Laude and J-F. Vautherot, eds.), *Adv. Exp. Med. Biol.* **342**: 81–82.

Carvajal A., Lanza I., de Arriba M.L., Rubio P., del Pozo M., Ackermann M., and Carmenes P., 1995, Preliminary characterisation of a Spanish isolate of porcine epidemic diarrhoea coronavirus. IVth National Virology Congress, Madrid, 21–23 September 1995.

Duarte M. and Laude H., 1994, Sequence of the spike protein of the porcine epidemic diarrhoea virus, *J. Gen. Virol.* **75**: 1195–1200.

Duarte M., Tobler K., Bridgen A., Rasschaert D., Ackermann M. and Laude H., 1994, Sequence analysis of the porcine epidemic diarrhea virus genome between the nucleocapsid and spike protein genes reveals a polymorphic ORF, *Virology* **198**: 466–476.

Eleouet J.F., Rasschaert D., Lambert P., Levy L., Vende P. and Laude H., 1995, Complete sequence (20 kilobases) of the polyprotein-encoding gene 1 of transmissible gastroenteritis virus, *Virology* **206**: 817–822.

Herold J., Raabe T., Schelle-Prinz B. and Siddell S.G., 1993, Nucleotide sequence of the human coronavirus 229E RNA polymerase locus, *Virology* **195**: 680–691.

Tobler K. and Ackermann M., 1995, PEDV leader and junction sites. In: Corona- and Related Viruses (P.J. Talbot and G.A. Levy, eds.), *Adv. Exp. Med. Biol.* **380**: 541–542.