A fast and robust method for full genome sequencing of Porcine Reproductive and Respiratory Syndrome Virus (PRRSV) Type 1 and Type 2

Kvisgaard, Lise Kirstine; Hjulsager, Charlotte Kristiane; Fahnøe, Ulrik; Breum, Solvej Østergaard; Ait-Ali, Tahar; Larsen, Lars Erik

Published in:
Journal of Virological Methods

Link to article, DOI:
10.1016/j.jviromet.2013.07.019

Publication date:
2013

Document Version
Publisher's PDF, also known as Version of record

Journal of Virological Methods 193 (2013) 697–705

Contents lists available at ScienceDirect

Journal of Virological Methods

journal homepage: www.elsevier.com/locate/jviromet

A fast and robust method for full genome sequencing of Porcine Reproductive and Respiratory Syndrome Virus (PRRSV) Type 1 and Type 2☆

Lise K. Kvisgaard[a,*], Charlotte K. Hjulsager[a], Ulrik Fahnøe[b], Solvej Ø. Breum[a], Tahar Ait-Ali[c], Lars E. Larsen[a]

[a] National Veterinary Institute, Technical University of Denmark, Frederiksberg C, Denmark
[b] National Veterinary Institute, Technical University of Denmark, Lindholm, Denmark
[c] The Roslin Institute and Royal (Dick) School of Veterinary Studies, The University of Edinburgh, UK

Article history:
Received 15 April 2013
Received in revised form 3 July 2013
Accepted 12 July 2013
Available online 24 July 2013

Keywords:
PRRSV
Full genome sequencing
Next generation sequencing
Illumina HiSeq2000
Roche 454 FLX
Ion Torrent PGM sequencer

A B S T R A C T

PRRSV is a positive-sense RNA virus with a high degree of genetic variability among isolates. For diagnostic sensitivity and vaccine design it is essential to monitor PRRSV genetic diversity. However, to date only a few full genome sequences of PRRSV isolates have been made publicly available. In the present study, fast and robust methods for long range RT-PCR amplification and subsequent next generation sequencing (NGS) were developed and validated on nine Type 1 and nine Type 2 PRRSV viruses. The methods generated robust and reliable sequences both on primary material and cell culture adapted viruses and the protocols performed well on all three NGS platforms tested (Roche 454 FLX, Illumina HiSeq2000, and Ion Torrent PGM™ Sequencer). These methods will greatly facilitate the generation of more full genome PRRSV sequences globally.

1. Introduction

Porcine Reproductive and Respiratory Syndrome (PRRS), a potent viral disease of pigs, has a major impact on the health and welfare of pigs throughout the world. PRRSV causes reproductive failures in sows and gilts and respiratory diseases in growing pigs (Collins et al., 1992; Terpstra et al., 1991; Wensvoort et al., 1991). Different PRRSV strains have been claimed to differ in virulence, but the underlining molecular determinants are still unknown despite decades of research.

PRRS virus belongs to the Arteriviridae family, within the order Nidovirales (Cavanagh, 1997). It is a small enveloped virus containing a positive-sense single-stranded RNA genome with a 5′ cap and a polyadenylated 3′-end (Benfield et al., 1992; Meulenberg et al., 1998). The genome is 15–15.5 kb in length and encodes 10 ORFs including the recently discovered ORF5a (Firth et al., 2011; Johnson et al., 2011; Snijder and Meulenberg, 1998; Wu et al., 2001). Based on nucleotide sequence comparison of European and North American isolates, PRRSV is divided into two genotypes, Type 1 and Type 2 respectively, with only 50–60% nucleotide identity (Allende et al., 1999). Furthermore, Type 1 PRRSV genotype has been found to be more diverse, and can be categorized into at least four subtypes, where the protogenotype 1, the Lelystad virus (Genbank: M96262), belongs to subtype 1 (Stadejek et al., 2002, 2006, 2008).

Most of the molecular analyses of the PRRSV genome have been performed on the genes coding for the structural proteins (ORF2-7) only. Of the seven structural proteins, glycoprotein 5 (GP5) encoded by ORF5 and the nucleocapsid (N) protein encoded by ORF7 have been the most examined. GP5 has been shown to contain a B-cell neutralizing epitope and is regarded as the most immunogenic viral protein, whereas the N protein has shown size polymorphism (Ansari et al., 2006; Oleksiewicz et al., 2002; Ostrowski et al., 2002; Stadejek et al., 2002). Previous reports suggested that the length of ORF7 could be used as a subtype maker since its length varied between Type 1 strains originating from Eastern Europe (Stadejek et al., 2002, 2006, 2008). However, recently a PRRSV isolate from Slovakia encoded a 132 amino acid N protein (compared to 128

* Corresponding author at: National Veterinary Institute, Technical University of Denmark, Bülowsvej 27, DK-1870 Frederiksberg C, Denmark. Tel.: +45 35886606; fax: +45 35886340.
E-mail addresses: likik@vet.dtu.dk (L.K. Kvisgaard), ckhj@vet.dtu.dk (C.K. Hjulsager), ulfah@vet.dtu.dk (U. Fahnøe), sbre@vet.dtu.dk (S.Ø. Breum), Tahar.aitali@roslin.ed.ac.uk (T. Ait-Ali), lael@vet.dtu.dk (L.E. Larsen).

0166-0934/$ – see front matter © 2013 The Authors. Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jviromet.2013.07.019

Table 1
Overview of sample origin and material.

Virus sample	Year of isolation	Genotype	Material	Origin
DK-1992-PRRS-111_92	1992	Type 1	PAM isolate	Denmark
DK-2010-10-10-3	2010	Type 1	PAM isolate	Denmark
DK-2011-05-11-14	2011	Type 1	PAM isolate	Denmark
DK-2011-05-23-9	2011	Type 1	PAM isolate	Denmark
ESP-1991-Olot91	1991	Type 1	Marc-145 isolate	Spain
DK-2003-6-5	2003	Type 1	PAM isolate	Denmark
DK-2003-7-2	2003	Type 1	PAM isolate	Denmark
DK-2008-10-5-2	2008	Type 1	Lung homogenate	Denmark
DK-2012-01-05-2	2012	Type 1	Serum	Denmark
DK-2010-10-1-2	2010	Type 2	Lung homogenate	Denmark
DK-2003-2-3	2003	Type 2	Marc-145 isolate	Denmark
DK-2004-2-1	2004	Type 2	Marc-145 isolate	Denmark
DK-2008-10-1-3	2008	Type 2	Lung homogenate	Denmark
DK-2012-01-11-3	2012	Type 2	Lung homogenate	Denmark
DK-2010-10-4-1	2010	Type 2	Lung homogenate	Denmark
DK-1997-19407B	1997	Type 2	Marc-145 isolate	Denmark
DK-2004-1-7-PI	2004	Type 2	Marc-145 isolate	Denmark
DK-2011-030311-1	2011	Type 2	Lung homogenate	Denmark

amino acids N protein in the Lelystad virus), but otherwise resembled the subtype 1 PRRSV (Jackova et al., 2012). These findings emphasized that the diversity of PRRSV are more complex than initially anticipated and that examination of small genetic regions is not sufficient for understanding PRRSV sequence heterogeneity. This is further supported by the emergence of a highly pathogenic strain of PRRSV in China in 2006, which was highly similar to other Type 2 PRRSV strains in ORF5 and ORF7, but had indeed unique feature differences in other genes (An et al., 2010; Tian et al., 2007). Until now, only few full genome sequences of European PRRSV isolates have been characterized (Darwich et al., 2011) and similarly little is known on the diversity of European Type 2 strains circulating in Europe. To accelerate the generation of full genome sequences of PRRSV, application of next generation sequencing (NGS) technology is the method of choice. In recent years the cost of using this technology has been reduced substantially and hereby making it more accessible (Glenn, 2011). The aim of the present study was to develop and validate a method for robust and reliable full length sequencing of PRRSV Type 1 and 2 based on long range PCR performed on viral RNA extracted from both primary material and cell culture isolates.

2. Material and methods

2.1. Sample material

Samples included in this study were collected in Danish pig farms during the years 1992–2012. Details on the viruses and their origin are listed in Table 1. Some of the viruses were proliferated in Marc-145 cells or porcine alveolar macrophages (PAM) following the general procedure (Kim et al., 1993; Wensvoort et al., 1991). The ESP-1991-Olot91 "third passage MA-104" isolate was obtained from Dr. Luis Enjuanes (Department of Molecular and Cell Biology, CSIC, Madrid, Spain). One further passage in Marc-145 was carried out prior to sequencing.

2.2. Viral RNA purification

Viral RNA was purified from serum, lung tissues or cell culture supernatants. Lung tissue was prepared as a 5% homogenate in RLT buffer (RNeasy® Mini Kit, QIAGEN, cat. no. 74106) with 1% β-mercaptoethanol (Sigma–Aldrich, cat. no. M3148). One steel bead (QIAGEN, cat. no. 69989) was added to each sample and the samples were homogenized in a Tissuelyser II (QIAGEN, cat. no. 85300) for 3 min at 30 Hz. The homogenate was centrifuged for 3 min at 12,000 × *g* 600 μl lung tissue homogenate was mixed with an equal volume of 70% ethanol by pipetting and then transferred to an RNeasy® Mini column. The rest of the purification procedure was performed as described by the manufacturer's instructions for purification of total RNA from animal tissues with RNeasy Mini Kit (QIAGEN). RNA from serum and cell culture supernatant was purified using QIAamp® Viral RNA Mini Kit (QIAGEN, cat. no. 52906) using the conditions recommended by the manufacturer. The RNA was stored at −80 °C until use.

2.3. Screening for PRRSV with real-time RT-PCR

PRRSV positive samples we selected for sequencing by screening of purified RNA using a modification of the Primer Probe Energy Transfer RT-PCR (PriProET-RT-PCR) OneStep real-time RT-PCR assay previously described (Balka et al., 2009).

2.4. cDNA synthesis

Full-length cDNA synthesis was performed with SuperScript® III First-Strand Synthesis System (Invitrogen, cat. no. 18080-51) using 3′-end poly(dT) reverse transcription (RT)-primer: 5′-CAG GAA ACA GCT ATG ACA CCT GAT CTC TAG AAA CGT $T(T)_{38}$-3′ (Nielsen et al., 2003). The method for the synthesis of full-length cDNA was as previously described for pestiviruses (Rasmussen et al., 2008, 2010). In short, 8 μl of purified RNA (approx. 0.7 μg) was mixed with 1 μl of 10 μM RT-primer and 1 μl of 10 nM dNTPs. The mixture was incubated for 5 min at 65 °C in a thermal cycler with heated lid and then placed on ice for at least 1 min. A second mixture containing 2 μl 10× RT buffer, 4 μl 25 mM $MgCl_2$, 2 μl 0.1 M DTT, 1 μl 40 U/μl RNaseOUT™, and1 μl SuperScript III RT™ was prepared and mixed gently with the RNA/primer mixture. The cDNA synthesis was performed in a thermal cycler with heated lid at 50 °C for 90 min followed by incubation for 5 min at 85 °C to inactivate the reverse transcriptase. To collect the samples a brief centrifugation step was applied and the cDNA was treated with 1 μl of 2 U/μl *Escherichia coli* RNase H for 20 min at 37 °C to remove the RNA template. The cDNA was stored at −20 °C until use.

2.5. Primers for long range PCR amplification

All the primer sequences used for long range PCR amplification are listed in Table 2. Each primer sequence was aligned against the sequence of the two PRRSV proto-genotypes, Lelystad virus (M96262) or VR2332 (U87392.3) using CLC Main Workbench v.6.6.2 (CLC BIO, Aarhus, Denmark). The primer sequences were adjusted to match the sequence of the two proto-genotype strains.

Table 2
Primer sequences for long range PCR amplification and expected amplicon sizes.

Primer name	Primer sequence (5′–3′)	Expected size (bp)
Type 1 PRRSV		
Fragment-1-EU		7278
5′-UTR-1-35-EU-Fw[a]	GGC GCG CCT AAT ACG ACT CAC TAT AGA TGA TGT GTA GGG TAT TCC CCC TAC ATA CAC GAC A	
ORF1-7278-EU-Rev[b]	CAG CTT CAA GGC AGT TGT CA	
Fragment-2-EU		8178
ORF1-6920-EU-Fw[c]	CCC CTC TTT TTG AGA ATG GT	
Poly(dT)-RT[d]	CAG GAA ACA GCT ATG ACA CCT GAT CTC TAG AAA CGT T(T)38	
Fragment A-EU		3019
5′-UTR-1-35-EU-Fw[a]	GGC GCG CCT AAT ACG ACT CAC TAT AGA TGA TGT GTA GGG TAT TCC CCC TAC ATA CAC GAC	
ORF1-3019-EU-Rev[a]	CGC GGG CGC TTG AGT TCG GCA AAT T	
Fragment B-EU		4529
ORF1-2749-EU-Fw[e]	CCT GGA CCA GCC TTT AAA TC	
ORF1-7278-EU-Rev[b]	CAG CTT CAA GGC AGT TGT CA	
Fragment C-EU		5500
ORF1-6920-EU-Fw[c]	CCC CTC TTT TTG AGA ATG GT	
ORF2-EU-Rev[f]	GCA CAC TGA TGA GCC ATT GT	
Fragment D-EU		
ORF2-EU-Fw[f]	CTG GCA CAG AAT TGC AGG TA	3375
Poly(dT)-RT[d]	CAG GAA ACA GCT ATG ACA CCT GAT CTC TAG AAA CGT T(T)38	
Type 2 PRRSV		
Fragment-1-US		
Fragment-A-US-Fw[g]	GGA GGG CCA AGT CTA CTG CAC ACG A	7588
Fragment-B-US-Rev[d]	TGG TTG TGC TCA ACC GCG T	
Fragment-2-US:		7939
Fragment-C-US-Fw[d]	TCT CAG AGT TGG CGA CCC T	
Poly(dT)-RT[d]	CAG GAA ACA GCT ATG ACA CCT GAT CTC TAG AAA CGT T(T)38	
Fragment-A-US		4541/4811
Fragment-A-1-35-US-Fw	CTCGAGGGCGCGCCTAATACGACTCACTATAGGATGA CGTATAGGTGTTGGCTCTATGCCTTGGCATT	
Fragment-A-US-Fw[g]	GGA GGG CCA AGT CTA CTG CAC ACG A	
Fragment-A-US-Rev[d]	GTG TCA GGG TCA ACC ACG A	
Fragment-B-US		3506
Fragment-B-US-Fw[d]	ATC TTG GCT GGA GCT TAC GT	
Fragment-B-US-Rev[d]	TGG TTG TGC TCA ACC GCG T	
Fragment-C-US		5500
Fragment-C-US-Fw[d]	TCT CAG AGT TGG CGA CCC T	
Fragment-C-US-Rev[d]	ATC CTG CAC CAA AGA GAC CT	
Fragment-D-US		2900
Fragment-D-US-Fw[d]	TTT CAG CAT CTA GCC GCC A	
Poly(dT)-RT[d]	CAG GAA ACA GCT ATG ACA CCT GAT CTC TAG AAA CGT T(T)38	

Fw, forward primer, Rev, reverse primer.
[a] Modified from Fang et al. (2006).
[b] Fang et al. (2006).
[c] Darwich et al. (2011).
[d] Nielsen et al. (2003).
[e] Darwich et al. (2011).
[f] Diaz et al. (2006).
[g] Nielsen et al. (2003).

2.6. Long range PCR amplification for sequencing

PCR amplifications were performed with AccuPrime™ *Taq* DNA Polymerase High Fidelity (Invitrogen, Cat. no. 12346-086). For amplification of PCR products covering the PRRSV genome in 4 fragments, 2 μl of cDNA template was mixed with 5 μl 10× AccuPrime™ PCR buffer I, 1 μl 10 μM specific forward primer (Table 2), 1 μl 10 μM specific reverse primer (Table 2), 0.5 μl AccuPrime™ *Taq* DNA Polymerase High Fidelity, and nuclease free water (AMRESCO®, cat. no. E476) in a final reaction volume of 50 μl. For amplification of PCR products covering the PRRSV genome in 2 fragments, 1 μl of AccuPrime™ *Taq* DNA Polymerase High Fidelity was used. The PCR amplification was carried out in a T3 Thermocycler (Biometra) under the following conditions: 94 °C 15 s [45 cycles: 94 °C for 15 s, 55 °C for 30 s, and 68 °C for 60 s per kb amplification], 68 °C for 2 min per kb amplification, hold at 4 °C. The PCR products were analyzed by agarose gel electrophoresis using E-gel® 0.8% agarose gels (Invitrogen, cat. no G501808). Five microliters of PCR product mixed with 7 μl 1:40 TrackIt™ loading buffer (Invitrogen, cat no. 10482-028) were loaded onto the gel. 10 μl 1 kb Plus DNA ladder as marker (Invitrogen, cat. no. 10787-018) was used as size marker. To remove primers and buffers, the PCR products were purified using Roche's High Pure PCR Product Purification kit (cat. no. 11 732 676 001) following the manufacturer's protocol, with the exception that elution was done in 30 μl nuclease free water (Amresco, cat. no. E476).

2.7. Full genome sequencing

Three different platforms were used for next generation sequencing. For the sequencing of PCR amplicons covering the full genome of PRRSV in two fragments, 2.7 μg in total of equimolar

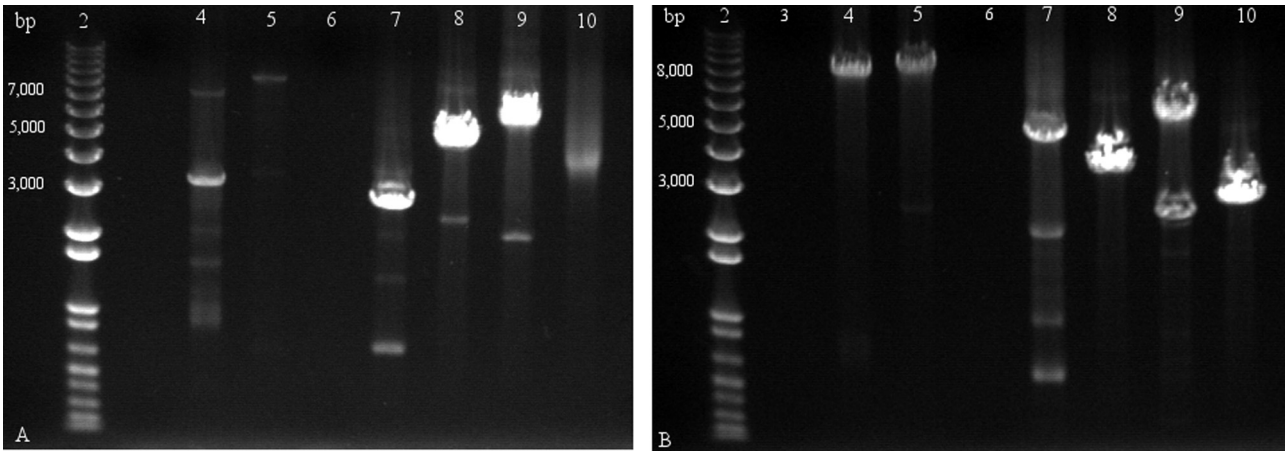

Fig. 1. 0.8% agarose gel electrophoreses analysis of long range PCR amplicons. (A) PCR amplicons for Type 1 PRRSV. Lanes 4 and 5: Frag-1-EU and Frag-2-EU, lanes 7–10: Frag-A-EU, Frag-B-EU, Frag-C-EU, and Frag-D-EU. (B) PCR amplicons for Type 2 PRRSV. Lanes 4 and 5: Frag-1-US and Frag-2-US, lanes 7–10: Frag-A-US, Frag-B-US, Frag-C-US, and Frag-D-US.

concentration was sent to LGC Genomics GmbH (Berlin, Germany) for sequencing on the platform Roche 454 FLX. The two samples were sequenced at separate occasions on a 1/16 Roche/454 Pico Titer Plate pooled together with 10 and 11 other libraries prepared from PCR amplicons of another virus with approximately the same size. The sequencing of PCR amplicons covering the PRRSV genome in four fragments was performed on the Illumina® HiSeq2000 and Ion Torrent PGM™ sequencer platforms. The four PCR products representing one PRRSV genome were pooled in equimolar quantities to a final amount of 1 μg. The sequencing on the platform Ion Torrent PGM™ sequencer was carried out by the DTU in-house facility (DTU Multi-Assay Core (DMAC), Technical University of Denmark). An Ion Torrent PGM™ chip 316 was used for 32 libraries of approximately the same size. Sequencing on Illumina® HiSeq2000 took place at ARK Genomics (The Roslin Institute, University of Edinburgh, Scotland, UK). A total of 22 libraries of the same genome size were pooled and sequenced on one channel flow cell. The preparation of these libraries was performed following the Illumina® TruSeq™ DNA Sample Preparation v2 Kit A and B with minor exceptions (Illumina®, cat. no. FC-121-2001-1 and FC-121-2002). Instead of purifying the samples using Agencourt AMPure XP beads as stated in the manufactures protocol, column purifications was performed using MinElute PCR Purification Kit (QIAGEN, cat. no. 28004).

2.8. 5′-RACE

For Type 2 viruses amplified using the forward primer 'Fragment-A-US-FW' (Table 2), the 5′ end nucleotides were not amplified and therefore 5′ Rapid Amplification of the cDNA Ends was performed to complete the sequences. This was performed using 5′ RACE System for Rapid Amplification of cDNA Ends, version 2.0 (Invitrogen, cat. no. 18374-058) according to instructions from the manufacturer. In short, one gene-specific RT-primer was designed for the cDNA synthesis, one for the first PCR amplification and one for the final nested PCR using Primer3Plus web utility (Untergasser et al., 2007). The nested PCR products of approximately 500 bp were purified directly from the reactions using Roche's High Pure PCR Product Purification kit (Roche, cat. no. 11 732 676 001) or they were excised from E-gel 0.8% agarose gels and extracted using MinElute Gel Extraction Kit (QIAGEN, cat. no. 28604). The final RACE products were cloned into pCR®4-TOPO® vectors and transformed into *E. coli* One Shot® TOP10 competent cells following the protocol for TOPO® TA Cloning® Kit for Sequencing (Invitrogen, K4575-02). Sequences were obtained with cycle sequencing performed at LGC Genomics using M13 primers.

2.9. Sequence data analysis

After initial removal of adaptors and low quality sequences, the quality of the FastQC files was examined using the applicant FastQC (version 0.10.1). The reads were trimmed in regards to the FastQC report. Trimming was done by the Prinseq-lite tool and mapping of the reads was performed by the Burrows–Wheeler aligner (BWA) using the Aln algorithm for the Illumina® HiSeq2000 data and the bwasw algorithm for Roche 454 FLX and Ion Torrent PGM™ sequencer data. Coverage depth was calculated by a combination of Samtools and Bedtools.

The full length nucleotide sequences were aligned using MUSCLE (MUltiple Sequence Comparison by Log-Expectation). Phylogenetic trees of nucleotide sequences were constructed using Maximum Likelihood Phylogeny with the following parameter settings: Starting tree: Neighbor Joining, Substitution model: HKY, Transition/transversion ratio = 2.0, Rate variation: none, Estimate substitution rate parameter(s) = yes, Estimate topology = yes. All the phylogenetic analyses were carried out using CLC Main Workbench v.6.6.2 (CLC BIO, Aarhus, Denmark).

3. Results

3.1. Specificity of the PCR products

The long range PCR amplicons generated in this study covered the complete PRRSV genome in two or four fragments, with sequence overlap of 270–677 bp between the fragments (Table 2). To confirm the size, the purified PCR products were run on a 0.8% agarose gel (Fig. 1). Most of the PCRs produced a single amplicon with the expected size, however, in few cases an unspecific band was in excess (Fig. 1, panel A, lane 4). In other cases, the PCR product of interest appeared in equal amount as unspecific products (Fig. 1, panel B, lane 9). In those cases with ambiguous products, the PCR amplicon of interest was gel extracted prior to sequencing.

3.2. Sequence quality and trimming

To evaluate the quality of the NGS data, adaptors and low quality sequences were first removed prior to further analysis with the FastQC application. This application helped visualize the quality of the reads by examining different parameters such as

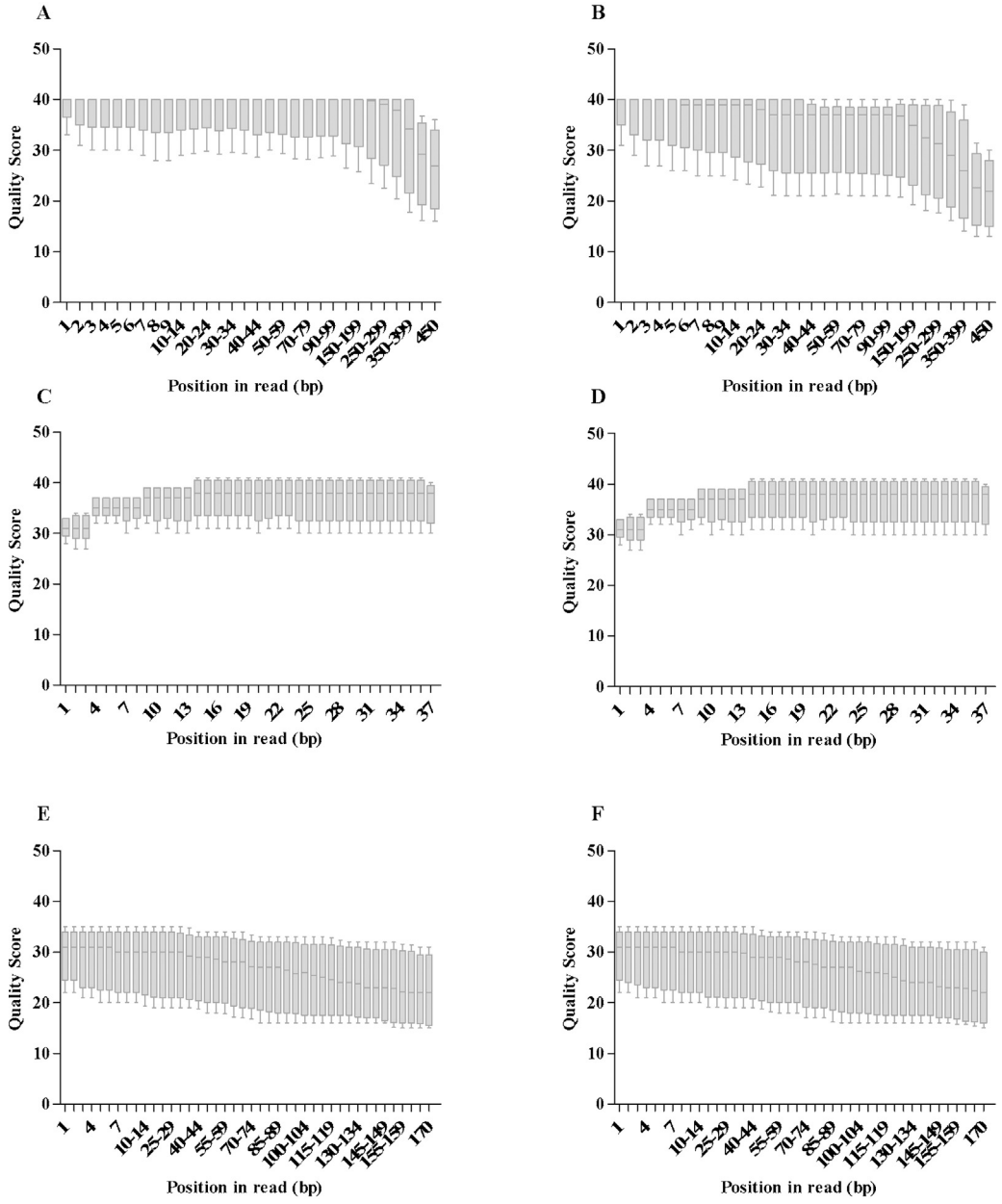

Fig. 2. Box plots of Phred quality scores generated by FastQC. The gray box plots (box: interquartile ranges 25–75%, whiskers: 10–90%, and dark bar inside box: median) show the nucleotide-calling Phred quality score across all reads. (A) DK-PRRS-111_92 (Roche 454 FLX), (B) DK-2010-10-1-2 (Roche 454 FLX), (C) DK-2011-05-11-14 (Illumina HiSeq 2000), (D) Olot91(Illumina HiSeq 2000), (E) DK-2011-10-4-1 (Ion Torrent PGM™ Sequencer), and (F) DK-2012-01-05-2 (Ion Torrent PGM™ Sequencer).

the distribution of the nucleotides in each read, read lengths, and calculated the Phred Quality Score of each nucleotide in each read. The Phred Quality Score described the probability of incorrect base calls and allowed to compare the efficacy of the different sequencing platforms. A Phred Quality Score of 30 (Q30) is equivalent to the probability of an incorrect base call in 1 out of a 1000 times, which gives the base-call accuracy of 99.9% (http://www.illumina.com/Documents/%5Cproducts%5Ctechnotes%5Ctechnote_Q-Scores.pdf). A box plot of the Phred Quality Score distribution is shown in Fig. 2 and in supplementary material Fig. 1. The FastQC report obtained for Roche 454 FLX sequencing showed a decline in 3′ end reads quality and thus the major read lengths were trimmed from 592 bp to 450 bp. The box plot of the Phred Quality Score distribution is shown in Fig. 2(A) and (B). The average Phred Quality Score of the HiSeq2000 Illumina data was Q36.5 and which was highly acceptable. Although, a small decrease in the Phred Quality Score was observed at the 5′ end of the reads (Fig. 2, panels C and D, and supplementary Fig. 1(A) and (B)), the mean *Q* values were still above Q30, thus there was no reason for further trimming of these data. The majority of the full genome sequences of PRRSV were generated using the Ion Torrent PGM sequencer. The analysis of these data by FastQC showed "warnings" in both the 5′ and 3′ ends of the reads with regard to the distribution of the four nucleotides. Thus, the reads were trimmed to a major length distribution of 170 bp to reach a minimal Phred Quality Score of Q20 for both the 5′ and 3′ ends. This strategy resulted in a mean Phred Quality Score of Q24 or Q25 (Fig. 2, panels E and F and Supplementary Fig. 1(C)–(L)), which gave an acceptable base call accuracy between 99% and 99.9%. An overview of the average Phred Quality Scores and major read lengths is listed in Table 3.

Table 3
Read quality and output yield.

Virus sample	Platform	No. of PCR fragments for sequencing	Avg. Phred Quality Score	Major read length (bp)	Total reads[a]	Mapped reads	Mapped reads %	Bases mapped
DK-PRRS-111_92	FLX	2	37	450	5679	5633	99.2	2,534,850
DK-2010-10-1-2	FLX	2	29	450	13,696	13,504	98.6	6,076,800
DK-2010-10-10-3	Illumina	4	36,5[b]	37	3,151,310[c]	2,691,058[b]	85.4	99,569,146
DK-2011-05-11-14	Illumina	4	36,5[b]	37	5,560,952[c]	5,409,041[b]	97.3	200,134,517
DK-2011-05-23-9	Illumina	4	36,5[b]	37	4,810,154[c]	1,145,373[b]	23.8	42,378,801
ESP-1992-Olot91	Illumina	4	36,5[b]	37	3,792,606[c]	3,252,661[b]	85.8	120,348,457
DK-2003-6-5	Ion Torrent	4	24	170	70,875	55,844	78.8	9,493,480
DK-2003-7-2	Ion Torrent	4	24	170	63,656	32,276	50.7	5,486,920
DK-2008-10-5-2	Ion Torrent	4	24	170	75,357	51,011	67.7	8,671,870
DK-2012-01-05-2	Ion Torrent	4	25	170	55,378	50,193	90.61	8,532,810
DK-2003-2-3	Ion Torrent	4	24	170	43,915	38,252	87.1	6,502,840
DK-2004-2-1	Ion Torrent	4	25	170	41,549	36,998	89.0	6,289,660
DK-2008-10-1-3	Ion Torrent	4	24	170	74,948	55,303	73.8	9,401,510
DK-2012-01-11-3	Ion Torrent	4	25	170	77,893	30,033	38.5	5,105,610
DK-2010-10-4-1	Ion Torrent	4	24	170	52,658	41,554	78.9	7,064,180
DK-1997-19407B	Ion Torrent	4	24	170	56,171	53,026	94.4	9,014,420
DK-2004-1-7-Pl	Ion Torrent	4	24	170	65,900	62,542	94.9	10,632,140
DK-2011-030311-1	Ion Torrent	4	24	170	85,327	66,526	78.0	11,309,420

[a] After removal of adaptor and low quality sequences and examination by FastQC.
[b] For Illumina the average Phred Quality Score is a mean value of the Pred quality score of R1 and R2.
[c] Paired end reads.

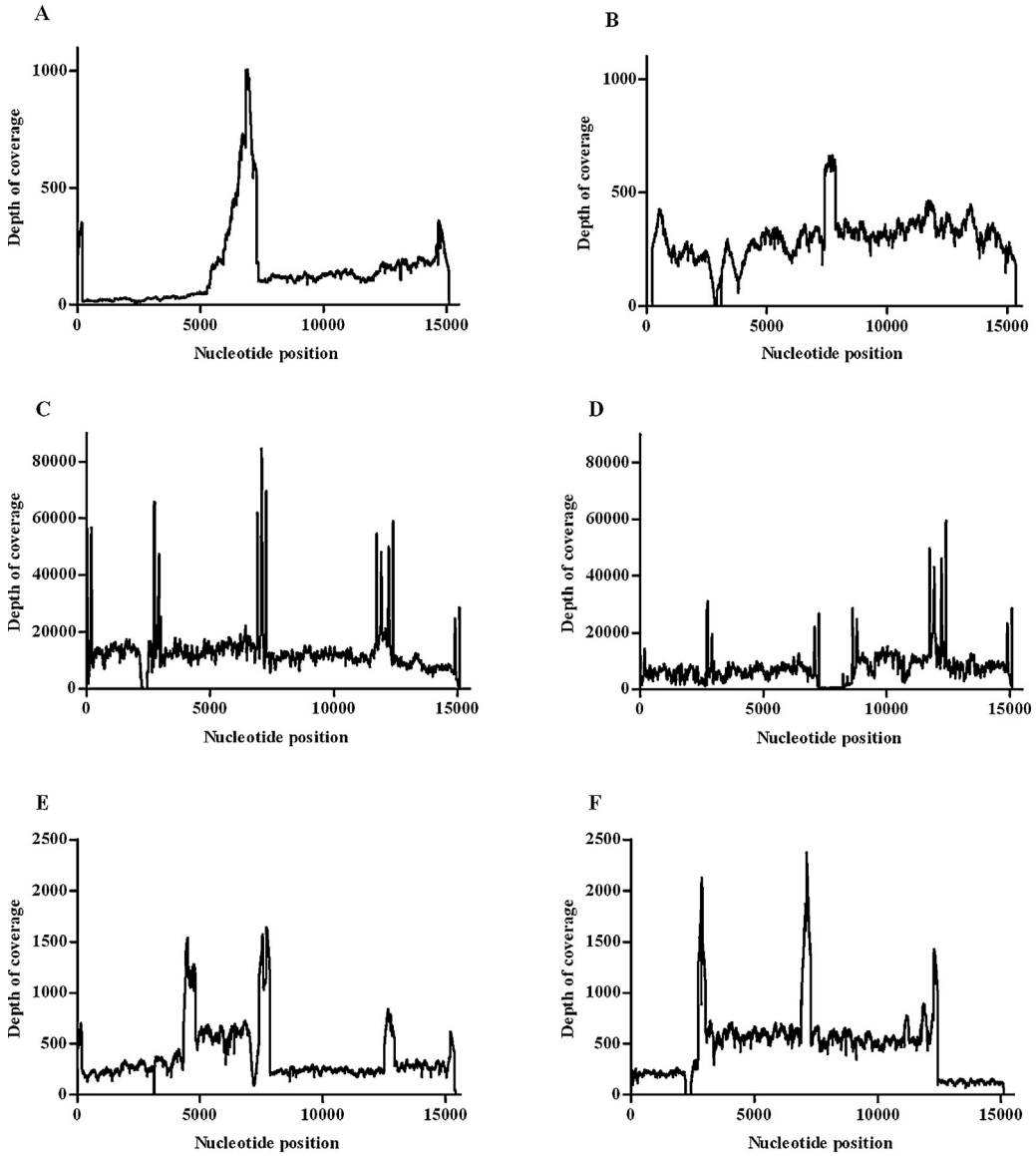

Fig. 3. Depth of coverage plot. The *x*-axes represent the coverage of every single nucleotide position in respect to the reference sequence and the *y*-axes represent the sequence depth of every single nucleotide. (A) DK-PRRS-111_92, Roche 454 FLX, (B) DK-2010-10-1-2, Roche 454 FLX, (C) DK-2011-05-11-14, Illumina HiSeq 2000, (D) Olot91, Illumina HiSeq 2000, (E) DK-2011-10-4-1, Ion Torrent PGM™ Sequencer, and (F) DK-2012-01-05-2, Ion Torrent PGM™ Sequencer.

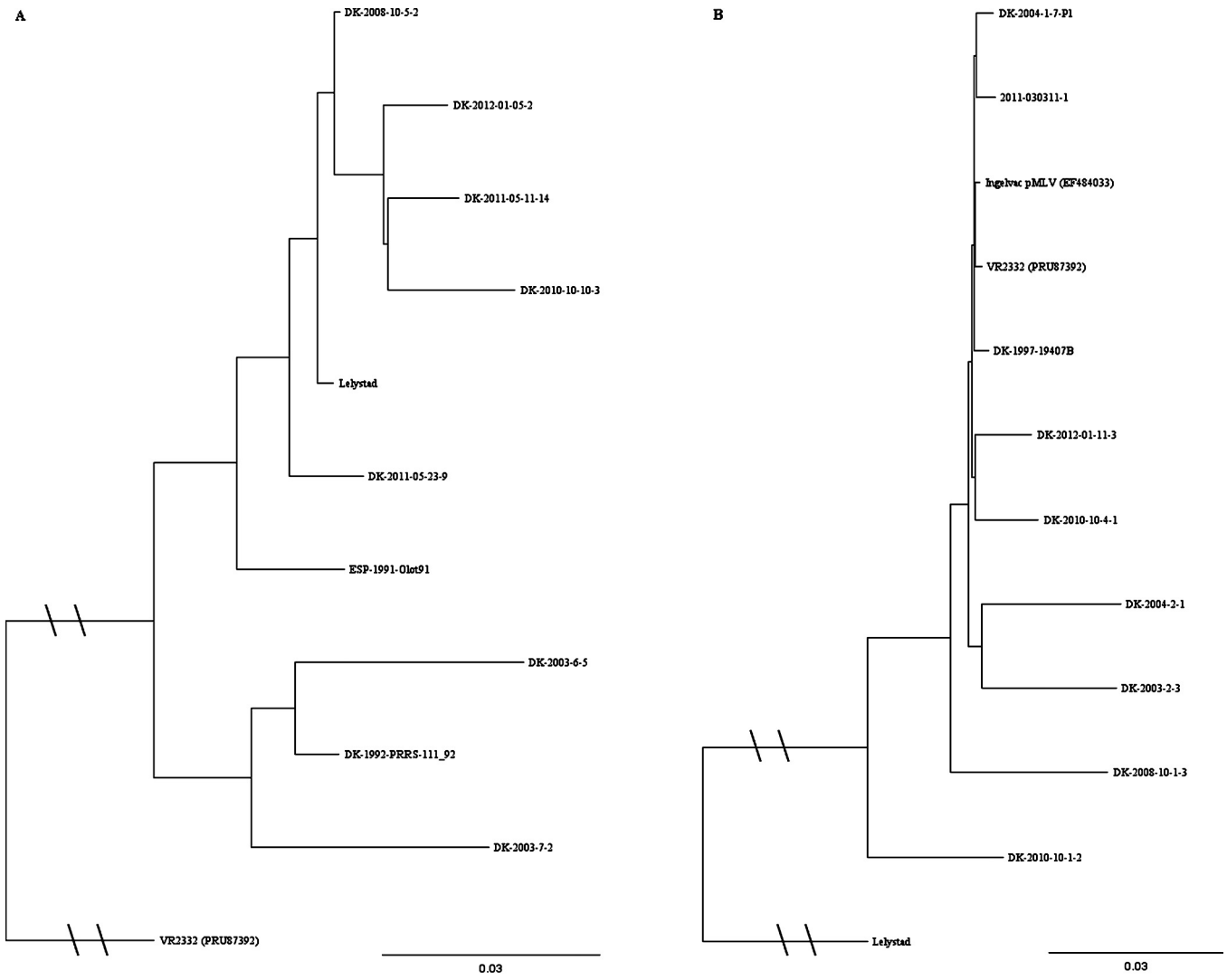

Fig. 4. Phylogenetic trees of full length genomes. (A) Type 1 PRRSV viruses and (B) Type 2 PRRSV viruses. Scale bar represents 3 nucleotide changes per hundred.

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jviromet.2013.07.019.

3.3. Sequence output and coverage

The data output varied among the different NGS platforms used, with the highest yield for the Illumina HiSeq2000 platform. In Table 3, total reads and mapped reads are listed for all samples. The amount of mapped reads in percent ranged from 23.8% to 99.2%. The outcome of mapped reads below 40% may be result of low purity of the PCR amplicons in the sample or low quality of the library preparation. The distribution of the mapped reads throughout the genomes is shown in Fig. 3 and supplementary material Fig. 2 as a depth-of-coverage graph. Peaks with a substantially larger depth of coverage are present in all graphs. These peaks correspond to the overlap in sequence of the PCR amplicons. Variations in the coverage depths were seen between the PCR amplicons from the same sample, which indicate different quality and purity of the PCR amplicons or imprecise preparation of equimolar pools. In some cases regions without coverage was observed (Fig. 3) by that indicating gaps compared to the reference sequence. The presence of gaps were either confirmed or rejected by cycle sequencing.

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jviromet.2013.07.019.

The complete full genome sequences had the lengths of 14,876–15,098 and 15,342–15,408 nucleotides, for the Type 1 and Type 2 viruses, respectively (excluding the poly(A)-tail). Phylogenetic trees of the obtained full length sequences are shown in Fig. 4. The clustering of the viruses in the trees based on the full genome sequences was as expected and resembled the clustering of the isolates when the tree was based on ORF5 sequences only (data not shown).

4. Discussion

The main focus of this study was to develop fast and robust methods for full length sequencing of PRRSV directly from infected tissues using next generation sequencing platforms. The protocols generated sequences with the expected lengths and quality as illustrated by the preliminary phylogenetic analysis which showed the expected clustering of the analyzed isolates. Further analyses of the full genome sequences are in progress.

The protocols described in the present study were based on the generation of long range PCR products using a high fidelity polymerase which has a low error-rate. This high fidelity polymerase

is a mixture of recombinant *Taq* DNA polymerase, *Pyrococcus* species *GB-D* polymerase, and Platinum *Taq* Antibody which results in a nine-fold increase in fidelity. Furthermore, this enzyme blend is effective over a wide range of target sizes up to 20 kb with some optimization (http://tools.invitrogen.com/content/sfs/manuals/accuprimetaqhifi_man.pdf). The protocol describes the sequencing of PCR amplicons covering the PRRSV genome in two and four fragments. Sequencing from two fragments has the advantage, besides the reduced labor handling only two amplicons, that only one sequence overlap will occur and less mismatch/bias from the primer sequencing will be incorporated in the PCR amplicons as only two primer pairs are used in the production of two PCR amplicons. On the other hand, the disadvantage when amplifying the PCR genome in two PCR fragments was that the concentration of the specific PCR amplicons often was of low concentration as they often needed to be extracted from the gel instead of direct purification from the PCR mix due to the present of unspecific products. The advantage of sequencing from PCR amplicons covering the PRRSV genome in four fragments was that the PCR products were of relatively high concentrations and specificity hence the amplicons could be purified directly from the PCR mix without the need for gel purification. The disadvantage of sequencing on four fragments is that two sites of overlapping sequences will occur. The sequence primer bias is not problematic when sequencing is performed to obtain a consensus sequence for phylogenetic and amino acid analyses, but should be kept in mind if the PCR products are sequenced for quasispecies and de novo analyses.

The main advantage of this method is that it enables viral RNA extracted directly from the primary material to be used as template for the cDNA synthesis. Such an approach bypasses the needs for propagation of PRRSV in cell culture which can be tedious and time-consuming. The use of primary material for sequencing furthermore prevents the introduction of mutations caused by cell culture adaptation and maintains the original level of quasispecies distribution. This new, innovative method offers a tool for a faster and quicker production of full genome PRRSV sequences compared to previous published protocol (Darwich et al., 2011; Van Doorsselaere et al., 2011).

Since only one NGS platform is available on most institutions, the applicability of the methods described here is tempting because all three tested NGS platforms were shown to generate PRRSV sequences of satisfactory quality. The selection of the optimal platform should be based on a variety of considerations including price, run time, read lengths, depth of coverage (yield data output). In the present set-up, the coverage depth was considerably higher for the Illumina HiSeq2000 compared to Roche 454 FLX and Ion Torrent PGM sequencer. Since PRRSV has a relative short genome of 15 kb, the coverage depth obtained in this study from all three NGS platforms was satisfying. The run time for Illumina HiSeq2000 was 11 days compared to the run time for Roche 454 FLX and Ion Torrent PGM sequencer of 10 and 2 h respectively. The short run times give these latter two platforms an advantage in applications where a fast result is crucial. For de novo assembly, read lengths should be a factor to consider. In this study, 37 bp read lengths were obtained from the Illumina and reads from the Roche 454 FLX and Ion Torrent PGM sequencer were trimmed to major read lengths of 450 and 170 bp, respectively. The reads were mapped using a reference sequence and therefore the short read length was not a problem.

In conclusion, the long-range PCR protocols of the present study offer a robust, cheap and fast method for the generation of full length PRRSV sequences directly from serum, tissues and cell cultures. The results of this work will be an added value for the development of new diagnostics and vaccine designs.

Acknowledgements

Dr. Anette Botner is gratefully thanked for the provision of PRRSV cell culture isolates.

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no 245141 (New tools and approaches to control Porcine Reproductive and Respiratory Syndrome in the EU and Asia (PoRRSCon) coordinated by Prof. H. Nauwynck) and the COST Action FA902: Understanding and combating porcine reproductive and respiratory syndrome in Europe (EuroPRRS.net).

References

Allende, R., Lewis, T.L., Lu, Z., Rock, D.L., Kutish, G.F., Ali, A., Doster, A.R., Osorio, F.A., 1999. North American and European porcine reproductive and respiratory syndrome viruses differ in non-structural protein coding regions. Journal of General Virology 80 (Pt 2), 307–315.

An, T.Q., Tian, Z.J., Xiao, Y., Li, R., Peng, J.M., Wei, T.C., Zhang, Y., Zhou, Y.J., Tong, G.Z., 2010. Origin of highly pathogenic porcine reproductive and respiratory syndrome virus, China. Emerging Infectious Diseases 16, 365–367.

Ansari, I.H., Kwon, B., Osorio, F.A., Pattnaik, A.K., 2006. Influence of N-linked glycosylation of porcine reproductive and respiratory syndrome virus GP5 on virus infectivity, antigenicity, and ability to induce neutralizing antibodies. Journal of Virology 80, 3994–4004.

Balka, G., Hornyak, A., Balint, A., Benyeda, Z., Rusvai, M., 2009. Development of a one-step real-time quantitative PCR assay based on primer-probe energy transfer for the detection of porcine reproductive and respiratory syndrome virus. Journal of Virological Methods 158, 41–45.

Benfield, D.A., Nelson, E., Collins, J.E., Harris, L., Goyal, S.M., Robison, D., Christianson, W.T., Morrison, R.B., Gorcyca, D., Chladek, D., 1992. Characterization of swine infertility and respiratory syndrome (SIRS) virus (isolate ATCC VR-2332). Journal of Veterinary Diagnostic Investigation 4, 127–133.

Cavanagh, D., 1997. Nidovirales: a new order comprising Coronaviridae and Arteriviridae. Archives of Virology 142, 629–633.

Collins, J.E., Benfield, D.A., Christianson, W.T., Harris, L., Hennings, J.C., Shaw, D.P., Goyal, S.M., McCullough, S., Morrison, R.B., Joo, H.S., 1992. Isolation of swine infertility and respiratory syndrome virus (isolate ATCC VR-2332) in North America and experimental reproduction of the disease in gnotobiotic pigs. Journal of Veterinary Diagnostic Investigation 4, 117–126.

Darwich, L., Gimeno, M., Sibila, M., Diaz, I., de la Torre, E., Dotti, S., Kuzemtseva, L., Martin, M., Pujols, J., Mateu, E., 2011. Genetic and immunobiological diversities of porcine reproductive and respiratory syndrome genotype I strains. Veterinary Microbiology 150, 49–62.

Diaz, I., Darwich, L., Pappaterra, G., Pujols, J., Mateu, E., 2006. Different European-type vaccines against porcine reproductive and respiratory syndrome virus have different immunological properties and confer different protection to pigs. Virology 351, 249–259.

Fang, Y., Rowland, R.R., Roof, M., Lunney, J.K., Christopher-Hennings, J., Nelson, E.A., 2006. A full-length cDNA infectious clone of North American type 1 porcine reproductive and respiratory syndrome virus: expression of green fluorescent protein in the Nsp2 region. Journal of Virology 80, 11447–11455.

Firth, A.E., Zevenhoven-Dobbe, J.C., Wills, N.M., Go, Y.Y., Balasuriya, U.B., Atkins, J.F., Snijder, E.J., Posthuma, C.C., 2011. Discovery of a small arterivirus gene that overlaps the GP5 coding sequence and is important for virus production. Journal of General Virology 92, 1097–1106.

Glenn, T.C., 2011. Field guide to next-generation DNA sequencers. Molecular Ecology Resources 11, 759–769.

Jackova, A., Vlasakova, M., Leskova, V., Vilcek, S., 2012. Identification of a new unusual length polymorphism of the nucleocapsid protein in porcine reproductive and respiratory syndrome virus. Virus Genes 45, 590–592.

Johnson, C.R., Griggs, T.F., Gnanandarajah, J., Murtaugh, M.P., 2011. Novel structural protein in porcine reproductive and respiratory syndrome virus encoded by an alternative ORF5 present in all arteriviruses. Journal of General Virology 92, 1107–1116.

Kim, H.S., Kwang, J., Yoon, I.J., Joo, H.S., Frey, M.L., 1993. Enhanced replication of porcine reproductive and respiratory syndrome (PRRS) virus in a homogeneous subpopulation of MA-104 cell line. Archives of Virology 133, 477–483.

Meulenberg, J.J., Bos-de Ruijter, J.N., van de Graaf, R., Wensvoort, G., Moormann, R.J., 1998. Infectious transcripts from cloned genome-length cDNA of porcine reproductive and respiratory syndrome virus. Journal of Virology 72, 380–387.

Nielsen, H.S., Liu, G., Nielsen, J., Oleksiewicz, M.B., Botner, A., Storgaard, T., Faaberg, K.S., 2003. Generation of an infectious clone of VR-2332, a highly virulent North American-type isolate of porcine reproductive and respiratory syndrome virus. Journal of Virology 77, 3702–3711.

Oleksiewicz, M.B., Botner, A., Normann, P., 2002. Porcine B-cells recognize epitopes that are conserved between the structural proteins of American- and

European-type porcine reproductive and respiratory syndrome virus. Journal of General Virology 83, 1407–1418.
Ostrowski, M., Galeota, J.A., Jar, A.M., Platt, K.B., Osorio, F.A., Lopez, O.J., 2002. Identification of neutralizing and nonneutralizing epitopes in the porcine reproductive and respiratory syndrome virus GP5 ectodomain. Journal of Virology 76, 4241–4250.
Rasmussen, T.B., Reimann, I., Hoffmann, B., Depner, K., Uttenthal, A., Beer, M., 2008. Direct recovery of infectious pestivirus from a full-length RT-PCR amplicon. Journal of Virological Methods 149, 330–333.
Rasmussen, T.B., Reimann, I., Uttenthal, A., Leifer, I., Depner, K., Schirrmeier, H., Beer, M., 2010. Generation of recombinant pestiviruses using a full-genome amplification strategy. Veterinary Microbiology 142, 13–17.
Snijder, E.J., Meulenberg, J.J., 1998. The molecular biology of arteriviruses. Journal of General Virology 79 (Pt 5), 961–979.
Stadejek, T., Stankevicius, A., Storgaard, T., Oleksiewicz, M.B., Belak, S., Drew, T.W., Pejsak, Z., 2002. Identification of radically different variants of porcine reproductive and respiratory syndrome virus in Eastern Europe: towards a common ancestor for European and American viruses. Journal of General Virology 83, 1861–1873.
Stadejek, T., Oleksiewicz, M.B., Potapchuk, D., Podgorska, K., 2006. Porcine reproductive and respiratory syndrome virus strains of exceptional diversity in eastern Europe support the definition of new genetic subtypes. Journal of General Virology 87, 1835–1841.
Stadejek, T., Oleksiewicz, M.B., Scherbakov, A.V., Timina, A.M., Krabbe, J.S., Chabros, K., Potapchuk, D., 2008. Definition of subtypes in the European genotype of porcine reproductive and respiratory syndrome virus: nucleocapsid characteristics and geographical distribution in Europe. Archives of Virology 153, 1479–1488.
Terpstra, C., Wensvoort, G., Pol, J.M., 1991. Experimental reproduction of porcine epidemic abortion and respiratory syndrome (mystery swine disease) by infection with Lelystad virus: Koch's postulates fulfilled. Veterinary Quarterly 13, 131–136.
Tian, K., Yu, X., Zhao, T., Feng, Y., Cao, Z., Wang, C., Hu, Y., Chen, X., Hu, D., Tian, X., Liu, D., Zhang, S., Deng, X., Ding, Y., Yang, L., Zhang, Y., Xiao, H., Qiao, M., Wang, B., Hou, L., Wang, X., Yang, X., Kang, L., Sun, M., Jin, P., Wang, S., Kitamura, Y., Yan, J., Gao, G.F., 2007. Emergence of fatal PRRSV variants: unparalleled outbreaks of atypical PRRS in China and molecular dissection of the unique hallmark. PLoS One 2, e526.
Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., Leunissen, J.A., 2007. Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Research 35, W71–W74.
Van Doorsselaere, J., Geldhof, M., Nauwynck, H.J., Delputte, P.L., 2011. Characterization of a circulating PRRSV strain by means of random PCR cloning and full genome sequencing. Virology Journal 8, 160.
Wensvoort, G., Terpstra, C., Pol, J.M., ter Laak, E.A., Bloemraad, M., de Kluyver, E.P., Kragten, C., van Buiten, L., den Besten, A., Wagenaar, F., 1991. Mystery swine disease in The Netherlands: the isolation of Lelystad virus. Veterinary Quarterly 13, 121–130.
Wu, W.H., Fang, Y., Farwell, R., Steffen-Bien, M., Rowland, R.R., Christopher-Hennings, J., Nelson, E.A., 2001. A 10-kDa structural protein of porcine reproductive and respiratory syndrome virus encoded by ORF2b. Virology 287, 183–191.