Journal of General Virology (2009), **90**, 1423–1432 DOI 10.1099/vir.0.009381-0

Improved strategies for sequence-independent amplification and sequencing of viral double-stranded RNA genomes

A. C. Potgieter,[1] N. A. Page,[2,3] J. Liebenberg,[4] I. M. Wright,[1] O. Landt[5] and A. A. van Dijk[6]

Correspondence
A. C. Potgieter
potgieterc@arc.agric.za

[1]OIE reference laboratories for African horsesickness and bluetongue, Virology Division, Onderstepoort Veterinary Institute, Private Bag X5, Onderstepoort 0110, South Africa

[2]Viral Gastroenteritis Unit, National Institute for Communicable Diseases, Private Bag X4, Sandringham 2131, South Africa

[3]Diarrhoeal Pathogens Research Unit, University of Limpopo, Medunsa Campus, PO Box 173, Medunsa 0204, South Africa

[4]Molecular Biology Division, Onderstepoort Veterinary Institute, Private Bag X5, Onderstepoort 0110, South Africa

[5]Tib Molbiol GmbH, Eresburgstrasse 22–33, D12103 Berlin, Germany

[6]Biochemistry Division, North-West University, Private Bag X6001, Potchefstroom 2520, South Africa

This paper reports significant improvements in the efficacy of sequence-independent amplification and quality of sequencing of viruses with segmented double-stranded RNA (dsRNA) genomes. We demonstrate that most remaining bottlenecks in dsRNA virus genome characterization have now been eliminated. Both the amplification and sequencing technologies used require no previous sequence knowledge of the viral dsRNA, there is no longer a need to separate genome segments or amplicons and the sequence-determined bias observed in cloning has been overcome. Combining very efficient genome amplification with pyrophosphate-based 454 (GS20/FLX) sequencing enabled sequencing of complete segmented dsRNA genomes and accelerated the sequence analysis of the amplified viral genomes. We report the complete consensus sequence of seven viruses from four different dsRNA virus groups, which include the first complete sequence of the genome of equine encephalosis virus (EEV), the first complete sequence of an African horsesickness virus (AHSV) genome determined directly from a blood sample and a complete human rotavirus genome determined from faeces. We also present the first comparison between the complete consensus sequence of a virulent and an attenuated strain of AHSV1. Ultra-deep sequencing (>400-fold coverage) of the AHSV1 reference and attenuated strains revealed different ratios of reassortants in the reference strain and allowed quasispecies detection in the plaque-purified attenuated strain of AHSV1. This approach amounts to a paradigm shift in dsRNA virus research, since it is sensitive and specific enough for comprehensive investigations of the evolution and genetic diversity in dsRNA virus populations.

Received 3 December 2008
Accepted 18 February 2009

INTRODUCTION

The double-stranded RNA (dsRNA) viruses are a large group of viruses that are classified in seven virus families: *Birnaviridae*, *Chrysoviridae*, *Cystoviridae*, *Hypoviridae*, *Partitiviridae*, *Reoviridae* and *Totiviridae*. These viruses infect a large variety of hosts including humans, animals, plants, fungi and bacteria (Mertens *et al.*, 2005). Infection

The GenBank/EMBL/DDBJ accession numbers for the sequences reported in this paper are AM883164–AM883173, FJ011107–FJ011116, FJ196584–FJ196593 and FJ183353–FJ183393 (given in Table 2).

A supplementary figure and the full technical protocol for sequence-independent amplification of viral dsRNA genomes are available with the online version of this paper.

with pathogenic dsRNA viruses can lead to human and animal deaths and often cause great financial losses. The viral dsRNA genomes are composed of 1–12 segments that range in size from about 200 to 6800 bp. The size of their genomes generally varies between 5000 and 20 000 bp (Mertens, 2004; Mertens *et al.*, 2005; Roy *et al.*, 1994).

Over the past 25 years there have been steady advances in techniques for the cloning, amplification and sequencing of the genomes of dsRNA viruses (Attoui *et al.*, 2000; Bigot *et al.*, 1995; Cashdollar *et al.*, 1982; Lambden *et al.*, 1992; Maan *et al.*, 2007; Potgieter *et al.*, 2002; Rao *et al.*, 1983; Vreede *et al.*, 1998). The most recent improvements achieved cDNA synthesis of the large (>2000 bp) dsRNA genome segments, the preparation of cDNA and cloning of genome sets using single one-tube reactions for oligo-ligation, cDNA synthesis and PCR (Potgieter *et al.*, 2002), as well as the increase of specificity by the introduction of 'anchor primers' which prime themselves for cDNA synthesis (Maan *et al.*, 2007). To date, sequencing of amplified genomes and genome segments has only been achieved by sequencing either individual cloned single genome segments (Attoui *et al.*, 2000; Lambden *et al.*, 1992; Potgieter *et al.*, 2002; Vreede *et al.*, 1998) or purified amplicons of individual genome segments (Maan *et al.*, 2007). Both approaches require the separation, purification and 'primer walking' of clones, amplicons or individual genome segments and primers specific for known sequences or conserved terminal ends (Maan *et al.*, 2007).

This paper reports the first experimental evidence that complete sets of cDNA amplicons and the full-length sequence of viral dsRNA genomes (approx. 20 000 bp) can be obtained directly from field and clinical samples (organs, blood and faeces) without any prior virus propagation, knowledge of sequence information, cloning or separation of amplified cDNA. This is achieved by virtue of significant improvement in the specificity and sensitivity of cDNA amplification of dsRNA viral genomes combined with sequencing in microfabricated high-density picolitre reactors on a massive parallel scale (Margulies *et al.*, 2005), by using GS20/FLX technology (Roche Applied Science). Massive parallel sequencing (more than 400-fold coverage) of the African horsesickness virus (AHSV)-1 reference and attenuated strains demonstrated that these technologies are sensitive and specific enough to reveal sequences and ratios of mixtures of reassortants containing different segments in viral populations. It was also possible to detect viral quasispecies. Finally, we discuss the use of consensus and quasispecies sequence information from virulent and attenuated AHSV populations to assess which factors are involved in viral tropism and virulence of AHSV. This is the first truly robust, generally applicable approach which is sensitive and specific enough to start comprehensive investigations into the genetic diversity in dsRNA virus populations.

METHODS

Viruses. The viruses used in this study and their passage history are given in Table 1. The AHSV, bluetongue virus (BTV), equine encephalosis virus (EEV) and epizootic hemorrhagic disease virus (EHDV) strains used in this study were propagated in baby hamster kidney (BHK)-21 cells grown in Eagle's minimal essential medium (BioWhittaker) in 75 cm^2 flasks, unless indicated otherwise in Table 1. The AHSV and EEV strains, as well as the organs and blood from horses that died of AHSV and the faeces of a calf suffering from rotavirus diarrhoea, were obtained from the OIE World Reference Centre for AHSV and BTV at the Agricultural Research Council–Onderstepoort Veterinary Institute, South Africa. The human rotavirus G9P[6] strain used in this study was kindly provided by Mrs Ina Peenze from the virus collection of the Diarrhoeal Pathogens

Table 1. Origin, passage history and amount of sequence data generated for the viruses used in this study

Eq Spln, Equine spleen; A, adult mice; S, suckling mice; LP, large plaque isolation; KC, KC cells. The raw sequence data are available on request.

Virus	Country of origin	Passage history	Genome sequencer	Company	Area on large PTP	Sequence generated (bp)	Average read length (bp)
AHSV1 reference strain [HS29/62]	South Africa	Eq Spln, 3 S, 2 BHK	GS20	Roche	2 × 1/4	10 024 679	102
AHSV1 attenuated strain	South Africa	Eq Spln, 3 S, 10 BHK, 3 LP, 5 Vero, 1 BHK	GS20	Roche	2 × 1/4	9 448 801	102
Low passage of AHSV1 reference strain [HS29/62]	South Africa	Eq Spl, 1 A, 1 S	GSFLX	Inqaba Biotec	1 × 1/16th	2 514 441	201
AHSV2 Lagos	Nigeria	None, from blood	GSFLX	Inqaba Biotec	1 × 1/16th	2 989 035	232
BTV8 Netherlands	Netherlands	1 KC, 3 BHK	GS20	DYN	1 × 1/16th	901 109	101
EEV Bryanston (HS103/06)	South Africa	1 Vero, 1 BHK	GS20	Inqaba Biotec	1 × 1/8th	2 182 244	101
Human rotavirus G9P[6]	South Africa	None, from faeces	GS20	Inqaba Biotec	2 × 1/16th	951 275	105

Research unit, University of Limpopo, Medunsa, South Africa. The BTV strain used in this study was kindly provided by Professor Peter Mertens from the OIE Reference Centre for BTV, Pirbright, UK.

dsRNA preparation. Total RNA was extracted from cell culture, lung, spleen, faeces or blood using the commercial guanidinium isothiocyanate reagent TRI-REAGENT–LS (Molecular Research Centre), according to the manufacturer's protocol. Single stranded RNA (ssRNA) was removed by precipitation with 2 M LiCl (Sigma) at 4 °C for 16 h followed by centrifugation at 16 000 ***g*** for 30 min. dsRNA was purified from the resulting supernatant using a MinElute gel extraction kit (Qiagen). Where visible, the integrity of dsRNA was evaluated after separation on a 1 % agarose gel (TBE) stained with ethidium bromide. dsRNA of the bovine rotavirus was subjected to SDS-PAGE using 9 % acrylamide for separation (Laemmli, 1970) and silver stained.

Oligo design. An 'anchor primer', PC3-T7 loop, similar to that described by Maan *et al.* (2007), was used in ligation. PC3-T7 loop (5′-p–GGATCCCGGGAATTCGGTAATACGACTCACTATATTTTTAT-AGTGAGTCGTATTA–OH-3′) was synthesized by Tib Molbiol.

Oligo-ligation. PC3-T7 loop (200 ng) was ligated to dsRNA (0.4–200 ng) in 50 mM HEPES/NaOH, pH 8.0 (Sigma), 18 mM $MgCl_2$ (Separations), 0.01 % BSA (TaKaRa), 1 mM ATP (Roche), 3 mM DTT (Roche), 10 % DMSO (Sigma), 20 % polyethyleneglycol $(PEG)_{6000}$ (BDH) and 30 U T4 RNA ligase (TaKaRa) in a final volume of 30 µl. Ligation was performed at 37 °C for 16 h. Ligated dsRNA was purified using MinElute Gel extraction columns following the manufacturer's recommendations (Qiagen).

Sequence-independent cDNA synthesis and PCR amplification. Purified ligated dsRNA was denatured by the addition of 300 mM methyl mercury hydroxide (MMOH; Alfa Aesar) to a final concentration of 30 mM. Alternatively, dsRNA was denatured by the addition of DMSO to a final concentration of 15 % (v/v), heating in a thermal cycler at 95 °C for 2 min and snap-freezing in an ice-water slurry. However, denaturation with MMOH is a lot more efficient than with DMSO and heat, so it is, therefore, the method of choice when only very small amounts of starting material are available. cDNA was reverse transcribed in a cDNA reaction containing 50 mM Tris/HCl, pH 8.3 (Sigma), 10 mM $MgCl_2$ (Separations), 70 mM KCl (Sigma), 30 mM β-mercaptoethanol (Sigma), 1 mM dNTPs (TaKaRa) and 15 U cloned AMV reverse transcriptase (Invitrogen). The reaction was incubated in a thermal cycler at 42 °C for 45 min followed by 55 °C for 15 min. After cDNA synthesis, the excess RNA was removed by adding NaOH (Sigma) to a final concentration of 0.1 M and incubation in a thermal cycler at 65 °C for 30 min. Before cDNA annealing, Tris/HCl, pH 7.5 (Sigma), was added to a final concentration of 0.1 M followed by the addition of HCl (Sigma) to a final concentration of 0.1 M. The cDNA was annealed at 65 °C for at least 1 h.

Amplification of cDNA was performed using primer PC2 (5′-p–CCGAATTCCCGGGATCC-3′) which contains the restriction enzyme sites for *Eco*RI, *Sma*I/*Xma*I and *Bam*HI to facilitate cloning and subcloning of amplified cDNA. The PCR mixture contained 1 × Ex Taq buffer, 0.2 mM dNTPs (TaKaRa), 5 µl cDNA and 2.5 U TaKaRa Ex Taq. The first step during cycling was 72 °C for 1 min to fill incomplete cDNA ends to produce intact DNA. This was followed by an initial denaturation step of 94 °C for 2 min followed by 15–25 cycles of 94 °C for 30 s, 67 °C for 30 s and 72 °C for 4 min (or 1 min per kb of the largest segment). A final extension step of 72 °C for 5 min was included. For increased fidelity, Phusion polymerase (Finnzymes) was used instead of TaKaRa Ex Taq. Amplified cDNA products were viewed after separation on 1 % agarose gels (TBE) containing ethidium bromide. The complete technical protocol for the sequence-independent amplification of viral dsRNA genomes and a schematic representation of this (Supplementary Fig. S1) are available in JGV Online.

Sequencing using GS20/FLX technology. Prior to sequencing using GS20 or GSFLX technology, the amplified cDNA was purified using a QIAquick PCR purification kit (Qiagen). The preparation of DNA libraries, titrations, emPCR and sequencing on the GS20/FLX sequencers, were performed by various companies that provide commercial sequencing services using the GS20/FLX sequencers, with the exception of RochePenzberg, where proof of principle studies were conducted on the AHSV1 genomes. The regions on a large Pico Titre Plate (PTP) used for sequencing each of the genomes and the company that performed the sequencing are listed in Table 1.

Sequence analysis. The initial assembly of sequences using GS20 software (Roche) was insufficient for our purposes. Therefore, the raw sequence data of the genomes were assembled *de novo* in GAP4 (Bonfield *et al.*, 1995) using the normal shotgun assembly. The assembly was confirmed by aligning it with known sequences (where available) and each segment was manually checked and edited. Subsequently, Lasergene7 software from DNASTAR was used for *de novo* assembly of the contigs. Files containing the sequence information, quality values and flowgrams (sff files) were loaded into the Seqman 7 programme of the Lasergene software. Default assembly parameters were used except for the minimum read length which was set to 30 bp for GS20 reads and 50 bp for GSFLX reads. Contigs resulting from the assembly were checked manually and their consensus sequences were exported as FASTA files. Consensus sequences were aligned to known sequences using MEGALIGN (Lasergene7). Finally, sequences were subjected to BLASTN analysis using the National Center for Biotechnology Information website. The consensus sequence of the seven complete dsRNA virus genome sets that were generated during this investigation have been deposited in GenBank under the accession numbers listed in Table 2.

RESULTS AND DISCUSSION

Improving oligo-ligation efficiency by macromolecular crowding

Since it has been reported that the efficacy of enzymes used for ligation of RNA and DNA segments can be improved substantially using macromolecular crowding with high molecular mass chemicals such as PEG (Harrison & Zimmerman, 1984), we investigated whether oligo-ligation with T4 RNA ligase could also be improved by the inclusion of PEG and hexamine cobalt chloride (HCC) in our reactions. Four ligation reactions were performed, as described above, each with 100 ng AHSV4 dsRNA. One ligation reaction was carried out using HEPES ligation buffer (Potgieter *et al.*, 2002) without PEG_{6000}, a second was carried out in the presence of 20 % (w/v) PEG_{6000}, a third contained 1 mM HCC and the fourth reaction contained both 20 % (w/v) PEG_{6000} and 1 mM HCC. cDNA synthesis was carried out as described previously (Potgieter *et al.*, 2002). The cDNA was amplified in 15 PCR cycles and the final amplification products were separated on 1 % TBE agarose gels. The presence of 20 % PEG_{6000} in the oligo-ligation reaction significantly increased the efficiency (Fig. 1b, lane 3). Analysis and quantification of the electrophoresis data using Vision Works Image

Table 2. Gene assignments and GenBank accession numbers of the complete consensus nucleic acid genome sequences of seven dsRNA viruses sequenced using Roche's 454 pyrosequencing GS technology

The raw sequence data are available on request.

Virus	Genome segment number*										
	1	2	3	4	5	6	7	8	9	10	11
AHSV1 reference strain	VP1	VP2	VP3	VP4	NS1	VP5	VP6	VP7	NS2	NS3	–
	AM883164	AM883165	AM883166	AM883167	AM883168	AM883169	AM883170	AM883171	AM883172	AM883173	
AHSV1 attenuated strain	VP1	VP2	VP3	VP4	NS1	VP5	VP6	VP7	NS2	NS3	–
	FJ183364	FJ183365	FJ183366	FJ183367	FJ183368	FJ183369	FJ183370	FJ183371	FJ183372	FJ183373	
AHSV1, 1 A, 1 S†	VP1	VP2	VP3	VP4	NS1	VP5	VP6	VP7	NS2	NS3	–
	FJ011107	FJ011108	FJ011109	FJ011110	FJ011111	FJ011112	FJ011113	FJ011114	FJ011115	FJ011116	
AHSV2 Lagos	VP1	VP2	VP3	VP4	NS1	VP5	VP6	VP7	NS2	NS3	–
	FJ196584	FJ196585	FJ196586	FJ196587	FJ196588	FJ196589	FJ196590	FJ196591	FJ196592	FJ196593	
BTV8 Netherlands	VP1	VP2	VP3	VP4	NS1	VP5	VP7	NS2	VP6	NS3	–
	FJ183374	FJ183375	FJ183376	FJ183377	FJ183378	FJ183379	FJ183380	FJ183381	FJ183382	FJ183383	
EEV Bryanston	VP1	VP2	VP3	VP4	NS1	VP5	NS2	VP7	VP6	NS3	–
	FJ183384	FJ183385	FJ183386	FJ183387	FJ183388	FJ183389	FJ183390	FJ183391	FJ183392	FJ183393	
Human rotavirus G9P[6]	VP1	VP2	VP3	VP4	NSP1	VP6	NSP3	VP7	NSP2	NSP5	NSP4
	FJ183353	FJ183354	FJ183355	FJ183356	FJ183357	FJ183358	FJ183359	FJ183360	FJ183361	FJ183362	FJ183363

*Genome segment number is solely based on the size of the segment and not on SDS-PAGE profiles.
†A, Adult mice; S, suckling mice.

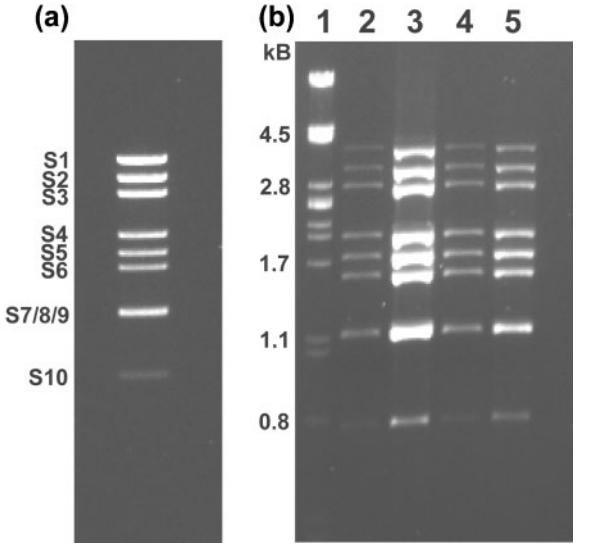

Fig. 1. Ethidium bromide-stained agarose gel (1 %) showing the effect of PEG_{6000} and HCC on oligo-ligation. (a) Purified dsRNA of the AHSV4 reference strain. Amplicons are numbered according to the cognate genome segment. (b) cDNA amplification products after oligo-ligation with different buffers, cDNA synthesis and 15 cycles of cDNA amplification. Lanes: 1, *Pst*I-digested phage λ DNA size marker; 2, cDNA amplified after ligation in normal HEPES-buffered oligo-ligation reaction mixture; 3, cDNA after ligation in HEPES-buffered oligo-ligation reaction mixture containing 20 % PEG_{6000}; 4, cDNA after ligation in HEPES-buffered oligo-ligation reaction mixture containing 1 mM HCC; 5, cDNA after ligation in HEPES-buffered oligo-ligation reaction mixture containing both 20 % PEG_{6000} and 1 mM HCC.

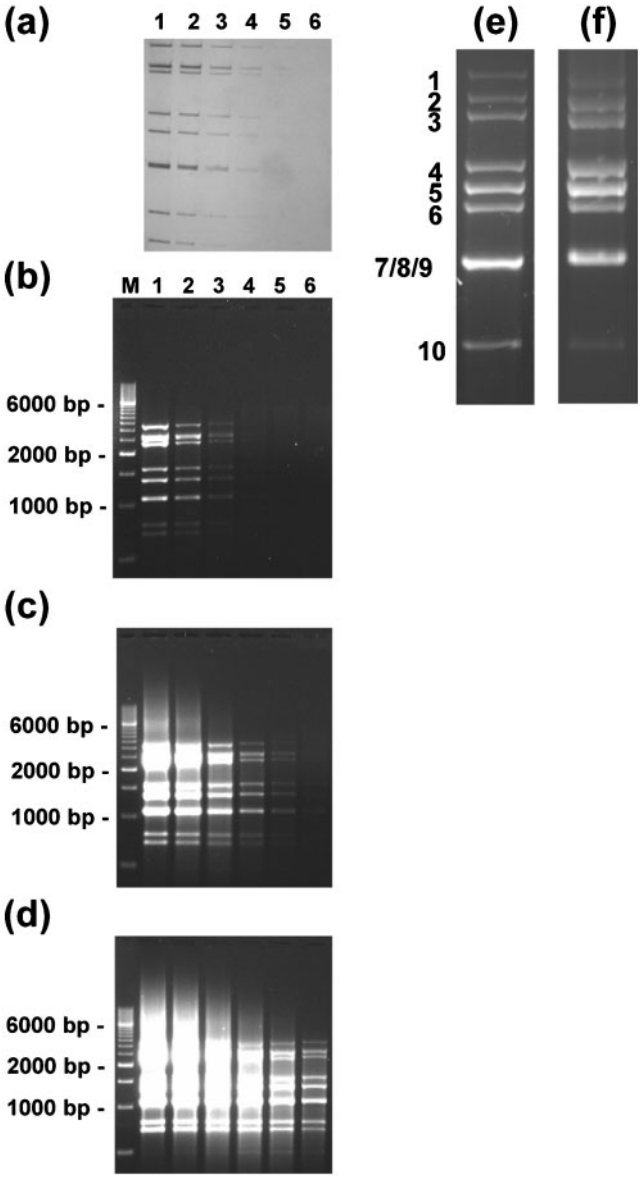

Fig. 2. Determining the sensitivity of the improved complete genome amplification procedure. (a) Silver-stained SDS-PAGE gel of a dilution series of calf rotavirus dsRNA. (b–d) Ethidium bromide-stained 1 % agarose gel of cDNA amplification products from 100, 33, 11, 3.7, 1.2 and 0.4 ng (lanes 1–6, respectively) dsRNA after 15 (b), 20 (c) and 25 (d) PCR cycles. Lane M, 500 bp ladder (Fermentas SM0643). (e, f) Agarose gel analysis of amplicons generated from AHSV dsRNA extracted directly from clinical samples. Genome amplified from dsRNA extracted from the spleen of a horse that died after infection with AHSV9 (e) and a blood clot from a horse that died after infection with AHSV2 (Lagos) (f). Amplicons are numbered according to the cognate genome segment.

Analysis software (UVP), indicated that the addition of PEG_{6000} to the ligation reaction increased the yield of amplicons approximately fivefold from 100 ng of starting dsRNA and 15 cycles of PCR. The addition of HCC to ligation reactions did not improve the ligation; in fact, addition of HCC to ligation buffer containing PEG_{6000} reduced the efficiency (Fig. 1b, compare lanes 3 and 5).

Determining the amplification sensitivity

To determine the impact of the increased efficiency of ligation on the sensitivity of cDNA synthesis, we compared the sensitivity to that of our previous method (Potgieter *et al.* 2002) using dsRNA purified from the faeces of a calf suffering from diarrhoea caused by a group A rotavirus as template. The dsRNA was diluted 1 : 3 with 10 mM Tris, pH 8.0, from a concentration of 10 ng μl^{-1} down to 40 pg μl^{-1}. Aliquots of 10 µl from each dilution containing approximately 100, 33.3, 11.1, 3.7, 1.2 and 0.4 ng were separated by SDS-PAGE and the dsRNA was visualized after silver staining (Fig. 2a). Another 10 µl aliquot from each dilution was subjected to the improved oligo-ligation and cDNA amplification in the presence of 20 % PEG_{6000}, as described above. The cDNA amplification was performed from 10 µl cDNA in a 30 µl PCR with TaKaRa Ex Taq. After 15, 20 and 25 PCR cycles, a 5 µl aliquot from each amplification reaction was separated on a 1 % agarose

gel (TBE) and visualized with ethidium bromide staining. The genomes amplified from different amounts of bovine rotavirus dsRNA and different numbers of PCR cycles are shown in Fig. 2. The cDNA of the complete genome of a bovine rotavirus could be amplified from approximately 400 pg starting material in 25 PCR cycles (Fig. 2d, lane 6). Previously, the amplification of a similar rotavirus genome was reported from 1 ng dsRNA and needed 30 PCR cycles (Potgieter *et al.*, 2002). Although about the same total amount of amplification product from ligation reactions of 100 and 0.4 ng dsRNA could be obtained by increasing the number of PCR cycles, the yields of the larger segments were not equal (compare Fig. 2b, lane 1, after 15 PCR cycles and Fig. 2d, lanes 4–6, after 25 PCR cycles). Therefore, the cDNA of the large genome segments of dsRNA viruses still does not amplify as well as that of the medium and small genome segments.

Encouraged by these results, we attempted to amplify the genomes of AHSV with dsRNA purified from field samples (spleen and blood). The successful amplification of AHSV cDNA from RNA extracted from the spleen and a blood clot from separate field cases are shown in Fig. 2 (e, f). The total amount of dsRNA from these samples is not known, since the yield of purified dsRNA was too low to be visible on ethidium bromide-stained agarose gels. However, the yield of the amplicons of the genome amplified from the blood sample was sufficient to determine the complete genome sequence using GSFLX sequencing (Table 2). This is the first report of sequence-independent genome amplification and sequencing of the genome of an orbivirus directly from a field sample. In other cases, we were able to sequence the terminal ends of genome segment S2 using 'phased primers' similar to those described by Maan *et al.* (2007) and determine the genotype of the viruses (Fasina *et al.*, 2008). In certain cases, the methods described here yielded only partial sequences, representing only some of the ten segments (results not shown). Therefore, the sensitivity of the methods depends on the level of viraemia and the amount of dsRNA extracted from the sample.

Sequencing amplified genomes using GS20/FLX technology

So far, we have amplified and sequenced 52 dsRNA virus genomes of seven different dsRNA viruses (AHSV, BTV, *Cryptosporidium* virus, EEV, EHDV, picobirnavirus and rotavirus) using pyrophosphate-based 454 sequencing technology. Here, as proof of principle of the improved amplification protocol, we report the consensus sequences of seven virus genome sets of four high profile dsRNA viruses (AHSV, BTV, EEV and human rotavirus). The results are summarized in Tables 1 and 2.

The average length of reads generated with the GS20, as expected, was 100 bp. The total sequence data generated per genome on 1/16th region of large PTP on the GS20 (sequenced by DYN) varied between 0.26 and 1.08 Mb (results not shown) and that on 1/8th regions on the GS20 (sequenced at Inqaba Biotec) between 1 and 3 Mb. On the GSFLX, reads were twice as long as on the GS20 by virtue of improvements in the sequencing technology. Total sequence data that were obtained on 1/16th regions on the GSFLX varied between 1 and 3 Mb. Overall, the coverage that was achieved varied between 12- and 150-fold. In our experience, a 40-fold coverage of dsRNA genomes amplified as described here, allows the determination of the complete consensus sequences of the 18–20 kb viral dsRNA genomes (Table 2). A lower coverage does not allow full genome sequence determination, since the larger segments amplify less efficiently and are present in smaller molar amounts. The larger the amount of dsRNA used for amplification, the better the amplification of the large segments (Fig. 2), as measured by the sequence coverage. An inherent technical problem of pyrophosphate-based GS20 sequencing appeared to be the inability to resolve multiple homopolymer base pair repeats accurately (Fig. 3). Manual checking of alignments of homopolymer regions resolves this problem quite easily, as extra base pairs or deletions are usually present in smaller amounts than the correct consensus sequences. The original flowgrams of a particular region can be viewed in Seqman (Lasergene7), which allows visual confirmation of the number of bases that actually occur in a specific homopolymer region. Although the resolution has been improved for the GSFLX, it is still difficult to detect deletion mutations correctly.

Despite these challenges, the experimental approach and methods used here for dsRNA sequencing are significantly more efficient than Sanger cycle sequencing, less expensive, less time consuming and less labour intensive overall. Amplicons of single genome segments do not need to be separated by gel electrophoresis and purified. No sequence information on the terminal ends is needed. These properties make the methods described here suitable for sequencing of known as well as new dsRNA viruses and even those without conserved terminal ends. Roche recently improved the system, allowing several different genomes to be sequenced in a single lane on PTPs (Meyer *et al.*, 2007) and to increase the read lengths to 400 bp and more, resulting in a decrease in the costs for complete viral genome sequencing.

Ultra-deep sequencing AHSV1 reference and attenuated strains, detecting quasispecies and evidence of reassortment with the AHSV3 reference strain

In contrast with the relatively low genome coverage obtained on 1/8th and 1/16th lanes, sequencing of the AHSV1 reference and attenuated strains on two regions each of a four region PTP on the GS20 (Roche, Penzberg) yielded approximately 10 Mb raw data per genome, corresponding to a coverage of >400-fold of the 20 kb genomes (Table 1). The average length of reads was 102 bp. Curiously, assembly of these sequences with GS20 software

did not yield the complete consensus sequences of each virus. When the alignments were repeated using manual GAP4 assembly, 10 large contigs were obtained per genome. Each of the contigs contained the complete consensus sequence of one of the genome segments of the two AHSV1 strains. The accession numbers of the consensus sequences are listed in Table 2.

Further analysis of alignments from the assembly of the AHSV1 reference and attenuated viruses revealed random mutations at various sites in each of the 10 genome segments of both viruses. These differences and deletions were observed in all genome segments. Whether these are true quasispecies or errors introduced due to the low fidelity of the reverse transcriptase and/or DNA polymerases used for cDNA amplification is not known.

Close scrutiny of sequence alignments from the AHSV1 reference strain also revealed mutations with distinct repetitive patterns at the same sites within four of the 10 genome segments, namely S5 (NS1), S8 (VP7), S7 (VP6) and S10 (NS3). An example of this from the alignments of the AHSV1 reference strain S7 (VP6) alignment is shown in Fig. 3. We refer to these repetitive changes as subpopulations. The subpopulations in the AHSV1 reference strain occurred at different frequencies in each of the four genome segments. At first it was thought that these were naturally occurring quasispecies. However, close scrutiny of the sequences revealed that these were in fact a mixture of AHSV1 reference strain and AHSV3 reference strain sequences (data not shown). The ratios of AHSV1 sequences to AHSV3 sequences were as follows: S5 (NS1),

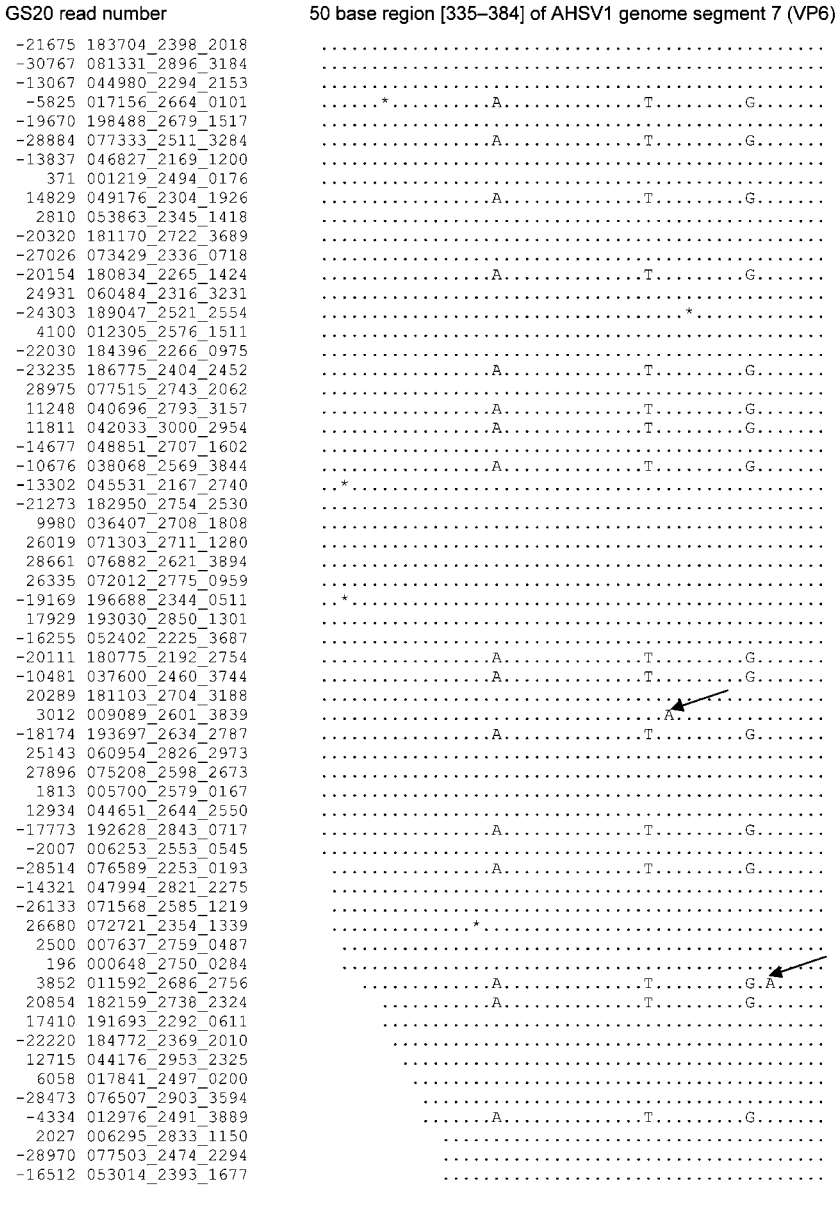

Fig. 3. Detecting reassortment between genome segment 7 (VP6) from AHSV1 and AHSV3. The aligned sequences were generated from the amplified genome of the AHSV1 reference strain on the GS20 sequencer. All sequences identical to the consensus sequence are indicated with a dot (.); sequences that differ from the consensus sequence are indicated with the letter corresponding to the base pair (A, T or G); random deletion mutations are indicated with an asterisk (*); other random mutations are indicated with arrows.

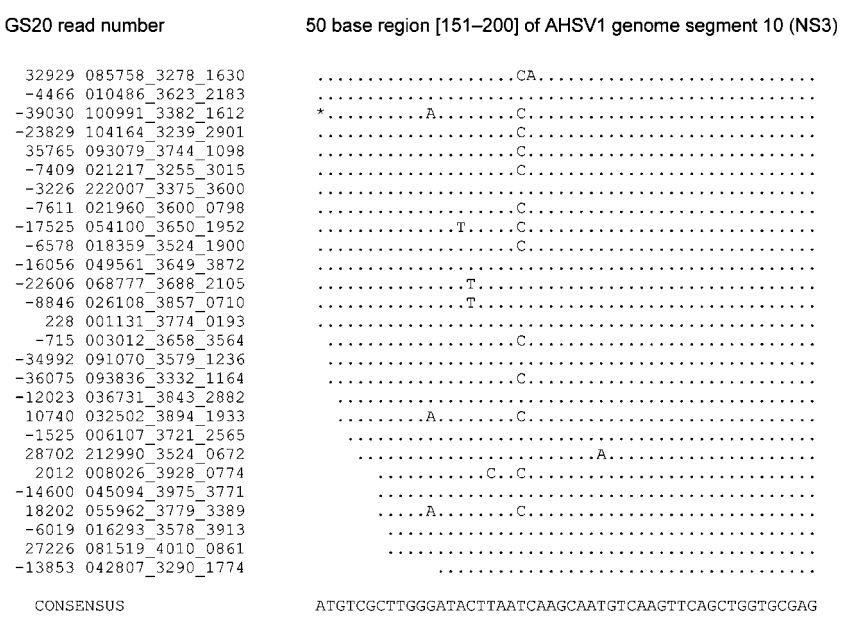

Fig. 4. Detection of quasispecies in genome segment 10 (NS3) of the AHSV1 attenuated (plaque-purified) strain. Sequences and random deletion mutation are labelled with dots, letters or asterisks as in Fig. 3.

15 : 85, S8 (VP7), 95 : 5; S7 (VP6), 70 : 30; and S10 (NS3), 85 : 15. Therefore, it was speculated that the AHSV1 reference strain used in this study (AHSV1 Equine spleen, 3 S, 2 BHK; Table 1) is in the process of reassortment with the AHSV3 reference strain. To prove this, we sequenced the lowest passage of the AHSV 1 reference strain that we have in our virus bank (AHSV1 Equine spleen, 1 A, 1 S; Table 1) from which this reference strain was derived. The consensus sequences of nine of the 10 genome segments of the AHSV1 reference strain and the low passage strain from which it was derived were identical. Only the sequence of genome segment 5, encoding NS1, of the reference strain was found to be almost identical to the AHSV3 reference strain.

It was thus possible to conclude that the AHSV1 reference strain must have been in contact with the AHSV3 reference strain during passage on cell culture and that the two viruses started to reassort. Reassortment was, however, not complete, since the virus population of the AHSV1 reference strain contained a mixture of AHSV1 and AHSV3 segments, indicating the presence of different ratios of reassorted segments. The plaque-purified, attenuated strain of AHSV1 also contains the NS1 protein of AHSV3. Our results indicate that the technologies described here allowed us to follow reassortment events as they take place.

We did not detect any significant repeated quasispecies in nine of the 10 genome segments (S1–S9) of the plaque-purified, attenuated AHSV1 virus population. In genome segment S10 (NS3) we did detect quasispecies that were acquired during the passage of the virus on Vero and BHK cells after it was plaque-purified three times (Fig. 4). These changes were not detected in either of the AHSV1 reference strains. We speculate that the mutation 171T(U)→C, which is shown in Fig. 4 and present in almost 50 % of the reads, occurred first and was followed by the mutation 162G→A, which is present in a much smaller frequency and only in reads where the first mutation is present. However, we do not have sequence information from earlier passages that could confirm this. Nevertheless, in this small section of sequence, we could detect three different varieties of NS3 which were acquired during only six passages in cell culture (5 Vero, 1 BHK). In our view, the other single random changes are not significant and may represent errors due to the low fidelity of the enzymes used.

These results are in agreement with those reported for BTV that was limited to genome segments S2 (VP2) and S10 (NS3) upon passage between sheep, cattle and the insect vector, *Culicoides sonorensis* (Bonneau *et al.*, 2001). In contrast with the latter investigation in which sequence data had to be generated from a limited number of genome segments using genome segment-specific primers or cloning, with our new approach we could screen the complete genome for quasispecies without the need for any genome segment-specific primers or cloning.

We anticipate that the combination of this improved sequence-independent dsRNA genome amplification and ultra-deep pyrosequencing will become an extremely useful tool to study the evolution in dsRNA viruses during either natural passages during virus outbreaks in the field, attenuation or adaptation to cell culture. It should also greatly speed up the development and general implementation of genome-based classification systems for dsRNA viruses for surveillance and epidemiological purposes (Matthijnssens *et al.*, 2008).

Comparing consensus genome sequences from a virulent and an attenuated AHSV1 strain

To initiate investigations aimed at identifying factors that determine virulence in AHSV, we compared the consensus

sequences of the genome of the attenuated strain of AHSV1 and its parental strain. Although the parental strain of the AHSV1 attenuated strain contained a mixture of AHSV1 and AHSV3 sequences, as shown above, AHSV1 low and high passage strains share an identical consensus, except for genome segment S5 (NS1), reflecting true changes in the consensus sequence of the AHSV1 attenuated strain.

The consensus sequence of three of the genome segments, S1 (VP1), S5 (NS1) and S9 (NS2), were identical between the virulent and attenuated strains of AHSV1 (Table 3). Overall, there were 16 nucleotide differences in the consensus sequences of the seven genome segments in which variations occurred, of which seven resulted in amino acid changes. Only two nucleotide changes occurred in the non-coding regions, one in S8 (VP7) and the other in S10 (NS3). Most nucleotide changes were transitions (12 of 16). Since the proportion of the non-coding regions in the genome is small, two of 16 nucleotide differences in this region seems high. However, when the differences were quantified, the proportion of changes is actually much less in the non-coding regions and is, therefore, not more significant. This is generally in agreement with several sequencing studies on other orbiviruses (A. C. Potgieter, unpublished data). We hypothesize that these are the most frequent errors that the viral RNA-dependent RNA polymerase makes during replication.

In our view, the most significant amino acid changes related to virulence might be those in the outer capsid proteins, namely VP2 and VP5. The reason for this hypothesis stems from extensive studies of AHSV3 and AHSV4 in horses, from which it was concluded that virulence of AHSV is related to the affinity of the virus for certain tissues (tissue tropism) and that attenuation of the virus is simply the selection of viruses that do not have selective affinity for vital organs such as the lung (Erasmus, 1973). This also correlates with findings that the outer capsid proteins of orbviruses are involved in cell entry and triggering of apoptosis by the virus (Bhattacharya *et al.*, 2007; Mortola *et al.*, 2004). However, it is also possible that any of the other changes in the genome sequence may be associated with the attenuation. For example, in this specific case, the reassortment of AHSV1 S5 (NS1) with AHSV3 S5 (NS1) may have caused the attenuation. To pinpoint which changes are responsible for attenuation, a reverse genetics system for AHSV would be required. The fact that the AHSV1 sequence data presented here are in agreement with the early AHSV3 and AHSV4 virological and clinical results (Erasmus, 1973) is exciting, since they demonstrate how the different disciplinary approaches support each other by (re)confirming, complementing and extending the findings. It is anticipated that the work started here will give an indication of which factors are involved in tissue tropism and virulence of AHSV.

In conclusion, this report on the significant improvement in the efficiency of generating cDNA from dsRNA viral genomes, combined with the high throughput and coverage of pyrosequencing, introduces a paradigm shift for dsRNA virus research. Viral dsRNA genomes can now be completely sequenced, analysed and cloned directly from field and clinical samples without requiring any prior sequence information. The technology also allows the detection of quasispecies in complete genome sets as opposed to single genome segments. It is envisaged that this progress will facilitate a host of qualitative and quantitative investigations of dsRNA virus population dynamics and viral evolution. It should now be possible to begin to rationally and systematically identify and investigate the factors that are involved in tissue tropism, virulence and virus/host and virus/vector interactions of many dsRNA viruses.

Table 3. Comparison of consensus sequences of the genome segments of the AHSV1 reference and attenuated strains

The consensus nucleic acid sequences of AHSV1 (3 S, 2 BHK) is identical to that of AHSV1 (1 A, 1 S) with the exception of genome segment 5, encoding NS1, which reassorted with AHSV3.

Genome segment	Nucleic acid change	Amino acid change
1 (VP1)	–	–
2 (VP2)	1083 U→A	–
	1365 G→A	357 N→K
	2814 A→G	–
3 (VP3)	720 U→C	232 Y→H
	1589 A→G	–
	1832 G→A	–
4 (VP4)	1748 A→G	–
5 (NS1)	–	–
6 (VP5)	724 C→U	
	1284 G→A	422 S→N
	1320 C→U	434 T→I
7 (VP6)	259 C→U	81 A→V
	523 G→A	69 R→Q
8 (VP7)	3 U→A	–
9 (NS2)	–	–
10 (NS3)	4 U→A	–
	170 C→U	–
	619 U→A	201 M→K

ACKNOWLEDGEMENTS

The authors would like to thank Shujing Rao for sharing his ideas on the design of the ligation primer. We are most grateful to Dr Baltus Erasmus (Deltamune, South Africa) for providing information on the attenuation of AHSV. We thank Dr Truuske Gerdes and Mr Roelf Greyling (OIE World Reference Laboratories for AHSV and BTV at the ARC–OVI) for providing the viruses used in this study and Mrs Rencia van der Sluis (North-West University, Potchefstroom, South Africa) for technical assistance. We thank Mr Austen Cohen and Mrs Anna Capovilla from Roche Applied Science in South Africa and Olaf Kaiser from Roche Applied Science in Penzberg for their expert assistance. We express our gratitude towards Relly Forer and Dena Leshkowitz from DYN in Israel. Finally we thank Mr Arshad Ismail and Dr Oliver Preisig from Inqaba Biotec in South Africa for the excellent and professional service they provide. This study was

partially funded by the South African National Department of Agriculture (DoA) and the South African National Research Foundation (FA2005031700015 grant to A. A. v. D.).

REFERENCES

Attoui, H., Billoir, F., Cantaloube, J. F., Biagini, P., de Micco, P. & de Lamballerie, X. (2000). Strategies for the sequence determination of viral dsRNA genomes. *J Virol Methods* **89**, 147–158.

Bhattacharya, B., Noad, R. & Roy, P. (2007). Interaction between Bluetongue virus outer capsid protein VP2 and vimentin is necessary for virus egress. *Virol J* **4**, 7.

Bigot, Y., Drezen, J. M., Sizaret, P. Y., Rabouille, A., Hamelin, M. H. & Periquet, G. (1995). The genome segments of DpRV, a commensal reovirus of the wasp *Diadromus pulchellus* (*Hymenoptera*). *Virology* **210**, 109–119.

Bonfield, J. K., Smith, K. F. & Staden, R. (1995). A new DNA sequence assembly program. *Nucleic Acids Res* **23**, 4992–4999.

Bonneau, K. R., Mullens, B. A. & MacLachlan, N. J. (2001). Occurrence of genetic drift and founder effect during quasispecies evolution of the VP2 and NS3/NS3A genes of Bluetongue virus upon passage between sheep, cattle, and *Culicoides sonorensis*. *J Virol* **75**, 8298–8305.

Cashdollar, L. W., Esparza, J., Hudson, G. R., Chmelo, R., Lee, P. W. & Joklik, W. K. (1982). Cloning the double-stranded RNA genes of reovirus: sequence of the cloned S2 gene. *Proc Natl Acad Sci U S A* **79**, 7644–7648.

Erasmus, B. J. (1973). The pathogenesis of African horsesickness. In *Proceedings of the Third International Conference on Equine Infectious Diseases*. Basel, Switzerland: Karger.

Fasina, F., Potgieter, A. C., Ibironke, A., Bakod, B., Bwala, D. & Kumbish, P. (2008). First report of an outbreak of African horsesickness virus serotype 2 in the Northern hemisphere. *J Equine Vet Sci* **28**, 167–170.

Harrison, B. & Zimmerman, S. B. (1984). Polymer-stimulated ligation: enhanced ligation of oligo- and polynucleotides by T4 RNA ligase in polymer solutions. *Nucleic Acids Res* **12**, 8235–8251.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685.

Lambden, P. R., Cooke, S. J., Caul, E. O. & Clarke, I. N. (1992). Cloning of noncultivatable human rotavirus by single primer amplification. *J Virol* **66**, 1817–1822.

Maan, S., Rao, S., Maan, N. S., Anthony, S. J., Attoui, H., Samuel, A. R. & Mertens, P. P. (2007). Rapid cDNA synthesis and sequencing techniques for the genetic study of bluetongue and other dsRNA viruses. *J Virol Methods* **143**, 132–139.

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y. J. & other authors (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376–380.

Matthijnssens, J., Ciarlet, M., Heiman, E., Arijs, I., Delbeke, T., McDonald, S. M., Palombo, E. A., Iturriza-Gómara, M., Maes, P. & other authors (2008). Full genome-based classification of rotaviruses reveals a common origin between human Wa-like and porcine rotavirus strains and human DS-1-like and bovine rotavirus strains. *J Virol* **82**, 3204–3219.

Mertens, P. (2004). The dsRNA viruses. *Virus Res* **101**, 3–13.

Mertens, P. P. C., Duncan, R., Attoui, H. & Dermody, T. S. (2005). Reoviridae. In *Virus Taxonomy, VIIIth Report of the ICTV*, pp. 447–454. Edited by C. M. Fauquet, M. A. Mayo, J. Maniloff, U. Desselberger & L. A. Ball. London: Elsevier.

Meyer, M., Stenzel, U., Myles, S., Prufer, K. & Hofreiter, M. (2007). Targeted high-throughput sequencing of tagged nucleic acid samples. *Nucleic Acids Res* **35**, e97.

Mortola, E., Noad, R. & Roy, P. (2004). Bluetongue virus outer capsid proteins are sufficient to trigger apoptosis in mammalian cells. *J Virol* **78**, 2875–2883.

Potgieter, A. C., Steele, A. D. & van Dijk, A. A. (2002). Cloning of complete genome sets of six dsRNA viruses using an improved cloning method for large dsRNA genes. *J Gen Virol* **83**, 2215–2223.

Rao, C. D., Kiuchi, A. & Roy, P. (1983). Homologous terminal sequences of the genome double-stranded RNAs of bluetongue virus. *J Virol* **46**, 378–383.

Roy, P., Mertens, P. P. & Casal, I. (1994). African horse sickness virus structure. *Comp Immunol Microbiol Infect Dis* **17**, 243–273.

Vreede, F. T., Cloete, M., Napier, G. B., van Dijk, A. A. & Viljoen, G. J. (1998). Sequence-independent amplification and cloning of large dsRNA virus genome segments by poly(dA)-oligonucleotide ligation. *J Virol Methods* **72**, 243–247.