Bovine viral diarrhoea virus loses quasispecies diversity rapidly in culture

George C. Russell, Ruth N. Zadoks[1], Kim Willoughby, Claudia Bachofen[2]

Moredun Research Institute, Pentlands Science Park, Midlothian, EH26 0PZ, UK

Present address (1) Sydney School of Veterinary Science, University of Sydney, Camden, NSW, Australia;

(2) Institute of Virology, Vetsuisse Faculty, University of Zurich, Winterthurerstr. 266a. CH-8057 Zürich, Switzerland.

Corresponding author:

George C. Russell.

Email: george.russell@moredun.ac.uk

ORCiD: 0000-0002-1951-3215

Keywords: pestivirus; next-generation-sequencing; genome variation; evolution; quasispecies

Sequence data associated with this manuscript has been submitted to the European Nucleotide Archive with accession numbers as follows:

Consensus genome sequences: LR699799-LR699803

NGS raw data: ERR3624580, ERR3624581

Abstract

Bovine viral diarrhoea (BVD) is an important disease of cattle with significant impacts on animal health and welfare. The wide host range of the causative pestiviruses may lead to formation of virus reservoirs in other ruminant or wildlife species, presenting a concern for the long-term success of BVD eradication campaigns. It is likely that the quasispecies nature of these RNA viruses contributes to their interspecies transmission by providing genetic plasticity. Understanding the spectrum of sequence variants present in BVD persistently infected (PI) animals is therefore essential for studies of virus transmission. To analyse quasispecies diversity without amplification bias, we extracted viral RNA from serum of a PI cow, and from cell culture fluid after three passages of the same virus in culture, to produce cDNA without amplification. Sequencing of this material using Illumina 250bp paired-read technology produced full-length virus consensus sequences from both sources and demonstrated the quasispecies diversity of this Pestivirus A type 1a field strain within serum and after culture. We report the distribution and diversity of over 800 single nucleotide polymorphisms and provide evidence for a loss of diversity after only three passages in cell culture, implying that cultured viruses cannot be used to understand quasispecies diversity and may not provide reliable molecular markers for source tracing or transmission studies.

Additionally, both serum and cultured viruses could be sequenced as a set of 25 overlapping PCR amplicons that demonstrated the same consensus sequences and the presence of many of the same quasispecies variants. The observation that aspects of the quasispecies structure revealed by massively parallel sequencing are also detected after PCR and Sanger sequencing suggests that this approach may be useful for small or difficult to analyse samples.

Impact statement

Bovine viral diarrhoea viruses are globally important cattle pathogens, which impact performance due to acute infection and BVD-induced immunosuppression. Eradication of BVD in cattle is widely pursued but is hampered by the production of persistently infected (PI) calves – the offspring of cows infected in early pregnancy – which shed virus constantly and drive BVD spread. Genetic variation in BVD viruses is an important feature of their biology, allowing them to adapt to changing conditions and to infect different hosts. Inaccurate virus replication produces a population of viruses with slightly different sequences, a quasispecies, some of which may grow better in other hosts or in culture. Analysing virus sequence variation may help us understand how the virus evolves within and between its hosts. In this paper we show that a BVD virus strain loses quasispecies diversity quickly when cultured and that these changes can be detected even in small diagnostic samples, implying that cultured viruses do not perfectly represent the field strains they were isolated from and therefore may not provide reliable molecular markers for source tracing or transmission studies.

Data Summary

Pestivirus A genome sequences used in this article are as follows:

strain	accession number
180	HQ174292
8844	HQ174293
GS5	KJ541471
NADL	M31182
SD1	M96751
SingerArg	DQ088995
MRI103	LR699799
526	HQ174294
8824	HQ174295
CC13B	KF772785
CP7-5A	AF220247
IBSP4ncp	KJ620017
KE9	EF101530
Osloss	M96687
ILLNC	U86600
10JJ-SKR	KC757383
Carlito	KP313732

Sequence data associated with this manuscript has been submitted to the European Nucleotide Archive (www.ebi.ac.uk/ena/) with accession numbers as follows:

Consensus genome sequences:

MRI103 serum NGS: LR699799

MRI103 culture NGS: LR699800

MRI103 serum Sanger: LR699801

MRI103 culture P3 Sanger: LR699802

MRI103 culture P5 Sanger: LR699803

NGS raw data

Serum dataset: ERR3624580

Culture dataset: ERR3624581

Introduction

Bovine viral diarrhoea (BVD), caused by viruses of the genus *Pestivirus,* family *Flaviviridae*, is an important cattle pathogen causing immunosuppression, ill-thrift and a lethal form of diarrhoea called mucosal disease (1). The BVD viruses have a positive sense single-stranded RNA genome of about 12500 nucleotides that is replicated in an error-prone manner, generating a wide range of related virus sequences within an infected individual – a “quasispecies cloud” (2, 3). BVD viruses exist as two major types, of which in Britain type 1 prevails and type 2 is only rarely identified (4, 5). Recently, however, pestivirus nomenclature has been revised so that BVDV type 1 (now Pestivirus A) and BVDV type 2 (Pestivirus B) are considered distinct virus species (6). Pestivirus A genotypes are defined by sequence similarity, particularly in the 5’ untranslated region (5’-UTR) with at least 23 genotypes of Pestivirus A being recognised (4, 7-10).

Due to the economic impact of this infection, several European nations, including Scotland, have ongoing control and/or eradication programmes. During such programmes, which mainly target bovine species, re-infection of previously BVD-free herds may occur, depending on the degree of biosecurity that can be maintained and the number of persistently infected (PI) animals remaining in the national cattle population. However, it is also possible that reinfection may occur from infected wildlife or non-bovine livestock species, particularly deer and small ruminants (11-14). Thus, the broad host range of BVD viruses (all even-toed ungulate species may be infected) may hinder the long-term success of BVD eradication. Indeed, with prevalence in pigs reported to be increasing internationally in recent years (15-17), the risk of spread among a wider range of livestock species should also be considered. In this situation, molecular epidemiology is a useful approach to finding sources of infection and routes of virus transmission (18-21).

The quasispecies nature of this RNA virus may contribute to interspecies transmission by providing genetic plasticity. It was shown previously that transmission of BVD from cattle to goats and between cattle and sheep led to nucleotide changes in the coding region of E2, the major receptor binding protein and host species determinant, and changes to the binding of monoclonal antibodies (22-24). Furthermore, in vitro experiments using cell cultures of bovine and ovine origin showed that such adaptive mutations were reversible within just 3 cell culture passages and were therefore most likely due to selection of pre-existing variants (25). Analysis of the spectrum of virus variants present in PI animals may reveal important information on the quasispecies dynamics involved in interspecies transmissions, and may help to reveal unusual reservoirs and transmission routes of BVD during the final stages of eradication campaigns (11,12). Thus, understanding quasispecies dynamics of pestiviruses and how this may give rise to variants or emerging viruses would enhance our understanding of interspecies transmission between wildlife and livestock species. Genome sequencing of BVD viruses has been possible for at least 20 years. Initial strains were sequenced using cDNA cloning and PCR (26-29), whilst more recent work has used reverse transcription (RT-) PCR and direct sequencing or next generation sequencing (NGS) (30-41). However, many pestivirus genome sequences were derived from isolates that were cultured before analysis and therefore do not necessarily represent the sequence diversity of the virus strain infecting its host. Few studies have taken advantage of NGS technology to look directly at quasispecies diversity in culture or *in vivo* (42) and none have looked at how the quasispecies diversity of BVD viruses changes during the initial passages of virus culture. In this paper we describe the direct sequencing of a Pestivirus A strain from serum and of the same strain after three passages in culture. In each case care was taken to remove potential sources of bias in the dataset, particularly avoiding amplification of the source material before construction of the sequencing libraries.

Methods

Viruses and culture

Laboratory strain NADL was obtained from the Moredun Virus Surveillance Unit and was grown on bovine turbinate cells in Iscove's modified Dulbecco's medium (IMDM), by inoculation at a multiplicity of infection (MOI) of 0.1. Viral cultures were expanded by serial subculture through three passages to produce sufficient material for virus concentration and RNA extraction.

A BVD PI cow (MRI103), which was held at the Scottish Centre for Production Animal Health & Food Safety, University of Glasgow for teaching purposes, was humanely euthanized by intravenous injection to prevent the spread of disease after confirmation of PI status. As part of the necropsy, samples of whole blood were collected and serum was stored at -20°C in aliquots of up to 50 ml. Serum samples were used as the source of virus for these studies after about two years of storage without thawing. The virus strain was characterized by RT-PCR and sequencing of 5'UTR and Npro gene fragments, which confirmed that this isolate was a BVDV1a genotype (43). This strain had also been used to infect rabbits, demonstrating that it could propagate *in vivo* and induce virus neutralizing antibodies (44). In order to define the effects of low-passage culture on the quasispecies diversity of BVD viruses, this isolate was passaged up to five times on bovine turbinate cells at MOI of 0.1.

RNA extraction and purification

Prior to RNA extraction, virus particles from serum and culture medium samples were concentrated. The uncultured MRI103 serum sample was concentrated by ultracentrifugation of 50 ml of serum at

120 000 ×g for 3h at 4 °C. The pelleted virus was resuspended in 1.3 ml of phosphate buffered saline (PBS).

The medium from cultured virus was first clarified by low speed centrifugation before concentration of virus particles by ultrafiltration using Centricon Plus-70 Centrifugal Filter Units, (Merck, 100 KDa cutoff). Approximately 350 ml of culture fluid was concentrated 100-fold before RNA extraction.

Concentrated virus particles were treated with RNase A before extraction of RNA using Trizol LS (Ambion, Life Technologies). The aqueous fraction of the Trizol lysate was precipitated with ethanol and was resuspended in 10 mM Tris pH 8.0 before purification using the QIAamp viral RNA mini kit (QIAGEN, Manchester, UK) in the absence of carrier RNA. Purified RNA was treated with RTS DNase (Cambio, Cambridge, UK) before cDNA synthesis.

Analysis of viral RNA

The degree of contamination of the viral RNA preparations with residual host or cell-culture material was analysed by real-time PCR and RT-PCR using a triplex assay for BVD (45), β-actin (46) and mitochondrial DNA (Table 1). These assays allowed the yield of viral RNA to be compared with potential host genomic or mitochondrial DNA contamination, respectively. The recovery of the full-length NADL viral genome was analysed by a set of long RT-PCR assays that covered the full length of the virus in either three or five fragments (not shown). The primer sequences for this analysis were not compatible with amplification of BVD strain MRI103, so the likely recovery of full-length RNA for this genome was inferred from NADL RNA purified in parallel.

cDNA synthesis

Initial experiments showed that priming cDNA synthesis with random primers led to a poorer yield of full-length cDNA, with clear reduction in representation of the 3'-end of the genome. In order to improve coverage of full-length genomes, two approaches were used. In the first approach, a Pestivirus A genotype 1a-specific primer (B12; Table 1) was used to prime cDNA synthesis from the 3'-end of the viral genome, whilst in the second approach, viral RNA samples were polyadenylated using polyA polymerase before cDNA synthesis primed by anchored oligo dT primers. In each case, synthesis of double-stranded cDNA was done using the Superscript III cDNA synthesis system (Life Technologies), with RNAseH, *E. coli* DNA polymerase and *E. coli* DNA ligase in the second strand mix.

The presence of double-stranded BVD-specific cDNA after synthesis was assayed qualitatively by end-point PCR before and after restriction endonuclease digestion of the cDNA, using primers that amplified either conserved regions of the virus (all samples) or the entire NADL genome as a set of overlapping amplicons. Restriction enzyme treatment of cDNA clearly reduced the yield of BVD PCR products (not shown).

Sequencing

Six viral cDNA samples were sequenced by The Genome Analysis Centre (TGAC; now Earlham Institute) as part of their Capacity and Capability Challenge (CCC) programme. The samples were extracted from the two viruses, NADL and MRI103, with MRI103 purified from serum and after three passages in culture (P3), and each cDNA was primed either with oligo-dT or with the BVD-specific 3'-UTR primer B12 (47). The cDNA samples were therefore named: NADL-dT; NADL-B12; MRI103-dT; MRI103-B12; MRI-P3-dT and MRI-P3-B12.

Library generation used the TruSeq CHiP protocol (Illumina) because of the small amounts of cDNA available (< 15 ng DNA per sample) and each library was tagged with barcoding primers to allow multiplex analysis of all six samples in a single MiSeq run with 250 base paired-end reads. Data was returned as FastQ format files.

Sequence assembly

The individual datasets were mapped against the bovine and pestivirus genomes to get an estimate of read coverage before assembly. An initial mapping of reads from the oligo-dT primed serum derived sequences to the Pestivirus A isolate SD-1 genome (26) was performed at low stringency using the program BWA (48). The preliminary consensus sequence derived from this analysis allowed the reference-led assembly of the MRI103 virus genome datasets.

Assembly of the sequence data for each cDNA sample was done using the DNASTAR Lasergene SeqMan Ngen program, using the preferences for a reference-led assembly of paired-end reads. Each MiSeq sequence assembly was then transferred from SeqMan Ngen to Seqman Pro for final quality control and SNP calling.

SNP analysis

Polymorphic positions within each consensus were called using the SNP Report feature within Seqman Pro and validated by inspection of the assembly. SNPs represented by a single base call within each assembly were discarded while all other SNP were retained for further analysis of diversity. Areas of low coverage (<100-fold) were not included in the analysis and SNPs within homopolymeric tracts at the end of reads were disregarded. The SNP reports were copied into MS Excel spreadsheets for further analysis. SNP positions were reported based on the numbering of the respective MRI103 serum or NADL (M31182) consensus sequence (29).

Sequence diversity within each assembly was measured as the Shannon Entropy (H) at each polymorphic base position, calculated as the negative sum of the product of each base call frequency and its natural log, as described by Milani and coworkers (49). Diversity for each sequence assembly was plotted as the distribution of single base entropy values or as the mean single nucleotide Shannon entropy within each of the individual coding regions.

Error rates within the sequence data were estimated from the base call quality (Phred) scores in the final edited assemblies. Average read length in the assembled data was 235 and mean read quality score was 36 with range 22-39. The minimum mean read quality score was therefore used to calculate a conservative threshold entropy score below which SNP were excluded as potential read errors. For a read quality value of 22, the estimated error rate was 0.63 %, with a corresponding H value of 0.0384. Thus only polymorphic positions with H >0.0384 were considered to be SNP in this analysis. In addition, overall entropy for each genome was calculated as the sum of all H values above the threshold value.

Genome analysis by PCR and Sanger sequencing

As an approach to deriving close to full-length Pestivirus A genome sequences from small amounts of virus material without propagation of virus, a panel of 24 primer pairs was designed based on an alignment of 17 genome sequences representing BVD virus genotypes found in the UK (7 type 1a including MRI103, 8 type 1b, 1 type 1d, 1 type 1e; Table 1). Degenerate primers were designed using the program PrimerDesign-M (50), with the following parameters: maximum degeneracy 32-fold, primer length 20-25, overlap mid, Tm difference 5 °C. Primers were synthesised by Eurofins-MWG tagged with the SP6 promoter sequence on all forward primers and the T7 promoter sequence on all reverse primers as sequencing adaptors (Table 1).

In addition, sequences derived from amplification with the pan-pestivirus 5'UTR primers 324 and 326 (51) added approximately 56 nucleotides (nt) at the 5'end of each sequence. RNA from the original material used for next-generation sequencing of the serum and passage 3 cultured MRI103 virus was used as the template for amplification using the One-Step Ahead RT-PCR kit (Qiagen).

To demonstrate genome sequencing from small samples, RNA purified from 100 µl of MRI103 passage 5 (P5) culture medium was amplified as above. This also allowed the analysis of sequence changes introduced by continued culture.

The PCR products were sequenced using SP6 and T7 promoter primers (Eurofins MWG) and trace files were assembled using the DNASTAR Lasergene SeqMan program (DNASTAR Inc, Madison, WI, USA). All potential SNP sites were assessed visually and were considered valid if there was clear evidence of multiple base calls at the candidate site in two or more sequencing reads. Variant bases were recorded as UPPERCASE/lowercase (e.g. G/a) if all relevant reads showed the same majority base call and as International Union of Pure and Applied Chemistry (IUPAC) ambiguity codes (e.g. R for G/A) if the majority base could not be identified from the available reads.

Results

Viral RNA purification and cDNA synthesis

MRI103 viral RNA was purified from approximately 50 ml of serum or 330 ml of culture medium and double stranded cDNA was synthesised using either anchored oligo dT primers or a BVD-specific primer (B12, Table 1). A triplex real-time PCR, using primer-probe sets for BVD, mitochondrial DNA and genomic B-actin (Table 1), demonstrated that all of the cDNA samples contained BVD-specific cDNA which was enriched compared to genomic or mitochondrial DNA (data not shown).

Multiple end point PCR assays (with and without prior treatment by restriction enzyme *Mse*I) demonstrated the presence of NADL-specific double-stranded cDNA representing the complete genome (data not shown) and suggested that the MRI103 samples (for which the NADL-specific primers did not work) were also likely to contain full-length viral cDNA.

Quality control of six cDNA samples submitted to the UK Biotechnology and Biological Sciences Research Council (BBSRC) Genome Analysis Centre (TGAC; now Earlham Institute) suggested that the samples contained between 2 and 14 ng cDNA. All of each cDNA sample was used for library construction and after barcoding were combined into a single MiSeq run which yielded a minimum of 1.7 million read-pairs per sample.

Virus genome Sequencing

Sequencing reads from TGAC were obtained as FastQ files of reads that passed initial quality thresholds. Per-base quality showed that many reads dropped out of the optimal quality range (>28) after about position 200, so all reads were quality trimmed before assembly. Reads that were mainly of poor quality or which contained long homopolymeric stretches were also excluded from further analysis. The reads were analysed to ascertain source (bovine, BVD or unknown) and to estimate the number of reads available for genome assembly (Table 2). This showed that only a small proportion of reads were homologous to BVD viruses in each library (0.2 % to 5.3 % for MRI103) and that the sequence data obtained from the serum-derived virus had the lowest proportion of virus-specific reads. Further analysis was therefore done to establish whether the BVD-specific data was sufficient to produce a whole genome consensus.

As an initial test of assembly and to determine whether a consensus sequence could be produced for reference-led assembly of the MRI103 genome, each set of reads was aligned with the most relevant reference genome sequence using the BWA read alignment package (48). These were NADL (Accession M31182) for the NADL reads and strain SD-1 (M96751) for the MRI103 reads, based on analysis of the known 5'UTR and Npro sequences of MRI103. Alignment to NADL was done at high stringency, while alignment of MRI103 reads to the SD-1 sequence was done at lower stringency to obtain optimal coverage.

These data showed that coverage of NADL was complete with average read-depth >5000 from the oligo-dT primed cDNA, whilst read depth in the data from the B12-primed cDNA was approximately 100-fold lower. When the MRI103 reads were aligned with BVD isolate SD-1, the coverage was essentially complete in the oligo-dT primed cDNA but both coverage and read-depth were lower in the B12-primed cDNA. These data also showed that only the oligo-dT primed cDNA produced enough sequence data to cover either virus genome at read depth >100. A consensus sequence was derived from each alignment, representative of the sequencing data for NADL and MRI103. The NADL consensus was identical to the NADL reference genome and the MRI103 consensus (of 12,292 nucleotides) was 90 % identical to the SD-1 genome sequence.

All six datasets (NADL-dT; NADL-B12; MRI103-dT; MRI103-B12; MRI-P3-dT and MRI-P3-B12) were then used in reference-led assembly of the respective NADL or MRI103 genome sequences using DNASTAR Seqman Ngen. All datasets assembled to their respective reference sequence with high quality and good consistency in the placing of read pairs (supplementary data table 1). However, once again, the B12-primed cDNA datasets provided five- to ten-fold lower read depth than the oligo-dT primed cDNA data. All subsequent analyses were therefore performed using the data obtained from the oligo-dT primed cDNA samples. For each of these datasets, (NADL-dT; MRI103-dT; MRI-P3-dT), a consensus sequence was derived from the assembled reads and this was used to assess the degree of polymorphism within the virus populations assayed. Sequence coverage within the oligo-dT-primed datasets was as follows: MRI103-dT had median read depth of 190 but about 10% of the genome had read depth 50-100; MRI-P3-dT had median read depth of about 1800 and minimum read depth of just over 100. In contrast, the NADL-dT sequence dataset had read depth of over 5000 across almost the complete genome, except for the region corresponding to bases 1-24 of the reference sequence, within which read depth rose from 4 to >1000.

Comparison of MRI103 serum and culture derived sequences

The consensus MRI103 sequences derived by next-generation sequencing (NGS) from serum and cell culture were almost identical. The serum virus consensus was 12292 nt in length and the cultured virus was 12296 nt, excluding a 3' poly-A tail which started at the same position in both sequences and resulted from in vitro polyadenylation for oligo dT-primed cDNA synthesis. Among the 12292 aligned bases, the consensus sequence derived from the cultured virus differed at only 4 positions, compared with the serum consensus sequence (A/T, 1485; G/A, 1504; G/A, 2532; C/A, 2712; supplementary data table 2) all of which were non-synonymous (NS) changes, encoding distinct amino acids in glycoproteins Erns (1485, 1504) and E2 (2532, 2712).

To compare the NGS MRI103 genome sequences with genome sequences produced from PCR products, the first-strand cDNA used for NGS of the MRI103 serum and P3 viruses was used for genome amplification by RT-PCR using primers designed to amplify a range of Pestivirus A genotypes as a set of 24 overlapping PCR products (Table 1). The assembly of sequence data from overlapping PCR products, including the 5'UTR sequence amplified by primers 324-326, produced double-strand coverage of length 12043 nt, equivalent to nt 116-12159 of the MRI103 serum consensus. To further test the reliability of this approach, 100 µl of culture supernatant from passage 5 (P5) of MRI103 virus propagation was used as a source of viral RNA and 1 µl aliquots were used for RT-PCR. The P5 virus double-stranded sequence coverage was 12023 nt, twenty bases shorter than the serum and P3 sequences. A consensus sequence was derived from each RNA source for comparison with the NGS-derived data. Both PCR-derived consensus sequences were identical to the respective NGS-derived consensus and the P5 PCR consensus was identical to the P3 NGS consensus. The length of each PCR sequence was shorter than the NGS equivalent due to the positions of the leftmost and rightmost primer sites and the extent of reliable (double-stranded) coverage of the terminal fragments but each amplicon-derived consensus sequence included the entire polyprotein CDS.

Analysis of the MRI103 genome sequence by BLAST identified Pestivirus A strain Nose (AB078951) as the most similar complete genome currently available, with 93 % identity across 99 % of the genome. Translation of the MRI103 genome indicated a single continuous open reading frame (nt 372 - 12068 in the serum reference dataset) encoding 3898 amino acid residues which also had greatest similarity to strain Nose with 95 % amino acid identity.

SNP analysis

SNP calling was done using Seqman Pro and all SNP were verified by visual inspection. This approach has been used previously to improve the reliability of SNP verification, and is especially applicable to short genomes such as RNA viruses (52). Diversity within each NGS genome assembly was measured using Shannon entropy (H), which was calculated at each candidate polymorphic base position in each sequence assembly (serum, culture and NADL). The MRI103-dT (serum) sequence had 824 SNP positions, with H ranging from the calculated error threshold of 0.0384 to 0.73. In comparison, the

MRI-P3-dT culture assembly had 407 SNP positions, with values of H between 0.05 - 0.69; while the NADL-dT sequence had only 12 SNP positions with a much narrower H range of 0.10 to 0.23. Analysis of the distribution of H values in each sequence (Fig. 1) suggests that the number of low entropy variants reduced during culture, while the diversity of the remaining variants increased. This is supported by the observation that median entropy values increased (Fig. 1) from MRI103-dT serum (0.07), to MRI-P3-dT (0.11) to NADL-dT (0.14). In addition, Mann-Whitney U-test analysis of the two MRI103 distributions in Fig. 1 showed that they were significantly different from each other (P<0.0001). In addition, the total Shannon entropy in each genome (based on all positions above the error threshold) decreased with culture, from 92.9 in the serum virus, to 64.1 in the passage 3 culture virus and 1.8 in the NADL strain.

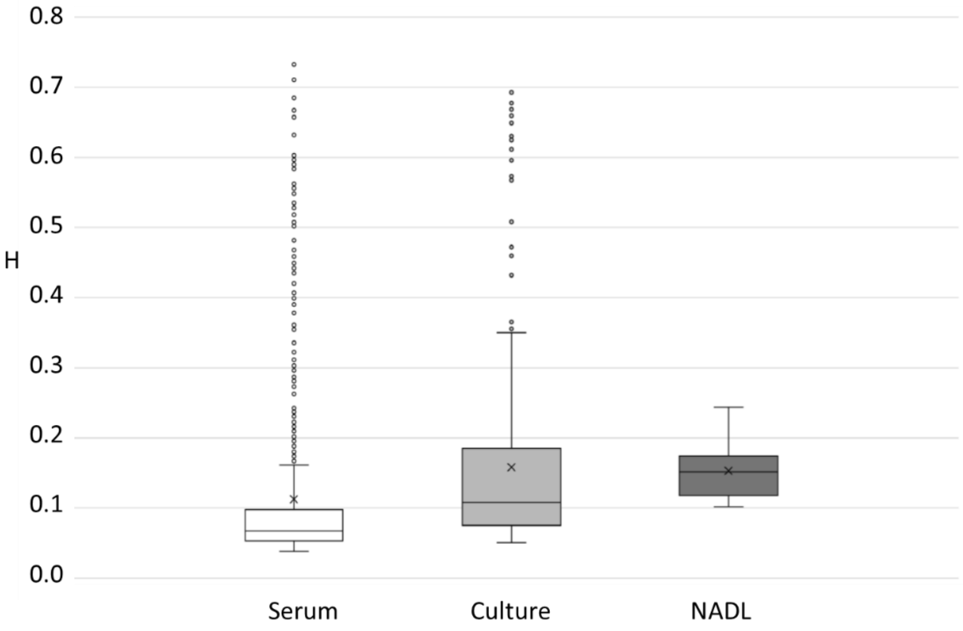

Figure 1. Boxplots of the Shannon Entropy values (H) for qualifying SNPs in each of the three genome sequences: MRI103-serum (open boxes, 824 SNP); MRI103-culture (light grey, 407 SNP) and NADL (dark grey, 12 SNP). For each plot the first and fourth quartiles are given as bounded vertical lines, the second and third quartiles are boxes separated by a line representing the median value, outliers are included as circles and the mean value is marked with a cross (x).

In these datasets, allele frequencies paralleled entropy values (supplementary data table 2), with MRI103-dT SNP ranging from 0.6 % to 46 %; MRI-P3-dT SNP ranging from 1 % to 41 % and NADL-dT ranging from 2.1 % to 6.3 %, in comparison with the respective consensus sequences. Indeed, where the consensus base changed in the MRI103 culture virus (positions 1485, 1504, 2532 and 2712), the frequencies of the alternate base were between 75 and 83 % (supplementary data table 2) relative to the serum consensus sequence.

Among the NGS data for MRI103, 681 SNP were found only in the serum sequence, 143 SNP positions were shared and 264 were found only in the culture sequence. Among the shared SNP, neither serum nor culture dataset had consistently higher allele frequencies suggesting that there was no consistent effect of culture passage on allele frequency.

The distribution of SNPs also varied across the MRI103 genome. Among the protein coding regions, SNPs were more frequent in the genes encoding the envelope proteins Erns, E1 and E2, compared to the non-structural genes (Fig. 2), in both the serum-derived virus and following propagation. Comparison of the Erns, E1 and E2 coding regions with the non-structural genes (NS2-NS5B) showed the glycoprotein gene sequences had significantly higher average single-base entropy values ($p<0.01$; Fig. 2) demonstrating a higher degree of polymorphism.

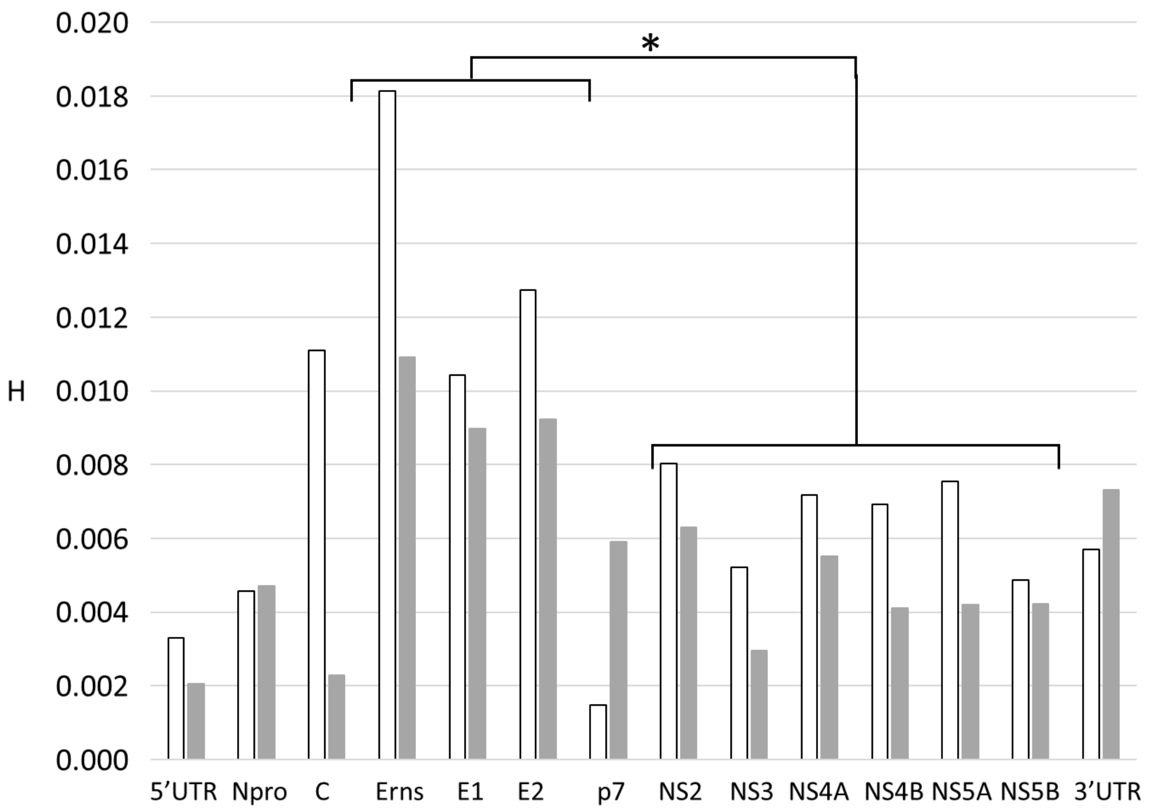

Figure 2. Distribution of sequence variation across the MRI103 genome. For each gene, the average single nucleotide entropy (H) was plotted. The columns are coloured as Fig 1: open = MRI103-serum; light grey = MRI103-culture. Student's T-test comparison of the glycoprotein genes (Erns, E1, E2) and the non-structural genes (NS2-NS5B) showed that average entropy among the glycoprotein genes was significantly higher than the NS genes for both the serum and the cultured virus (p<0.01) *.

It was also clear that both synonymous (S) and non-synonymous (NS) SNP were more frequent in the MRI103 serum dataset than the culture dataset (p<0.005; Student's T-test). There was also evidence to suggest that NS SNP in Erns, E1 and E2 were more frequent than in non-structural genes (NS2-NS5B). In the P3 culture dataset, NS SNP were more than twice as frequent in the envelope genes (p<0.01; Student's T-test), while in the serum dataset NS SNP were also more frequent but this difference was not significant (p=0.055; Student's T-test).

The SNP within the MRI103 datasets included 23 that encoded stop codons (19 in MRI103-dT and 4 in MRI-P3-dT, none shared; supplementary data table 2). All had minor allele frequencies below 3%, except for a single polymorphism (GAG>TAG nt 2283, within E1) that was found at 29% in the serum dataset.

Quasispecies variation observed in PCR derived sequence

Examination of the sequence trace data from the PCR-derived genome assemblies also suggested that some quasispecies variation could be detected. Although each Sanger-derived consensus PCR sequence was identical to the respective NGS consensus, the PCR data also contained evidence of multiple base calls at many positions in the trace data.

The trace data was therefore analysed more closely to identify sites that exhibited clear quasispecies variation. Each base position within the PCR derived sequences that contained evidence of multiple base calls in at least two reads was recorded as either ambiguous (two base calls of approximately equal peak height, denoted by IUPAC ambiguity codes in supplementary data table 2); or as one major and one minor base call (one base having a higher peak in all reads, denoted as uppercase/lowercase (e.g. G/a) in supplementary data table 2).

Examples of sequencing traces from polymorphic positions (including two that represent consensus base changes between serum and cultured viruses) are shown in Fig. 3 and relate to SNP positions detailed in supplementary data table 2. The traces show either reducing diversity (nt 1504), diversity throughout (nt 2532) or increasing diversity (nt 4263) with culture passage. Position 1504 was called as R (G/A) within the serum PCR sequence but A was called as the majority base in both P3 and P5 sequences. At position 2532, all three virus sequences had evidence of distinct base calls but the majority base call appeared to change from G in serum to A in culture. In contrast, at position 4263, the serum virus had A as the majority base call, while the minority G call became more prominent in P3 and was called as R in P5.

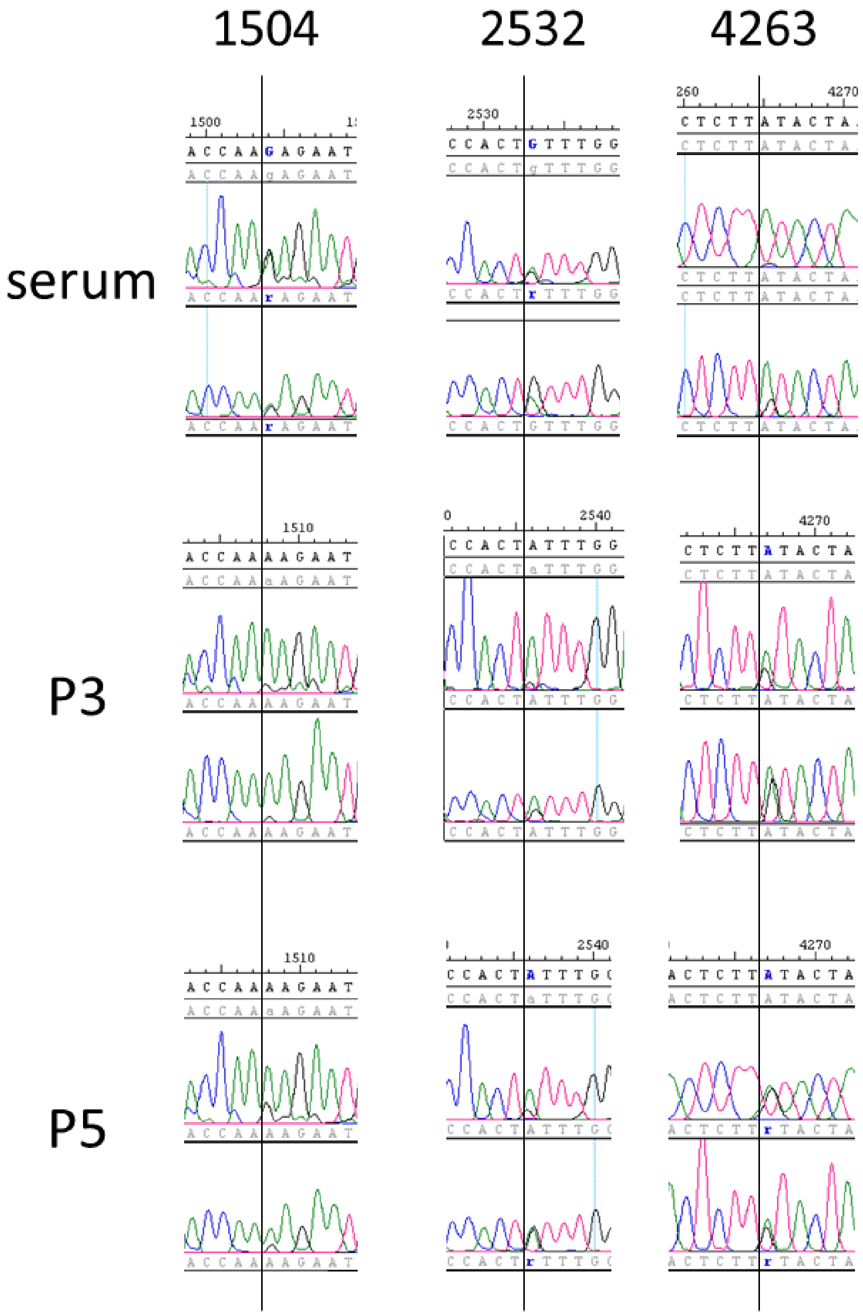

Figure 3. Example traces from PCR amplicon sequencing of MRI103 from serum, passage 3 (P3) and passage 5 (P5) of culture. In each example two traces from the equivalent position in each virus sequence are shown. The number above each set of traces refers to the SNP coordinates given in supplementary data table 2. The local consensus sequence and numbering is given in the top line of each panel, while base calls for each trace are given below the trace. Software-assigned potential SNP in each read and in the consensus are highlighted in blue. Additional lines containing read data were collapsed to simplify the figure.

Overall, in the serum PCR sequence, 58 positions were called as polymorphic, including 10 that were not seen in the serum NGS dataset. Among the remaining 48 positions, the corresponding SNP frequency in the serum NGS dataset varied from 4.9% to 46%. These included 68% of serum NGS SNP with frequency >10%.

Within the P3 culture PCR sequence, 33 positions were called as polymorphic, all of which were variable in the NGS data, with SNP frequency ranging from 5% to 80%. These positions included 80% of P3 culture NGS SNP with frequency >10%.

For the P5 culture PCR sequence, 57 positions were called as polymorphic. In the absence of a P5 NGS dataset, these were compared with the P3 NGS data. This showed that 12 SNP were not found in the P3 culture NGS data. Corresponding SNP frequencies for the remaining 45 SNP positions in the MRI-P3-dT dataset varied from 1.4% to 91%, including 73% of P3 culture NGS SNP with frequency >10%. The P5 PCR dataset was clearly distinct from the serum and P3 sequences and contained 17 SNP at positions that had variant frequencies below 5% in the P3 culture NGS dataset. This may reflect further changes in quasispecies structure between P3 and P5, with some variant bases becoming more highly represented in the P5 virus population.

Discussion

The number of complete genome or polyprotein sequences for pestiviruses that are publicly available now exceeds 100 for each of Pestivirus A and B (European Nucleotide Archive: www.ebi.ac.uk/ena). Of these, most were sequenced by Sanger methods after cloning or amplification of genome segments while many of the more recently determined sequences were defined by NGS approaches. In almost all cases, however, only the consensus sequences have been published and quasispecies variation has seldom been examined. For example, Workman and co-workers (36) defined genome sequences of 19 BVD virus strains directly from plasma. These strains were genotyped within Pestivirus A and B species but information on quasispecies variation within these strains was not presented or discussed. Similarly, Neil and co-workers (40) used NGS approaches to sequence 36 virus isolates representing a range of Pestivirus B (BVDV type 2) strains but this work did not examine quasipecies diversity.

The quasispecies nature of RNA viruses may facilitate their transmission between host species by providing a reservoir of variants for selection to operate on (53, 54). This process, whereby a subpopulation of viruses can infect and adapt to a different host, is extremely important for the emergence of novel or more virulent viruses and also may lead to vaccine or diagnostic “escape mutants”. Understanding the baseline quasispecies structure of RNA virus populations may allow the risks of acquiring a wider host range or increased virulence to be assessed in advance.

Specific studies of BVD virus quasispecies variation have generally focused on individual genes, by counting clones of PCR products reflecting sequence variation within an E1/E2 fragment (55) or in the E2 and NS5B genes (56). Other studies have compared the virus consensus sequences between sequentially-infected individuals, inferring the nature and rates of quasispecies variation between hosts (39).

Recent work described a PCR and NGS approach to define consensus sequences of major genes and infer detailed phylogenetic relationships among Canadian BVD virus strains, but without analysing quasispecies variation (57). The same group used RT-PCR amplification of complete genomes from lymph node tissue to study quasispecies diversity in different BVD strains (58). This work found more variation in structural genes, in accord with results published here (Fig. 2). However, this study did not include unamplified control material or examine the influence of PCR amplification on the datasets obtained.

In this study we used a PCR-free approach to derive NGS datasets for the MRI103 strain of Pestivirus A from serum and after three passages in culture, obtaining both consensus sequences and genome level quasispecies data. We also analysed the laboratory strain NADL. The serum MRI103 virus had the largest number of variant sites within its genome, while the P3 cultured and NADL genomes showed decreasing numbers of polymorphic sites. However, the median Shannon entropy value increased with culture (Fig. 2), suggesting that the remaining variants in cultured viruses had higher individual entropy values and were represented at higher frequency within the populations. The view that overall virus diversity decreases with culture was supported both by the number of variant sites in each genome (824 in serum, 407 in P3 and 12 in NADL) and by the total entropy in each genome (sums of 93 in serum, 64 in P3 and 1.8 in NADL). The summed entropy figures may give a more accurate picture of the loss of diversity in culture, reflecting both the number of sites above the error cutoff and the level of diversity at each site.

The observation of multiple stop codons at low frequency within the MRI103 NGS datasets and a single high frequency mutation within E1 (nt 2283; 29% T) in the serum virus, raises the question of their potential function. Genomes carrying stop codons may be defective or may express truncated gene products. Genomes with polyprotein truncation may enhance the survival of the population as a whole, or be rescued by translational frameshifting, the use of alternative start codons or by complementation in cells infected with multiple viruses. Some stop codon mutations may benefit persistent infection by reducing the expression of viral antigens (59, 60), although increasing the frequency of stop codon mutations in RNA viruses was associated with significant losses in viral fitness both *in vivo* and *in vitro* (61). The observations here, that 19 stop codon variants were found in the serum virus sequence, while only four stop codon variants were seen in the P3 culture virus sequence and that none of these were shared (supplementary data table 2), support the view that the stop codon variants did not confer an advantage in culture and were lost. It will be of interest to analyse further samples *ex vivo* to determine whether high frequency stop codon variants are a common feature of pestivirus quasispecies populations.

The four consensus base changes in the cultured virus sequence were also those that had the greatest change in SNP frequency across the entire genome (supplementary data table 2). The changed consensus bases encoded amino acid changes to the virus glycoproteins Erns (nt 1485=T102S; 1504=R108K) and E2 (2532=V62I; 2712=Q122K) and in each case the substitutions lay within the exposed domains of the proteins (reviewed in 62). For Erns, the altered residues were located between the catalytic domain (residues 32-81), which has RNase activity, and the C-terminal region (residues 167-227) which mediates membrane retention and processing. These residues have not been recorded as being associated with any recognised antibody epitope (62) and the nature of the changes (Thr-Ser; Arg-Lys) maintained the respective hydrophilic and basic character of the residues and therefore may not have significant effects on Erns structure-function.

In E2 the substitutions were in the two outermost (membrane-distal) domains that possess Ig-like folds and the substitution in the outermost domain (V61I) is close to a recognised neutralising motif in E2 of Classical Swine Fever Virus (residues 64-76), which has analogous structure (62). However, like Erns, the substitutions in E2 appeared to be conservative in nature and may not cause any clear change in structure or function of E2.

However the observation that mutations in the Erns glycoprotein gene can lead to failure of diagnostic assays implies that serum-derived and culture-derived samples of the same virus could give different results in such assays (63).

Although the overall degree of variation within the cultured virus was lower than the serum virus genome, it was clear from analysis of shared SNP that culture passage did not lead to a consistent reduction in the degree of variation of individual SNP, suggesting that culture may select for increased frequency of some SNP variants but not others. In this context we also note that the twelve variant positions within the NADL genome, with SNP frequencies of 2.1 - 6.3 %, did not appear to correspond with SNP positions in the MRI103 genome, when the two genome sequences were aligned (supplementary data table 2). This may suggest that extended culture exerts selective pressure *in vitro* on sites that are not under such pressure *in vivo*.

In addition, Sanger sequencing of MRI103 RT-PCR products produced the same virus consensus sequences as the NGS analysis and identified many polymorphic sites, suggesting that PCR amplification has potential for use in quasispecies analysis. While this may be more error-prone that an amplification-free approach, it clearly has value where the amount or complexity of input material makes direct sequencing difficult. The use of PCR and clone counting approaches (55, 56) to assess quasispecies variation supports the view that read-counting within NGS data from such PCR products may yield accurate quantitative data. This suggests that analysis of quasispecies variation in pestiviruses (and potentially other RNA viruses) could focus on uncultured samples from infected individuals rather than after virus isolation, potentially taking advantage of amplification methods to derive genome level quasispecies data.

From the Sanger data, it appeared that the quasispecies variants detected in the MRI103 P5 virus included a number that were derived from low-frequency variants in the P3 virus and others that were not detected in the serum or P3 datasets. This appears to be in keeping with the suggestion above that culture exerts selective pressure on different sites from those selected *in vivo*, in addition to reducing the overall quasispecies variation in the cultured virus population.

Drivers that may act to fix sequence changes include selection pressures resulting from infection of new cell types within a host, of new individuals or of new species. Where available, data on cross-species infection suggests that success may be facilitated by the presence of ‘pre-adapted’ variants that could be selected and rapidly established in the new host (23, 54). The appearance of host-specific sequence changes in goats, sheep and cattle serially infected with BVD viruses supports this view (22-24). In addition, NGS studies of bird-to-mammal transmission of avian influenza viruses indicates that transmission to a new host creates genetic bottlenecks that select variants with suitable adaptations (53, 54). The existence of virus sub-populations within individuals is also supported by studies of HIV, where distinct, organ-specific populations were found and the differences between populations increased with time (64).

The sequencing of the NADL strain here showed that it had identical consensus sequence to the reference genome sequence determined in 1988 (29). Additionally, the NADL dataset had a lower number of variant base positions than MRI103 P3 (12 compared to 407), but these few variants had higher median diversity than either MRI103 dataset (Fig. 1; supplementary data table 2). Thus, this highly culture-adapted virus had few variant positions but these appeared to be maintained stably.

The PCR-based Sanger sequence data obtained here reflected the quasispecies variation detected by NGS, but SNPs with frequency <10% in the corresponding NGS dataset appeared to be more difficult to detect, reflecting the more qualitative nature of Sanger sequencing. The application of NGS analysis to PCR amplicon sequencing could be an easy and cost-effective way to gather information on quasispecies variation and allow evolution within chains of infection to be monitored from even small serum or tissue samples. Such a forensic use of sequencing could also be applied in situations such as the late stages of BVD eradication campaigns, where single locus sequencing may not distinguish similar viruses shared across multiple premises from true transmission events. However, it is clear that for such forensic epidemiology we will also need to improve our understanding of virus sequence variation within a chain of infection and between animals infected by the same source. The recently described Scottish BVD Biobank (43) includes such samples that could be analysed to improve our understanding of what defines a single Pestivirus A strain.

While it is clear that the amplification-free process used here, offering a no- or low-bias protocol for producing RNA virus templates for NGS, may provide a gold-standard approach for analysis of quasispecies diversity, it has the disadvantage of requiring a high volume of input material. This may however be alleviated by ongoing technical improvements in the generation of quantitative NGS datasets from small amounts of RNA. In addition, the use of PCR as a first step in quasispecies analysis needs further validation and comparison with direct NGS. Analysis of potential problems, such as uneven coverage, especially at the ends of target viruses, will help critically evaluate this potentially valuable approach. Previous work (65) used PCR to amplify the FMDV genome prior to NGS library preparation to investigate quasispecies variation in the inoculum virus and in lesions from same animal. While this approach accurately replicated the consensus inoculum reference sequence, it did not address potential limitations of the approach and there was no NGS-only dataset for comparison. It also demonstrated that the high read-depth achievable by using NGS of PCR products allowed definition of quasispecies variants with frequencies below 1%.

Conclusion

This work provides evidence that initial culture of a BVD virus isolate drives virus evolution, producing changes in the virus consensus sequence. Here, four clear changes in MRI103 virus consensus sequence, encoding amino acid sequence changes in the Erns and E2 glycoproteins, were found after only 3 passages in culture, while 15 novel variant sites were detected in the virus after 5 passages. This was accompanied by a clear reduction in quasispecies diversity with culture, as defined by number of variant sites and total entropy, suggesting that genome sequences derived after virus isolation may not present an accurate picture of the quasispecies variation *in vivo* or the true virus consensus sequence. The very low total entropy of variant sites in the NADL genome analysed here suggests that extensive culture works to maintain a genome sequence that is well-adapted to the tissue culture environment.

While this study has demonstrated clear changes in quasi-species structure after only three passages in culture, additional experiments could enhance the future value of this work. Repetition of the culture passages and quasispecies analysis would provide useful information on the reproducibility of the changes induced by culture, while infection of BVD-naive animals with the cultured virus could help define whether the changes were reversible *in vivo*. Further analysis of genetic variation within culture-adapted viruses, in homologous and heterologous culture systems, may also help us understand how this low level of genetic variation is maintained. In any case, the clonal nature of the NADL strain compared to the impressive diversity of the virus in a BVD PI animal implies that using highly passaged lab strains to model natural infection may not be representative.

We also showed that PCR amplification of viral RNA produces amplicons in which many viral polymorphisms appear to be conserved. Further quantitative analysis, with and without PCR amplification, will help define how closely genome sequence diversity in PCR amplicons reproduces that obtained directly from viral RNA. However, the data produced here suggest that even Sanger sequence analysis of PCR-amplified virus genome fragments can reflect the quasispecies variation found in the source material, at least for more frequent variants. The combination of PCR amplification and NGS analysis of the quasispecies structure of virus genomes may be a powerful approach to allow investigation of samples that would otherwise be difficult or impossible to study.

Author contributions

George C. Russell: conceptualisation, investigation, validation, formal analysis, resources, data curation, writing (original draft preparation; review and editing), visualisation, supervision, project administration, funding

Ruth N. Zadoks: conceptualisation, writing (review and editing), visualisation, supervision

Kim Willoughby: conceptualisation, resources, writing (review and editing), supervision

Claudia Bachofen: conceptualisation, investigation, validation, formal analysis, resources, data curation, writing (original draft preparation; review and editing), visualisation, funding

Conflicts of interest

The authors declare that there are no conflicts of interest.

Funding information

This work was funded by an early mobility grant to CB from the Swiss National Science Foundation and by a Moredun Innovation Fund grant to GCR. RNZ and GCR were additionally funded by The Scottish Government as part of the EPIC Centre of Expertise on Animal Disease Outbreaks.

Acknowledgements

We thank The Genome Analysis Centre (now Earlham Institute) for the provision of sequencing facilities under their Capacity and Capability Challenge programme (project CCC-123). We are grateful to staff of the Scottish Centre for Production Animal Health & Food Safety, University of Glasgow for facilitating the collection of BVD-positive blood samples. We are also grateful to the Moredun Virus Surveillance Unit for the provision of strains and support.

References

1. Peterhans E, Schweizer M. BVDV: A pestivirus inducing tolerance of the innate immune response. Biologicals. 2013;41(1):39-51.
2. Eigen M, Schuster P. Hypercycle - Principle of Natural Self-Organization .A. Emergence of Hypercycle. Naturwissenschaften. 1977;64(11):541-65.
3. Andino R, Domingo E. Viral quasispecies. Virology. 2015;479-480(Supplement C):46-51.
4. Booth R, Thomas C, El-Attar L, Gunn G, Brownlie J. A phylogenetic analysis of Bovine Viral Diarrhoea Virus (BVDV) isolates from six different regions of the UK and links to animal movement data. Veterinary Research. 2013;44(1):43.
5. Courtenay AE, Henderson RG, Cranwell MP, Sandvik T. BVD virus type 2 infection and severe clinical disease in a dairy herd. Veterinary Record. 2007;160(20):706-7.
6 Smith DB, Meyers G, Bukh J, Gould EA, Monath T, Muerhoff AF, Pletnev A, Rico-Hesse R, Stapleton JT, Simmonds P, Becher P. Proposed revision to the taxonomy of the genus Pestivirus, family Flaviviridae. Journal of General Virology. 2017; 98(8):2106-2112.
7. Vilcek S, Durkovic B, Kolesarova M, Paton DJ. Genetic diversity of BVDV: Consequences for classification and molecular epidemiology. Preventive Veterinary Medicine. 2005;72(1-2):31-5.
8. Vilcek S, Paton DJ, Durkovic B, Strojny L, Ibata G, Moussa A, et al. Bovine viral diarrhoea virus genotype 1 can be separated into at least eleven genetic groups. Archives of Virology. 2001;146(1):99-115.
9. Yeşilbağ K, Alpay G, Becher P. Variability and Global Distribution of Subgenotypes of Bovine Viral Diarrhea Virus. Viruses. 2017;9(6):128.
10. Bazzucchi M, Bertolotti L, Giammarioli M, Rossi E, Petrini S, Rosati S, et al. Complete Genome Sequence of a Bovine Viral Diarrhea Virus Subgenotype 1g Strain Isolated in Italy. Genome announcements. 2017;5(17):e00263-17.
11. Nettleton PF, Herring A, Corrigall W. Isolation of bovine virus diarrhoea virus from a Scottish red deer. Veterinary Record. 1980;107:2.
12. Fernandez-Aguilar X, Lopez-Olvera JR, Marco I, Rosell R, Colom-Cadena A, Soto-Heras S, et al. Pestivirus in alpine wild ruminants and sympatric livestock from the Cantabrian Mountains, Spain. Veterinary Record. 2016;178(23):586-U57.

13. Frolich K, Thiede S, Kozikowski T, Jakob W. A review of mutual transmission of important infectious diseases between livestock and wildlife in Europe. Domestic Animal/Wildlife Interface: Issue for Disease Control, Conservation, Sustainable Food Production, and Emerging Diseases. 2002;969:4-13.

14. Ridpath JF, Neill JD. Challenges in Identifying and Determining the Impacts of Infection with Pestiviruses on the Herd Health of Free Ranging Cervid Populations. Frontiers in Microbiology. 2016;7.

15. Deng Y, Sun C-Q, Cao S-J , Lin T, Yuan S-S, Zhang H-B, Zhai S-L, Huang L, Shan T-L, Zheng H, Wen X-T, Tong G-Z. High prevalence of bovine viral diarrhea virus 1 in Chinese swine herds. Veterinary Microbiology. 2012; 159(3–4);490-493.

16. Tao J, Liao J, Wang Y, Zhang X, Wang J, Zhu G. Bovine viral diarrhea virus (BVDV) infections in pigs. Veterinary Microbiology. 2013; 165(3–4);185-189.

17. Chakraborty AK, Mukherjee P, Karam A, Das S, Lit Barkalita L, Puro K, Sanjukta R, Ghatak S, Sakuntala I, Laha RG, Borah P, Ngachan SV, Sharma I, Sen A. Evidence of BVDV in Pigs from North Eastern Part of India - Genetic Profiling and Characterisation. Open Virology Journal. 2018; 12; 110-120.

18. Rossmanith W, Jackova A, Appel F, Wilhelm E, Vilcek S. Analysis of BVDV isolates and factors contributing to virus transmission in the final stage of a BVDV eradication program in lower Austria. Berl Munch Tierarztl Wochenschr. 2014;127(1-2):12-8.

19. Ståhl K, Kampa J, Baule C, Isaksson M, Moreno-López J, Belák S, et al. Molecular epidemiology of bovine viral diarrhoea during the final phase of the Swedish BVD-eradication programme. Preventive Veterinary Medicine. 2005;72(1–2):103-8.

20. Stalder H, Hug C, Zanoni R, Vogt H-R, Peterhans E, Schweizer M, et al. A nationwide database linking information on the hosts with sequence data of their virus strains: A useful tool for the eradication of bovine viral diarrhea (BVD) in Switzerland. Virus Research. 2015;218:49-56.

21. Giangaspero M, Yesilbag K, Apicella C. Who's who in the Bovine viral diarrhea virus type 1 species: Genotypes L and R. Virus Research. 2018; 256: 50-75

22. Paton D, Gunn M, Sands J, Yapp F, Drew T, Vilcek S, et al. Establishment of serial persistent infections with bovine viral diarrhoea virus in cattle and sheep and changes in epitope expression related to host species. Archives of Virology. 1997;142(5):929-38.

23. Bachofen C, Vogt H, Stalder H, Mathys T, Zanoni R, Hilbe M, et al. Persistent infections after natural transmission of bovine viral diarrhoea virus from cattle to goats and among goats. Veterinary Research. 2013;44(32):15-.

24. Kuca T, Passler T, Newcomer BW, Neill JD, Galik PK, Riddell KP, et al. Identification of Conserved Amino Acid Substitutions During Serial Infection of Pregnant Cattle and Sheep With Bovine Viral Diarrhea Virus. Frontiers in Microbiology. 2018;9.

25. Meier P. Studies on the Molecular Evolution of Bovine Viral Diarrhea Virus (BVDV): University of Berne; 1998.

26. Deng R, Brock KV. Molecular cloning and nucleotide sequence of a pestivirus genome, noncytopathic bovine viral diarrhea virus strain SD-1. Virology. 1992;191(2):867-79.

27. Moerlooze LD, Lecomte C, Brown-Shimmer S, Schmetz D, Guiot C, Vandenbergh D, et al. Nucleotide sequence of the bovine viral diarrhoea virus Osloss strain: comparison with related viruses and identification of specific DNA probes in the 5′ untranslated region. Journal of General Virology. 1993;74(7):1433-8.

28. Meyers G, Tautz N, Becher P, Thiel HJ, Kümmerer BM. Recovery of cytopathogenic and noncytopathogenic bovine viral diarrhea viruses from cDNA constructs. Journal of Virology. 1996;70(12):8606-13.

29. Colett MS, Larson R, Gold C, Strick D, Anderson DK, Purchio AF. Molecular cloning and nucleotide sequence of the pestivirus bovine viral diarrhea virus. Virology. 1988;165(1):191-9.

30. Xie Z, Fan Q, Xie Z, Liu J, Pang Y, Deng X, et al. Complete Genome Sequence of a Bovine Viral Diarrhea Virus Strain Isolated in Southern China. Genome Announcements. 2014;2(3).

31. Oem J-K, Joo S-K, An D-J. Complete Genome Sequences of Two Bovine Viral Diarrhea Viruses Isolated from Brain Tissues of Nonambulatory (Downer) Cattle. Genome Announcements. 2013;1(5).
32. Gao S, Du J, Shao J, Lang Y, Lin T, Cong G, et al. Genome analysis reveals a novel genetically divergent subgenotype of bovine viral diarrhea virus in China. Infection, Genetics and Evolution. 2014;21:489-91.
33. Soltan MA, Wilkes RP, Elsheery MN, Elhaig MM, Riley MC, Kennedy MA. Complete Genome Sequence of Bovine Viral Diarrhea Virus-1 Strain Egy/Ismailia/2014, Subtype 1b. Genome Announcements. 2015;3(6).
34. Stalder H, Schweizer M, Bachofen C. Complete Genome Sequence of a Bovine Viral Diarrhea Virus Subgenotype 1e Strain Isolated in Switzerland. Genome Announcements. 2015;3(3).
35. Workman AM, Harhay GP, Heaton MP, Grotelueschen DM, Sjeklocha D, Smith TPL. Full-Length Coding Sequences for 12 Bovine Viral Diarrhea Virus Isolates from Persistently Infected Cattle in a Feedyard in Kansas. Genome Announcements. 2015;3(3):e00487-15.
36. Workman AM, Heaton MP, Harhay GP, Smith TPL, Grotelueschen DM, Sjeklocha D, et al. Resolving Bovine viral diarrhea virus subtypes from persistently infected U.S. beef calves with complete genome sequence. Journal of Veterinary Diagnostic Investigation. 2016;28(5):519-28.
37. Neill JD, Bayles DO, Ridpath JF. Simultaneous rapid sequencing of multiple RNA virus genomes. Journal of Virological Methods. 2014;201:68-72.
38. Neill JD, Newcomer BW, Marley SD, Ridpath JF, Givens MD. Genetic change in the open reading frame of bovine viral diarrhea virus is introduced more rapidly during the establishment of a single persistent infection than from multiple acute infections. Virus Research. 2011;158(1-2):140-5.
39. Neill JD, Newcomer BW, Marley SD, Ridpath JF, Givens MD. Greater numbers of nucleotide substitutions are introduced into the genomic RNA of bovine viral diarrhea virus during acute infections of pregnant cattle than of non-pregnant cattle. Virol J. 2012;9:150.
40. Neill JD, Workman AM, Hesse R, Bai J, Porter EP, Meadors B, et al. Identification of BVDV2b and 2c subgenotypes in the United States: Genetic and antigenic characterization. Virology. 2019;528:19-29.
41. Sato A, Tateishi K, Shinohara M, Naoi Y, Shiokawa M, Aoki H, et al. Complete Genome Sequencing of Bovine Viral Diarrhea Virus 1, Subgenotypes 1n and 1o. Genome Announcements. 2016;4(1).
42. Jenckel M, Hoper D, Schirrmeier H, Reimann I, Goller KV, Hoffmann B, et al. Mixed Triple: Allied Viruses in Unique Recent Isolates of Highly Virulent Type 2 Bovine Viral Diarrhea Virus Detected by Deep Sequencing. Journal of Virology. 2014;88(12):6983-92.
43. Russell GC, Grant DM, Lycett S, Bachofen C, Caldow GL, Burr PD, et al. Analysis of bovine viral diarrhoea virus: Biobank and sequence database to support eradication in Scotland. Veterinary Record. 2017;180(18):447-.
44. Bachofen C, Grant DM, Willoughby K, Zadoks RN, Dagleish MP, Russell GC. Experimental infection of rabbits with bovine viral diarrhoea virus by a natural route of exposure. Veterinary Research. 2014;45:34.
45. Willoughby K, Valdazo-Gonzalez B, Maley A, Gilray J, Nettleton R. Development of a real time RT-PCR to detect and type ovine pestiviruses. Journal of Virological Methods. 2006;132(1-2):187-94.
46. Thonur L, Maley M, Gilray J, Crook T, Laming E, Turnbull D, et al. One-step multiplex real time RT-PCR for the detection of bovine respiratory syncytial virus, bovine herpesvirus 1 and bovine parainfluenza virus 3. BMC Veterinary Research. 2012;8:9.
47. Vilcek S, Paton D, Lowings P, Bjorklund H, Nettleton P, Belak S. Genetic analysis of pestiviruses at the 3 ' end of the genome. Virus Genes. 1999;18(2):107-14.
48. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 2009;25(14):1754-60.
49. Milani A, Fusaro A, Bonfante F, Zamperin G, Salviato A, Mancin M, et al. Vaccine immune pressure influences viral population complexity of avian influenza virus during infection. Veterinary Microbiology. 2017;203:88-94.

50. Yoon H, Leitner T. PrimerDesign-M: a multiple-alignment based multiple-primer design tool for walking across variable genomes. Bioinformatics. 2015;31(9):1472-4.

51. Vilček Š, Herring AJ, Herring JA, Nettleton PF, Lowings JP, Paton DJ. Pestiviruses isolated from pigs, cattle and sheep can be allocated into at least three genogroups using polymerase chain reaction and restriction endonuclease analysis. Archives of Virology. 1994;136(3):309-23.

52. Kong HR, Anthony NB, Rowland KC, Khatri B, Kong BC. Genome re-sequencing to identify single nucleotide polymorphism markers for muscle color traits in broiler chickens. Asian-Australasian Journal of Animal Sciences 2018; 31(1): 13-18.

53. Moncla Louise H, Zhong G, Nelson Chase W, Dinis Jorge M, Mutschler J, Hughes Austin L, et al. Selective Bottlenecks Shape Evolutionary Pathways Taken during Mammalian Adaptation of a 1918-like Avian Influenza Virus. Cell Host & Microbe. 2016;19(2):169-80.

54. Varble A, Albrecht Randy A, Backes S, Crumiller M, Bouvier Nicole M, Sachs D, et al. Influenza A Virus Transmission Bottlenecks Are Defined by Infection Route and Recipient Host. Cell Host & Microbe. 2014;16(5):691-700.

55. Ridpath JF, Bayles DO, Neill JD, Falkenberg SM, Bauermann FV, Holler L, et al. Comparison of the breadth and complexity of bovine viral diarrhea (BVDV) populations circulating in 34 persistently infected cattle generated in one outbreak. Virology. 2015;485:297-304.

56. Dow N, Chernick A, Orsel K, van Marle G, van der Meer F. Genetic Variability of Bovine Viral Diarrhea Virus and Evidence for a Possible Genetic Bottleneck during Vertical Transmission in Persistently Infected Cattle. PLoS ONE. 2015;10(7):e0131972.

57. Chernick A, van der Meer F. Evolution of Bovine viral diarrhea virus in Canada from 1997 to 2013. Virology. 2017;509:232-8.

58. Chernick A, Ambagala A, Orsel K, Wasmuth JD, van Marle G, van der Meer F. Bovine viral diarrhea virus genomic variation within persistently infected cattle. Infection, Genetics and Evolution. 2018;58:218-23.

59. Tong S-P, Li J-S, Vitvitski L, Trépo C. Replication capacities of natural and artificial precore stop codon mutants of hepatitis B virus: Relevance of pregenome encapsidation signal. Virology. 1992;191(1):237-45.

60. Cassini R, De Mitri MS, Gibellini D, Urbinati L, Bagaglio S, Morsica G, et al. A novel stop codon mutation within the hepatitis B surface gene is detected in the liver but not in the peripheral blood mononuclear cells of HIV-infected individuals with occult HBV infection. J Viral Hepatitis. 2013;20(1):42-9.

61. Moratorio G, Henningsson R, Barbezange C, Carrau L, Bordería AV, Blanc H, et al. Attenuation of RNA viruses by redirecting their evolution in sequence space. Nature Microbiology. 2017;2:17088.

62. Wang FI, Deng MC , Huang YL, Chang CY. Structures and Functions of Pestivirus Glycoproteins: Not Simply Surface Matters. Viruses. 2015; 7: 3506-3529.

63. Gripshover EM, Givens MD, Ridpath JF, Brock KV, Whitley EM, Sartin EA. Variation in Erns viral glycoprotein associated with failure of immunohistochemistry and commercial antigen capture ELISA to detect a field strain of bovine viral diarrhea virus. Veterinary Microbiology. 2007;125(1):11-21.

64. Sanjuan R, Codonnter FM, Moya A, Elena SF. Natural selection and the organ-specific differentiation of HIV-1 v3 hypervariable region. Evolution. 2004;58(6):1185-94.

65. Wright CF, Morelli MJ, Thebaud G, Knowles NJ, Herzyk P, Paton DJ, et al. Beyond the consensus: dissecting within-host viral population diversity of foot-and-mouth disease virus by using next-generation genome sequencing. J Virol. 2011;85(5):2266-75.

Data Bibliography

1 European Nucleotide Archive (www.ebi.ac.uk/ena/)

Figure legends

Figure 1. Boxplots of the Shannon Entropy values (H) for qualifying SNPs in each of the three genome sequences: MRI103-serum (open boxes, 824 SNP); MRI103-culture (light grey, 407 SNP) and NADL (dark grey, 12 SNP). For each plot the first and fourth quartiles are given as bounded vertical lines, the second and third quartiles are boxes separated by a line representing the median value, outliers are included as circles and the mean value is marked with a cross (x).

Figure 2. Distribution of sequence variation across the MRI103 genome. For each gene, the average single nucleotide entropy (H) was plotted. The columns are coloured as Fig 1: open = MRI103-serum; light grey = MRI103-culture. Student’s T-test comparison of the glycoprotein genes (Erns, E1, E2) and the non-structural genes (NS2-NS5B) showed that average entropy among the glycoprotein genes was significantly higher than the NS genes for both the serum and the cultured virus ($p<0.01$) *.

Figure 3. Example traces from PCR amplicon sequencing of MRI103 from serum, passage 3 (P3) and passage 5 (P5) of culture. In each example two traces from the equivalent position in each virus sequence are shown. The number above each set of traces refers to the SNP coordinates given in supplementary data table 2. The local consensus sequence and numbering is given in the top line of each panel, while base calls for each trace are given below the trace. Software-assigned potential SNP in each read and in the consensus are highlighted in blue. Additional lines containing read data were collapsed to simplify the figure.

Table 1. Primers used in this work

Primer name	Sequence	Target	Source
real-time PCR primers			
mitochondrial DNA analysis			
BOVmit-f	ATTACCATGCCGCGTGAAA	bovine mitochondrial genome, EN006853	this work
BOVmit-r	TAGCGACCCCCACGGTTTA	bovine mitochondrial genome, EN006854	this work
BOVmit-MGBp	VIC-CCCGCTAGGCAGGGA	bovine mitochondrial genome, EN006855	this work
actin			
Bac1F_uni	GACAGGATGCAGAARGAGATCAC	mammalian beta actin	(59)
Bac2R_uni	TCCACATCTGCTGGAAGGTG	mammalian beta actin	(59)
BacMGB_mammal	NED-TGAAGATCAAGATCATCG	mammalian beta actin	(59)
BVDV			
BVDV1_2F	CCATGCCCTTAGTAGGACTAGC	BVDV1 + 2	(40)
BVDV1_2R	TGACGACTACCCTGTACTCAGG	BVDV1 + 2	(40)
BVDV1MGB	6FAM-AACAGTGGTGAGTTCGT	BVDV1	(40)
Generic BVDV primers			
324	ATGCCCWTAGTAGGACTAGCA	pan-pestivirus	(60)
326	TCAACTCCATGTGCCATGTAC	pan-pestivirus	(60)
B12	TAGACGTCGGGTRTATCCTC	NADL, BVDV 1a	(41)

BVDV1 genome primers $

Primer name	Sequence	Target	Source
BVDV1-F1	*GAGTTCGTTGGATGGCTKAAGC	BVDV1	this work
BVDV1-R1	*GCRTCCATCATACCRTGRATTG	BVDV1	this work
BVDV2-F1	*TACCACATTTAYGTGTGYATWGATG	BVDV1	this work
BVDV2-R1	*GCTATTTCAAARTTGCARGG	BVDV1	this work
BVDV3-F1	*TACAAGAYAATGGRACRGAAGG	BVDV1	this work
BVDV3-R1	*TCATAAGGCCACCAATYYGGTC	BVDV1	this work
BVDV4-F1	*CCYTGCAAYTTTGAAATAGC	BVDV1	this work
BVDV4-R1	*TCAAAGTYRTCRTYCATTTC	BVDV1	this work
BVDV5-F1	*ATTGGTGGCCTTATGARACARC	BVDV1	this work
BVDV5-R1	*GTCTTTTCYACHGGCCCCTCAC	BVDV1	this work
BVDV6-F1	*ACTTTGARTTYGGWCTYTGCCC	BVDV1	this work
BVDV6-R1	*ARYARGTTGCCCATCATCAC	BVDV1	this work
BVDV7-F2	*AGTGAGGGGCCDGTRGAAAAGAC	BVDV1	this work
BVDV7-R2	*TYARGAAGTCRGCCCAYARGG	BVDV1	this work
BVDV8-F1	*GATGATGGGCAACYTRYTRAC	BVDV1	this work

BVDV8-R1	*CTRTGCATGTAGTAYATAAAGTC	BVDV1	this work
BVDV9-F1	*CTGATGRYYAGYTATGTRACAG	BVDV1	this work
BVDV9-R1	*CCYACCATRAKCATYTTTAC	BVDV1	this work
BVDV10-F1	*GACTTTATRTACTACATGCAYAG	BVDV1	this work
BVDV10-R1	*ACRCCATAYTCTGTCTCATC	BVDV1	this work
BVDV11-F1	*AAAGTAAARATGMTYATGGTRGG	BVDV1	this work
BVDV11-R1	*GCTATRAATTCYTCTATYGGGTG	BVDV1	this work
BVDV12-F1	*ATGGTYAAGAAGATWACCAG	BVDV1	this work
BVDV12-R1	*TAGTGRWAGTCCTTTGACCC	BVDV1	this work
BVDV13-F1	*CACCCRATAGARGAATTYATAGC	BVDV1	this work
BVDV13-R1	*GGTGGGTKGTGTCYTCTAGTC	BVDV1	this work
BVDV14-F1	*GGTCAAAGGACTWYCACTAYGAC	BVDV1	this work
BVDV14-R1	*GGGGTRAACATTTTYARTGC	BVDV1	this work
BVDV15-F1	*AGACTAGARGACACMACCCAC	BVDV1	this work
BVDV15-R1	*CCAATTGTCTGDACTGCTTC	BVDV1	this work
BVDV16-F1	*TATTATGTKATSAAYAARCCYTC	BVDV1	this work
BVDV16-R1	*CTTTCATCTSATARCCACATGG	BVDV1	this work
BVDV17-F2	*AAARYCCAGARAAAATHATAATGGC	BVDV1	this work
BVDV17-R2	*ACCTCCCAYTTRTCAGTRGC	BVDV1	this work
BVDV18-F1	*CATGTGGYTATSAGATGAAAGC	BVDV1	this work
BVDV18-R1	*AARGTCCCCYTGAARTCRAC	BVDV1	this work
BVDV19-F1	*AAAATGAARAARGGSTGTGC	BVDV1	this work
BVDV19-R1	*TTGGTARGCTGATGCSATRTG	BVDV1	this work
BVDV20-F1	*TTCYATRTCAAATAGRGCWAARAC	BVDV1	this work
BVDV20-R1	*TCTATCTTATCYCTDATYGCCTC	BVDV1	this work
BVDV21-F1	*GGGCAYATSGCATCAGCYTACC	BVDV1	this work
BVDV21-R1	*GCACTKGTGTCTGGYTGRCC	BVDV1	this work
BVDV22-F1	*GCWATACCAAARAAYGARAARAGAG	BVDV1	this work
BVDV22-R1	*CTTGAATCCARTCTTGTWGCC	BVDV1	this work
BVDV23-F1	*GGCACAARTTCATYGAYACCATYAC	BVDV1	this work
BVDV23-R1	*TGGCCRGTTTTGCTRGATAYC	BVDV1	this work
BVDV24-F1	*ATGGCWACAAGAYTGGATTCAAG	BVDV1	this work
BVDV24-R1	*ACAGCTAAAGTGCTGTGTGC	BVDV1	this work

$ Pestivirus A type 1 primer sets were designed based on an alignment of BVDV genome sequences as follows: type 1a strains 180, 8844, GS5, NADL, SD1, SingerArg and MRI103; accession numbers HQ174292, HQ174293, KJ541471, M31182, M96751, DQ088995 & LR699799; type 1b strains 526, 8824, CC13B, CP7-5A, IBSP4ncp, KE9, Osloss, ILLNC, accession numbers HQ174294, HQ174295, KF772785, AF220247, KJ620017, EF101530, M96687 and U86600; type 1d strain 10JJ-SKR, KC757383 and type 1e strain Carlito, KP313732.

* Primers were 5'-tagged with either SP6 (F primers) or T7 (R primers) promoter sequences for plate sequencing

SP6 promoter	CATTTAGGTGACACTATAG
T7 promoter	TAATACGACTCACTATAGGG

Table 2. BVD virus NGS measures.

Number of reads, quality control data and mapping to *Bos taurus* and BVD virus genomes. Total read numbers and percentages are given for each sequencing library.

Library	reads	homopolymeric	failed quality	mapped B. taurus	mapped BVDV	unmapped
NADL-dT	3.6 x10^6	7.5 x10^5	2.6 x10^5	1.7 x10^6	6.9 x10^5	1.6 x10^5
NADL-B13	4.7 x10^6	1.0 x10^4	2.5 x10^5	3.2 x10^6	4.2 x10^5	8.2 x10^5
MRIserum-dT	3.6 x10^6	7.8 x10^5	2.3 x10^5	2.4 x10^6	1.2 x10^4	2.0 x10^5
MRIserum-B12	5.1 x10^6	4.7 x10^3	2.7 x10^5	4.5 x10^6	9.3 x10^3	2.5 x10^5
MRI-P3-dT	3.5 x10^6	1.2 x10^6	2.2 x10^5	1.6 x10^6	1.8 x10^5	2.0 x10^5
MRI-P3-B12	5.3 x10^6	2.3 x10^4	2.9 x10^5	4.0 x10^6	2.2 x10^5	8.0 x10^5

Library	reads	% homopol	% failed qual	% B. taurus	%BVDV
NADL-dT	3.59 x10^6	21.0	7.4	48.1	19.1
NADL-B13	4.74 x10^6	0.2	5.3	68.2	8.8
MRIserum-dT	3.58 x10^6	21.8	6.3	65.9	0.3
MRIserum-B12	5.06 x10^6	0.1	5.4	89.4	0.2
MRI-P3-dT	3.49 x10^6	35.6	6.3	47.2	5.3
MRI-P3-B12	5.32 x10^6	0.4	5.5	74.8	4.2