Near-complete genome sequencing of swine vesicular disease virus using the Roche GS FLX sequencing platform

Nielsen, Sandra Cathrine Abel; Bruhn, Christian Anders Wathne; Samaniego Castruita, Jose Alfredo; Wadsworth, Jemma; Knowles, Nick J.; Gilbert, Tom

Published in:
PLoS ONE

DOI:
10.1371/journal.pone.0097180

Publication date:
2014

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

Citation for published version (APA):
Nielsen, S. C. A., Bruhn, C. A. W., Samaniego Castruita, J. A., Wadsworth, J., Knowles, N. J., & Gilbert, M. T. P. (2014). Near-complete genome sequencing of swine vesicular disease virus using the Roche GS FLX sequencing platform. DOI: 10.1371/journal.pone.0097180
Near-Complete Genome Sequencing of Swine Vesicular Disease Virus Using the Roche GS FLX Sequencing Platform

Sandra C. Abel Nielsen1*, Christian A. W. Bruhn1*, Jose Alfredo Samaniego1, Jemma Wadsworth2, Nick J. Knowles2*, M. Thomas P. Gilbert1*

1 Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark, 2 Vesicular Disease Reference Laboratory Group, The Pirbright Institute, Pirbright, Surrey, United Kingdom

Abstract

Swine vesicular disease virus (SVDV) is an enterovirus that is both genetically and antigenically closely related to human coxsackievirus B5 within the Picornaviridae family. SVDV is the causative agent of a highly contagious (though rarely fatal) vesicular disease in pigs. We report a rapid method that is suitable for sequencing the complete protein-encoding sequences of SVDV isolates in which the RNA is relatively intact. The approach couples a single PCR amplification reaction, using only a single PCR primer set to amplify the near-complete SVDV genome, with deep-sequencing using a small fraction of the capacity of a Roche GS FLX sequencing platform. Sequences were initially verified through one of two criteria; either a match between a de novo assembly and a reference mapping, or a match between all of five different reference mappings performed against a fixed set of starting reference genomes with significant genetic distances within the same species of viruses. All reference mappings used an iterative method to avoid bias. Further verification was achieved through phylogenetic analysis against published SVDV genomes and additional Enterovirus B sequences. This approach allows high confidence in the obtained consensus sequences, as well as provides sufficiently high and evenly dispersed sequence coverage to allow future studies of intra-host variation.

Citation: Nielsen SCA, Bruhn CAW, Samaniego JA, Wadsworth J, Knowles NJ, et al. (2014) Near-Complete Genome Sequencing of Swine Vesicular Disease Virus Using the Roche GS FLX Sequencing Platform. PLoS ONE 9(5): e97180. doi:10.1371/journal.pone.0097180

Editor: Junwen Wang, The University of Hong Kong, Hong Kong

Received December 17, 2013; Accepted April 16, 2014; Published May 9, 2014

Copyright: © 2014 Nielsen et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This study was supported by Danish National Research Foundation grant DNRF94 and Lundbeck Foundation grant R52-A5062. Furthermore, The Pirbright Institute was funded by the European Reference Laboratory for Swine Vesicular Disease (NJK), core funding under the Livestock Viral Diseases Programme provided by the Biotechnology and Biological Sciences Research Council (NJK) and a Department for Environment, Food and Rural Affairs, UK, project SE2940 (NJK & JW). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: mtgilbert@gmail.com (MTPG); nick.knowles@pirbright.ac.uk (NJK)
† These authors contributed equally to this work.

Introduction

Swine vesicular disease (SVD) is a highly contagious viral disease of pigs. The first outbreak of SVD was observed in 1966 in Italy, although at the time, it was misidentified as foot-and-mouth disease (FMD) [1]. Shortly after, the causative agent, SVD virus (SVDV), was shown to be an enterovirus [1–3] and classified as a member of the Enterovirus B species within the Picornaviridae family. It is a non-enveloped virus with a single-stranded RNA genome of positive polarity of approximately 7.4 kb in length [4,5], and shares the same physico-chemical properties of other enteroviruses [1,3]. Previous studies have confirmed that SVDV is antigenically related to the human pathogen coxsackievirus B5 (CV-B5) [6,7], leading Graves to propose that SVD and its causative agent most originated from infection of pigs with a human CV-B5. A study comparing the complete nucleotide sequence of two SVDVs and one CV-B5, reported that CV-B5 is more closely related to SVDV than to other coxsackie B viruses in the regions encoding the major capsid proteins 1B, 1C and 1D (also known as VP2, VP3 and VP1, respectively) [8]. Another study by Zhang et al. [9] investigated the molecular evolution of SVDV. Here, they analyzed sequence data from 42 SVDV isolates and seven CV-B5 isolates in two shorter regions of the viral genomes; the regions encoding the capsid protein VP1 and the non-structural 3BC corresponding to 283 and 205 amino acids, respectively. They showed that SVDV formed a single monophyletic group, which was clearly distinct from any included CV-B5 as well as other coxsackieviruses included in the study. Furthermore, for each of these regions, the date for which SVDV is believed to have originated in pigs was estimated to have been between 1945 and 1965.

Due to improvements in DNA sequencing technologies [10], and hence the ease with which sequencing data can be generated, sequencing and characterization of viral genomes has accelerated. Because SVDV is of ongoing veterinary concern, we anticipate future studies will have an interest in obtaining increased numbers of SVDV coding-region sequences. Therefore, we developed a simple method to reconstruct near-complete SVDV genomes (encompassing the complete coding region) that is suitable for application to SVDV isolates in which RNA is largely intact. The method is based on PCR amplification using only a single primer set to generate the near-complete SVDV genome, which is performed after first-strand cDNA synthesis and coupled to
second-generation sequencing using the Roche GS FLX series of platforms, and is demonstrated on five SVDV isolates originating from Hong Kong during the 1970s. Sequence authenticity was validated in a multi-tiered approach (see Figure 1). Firstly through one of two criteria; either a match between a de novo assembly and a reference mapping, or a match between all of five different reference mappings performed against a fixed set of starting reference genomes with significant genetic distances within the same species of viruses. All reference mappings used an iterative method to avoid bias. Secondly, final validation was achieved through phylogenetic analysis against published SVDV sequences and additional Enterovirus B viruses using the VP2-VP3-VP1 genome region. The success of the method suggests it will be a useful tool for future analyses into the origin and spread of SVDV.

Figure 1. Decision diagram for validation of consensus sequences. The diagram illustrates the logic of the applied methodology for obtaining consensus sequences validated for further analysis. The process begins with de novo assembly as described in the methods section. The starting point from there is the blue-outline box in the top left hand corner. Positive answers follow the green "YES" arrows, negative ones follow the red "NO" arrows, and grey arrows are followed in all cases. Termination in a red box should lead to thorough analysis of upstream sources of error, including everything from contaminated samples to late stage in silico problems. Arriving at the first green box means that the consensus sequence of assembly/mapping has been verified. Arriving at the final green box means that the sample sequence is fully validated, now also with regard to sample provenance. The sample sequence is now ready to be used for further scientific analysis.

doi:10.1371/journal.pone.0097180.g001
Table 1. Results of assemblies and mappings of five Hong Kong swine vesicular disease virus (SVDV) isolates.

| SVDV isolate | HKN/8/73 | HKN/18/74 | HKN/25/75 | HKN/1/77 | HKN/5/77 |
|--------------|----------|-----------|-----------|----------|----------|
| Passage history | IB-RS-2 passage 1 | IB-RS-2 passage 1 | IB-RS-2 passage 1 | IB-RS-2 passage 2 | IB-RS-2 passage 2 |
| Harvest date | 13 May 1975 | 29 Oct 1974 | 24 March 1975 | 7 March 1977 | 7 March 1977 |
| GenBank acc.no. | KF963276 | KF963277 | KF963278 | KF963274 | KF963275 |
| Total # reads | 4287 | 5972 | 7078 | 7095 | 9985 |
| Total # unique reads | 4003 | 5489 | 6515 | 6497 | 9190 |

**Unique read lengths**
- Mean: 164.2
- Std. dev.: 111.2
- Max.: 526

**De novo reads assembled**
- Coverage: 3217
- Min.: 28
- Max.: 152
- Mean: 75.4
- Std. dev.: 20.8

**Reads mapped against UKG/27/72 (Acc. no. X54521)**
- Coverage: 3981
- Min.: 50
- Max.: 172
- Mean: 93.5
- Std. dev.: 18.8

**Reads mapped against SVDV HK’70 (Acc. no. AY429470)**
- Coverage: -
- Min.: -
- Max.: -
- Mean: -
- Std. dev.: -

**Reads mapped against CV-B5 ‘FAULKNER’ (Acc. no. AF114383)**
- Coverage: -
- Min.: -
- Max.: -
- Mean: -
- Std. dev.: -

**Reads mapped against CV-B3 ‘Woodruf’ (Acc. no. CXUS7056)**
- Coverage: -
- Min.: -
- Max.: -
- Mean: -
- Std. dev.: -

**Reads mapped against CV-B6 ‘Schmitt’ (Acc. no. AF105342)**
- Coverage: -
- Min.: -
- Max.: -
- Mean: -
- Std. dev.: -

**Number of polymorphic sites across all consensus sequences**
(1: 1)
0 (-)
0 (-)
0 (-)
The following reagents were added to a total of 20 μL of RNA/primer/dNTP mix previously published with forward primer: 5'-CACC-GAGTTGAGGAGATTTAGGC-3' and reverse primer: 5'-CACC-GAAAYGCGGAKAATTTACCAC-3' [12].

Sample fragmentation and preparation for export from the Pirbright Laboratory

DNA concentration was quantified using a ND-1000 spectrophotometer (NanoDrop Technologies, Thermo Scientific, Wilmington, DE) prior to fragmentation using NEBNext dsDNA Fragmentase (New England Biolabs, Ipswich, MA, USA). Fragmentation was performed as follows: 3 μg DNA from each sample was added to 1× Fragmentase Reaction buffer, 1× BSA, and nuclease-free water ad hoc to 54 μL. The reaction mix was subsequently vortexed thoroughly and incubated on ice for 5 min. Three units of dsDNA Fragmentase were added to the reaction and incubated 15 min at 37°C in order to generate fragments sizes of 600–800 bp. The incubation was stopped by addition of 5 μL 0.5 M EDTA. Samples were purified using Qiagen's PCR purification kit (Qiagen, Valencia, CA, USA) according to manufacturer’s guidelines and eluted in 30 μL EB buffer.

Correct fragment sizes were verified on the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) using a DNA7500 chip. Prior to sample export, 1/10 volume of sodium acetate (3 M) was added as well as 2.5× volumes of absolute ethanol (calculated after addition of sodium acetate). Samples were incubated for 2 hours at 56°C and the tubes washed with disinfectant (FAM diluted 1/100).

Ethanol precipitation

Amplions from Pirbright containing sodium acetate and absolute ethanol were centrifuged at 14,000 g for 1 hour at 4°C. Supernatant was carefully removed and discarded, leaving a DNA pellet. Pellets were dissolved and rinsed in 150 μL ice-cold ethanol (70%) and afterwards centrifuged again for 15 mins. Supernatant was discarded and pellet was dried for 10 min at 65°C before being dissolved in 85 μL EB buffer.

High-throughput deep sequencing

Independent sequencing libraries were produced on each sample using New England Biolabs’ NEBNext DNA Sample Prep Master Mix Set 2. Samples were subsequently pooled and sequenced as just under half of 1/8 of a FLX 454 Titanium plate.

Processing of data and sequence assembly/mapping

Sequence reads and their corresponding quality scores were imported and merged into Geneious (v.6.0.3 created by Biomatters, available from http://www.geneious.com). Reads were separated by Multiplex Identifier (MID) barcodes allowing no mismatches. Further, only unique reads were kept (see Table 1 for the number of total and unique reads for each sample following separation by barcode).

For the sake of validation and to avoid bias being introduced into the mapping process by the choice of reference sequence, a multi-tiered approach was chosen to obtain consensus sequences, and is illustrated in detail in Figure 1. For each sample de novo assembly was attempted using the Geneious assembler (v.6.0.3) [13] starting with the lowest standard sensitivity setting, and if this failed to deliver a full-length contig the sensitivity was incremented following the standard settings. If a full-length contig was obtained, the veracity of the de novo consensus sequence was substantiated by performing an iterative mapping against one or more starting reference sequences (the starting reference sequence is only used in the first round in the iterative mapping process, whereas all subsequent rounds use the consensus from the contig of the previous round as reference), and ensuring that the two consensus sequences matched (see Table 1 and Figure 1). If no full-length contigs were obtained in the de novo assembly, multiple iterative mappings against starting references of varying genetic distance (within the same species but including different serotypes, see Table 1) were performed using the Geneious Read Mapper.
Figure 2. Phylogenetic validation of samples. Using a maximum likelihood approach as described in Methods on the 1B-1C-1D genome region, corresponding to the outer capsid proteins VP2, VP3 and VP1, final validation was obtained for the samples. The five Hong Kong SVDVs isolated in the 1970s and which were sequenced in this study, are shown in red and all other SVD virus isolates are shown in pink. Coxsackievirus B5 isolates are shown in blue, and other Enterovirus B serotypes are shown in black. As expected from the previous literature, all CV-B5 together with all SVDV form a monophyletic cluster. This is supported with a bootstrap proportion of 100/100. Within this cluster all the SVD virus isolates, including those sequenced in this study, form a monophyletic cluster with bootstrap support of 96/100. Additionally, none of the five presently sequenced Hong Kong isolates show any obvious aberrations with regards to their position in the topology concerning either geographical information or branch lengths (vs. age). The tree is rooted on the branch leading from CV-B5 to all other CV-B sequences, and all branch labels show support in the form of bootstrap proportions.
bootstrap proportions out of one hundred. Support values for nodes with minor importance regarding the verification are not shown.

consensus sequences, the differences between consensus sequence. All consensus sequences were called using a allowed for mappings to converge to a stable (and thus unbiased) iterative mapping feature, ensuring that sufficient iterations were resulting contig of the iterative mapping process. Thus they are a strength of the method, as they represent the statistics of the final resulting assembly. The consensus sequences have been deposited in NCBI's GenBank under reference accessions KF963274- KF963278. Further, sequence data are accessible through EBI's European Nucleotide Archive (PRJEB5810). For direct access: http://www.ebi.ac.uk/ena/data/view/PRJEB5810.

Assembly and mapping statistics are given in Table 1. The strong similarity for the statistics between mappings against different starting references for the same sample is a sign of the strength of the method, as they represent the statistics of the final resulting contig of the iterative mapping process. Thus they are a reflection of the fact that the mappings have all converged. This is further reflected in the lack of difference in the consensus sequences. For the samples where de novo assembly was successful at full-length assembly, there are very few differences between the consensus sequence from these and those from the mappings (see Table 1) with a maximum of two polymorphic sites between all assembly/mappings for any sample. This level of accord should serve as a guide when applying the current methodology.

Phylogenetic reconstruction using a Maximum Likelihood approach

At nearly three times the length of alignments typically employed for VP1 based phylogenies, our alignment can be considered highly informative. The phylogenetic tree (Figure 2), rooted on the branch between the CV-B5’s and all other CV-B’s, shows all CV-B5 sequences and all SVDV sequences forming a highly supported (bootstrap proportion 96/100) monophyletic cluster, with the SVDV sequences themselves forming a highly supported (bootstrap proportion 100/100) monophyletic cluster within this. This conforms to the currently accepted evolutionary relationship of the strains for this genomic region. Furthermore, the topology shows no apparent aberrations with regards to the geographical location or age of isolates vs. branch lengths with regards to the five SVDV samples.

Discussion

The principal aim of this study was to develop a simple and rapid workflow for the sequencing of near-complete SVDV genome sequences, under the assumption that the samples contain suitable quality RNA for initial first-strand cDNA synthesis. With many samples available in cultures in reference banks such as that held at the Pirbright Institute, this is not an unlikely expectation. A limitation of the method, however, is that as it is based upon a single long-range PCR reaction, it will not be suitable for samples in which the SVDV RNA molecules are not near-complete length, thus in those situations other methods should be considered. Nevertheless, for isolates were the genome is intact, the primer set presented here will allow for the rapid generation of the complete coding genome.

The aim of this study was purely methodological, thus we do not wish to dwell on the phylogenetic results. However, the analysis demonstrates that all sequences cluster, as expected, among published SVDV sequence data, and overall the phylogenetic relationship for the full dataset is consistent with that previously generated using the VP1 gene sequences by Zhang et al. [9]. The data furthermore shows that assembly and mapping of the viral genomes and use of the validation strategy employed provides unbiased consensus sequences and is thus suitable for obtaining reliable results. Furthermore, contigs and assemblies have a sufficiently high and evenly distributed coverage to make them suitable for intra-host studies, should this be desired.

Concluding remarks

In this study, near-complete genomes, encompassing complete coding regions, of five cultured SVDV isolates were generated in a simple and rapid manner. Given the known existence of large numbers of similarly cultured samples in archives, and known questions about the time and source of origin of SVDV as a livestock pathogen, we anticipate that our method may be of use for future studies that wish to address this question in more detail.

Acknowledgments

We thank the Danish National High-throughput DNA Sequencing Centre for technical assistance.
**Author Contributions**
Conceived and designed the experiments: SCAN CAWB NJK MTPG.  
Performed the experiments: SCAN JW.  
Analyzed the data: CAWB JAS  
Contributed reagents/materials/analysis tools: NJK MTPG.  
Wrote the paper: SCAN CAWB NJK MTPG.  
Obtained permission for use of cell line: NJK.

**References**
1. Nardelli I, Lodetti E, Gualandi GL, Burrows R, Goodridge D, et al. (1968) A foot and mouth disease syndrome in pigs caused by an enterovirus. Nature 219: 1275–1276.
2. Davison GN, Darbyshire JH, Humfrey JF (1972) Differentiation of a vesicular disease of pigs in Hong Kong from foot-and-mouth disease. Veterinary Record 90: 618–621.
3. Dawe PS, Forman AJ, Smaile CJ (1973) A preliminary investigation of the swine vesicular disease epidemic in Britain. Nature 241: 540–542.
4. Inoue T, Suzuki T, Sekiguchi K (1989) The complete nucleotide sequence of swine vesicular disease virus. J Gen Virol 70: 919–934.
5. Seechurn P, Knowles NJ, McCauley JW (1990) The complete nucleotide sequence of a pathogenic swine vesicular disease virus. Virus Res 16: 255–274.
6. Graves JH (1973) Serological relationship of swine vesicular disease virus and Coxsackie B3 virus. Nature 245: 314–315.
7. Brown F, Talbot P, Burrows R (1973) Antigenic differences between isolates of swine vesicular disease virus and their relationship to Coxsackie B3 virus. Nature 245: 315–316.
8. Zhang G, Wilsden G, Knowles NJ, McCauley JW (1993) Complete nucleotide sequence of a coxsackie B3 virus and its relationship to swine vesicular disease virus. J Gen Virol 74: 845–853.
9. Zhang G, Haydon DT, Knowles NJ, McCauley JW (1999) Molecular evolution of swine vesicular disease virus. J Gen Virol 80: 639–651.
10. Metzker ML (2009) Sequencing technologies — the next generation. Nat Rev Genet 11: 31–46. doi:10.1038/nrg2626.
11. De Castro MP (1964) Behaviour of the foot-and-mouth disease virus in cell cultures: susceptibility of the IB-RS-2 cell line. Arq Inst Biol Sao Paulo: 63–78.
12. Nielsen SCA, Mourier T, Baandrup U, Soland TM, Bertelesen MF, et al. (2012) Probable Transmission of Coxsackie B3 Virus from Human to Chimpanzee, Denmark. Emerg Infect Dis 18: 1163–1165. doi:10.3201/eid1807.111609.
13. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, et al. (2012) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28: 1647–1649. doi:10.1093/bioinformatics/bts199.
14. Katoh K, Misawa K, Kuma K-I, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30: 3052–3060.
15. Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. Nat Methods 9: 772–772. doi:10.1038/nmeth.2109.
16. Akaike H (1974) A new look at the statistical model identification. IEEE Trans Automat Contr 19: 716–723. doi:10.1109/TAC.1974.1100705.
17. Hurvich CM, Tsai C-L (1989) Regression and time series model selection in small samples. Biometrika 76: 297–307. doi:10.1093/biomet/76.2.297.
18. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic Biol 52: 696–704. doi:10.1080/10635150390235520.