BMC Genomics

RESEARCH

Open Access

Comparison of whole genome sequencing approaches for *Capripox* viruses

Floris C. Breman[1*], Stefan Hoffman[2], Andy Haegeman[1], Wannes Philips[1], Sigrid de Keersmaecker[2], Patrick Mileto[3], Matthew J. Neave[3], Timothy R. Bowden[3], Tirumala B.K. Settypalli[4], Charles E. Lamien[4], Carrie Batten[5], Caroline Wright[5] and Nick De Regge[1]

Abstract

Background Direct whole genome sequencing of *Capripox* virus genomes from diagnostic samples is not always straightforward. Low viral content in a sample, biased sequencing and subsequent assembly and mapping methods may all influence the outcome.

Methods In this study we have tested and compared six next generation sequencing approaches on a homogenized skin sample from a bull infected with LSDV. We compared enrichment vs. non-enrichment strategies, different library preparation methods, short read Illumina sequencing with long read sequencing methods.

Results We found that methods that use an unbound transposon during tagmentation produced unbalanced results and lower target read yield versus methods that use other approaches to the tagmentation step. We further find that the use of hybrid capture probes increased the number of target reads. The result of subsequent mapping and assembly steps are influenced by the choice of reference when using reference-based assembly approaches.

Conclusions When using a short read sequencing approach we advise to use a transposon free method or a method with bound transposons for DNA fragmentation. These methods outperform kits that employ free transposons for DNA fragmentation when targeting AT-rich genomes. When mapping the reads it is best to use a reference for assembly that is as closely related as possible to the sample under study. Mapping problems can be resolved by long read sequencing which we recommend for denovo whole genome sequencing. Pacific Bioscience based long read sequencing outperforms Oxford Nanopore sequencing because it is less error prone. The ONT approach used, displays the same bias as the transposon based approach from Illumina and is therefore less suitable when attempting (Capri)pox whole genome sequencing.

Keywords Capripox, Lumpy skin disease virus, Whole genome sequencing, AT-bias, Direct sequencing

*Correspondence:
Floris C. Breman
floris.breman@sciensano.be
[1]Sciensano (Belgium), Service of 'Exotic and Vector borne diseases' (ExoVec), European reference laboratory for Capripox viruses, Brussels, Belgium
[2]Sciensano (Belgium), unit 'Transversal activities in Applied Genomics' (TAG), Brussels, Belgium
[3]CSIRO Australian Centre for Disease Preparedness (ACDP), 5 Portarlington Rd, East Geelong, VIC 3219, Australia
[4]Animal Production and Health Laboratory, Joint FAO/IAEA Centre of Nuclear Techniques in Food and Agriculture, Department of Nuclear Sciences and Applications, International Atomic Energy Agency, 100, A-1400, Vienna, Austria
[5]The Pirbright Institute (UK), Ash Road, Pirbright, Woking GU24 0NF, United Kingdom

Introduction[1]

LSDV and context

Whole genome sequencing is very important in establishing relationships within Poxvirus studies as well as reliably establishing the precise strain responsible for outbreaks [1–7]. However, many publications have rather short material and methods sections with regards to their sequencing approach. Often effective read depth (reads per position) of the target virus is lacking, the distribution of viral reads across a reference genome (breadth) is also lacking as are the particular kits and conditions applied. This makes the quality of the work hard to asses and replicate. With the recent outbreaks of poxviruses such as *Mpox* [8], *Suipox* [9], avian pox [10] and *Myxoma*, [11] there is a need for clear and reliable methods of genome sequence recovery.

Two circulating poxviruses that recently caused outbreaks and have a large economic impact [12] for livestock holders are Lumpy Skin Disease Virus (LSDV) and Sheeppox Virus (SPPV), both belonging to the genus *Capripox*. Sheeppox is endemic in Asia and Africa and recent outbreaks occurred in the European Union (EU) [6]. Lumpy skin disease virus is currently distributed and dispersing across Asia and Africa (described in [13]). and additional reports of LSDV outbreaks or new strains have been published [14–17]. The SPPV and especially the LSDV outbreaks have led to a surge in research into these viruses. One domain that has especially received increased attention is the use of whole genome sequencing to gain more comprehensive insight in circulating strains and disease epidemiology. Analyses of WGS's has led to the establishment of a phylogenetic framework for LSDV and SPPV [5, 12]. These frameworks will enable research to be done in an evolutionary context and to start looking for genomic changes which influence important traits such as virulence, transmission capacity, host specificity and others. *Capripox* viruses are large dsDNA viruses of 150 kb and obtaining WGS can be difficult. Difficulties in obtaining complete sequences from tissue involve the high amount of non-target (host) DNA in the samples compared to the low number of target reads, and the fact that the viral genomes are AT-rich [18], which may lead to problems during sequencing when using DNA library kits with unbound transposons for DNA fragmentation [19, 20]. The problem of a low number of target reads in a tissue sample can be overcome by growing the virus on a cell culture, deeper sequencing or the use of (large amplicon) PCR based methods such as developed by [21]. All of these approaches have their drawbacks. Deeper sequencing is more expensive and will not result in a better distribution of reads across a reference genome. Growing the virus in a cell culture requires a specialized high containment lab, specialized personnel, and is associated with high costs and loss of time. Growth on a culture also bears with it the risk of introducing mutations although this aspect is less important for the relatively slowly evolving poxviruses. Using PCR enrichment can be very effective, but requires multiple reactions which take time, increase costs and require skilled people to perform them successfully. Additionally, a PCR may fail if the primers site has undergone mutations, a problem more pronounced when analysing RNA viruses, but it cannot be overlooked in poxviruses either. Ideally, a method would be available that can be directly performed on a tissue sample, is fast and comes at relatively low cost. There are two sequencing approaches currently in use: 'long read sequencing' (LRS) and 'short read sequencing' (SRS) are used. The SRS methods generate sequences of no more than 300–500 bp in length. Sequencing both ends of a larger fragment (so called 'paired end sequencing') allows for slightly longer fragments to be recovered and may convey extra information, after mapping, on the location of each read relative to a reference genome. Typically, SRS results are mapped against a reference genome after which the resulting map is evaluated and variations relative to a reference can be observed. The LRS methods are characterized by the fact that they generate reads of well over 1000 bp in length. This is highly useful for reconstruction of repetitive parts of the genome as well as the terminal repeat regions present in poxviruses [22]. These methods are also more suitable for denovo whole genome reconstruction. These LRS methods however, also suffer from drawbacks such as low output of data, high input requirements (in terms of nanograms or even micrograms of DNA), specialized protocols for sample preparation, highly specific and complex data processing methods and finally high error rates [22]. These biases and technical challenges become highly relevant when attempting to detect low copy variants, to characterize viral communities, and assess vaccine purity [23]. Samples for both LRS and SRS can be prepared in two ways. The first is shearing using sonication, the second is enzymatic shearing ('tagmentation') using unbound transposons. The latter method is employed in the 'rapid kit for Oxford Nanopore sequencing and in the 'Nextera kit' manufactured by Illumina. The transposon-based methods are widely used but they perform sub-optimal in shearing AT-rich genomes [19, 20]. The Illumina 'DNA prep' kit also uses transposons, but these are bound to a bead.

In an effort to streamline the methods and provide guidelines for successful *Capripox* virus whole genome sequencing, we have established a collaboration between several reference laboratories which compared seven different next generation sequencing methods. We provide

[1] Taxonomy statement: we follow the names and acronyms as proposed by [61].

an overview of the approximate effort needed to reach a reliable and (near) complete WGS for (Capri)pox viruses and guidelines for good sequencing practices.

Materials and methods

Experimental setup

Collaborative test

The four participating laboratories were: (1) National reference laboratory of the United Kingdom (the Pirbright Institute) located in Pirbright, United Kingdom; (2) The Laboratory of the United Nations' International Atomic Energy Agency (IAEA) located in Vienna, Austria; (3) The national reference laboratory of Australia (CSIRO) located in Canberra, Australia; and (4) the national reference laboratory of Belgium (Sciensano) located in Brussels, Belgium. Sciensano additionally serves as the European and World reference laboratory for *Capripox*. Each laboratory independently attempted to reconstruct a whole genome sequence from the same sample (see below) using their own methods and experience. Afterwards, the mapping and assembly of the genome sequence was redone using the results from the denovo sequencing as reference.

Sample used

The sample used in this study was derived from a bull infected during a previous study [24] with a clade 1.2 LSDV strain which was sequenced previously and published under GenBank acc. nr. KX894508. No animal experiments were thus performed for this study. The original inoculum used for infection was grown on cell cultures and underwent four passages in OA3-T cells before being used in an animal trial [24]. The animal developed skin nodules and one of these was collected and stored We obtained one of these nodules from two of the co-authors (AH and WP). The homogenate was aliquoted in 500 µl tubes and inactivated by heating the sample for 4:30 h at 56 °C. These aliquots were subsequently shipped to the participating laboratories.

Experiment A: Pacific bio LRS run

This experiment used the Pacific Bioscience long read sequencing approach to create a denovo WGS for the strain used in this study.

DNA was extracted from four subsamples of the homogenate using the Puregene tissue core kit (Qiagen). The concentration of the combined samples was assessed using ScreenTape on Tapestation 2200. Cycle threshold (C_T) values were determined using the method developed by [25], which targets the ORF074 gene.

The extract was then used as starting material for NGS and prepared according to the following methods: SMRTbell libraries were prepared with the SMRTbell prep kit 3.0, including shearing, repair, A-tailing, ligation with barcodes and adapters, and size selection with SMRTbell clean-up beads. Pooled libraries (≥ 300 ng) were converted to SMRTbell libraries and sequenced on the Sequel II platform under CCS mode for 30 h.

The thus obtained read library was assembled de-novo according to the following procedure. Reads were classified using the Centrifuge classifier [26]. LSDV-matching reads were extracted and assembled de novo using Canu v2.3 [27], Flye v2.2.2 [28], and Unicycler v0.50 [29]. Alternatively, de-novo assembly was performed directly with the Improved Phased Assembly software IPA (https://github.com/PacificBiosciences/pbipa) without host removal or with Canu after removing cattle reads. Assemblies were compared, and discrepancies were assessed through mapping using minimap2 v2.26 [29–31] and visualization with IGV [32].

Experiment B: Oxford nanopore LRS run

In this experiment we employed Oxford Nanopore technologies long read sequencing approach to obtain a sequence.

DNA was extracted from one subsample of the homogenate using the Machery Nagel (MN) DNA tissue kit. The concentration of the sample was assessed using the Qubit 4.0 HS dsDNA kit. C_T values were determined using the method developed by [33], which targets the D5R gene.

The extracted nucleic acids were amplified using a metagenomics amplification protocol as described previously [34, 35]. The library was prepared using the rapid library preparation (SQK-RBK110-96; ONT) protocol and sequenced on an R9 MinION flow cell with a GridION device. Super-accurate base calling and demultiplexing was performed using Dorado v0.7.1 [36]. Reads were filtered on a minimum Phred average quality score of 7 using Chopper v0.6.0 50].

There was insufficient data for de-novo assembly, so a reference-based assembly method was used. For this, reads were mapped to the closest matching GenBank entry (MN995838.1) using minimap2 v2.24 [37]. A consensus genome was generated using Samtools and Medaka [38]. All positions with a depth below 3X were masked.

Experiment C) low pass illumina nextera XT run

In this experiment we attempted to recover the whole genome sequence using low-pass Illumina Nextera based paired end sequencing and subsequent mapping to the reference.

As a pre-treatment, one subsample of the homogenate was pre-treated with Baseline Zero DNAse for 20 min. Subsequently the DNA was extracted from the treated subsample of the homogenate using the Machery Nagel (MN) NA tissue kit. The concentration of the sample was assessed using the Qubit 4.0 HS dsDNA kit. C_T values

were determined using the method developed by [33], which targets the D5R gene.

Library preparation was done using the Nextera XT kit (Illumina) according to the manufacturer's instructions and whole genome sequencing was done using the Illumina MiSeq instrument. Paired-end reads of 250 bp length were generated (MiSeq V3 chemistry, i.c. Miseq v3 600c reagent cassette). Resulting data was de-multiplexed and adapters and Illumina indices were removed before continuing with downstream analyses.

Read mapping was performed using the LSDV GenBank reference sequence NC_003027.1 complete genome using the BWA mapper [39, 40]. Because the viral read yield was low (see the results section) no quality filtering on the reads was applied. All reads from the raw data were included in the assembly. Samtools [41] was used to create a consensus and calculate read depth and genome coverage The resulting assembly was evaluated in the program 'Tablet' [42] and a majority rule was applied when assessing variable indel positions. A threshold of 10% read variation per base position was used to call ambiguous (if present), non indel, sites (i.e., if > 10% of the mapped reads show a consistent variant, the base was called as being ambiguous). The consensus was manually assessed and curated to verify polymorphisms and indel sites.

Experiment D) low pass illumina TruSeq DNA nano run

In this experiment we attempted to recover the whole genome sequence using low-pass Illumina based paired end sequencing using the TruSeq approach with subsequent mapping to the reference.

As a pre-treatment the homogenate was pre-treated with Baseline Zero DNAse for 20 min. Subsequently the DNA was extracted from one subsample of the homogenate using the Machery Nagel (MN) NA tissue kit. The concentration of the sample was assessed using the Qubit 4.0 HS dsDNA kit. C_T values were determined using the method developed by [33], which targets the D5R gene.

Library preparation was done using the TruSeq DNA Nano sequencing kit (Illumina) in combination with the IDT for Illumina – TruSeq DNA UD Indexes v2 (Illumina) according to the manufacturer's instructions following the protocol for mechanical fragmentation into 350 bp fragments (100ng input). Paired-end reads of 150 bp length were generated on an iSeq Sequencer (iSeq100 v2 chemistry, i.c. iSeq100 i1 Reagent v2 (300-cycle) reagent cassette.).

Resulting data was de-multiplexed and adapters and Illumina indices were removed before continuing with downstream analyses. Read mapping was performed using the LSDV GenBank reference sequence NC_003027.1 complete genome using the BWA mapper [39, 40]. Because the viral read yield was low (see the results section) no quality filtering on the reads was applied. All reads from the raw data were included in the assembly. A first draft consensus sequence was generated using Samtools. Read depth and coverage was also evaluated using Samtools [41]. The total assembly was evaluated in the program 'Tablet' [42] and a majority rule was applied when assessing variable indel positions. A threshold of 10% read variation per base position was used to call ambiguous (if present), non indel, sites (i.e., if > 10% of the mapped reads show a consistent variant, the base was called as being ambiguous). The consensus was manually assessed and curated to verify polymorphisms and indel sites.

Experiment E) illumina nextera XT run with hybrid probe capture

In this experiment we attempted hybrid probe capture of the LSDV genome and subsequent paired end sequencing using the Nextera/TruSeq based Illumina chemistry and reference based read mapping.

DNA was extracted using the MagMAX Viral RNA Isolation kit (ThermoFisher) on an automated MagMAX Express 24 sample processer. C_T values were determined using the capripoxvirus real-time PCR developed by [39], which targets ORF074 of the viral genome. The DNA concentration was determined using the Qubit dsDNA HS Assay Kit (Thermofisher Scientific).

A probe hybridization procedure for LSDV was developed using the MyBaits Target Capture Kit (Arbor Biosciences), which uses complementary oligonucleotides for selectively binding target DNA in favour of host or environmental DNA. The oligonucleotide hybrid capture probes for enrichment were designed using representative LSDV genomes from GenBank. The final design consisted of 18,943 probes of 90 bp each with 3x depth over the LSDV genomes. Sequencing libraries for the samples were prepared using the Nextera XT DNA Library Prep Kit (Illumina) and enriched for LSDV fragments using the probes as per the manufacturer's recommendations (Arbor Biosciences). The libraries were reamplified and sequenced using a P1 300-cycle cartridge on a NextSeq2000 instrument (Illumina). The raw reads were cleaned using Trimmomatic v.0.3911 and mapped to the Neethling Warmbaths LSDV genome (GenBank Acc. AF409137.1) using Bowtie v.2.4.412 [43]. The mapping was manually examined and edited with Geneious Prime v.2023.1.2.

Experiment F) illumina nextera run high output

In this experiment we attempted to recover the whole genome sequence using deep, Illumina Nextera based, paired end sequencing and subsequent mapping to the reference.

DNA was extracted from one subsample of the homogenate using the MagMAX CORE Nucleic Acid Purification Kit (Applied Biosystems), this was then treated with BaselineZero DNAse for fifteen minutes. The concentration of the sample was assessed using Qubit 4.0 HS dsDNA kit. C_T values were determined using an in house developed method, which targets the P32 gene.

A total of 1ng of DNA was used as input of the Nextera XT DNA library preparation kit. Sequencing was performed using the P1 (300 cycles/100 M reads) sequencing cartridge for the Illumina NextSeq 2000 sequencing platform. Initial quality control (QC) of Illumina paired-end reads was performed using FastQC, followed by trimming of low-quality bases and adapter sequences using fastp. An iterative mapping and assembly approach was then applied. Reads were first filtered by mapping to the reference genome sequence OM033705 using BBMap v39.06 (https://sourceforge.net/projects/bbmap/, 2024). The filtered reads were then de novo assembled using SPAdes, and assembly quality was evaluated with QUAST. The resulting contigs were queried against the NCBI virus database using BLASTn [44, 45] to identify the closest reference sequence. The best matching reference (KY829023.3) was selected for subsequent read alignment using BWA and the resulting Alignment quality was first assessed with Qualimap, after which duplicate reads were removed using samtools markdup. The deduplicated reads were then realigned to the reference using BWA for improved accuracy, followed by a second round of alignment quality assessment with Qualimap. Variants were called with bcftools, mpileup and bcftools call, and consensus sequences were generated accordingly. All alignments, variants, and consensus sequences were visually inspected using the IGV [32].

Experiment G) illumina DNA-prep run

In this experiment we attempted to recover the whole genome sequence using low-pass Illumina based paired end sequencing using the DNA prep approach with subsequent mapping to the reference.

One subsample of the homogenate was pre-treated with Baseline Zero DNAse for 20 min. Subsequently the DNA was extracted from the treated subsample of the homogenate using the Machery Nagel (MN) NA tissue kit. The concentration of the sample was assessed using the Qubit 4.0 HS dsDNA kit. C_T values were determined using the method developed by [33], which targets the D5R gene.

Library preparation was done using the Illumina DNA Prep Library preparation kit (Illumina) according to the manufacturer's instructions. Paired-end reads of 150 bp length were generated on an iSeq™ Sequencer (iSeq100 v2 chemistry, i.c. iSeq100 i1 Reagent v2 (300-cycle)reagent cassette).

Resulting data was de-multiplexed and adapters and Illumina indices were removed before continuing with downstream analyses. Read mapping was performed using the LSDV GenBank reference sequence NC_003027.1 complete genome using the BWA mapper [39, 40]. Because the viral read yield was low (see the results section) no quality filtering on the reads was applied. All reads from the raw data were included in the assembly. A first draft consensus sequence was generated using Samtools. Read depth and coverage were also evaluated using Samtools [41]. The total assembly was evaluated in the program 'Tablet' [42] and a majority rule was applied when assessing variable indel positions. A threshold of 10% read variation per base position was used to call ambiguous (if present), non indel, sites (i.e., if > 10% of the mapped reads show a consistent variant, the base was called as being ambiguous). The consensus was manually assessed and curated to verify polymorphisms and indel sites.

Detailed methodology of sequencing methods

Seven different sequencing experiments (A to G) were performed to determine the WGS of the LSDV strain present in the sample. Detailed protocols for all of the sequencing methods can be found in supplementary data (OSM1). An overview of the key steps and kits used in each of the sequencing protocols can be found in Table 1; Fig. 1.

Genome re-mapping of datasets B-G

Raw read datasets obtained by methods C-G were re-mapped using the same methods as described below against the dataset obtained by method A to remove any potential effect of different assembly and mapping methods. For the SRS datasets (C-G) the BWA mapper was used. For ONT LRS dataset (B) minimap2 v2.24 was used and no minimum value was set for read depth used to report the consensus. All assemblies were manually evaluated in Tablet as well as with Quast 5.3.0 [47]. For the Quast results we refer to the supporting materials 2. The visualization of the read depth was done via Coverage and Depth (or "CoDe") plots prepared in R v4.2.0 [23]. using ggplot2 v3.5.1 [48].

Mapping evaluation and genome completeness

We evaluated the final re-mapped results and compared these with regards to minimum, maximum and average read depth. We also evaluated to which extent each assembly contained regions of low (< 10x), very low (< 5x) or no (0) coverage. This allowed us to detect regions of low or no coverage and ignore the effect of the zero reads mapping on the flanking regions and asses the result in more detail. We then determined the standard deviation

Table 1 DNA extraction methods and pre sequencing evaluation methods

	A	B	C	D	E	F	G
DNA extraction method	Puregene tissue core kit (Qiagen)	Machery Nagel DNA tissue	Machery Nagel DNA tissue	Machery Nagel DNA tissue	MagMAX Viral RNA Isolation	MagMAX CORE Nucleic Acid Purification	Machery Nagel DNA tissue
DNAse yes/no	N	N	Y	Y	N	Y	Y
concentration assesment method	ScreenTape on Tapestation 2200	Qubit 4.0 HS dsDNA kit	Qubit 4.0 HS dsDNA kit	Qubit 4.0 HS dsDNA kit	Qubit 4.0 HS dsDNA kit	Qubit 4.0 HS dsDNA kit	Qubit 4.0 HS dsDNA kit
Assement of C_T values	[39]	[46]	[46]	[46]	[39]	[*]	[46]
Enrichment step	-	-	-	-	hybrid probe capture	-	-
Library prep kit	SMRTbell prep kit 3.0	rapid library preparation	Nextera XT kit	TruSeq Nano	Nextera XT kit	Nextera XT kit	DNA prep
Sequencing platform	Pacific Bioscience	R9 MinION GridION	MiSeq	iSeq	NextSeq2000	NextSeq2	iSeq
Bio-informatic analysis	De novo	reference based	reference based	reference based	reference based	reference based	reference based

[*] unpublished method, targeting the P32 gene

A) denotes the denovo PaCBio LRS experiment, B) denotes the ONT based LRS experiment, C) denotes Nextera SRS experiment, D) denotes the TruSeq based SRS experiment, E) denotes the Nextera SRS experiment with the BCM employed, F) denotes the Nextera SRS experiment with much deeper sequencing, G) denotes the DNA prep SRS based experiment

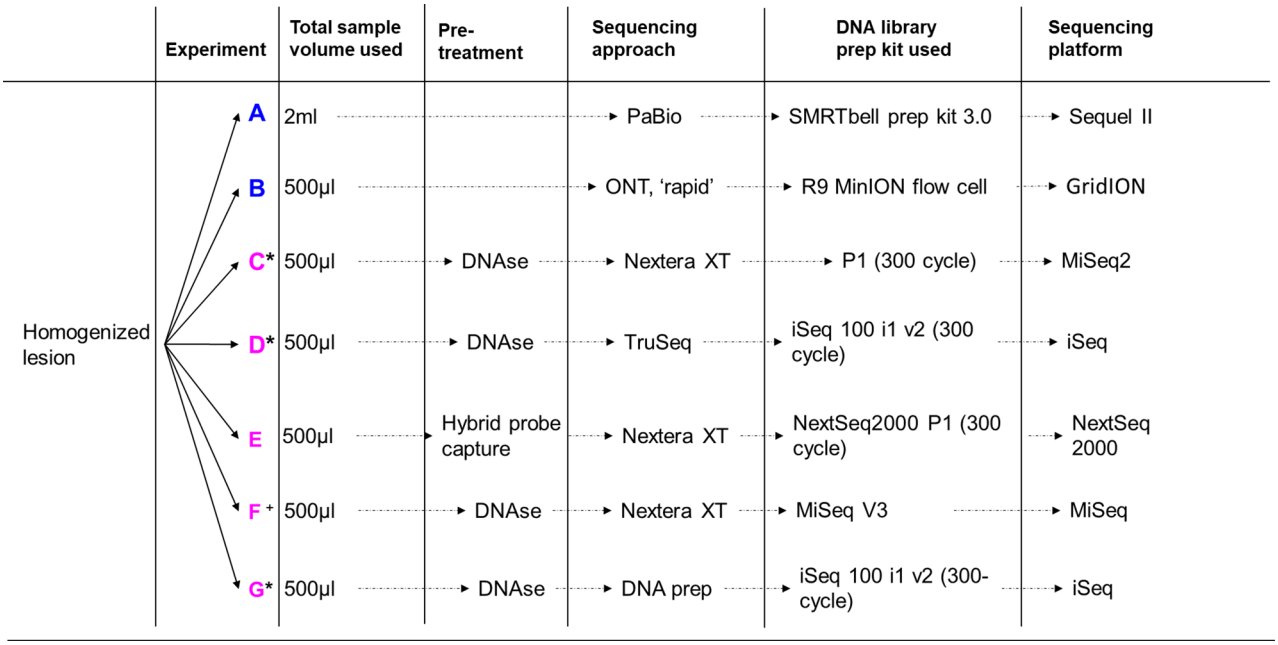

Fig. 1 Flow of samples per experiment. Samples indicated with a * are from the same individual tube. For the sample indicated with a + a full MiSeq cartridge was used

(sd) of the read depth as a reflection of the differences of depth across each genome position.

To determine the completeness of our denovo reconstructed genome we evaluated the length of the terminal repeat region (TR-region). Dotplots generated by Mafft [49] and an alignment with the LSDV reference (GenBank acc. nr. NC_003027.1) were used to determine the length of the TR. Our genome reconstructions were considered complete when the length of the TR-regions reached 99%, were as long as, or longer than what is reported for the reference sequence.

The consensus of experiment A was submitted to GenBank with accession number: PX492334.1.

Results

DNA extraction, viral load and sequencing output

An overview of the DNA extraction methods, concentrations and C_T measurements, enrichment methods, sequencing methods as well as results of each separate WGS reconstruction as described in the paragraphs below can be found in Table 2.

Table 2 DNA concentration and C_T va:lues for each extract and subsequent NGS sequencing read yields for each resulting NGS dataset

	A	B	C	D	E	F	G
Concentration (ng/µl)	3.50	5.20	0.25	0.25	6.71	0.50	0.20
C_T virus	20.96	18.8	20.44	20.44	18.35	25	20.44
C_T host	-	24.4	29.62	29.62	-	-	29.62
C_T ratios (target/non target)	-	0.77	0.69	0.69	-	-	0.69
complete genome reported	Y	Y	N*	Y	Y	Y	Y
genome coverage (%)	100	100	99	100	100	100	100
output file (10^6 reads)	8.91	6.72	3.60	3.70	30.10	9.15	7.53
genome length recovered after re-mapping	150,896	150,896	150,855	150,886	150,680	150,728	150,891
% of LSDV associated reads	0.016	0.15	3.6	7.2	34.1	0.3	3.6
count of LSDV associated reads	1426	9937	130,410	269,965	3,120,596	176,980	271,316

* This sequence was not complete because there were gaps in the regular, non terminal, genomic regions. The terminal repeats for this sequence were nearly complete in a way similar to the other reconstructions

A) denotes the denovo PaCBio LRS experiment, B) denotes the ONT based LRS experiment, C) denotes Nextera SRS experiment, D) denotes the TruSeq based SRS experiment, E) denotes the Nextera SRS experiment with the BCM employed, F) denotes the Nextera SRS experiment with much deeper sequencing, G) denotes the DNA prep SRS based experiment

Table 3 Mapping details for each experiment

	A	B	C	D	E	F	G
average read depth	58	48	268	179	2959	134	246
standard deviation	13.13	50.94	325.57	63.6	1659.55	155.11	30.8
standard deviation as % of the average	22.6	106.1	121.5	35.5	56.1	115.8	12.5
max. depth	107	575	2930	566	12,756	1604	400
min. depth	17	1	0	0	0	0	0
5' missing	0	5	5	36	5	5	1
3' missing	0	5	5	5	5	5	5
number of positions at zero	0	0	5	31	0	0	1
number of positions at < 5	0	653	1367	147	0	75	5
number of positions at < 10	0	6092	6087	181	1	262	1
min. depth excluding the first and last 5 bp	-	1	0	1	124	2	16

Different extraction kits were used, but the DNA concentration recovered was generally low. For the DNase treated samples the concentrations ranged from 0.20 to 0.5 ng/µl. for the untreated samples this ranged from 3.5 to 6.71 ng/µl. Obtained C_T values for the viral target varied in the DNase treated samples were between 20.44 (dataset B & C) and 25 (dataset F). For the untreated samples the range was between 18.35 and 20.96. Experiments B and C, D and G also determined the C_T values of a host gene, making that the ratio of host/virus could be determined. This ratio was 0.77 in experiment B and 0.69 in experiments C, D and G (as these datasets were all derived from the same sample (Table 2).

The resulting read libraries varied in size. The smallest read-library was obtained for experiment A with $3.6 * 10^6$ reads and the largest read-library was obtained for experiment E with $\sim 30 * 10^6$ reads. The total read library sizes are listed in Table 2.

Target read yield and mapping results

Depending on the mapping approach and reference genome used, some of the obtained consensus genomes contained gaps when compared to the sequence obtained by method A. To be able to compare the results obtained by the different methods and by the different participating laboratories we remapped the datasets against the sequence obtained via method A.

The yield of target (viral) reads per dataset differed (Table 3). The lowest percentage of target reads was recovered for both LRS datasets with PacBio and ONP yielding 0.016% for the PaCBio LRS based method (A) and 0.15% for the ONT LRS based method (B) of viral reads, respectively. The difference in target read output between the same sample processed with the Nextera (method C) and DNA-prep (method G) kits vs. the TruSeq kit (method D) is twofold as they yielded 3.6% (C, G) vs. 7.2% (D) target reads, respectively. Method F, representing the deep sequencing effort using the Nextera kit, yielded 0.3% of viral reads. Dataset E, combining TruSeq library preparation with target enrichment via hybrid capture probes, yielded 34.1% of viral reads.

The recovered LSDV sequence contained an AT-percentage of 74.12% and a GC content of 25.88%. There were no ambiguous base positions in any assembly

except in the assembly for dataset G which contains two ambiguous positions and one deletion when compared to the sequence from experiment A. The terminal repeat was 2471 bp in length and was present on both the 5' and 3' side of the reconstructed WGS from dataset A (Fig. 2 & OSM2).

All datasets, except G missed 5 bp of the terminal repeats compared to the sequence obtained by the PACBIO LRS in method A. Method C which used the Nextera kit even missed 36 bp at the 3' end of the genome. Method C was also the only which resulted in a dataset with five unmapped positions in the core part of the genome and furthermore had a large number of positions (1367) only covered by less than five reads.

Read distribution and read depth

Table 3 provides an overview of the obtained read depth. The read distribution over the genome is visualized in the CoDe plots in Fig. 3. The results show that datasets C-G all had an average read depth of > 100 reads per position. The probe capture based dataset (E) showed the highest average read depth of 2559 reads per position and the long read ONT based dataset (B) the lowest with 48 reads per position. The variation of the read depth over the genome however strongly differs between the methods. The read distribution for datasets generated with the transposon-based kits (B, C and F) are highly unequal (Figs. 3 and 4). The sd's for these sets are > 100% of the average (Table 3). The distribution of reads for datasets A, D and G is more homogenous and the corresponding sd's of distribution are lower. For these three datasets there is also an increased read mapping at the flanks of the genome. Dataset E also has a high sd (1659, 56.1% of the average). Dataset G also has a transposon component involved but it is treated differently during sample preparation (see the discussion). When comparing the mapping results for all datasets no polymorphism between them were found.

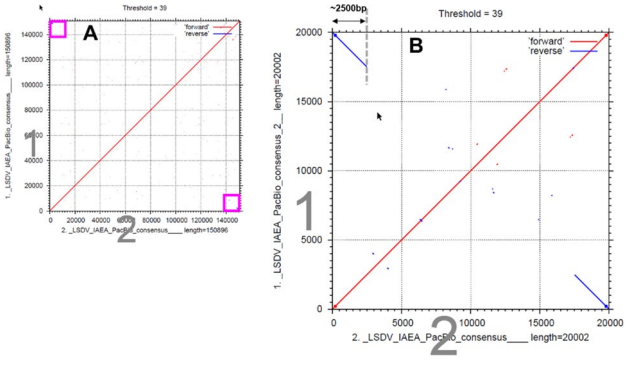

Fig. 2 Dot plots of the WGS based on dataset alpha. **A** displays a comparison of the sequence with itself. The blue lines in the topleft and lower richt corners indicate the presence of the terminal repeat region. **B** displays a representation of the first and last 10,000 basepairs of dataset **A**. The blue lines indicate the terminal repeat region. The length of the region is displayed at the top left corner

Discussion

The use of whole genome sequences in virus studies is of increasing importance as it provides insights into the changes of the whole virus genome, not just a separate gene. Especially with relatively slow evolving viruses such as the poxviruses [50] there is need to obtain more comprehensive overview of the virus under study. But poxviruses show skewed AT-GC ratios [18] which may impact sequencing success and therefore the reliability of the WGS reconstruction.

The first step of each sequencing project is the extraction of the DNA from any given sample. While this is critical there appears to be not much difference between the methods used in this study. All commercially available kits will yield approximately the same amount of DNA. The application of DNase to increase the virus to host ratio is based on the idea that the encapsulated virus (prior to extraction) may be more resistant to the treatment than the background DNA and also that there is much more background DNA than the target DNA which will thus be, relatively speaking, less affected. While this may be the case, given the slight improvement in virus to host C_T ratios observed (0.77 to 0.69 experiments C, D,G, there is a considerable loss of both target and non-target DNA what may impair downstream applications such as hybrid probe capture of perhaps even LRS methods.

Our results show that truly random DNA fragmentation using sonication as applied in the TruSeq (experiment D) kit results in a more even distribution of reads over the genome when compared with the non-random tagmentation as employed in the Nextera Kit (experiments C, E,F). Tagmentation is achieved by the use of a transposon. We hypothesize that the transposon used to fragment the DNA has a preference for GC-rich regions where it will attach and subsequently catalyse the fragmentation. These regions are therefore more represented in the resulting sequence read libraries. This is not a problem if AT/GC content is roughly equal within a target genome (although GC-poor regions may still be missed there), but it is not efficient in the case of *Capripox* genomes. When leaving out the transposon-based steps, both depth (yield) and breadth (distribution across the genome) of reads increased. This can be seen in the case of datasets C, D and G where the yield of virus associated reads vs. host reads is almost doubled when using the TruSeq kit vs. the Nextera and DNA-prep kits (in terms of read count and percentage). This further underscores the bias introduced by the transposon-based methods and confirms what was found by [19, 20]. While [51] report a minor negative AT-bias when using the Nextera kit, our results suggest a more profound effect.

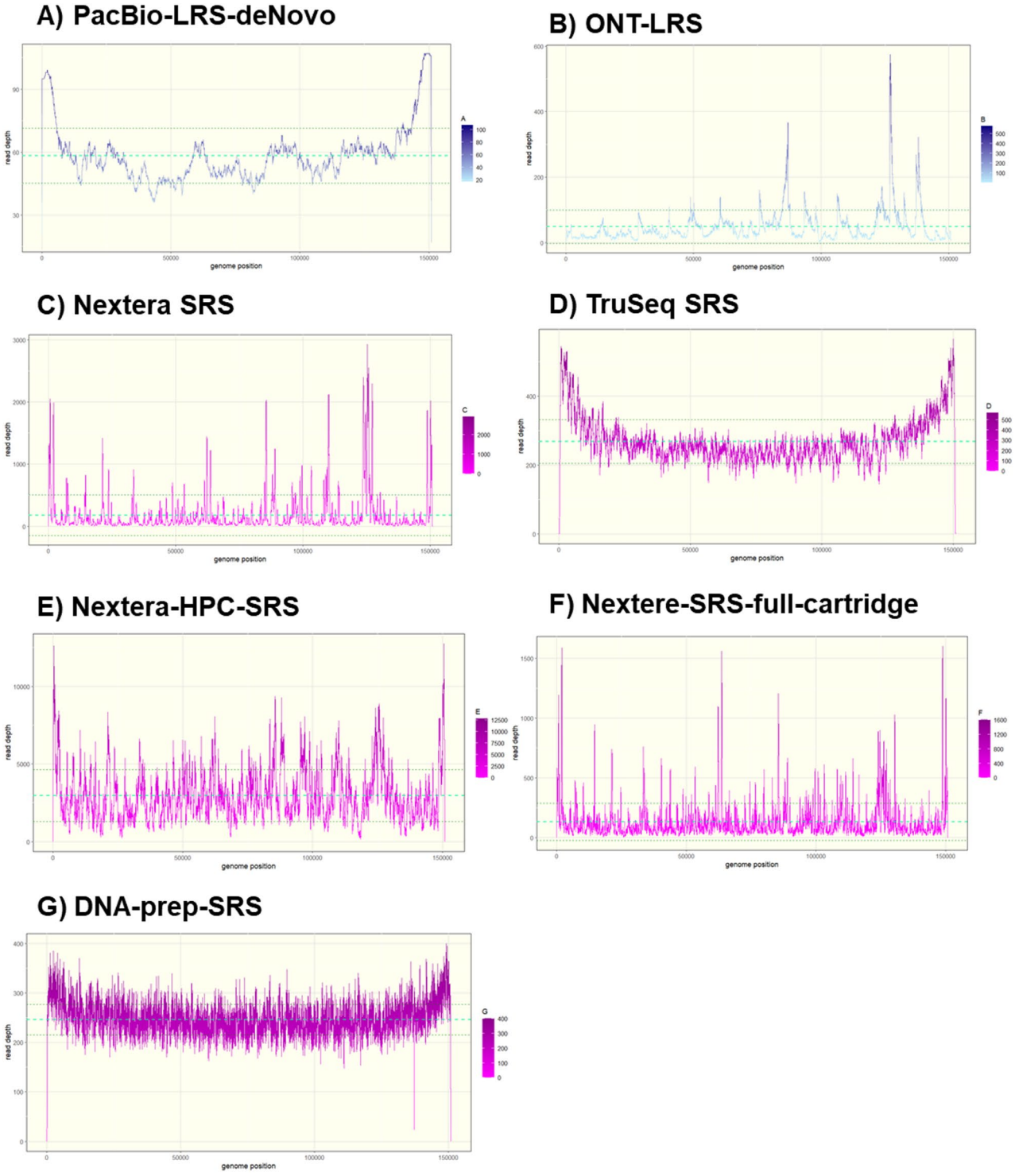

Fig. 3 Coverage and Depth plots ("CoDe") plots for each dataset. Panels **A** & **B** represent the long-read sequencing based datasets **A** and **B**. Panels (**C**,**B**-**G**) represent the Illumina based datasets (**C**-**G**) respectively. The dashed green lines denote the average read depth for each dataset based on all positions. *LRS* denotes Long Range Sequencing methods. *SRS* denotes Short Read Sequencing methods. *HPC* means Hybrid Probe Capture was employed

In this study we also tested the DNA-prep Illumina kit (experiment G), which claims to avoid the negative AT-bias by binding the transposon to beads forcing a maximum of DNA to bind to the beads irrespective of the AT/GC-ratio of the DNA [52]. Our results show that this is indeed the case, but the yield of target reads was lower than with the TruSeq based experiment (Experiment D) (Table 1; Figs. 3 and 4). The use of hybrid capture probes

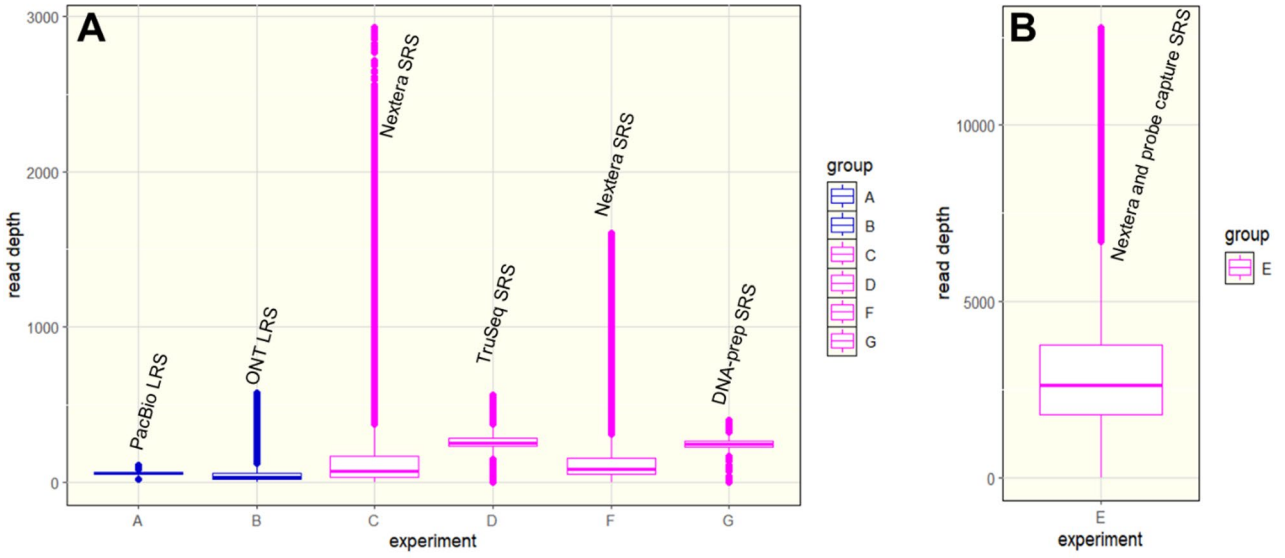

Fig. 4 Boxplots displaying distribution of read depth around the average for datasets **A**, **B**,**C**, **D**, **F**, **G**. Mapping results for LRS are in blue, for SRS in pink (**A**). Boxplot displaying distribution of read depth around the average for dataset E with a different scale (**B**). *LRS* denotes Long Range Sequencing methods. *SRS* denotes Short Read Sequencing methods. *HPC* means Hybrid Probe Capture was employed

proved to be very effective. The amount of target reads recovered increased more than 4.5 times, to > 30% of the read library, when compared to the next best result of obtained with the TruSeq experiment (D). In the case of important samples or samples with low virus yield this approach may be useful and effective.

Another way to overcome the low virus yield would be to use a PCR based approach. The long-range PCR protocol developed by [21] overcomes this partially. This is a very robust protocol resulting in a large amount of virus derived sequences. The downside of this procedure is that it is also time consuming and requires experienced personnel to perform. Further this protocol does not allow to reliably detect sequence variants that may occur in a virus population, which is especially of concern when evaluating vaccine sequences of live attenuated vaccines.

The data processing and genome assembly is relatively straightforward, but in the case of LRS data requires powerful computing units and a good understanding of the various tools needed. The main pitfall when using the SRS data is the correct selection and use of the mapping reference sequence and proper inspection of the resulting assembly. Use of a reference that is too variable, especially when large indels are present, may influence the result. In the case of this study, we encountered this in the case of the hybrid probe capture procedure and the Nextera deep sequencing method whereby the initially reported WGS differed from the results of the other experiments (see OSM2). Remapping revealed that this was caused by the use of a different, less related reference. Another known problem is caused by long repeat regions (> 1000 bp) in a genome. These cannot be covered by single reads from SRS methods and mapping will be less reliable.

Long read sequencing methods hold great promise for virus studies [53] but have not been done routinely in poxviruses (e.g., 46, 54]. Perhaps the fact that LRS methods might also have a bias in AT-rich genomes [55] may have discouraged people from using them. However, LRS methods avoid problems that arise from the presence of repeats in a genome and the longer reads that are generated may traverse otherwise hard to assemble parts of the genome. Further, the length of the reads also makes these methods more suitable for denovo genome reconstruction. Our results show that, while suited for denovo sequencing, LRS methods result in low virus associated read recovery. That being said, the length of the reads more than compensates for this in our results with a comprehensive mean coverage of 58 (PacBio) and 40 (ONP) reads per position. This seems sufficient to us for a reliable recovery of the virus genome, provided the coverage of the genome is homogenous [56]. While the read distribution of the PacBio result is homogenous (dataset A, Figs. 3A and 4A), this was less for the ONP LRS (dataset B, Figs. 3B and 4B) where the results showed similarities with those observed in the Nextera based protocols (compare Fig. 3B with Fig. 3C). We attribute this to the fact that the ONP procedure also involves the use of a transposon during the tagmentation step [57]. This may indicate that the employed ONP approach also suffers from a sequencing bias in AT-rich genomes and it may therefore be less suited than the PacBio technology for sequencing genomes with skewed AT/GC ratios.

Based on the occurrence of restriction enzyme sites, the TR regions of *Capripox* were historically determined to be between 2.25 and 3.4 kb [58]. Research from [59] described the reference LSDV genome having a 2418 bp inverted repeat. In the reconstruction of the LSDV genome in our sample, the TR region was ~ 2471 bp (Fig. 2 and supporting materials 1), which is 53 bp longer than the reference. Our denovo reconstructed genome therefore probably represents the complete genome. While the TruSeq and PacBio read distributions are considered the most homogenous, an increased number of reads mapping at both flanks in the TR-region of the reconstructed genomes was observed (Fig. 3A, D & G). This can be explained by unequal mapping of reads across the TR-region relative to the reference, which happens when multiple optimal mapping coordinates occur in a reference genome.

The terminal repeats of LSDV are currently reported to be very variable (see the alignments produced by [13]), we are not sure whether this is really the case or if this is an artefact of low coverage of these regions in the studies publishing them or as a result of sequencing procedures. The CoDe plots used in this study are a simple to create and visually easy way to evaluate and present this. Finally, to asses genome completeness a dot plot can be reconstructed and the length of the TR-regions can be compared with the known literature and confirmed full length genome sequences present in public databases (i.e., GenBank).

Recommendations

Based on the sequencing breadth and depth obtained for the different sequencing protocols tested we recommend to use the TruSeq kit for whole genome sequencing of *Capripox* viruses in samples with sufficient viral load and high enough DNA concentration as the TruSeq kit does need more input material. When it is needed to sequence a sample with very low virus content the hybrid probe capture method will be valuable to use in combination with DNA prep-based sequencing as it allows for lower concentrations of start material. Both methods can be easily implemented by any laboratory already using Illumina sequencing. For mapping and assembling we recommend to use an iterative strategy with multiple reference genomes when SRS is used.Iterative mapping can be used to compensate for mis-mapping in regions that vary in comparison with the reference.

In order to be able to report reliable sequencing results, we recommend to strive for a read depth of 30, especially in the case of error prone methods such as ONP, as this will allow to safely capture both INDEL and nucleotide variation [60]. This, furthermore, strongly reduces the chance of miscalling bases due to base calling errors [56]. This is already true at 10 times coverage, but indel variation is often more difficult to capture. We also recommend to report information on read coverage and depth for each reconstructed genome so the readers can have an idea about the robustness of the reported sequence information. The latter is currently often lacking and might be the cause of difficulties during the alignment of the terminal repeat regions or regions with indels when such sequences are used as reference themselves.

Abbreviations

LSDV	Lumpy skin disease virus
WGS	Whole genome sequence
LRS	Long read sequencing
BCM	Bait capture methods
SRA	Sequence read archive
SPPV	Sheeppox virus
SRS	Short read sequencing
.	Long read sequencing: Sequencing methods that capable of generating reads longer than 1000bp
.	Short read sequencing: sequencing methods that generate reads shorter than a 1000bp but typically with a maximum of 300bp
.	Shallow sequencing: sequencing of a sample not using the entire cassette, e.g; as part of a multiplexed sequencing run. Typically output is less than 1GB
.	Deep sequencing: sequencing with a data output in excess of 1GB

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-025-12463-3.

Supplementary Material 1.

Supplementary Material 2.

Acknowledgements

The authors are grateful to the laboratory teams at Sciensano, CSIRO, IAEA and Pirbright for their help with the molecular analyses as well as to the field workers for collecting the samples and the animals' owners for generously sharing these. We are also grateful to the Bill & Melinda gates foundation for financial support.

Authors' contributions

Conceptualization : FCB NDR.; FCB, NDR, SH. —Original draft preparation : FCB, NDR. —Review and editing FCB, AH, WP, NDR, SH, SdK, TRB, CB, CW, CEL. NGS : SH, SdK, PM MJW CB, CW, TBKS. Data analyses : FCB, MJW, CB, TRB, TBKS, CEL, CW, CB. Visualization: FCB. All authors have read and agreed to the published version of the manuscript.

Funding

This study was financed by a grant from the Bill and Melinda Gates foundation, grant nr: INV-055082 awarded Sciensano. PacBio sequencing was funded by the IAEA laboratory in Vienna.

Data availability

Sequence alignments are available on request. The long read sequencing result has been deposited in GenBank with accession numbers PX492334. The sequence read archives are available on request.

Declarations

Competing interests

The authors declare no competing interests.

Received: 15 October 2025 / Accepted: 17 December 2025
Published online: 08 January 2026

References

1. Mathijs E, Vandenbussche F, Nguyen L, Aerts L, Nguyen T, De Leeuw I, Quang M, Nguyen HD, Philips W, Dam TV, et al. Coding-complete sequences of Recombinant lumpy skin disease viruses collected in 2020 from four outbreaks in Northern Vietnam. Microbiol Resour Announc. 2021;10(48):e0089721. https://doi.org/10.1128/MRA.00897-21.
2. van Schalkwyk A, Kara P, Ebersohn K, Mather A, Annandale CH, Venter EH, Wallace DB. Potential link of single nucleotide polymorphisms to virulence of vaccine-associated field strains of lumpy skin disease virus in South Africa. Transbound Emerg Dis. 2020;67(6):2946–60. https://doi.org/10.1111/tbed.13670.
3. van Schalkwyk A, Byadovskaya O, Shumilova I, Wallace DB, Sprygin A. Estimating evolutionary changes between highly passaged and original parental lumpy skin disease virus strains. Transbound Emerg Dis. 2021;69:e486–96.
4. van Schalkwyk A, Kara P, Last RD, Romito M, Wallace DB. Detection and genome sequencing of lumpy skin disease viruses in wildlife game species in South Africa. Viruses. 2024;16(2):172. https://doi.org/10.3390/v16020172.
5. Kumar N, Chander Y, Kumar R, Khandelwal N, Riyesh T, Chaudhary K, Shanmugasundaram K, Kumar S, Kumar A, Gupta MK, et al. Isolation and characterization of lumpy skin disease virus from cattle in India. PLoS ONE. 2021;16(1):e0241022. Doi: 1371/journal.pone.0241022. PMID: 33428633; PMCID: PMC7799759.
6. Breman FC, Haegeman A, Philips W, Krešić N, Hoffman S, De Keersmaecker SCJ, Roosens NHC, Agüero M, Villalba R, Miteva A, Ivanova E, Tasioudi KE, Chaintoutis SC, Kirtzalidou A, De Regge N. Sheeppox virus genome sequences from the European outbreaks in Spain, Bulgaria, and Greece in 2022–2024. Arch Virol. 2024;169(11):234. https://doi.org/10.1007/s00705-024-06165-6. PMID: 39476278; PMCID: PMC11525412.
7. Haga IR, Shih BB, Tore G, Polo N, Ribeca P, Gombo-Ochir D, Shura G, Tserenchimed T, Enkhbold B, Purevtseren D, et al. Sequencing and analysis of lumpy skin disease virus whole genomes reveals a new viral subgroup in West and central Africa. Viruses. 2024;16:557. https://doi.org/10.3390/v16040557.
8. Centre for Disease Control (CDC). Monkeypox Outbreak Global Map. Available online: https://www.cdc.gov/poxvirus/monkeypox/response/2022/world-map.html update 05 Mar 2024, viewed/accessed 24 April 2024.
9. Koltsov A, Sukher M, Kholod N, Namsrayn S, Tsybanov S, Koltsova G. Isolation and characterization of swinepox virus from outbreak in Russia. Animals. 2023;13(11):1786. https://doi.org/10.3390/ani13111786.
10. Mohamed RI, Elsamadony HA, Alghamdi RA, Eldin ALAZ, El-Shemy A, Abdel-Moez Amer S, Bahshwan SMA, El-Saadony MT, El-Sayed HS, El-Tarabily KA, Saad ASA. Molecular and pathological screening of the current circulation of fowlpox and pigeon pox virus in backyard birds. Poult Sci. 2024;103(12):104249. https://doi.org/10.1016/j.psj.2024.104249. PMID: 39418793; PMCID: PMC11532475.
11. García-Bocanegra I, Camacho-Sillero L, Caballero-Gómez J, Agüero M, Gómez-Guillamón F, Manuel Ruiz-Casas J, Manuel Díaz-Cao J, García E, José Ruano M, de la Haza R. Monitoring of emerging Myxoma virus epidemics in Iberian hares (*Lepus granatensis*) in Spain, 2018–2020. Transbound Emerg Dis. 2021;68(3):1275–82. https://doi.org/10.1111/tbed.13781.
12. Mahfuza A, Akter SH, Sarker S, Aleri JW, Annandale H, Abraham S, Uddin JM. Global burden of lumpy skin Disease, outbreaks, and future challenges. Viruses. 2023;15(9):1861. https://doi.org/10.3390/v15091861.
13. Breman FC, Haegeman A, Krešić N, Philips W, De Regge N. Lumpy skin disease virus genome sequence analysis: putative spatio-temporal epidemiology, single gene versus whole genome phylogeny and genomic evolution. Viruses. 2023;28(7):1471. https://doi.org/10.3390/v15071471. PMID: 37515159; PMCID: PMC10385495.
14. Moudgil G, Chadha J, Khullar L, Chhibber S, Harjai K. Lumpy skin disease: insights into current status and geographical expansion of a transboundary viral disease. Microb Pathog. 2024;186:106485. https://doi.org/10.1016/j.micpath.2023.106485.
15. Sendow I, Meki IK, Dharmayanti NLPI, et al. Molecular characterization of Recombinant LSDV isolates from 2022 outbreak in Indonesia through phylogenetic networks and whole-genome SNP-based analysis. BMC Genomics. 2024;25:240. https://doi.org/10.1186/s12864-024-10169-6.
16. Sprygin A, van Schalkwyk A, Mazloum A, et al. Genome sequence characterization of the unique Recombinant vaccine-like lumpy skin disease virus strain Kurgan/2018. Arch Virol. 2024;169:23. https://doi-org.vdic.idm.oclc.org/10.1007/s00705-023-05938-9.
17. Sudhakar SB, Mishra N, Kalaiyarasu S, et al. Emergence of lumpy skin disease virus (LSDV) infection in domestic Himalayan Yaks (*Bos grunniens*) in Himachal Pradesh, India. Arch Virol. 2024;169:51. https://doi.org/10.1007/s00705-024-05994-9.
18. Roychoudhury S, Pan A, Mukherjee D. Genus specific evolution of codon usage and nucleotide compositional traits of poxviruses. Virus Genes. 2011;42(2):189 – 99. https://doi.org/10.1007/s11262-010-0568-2. PMID: 21369827.
19. Gunasekera S, Abraham S, Stegger M, Pang S, Wang P, Sahibzada S, et al. Evaluating coverage bias in next-generation sequencing of *Escherichia coli*. PLoS ONE. 2021;16(6):e0253440. https://doi.org/10.1371/journal. pone.0253440.
20. Sato MP, Ogura Y, Nakamura K, Nishida R, Gotoh Y, Hayashi M, Hisatsune J, Sugai M, Takehiko I, Hayashi T. Comparison of the sequencing bias of currently available library Preparation kits for illumina sequencing of bacterial genomes and metagenomes. DNA Res. 2019;26(5):391–8. https://doi.org/10.1093/dnares/dsz017.
21. Mathijs E, Haegeman A, De Clercq K, Van Borm S, Vandenbussche F. A robust, cost-effective and widely applicable whole-genome sequencing protocol for capripoxviruses. J Virol Methods. 2022;301:114464. https://doi.org/10.1016/j.jviromet.2022.114464. PMID: 35032481; PMCID: PMC8872832.
22. Sloan DB, Broz AK, Sharbrough J, Wu Z. Detecting rare mutations and DNA damage with sequencing-based methods. Trends Biotechnol. 2018;36(7):729–40. https://doi.org/10.1016/j.tibtech.2018.02.009. PMID: 29550161; PMCID: PMC6004327.
23. R Core Team. (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/
24. Haegeman A, De Leeuw I, Mostin L, Campe WV, Aerts L, Venter E, Tuppurainen E, Saegerman C, De Clercq K. Comparative evaluation of lumpy skin disease Virus-Based live attenuated vaccines. Vaccines (Basel). 2021;9(5):473. https://doi.org/10.3390/vaccines9050473. PMID: 34066658; PMCID: PMC8151199.
25. Bowden TR, Babiuk SL, Parkyn GR, Copps JS, Boyle DB. *Capripox* virus tissue tropism and shedding: A quantitative study in experimentally infected sheep and goats. Virology. 2008;371:380–93. https://doi.org/10.1016/j.virol.2007.10.002.
26. Kim D, Song L, Breitwieser FP, Salzberg SL. Centrifuge: rapid and sensitive classification of metagenomic sequences. Genome Res. 2016;26(12):1721–9. https://doi.org/10.1101/gr.210641.116. PMID: 27852649; PMCID: PMC5131823.
27. Koren S, Walenz BP, Berlin K, Miller JR, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 2017;27:722–36. https://doi.org/10.1101/gr.215087.116.
28. Kolmogorov M, Bickhart DM, Behsaz B, Gurevich A, Rayko M, Shin SB, Kuhn K, Yuan J, Polevikov E, Smith TPL, Pevzner PA. MetaFlye: scalable long-read metagenome assembly using repeat graphs. Nat Methods, 17(11):1103–10. https://doi.org/10.1038/s41592-020-00971-x. Epub 2020 Oct 5. PMID: 33020656; PMCID: PMC10699202.
29. Wick RR, Judd LM, Gorrie CL, Holt KE, Unicycler. Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol. 2017;13(6):e1005595. https://doi.org/10.1371/journal.pcbi.1005595.
30. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34:3094–100. https://doi.org/10.1093/bioinformatics/bty191.
31. Li H. New strategies to improve minimap2 alignment accuracy. Bioinformatics. 2021;37:4572–4. https://doi.org/10.1093/bioinformatics/btab705.
32. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative genomics viewer. Nat Biotechnol. 2011;29(1):24–6. https://doi.org/10.1038/nbt.1754.
33. Haegeman A, Zro K, Vandenbussche F, Demeestere L, Van Campe W, Ennaji MM, De Clercq K. Development and validation of three capripoxvirus real-time PCRs for parallel testing. J Virol Methods. 2013;193:446–51. https://doi.org/10.1016/j.jviromet.2013.07.010.
34. Vereecke N, Kvisgaard LK, Baele G, Boone C, Kunze M, Larsen LE, Theuns S, Nauwynck H. Molecular epidemiology of Porcine parvovirus type 1 (PPV1) and the reactivity of vaccine-induced antisera against historical and current PPV1 strains. Virus Evol. 2022;8(1):veac053. https://doi.org/10.1093/ve/veac053. PMID: 35815310; PMCID: PMC9252332.
35. Vereecke N, Zwickl S, Gumbert S, Graaf A, Harder T, Ritzmann M, Lillie-Jaschniski K, Theuns S, Stadler J. Viral and bacterial profiles in endemic influenza A virus infected swine herds using nanopore metagenomic sequencing on tracheobronchial swabs. Microbiol Spectr. 2023;28(2):e0009823. https://doi.org/10.1128/spectrum.00098-23. PMID: 36853049; PMCID: PMC10100764.
36. Oxford Nanopore Technologies Ltd. 2024. https://store.nanoporetech.com/eu/productDetail/?id=rapid-sequencing-kit-v14

37. De Coster W, Rademakers R. NanoPack2: population-scale evaluation of long-read sequencing data. Bioinformatics. 2023;39:5:btad311. https://doi.org/10.1093/bioinformatics/btad311.
38. Oxford Nanopore Technologies Ltd. 2018. https://github.com/nanoporetech/medaka
39. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754–60. https://doi.org/10.1093/bioinformatics/btp324.
40. Li H, Durbin R. Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics. 2010;26(5):589–95. https://doi.org/10.1093/bioinformatics/btp698.
41. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H. Twelve years of SAMtools and BCFtools, GigaScience. 2021;10(2). https://doi.org/10.1093/gigascience/giab008
42. Milne I, Stephen G, Bayer M, Cock PJA, Pritchard L, Cardle L, Shaw PD, Marshall D. Using tablet for visual exploration of second-generation sequencing data. Brief Bioinform. 2013;14(2):193–202. https://doi.org/10.1093/bib/bbs012.
43. Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. Nat Methods. 2024;9(4):357–9. https://doi.org/10.1038/nmeth.1923.
44. Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, Connor R, Funk K, Kelly C, Kim S, Madej T, Marchler-Bauer A, Lanczycki C, Lathrop S, Lu Z, Thibaud-Nissen F, Murphy T, Phan L, Skripchenko Y, Tse T, Wang J, Williams R, Trawick BW, Pruitt KD, Sherry ST. Database resources of the National centre for biotechnology information. Nucleic Acids Res. 2022;750(D1):D20–6. https://doi.org/10.1093/nar/gkab1112.
45. Camacho C, Coulouris G, Avagyan V, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421. https://doi.org/10.1186/1471-2105-10-421.
46. Croville G, Le Loc'h G, Zanchetta C, Manno M, Camus-Bouclainville C, Klopp C, Delverdier M, Lucas MN, Donnadieu C, Delpont M, Guérin JL. Rapid whole-genome based typing and surveillance of avipoxviruses using nanopore sequencing. J Virol Methods. 2018;261:34–9. Epub 2018 Aug 4. PMID: 30086381.
47. Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A. July. Versatile genome assembly evaluation with QUAST-LG, *Bioinformatics*, Volume 34, Issue 13, 2018, Pages i142–i150. https://doi.org/10.1093/bioinformatics/bty266
48. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. 2016. ISBN 978-3-319-24277-4. https://ggplot2.tidyverse.org
49. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform. 2019;20(4):1160–6. https://doi.org/10.1093/bib/bbx108.
50. Sanjuán R, Nebot MR, Chirico N, Mansky LM, Belshaw R. Viral mutation rates. J Virol. 2010;84(19):9733–48. https://doi.org/10.1128/JVI.00694-10. PMID: 20660197; PMCID: PMC2937809.
51. Ribarska T, Bjørnstad PM, Sundaram AYM, Gilfillan GD. Optimization of enzymatic fragmentation is crucial to maximize genome coverage: a comparison of library Preparation methods for illumina sequencing. BMC Genomics. 2022;1(1):92. https://doi.org/10.1186/s12864-022-08316-y. PMID: 35105301; PMCID: PMC8805253.
52. Bruinsma S, Burgess J, Schlingman D, et al. Bead-linked transposomes enable a normalization-free workflow for NGS library Preparation. BMC Genomics. 2018;19:722. https://doi.org/10.1186/s12864-018-5096-9.
53. Cook R, Brown N, Rihtman B, Michniewski S, Redgwell T, Clokie M, Stekel DJ, Chen Y, Scanlan DJ, Hobman JL, Nelson A, Jones MA, Smith D, Millard A. The long and short of it: benchmarking viromics using Illumina, nanopore and PacBio sequencing technologies. Microb Genom. 2024;10(2):001198. https://doi.org/10.1099/mgen.0.001198. PMID: 38376377; PMCID: PMC10926689.
54. Graf A, Rziha HJ, Krebs S, Wolf E, Blum H, Büttner M. *Parapoxvirus* species revisited by whole genome sequencing: A retrospective analysis of bovine virus isolates. Virus Res. 2024;346:199404. https://doi.org/10.1016/j.virusres.2024.199404.
55. Quail MA, Smith M, Coupland P, et al. A Tale of three next generation sequencing platforms: comparison of ion Torrent, Pacific biosciences and illumina miseq sequencers. BMC Genomics. 2012;13:341. https://doi.org/10.1186/1471-2164-13-34.
56. Sims D, Sudbery I, Ilott NE, Heger A, Ponting CP. Sequencing depth and coverage: key considerations in genomic analyses. Nat Rev Genet. 2014;15(2):121 – 32. https://doi.org/10.1038/nrg3642. PMID: 24434847.
57. Oxford Nanopore Technologies PLC. 2024. https://github.com/nanoporetech/dorado/
58. Gershon PD, Black DN. A comparison of the genomes of Capripoxvirus isolates of sheep, goats, and cattle. Virology. 1988;164(2):341-9. doi https://doi.org/10.1016/0042-6822(88)90547-8. PMID: 2835855.
59. Tulman ER, Afonso CL, Lu Z, Zsak L, Kutish GF, Rock DL. Genome of lumpy skin disease virus. J Virol. 2001;75:7122–30. https://doi.org/10.1128/JVI.75.15.7122-7130.2001.
60. Fang H, Wu Y, Narzisi G, O'Rawe JA, Barrón LT, Rosenbaum J, Ronemus M, Iossifov I, Schatz MC, Lyon GJ. Reducing INDEL calling errors in whole genome and exome sequencing data. Genome Med. 2014;6(10):89. https://doi.org/10.1186/s13073-014-0089-z. PMID: 25426171; PMCID: PMC4240813.
61. McInnes CJ, Damon IK, Smith GL, McFadden G, Isaacs SN, Roper RL, Evans DH, Damaso CR, Carulei O, Wise LM, Lefkowitz EJ. ICTV virus taxonomy profile: poxviridae 2023. J Gen Virol. 2023;104(5):001849. https://doi.org/10.1099/jgv.0.001849. PMID: 37195882; PMCID: PMC12447285.

Publisher's Note