A high-quality telomere-to-telomere LSDV genome assembly

Authors:
Caroline Wright[1] (ORCID: 0000-0002-1770-4325), Noemi Polo[1] (ORCID: 0000-0002-0470-2611), Sarwar Azam[2] (ORCID: 0000-0001-7834-483X), Graham Freimanis[1] (ORCID: 0000-0001-8186-2638), G. Taru Sharma[2,3] (ORCID: 0000-0002-5052-9477), Bryan Charleston[1] (ORCID: 0000-0002-6952-9338), Carrie Batten[1] (ORCID: 0000-0003-4274-7375), Tim Downing[1,*] (ORCID: 0000-0002-8385-6730), Jonas Albarnaz[1,*] (ORCID: 0000-0002-8792-813X).

Affiliations:
[1]The Pirbright Institute, Surrey, UK.
[2]National Institute of Animal Biotechnology (NIAB), Hyderabad, India.
[3]Indian Veterinary Research Institute (IVRI), Izatnagar, India.
*These authors contributed equally to this work. To whom correspondence should be addressed: jonas.albarnaz@pirbright.ac.uk, tim.downing@pirbright.ac.uk

Abstract

Lumpy skin disease virus (LSDV) is an emerging livestock capripoxvirus (CaPV) that continues to cause substantial economic losses across Africa, Asia and Europe. However, important uncertainties remain regarding LSDV genome structure, particularly at the telomeric inverted terminal repeats (ITRs) that are central to host interaction, replication and adaptive evolution. Structural variation at the ITRs is driven by recombination during virus replication. Previous assemblies have relied primarily on short-read sequencing, which provides limited resolution of repeat-rich telomeric regions and propagates structural ambiguities into downstream annotation and comparative analyses. Here, we present a high-quality telomere-to-telomere (T2T) assembly of the LSDV Oman 2009 isolate, generated using a hybrid approach that integrates short Illumina reads with higher-accuracy long Nanopore reads. The resulting 151,091 bp genome contains 157 annotated open reading frames and fully resolves complex repeat-rich structures at both ITRs. This new reference corrects misassemblies present in earlier LSDV genomes and confirms clade-specific gene truncations. This genome provides a robust foundation for improved genomic surveillance, accurate read mapping, mutation detection and evolutionary inference of LSDV, and demonstrates the value of long-read approaches for resolving complex CaPV genome structures.

Keywords: lumpy skin disease, LSDV, poxvirus, hybrid assembly, genome.

Introduction

Lumpy skin disease virus (LSDV, *Capripoxvirus lumpypox*) is an endemic livestock disease in Africa and an emerging threat to bovine health, predominantly in Asia and southern Europe. Owing to its substantial economic impact and rapid transboundary spread, lumpy skin disease is a notifiable disease by the World Organization for Animal Health (WOAH) (WOAH 2010). Although genomic variation clearly influences LSDV virulence and transmission dynamics, large gaps remain in our understanding of its genome structure, particularly within the highly repetitive terminal regions. These limitations can now be addressed using new technological developments in long-read sequencing approaches.

LSDV is a dsDNA virus that belongs to the *Capripoxvirus* (CaPV) genus of the *Poxviridae* family and is typically classified into three main clades (1.1, 1.2 and 2), with clade 2 representing a recombinant lineage derived from ancestors of clades 1.1 and 1.2 (Biswas et al 2021, Haga et al 2024). Currently circulating field strains are predominantly associated with clade 1.1, whereas clade 1.2 is linked to live attenuated vaccine strains. Most existing LSDV genomes are assembled solely from short-read data, leading to fragmented contigs, incomplete annotation, and uncertainty at key loci with high structural complexity. This is especially true for the genes at the terminal 5’ and 3’ ends, e.g. LSDV001 and LSDV156, which are important for host immune modulation (Zhang et al 2025). Short-read assemblies may obscure true biological variation in these regions and introduce artefacts that confound phylogenetic interpretation.

The linear 151 kb LSDV genome has a structure typical to poxviruses. It contains 156 annotated ORFs and is flanked by homologous inverted terminal repeats (ITRs) that harbour four ORFs each (ORFs 1-4 and 153-156, respectively) (Tulman et al 2001) and contain regulatory sequences essential during virus replication. Poxvirus genomes have the two DNA strands covalently linked at the termini forming an incompletely base-paired hairpin analogous to telomeres at the end of each ITR. Poxviruses replicate their genomes through head-to-head and tail-to-tail concatemeric intermediates, and the conserved sequence that is necessary for concatemer resolution is localised between the hairpins and a series of short tandem repeats (Moyer & Graves 1981, Moss 2013). The ITRs and the adjacent 5’ and 3’ regions spanning ORFs 5-23 and 124-152 are among the most variable and encode functions associated with virulence and immune modulation, whereas the central region is highly conserved and encodes genes involved in transcription, DNA replication, and virion assembly. Despite their biological importance, the structure of LSDV ITRs remains poorly defined. By contrast, the ITRs of other poxviruses, such as vaccinia virus (Moss 2013), have been resolved in detail, including long A/T-rich repeats, suggesting a potentially analogous but uncharacterised architecture in LSDV.

Here, we generate a telomere-to-telomere (T2T) annotated reference genome for the Oman 2009 isolate of LSDV by integrating short Illumina reads with long Oxford Nanopore Technology (ONT) reads. This approach used these long reads as the main scaffold for the resulting assembly. It provided an enhanced resolution of the genome, particularly at the repetitive terminal regions, enabling the resolution of complex ITR structures that were previously inaccessible. The resulting high-quality genome provides an improved reference for accurate read mapping, comparative genomics, and evolutionary analyses of LSDV, and as a template for future CaPV hybrid assemblies.

Methods

Sample collection

The LSDV/OMAN/2009/MDBKP2 isolate (‘Oman 2009’) was obtained from the Non-Vesicular Reference Laboratory (NVRL) at The Pirbright Institute. The virus was isolated from a cattle skin nodule biopsy after passaging twice in Madin-Darby bovine kidney (MDBK) cells confirmed free of bovine viral diarrhoeal virus (BVDV) by qRT-PCR, following WOAH protocol guidelines (WOAH Terrestrial Manual 2024).

Read library preparation and sequencing

For short-read Illumina sequencing, total DNA was extracted from 100 µL of cell culture lysate using a MagMax CORE nucleic acid purification kit and a KingFisher Flex automated extraction robot (Thermo Fisher Scientific) according to the manufacturer's instructions. The extracted DNA was tagmented and indexed using an Illumina DNA Prep with Enrichment kit (Illumina), with 11 cycles of amplification used during DNA tagmentation. The indexed library underwent targeted enrichment with a custom myBaits panel (Daicel Arbor Biosciences), also used in Breman et al 2026 (sequences available as supplementary material). Manufacturer’s instructions were followed for the High Sensitivity Protocol, which included two rounds of probe hybridisation at 63°C. The final library size distribution was determined using a D1000 High Sensitivity ScreenTape on a TapeStation 4200 (Agilent) and quantitated using the Qubit HS kit (Thermo Fisher Scientific). The library was then diluted and re-quantified using the NEBNext Library Quant Kit for Illumina (New England Biolabs) assay. The diluted library was loaded onto an Illumina Nextseq 2000 at a final concentration of 1.6 pM for paired-end sequencing (150 cycles) with a 2% PhiX spike-in.

For long-read ONT sequencing, library preparation was performed using the ONT ligation by sequencing kit (SQK-LSK114) and an input of 340 ng genomic DNA. Final library size distribution was determined using the D1000 genomic DNA ScreenTape on a TapeStation 4200 (Agilent) and quantified using the Qubit DNA BR kit (Thermo Fisher Scientific). Thirty femtomoles of the library were loaded onto a MinION R10.4.1 (FLO-MIN114) flow cell and sequenced for 42 hours on a MinION Mk1B device. Basecalling was performed using Dorado 7.3.11.

Read library quality control and host read removal

Adapter trimming, removal of both short and long reads with low base quality scores (phred score <30), exclusion of ambiguous (N) bases, mismatched base pair correction in overlapping regions, and cutting of poly-G tracts at 3′ ends was completed using Fastp v0.23.4 (Chen et al 2018). Low-quality ONT bases were removed with Nanoq including trimming the 5’ 11 and 3’ 12 bases from all reads (Steinig & Coin 2022). Low-quality Illumina bases with base qualities <30 and reads with lengths <110 bp were removed using the FASTX-Toolkit v0.0.13 (http://hannonlab.cshl.edu/fastx_toolkit/). FastQC v0.12.1 (www.bioinformatics.babraham.ac.uk/projects/fastqc/) was used to verify the effectiveness of Fastp, the FASTX-Toolkit and Nanoq. We estimated species-level abundances for each read using the k-mer classifier Kraken2 (Wood et al 2019).

The initial ONT read library lengths ranged 35-82,438 bp (median 557 bp) (Table 1). A total of 209,203 LSDV reads had lengths 35-80,952 bp (median 533 bp). This corresponded to 1.9% of the post-QC reads (or 187.06 Mbp of data), reflecting the absence of ONT library target enrichment. This generated a mean genome-wide read depth of 953-fold when mapped to the final assembly with a mean base quality (BQ) of 34.3. The paired-end Illumina data was 99% LSDV because probes were used to enrich that library for LSDV. This resulted in 39.3 million 151 bp LSDV reads, resulting in a mean genome-wide read depth of 10,627-fold with a mean BQ of 39.3.

Technology	Cow	Human	Deer	Sheep	LSDV	Cow %	Human %	Deer %	Sheep %	LSDV %

ONT	7,632,866	175,640	1,940,662	623,899	202,760	72.2%	1.7%	18.3%	5.9%	1.9%
Illumina	100,795	98,292	98,340	98,009	39,323,196	0.3%	0.2%	0.2%	0.2%	99.0%

Table 1. The numbers and percentages of reads classified across organisms by Kraken2 for our ONT and Illumina read libraries. The Illumina reads were enriched for LSDV prior to sequencing.

De novo genome assembly and evaluation

Flye v2.9.1-b1784 was used to assemble the genome from the ONT reads (Kolmogorov et al 2019). This used a subset of the longest reads for initial disjointig assembly, ten polishing iterations, an expected genome size of 151 kb, and high-quality ONT reads. No Flye polishing improvements were observed beyond four iterations. We created assemblies with ONT reads subject to a size requirement of >500 bp, >1,000 bp, >2,000 bp and >4,000 bp that either had no effect or assembled less well. Additionally, reducing the coverage for the initial disjointig assembly was attempted but did not improve assembly quality. Assemblies with alternative contigs were excessively long and impaired overall quality. SPAdes v3.13.1 (Prjibelski et al 2020) was run in careful mode using a range of input: ONT reads, Illumina reads, both read sets, and the same combinations again using contigs from the optimal Flye assembly. We also used Unicycler v0.4.8 (Wick et al 2017) to create Illumina-only, ONT-only and hybrid assemblies using conservative, normal and bold modes.

Assembly quality was measured with Quast based on the assembly length, number of contigs, largest contig size, GC content, N50 and number of N bases. To explore local assembly quality, read pileups and breakpoints, we used Minimap2 v2.24-r1122 (Li et al 2018) to index the assembly and map the reads to it. The resulting Sequence Alignment Map (SAM) files were processed with SAMtools v1.19.2 (Li et al 2009) to process, sort and estimate read depth per site. The Binary Alignment Map (BAM) files stemming from this were examined with IGV v2.16.0 (Robinson et al 2011) for read alignment and paired-end read consistency inspection. Read-depth variation was visualised using R v4.3.2 (R Core Team 2025) packages ggplot2 v3.5.1 (Wickham 2016) and gridExtra v2.3 (Auguie & Antonov 2017).

Post assembly polishing and genome annotation

The assembly was polished using four steps with Medaka, Pilon, Polypolish and Pypolca. Firstly, Medaka v2.0.1 (https://github.com/nanoporetech/medaka) was used to polish the optimal Flye assembly to correct large indel and homopolymer errors using the cleaned ONT reads with a r1041_e82_400bps_sup_v5.0.0 model. The assemblies were compared using Nucmer v3.1 (Marçais et al 2018). Secondly, we used Pilon v1.24 (Walker et al 2014) to fix SNPs, indels, gaps, local misassemblies, and ambiguous bases over five iterations using the assembly resulting from Medaka. Thirdly, Polypolish v0.6.0 corrected errors in repeats in the Pilon output (Wick & Holt 2022). To run Polypolish, BWA v0.7.17 mapped the Illumina reads to all possible matches where forward and reverse reads were mapped separately. These reads were filtered by insert size and used for polishing. Fourthly, Pypolca v0.3.1 in careful mode was used to do a final correction step on the Polypolish assembly using the Illumina reads (Bouras et al 2024). Canu (Koren et al 2017) was used to error-correct and assemble the ONT reads but this provided no improvement.

The assembly versions were annotated using Prokka v1.14.6 (Seeman 2014) with the --kingdom Viruses option specified, with gene annotation guided by the reference genome KX894508 (155920/2012; 156 ORFs). Annotation was also mapped from the reference strain LSDV NI-2490 (AF325528.1) using Genome Annotation Transfer Utility (GATU) (Tcherepanov et al 2006). The annotation was manually inspected by visualising long- and short-read alignments in IGV v2.16.0 (Robinson et al 2011) to ensure the gene boundaries were correct. Genome completeness and quality, including ITRs within the terminal regions, was further assessed using CheckV v1.0.3 (Nayfach et al 2021).

Results and Discussion

A more accurate LSDV reference genome

We used a combination of long-read ONT and short-read Illumina sequencing data to *de novo* assemble a new reference genome sequence for the LSDV Oman 2009 isolate. The final assembly comprised a single contig of 151,091 bp (N50 151,091 bp), with 157 predicted ORFs and a mean GC content of 25.9% (median 25.6%, Figure S1). This contig was generated from long reads using Flye v2.9.1-b1784 (Kolmogorov et al 2019) because Flye has a low chance of making large-scale errors (Wick et al 2023). It was refined through iterative polishing with Medaka v2.0.1 that made 15 changes, Pilon v1.24 (Walker et al 2014) that fixed two SNP errors, and correction with Polypolish v0.6.0 (Bouras et al 2024) that fixed nine indel and two substitution errors. Alternative assembly strategies resulted in fragmented assemblies with inflated total lengths and numbers of contigs. Overall, polishing with Illumina reads corrected only 28 bp (0.019%) of the 151,091 bp genome assembled from long ONT reads.

The genomic termini of the Oman 2009 reference genome were verified using a known 123 bp hairpin sequence (MN072625.1, Biswas et al 2020), which had T-A differences at bases 30 and 33. The terminal 23 bp of the hairpin contained concatemer resolution sequences (CRSs), which mediate site-specific genome resolution during viral replication (Moss 2013). The 5’ and 3’ ITRs were sequence-symmetrical and contained CRSs at homologous positions, matching a previously described GTPV CRS (Biswas et al 2020).

Sequence	Position
TACTTTTTTTTTAGCCTATAAAT	1-23, 150,969-150,991
ATTTATAGGCTAAAAAAAAAGTA	101-123, 151,068-151,091

Table 2. CRS nucleotide sequence at the ITRs in the LSDV Oman 2009 genome assembly.

To assess assembly quality, we visualised per-base read depth for both short- and long-read data across the genome (Figure 1). Long-read depth was relatively uniform, with a standard deviation of 66 (7% of the median depth of 962), compared with greater variability in short-read depth (standard deviation 2,433; 23% of the median depth of 10,379). Examination of the genome annotation additionally identified two truncated genes, LSDV019 and LSDV026, which are also truncated in other wild-type LSDV isolates from clade 1.1 (Bamouh et al 2021, Kumar et al 2023, Chang et al 2025). Gene fragments LSDV019a/b and LSDV026a/bwere present as truncated coding sequences matching LSDV019 and LSDV026, respectively, in the Neethling isolate NI-2490 from clade 1.2 (AF325528.1), where the genes are full length. LSDV019 is homologous to kelch-like virulence factors characterised in orthopoxviruses (Pires de Miranda et al 2003, Beard et al 2006, Froggatt et al 2007). LSDV026 is also truncated in SPPV and GTPV and is an orthologue of vaccinia virus F11, a virulence factor involved in regulation of virus spread (Cordeiro et al 2009). Furthermore, we annotated a new ORF, tentatively named LSDV042.5, which is predicted to be an orthologue of vaccinia virus O3L. O3L encodes a 35-amino-acid protein that is a component of the entry fusion complex required for virus penetration into host cells and conserved across poxviruses of vertebrates (Satheshkumar & Moss 2009). LSDV042.5 encodes a predicted polypeptide of 29 amino acid residues and therefore, was not annotated in the first LSDV genome sequenced (AF325528.1) even when a sensitive cutoff (ORFs >30 amino acid residues) was applied (Tulman et al 2001). Nonetheless, LSDV042.5 is likely to be functional because the O3L-like gene from SPPV (AY077832.1), which is 93% identical to LSDV042.5, can complement O3L deficiency in a recombinant vaccinia virus (Satheshkumar & Moss 2009).

The reference genome contained a 5’ ITR spanning positions 1-2,547 bp and a corresponding 3’ ITR spanning positions 148,557-151,091 bp, yielding ITR lengths of approximately 2.5 kb at each terminus. This annotation was supported by the genomic positions of conserved flanking genes, including LSDV005 (2,599-3,111 bp) at the 5’ end and an ankyrin repeat protein (LSDV152; 146,929-148,398) at

the 3’ end. Analysis of read depth revealed abrupt changes in short-read coverage at approximately 1.2, 3.3, 5.8 and 6.4 kb from the 5’ end, with corresponding features observed at the 3’ end of the genome (Figure 1). These coverage transitions were present when short reads, but not long reads, were mapped to the reference. Median short-read depth across the 5’and 3’ ITRs (12,580 and 12,592, respectively) exceeded the genome-wide median short-read depth of 10,379, indicating overrepresentation of these regions due to their repetitive structure. This suggests that short reads alone overestimate ITR length by 22%, whereas long reads spanned the repetitive regions and enabled accurate estimation of ITR structure and length.

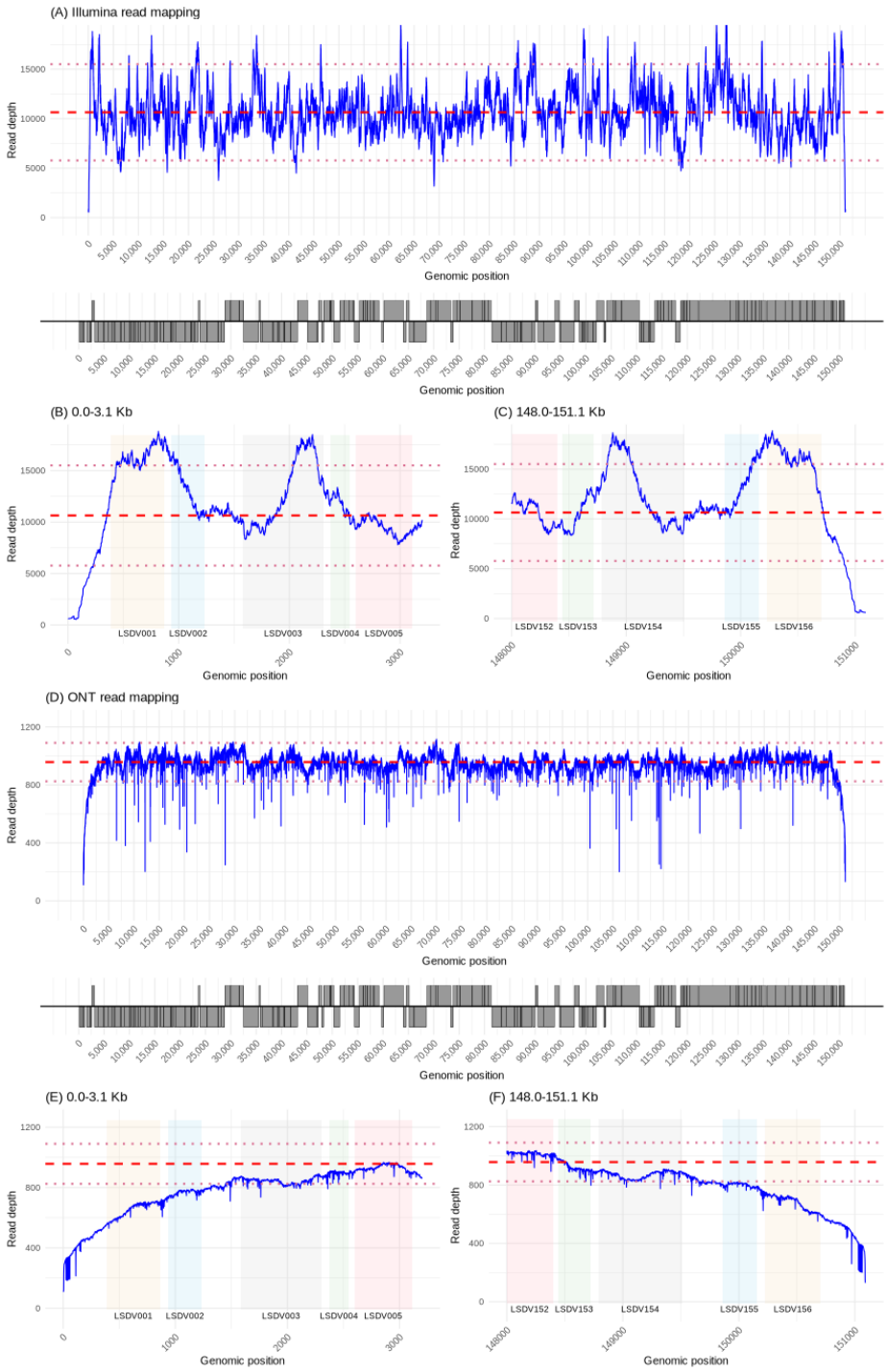


Figure 1. Per-base read depth across the LSDV Oman 2009 assembly for Illumina data (A) genome-wide, (B) at 0-3.1 kb and (C) at 148-151.1 kb, and equivalent ONT data within the same regions (D) genome-wide, (E) at 0-3.1 kb and (F) at 148-151.1 kb. The blue lines show read depth per site. The red dashed

lines show genome-wide average read depth. The red dotted lines indicate twice the SD above and below the average. The small black boxes in (E) indicates the CRS. The ITRs are at 1-2,547 bp and 148,557-151,091 bp. The grey panels below (A) and (D) represent locations of annotated CDSs across the genome. Coloured shading in (B), (C), (E) and (F) demarcates LSDV001 (beige), LSDV002 (pale blue), LSDV003 (pale purple), LSDV004 (pale green), LSDV005 (pink), LSDV152 (pink), LSDV153 (pale green), LSDV154 (pale purple), LSDV155 (pale blue) and LSDV156 (beige).

Alignment of the Oman 2009 hybrid assembly with the PacBio-based genome assembly PX492334 (Breman et al 2026) revealed only 12 mismatches across the full genome length. This high level of concordance indicates that both ONT and PacBio sequencing approaches can generate highly accurate assemblies for LSDV. Comparison of the terminal regions showed that the first 67 bp of the Oman 2009 assembly, including the first CRS, were not present in the 5' ITR of PX492334, where only a partial 11 bp match was observed; an identical pattern was detected at the 3' end. Importantly, the presence of this additional sequence symmetrically at both genome termini in the Oman 2009 assembly supports its authenticity and suggests that the hybrid ONT-Illumina approach resolved an additional 56 bp at each end of the genome relative to the PacBio-only assembly.

Summary

This study used a hybrid short- and long-read sequencing strategy to generate an annotated reference genome for the LSDV Oman 2009 isolate, resolving the architecture of the ITRs more accurately than previously available assemblies. The results provide a framework for future hybrid genome assemblies of CaPVs and demonstrate the effectiveness of ONT R10 long-read sequencing for producing high-quality CaPV genomes. Consistent with observations in bacterial systems (Sereika et al 2022), our findings suggest that long-read data alone may be sufficient for accurate CaPV genome assembly, while short-read data are more effectively applied to resequencing and read mapping (Wright et al 2025). The improved resolution of ITRs and truncated genes in the Oman 2009 genome supports its use as a reliable LSDV reference and facilitates more accurate clade-specific analyses and diagnostic assay development (Tennakoon et al 2025). Continued improvements in sequencing technologies, including ultra-long reads, alterative long read platforms and improved polymerase accuracy, are likely to enable further refinement of LSDV and CaPV genome assemblies.

Data Availability

The Illumina and ONT reads are on the SRA under the BioProject accession PRJNA1366067. The ONT reads are available under the accession SRX31142208. The Illumina reads are available under the accession SRX31141470. The genome assembly file and its annotation files are on the Nucleotide database under the accession PV877838. The code used to produce these analyses is publicly available at https://github.com/downingtim/LSDV_hybrid_assembly/

Acknowledgements

We would like to acknowledge the Pirbright Institute's Bioinformatics, and High-Throughput Sequencing STPs and Non-Vesicular Reference Laboratory (NVRL). We also acknowledge the use of generative artificial intelligence tools (ChatGPT, Copilot) to write code used for this work, and to proofread the text. All outputs were reviewed, tested, and/or validated by the authors.

Funding

This work was supported by funding from UK Research and Innovation (UKRI) Biotechnology and Biological Sciences Research Council (BBSRC) grants BBS/E/PI/230002A, BBS/E/PI/230002B, BBS/E/PI/230002C and BBS/E/PI/23NB0003. CW was funded by UK Department for Environment, Food & Rural Affairs (Defra) (grant C17508; Provision of National Reference Laboratory, Disease Surveillance and Outbreak Response Function for Vesicular and Non-Vesicular Disease; awarded to CB),

and NP was supported by Defra (grant SE2623; awarded to CB). This material was also partially funded by UK International Development from the UK government (Lumpy Skin Disease – developing a synthetic genome to produce safe and effective vaccines; INT IND 2324 001); however, the views expressed do not necessarily reflect the UK government's official policies.

Conflicts of Interest

The authors declare no conflicts of interest.

Figure legends

Figure 1. Per-base read depth across the LSDV Oman 2009 assembly for Illumina data (A) genome-wide, (B) at 0-3.1 kb and (C) at 148-151.1 kb, and equivalent ONT data within the same regions (D) genome-wide, (E) at 0-3.1 kb and (F) at 148-151.1 kb. The blue lines show read depth per site. The red dashed lines show genome-wide average read depth. The red dotted lines indicate twice the SD above and below the average. The small black boxes in (E) and (F) indicate the CRSs. The ITRs are at 1-2,547 bp and 148,557-151,091 bp. The grey panels below (A) and (D) represent locations of annotated CDSs across the genome. Coloured shading in (B), (C), (E) and (F) demarcates LSDV001 (beige), LSDV002 (light blue), LSDV003 (pale purple), LSDV004 (pale green), LSDV005 (pink), LSDV152 (pink), LSDV153 (pale green), LSDV154 (pale purple), LSDV155 (pale blue) and LSDV156 (beige).

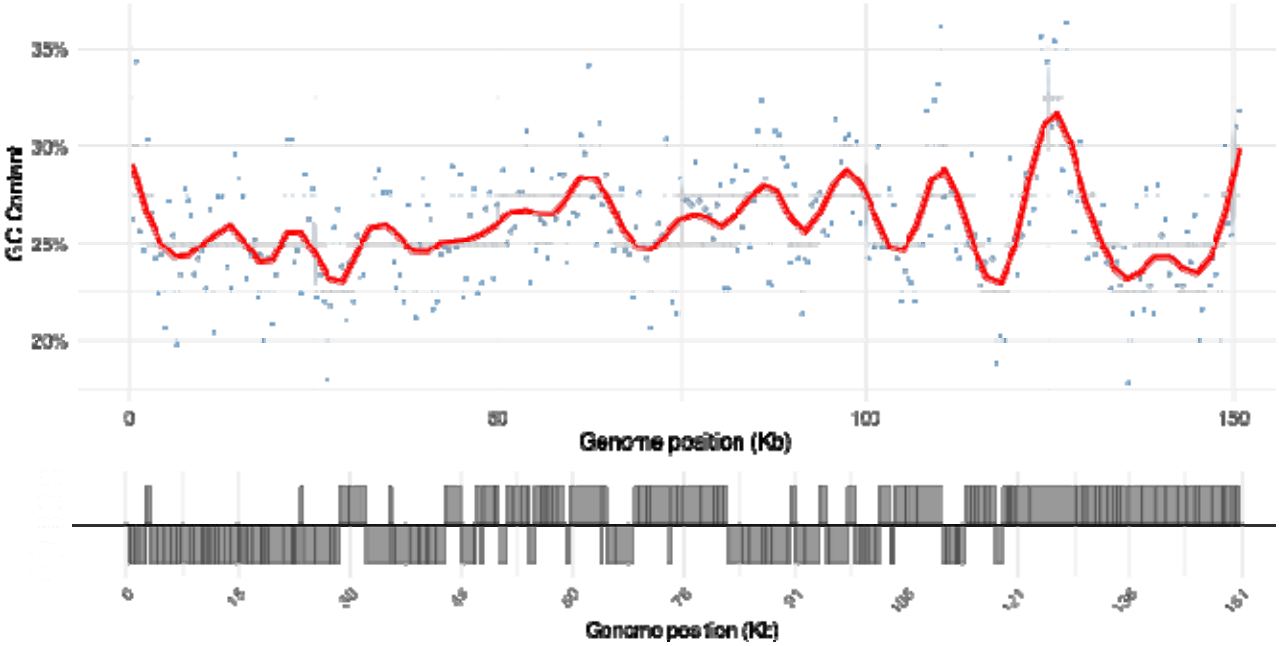


Figure S1. The genome-wide variation in GC content in the LSDV Oman 2009 hybrid genome assembly. The red line shows the smooth pattern in 500 bp bins and the grey area shows the 95% confidence interval. The grey panel represents the locations of the annotated CDSs across the genome. The y-axis has been cut.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990 215(3):403-10. doi: 10.1016/S0022-2836(05)80360-2

Auguie B, Antonov A. gridExtra: Miscellaneous Functions for “Grid” Graphics. 2017. https://CRAN.R-project.org/package=gridExtra

Bamouh Z, Fellahi S, Khayi S, Hamdi J, Omari Tadlaoui K, Fassi-Fihri O, Elharrak M. Draft Genome Sequence of the Capripoxvirus Vaccine Strain KSGP 0240, Reisolated from Cattle. Microbiol Resour Announc. 2021 10(30):e0044021. doi: 10.1128/MRA.00440-21.

Beard PM, Froggatt GC, Smith GL. Vaccinia virus kelch protein A55 is a 64 kDa intracellular factor that affects virus-induced cytopathic effect and the outcome of infection in a murine intradermal model. J Gen Virol. 2006 87(Pt 6):1521-1529. doi: 10.1099/vir.0.81854-0.

Biswas S, Noyce RS, Babiuk LA, Lung O, Bulach DM, Bowden TR, Boyle DB, Babiuk S, Evans DH. Extended sequencing of vaccine and wild-type capripoxvirus isolates provides insights into genes modulating virulence and host range. Transbound Emerg Dis. 2020 67(1):80-97. doi: 10.1111/tbed.13322.

Breman FC, Hoffman S, Haegeman A, Philips W, de Keersmaecker S, Mileto P, Neave MJ, Bowden TR, Settypalli TBK, Lamien CE, Batten C, Wright C, De Regge N. Comparison of whole genome sequencing approaches for Capripox viruses. BMC Genomics. 2026 27:174. doi: https://doi.org/10.1186/s12864-025-12463-3

Bouras G, Judd LM, Edwards RA, Vreugde S, Stinear TP, Wick RR. How low can you go? Short-read polishing of Oxford Nanopore bacterial genome assemblies. Microbial Genomics. 2024. doi: https://doi.org/10.1099/mgen.0.001254.

Chang W, Fang J, Zhai T, Han S, Fan W, Lei C, Wang L, Qi X, Xue Q, Wang J. Genetic Evolutionary Analysis of Lumpy Skin Disease Virus Strain Under Immune Pressure Exerted by Heterologous Goat Poxvirus Vaccines. Transbound Emerg Dis. 2025 2025:2883245. doi: 10.1155/tbed/2883245.

Cordeiro JV, Guerra S, Arakawa Y, Dodding MP, Esteban M, Way M. F11-mediated inhibition of RhoA signalling enhances the spread of vaccinia virus in vitro and in vivo in an intranasal mouse model of infection. PLoS One. 2009 4(12):e8506. doi: 10.1371/journal.pone.0008506. Erratum in: PLoS One. 2010;5(6). doi: 10.1371/annotation/0c8d44be-4bff-49b3-855a-4cba76160a03.

Froggatt GC, Smith GL, Beard PM. Vaccinia virus gene F3L encodes an intracellular protein that affects the innate immune response. J Gen Virol. 2007 88(Pt 7):1917-1921. doi: 10.1099/vir.0.82815-0.

Haga IR, Shih BB, Tore G, Polo N, Ribeca P, Gombo-Ochir D, Shura G, Tserenchimed T, Enkhbold B, Purevtseren D, Ulziibat G, Damdinjav B, Yimer L, Bari FD, Gizaw D, Adedeji AJ, Atai RB, Adole JA, Dogonyaro BB, Kumarawadu PL, Batten C, Corla A, Freimanis GL, Tennakoon C, Law A, Lycett S, Downing T, Beard PM. Sequencing and Analysis of Lumpy Skin Disease Virus Whole Genomes Reveals a New Viral Subgroup in West and Central Africa. Viruses. 2024 16(4):557. doi: 10.3390/v16040557

Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol. 2019 37(5):540-546. doi: 10.1038/s41587-019-0072-8

Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 2017 27(5):722-736. doi: 10.1101/gr.215087.116

Kumar A, Venkatesan G, Kushwaha A, Poulinlu G, Saha T, Ramakrishnan MA, Dhar P, Kumar GS, Singh RK. Genomic characterization of Lumpy Skin Disease virus (LSDV) from India: Circulation of Kenyan-like LSDV strains with unique kelch-like proteins. Acta Trop. 2023 241:106838. doi: 10.1016/j.actatropica.2023.106838.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009 25(16):2078-9. doi: 10.1093/bioinformatics/btp352

Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018 34(18):3094-3100. doi: 10.1093/bioinformatics/bty191

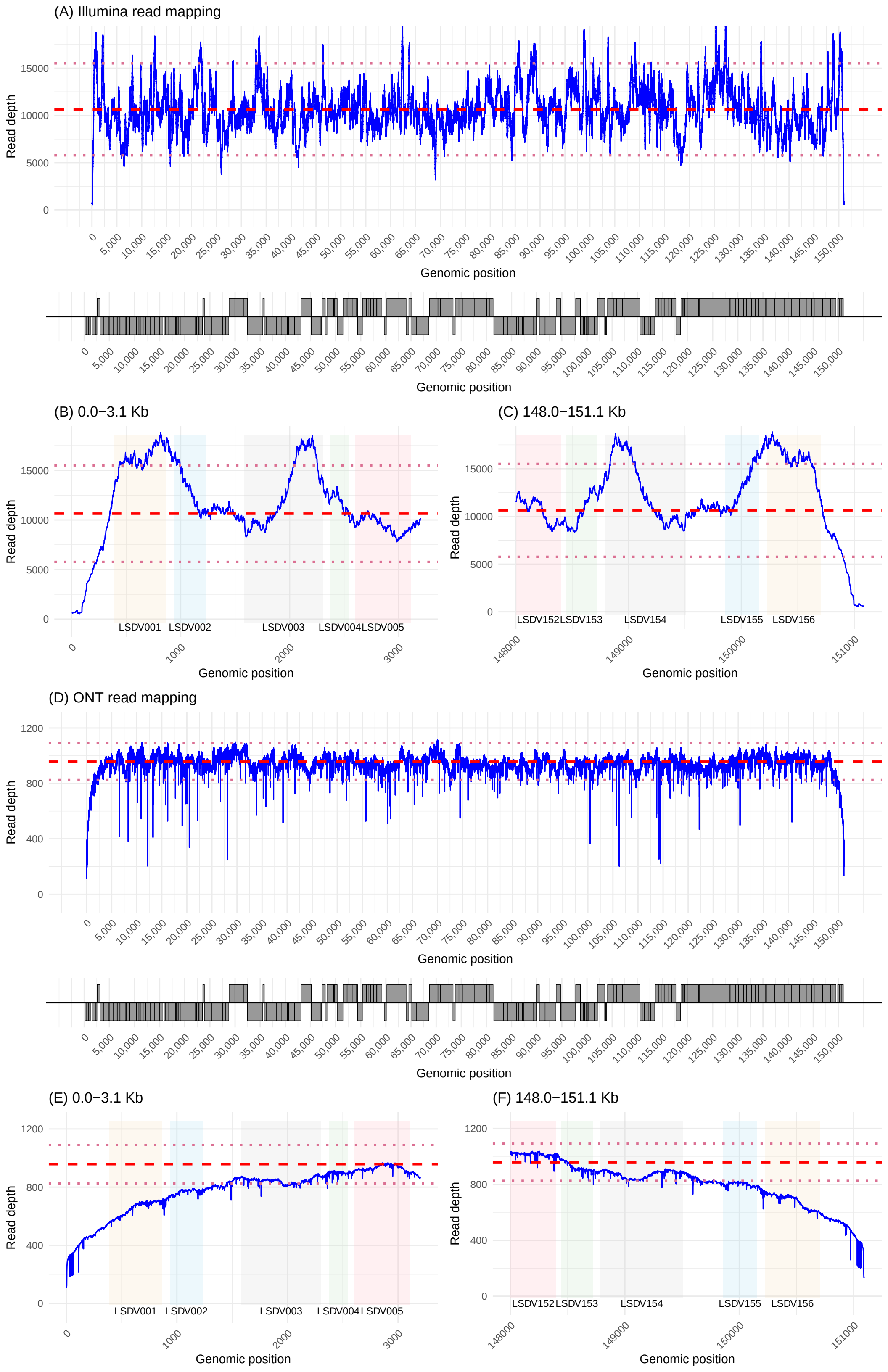
Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. MUMmer4: A fast and versatile genome alignment system. PLoS Comput Biol. 2018 14(1):e1005944. doi: 10.1371/journal.pcbi.1005944
Moss B. Poxvirus DNA replication. Cold Spring Harb Perspect Biol. 2013 5(9):a010199. doi: 10.1101/cshperspect.a010199
Moyer RW, Graves RL. The mechanism of cytoplasmic orthopoxvirus DNA replication. Cell. 1981 27(2 Pt 1):391-401. doi: 10.1016/0092-8674(81)90422-0
Nayfach S, Camargo AP, Schulz F, Eloe-Fadrosh E, Roux S, Kyrpides NC. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. Nat Biotechnol. 2021 39(5):578-585. doi: 10.1038/s41587-020-00774-7
Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using SPAdes De Novo Assembler. Curr Protoc Bioinformatics. 2020 70(1):e102. doi: 10.1002/cpbi.102
R Core Team (2025). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-pro-ject.org/
Pires de Miranda M, Reading PC, Tscharke DC, Murphy BJ, Smith GL. The vaccinia virus kelch-like protein C2L affects calcium-independent adhesion to the extracellular matrix and inflammation in a murine intradermal model. J Gen Virol. 2003 84(Pt 9):2459-2471. doi: 10.1099/vir.0.19292-0.
Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative genomics viewer. Nat Biotechnol. 2011 29(1):24-6. doi: 10.1038/nbt.1754
Satheshkumar PS, Moss B. Characterization of a newly identified 35-amino-acid component of the vaccinia virus entry/fusion complex conserved in all chordopoxviruses. J Virol. 2009 83(24):12822-32. doi: 10.1128/JVI.01744-09.
Satheshkumar PS, Moss B. Sequence-divergent chordopoxvirus homologs of the o3 protein maintain functional interactions with components of the vaccinia virus entry-fusion complex. J Virol. 2012 86(3):1696-705. doi: 10.1128/JVI.06069-11.
Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014 30(14):2068-9. doi: 10.1093/bioinformatics/btu153
Sereika M, Kirkegaard RH, Karst SM, Michaelsen TY, Sørensen EA, Wollenberg RD, Albertsen M. Oxford Nanopore R10.4 long-read sequencing enables the generation of near-finished bacterial genomes from pure cultures and metagenomes without short-read or reference polishing. Nat Methods. 2022 19(7):823-826. doi: 10.1038/s41592-022-01539-7
Shen W, Sipos B, Zhao L. SeqKit2: A Swiss army knife for sequence and alignment processing. Imeta. 2024 3(3):e191. doi: 10.1002/imt2.191
Steinig E, Coin L. Nanoq: ultra-fast quality control for nanopore reads. Journal of Open Source Software 2022 https://doi.org/10.21105/joss.02991
Tcherepanov V, Ehlers A, Upton C. Genome Annotation Transfer Utility (GATU): rapid annotation of viral genomes using a closely related reference genome. BMC Genomics. 2006 7:150. doi: 10.1186/1471-2164-7-150
Tennakoon C, Freville T, Downing T. Panalyze: automated virus pangenome variation graph construction, analysis and annotation. BioRxiv 2025 doi: https://doi.org/10.1101/2025.04.10.646565
Tulman ER, Afonso CL, Lu Z, Zsak L, Kutish GF, Rock DL. Genome of lumpy skin disease virus. J Virol. 2001 75(15):7122-30. doi: 10.1128/JVI.75.15.7122-7130.2001.
Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One. 2014 9(11):e112963. doi: 10.1371/journal.pone.0112963
Wick RR, Holt KE. Polypolish: Short-read polishing of long-read bacterial genome assemblies. PLoS Comput Biol. 2022 18(1):e1009802. doi: 10.1371/journal.pcbi.1009802.
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. doi: 10.1371/journal.pcbi.1005595
Wick RR, Judd LM, Holt KE. Assembling the perfect bacterial genome using Oxford Nanopore and Illumina sequencing. PLoS Comput Biol. 2023 19(3):e1010905. doi: 10.1371/journal.pcbi.1010905

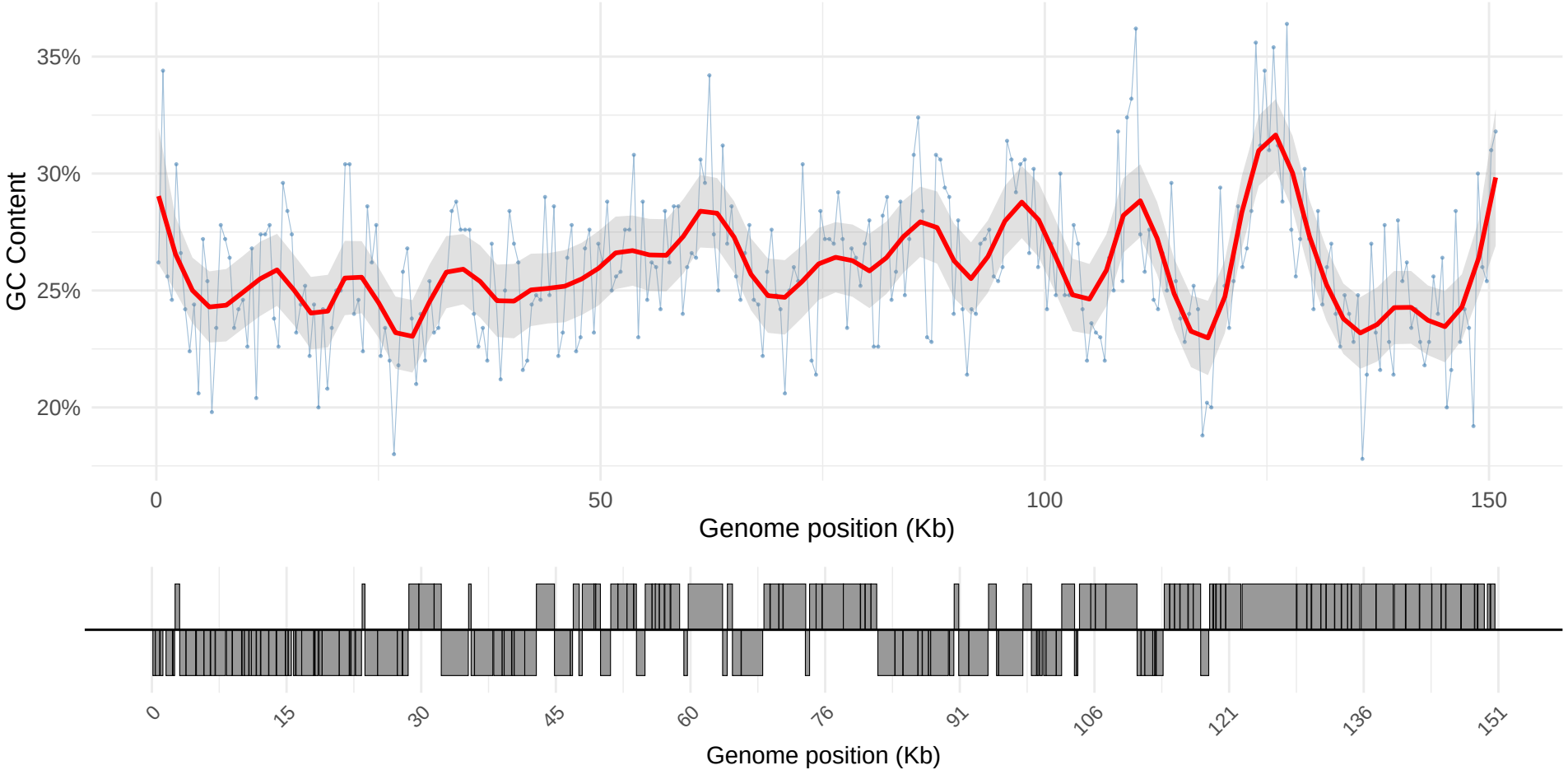
Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org

WOAH. Version adopted by the World Assembly of Delegates of the OIE in 2010, OIE, Paris.: Terrestrial Manual of Lumpy Skin Disease; 2010 [Available from: https://www.woah.org/app/uploads/2021/03/a-reso-2016-public.pdf

Wright C, Tennakoon C, Lasecka-Dykes D, Downing T. Using pangenome variation graphs to improve mutation detection in a large DNA virus. BioRxiv 2025 doi:10.1101/2025.11.26.690900

Zhang M, Shi Y, Lu X, Zhang Q, Zhao Y, Li S, Wen Z, Ge J, Wang X, Li J, Bu Z, Yin X. Lumpy skin disease virus 001/156 protein is a virulence factor that suppresses interferon production through impairing IRF3 dimerization. PLoS Pathog. 2025 21(7):e1013362. doi: 10.1371/journal.ppat.1013362.


(A) Illumina read mapping
Read depth
Genomic position
(B) 0.0−3.1 Kb
LSDV001
LSDV002
LSDV003
LSDV004
LSDV005
(C) 148.0−151.1 Kb
LSDV152
LSDV153
LSDV154
LSDV155
LSDV156
(D) ONT read mapping
(E) 0.0−3.1 Kb
(F) 148.0−151.1 Kb


GC Content
35%
30%
25%
20%
0
50
100
150
Genome position (Kb)
0
15
30
45
60
76
91
106
121
136
151
Genome position (Kb)