AMERICAN SOCIETY FOR MICROBIOLOGY

Microbiology®
Resource Announcements

Whole-Genome Sequence of an Orf Virus Isolate Derived from a Cell Culture Infected with Contagious Ecthyma Vaccine

Daniel L. Heare,[a,b] Sara V. Little,[a] Dale W. Weise,[b] James R. Harris,[b] Andrew E. Hillhouse,[a,b,c] Kranti Konganti,[c] Sara D. Lawhon[a]

aDepartment of Veterinary Pathobiology, College of Veterinary Medicine and Biomedical Sciences, Texas A&M University, College Station, Texas, USA
bTexas Vet Lab, Inc., San Angelo, Texas, USA
cTexas A&M Institute for Genome Sciences and Society, College of Veterinary Medicine and Biomedical Sciences, Texas A&M University, College Station, Texas, USA

ABSTRACT This is a draft genome of an orf virus (ORFV) vaccine strain assembled via long- and short-read hybrid assembly. ORFV is a zoonotic pathogen that affects sheep and goats. The genome of the virus contained in the vaccine was found to have high similarity (98%) to those of other published strains.

Contagious ecthyma is caused by the orf virus (ORFV), a member of the genus *Parapoxvirus* within the family *Poxviridae* that infects the skin and oral mucosa of ruminants, primarily sheep and goats, and causes painful vesicles, pustules, and proliferative lesions, typically around the mouth and nose (1, 2). The morbidity rate is approximately 89% in sheep, and the disease has a substantial economic impact (1, 3). The virus survives in scab material from ORFV lesions and can remain infectious for up to 15 years (4, 5).

A commercially available ovine ecthyma vaccine live virus (USDA product serial number 1821.51) was selected for sequencing because the virus is typically propagated in animals, and there is the potential for the vaccine strains to vary in sequence over time. Sequencing of the virus provides a reference point for future investigations if needed. The vaccine virus was used to inoculate ovine testicular cells (ATCC OA3.Ts) suspended in Dulbecco's modified Eagle's medium (Sigma-Aldrich). The infected cells were harvested after an additional 48 h of incubation by repeat freeze-thaw cycles alternating between −80°C and 25°C. DNA was extracted using a DNeasy blood and tissue kit (Qiagen) according to the manufacturer's protocol. DNA quality was verified on a genomic DNA TapeStation (Agilent) prior to sequencing. Default parameters were used for all programs except where otherwise noted.

Illumina libraries were prepared using the Illumina Nextera DNA Flex library preparation kit following the manufacturer's protocol and were sequenced with an Illumina MiSeq v3 kit (2 × 300 bp). All data were uploaded to the Illumina cloud-based resource BaseSpace for run monitoring, fastq file generation, demultiplexing, and adapter trimming. The sequencing output of paired-end read sets contained 1,725,292 reads of 301 bp. Reads under 200 bp were filtered and trimmed using Trimmomatic v0.39 (6), resulting in 1.4 million retained reads.

Nanopore libraries were prepared following the manufacturer's protocol for 1D PCR barcoding of genomic DNA using the Nanopore SQK-RAD004 rapid sequencing kit. Sequencing data collection and base calling were performed by MinKNOW v3.4.8 software with real-time base calling enabled. Sequencing resulted in an output of read sets containing 1.6 Gbp of data. There were 956,309 Nanopore reads generated and 1,676,000,000 bp obtained from the library. Reads were corrected and trimmed using Canu v1.8 (7).

Ovine DNA Illumina reads were removed using the Texas A&M Institute for Genome Sciences and Society (TIGSS) Virus Identification Pipeline with *Ovis aries* ENSEMBL build

84. In brief, the paired-end reads were combined and mapped to the host genome assembly using Bowtie 2 v2.3.4.1 in local mapping mode (8, 9). The program next filtered out any bacterial reads by mapping to the GOTTCHA bacterial signature database at the species level (10), after which the final 31,626 filtered paired-end viral reads were deinterleaved and assembled using Unicycler v0.2.3 (11), resulting in a 10-contig assembly of 124,608 bp. The Nanopore reads were mapped to this assembly using minimap2 v2.14 (12). The mapped Nanopore reads ($n = 27,573$ reads) and the Illumina viral reads were assembled together, with a coverage of over 164×, using Unicycler in conservative mode to form a hybrid assembly of 7 contigs and 165,738 bases. This assembly was scaffolded to the complete ORFV reference genome OV-IA82 (GenBank accession number AY386263.1) using Ragout v2.1.1 (13, 14), resulting in a single scaffold of 2 contigs and 5 unplaced contigs. The unplaced contigs were confirmed by BLAST to be host contamination (15) and were subsequently discarded. The reference genome has 137,241 bp, compared to the genome described here, which has 134,882 bp, a difference of 2,359 bp.

The final draft genome consisted of a single scaffold containing 2 contigs with a total 134,882 bp and a GC content of 64.14%. The largest contig was 101,329 bp and the second largest was 33,553 bp long. The scaffold N_{50} value is 134,893 bp, and the contig N_{50} value is 101,329 bp. This assembly demonstrates high levels of identity to those of other published ORFV strains, including OV-IA82 (GenBank accession number AY386263.1; identity, 98.71%), B029 (KF837136.1; identity, 99.15%), OV-SA00 (AY386264.1; identity, 98.08%), and OV-HN3/12 (KY053526.1; identity, 98.37%), when compared using NCBI BLAST (16). Relative to the OV-IA82 sequence, there was a truncation at the beginning of the sequence, as the first base pair in accession number MN454854.1 paired with base pair 3313 of accession number AY386263.1.

Data availability. This genome sequence was deposited in GenBank under BioProject number PRJNA563624 and accession number MN454854. The raw reads are available under SRA accession numbers SRR10102957 (Nanopore fast5) and SRR10102956 (Illumina fastq). This announcement represents the first version of this genome.

REFERENCES

1. Bala JA, Balakrishnan KN, Abdullah AA, Mohamed R, Haron AW, Jesse FFA, Noordin MM, Mohd-Azmi ML. 2018. The re-emerging of orf virus infection: a call for surveillance, vaccination and effective control measures. Microb Pathog 120:55–63. https://doi.org/10.1016/j.micpath.2018.04.057.

2. Jia H, Zhan L, Wang X, He X, Chen G, Zhang Y, Feng Y, Wei Y, Zhang Y, Jing Z. 2017. Transcriptome analysis of sheep oral mucosa response to orf virus infection. PLoS One 12:e0186681. https://doi.org/10.1371/journal.pone.0186681.

3. Balakrishnan S, Venkataramanan R, Ramesh A, Roy P. 2016. Contagious ecthyma outbreak among goats at Nilgiri Hills. Indian J Anim Res 51:197–200. https://doi.org/10.18805/ijar.10277.

4. Spyrou V, Valiakos G. 2015. Orf virus infection in sheep or goats. Vet Microbiol 181:178–182. https://doi.org/10.1016/j.vetmic.2015.08.010.

5. Nandi S, De UK, Chowdhury S. 2011. Current status of contagious ecthyma or orf disease in goat and sheep: a global perspective. Small Rumin Res 96:73–82. https://doi.org/10.1016/j.smallrumres.2010.11.018.

6. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

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

8. Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25. https://doi.org/10.1186/gb-2009-10-3-r25.

9. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923.

10. Freitas TAK, Li PE, Scholz MB, Chain P. 2015. Accurate read-based metagenome characterization using a hierarchical suite of unique signatures. Nucleic Acids Res 43:e69. https://doi.org/10.1093/nar/gkv180.

11. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 13:e1005595. https://doi.org/10.1371/journal.pcbi.1005595.

12. Li H. 2018. minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094–3100. https://doi.org/10.1093/bioinformatics/bty191.

13. Kolmogorov M, Raney B, Paten B, Pham S. 2014. Ragout: a reference-

assisted assembly tool for bacterial genomes. Bioinformatics 30: i302–i309. https://doi.org/10.1093/bioinformatics/btu280.

14. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

15. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421. https://doi.org/10.1186/1471-2105-10-421.

16. Zhang Z, Schwartz S, Wagner L, Miller W. 2000. A greedy algorithm for aligning DNA sequences. J Comput Biol 7:203–214. https://doi.org/10.1089/10665270050081478.