Tracing Viral Transmission and Evolution of BLV through Long Read Oxford Nanopore Sequencing of the Proviral Genome

Laura A. Pavliscak[1], Jayaveeramuthu Nirmala[2], Vikash K. Singh[2], Kelly R. B. Sporer[1], Tasia M. Taxis[3], Pawan Kumar[2], Sagar M. Goyal[2], Sunil Kumar Mor[2], Declan C. Schroeder[2], Scott J. Wells[2], Casey J. Droscha[1,3]

[1]CentralStar, [2]University of Minnesota [3]Michigan State University

Abstract

Bovine leukemia virus (BLV) causes Enzootic Bovine Leukosis (EBL), a persistent life-long disease resulting in immune dysfunction and shortened lifespan in infected cattle. The shortened lifespan and decreased milk production of infected animals severely impacts the profitability of the US dairy industry. Our group has found that 94% of dairy farms in the USA are infected with BLV with an average in-herd prevalence of 46%. This is partly due to the lack of clinical presentation during the early stages of primary infection and the elusive nature of BLV transmission. This study sought to validate a near-complete genomic sequencing approach for reliability and accuracy before determining its efficacy in characterizing the sequence identity of BLV proviral genomes for the purposes of viral contact tracing. These BLV-infected animals were comprised of seven adult dam(mother)/daughter pairs, from one commercial dairy herd, that tested positive by ELISA and qPCR. The results demonstrate sequence identity or divergence of the BLV genome from the same samples tested in two independent laboratories, suggesting both vertical and horizontal transmission in this dairy herd. This study supports the use of Oxford Nanopore Technologies (ONT) sequencing for the identification of viral SNPs that can be used for retrospective genetic contact tracing of BLV transmission.

Methods and Materials

Seven adult dam/daughter pairs were tested by ELISA and qPCR to determine PVL followed by HMW DNA extraction and long-range PCR to amplify the BLV genome for Oxford Nanopore sequencing. The corresponding DNA extracts were amplified 2x in independent laboratories using different PCR chemistries and primer sets. Phylogenetic analysis of the sequences geographically shows the similarity of the generated sequences and can be used to determine relatedness. Sequencing controls were developed via the ligation of previously Sanger sequenced bacterial inserts into a well-known TOPO vector to validate the fidelity and accuracy of ONT sequencing. The Rapid Barcoding Sequencing kit was used for the library preparation and the pooled samples were loaded onto a R9.4 flow cell on a GridION device and run with the SQK-RBK004_plus_Bascalller script of MinKnow1.5.12. Reads were base called using *Guppy* and demultiplexed with *qcat* v1.1.0. Medaka v1.2.2 was used to ID consensus-level variant candidates. The methods pipeline is shown below.

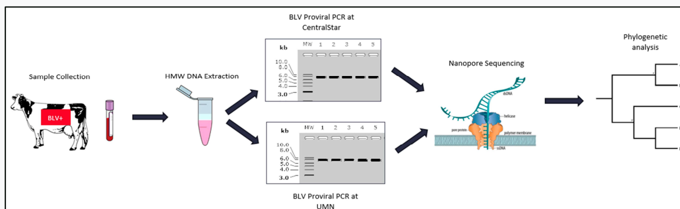
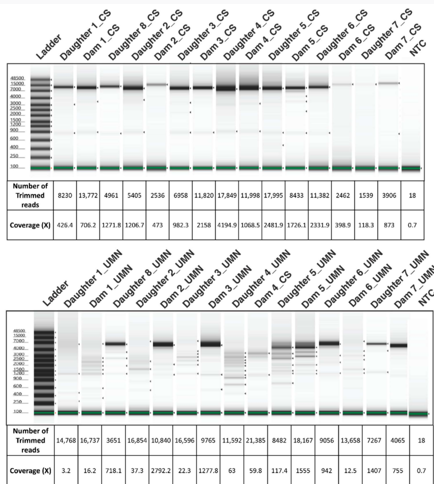


Figure 1. Project Methodology Pipeline



	Daughter 1_CS	Dam 1_CS	Daughter 8_CS	Daughter 2_CS	Dam 2_CS	Daughter 3_CS	Dam 3_CS	Daughter 4_CS	Dam 4_CS	Daughter 5_CS	Dam 5_CS	Daughter 6_CS	Dam 6_CS	Daughter 7_CS	Dam 7_CS	NTC
Number of Trimmed reads	8230	13,772	4961	5405	2536	6958	11,820	17,849	11,998	17,995	8433	11,382	2462	1539	3906	18
Coverage (X)	426.4	706.2	1271.8	1206.7	473	982.3	2158	4194.9	1068.5	2481.9	1726.1	2331.9	398.9	118.3	873	0.7

	Daughter 1_UMN	Dam 1_UMN	Daughter 8_UMN	Daughter 2_UMN	Dam 2_UMN	Daughter 3_UMN	Dam 3_UMN	Daughter 4_UMN	Dam 4_CS	Daughter 5_UMN	Dam 5_UMN	Daughter 6_UMN	Dam 6_UMN	Daughter 7_UMN	Dam 7_UMN	NTC
Number of Trimmed reads	14,768	16,737	3651	16,854	10,840	16,596	9765	11,592	21,385	8482	18,167	9056	13,658	7267	4065	18
Coverage (X)	3.2	16.2	718.1	37.3	2792.2	22.3	1277.8	63	59.8	117.4	1555	942	12.5	1407	755	0.7

Figure 2. Electrophoretograms of the BLV proviral amplicons generated by two independent laboratories.

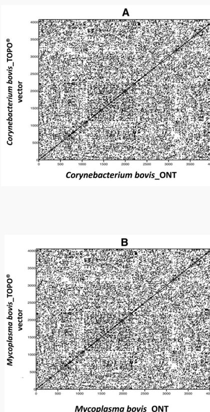


Figure 3. Dot plot matrices highlighting the 100% similarity between the references and the ONT sequences

Results

Sequencing Controls

The generated reference sequence, using the previous Sanger sequencing along with the known sequence of the vector, was found to have 100% identity to the sequences generated by ONT.

Oxford Nanopore Technologies Sequencing

Approximately 7.5kb PCR products were generated using two different sets of primer pairs from two independent laboratories. Electrophoretograms of the generated amplicons were evaluated to determine amplicon quality and infer the accuracy of subsequent analysis. The differences seen with a cross-laboratory comparison of the labs could be due to read depth variability caused by amplicon quality which can confound phylogenetic analysis.

Phylogenetic Analysis

Ten out of the 14 pairs of technical replicates from the two laboratories matched on taxa location, solidifying the accuracy of both PCR techniques and the sequencing of the generated amplicons. The outlying samples were concluded as resulting from poor coverage. When analyzing the samples based on familial relationships, the distance between proviral sequences varies. Pair 5 shows sequence identify based on clade location while pair 6 are significantly different. Pair 2 displayed viral evolution in just one generation.

Amino Acid Substitutions

A comparative amino acid analysis was performed to further compare the dam/daughter pair relationship. Amino acid profiles of 5 dams diverged from their daughters where 2 were identical. The Gag/Pol/Pro complex was the most polymorphic. The amino acid changes show a significant level of functional variation between the dam/daughter pairs investigated.

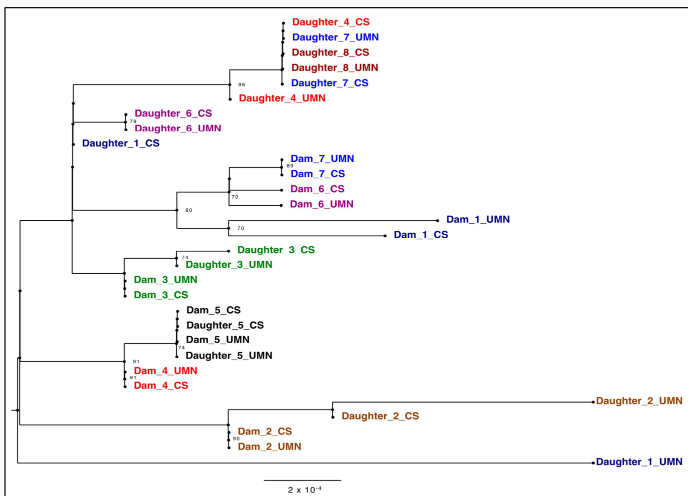


Figure 4. Phylogenetic tree of the BLV genomes derived from both laboratories displaying both vertical and horizontal transmission between dam-daughter pairs.
Lab #1: CentralStar Cooperative, Lansing, MI, USA Lab #2: University of Minnesota, Minneapolis, MN, USA

Discussion

Since a novel library preparation and bioinformatics pipeline were used, the fidelity of the generated reads were evaluated via the creation and sequencing of bacterial positive controls and the use of stepwise validation as a quality control check at each step along the pipeline. This phylogenetic analysis tool can help in tracing spatial and temporal modeling of infections within different herds and elucidate how cattle movement and farm management practices affect transmission. This approach could also be used to help resolve genetic relationships of dynamic viral populations of other pathogens, such as SARS-CoV-2. The use of the entire proviral genome would aid in the comparison of international strains instead of drawing conclusions on the analysis of a singular gene, which is historically how these comparisons have been done. Future research could investigate the impacts of the identified amino acid changes, along with the use of RNA transcripts. Due to the small sample size, a longitudinal study including neonate samples is needed to evaluate direct and temporally relevant transmission patterns.

Contact:

Laura Pavliscak
Research and Development Laboratory Manager
CentralStar Cooperative
Email: Laura.Pavliscak@mycentralstar.com
Website: www.linkedin.com/in/laura-pavliscak
Phone: 248.408.2640

References:

1. Green, M.R.; Sambrook, J. Isolation of High-Molecular-Weight DNA from Mammalian Blood Using Proteinase K and Phenol. Cold Spring Harb. Protoc. 2017.
2. Polat, M.; Takeshima, S.S.; Hosomichi, K.; Kim, J.; Miyasaka, T.; Yamada, K.; Arainga, M.; Murakami, T.; Matsumoto, Y.; Barra Diaz, V.; et al. A New Genotype of Bovine Leukemia Virus in South America Identified by NGS-Based Whole Genome Sequencing and Molecular Evolutionary Genetic Analysis. Retrovirology 2016, 13, 1–23
3. Yu, C.; Wang, X.; Zhou, Y.; Wang, Y.; Zhang, X.; Zheng, Y. Genotyping Bovine Leukemia Virus in Dairy Cattle of Heilongjiang, Northeastern China. BMC Vet. Res. 2019, 15.
4. Rola-Łuszczak, M.; Sakhawat, A.; Pluta, A.; Ryło, A.; Bomba, A.; Bibi, N.; Kuźmak, J. Molecular Characterization of the Env Gene of Bovine Leukemia Virus in Cattle from Pakistan with NGS-Based. Evid. Virus Heterog. 2021, 10, 910.
5. Li, J.; Wang, H.; Mao, L.; Yu, H.; Yu, X.; Sun, Z.; Qian, X.; Cheng, S.; Chen, S.; Chen, J.; et al. Rapid Genomic Characterization of SARS-CoV-2 Viruses from Clinical Specimens Using Nanopore Sequencing. Sci. Rep. 2020, 10, 17492.

Scan to read the published article!