Bait capture and enrichment followed by long-read whole-genome sequencing and viral variant discrimination through TELSVirus workflow using influenza A virus as a model

Mariana Meneguzzi[1*]; Montserrat Torremorell[1]; Noelle Noyes[1]
[1]Department of Veterinary Population Medicine, University of Minnesota, meneg009@umn.edu

Background

Viral co-infections are a frequent health challenge in swine worldwide, contributing to aggravated disease outcomes. Whole-genome sequence (WGS) of viral genomes can improve our understanding of viral co-circulation dynamics, ecology, evolution, and pathogenesis in swine herds. However, low viral titers and high levels of host genomic material are some of the technical challenges when applying sequencing to field samples. Furthermore, methods such as PCR amplicon sequencing usually target only single viruses, which limits the ability to understand viral co-infection and co-circulation dynamics. Finally, viral culture is often needed to enrich genomic material prior to sequencing, which can introduce bias into the resulting genomic sequences. To overcome these limitations, we have developed a workflow called "TELSVirus", or "Target-Enriched Long-read Sequencing of Virus". The workflow combines a "capture & enrichment" method with long-read, real-time sequencing technology (by Oxford Nanopore), followed by an ensemble bioinformatics pipeline for data analysis.

Materials and Methods

To initiate the TELSVirus workflow, we bioinformatically designed a panel of "probes" that selectively targets 50 RNA swine viruses. Probes were designed (length of probe = 120bp) using a custom bioinformatic algorithm to cover 100% of all known complete genomes included in the bait set (a total of 10,949 genomes). Probe binding can tolerate a probe-to-target mismatch rate of up to 30%. This means the probe set can recover novel virus variants even when the probe does not fully match the sequence of the known variant, and thus is well suited for rapidly evolving viruses such as influenza A virus (IAV). Then, a pilot study was performed to test the TELSVirus workflow protocol. Five nasal swabs from weaned pigs with known IAV status (n=3 IAV qPCR positive; n=2 IAV qPCR negative) were subjected to the TELSVirus workflow. First, IAV RNA was extracted by QIAamp® Viral RNA Mini Kit, followed by complementary DNA synthesis. Probe hybridization and subsequent enrichment were performed on all samples using the SureSelect XT HS2 DNA System (Agilent Technologies). Subsequently, library preparation (PCR barcoding kit SQK-PB004) was performed prior to loading the samples in the minION flow cell using the mk1c device. Once we obtain the sequence data (i.e., the reads), we then use alignment (minimap2 program) to identify the reads that originated from any of the 50 viruses;our reference database is the same set of genomes that we used to design the probes. For our pilot, we focused on reads aligning to influenza genomes. We then report the percentage of total sequenced reads aligned with the target of interest, defined as our "on-target reads". Also, we reported the proportion of reads that did not align with IAV, called "off-target reads". After alignment, depth coverage for each segment of the IAV was calculated. Out of the three qPCR-positive samples that were submitted to the TELSVirus workflow, one sample was used to pilot an ensemble bioinformatics pipeline for haplotype analysis; this pipeline performs haplotype calling on a segment-by-segment basis using RVHaplo for variant calling. RVHaplo has a threshold of 0.3% diversity and 0.3% abundance for haplotype reconstruction considering sequence data with 20x depth coverage and 100% breadth coverage. Error correction of the generated reads was not performed since RVHaplo accounted for it.

Results

Our preliminary data show that the designed probes can increase the percentage of on-target reads for IAV from 0.1% up to 66%. In contrast, there was no increase in the on-target read proportion for IAV qPCR-negative samples. The average depth coverage for the IAV positive samples was calculated as the ratio of total reads obtained for the target genome (IAV) by the length of each segment and a range of 6 to 48x were obtained. Hemagglutinin (HA) and neuraminidase (NA) gene segments had 20x and 15x depth coverage respectively and RVHaplo was performed. From RVHaplo analysis, we found 5 circulating haplotypes of IAV within a single sample based on the HA and NA gene segments. The abundance of the reconstruct haplotypes for the two segments (HA and NA) varied from 1.3% to 82%. Both segments had one of the haplotypes dominant at ~80% of the influenza reads (HA-Haplotype 1 at 82% and NA-Haplotype 1 at 80%).

Conclusions

In conclusion, our data showed that the TELSVirus workflow can effectively increase the proportion of sequenced reads for the target virus, and discriminated among different IAV variants that were simultaneously present in a single sample with the whole process taking 24-48 hours. Future work aims to streamline the TELSVirus workflow by applying it to a greater number of field samples from swine herds. Prospective results from the TELSVirus workflow will shed light on understanding the ecology and epidemiology of viral co-infections, assistance on more effective on-farm strategies for controlling and preventing viruses in swine populations. Consequently, this will help improve swine welfare, productivity, and food safety.

Acknowledgments: This work is supported by Pipestone MorriSTONE Research award, Education and Extension Technology Transfer Program, Minnesota Agricultural Research, USDA Hatch Capacity Grant Funding, National Pork Board, Morrison Fellowship.