BEEF INDUSTRY SAFETY SUMMIT
March 3-5, 2015
Dallas, TX

Project Title: **Use of Next-generation Sequencing and Robust Bioinformatics to Detect and Identify Foodborne Pathogen Genes in Cattle and Their Environments**

Presenter: **Xiang Yang**, N. R. Noyes, L. M. Linke, R. J. Magnuson, A. T. Dettenwanger, C. A. Boucher, I. Georgnaras, H. Yang, D. R. Woerner, K. L. Jones, P. S. Morley, and K. E. Belk
Presenters email address: yxxiang@rams.colostate.edu
Mailing address: Colorado State University
1171 Department of Animal Sciences
Fort Collins, CO, 80523

Category: Pre-harvest research

Published: Unpublished to date

Objective: To evaluate the potential for utilizing shotgun metagenomics and next-generation DNA sequencing to detect and identify pathogen genes in cattle feedlots.

Experimental Design & Analysis:

Samples of feces, water, and soil from two pens of cattle in each of four feedlots that were geographically dispersed were aseptically collected at the time of cattle placement in the feedlot and again at the time that the same cattle were shipped for harvest. Whole community DNA in each sample was extracted using Mo-Bio PowerFecal® and PowerSoil® DNA Isolation Kits, purified, and subjected to metagenomic shotgun sequencing on an Illumina HiSeq. After trimming the sequencing data, reads were filtered to exclude *Bos Taurus* DNA and then searched against the NCBI bacterial database using Kraken software. The foodborne pathogen groups of interest included: Shiga toxin-producing Escherichia coli (STEC; including *Escherichia coli* O157:H7 plus the "Big Six" non-O157 serogroups), *Salmonella enterica*, *Listeria monocytogenes*, *Campylobacter coli/jejuni*, *Staphylococcus aureus*, and *Clostridium botulinum/perfringens*. Relative abundance of genes of each pathogen group, per sample, was calculated as (number of reads matched to specific pathogen genes x 1,000,000) ÷ total number of filtered reads.

Key Results:

Average sequencing reads for arrival and exit samples were 48M (ranging from 24 to 69M) and 38M (ranging from 13 to 63M), respectively. The mean quality (Phred) score for all sequencing samples was Q35. Additionally, 90% of reads had a quality score above Q30. The average proportion of filtered reads assigned to pathogens-of-interest was 0.031% (ranging from 0 to 1.5%). Reads identified as belonging to pathogens of interest were detected in all samples collected at the feedlots; this did not necessarily suggest that living organisms or significant numbers of pathogen cells were present, but simply indicated that portions of their genome were present. Genes from *Clostridium botulinum* and *Clostridium perfringens* was the most predominant species across sample collection locations (arrival, exit) and sample matrices (feces, soil, and water). Relative abundance of STEC genes was lower ($P < 0.05$) in samples collected at time of cattle exit from the feedlots than in samples collected at time of cattle arrival at the feedlots..

Industry Application:

The results from this research indicated that metagenomics approach was able to identify pathogen genes/DNA and allow us to study the ecology of pathogens rather than individual cells. However, the limitation of current metagenomics with next generation sequencing was that the results from next generation sequencing without enrichment of samples could not distinguish between viable and non-viable pathogens but only their genes/DNA. Additionally, the method could not determine whether multiple matched reads were from DNA fragments of one pathogen cell or from DNA fragments of several cells of the same pathogen. Therefore, further research is needed to optimize sample preparation methods and bioinformatics analysis to overcome these limitations.

Table 1. Comparison of relative abundance (reads/million) of food-borne pathogens from samples collected in feedlots at time of cattle arrival and exit.

	Sample Matrix					
	Feces		Water		Soil	
Pathogens	**Arrival**	**Exit**	**Arrival**	**Exit**	**Arrival**	**Exit**
STEC	90.4[a]	1.7[b]	1.2[a]	0.7[b]	54.4[a]	1.6[b]
Salmonella enterica	118.4[a]	12.9[b]	38.9	57.8	57.1	17.8
Listeria monocytogenes	63.6	15.4	37.2	67.6	44.6	56.3
Campylobacter coli/jejuni	145.6	25.2	20.7	75.5	21.2	20.7
Staphylococcus aureus	144.1	30.7	25.6	58.1	271.6	182.3
Clostridium botulinum/perfringens	854.1	651.9	220.7	999.1	211.2	663.7

[a,b] Means bearing different superscript letters within a row and within a sample matrix (feces, water, soil) differed ($P < 0.05$).