A Genetic Data Portal to Enable Discovery of Deleterious Genetic Variants in Farmed Animals

Xiomara Arias[1], Jennifer G. Janes[2], Darren E. Hagen[3], Jessica L. Petersen[4], Brenda M. Murdoch[5], David Steffen[4], Elaine Norton[6], Fiona M. McCarthy[6], and Ted Kalbfleisch[2]

(1)University of Kentucky, Lexington, KY, (2)Department of Veterinary Science, University of Kentucky, Lexington, KY, (3)Oklahoma State University, Stillwater, OK, (4)University of Nebraska-Lincoln, Lincoln, NE, (5)Department of Animal, Veterinary and Food Sciences, University of Idaho, Moscow, ID, (6)University of Arizona, Tucson, AZ

xr.arias@uky.edu

Abstract

Data and samples are often siloed at the diagnostic labs where they are generated and collected, respectively. In this project we aim to provide an integrative service to genetically characterize samples and couple these data with their respective pathology reports to identify and publish putative deleterious alleles found in farmed animals. This project links scientists from 5 different research institutions, and 4 different veterinary diagnostic labs. The service we create will provide an interface that will allow diagnosticians to submit samples and reports for genetic analysis, and subsequent distribution of derived information to producers, and the research community.

Recessive lethal alleles exist benignly in breeding populations, until a sire and dam carrying them are mated. One quarter of the resulting pregnancies will be homozygous for the lethal allele and will result in an aborted pregnancy, or death soon after birth. In cattle, sheep, and horses, abortions are often necropsied. Although many have a known cause, such as being the result of a viral or bacterial infection, many do not. Those that do not may harbor a homozygous genotype for a lethal recessive allele. We currently have and are building sequence datasets for on the order of 100 healthy animals from each of these species. This project is collecting pathology reports for and will sequence 40 abortions, up to 15 each from cattle, sheep, and horses to look for alleles that are homozygous in these samples, but not in the larger population.

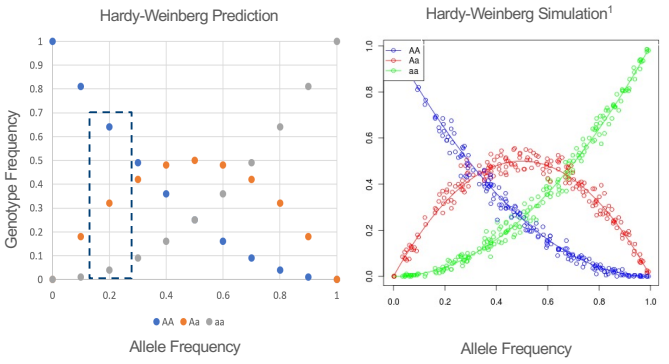

Figure 1. The Hardy-Weinberg equation suggests that in a population with a 20% minor allele frequency, 64% will be homozygous for the common allele, 32% will be heterozygotes, and 4% will be homozygous for the rare allele. If the homozygous rare genotype frequency is 0 or otherwise different, the polymorphism is either an assembly artifact or a polymorphism of interest.

Objectives

- Publish a catalog of putative homozygous lethal alleles identified from livestock fetuses
- Produce an online data portal accessible for livestock breeders
- Allow breeders to reduce fetal deaths while increasing desirable production traits

Livestock	Number to Sequence in Study	Data Set for Comparison
E. caballus -Thoroughbred	18	234
Ovis aries	11	96[2]
Bos taurus	11	96[2]

Table 1. Details of the livestock being studied, the number of aborted fetuses being sequenced, and the total Whole Genome Sequences (WGS) available for comparison. Sequenced samples will be analyzed for chromosomal abnormalities against data set from United States Meat Animal Research Center (USMARC). Homozygotes should be present for alleles in the 10%-20% minor allele frequency range in a population this large. Alleles for which there are not homozygotes in the data set but are in the sequenced set will be enumerated.

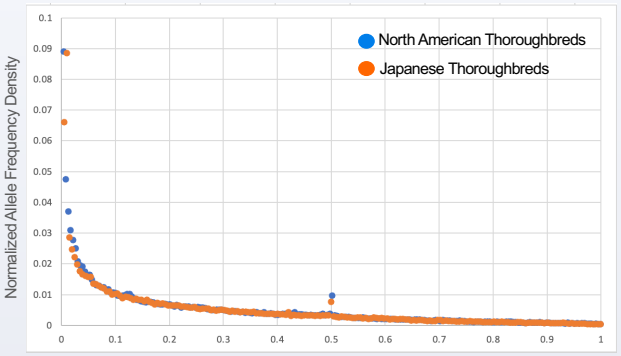

Figure 2. Allele frequency distribution graph [3] based on North American and Japanese Thoroughbred populations in Tozaki et al. [4]The spike at 0.5 is an artificial jump due to artifacts. A challenge to data analysis will be assembly artifacts. Analysis will focus on finding heterozygous alleles. Artifacts will be abundant as heterozygotes and trigger a false positive for a predicted lethal. As a measured lethal, the difference will be clear. Artifacts will never appear as homozygotes. Lethal recessive alleles will exist as homozygotes in the sequenced fetuses.

Parent	GT (Genotype)	AD (Allele Depth)	DP (Read Depth)
Both parents have duplication	0/1	16/24	40
One parent has duplication	0/1	12/18	30

Table 2. These are artifacts of collapsed duplications. Duplication events in a genome will present a challenge to data analysis. Duplication events will have the same Variant Call Format (VCF) signal as a deleterious recessive. Regardless of parent status, genotype will always indicate heterozygosity. Artifacts do not appear as homozygotes. The read depth will display as double for duplications.

Next Steps

- Whole Genome Shotgun Sequencing on 40 livestock tissue samples
- Analyze data for chromosomal abnormalities against database of livestock genotypes
- Map data

References

[1] Chivers, C. (2011). "Using Simulation to Demonstrate Theory: Hardy-Weinberg Equilibrium." *Bayesianbiologist*. https://bayesianbiologist.com/2011/06/13/using-simulation-to-demonstrate-theory-hardy-weinberg-equilibrium/. Accessed 14 December 2022.

[2] U.S. Meat Animal Research Center: Clay Center, NE (n.d.). "Cattle and Sheep Whole Genome Sequence (WGS)." *USDA Agricultural Research Service*. https://www.ars.usda.gov/plains-area/clay-center-ne/marc/wgs/main/. Accessed 14 December 2022

[3] Kalbfleisch, T., et al. (2022). "Creation of a public resource to establish the state of genetic variation in North American Thoroughbreds." [Poster].

[4] Tozaki, T., et al. "Rare and common variant discovery by whole-genome sequencing of 101 Thoroughbred racehorses." *Scientific Reports*, vol. 11, no. 1, pg. 16057, 2021, https://doi.org/10.1038/s41598-021-95669-1. Accessed 3 January 2023

Acknowledgements

2022-2023 Undergraduate Research Award (UK Office of Undergraduate Research) Arias PI
AG2PI Seed Grant: A genetic data portal to enable discovery of deleterious genetic variants in farmed animals". Source: Iowa State University, Parent grant USDA/NIFA (20217041235233)