

RAPID IDENTIFICATION OF HIGH-QUALITY BEEF PRODUCTS THROUGH ON-SITE
GENOMIC TESTING IN UNDER-RESEARCHED ASIAN CATTLE BREEDS
ORIGINATING FROM THE UNITED STATES

By

Hanna Ostrovski

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Animal Science – Doctor of Philosophy

2023

ABSTRACT

The Wagyu breed of cattle, known for its marbled and high-quality beef, commands a significant premium in global markets, underlining the importance of investigating the Wagyu population in the United States. This small population of Red Wagyu (Akaushi) and Black Wagyu were imported to the United States from Japan in the early 1990's and since then has not had any more live animals, semen, or embryos available. This strange introduction of this cattle breed to the US and the unique selection pressures on a relatively un-researched breed demands further investigation through genomic technologies.

In the U.S. market, Wagyu beef products are becoming increasingly commonplace. However, there is a notable absence of standardization to these products which carry high price tag due to the breed's reputation for superior quality and taste. Verifying products through genotype is the most straightforward approach, yet sequencing methods have been largely inaccessible, limited to specialists in molecular genetics. Oxford Nanopore's new mobile sequencing tools aim to increase sequencing capabilities for anyone. An initial trial run with inexperienced user conducted seven flow cell sequencing runs on the handheld MinION sequencer to sequence a single animal's genome. Results achieved good breadth and a low depth coverage across the genome, with each run generating more data. While ONT promises over 50 GB per run, the highest run achieved ~6 GB, signaling a significant gap between expected and actual output. Despite this difference, the technology's novelty and the user's inexperience didn't prevent successful sequencing. This emphasizes the potential of ONT's products for mobile sequencing, particularly for newcomers, extending beyond traditional lab settings.

Enabling the mobility of this sequencer for on-site product verification necessitated developing a mobile genomic sequencing kit for field use. Establishing an out-of-lab protocol was essential to swiftly identify breeds, with a specific focus on identifying Wagyu animals in this study. Sam Red (Akaushi) Wagyu and Black Wagyu animals were sequenced using the mobile kit. Breed verification of all animals was initially done with principal component analysis (PCA), but due to low output and sporadic coverage of the genome, PCA showed to be a poor way to identify breed of origin. Another method of directly matching haplotypes to a reference population was employed which was successful and boasted high correlation (0.55) and concordance rates (0.94) of sample haplotypes to the correct reference breed.

These identification methods successfully verified Wagyu samples and hold potential for broader application in field product verification. However, the genomic landscape of US Wagyu largely remains unexplored. While the traditional Wagyu breeds from Japan are well-documented, the genetic composition of American Wagyu is not yet fully grasped. Initial explorations into this breed revealed inbreeding and extensive linkage disequilibrium within the genome. A particularly intriguing finding emerged in Akaushi animals: they exhibited a close genetic relation to the Korean Hanwoo breed, as evidenced by a Principal Component Analysis (PCA). This correlation isn't entirely surprising, given the historical understanding that Japanese animals have roots tracing back to inland Korea. This connection sheds light on the genetic affinity between Red Wagyu and the Hanwoo breed, offering insights into their shared ancestry.

Further connections between the Black Wagyu, Red Wagyu, the RedBlack cross between the two Wagyu groups and Korean Hanwoo were tested through estimating the accuracy of predicting phenotype between the breeds. The accuracy between Red and Black Wagyu was low, approximately 0.10 and increased when using the RedBlack or the Korean Hanwoo, ranging from 0.23 to 0.27. To address unbalanced breed group sizes (~150 Black Wagyu versus ~5000 Red Wagyu), the total population was divided into 10 balanced groups based on animal relatedness via the first principal component. Testing prediction accuracies within these splits revealed higher accuracies, especially between closely related splits, reaching up to 0.45. Notably, the split involving Red Wagyu (1st PC split) and Korean Hanwoo (10th PC split) demonstrated this heightened accuracy, reinforcing the close genetic relationship between these breeds. Lastly, a comprehensive genome-wide association study across all breeds identified new genomic regions on chromosomes 6, 10, and 14 associated with growth in Asian cattle.

The uncovered genomic architecture of US Wagyu can aid in understanding the unusual introduction of US Wagyu and small number of animals available that have shaped the Wagyu population today. This understanding will pave the way for enhanced breeding programs, enabling producers to further refine and optimize the desirable traits of Wagyu beef, ultimately improving its quality and consistency. The exploration of US Wagyu's genomic landscape further contributes to the authenticity and traceability of Wagyu products in the market. By establishing a comprehensive genetic profile, it is easier to verify and certify the quality of Wagyu beef, thereby ensuring transparency and trust for consumers.

To my family, the one I was born into: my Mom and Dad, Megan and Claire, and to the one I was blessed with: my husband Alan, my son Isaak, and my emotional support, May. Your love, reassurance and encouragement are unwavering, and I will forever be grateful.

ACKNOWLEDGEMENTS

Firstly, I would like to acknowledge the many mentors that I have had in my life, especially those that have had a direct impact on my path in university, Dr. Phil Miller, Dr. Ron Lewis, Dr. Christian Maltecca, Dr. Francesco Tiezzi, and Dr. Cedric Gondro. I have been permanently shaped by your mentorship, and think of myself as a better researcher, communicator, and overall person because of you all. I would also like to thank my PhD committee members, Dr. Wen Huang, Dr. Juan Steibel, Dr. Robert Tempelman and Dr. Ana Vazquez, who have guided me in my journey and have given me invaluable insight and encouragement in my program. I thank you for your time and counsel in my growth at Michigan State and look forward to future collaboration and partnership.

I would like to thank and acknowledge the many fellow graduate students and post-docs that I have learned from at the University of Nebraska – Lincoln, North Carolina State University and Michigan State University. I am very lucky to call many brilliant professionals my friends.

I would like to thank the American Wagyu Association, it's staff and members, who fuel my passion for Wagyu in America. This thesis is just the first step into exploring, expanding and researching Wagyu. I would also like to thank the American Akaushi Association, who welcomed me in Texas and supported my research. I am looking forward to the endless possibilities in this breed and the positive impact new discoveries will have on cattle producers.

Lastly, I would like to acknowledge my family in friends, who many times, saw what I could not, and helped me to be the best version of myself. I would specifically like to acknowledge my Grandad, as my curiosity in life and learning is the result of listening to and watching you. I could not have accomplished anything in this journey without my husband Alan. You are my source of the most un-wavering support and love; I could not have accomplished much without you.

TABLE OF CONTENTS

CHAPTER 1: Introduction.....	1
CHAPTER 2: Literature Review.....	6
LITERATURE CITED.....	28
CHAPTER 3: Investigating New Technologies for On-Site Real-Time sequencing for any Animal Scientist.....	38
LITERATURE CITED.....	51
CHAPTER 4: Mobile, Rapid Beef Product Identification through 3 rd generation Sequencing Methods.....	53
LITERATURE CITED.....	73
CHAPTER 5: Genetic Characterization of the Akaushi Breed in the United States.....	76
LITERATURE CITED.....	92
CHAPTER 6: Estimation of Within and Across Breed Prediction Accuracies in the Wagyu Population in the United States and the Korean Hanwoo.....	94
LITERATURE CITED.....	108
CHAPTER 7: Conclusions.....	117
APPENDIX A.	119
APPENDIX B.	121
APPENDIX C.	122

CHAPTER 1: Introduction

The recent boom of high-quality beef in America is largely due to the expansion of use of Japanese breeds, Black Wagyu (Figure 1.1) and Akaushi (also known as Red or Brown Wagyu, Figure 1.2).



Figure 1.1 Black Wagyu bull from the American Wagyu Association



Figure 1.2 Red/Brown Wagyu bull from the American Wagyu Association

The first introduction of these breeds in the United States was in the 1970's, when 4 bulls were imported into Texas (American Wagyu Association). From these first males, some crossbred Wagyu started to emerge throughout the country until a larger group of Wagyu cattle were imported in the early 1990's which provided females for fullblood animals. This herd was made up of the two "sub-types" of Wagyu, the Japanese Black and the Japanese Red/Brown or Akaushi (Gotoh et al., 2018). After these animals were imported, Japan locked its borders to further exports of live animals, semen or embryos of Wagyu cattle. All Wagyu cattle that exist today in the United States have come from this initial group of animals and many can be traced back to Japan through pedigree and genotype.

The initial scarcity of available Wagyu animals outside of Japan for breeding purposes raises concerns about the breed's long-term sustainability. A limited genetic pool can have adverse effects on the current Wagyu population, potentially resulting in reduced variation and increased inbreeding which poses a threat of population collapse. Over time, selective breeding and possible genetic drift have influenced this population, making it crucial to delve into its genomic architecture. Delving into the US Wagyu population can unravel how forces of selection and genetic drift have shaped the breed's characteristics and diversity. By making

informed decisions that promote genetic diversity, breeders can safeguard the health and resilience of the Wagyu breeding stock available beyond Japan.

Further research in this breed, especially those animals in the United States is needed as no published research have used the whole American Wagyu population. This contrasts with other beef breeds, such as Angus or Hereford, which have thousands of research articles exploring the many facets of the breed. A quick search on google scholar shows ~914,000 results when searching “angus”, while a search for “wagyu” results in ~9,290 hits. This is a huge discrepancy in research articles and brings attention to a need of exploration into Wagyu and the genomic pressures the US population has been under.

Uncovering Wagyu population architecture is crucial as demand for Wagyu has exploded, and consumers have started searching out a higher-quality beef product. An increase of animals and products without proper monitoring of the rapid expansion of the population. The Wagyu breed is exploding onto the scene for F1 commercial producers, as a cross with a Fullblood Wagyu to produce a ~50% Wagyu animal can increase the grade of the animal. Quality grading in the US tops out at Prime, which is considered 12% IMF (Lonergan et al., 2019). International grading scales have much more specified scaling, with the top grades in Japan scoring up to 60% IMF (Gotoh et al., 2018). Japanese grading standards compared to IMF% can be seen in Fig 1.3 (Horii, 2009)

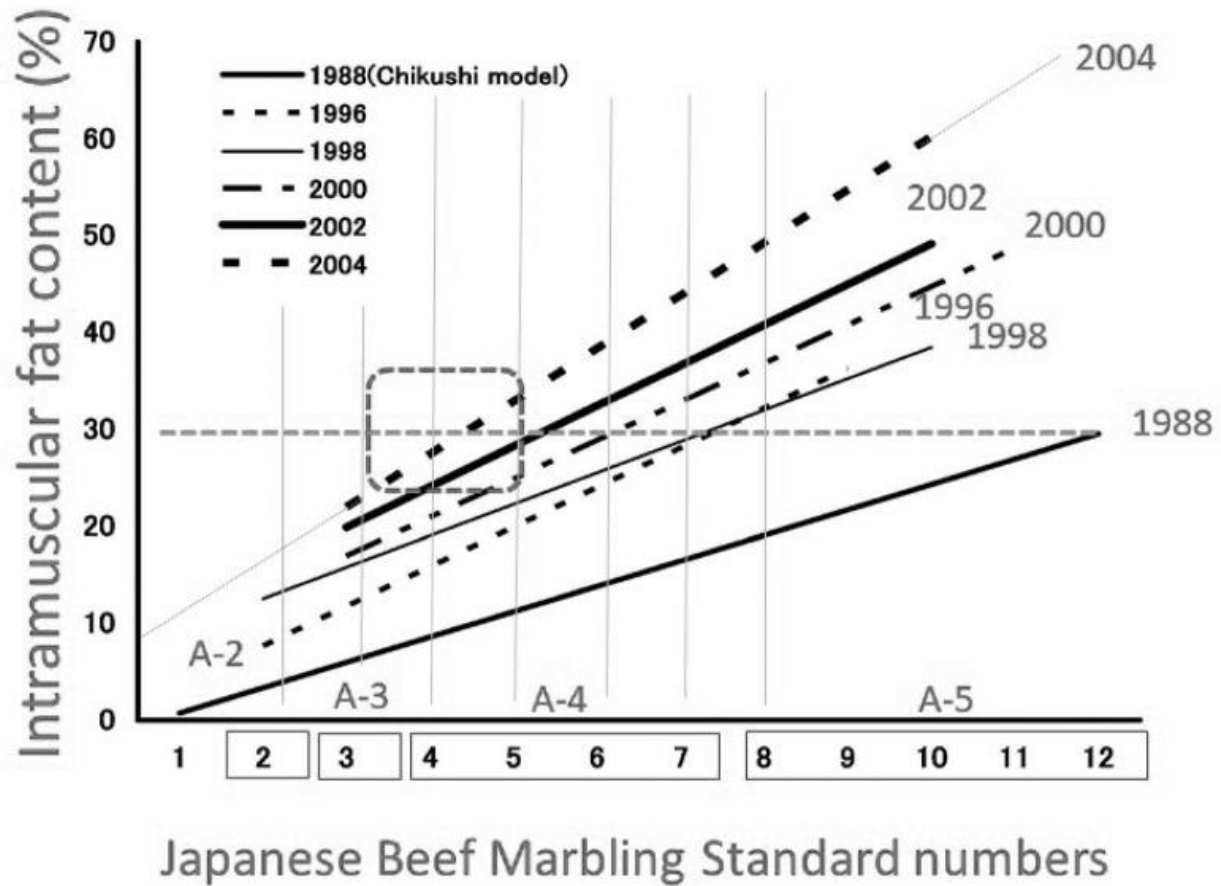


Figure 1.3 Intramuscular fat percent per Japanese marbling standards over time (Horii, 2009)

This huge discrepancy in marble grading has opened an avenue for the crossbred Wagyu animals to be sought after in the US, as many of these cattle grade Prime here, but may not grade as high in other countries. This has opened the door into the world of 50% F1 Wagyu crosses becoming a huge leader in the US beef market, as many of these F1 animals qualify for Prime on the US scale. These products have flooded the market because it made “Wagyu influenced” beef available for any budget.

The name “Wagyu” has now become commonplace and is synonymous to a beef product that has high-marbling and buttery texture. This has driven up the value of this breed in the US, which is rapidly growing, as the demand for higher quality beef has expanded (Forristall et al., 2002; Gonzalez & Phelps, 2018; Kempster, 1989). A higher quality beef product does come at a higher price, so correct breed identification of products is necessary to give consumers piece of mind when purchasing more expensive products that claim to be of higher quality. The identification of these animals in the US has now reached a tipping point, as many products now

boast a “Wagyu”-type name in advertising a product. Testing of verification can most rapidly be done through genotype, and new sequencing techniques recently boast out-of-lab protocols. Identification of Wagyu products via these protocols can be an avenue for rapid verification of these products.

The beef breeding industry outlook for Wagyu products is very positive, with many producers of Wagyu animals for breeding purposes boasting high prices. The obvious demand for this breed is apparent, and the costs associated with purchasing these animals are very high. An average private sale may sell Wagyu cows at an average of \$10k US dollars (see Appendix A), which is a large discrepancy behind the US average for other breeds at \$2k (see Appendix B). Estimating accurate breeding values for these expensive animals is crucial to keep and increase the value of these animals. More accurate breeding values can aid in selection decisions in Wagyu, especially targeting the highly coveted high-marbling animals.

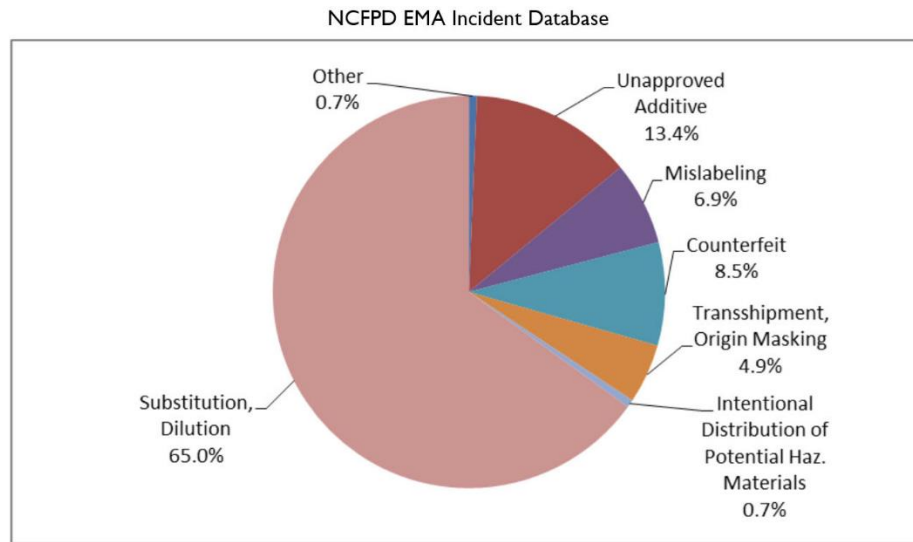
Wagyu cattle have swiftly garnered acclaim, coveted for their high-quality beef, noted for their unparalleled tenderness and rich marbling. However, with their rise in popularity, ensuring the authenticity of Wagyu products at the consumer level has become pivotal. Rapid and accurate identification of these products is crucial to preserve the integrity and reputation of authentic Wagyu. Deciphering the US Wagyu population structure will shed light on its origins, divergences, and unique attributes. These insights not only bolster the breed's authenticity and reliability but also pave the way for improved breeding practices and the preservation of Wagyu's distinguished qualities. Ultimately, this comprehensive understanding serves as a cornerstone in safeguarding the legacy and future of Wagyu in the United States.

CHAPTER 2: Literature Review

Feasibility of Wagyu Product Identification

The first hurdle to tackle in identification of animal products is how to obtain genomic information. There have been many attempts at tracing animal products without a genomic trail (Aung & Chang, 2014; Bosona & Gebresenbet, 2013; Schroeder & Tonsor, 2012; Souza-Monteiro & Caswell, 2004), which can only lead to uncertainty due to possible human error in processing or packaging of these products. The US has set up *Process Verified Programs* (PVP) since the 1990's that have tried to set a standard for safe food products and traceability.

Tracing cattle through the production system in the US can be difficult, as an animal can flow through many different hands before they are processed. A reliable traceability system from cow/calf producer to meat processor has still to be developed, but the implication of this system could help reduce the cost of disease outbreak problems (Blakebrough-Hall et al., 2020) and identify fraudulent labeling (Jo et al., 2021; Spink & Moyer, 2011). Both issues can weigh heavily on the beef market in the US if not identified. For example, bovine respiratory disease from 2011-2015 cost the cow/calf industry \$165 million dollars annually (M. Wang et al., 2018). Quick identification of outbreaks before they reach a full-blown endemic level can save the industry millions of dollars. Food fraud also costs the US industry millions of dollars per year (FDA), with the top fraudulent activities being dilution, unapproved additives, mislabeling and counterfeiting (Johnson, 2014).



Source: Compiled by NCFPD EMA Incident Database (January 6, 2014), based on 306 reported incidents.

Figure 2.1 Food fraud as reported by the National Center for Food Protection and Defense (Johnson, 2014)

The impact that product identification would have on the Wagyu market in the US would be large, as labeling of “Wagyu” product in the US is largely done on “Wagyu Influenced” cattle. Live animal specifications done by the USDA state that the animal must have at least one registered parent of 15/16 Wagyu breed (see Appendix C). This does not hinder other marketing schemes to use these high dollar definitions to label beef products. This is largely due to the name “Wagyu” having no protections in the United States but can be regulated closely in Japan due to protections of the word “Wagyu” available in Japan.

The identification of these products beyond the label can be most accurate if the products are identified through genomic sequence, as the sequence at the start of an animal’s life does not change. The easiest identification methods that would impact US Wagyu would be the identification of how much “Wagyu” is in products that claim to be Wagyu. This would be a simple breed identification that can be done through sequencing meat samples. The challenge lies in the cost of sequencing these samples and the time it takes to obtain the sequence. If the sample is sent to a 3rd party lab, the time it takes to get a sequence back could be weeks. By then, the sample of meat would have already been consumed, and the identification not known. The sequencing done for a sample must be quick, accurate and the protocol must be user friendly to obtain correct identification during the time of consumption.

Genomic Sequencing and Breed Identification

Fred Sanger was one of the first to crack the code of DNA sequencing (Sanger et al., 1965), with his method becoming one of the most common ways to sequence DNA in the early years of first-generation sequencing. Allan Maxam and Walter Gilbert also pioneered their own method, Maxam-Gilbert method (Heather & Chain, 2016), which was commonly used, but was not as easily done as Sanger sequencing. These methods were chemical and mechanical in nature, with low sequencing pace. The second generation of sequencing methods utilized parallelization of DNA sequencing, with many strands of DNA being sequenced in one run. 454 Life Sciences (Gupta & Gupta, 2014) produced some of the first of these parallel sequencing machines, which lead to quicker turnaround times from DNA to fully sequenced output. Quickly after, Solexa, later acquired by Illumina, produced their own high throughput sequencing machines that improved accuracy and read depth in sequencing (*History of Illumina Sequencing & Solexa Technology*, n.d.). To date, Illumina sequencing is the front runner in the DNA sequencing world, with most industry and research efforts relying on these technologies for their efforts.

The traditional methods of sequencing are great for projects that have extended deadlines, but rapid identification needs a real-time, out-of-lab sequencing protocol. The most interesting technology in this space has surfaced from Oxford Nanopore Technologies (ONT). A small, easy-to-use machine was introduced in 2016 called the MinION (Lu et al., 2016). It is touted as a mobile sequencing device that has the capability to take sequencing out of the lab. Figure 2.2 shows the scale of this sequencer, which can fit into the palm of a human hand.



Figure 2.2 Oxford Nanopore Technology's MinION in a human hand.

Oxford Nanopore has been around the sequencing world for many years. They have many products including the GridION, PromethION (Deamer et al., 2016; *PromethION* | *Oxford Nanopore Technologies*, n.d.) which are large sequencers that are used in an in-lab setting and can obtain large amounts of sequence.

The sequencing method that ONT is known for is long read sequencing or taking the raw DNA without amplification and reading it through their namesake, the “nanopore”(Clamer et al., 2014; M. Jain et al., 2016). The MinION outputs sequence through the flow cell that contains these nanopores. This flow cell is considered a consumable, as it can only be used 2-3 times when washed and stored correctly (Michael et al., 2018). This cell is inserted into the MinION, which electrically charges the flow cell that houses the nanopores. When a nucleotide is read through the nanopore, the electrical charge that is pumped through the flow cell is disrupted. This disruption is recorded and is noted as a “squiggle”. See Figure 2.3 for a detailed flow of these steps from (Bhattaru et al., 2019) where A is the library preparation that includes attaching

an adapter, B is the nanopore reading the string of DNA, C is the MinION itself and D shows the output squiggles.

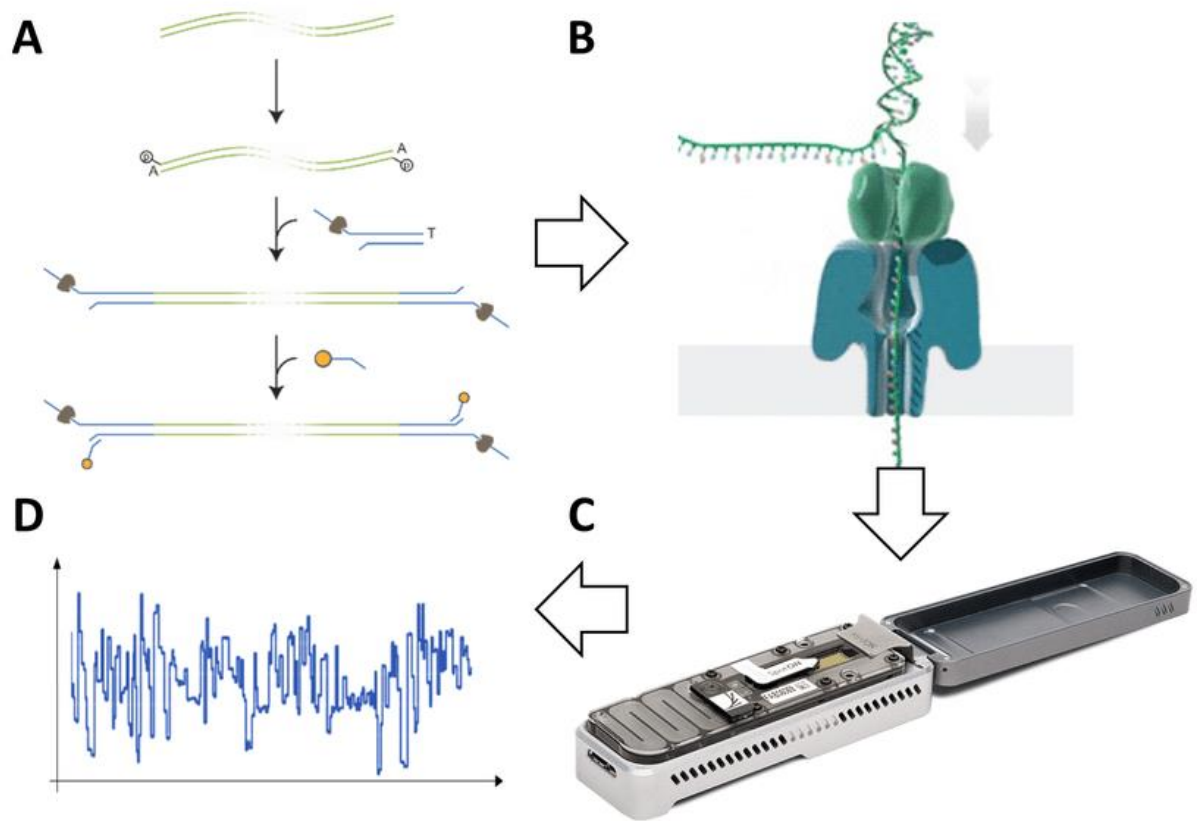


Figure 2.3 Flow of sequencing with the MinION (Bhatarru, 2019)

These squiggles are then basecalled through the ONT program “guppy” (Wick et al., 2019), which uses a learning algorithm to interpret these electrical outputs. The output is then in a form that can be aligned, indexed, sorted and variants identified through free 3rd party bioinformatic tools. The finalized sequence is then usable for population analysis, breed identification, and more in-depth genomic analyses.

These finalized sequences can be filtered at different depth, quality, and coverage at each position sequenced. This filtering can lead to differing outputs of accuracy and number of variants included in the filter. The cut off depth for high accuracy is usually at 40x for whole genome sequencing. Previous studies utilizing the MinION output have outlined a cut off for quality 7 (Delahaye & Nicolas, 2021) which lowers error rate in the reads. In most cases, the more stringent the filtering is, the lower amount of sequence will be available for analysis, but it can be assumed that this sequence is the most accurate.

When trying to identify breed composition with these sequencing methods, an efficient bioinformatic pipeline is needed as well as an already established reference population of target breeds as well as other common cattle breeds. This is needed to understand the genomic sample that was sequenced, as obtaining the sequence in an out-of-lab setting in real time is only the first hurdle. Without a genomic reference, there is no way to compare the sequence obtained, and thus no way to understand the composition of the sample.

Traditional imputation with short-read sequence data has been well researched (Yun et al., 2009) and is used to infer areas of the genome that were not genotyped. Imputation software is quite accurate when provided with short-read sequences, to the point where it is expected that lowpass sequencing coupled with imputation will soon replace traditional array genotyping as the standard genotyping method. Imputation of long-read sequencing data is, however, more complicated as they consist of longer uninterrupted stretches of the genome. This in effect means that for the same depth of coverage in comparison to the short-read data, there will be less broadness of coverage (i.e. reads will be more concentrated in some areas and consequently the distance between sequenced regions will be longer). This hinders imputation as the methods rely on the linkage disequilibrium (LD) structure of the population – the weaker the LD the lower the imputation accuracy will be. This is detrimental towards understanding breed composition as accurate genotypes are necessary to create genomic relationship matrices, which are the backbone into looking at the genetic relationships between animals. A recent imputation software QUILT (Davies et al., 2021) has been developed for long-read sequencing platforms such as ONT and could help mitigate some of these problems, although a higher depth of coverage in comparison to lowpass sequencing with short reads will probably still be needed.

Future prospects of ONT products are positive, as the tooting issues of the new technology are being addressed. Specifically, the introduction of the VolTRAX (Oxford Nanopore Technologies) will be crucial in eliminating human error in creation of the sequencing library. This product will remove all human errors in mixing and pipetting reagents to create libraries by automating the process. Other sequencing products aim for even smaller applications, such as a sequencer that can plug into a phone (Oxford Nanopore Technologies, n.d.-a). The practical application for out-of-lab sequencing that anyone can use will be through development of products that have easy protocols and less need for traditional lab requirements.

Application of new advanced technologies can be difficult to integrate, as scientists and researchers tend to be skeptical about the application of new tech, the technologies are usually high in price (Schaller, 1997), and changes in the status quo may be harder to adopt across the discipline. This is not something to be overlooked by cost or difficulty, as real-time sequencing can have huge effects if we get to the point where product certification and identification of breed composition in animals can be achieved in real-time during active consumption of the product in question.

Wagyu Origins and Population Structure

Modern cattle breeds stemmed from the middle east, *Bos Taurus* and *Bos Indicus*, with traditional European breeds arising from the *Bos Taurus* lineage. These animals were already used for meat consumption during thousands of years in Europe and even specifically reared for beef production during the last few centuries. Asian breeds only more recently started to be raised for beef production. The Wagyu breed was originally bred for work, as they were farming animals that helped with heavy plowing in Japan (Motoyama et al., 2016) at a time where it was illegal to consume meat in Japan. For many years, the Wagyu breed was not known as a food source until meat became commonplace in Japan, and the realization of the Wagyu breed as high-quality was established.

The original country of origin where the Wagyu breed started is somewhat debated, but many studies (Chen et al., 2018; Sasazaki et al., 2006) have shown that the Wagyu population in Japan originated from modern-day Korea. These animals were brought to Japan and crossed with traditional European breeds (Namikawa, 1980) to create the population today which consists of many sub-types of Wagyu: the Japanese Black, the Japanese Red/Brown or Akaushi, the Japanese Shorthorn and the Japanese Polled. The Wagyu available for production of meat products are the Japanese Black and the Japanese Red/Brown (MAFF, 2020). See the crossbred animals in Table 2.1 from Namikawa's paper.

Name of Modern breed	Prefecture	Crossed Foreign Breeds
Japanese Black	Kyoto	Brown Swiss
	Hyogo	Shorthorn, Devon, brown Swiss
	Okayama	Shorthorn, Devon
	Hiroshima	Simmental, Brown Swiss, Shorthorn, Ayrshire
	Tottori	Brown Swiss, Shorthorn
	Shimane	Devon, Brown Swiss, Simmental, Ayrshire
	Yamaguchi	Devon, Ayrshire, Brown Swiss
	Ehime	Shorthorn
	Ohita	Brown Swiss, Simmental
	Kagoshima	Brown Swiss, Devon, Holstein
Japanese Brown	Kochi	Simmental, Korean Cattle
	Kumamoto	Simmental, Korean Cattle, Devon
Japanese Poll	Yamaguchi	Aberdeen-Angus
Japanese Shorthorn	Aomori	Shorthorn
	Iwate	Shorthorn
	Akita	Shorthorn, Devon Ayrshire

Table 2.1. Composite breeds that make up the modern Wagyu breed per Japanese prefecture (Namikawa, 1980)

It wasn't until the Japanese government allowed the consumption of meat that the Wagyu breed burst onto the food scene. With many years of no artificial selection for meat production traits, these animals had developed into something quite different from their European counterparts, with increased marbling and fineness of fat strands throughout the meat. This type of fat is also very different from the traditional meat breeds and boasts a profile that is characterized as “healthy fats” or mono-unsaturated fats (Kohama et al., 2021). The meat also comes with a boost of Omega 3 and Omega 6 fatty acids, which are recognized as healthy fats as well (Shahidi & Ambigaipalan, 2018) .

The relationships between these breeds has been previously researched (Honda et al., 2004, 2006; Nomura et al., 2001), with many studies reporting a large variation between the European breeds and Asian breeds, even with the previous crossbreeding of native Japanese strains. This is to be expected, as differing selection pressures from human and environmental pressures have driven different phenotypic outcomes for these breeds. Most recent studies show the Wagyu, Akaushi and Hanwoo (Lee et al., 2014) to all be closely related in relation to other

European breeds, such as Angus or Milking breeds, such as Jersey or Holstein, which can be seen in Fig. 2.4 from Lee's 2014 paper.

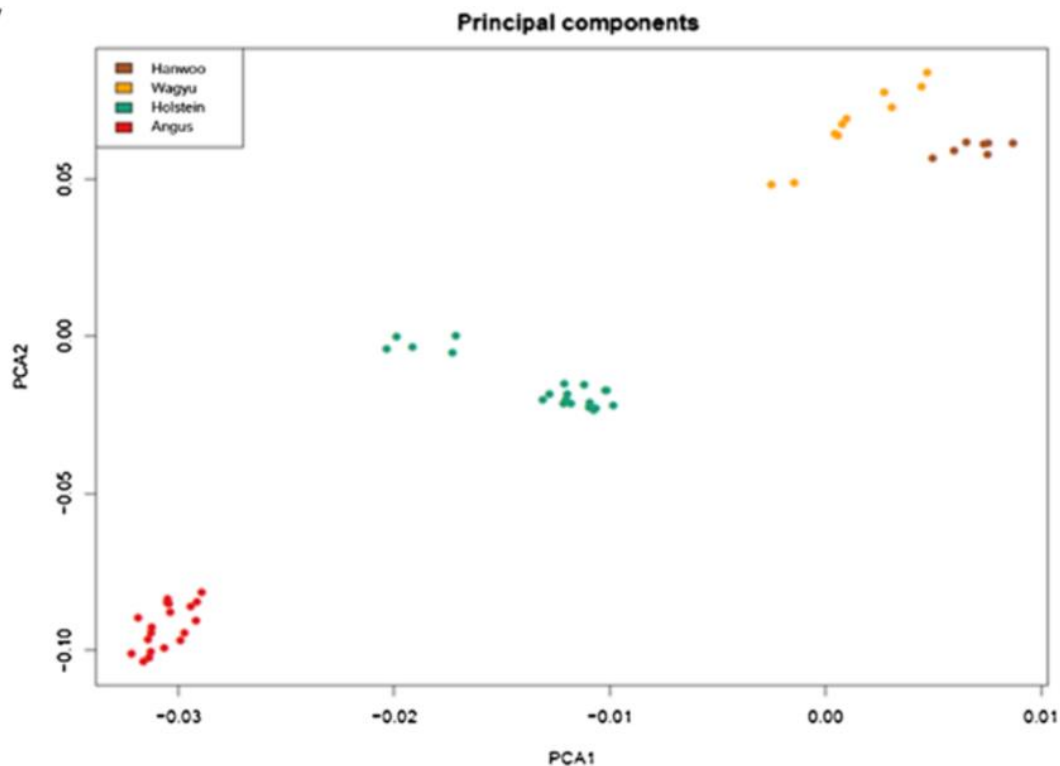


Figure 2.4 Principal Component Analysis of Wagyu, Hanwoo, Angus and Holstein; where the Asian breeds group most closely together (Lee, 2014)

The Wagyu breed outside of Japan is known to have originated from only a few animals that were exported to the United States in 1970 and in the 1990s. After the initial export, the Japanese government banned further export of Wagyu live animals, semen or embryos declaring the breed a “national treasure”. All Wagyu animals that exist outside of Japan are from the small number of animals that were allowed out of country. This is of some concern, as the genetic pool of animals to grow a large herd outside of Japan is limited. This has increased the level of inbreeding in the American population (Heffernan et al., 2021), which can lead to recessive diseases without careful consideration of breeding decisions. A strong selection in the Wagyu population has contributed to this genomic architecture in the US but can also be seen in other Asian cattle breeds (Z. Wang et al., 2019). Previous population studies of Wagyu in Japan have also shown a large decrease (Mukai et al., 1989; Uemoto et al., 2021) in variability and effective population size.

Population analyses help define the state of certain populations and are important within the animal industry, as artificial selection (Flori et al., 2009; Seo et al., 2022) and genetic drift (Brüniche-Olsen et al., 2012; Kidd & Cavalli-Sforza, 1974) can affect populations in dramatic ways. Such changes occurring in cattle due to selection are in the growth and performance, specifically the quality of the carcass, fertility traits for consistent calving, and efficient growth per animal. These changes affect the phenotype of the animal but can also change the genomic architecture of populations, as animals who are favorable in production traits are selected more often as parents than those who are not.

Population studies in Wagyu have previously tried to understand the changes in the population in the US. In particular, uncovering the structure of the relationships within the breed as well as the relationship to other breeds which is crucial for maintaining the breed with enough genomic variation and keeping inbreeding levels low. Previous studies used the numerator relationship matrix through pedigree, or the A-matrix, which utilizes numeric relationships between animals (i.e., the relationship between a parent and offspring is 0.50). These relationship matrices are built on the founders in a population, then estimating all relationships from those original animals onward. This relationship matrix only identifies those relationships that are Identity by Decent (IBD), or relationships that are assumed through parent-offspring relationships. Further identification of relationships through Identity By State (IBS), or areas of the genome that may be the same between animals that are not related due to genomic architecture of breeds. These relationships can be identified from genotypic data through the genomic relationship matrix, or G-matrix (VanRaden, 2008a):

$$G = \frac{Z'Z}{2 \sum p_i(1 - p_i)}$$

Inbreeding in a population can give the researcher a good sense of the breeding trends that have been occurring over the years. An increase in inbreeding could mean that this population was heavily selected in a line-breeding scheme, or this population has been isolated and was only able to breed among themselves. Previous studies on the American Wagyu population have been published (Heffernan et al., 2021; Scraggs et al., 2014), which state a very low effective population size (14) and large runs of homozygosity, which all point to a small gene pool and large levels of inbreeding.

Intensive selection can lead to a bottleneck in a population, as only a few animals will be selected for continuing generations. This will lead to a loss in genomic variability which can lead to inbreeding depression wreaking havoc on a population (Charlesworth & Charlesworth, 1987). Inbreeding depression can have an impact via low fertility, low fitness, or deficient performance phenotype (Brüniche-Olsen et al., 2012; González-Recio et al., 2007). The importance of understanding average inbreeding in a population is necessary to keep these negative effects at bay and to increase genetic variability in the population. Identifying highly inbred animals is also of importance, as utilizing them as breeding animals must be balanced with the need to decrease the inbreeding coefficient in future generations.

Another good indicator of inbreeding and genetic variability is the estimate of the effective population size. Effective population size can measure the amount of genetic drift that has occurred in the population from selection. In general, a small N_e is concurrent with a population that has been selected intensively, as a restricted number of animals are available for each generation as breeding animals. Effective population size can give insight into the genetic makeup of a population through estimating the number of animals it would take to make up the current population that is being analyzed.

Opposing homozygotes is also utilized in population analysis as it explores inconsistencies within the pedigree (Hayes, 2011). It can help identify animals that may be incorrectly recorded as related in the pedigree or identify unrecorded relationships. Opposing homozygotes in these animals (i.e., Parent 1 is AA and offspring is BB) show inconsistencies within the pedigree of animals that are related and can be solved through this test. This is important in populations that are pedigree dependent, such as breed associations, which register animals based on parentage verification.

All previous population analyses can identify structures within populations, but comparing different populations within one analysis is most done with principal component analysis (PCA). This is one of the most common ways to understand breed composition using genomic sequence which establishes breed grouping through eigenvectors and eigenvalues. The eigen decomposition of the genomic relationship matrix uncovers the variation that is in a population that is attributed to breed (McVean, 2009; Patterson et al., 2006). The eigenvectors help understand the grouping of animals per breed, the principal components, while eigenvalues explain the variance between the principal components, which explains the relationship between

the breeds (Karamizadeh et al., 2013). The top principal components that account for most variation between the breeds in the genomic matrix are then plotted against each other to visualize groupings of these breeds studied, which is usually the first and second principal component. For breed identification, inclusion of multiple populations and breeds is crucial for true identification of each sample. Without a good base representation of many breeds, a sample may not be grouped within any breed or may be incorrectly grouped within a principal component, and identification is then not possible.

PCA can also be used within populations to identify family lines. Family lines within Wagyu are known by prefecture that the exported animal originated from in Japan. Previous studies have outlined these prefecture lines and the phenotypic differences that are associated with these lines (Oikawa, 2018; Smith et al., 2001). The main differences lie in the expression of marbling and fatty acid profile. The most prevalent and well-known prefecture being that of Kobe, which is cherished for its extremely high marbling properties and authentication process that comes with its own certificate (Kobe Beef Marketing & Distribution Promotion Association, 2023).

Status of Wagyu worldwide shows an increase of countries that are utilizing Wagyu as a prominent source of beef (Fortune Business Insights, 2023). The population is still most highly concentrated in Japan, with the US and Australia (Gokey, 2018; Rouse et al., 2000) producing many Wagyu outside of Japan. Many of these animals are Fullblood animals, or animals that are considered 100% Wagyu, which trace their lineage back to the Japanese founders through pedigree and genomic analysis.

The recent large growth of Fullblood animals present in the United States is due to an explosion in embryo transfer use (Elsden et al., 1976; Tanabe et al., 1985). These reproductive technologies paved the way for a rapid expansion of this population without the need for many Fullblood cows. The growing population has been under selection pressures that are different from their Japanese counterparts, which may have contributed to genetic drift in this population. This phenomenon may lead to differing Wagyu populations worldwide not overlapping in a principal component analysis. Addressing all population measures in the Wagyu population in the US before further analysis can help identify the genetic structure of the current population. This can help identify parent-offspring relationships, help understand breed or family

composition through PCA or even understand if this population has a high inbreeding coefficient and define selection decisions to reduce inbreeding.

Significant QTL in Wagyu identified through GWAS

The introduction of genotyping animals through SNP chip technology opened the door for producers to utilize genomic information in a new way. Phenotypes could now be connected to QTL (Quantitative Trait Loci), more complex traits could be explored and explained, and fitting random genetic effects with a genomic relationship matrix was now possible in modeling. The utilization of genomics in the beef industry has allowed for selection of young bulls before the need of progeny testing. This has increased selection intensity, which can increase the rate of genetic gain, but can also have adverse effects (S. K. Jain & Allard, 1966; T. Meuwissen et al., 2016). This is seen in the Holstein population in the US in the early 2000s, as the introduction of genomics helped tremendously in milk output, but decreased genomic variability, which lead to a decrease in fertility rates (Lucy, 2007). The Wagyu population may be in danger of the same issues, as only a small population is available for breeding outside of Japan, and the effective population size is very small (Scraggs et al., 2014).

Identification of related animals in Wagyu was previously done with pedigree records only, but the introduction of genomic information in the form of SNPs transformed how animals are identified in a breeding population. The creation of the genetic relationship matrix (Gianola & de los Campos, 2008; González-Recio et al., 2008; VanRaden, 2008a) was a large step in utilizing genomic information in animal models. The G-matrix is best described as the relationship between animals based on the allelic frequency in the population being analyzed. These genetic markers explain the random genetic variation that occurs in each population. This can also be classified as the additive effect, or the purely “SNP” based effect throughout the genome that is contributing to phenotypic variation.

The introduction of the G-matrix did not include those older animals that may only have pedigree information and may not be available to collect a DNA sample (death, culling, harvested, etc.). A single-step approach (H-matrix) to combine the pedigree and genomic information in these populations was introduced by (Legarra et al., 2009) and is the current standard for many animal industries (J. C. Dekkers, 2004; Hutchison et al., 2014; Knol et al., 2016; Wolc et al., 2016) .

Genome-wide association studies (GWAS) are one of the most common ways to identify significant SNP in a population of genotyped animals. Identifying these SNP can be done through a BLUP model, or GBLUP if genomic information is being utilized. Obtaining SNP effects through this model is done through backsolving (Strandén & Garrick, 2009; VanRaden, 2008b; H. Wang et al., 2012). A more specific SNP-based model estimates the genomic EBVs through a snpGBLUP, which estimates the value of each SNP with fixed effects considered.

$$y = Xb + Wa + e$$

Where y is the phenotype, X is the incident matrix for fixed effects, b is the vector of fixed effects, W is the genotype matrix, a is the vector of regression coefficients for random SNP effects $N \sim (0, G\sigma_a^2)$ and e is the vector of residual effects, where we assume $N \sim (0, I\sigma_e^2)$.

To obtain p-values of each SNP effect we can utilize output from the snpBLUP into this equation:

$$pval_i = 2 \left(1 - \Phi \left(\left| \frac{\hat{a}_i}{sd(\hat{a}_i)} \right| \right) \right)$$

Which identifies the significance of the SNP effect \hat{a}_i as $pval_i$ through the density distribution (t-distribution). Visual identification of significant SNP throughout a genome can be done through traditional visualization of these p-values is by way of Manhattan plots, famously named after the skyline of Manhattan. The highest peaks are those with most significance towards the desired phenotype, whether that be a positive or negative response to the phenotype. The highest p-values of significance that cross the p-value threshold of $-\log_{10}(5 \times 10^{-8})$ are the SNP with the most influence on a phenotype. This log transformation is used due to the nature of very small p-values that are obtained to be easier to visualize.

Utilizing whole-genome sequence (WGS) in GWAS can help tease out these significant peaks and identify more rare variants (Onteiro et al., 2012; Wu et al., 2017) but may also introduce more noise into these association studies. Most traits that are being explored via GWAS are complex, in that they are controlled by many areas on the genome. There are important production traits such as polled/horned or coat color that are controlled by one locus and follow a simplistic dominance of appearance or also known as mendelian inheritance i.e., polled animals or black hided animals have at least one dominant copy of the gene to express the phenotype.

Some of the previous work done on Wagyu animals has been on the Japanese and Chinese populations (An et al., 2019; Mizoguchi et al., 2006; Mizoshita et al., 2004; Takasuga et al., 2007; Zhang et al., 2019) which have identified significant QTL in both Chinese and Japanese Wagyu. Specific QTL identified in An et al. related to growth traits in Wagyu can be found in Table 2.2.

Trait	SNP Name	BTA	MAF	Gene Name	P-Value
Body Height	Hapmap46986-BTA-34282	14	0.42	PENK	4.19E-06
Body Height	BovineHD1400007323	14	0.42	PENK	5.69E-06
Body Height	BovineH04100011295	14	0.42	PENK	6.97E-06
Body Height	BTB-00557532	14	0.4	XKR4	9.48E-06
Body Height	BovineH01400007377	14	0.48	IMPAD1	9.58E-06
Body Length	BovineHD1400015419	14	0.2	CSMD3	6.69E-06
Hip Height	Hapmap46986-BTA-34282	14	0.42	PENK	1.34E-07
Hip Height	BovineHD1400007259	14	0.45	PLAG1	4.41E-07
Hip Height	BovineH04100011295	14	0.42	PENK	5.36E-07
Hip Height	BovineHD1400007323	14	0.42	PENK	6.26E-07
Hip Height	BTB-00557532	14	0.4	XKR4	7.41E-07
Hip Height	ARS-BFGL-NGS-98420	5	0.1	CCND2	1.40E-06
Hip Height	BovineH01400006445	14	0.27	SNTG1	2.86E-06
Hip Height	BovineH01400007333	14	0.41	PENK	3.05E-06
Hip Height	BTB-01530836	14	0.4	XKR4	3.12E-06
Hip Height	BovineHD1400007377	14	0.48	IMPAD1	3.63E-06
Hip Height	Hapmap32552-BTA-129045	14	0.26	SNTG1	3.83E-06
Hip Height	BovineH01400007314	14	0.34	PENK	4.07E-06
Hip Height	Hapmap26308-BTC-057761	6	0.1	LAP3	5.10E-06
Hip Height	BovineHD1400007375	14	0.47	IMPAD1	5.52E-06
Hip Height	BovineHD0500034451	5	0.34	FAM19A5	6.14E-06
Hip Height	BovineHD0500020210	5	0.19	SYN3	7.02E-06
Hip Height	BovineHD0500020213	5	0.19	TIMP3	7.02E-06
Hip Height	BovineHD1400007373	14	0.48	IMPAD1	9.18E-06

Table 2.2 Significant QTL found in Chinese Wagyu for growth

Table 2.2 (cont'd)

Multi-Trait	Hapmap46986-BTA-34282	14	0.42	PENK	1.63E-06
Multi-Trait	BovineHD0500026837	5	0.49	STRAP	2.96E-06
Multi-Trait	BTB-00557532	14	0.4	XKR4	4.73E-06
Multi-Trait	ARS-BFGL-NGS-98420	5	0.1	CCND2	4.93E-06
Multi-Trait	Hapmap26308-BTC-057761	6	0.1	LAP3	6.54E-06
Multi-Trait	BovineHD4100011295	14	0.42	PENK	6.64E-06
Multi-Trait	BovineHD1400007259	14	0.45	PLAG1	8.19E-06
Multi-Trait	BovineHD1400007323	14	0.42	PENK	8.22E-06

Utilizing the most significant SNP identified through GWAS in genomic prediction has been proposed, but addition of these SNP seems to have little effect (Moser et al., 2009). In general, when more information is added to estimate breeding values, the more accurate the estimate will be. Adding these significant SNP may increase accuracy, but it may just be due to the addition of new information. Identifying significant areas of the genome for genomic prediction between breeds can be done (T. Meuwissen et al., 2021), but it will require high density genotypes, which may not be available to industry applications due to cost and computing space. Further exploration into addition of significant SNP in genomic prediction is by artificial intelligence (Li et al., 2018) to understand the best subset of SNP for an accurate genomic prediction through learning algorithms.

Consideration of epistatic effects can also be taken into account, as these effects may introduce a “double-dipping” effect of including multiple SNP that are influencing each other (S. K. Jain & Allard, 1966; Phillips, 2008). This can inflate or deflate the significance of SNP that are directly influencing the phenotype.

Genomic Prediction in Wagyu

Traditional modeling first used sire models to understand and estimate random genetic effects, while still including fixed effects. Use of these methods in the animal industry has shown large leaps in efficiency (Lourenco et al., 2020; Pocrnic et al., 2019) as selection of animals for increased production traits has led to less animals being needed while producing the same amount or more animal products with less animals.

The model used to obtain current breeding values is a BLUP, or best linear unbiased prediction. This model utilizes mixed effects (random and fixed) to estimate the effect a random genotype on a phenotype given fixed effects. The model is as follows:

$$y = Xb + Zu + e$$

Where y is the phenotype, X is the incident matrix for fixed effects, relating each animal to its fixed effect in b , which is the vector of fixed effects, Z is the genotype matrix, u is the vector of random breeding values where we assume $N \sim (0, G\sigma_u^2)$ and e is the vector of residual effects where we assume $N \sim (0, I\sigma_e^2)$. This is the most traditional and widely used model to obtain breeding values for animals that combines production information and animal relationships via genotype. Breeding values are the backbone for the cattle industry, as selection for EPDs (Estimated Progeny Differences) is the most efficient way to make significant change within a cattle population. Solutions can be obtained through mixed model equations:

$$\begin{bmatrix} \hat{b} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + G^{-1}\alpha \end{bmatrix} \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

Where α is defined as σ_e^2/σ_a^2 .

To fully understand the power of genomic prediction, a training and testing group of animals is used to test the accuracy of prediction. In most prediction studies, the training group is much larger than the testing group as it needs to capture the variation in the population sample that is being tested. This means that there needs to be a large sample of animals in this group to get a good understanding of the allele frequencies that are present in a specific population. In many cases, if there is a good representation of the population in the training group, the accuracy of prediction in the testing group will be high (Berry et al., 2016; T. H. E. Meuwissen et al., 2001). In this case, the underlying genomic connections between these two groups would be similar enough to obtain high genomic prediction accuracy of the testing group.

All Wagyu animals outside of Japan can be traced back to a handful of animals exported in the 1970s and 1990s. The commonality of ancestors in Wagyu are apparent, as all US Wagyu, red or black, originate from Japan. There are also some Red/Black crossed animals present in the American population, which may aid in more accurate genomic prediction between the Red and Black populations (Esfandiyari et al., 2015; Moghaddar et al., 2014; Van Grevenhof et al., 2019). If there are common ancestors that are present between training and testing groups, the accuracy

of prediction may increase between the two populations. This is due to the two groups having genomic connections or linkage, which aids in more accurate prediction of a breeding value.

Re-estimating breeding values on a regular basis is crucial to realize higher accuracies, as per generation break down of linkage within the genome can attribute to lower accuracies of breeding values (Cuyabano et al., 2019). Further exploration into reporting of accuracy within an industry setting of animals with predicted breeding values can be realized by this equation:

$$ACC = 1 - \sqrt{\frac{\text{Prediction Error Variance}}{\text{Additive Genetic Variance}}}$$

Where both the prediction error variance and the additive genetic variance can be estimated by solving the mixed model equation. Prediction error variance can be estimated by the inverse of the coefficient matrix (Harris & Johnson, 1998; Misztal et al., 2013), but more computational efficient methods have been proposed, such as an MCMC sampler to estimate these posterior effects (Hickey et al., 2009).

In research settings, where the model used or the population used is the focus of genomic prediction, the accuracy of the breeding value is just the Pearson Correlation between the estimated breeding value and the actual phenotype recorded. The higher this correlation, the more accurate the prediction is. A standardized accuracy of prediction can be found by (J. C. M. Dekkers, 2007):

$$ACC = \frac{p(y, \hat{y})}{\sqrt{h^2}}$$

Where the correlation of the predicted and actual breeding values is in the numerator and the square root of the heritability can be found in the denominator. Testing the accuracy of prediction of a model is large focus in research and industry settings, as increases in accuracy can directly impact the economic impact of selection for a trait.

Genomic prediction has changed the landscape of selection in the animal industry due to the shortening of the generation interval (García-Ruiz et al., 2016). Choosing animals with the greatest genetic potential before they are proven by progeny has increased rate of genetic gain in most animal industries. Identifying those animals before they are proved via their own phenotype or by progeny has been the main driver of selection in all animal industries as this decreases the generation interval and increases genetic gain, as seen in the selection response equation:

$$\Delta G = \frac{ir\sigma_A}{L}$$

Where i is the selection intensity, r is the accuracy, σ_A is the genetic standard deviation and L is the generation interval. The response we expect to see per generation or the difference between the mean phenotype of the offspring of selected parents and the whole parental generation before the selection occurred can be identified through:

$$R = h^2 S$$

Where h^2 is the narrow sense heritability (estimated by $h^2 = \frac{\sigma_a^2}{\sigma_p^2}$, or the proportion of phenotypic variability we can attribute to genotype) and S is the selection differential measured by the average mean of the selected parents comparatively to the population mean, which is usually inferior to those parents selected (Falconer & Mackay, 1996).

The effort into increasing accuracies may not be on the forefront of the average cattleman but must be a consistent drum in the heartbeat of the beef industry. Without high accuracies, the selection decisions made by industry professionals may not be realized, and significant genetic change will not occur. Increasing accuracy can be done by including more performance data on related animals (Quaas & Pollak, 1980), genotyping more animals within the population (Hayes, Visscher, et al., 2009), or improving the pedigree by identifying correct parentage (Geldermann et al., 1986). This is especially important to keep at the forefront of the American Wagyu population, as the pool of animals to select from is much smaller than other traditional breeds. Striving for larger accuracies through phenotype reporting and continuous genotype collection will be crucial for improving Wagyu in the US through genetic change.

Wagyu in America and Future Outlooks

Utilization of Wagyu animals, semen and embryos in the US has taken off due to crossing Wagyu with traditional breeds, such as Angus, Hereford, or dairy breeds, such as Holstein. This is due to the large discrepancy in marbling between Wagyu and traditional beef breeds, which yields a high marbling F1 animal cross. By using a high-marbling breed, the costs of these progeny can yield high grading outputs, as many of these animals will hit the “prime” grading that is highly sought after in the United States. On average, an animal graded prime will come at a \$18 premium as of August 11, 2023, (USDA Livestock Poultry & Grain Market News,

2023). This increased payout per prime graded animal has driven US producers towards Wagyu for a higher grading product.

Genomics is the most effective tool to identify these specialty products. This is because the genotype will not change, and identification of fraudulent labeling of products can be achieved if these genotyping methods can be brought out of the lab. Eventually, bringing genomic methodology into any farmer's hand will be possible. The advent of more mobile sequencing methods (Deamer et al., 2016; Delahaye & Nicolas, 2021; Lamb et al., 2021; Tyler et al., 2018) could lead to identification of animals without the wait of collecting and submitting samples to a 3rd party lab. This is especially important when discussing Wagyu animals, as these animals can fetch up to \$100,000 in private sales. If the animal was to be discovered as not a Fullblood Wagyu with an easy to use, pocket-sized sequencer, then mistakes in purchasing a fraudulent animal could be exposed. Issues may still arise in the bioinformatic pathways, as a straightforward and easy way to obtain usable data from these sequencers can be difficult. Some recent advancements in this area include Dusselpore (Vogeley et al., 2021) and QUILT (Davies et al., 2021), which introduce new bioinformatic platforms that boasts ease-of-use results from ONT sequencers for human DNA. This area of research, bioinformatic platforms that are easy to use, should be in stronger focus if this sequencer is to be used by an average consumer. The jump from sequence to breeding value has many steps in-between that may not be feasible for an everyday user, but is a driving force for rapid selection tools in beef.

Many producers are looking for cutting edge research that will be able to quickly influence their herds, either looking for positive production traits, such as high marbling in Wagyu or for identifying and straying from negative ones, such as recessive diseases. This is especially prominent in the Wagyu breed, as many producers are introducing this “new” breed into their herd. Understanding the breed history, composition and relatedness to other breeds can help industry professionals make more informed decisions in breeding their animals. This can also aid breed associations, as more accurate information on genetic architecture of the breed being evaluated will lead to more accurate breeding values.

Agricultural practices have been around for tens of thousands of years, with very crude “selection” practices starting with choosing the best animals just through physical or behavioral attributes. Many of these selection decisions, whether direct or indirect, led to the creation of our modern-day cattle breeds. Specifically, the indirect selection for East Asian breeds, such as

Wagyu has created an individual product for consumption today. Wagyu animals selected for work, such as pulling plows, were unknowingly developing a high-quality marbling product that we cherish and strive for today.

The evolution from the primitive selection of our very distant ancestors to the modern breeding program is a large jump. Breeding programs today are complex, taking many scientific disciplines to understand the “best” selection decisions for a herd. The many facets of creating a modern breeding program can lead to disconnect between all the specialized areas (molecular, bioinformatic, statistical, functional, etc.) and interpreting the output from each field can be a rather steep learning curve. Yet, all these fields must be considered to create a comprehensive breeding program, but many overlook how important each step is in making an overarching decision. A study into each specialized area and how they impact and interact with each other is needed to better understand the outwardly looking simple process of “selection”. An exploration into obtaining genomic information, creating bioinformatic pipelines, identification of animals, understanding population structure and how this structure can influence genomic prediction, prediction between breeds, and identification of influential SNP to industry traits is needed for full understanding of identifying the best modern breeding techniques for breeding programs.

Fleshing out modern breeding programs in relatively new and high-quality breeds such as Wagyu in the US market should be explored as deeply as possible. Such information that needs to be available is breed composition of these animals, the population structure of these animals comparatively to “traditional” European breeds, and identification of if these Asian breeds can be grouped together in genetic evaluations. This is important to the current state of Wagyu outside of Japan, as evaluations are done with Red (brown wagyu or Akaushi), Black, and Red/Black crossed animal combined in one single evaluation. Including all animals may help give the model more genomic information in the Wagyu population, but there must be enough animals from all groups (brown/red, black, cross) that are related to animals that you want predictions on to have accurate estimates. Accurate estimates of breeding values are of utmost importance in an industry setting, as these estimates directly relate to the monetary value of the animal and its products, whether that be meat or future semen or embryos for breeding.

The Wagyu population in the United States is still in its infancy, with the first Fullblood animals arriving in the 1990s. This has paved a pathway for the demand of high-marbling beef in the western world, which can be achieved even with a 50% Wagyu cross. Identifying those

Wagyu products in the market should also be well established, as a washing out of the Wagyu brand may occur, where fraudulent labeling will tarnish the Wagyu brand. Flushing these products from the marketplace may be in the hands of out of lab sequencing with the Nanopore's MinION. This technology is not foolproof yet, but these are the first steps of product identification through sequence. The importance of keeping genetic diversity in this population should be at the forefront of most selection decisions, as the inbreeding in this population is very high compared to other beef breeds in the US. Increasing accuracy of breeding value estimates as well as keeping inbreeding low will be the critical selection decisions for establishing US Wagyu as a permanent market frontrunner.

LITERATURE CITED

- An, B., Xia, J., Chang, T., Wang, X., Xu, L., Zhang, L., Gao, X., Chen, Y., Li, J., & Gao, H. (2019). Genome-wide association study reveals candidate genes associated with body measurement traits in Chinese Wagyu beef cattle. *Animal Genetics*, 50(4), 386–390.
- Aung, M. M., & Chang, Y. S. (2014). Traceability in a food supply chain: Safety and quality perspectives. In *Food Control* (Vol. 39, Issue 1, pp. 172–184). Elsevier BV.
- Berry, D. P., Garcia, J. F., & Garrick, D. J. (2016). Development and implementation of genomic predictions in beef cattle. *Animal Frontiers*, 6(1), 32–38.
- Bhattaru, S., Tani, J., Saboda, K., Borowsky, J., Ruvkun, G., Zuber, M., & Carr, C. (2019). Development of a Nucleic Acid- Based Life Detection Instrument Testbed. *IEEE Aerospace Conference*.
- Blakebrough-Hall, C., Mcmeniman, J. P., & González, L. A. (2020). An evaluation of the economic effects of bovine respiratory disease on animal performance, carcass traits, and economic outcomes in feedlot cattle defined using four BRD diagnosis methods. *Journal of Animal Science*, 1–11.
- Bosona, T., & Gebresenbet, G. (2013). Food traceability as an integral part of logistics management in food and agricultural supply chain. In *Food Control* (Vol. 33, Issue 1, pp. 32–48). Elsevier.
- Brüniche-Olsen, A., Gravlund, P., & Lorenzen, E. D. (2012). Impacts of genetic drift and restricted gene flow in indigenous cattle breeds: evidence from the Jutland breed. *Animal Genetic Resources/Resources Génétiques Animales/Recursos Genéticos Animales*, 50, 75–85.
- Charlesworth, D., & Charlesworth, B. (1987). Inbreeding Depression And Its Evolutionary Consequences. In *Ann. Rev. Ecol. Syst* (Vol. 18).
- Chen, N., Cai, Y., Chen, Q., Li, R., Wang, K., Huang, Y., Hu, S., Huang, S., Zhang, H., Zheng, Z., Song, W., Ma, Z., Ma, Y., Dang, R., Zhang, Z., Xu, L., Jia, Y., Liu, S., Yue, X., ... Lei, C. (2018). Whole-genome resequencing reveals world-wide ancestry and adaptive introgression events of domesticated cattle in East Asia. *Nature Communications*, 9(1).
- Christensen, O. F., & Lund, M. S. (2010). Genomic prediction when some animals are not genotyped. *Genetics Selection Evolution*, 42(3).
- Clamer, M., Höfler, L., Mikhailova, E., Viero, G., & Bayley, H. (2014). Detection of 3'-end RNA uridylation with a protein nanopore. *ACS Nano*, 8(2), 1364–1374.

- Cuyabano, B. C. D., Wackel, H., Shin, D., & Gondro, C. (2019). A study of genomic prediction across generations of two Korean pig populations. *Animals*, 9(9).
- Davies, R. W., Kucka, M., Su, D., Shi, S., Flanagan, M., Cunniff, C. M., Chan, Y. F., & Myers, S. (2021). Rapid genotype imputation from sequence with reference panels. *Nature Genetics*, 53(7), 1104.
- Deamer, D., Akeson, M., & Branton, D. (2016). Three decades of nanopore sequencing. *Nature Biotechnology* 2016 34:5, 34(5), 518–524.
- Dekkers, J. C. (2004). Commercial application of marker- and gene-assisted selection in livestock: Strategies and lessons,. *Journal of Animal Science*, 82(suppl_13), E313–E328.
- Dekkers, J. C. M. (2007). Prediction of response to marker-assisted and genomic selection using selection index theory. *Journal of Animal Breeding and Genetics*, 124(6), 331–341.
- Delahaye, C., & Nicolas, J. (2021). Sequencing DNA with nanopores: Troubles and biases. *PLoS ONE*, 16(10).
- Elsden, R. P., Hasler, J. F., & Seidel, G. E. (1976). Non-surgical recovery of bovine eggs. *Theriogenology*, 6(5), 523–532.
- Esfandyari, H., Sørensen, A. C., & Bijma, P. (2015). A crossbred reference population can improve the response to genomic selection for crossbred performance. *Genetics Selection Evolution*, 47(1), 1–12.
- Falconer, D., & Mackay, T. F. C. (1996). *Introduction to quantitative genetics* (4th ed.). Prentice Hall.
- Flori, L., Fritz, S., Jaffrézic, F., Boussaha, M., Gut, I., Heath, S., Foulley, J. L., & Gautier, M. (2009). The Genome Response to Artificial Selection: A Case Study in Dairy Cattle. *PLOS ONE*, 4(8), e6595.
- Forristall, C., May, G. J., & Lawrence, J. D. (2002). *Assessing the Cost of Beef Quality*.
- Fortune Business Insights. (2023). *Wagyu Beef Market Share*.
<https://www.fortunebusinessinsights.com/wagyu-beef-market-106905>
- García-Ruiz, A., Cole, J. B., VanRaden, P. M., Wiggans, G. R., Ruiz-López, F. J., & Van Tassell, C. P. (2016). Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. *Proceedings of the National Academy of Sciences of the United States of America*, 113(28), E3995–E4004.

- Geldermann, H., Pieper, U., & Weber, W. E. (1986). Effect of Misidentification on the Estimation of Breeding Value and Heritability in Cattle. *Journal of Animal Science*, 63(6), 1759–1768.
- Gianola, D., & de los Campos, G. (2008). Inferring genetic values for quantitative traits non-parametrically. *Genetics Research*, 90(6), 525–540.
- Gokey, M. (2018). Japan’s obsession with marbling seeps into U.S. *Progressive Cattle*. <https://www.progressivecattle.com/topics/beef-quality/japan-s-obsession-with-marbling-seeps-into-u-s>
- Gonzalez, J. M., & Phelps, K. J. (2018). United States beef quality as chronicled by the National Beef Quality Audits, Beef Consumer Satisfaction Projects, and National Beef Tenderness Surveys — A review. *Asian-Australasian Journal of Animal Sciences*, 31(7), 1036.
- González-Recio, O., Gianola, D., Long, N., Weigel, K. A., Rosa, G. J. M., & Avendaño, S. (2008). Nonparametric Methods for Incorporating Genomic Information Into Genetic Evaluations: An Application to Mortality in Broilers. *Genetics*, 178(4), 2305–2313.
- González-Recio, O., López De Maturana, E., & Gutiérrez, J. P. (2007). Inbreeding Depression on Female Fertility and Calving Ease in Spanish Dairy Cattle. *Journal of Dairy Science*, 90(12), 5744–5752.
- Gotoh, T., Nishimura, T., Kuchida, K., & Mannen, H. (2018). The Japanese Wagyu beef industry: Current situation and future prospects - A review. In *Asian-Australasian Journal of Animal Sciences* (Vol. 31, Issue 7, pp. 933–950). Asian-Australasian Association of Animal Production Societies.
- Gupta, A. K., & Gupta, U. D. (2014). Next Generation Sequencing and Its Applications. *Animal Biotechnology: Models in Discovery and Translation*, 345–367.
- Harris, B., & Johnson, D. (1998). Approximate Reliability of Genetic Evaluations Under an Animal Model. *Journal of Dairy Science*, 81, 2723–2728.
- Hayes, B. J. (2011). Technical note: Efficient parentage assignment and pedigree reconstruction with dense single nucleotide polymorphism data. *Journal of Dairy Science*, 94(4), 2114–2117.
- Hayes, B. J., Bowman, P. J., Chamberlain, A. J., & Goddard, M. E. (2009). Invited review: Genomic selection in dairy cattle: Progress and challenges. *Journal of Dairy Science*, 92(2), 433–443.

- Hayes, B. J., Visscher, P. M., & Goddard, M. E. (2009). Increased accuracy of artificial selection by using the realized relationship matrix. *Genetics Research*, 91(1), 47–60.
- Heather, J. M., & Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics*, 107(1), 1.
- Heffernan, K. R., Enns, R. M., Blackburn, H. D., Speidel, S. E., Wilson, C. S., & Thomas, M. G. (2021). Case study of inbreeding within Japanese Black cattle using resources of the American Wagyu Association, National Animal Germplasm Program, and a cooperator breeding program in Wyoming. *Translational Animal Science*, 5(Supplement_S1), S170–S174.
- Hickey, J. M., Veerkamp, R. F., Calus, M. P., Mulder, H. A., & Thompson, R. (2009). Estimation of prediction error variances via Monte Carlo sampling methods using different formulations of the prediction error variance. *Genetics Selection Evolution*, 41(1), 1–9.
- History of Illumina Sequencing & Solexa Technology*. (n.d.). Retrieved September 21, 2023, from <https://www.illumina.com/science/technology/next-generation-sequencing/illumina-sequencing-history.html>
- Honda, T., Fujii, T., Nomura, T., & Mukai, F. (2006). Evaluation of genetic diversity in Japanese Brown cattle population by pedigree analysis. *Journal of Animal Breeding and Genetics*, 123(3), 172–179.
- Honda, T., Nomura, T., Yamaguchi, Y., & Mukai, F. (2004). Monitoring of genetic diversity in the Japanese Black cattle population by the use of pedigree information. *Journal of Animal Breeding and Genetics*, 121(4), 242–252.
- Horii, M. (2009). Relationship between Japanese Beef Marbling Standard numbers and intramuscular liquid in *M. longissimus thoracis* of Japanese Black steers from 1994 to 2004. *Nihon Chikusan Gakkaiho*, 80, 55–61.
- Hutchison, J. L., Cole, J. B., & Bickhart, D. M. (2014). Short communication: Use of young bulls in the United States. *Journal of Dairy Science*, 97(5), 3213–3220.
- Jain, M., Olsen, H. E., Paten, B., & Akeson, M. (2016). The Oxford Nanopore MinION: Delivery of nanopore sequencing to the genomics community. *Genome Biology*, 17(1).
- Jain, S. K., & Allard, R. W. (1966). The Effects of Linkage, Epistasis, and Inbreeding on Population Changes under Selection. *Genetics*, 53(4), 633–659.
- Jo, M., Garc, J., Almeida, J. M. M. M. De, & Saraiva, C. (2021). Consumer Knowledge about Food Labeling and Fraud. *Foods*, 10, 1–12.

- Johnson, R. (2014). *Food Fraud and “Economically Motivated Adulteration” of Food and Food Ingredients*. www.crs.gov/R43358
- Karamizadeh, S., Abdullah, S. M., Manaf, A. A., Zamani, M., & Hooman, A. (2013). An Overview of Principal Component Analysis. *Journal of Signal and Information Processing*, 04(03), 173–175.
- Kempster, A. J. (1989). Carcass and meat quality research to meet market needs. *Animal Production*, 48(3), 483–496.
- Kidd, K. K., & Cavalli-Sforza, L. L. (1974). The Role of Genetic Drift in the Differentiation of Icelandic and Norwegian Cattle. *Evolution*, 28(3), 381.
- Knol, E. F., Nielsen, B., & Knap, P. W. (2016). Genomic selection in commercial pig breeding. *Animal Frontiers*, 6(1), 15–22.
- Kobe Beef Marketing & Distribution Promotion Association. (2023). *Kobe Beef*. <https://www.kobe-niku.jp/en/contents/council/index.html>
- Kohama, N., Yoshida, E., Masaki, T., Iwamoto, E., Fukushima, M., Honda, T., & Oyama, K. (2021). Estimation of genetic parameters for carcass grading traits, image analysis traits, and monounsaturated fatty acids in Japanese Black cattle from Hyogo Prefecture. *Animal Science Journal*, 92(1), e13664.
- Lamb, H. J., Hayes, B. J., Randhawa, I. A. S., Nguyen, L. T., & Ross, E. M. (2021). Genomic prediction using low-coverage portable Nanopore sequencing. *PLOS ONE*, 16(12), e0261274.
- Lande, R., & Thompson, R. (1990). Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics*, 124(3), 743–756.
- Lee, S.-H., Park, B.-H., Sharma, A., Dang, C.-G., Lee, S.-S., Choi, T.-J., Choy, Y.-H., Kim, H.-C., Jeon, K.-J., Kim, S.-D., Yeon, S.-H., Park, S.-B., & Kang, H.-S. (2014). Hanwoo cattle: origin, domestication, breeding strategies and genomic selection. *Journal of Animal Science and Technology*, 56(1), 2
- Legarra, A., Aguilar, I., & Misztal, I. (2009). A relationship matrix including full pedigree and genomic information. *Journal of Dairy Science*, 92.
- Li, B., Zhang, N., Wang, Y. G., George, A. W., Reverter, A., & Li, Y. (2018). Genomic prediction of breeding values using a subset of SNPs identified by three machine learning methods. *Frontiers in Genetics*, 9(JUL), 377541.

- Lonergan, S. M., Topel, D. G., & Marple, D. N. (2019). Fat and fat cells in domestic animals. *The Science of Animal Growth and Meat Technology*, 51–69.
- Lourenco, D., Legarra, A., Tsuruta, S., Masuda, Y., Aguilar, I., & Misztal, I. (2020). Single-Step Genomic Evaluations from Theory to Practice: Using SNP Chips and Sequence Data in BLUPF90. *Genes*, 11(170).
- Lu, H., Giordano, F., & Ning, Z. (2016). Oxford Nanopore MinION Sequencing and Genome Assembly. In *Genomics, Proteomics and Bioinformatics* (Vol. 14, Issue 5).
- Lucy, M. C. (2007). Fertility in high-producing dairy cows: reasons for decline and corrective strategies for sustainable improvement. *Society of Reproduction and Fertility Supplement*, 64, 237–254.
- MAFF. (2020). *Targets of domestic animal improvement*.
- McVean, G. (2009). A genealogical interpretation of principal components analysis. *PLoS Genetics*, 5(10).
- Meuwissen, T. H. E., Hayes, B. J., & Goddard, M. E. (2001). Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics*, 157(4), 1819–1829.
- Meuwissen, T., Hayes, B., & Goddard, M. (2016). Genomic selection: A paradigm shift in animal breeding. *Animal Frontiers*, 6(1), 6–14.
- Meuwissen, T., Van Den Berg, I., & Goddard, M. (2021). On the use of whole-genome sequence data for across-breed genomic prediction and fine-scale mapping of QTL. *Genet Sel Evol*, 53, 19.
- Michael, T. P., Jupe, F., Bemm, F., Motley, S. T., Sandoval, J. P., Lanz, C., Loudet, O., Weigel, D., & Ecker, J. R. (2018). High contiguity *Arabidopsis thaliana* genome assembly with a single nanopore flow cell. *Nature Communications* 2018 9:1, 9(1), 1–8.
- Misztal, I., Tsuruta, S., Aguilar, I., Legarra, A., VanRaden, P. M., & Lawlor, T. J. (2013). Methods to approximate reliabilities in single-step genomic evaluation. *Journal of Dairy Science*, 96(1), 647–654.
- Mizoguchi, Y., Watanabe, T., Fujinaka, K., Iwamoto, E., & Sugimoto, Y. (2006). Mapping of quantitative trait loci for carcass traits in a Japanese Black (Wagyu) cattle population. *Animal Genetics*, 37(1), 51–54.
- Mizoshita, K., Watanabe, T., Hayashi, H., Kubota, C., Yamakuchi, H., Todoroki, J., & Sugimoto, Y. (2004). Quantitative trait loci analysis for growth and carcass traits in a half-

- sib family of purebred Japanese Black (Wagyu) cattle. *Journal of Animal Science*, 82(12), 3415–3420.
- Moghaddar, N., Swan, A. A., & Van Der Werf, J. H. J. (2014). Comparing genomic prediction accuracy from purebred, crossbred and combined purebred and crossbred reference populations in sheep. *Genetics Selection Evolution*, 46(1).
- Moser, G., Tier, B., Crump, R., Khatkar, M., & Raadsma, H. (2009). A comparison of five methods to predict genomic breeding values of dairy bulls from genome-wide SNP markers. *Genetics Selection Evolution*, 41(1), 1–16.
- Motoyama, M., Sasaki, K., & Watanabe, A. (2016). Wagyu and the factors contributing to its beef quality: A Japanese industry overview. *Meat Science*, 120, 10–18.
- Mukai, F., Tsuji, S., Fukazawa, K., Ohtagaki, S., & Nambu, Y. (1989). History and population structure of a closed strain of Japanese Black Cattle. *Journal of Animal Breeding and Genetics*, 106(1–6), 254–264.
- Namikawa, K. (1980). *Breeding History Of Japanese Beef Cattle And Preservation Of Genetic Resources As Economic Farm Animals*.
- Nomura, T., Honda, T., & Mukai, F. (2001). Inbreeding and effective population size of Japanese Black cattle. *Journal of Animal Science*, 79(2), 366–370.
- Oikawa, T. (2018). Improvement of indigenous cattle to modern Japanese Black (Wagyu) cattle. *IOP Conference Series: Earth and Environmental Science*, 119(1).
- Onteru, S. K., Fan, B., Du, Z.-Q., Garrick, D. J., Stalder, K. J., & Rothschild, M. F. (2012). A whole-genome association study for pig reproductive traits. *Animal Genetics*, 43(1), 18–26.
- Oxford Nanopore Technologies. (n.d.-a). *SmidgION*.
<https://Nanoporetech.Com/Products/Smidgion>.
- Oxford Nanopore Technologies. (n.d.-b). *VolTRAX*. Retrieved February 6, 2021, from <https://nanoporetech.com/products/voltrax>
- Patterson, N., Price, A. L., & Reich, D. (2006). Population Structure and Eigenanalysis. *PLoS Genetics*, 2(12), 2074–2093.
- Phillips, P. C. (2008). Epistasis--the essential role of gene interactions in the structure and evolution of genetic systems. *Nature Reviews. Genetics*, 9(11), 855–867.

- Pocrnic, I., Lourenco, D. A. L., Masuda, Y., & Misztal, I. (2019). Accuracy of genomic BLUP when considering a genomic relationship matrix based on the number of the largest eigenvalues: A simulation study. *Genetics Selection Evolution*, 51(1), 1–10.
- PromethION / Oxford Nanopore Technologies. (n.d.). Retrieved April 7, 2021, from <https://nanoporetech.com/products/promethion>
- Quaas, R. L., & Pollak, E. J. (1980). Mixed Model Methodology For Farm And Ranch Beef Cattle Testing Programs. *Journal of Animal Science*, 51(6).
- Rouse, G. H., Ruble, M., Greiner, S., Tait, R. G., Hays, C. L., & Wilson, D. E. (2000). Growth and Development of Angus-Wagyu Crossbred Steers. *Iowa State University Animal Industry Report*, 1(1).
- Sanger, F., Brownlee, G. G., & Barrell, B. G. (1965). A two-dimensional fractionation procedure for radioactive nucleotides. *Journal of Molecular Biology*, 13(2).
- Sasazaki, S., Odahara, S., Hiura, C., Mukai, F., & Mannen, H. (2006). Mitochondrial DNA Variation and Genetic Relationships in Japanese and Korean Cattle. *Asian-Australasian Journal of Animal Sciences*, 19(10), 1394–1398.
- Schaller, R. R. (1997). Moore’s law: past, present, and future. *IEEE Spectrum*, 34(6), 52–55, 57.
- Schroeder, T. C., & Tonsor, G. T. (2012). International cattle ID and traceability: Competitive implications for the US. *Food Policy*, 37(1), 31–40.
- Scraggs, E., Zanella, R., Wojtowicz, A., Taylor, J. F., Gaskins, C. T., Reeves, J. J., de Avila, J. M., & Neiberghs, H. L. (2014). Estimation of inbreeding and effective population size of full-blood wagyu cattle registered with the American Wagyu Cattle Association. *Journal of Animal Breeding and Genetics*, 131(1), 3–10.
- Seo, D., Lee, D. H., Jin, S., Won, J. Il, Lim, D., Park, M., Kim, T. H., Lee, H. K., Kim, S., Choi, I., Lee, J. H., Gondro, C., & Lee, S. H. (2022). Long-term artificial selection of Hanwoo (Korean) cattle left genetic signatures for the breeding traits and has altered the genomic structure. *Scientific Reports 2022 12:1*, 12(1), 1–15.
- Shahidi, F., & Ambigaipalan, P. (2018). Omega-3 Polyunsaturated Fatty Acids and Their Health Benefits. *Annual Review of Food Science and Technology*, 9(1), 345–381.
- Smith, S. B., Zembayashi, M., Lunt, D. K., Sanders, J. O., & Gilbert, C. D. (2001). Carcass traits and microsatellite distributions in offspring of sires from three geographical regions of Japan. *Journal of Animal Science*, 79(12), 3041–3051.

- Souza-Monteiro, D. M., & Caswell, J. A. (2004). *The Economics of Implementing Traceability in Beef Supply Chains: Trends in Major Producing and Trading Countries*.
<http://www.umass.edu/resec/workingpapers>
- Spink, J., & Moyer, D. C. (2011). Defining the Public Health Threat of Food Fraud. *Journal of Food Science*, 76(9), R157–R163.
- Strandén, I., & Garrick, D. J. (2009). Technical note: Derivation of equivalent computing algorithms for genomic predictions and reliabilities of animal merit. *Journal of Dairy Science*, 92(6), 2971–2975.
- Takasuga, A., Watanabe, T., Mizoguchi, Y., Hirano, T., Ihara, N., Takano, A., Yokouchi, K., Fujikawa, A., Chiba, K., Kobayashi, N., Tatsuda, K., Oe, T., Furukawa-Kuroiwa, M., Nishimura-Abe, A., Fujita, T., Inoue, K., Mizoshita, K., Ogino, A., & Sugimoto, Y. (2007). Identification of bovine QTL for growth and carcass traits in Japanese Black cattle by replication and identical-by-descent mapping. *Mammalian Genome*, 18(2), 125–136.
- Tanabe, T. Y., Hawk, H. W., & Hasler, J. F. (1985). Comparative fertility of normal and repeat-breeding cows as embryo recipients. *Theriogenology*, 23(4), 687–696.
- Tyler, A. D., Mataseje, L., Urfano, C. J., Schmidt, L., Antonation, K. S., Mulvey, M. R., & Corbett, C. R. (2018). Evaluation of Oxford Nanopore’s MinION Sequencing Device for Microbial Whole Genome Sequencing Applications. *Scientific Reports*, 8(1), 10931.
- Uemoto, Y., Suzuki, K., Yasuda, J., Roh, S., & Satoh, M. (2021). Evaluation of inbreeding and genetic diversity in Japanese Shorthorn cattle by pedigree analysis. *Animal Science Journal*, 92(1), e13643.
- USDA Livestock Poultry & Grain Market News. (2023). *National Weekly Cattle and Beef Summary*.
- Van Grevenhof, E. M., Vandenplas, J., & Calus, M. P. L. (2019). Genomic prediction for crossbred performance using metafounders. *Journal of Animal Science*, 97(2), 548–558.
- VanRaden, P. M. (2008a). Efficient Methods to Compute Genomic Predictions. *Journal of Dairy Science*, 91(11), 4414–4423.
- VanRaden, P. M. (2008b). Efficient Methods to Compute Genomic Predictions. *Journal of Dairy Science*, 91(11), 4414–4423.
- Vogeley, C., Nguyen, T., Woeste, S., Krutmann, J., Haarmann-Stemmann, T., & Rossi, A. (2021). DuesselporeTM: a full-stack local web server for rapid and simple analysis of Oxford Nanopore Sequencing data. *BioRxiv*, 2021.11.15.468670.

- Wang, H., Misztal, I., Aguilar, I., Legarra, A., & Muir, W. M. (2012). Genome-wide association mapping including phenotypes from relatives without genotypes. *Genetics Research*, 94(2), 73–83.
- Wang, M., Schneider, L. G., Hubbard, K. J., & Smith, D. R. (2018). Cost of bovine respiratory disease in preweaned calves on US beef cow-calf operations (2011-2015). *Journal of the American Veterinary Medical Association*, 253(5), 624–631.
- Wang, Z., Ma, H., Xu, L., Zhu, B., Liu, Y., Bordbar, F., Chen, Y., Zhang, L., Gao, X., Gao, H., Zhang, S., Xu, L., & Li, J. (2019). Genome-wide scan identifies selection signatures in chinese wagyu cattle using a high-density SNP array. *Animals*, 9(6).
- Wick, R. R., Judd, L. M., & Holt, K. E. (2019). Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biology*, 20(1), 1–10.
- Wolc, A., Kranis, A., Arango, J., Settar, P., Fulton, J. E., O’Sullivan, N. P., Avendano, A., Watson, K. A., Hickey, J. M., de los Campos, G., Fernando, R. L., Garrick, D. J., & Dekkers, J. C. M. (2016). Implementation of genomic selection in the poultry industry. *Animal Frontiers*, 6(1), 23–31.
- Wu, Y., Zheng, Z., Visscher, P. M., & Yang, J. (2017). Quantifying the mapping precision of genome-wide association studies using whole-genome sequencing data. *Genome Biology*, 18(1), 1–10.
- Yun, L., Willer, C., Sanna, S., & Abecasis, G. (2009). Genotype Imputation. <https://doi.org/10.1146/annurev.genom.9.081307.164242>, 10, 387–406.
- Zhang, R., Miao, J., Song, Y., Zhang, W., Xu, L., Chen, Y., Zhang, L., Gao, H., Zhu, B., Li, J., & Gao, X. (2019). Genome-wide association study identifies the PLAG1-OXR1 region on BTA14 for carcass meat yield in cattle. *Physiological Genomics*, 51(5), 137–144.

CHAPTER 3: Investigating New Technologies for On-Site Real-Time sequencing for any Animal Scientist

Hanna Ostrovski, Rodrigo P Savegnago, Wen Huang and Cedric Gondro

Abstract

Animal breeding has been significantly impacted by genomic sequencing technologies, traditionally accessible only to lab professionals due to their complex nature requiring specialized laboratories and trained personnel. However, the emergence of the latest generation of sequencing instruments offers small, portable, real-time devices tailored for inexperienced users to effortlessly obtain genomic data. These new sequencing technologies can open a wealth of opportunities for livestock production systems by bringing testing for livestock directly at the sampling site. This study aims to be the initial exploration into mobile sequencing devices to obtain genomic information by a novice with no previous molecular experience. Sequencing was done with the MinION from Oxford Nanopore Technologies, which is a small portable sequencing device that can be brought out of the lab with an easy protocol for any level of researcher. Whole-genome sequence on one animal was achieved with multiple flow cell runs, with each run producing more data than the last, which points to an improvement in laboratory skills by the user. The maximum amount of genomic information achieved was 5GB, which is a large discrepancy from the possible 50GB per run that ONT states can be accomplished. The bioinformatic pipeline used combined all flow cell outputs into one aligned sequence and had a high breadth of coverage, above 97%, but low depth of coverage at ~8x. This is due to the nature of Nanopore long-read sequencing, as the protocol does not require amplification of the DNA and reads each strand directly through the flow cell nanopores. Results of this initial exploration into the ease of use of ONT technologies are the first steps in providing a roadmap for practical adoption of on-site sequencing applications in agricultural production systems which can improve traceability and livestock production efficiency.

Introduction

Advances within the world of genomics have brought about an ease of obtaining sequence information from any animal at a reasonable price. Today, the most recent generation of sequencers, the third generation, are defined by sequencing single molecules without PCR amplification all while doing this in real-time (Dijk et al., 2018). Sequencing has become a widespread practice in most biology disciplines and has defined the past two decades of animal breeding and genetics. By utilizing sequence information, identification of important QTL (Quantitative Trait Loci), connecting genomic variation to phenotypic variation and genomic prediction have become commonplace for research and industry animal applications.

One of the setbacks of obtaining any type of sequence information is the reliance on an external lab to obtain the genetic information. The current generation of sequencers is the most user-friendly, with the products from Oxford Nanopore Technologies (ONT) boasting a small sequencer and straight-forward protocol that can be used by even untrained personnel. This sequencer, the MinION from ONT (Jain et al., 2016; Lu et al., 2016), is to be the focus of this study which aims at understanding the “in’s-and-out’s” of this device as well as setting the groundwork for future work with this sequencing platform.

Previous studies have started to explore the sequencing possibilities of this device, as it has an easy protocol, set up, and utilizes common lab practices (Lamb et al., 2020). The size of the MinION is perfect for in-field sequencing as it is small enough to fit in your hand and is very light. Field work with this device has already been done in disease diagnostics, e.g., a study done to identify Ebola outbreaks (Quick et al., 2016), Zika outbreaks (Faria et al., 2016) and most recently the outbreak of Covid-19 (Bull et al., 2020). These studies successfully outlined a protocol for mobile sequencing using the MinION while identifying samples for diagnosis and tracing the disease to specific strains or regions of origin. Rapid and mobile analysis of the genome has shown to be possible in these studies and should be applied to the field of animal genetics and genomics.

Establishing usable sequence protocols for users of genomic information with no training in molecular biology can rapidly increase the usage of this technology in the animal industry, which leads to an increase in the use of genomic data. Recent studies have outlined ways to use this technology for genomic prediction (Lamb et al., 2021), which is the backbone to many

animal production groups. Understanding the full capabilities of the MinION is crucial in bringing genomic sequencing to the forefront of diagnosis, rapid testing, and traceability in the animal industry.

Materials and Methods

Development of Protocol

The focus of this study was to identify the protocol to obtain whole genome sequence of an Akaushi bull from the MinION from ONT (Fig. 3.1). Most importantly, this study aimed to understand if genomic information could be obtained with this device by someone with no training within a traditional lab setting. The methodologies of obtaining a DNA library and sequencing with the MinION will be the focus, as this is a recent technology which has some troubleshooting to be done within the lab protocol. This protocol should serve as a blueprint for other newcomers to the world of sequencing and should outline the pros and cons of this device in this context.



Figure 3.1 ONT MinION size relative to an average human hand.

Illumina Sequencing

The most laborious part of obtaining Illumina data was just sending off the blood sample to a service provider. This is a very hands-off way to get high-coverage and accurate data. The beauty of the Illumina platform is the complete hands-off approach, where the goal is to obtain the sequence, not to revolutionize the practice. These two sequencing platforms come with different approaches; ONT's (Oxford Nanopore Technologies) MinION is a low throughput portable device suitable for a hands-on approach without the need to send samples to external labs, while Illumina technologies are high throughput devices that require a dedicated lab setup.

MinION Initial Costs and Materials

The MinION was purchased along with flow cells and 3rd party reagents from NEBNext (NEBNext® Companion Module for Oxford Nanopore Technologies® Ligation Sequencing) that

are required for sequencing. All other lab materials, specifically mixers, pipettes, thermocyclers, computing systems, common reagents and AMPure beads needed for the ligation protocol were already available.

The MinION device along with 12 flow cells cost \$4,500, and the required 3rd party materials upwards of \$1,000 for 24 reactions. The cost of additional flow cells is \$900 per flow cell, but ONT does provide cheaper package options when purchasing more than one flow cell at a time.

This system is quite time sensitive, as the flow cells have a shelf life of only 3 months. This amount of time can be a stretch for some experiments and limits the user to have to purchase the flow cells in succession of need. Obtaining substantial amounts of flow cells at the beginning of an experiment runs the risk of that flow cell “going bad”, which can be a huge blow to the cost of an experiment. In this experiment, all flow cells were used within 3-5 months of receiving the materials from ONT. No issues were found with the integrity of the flow cell past its use-by date, but differences in output could have been due to differences in flow cell construction by ONT.

Library Preparation Protocol

All experiments run on the MinION were done using the LSK-109 protocol, which is the ligation protocol provided by ONT. Other needs of experiments can be fulfilled with the wide selection of protocols/reagent kits provided by ONT with purchase of a flow cell. This ligation protocol contained a control experiment, which comes with all protocols from ONT. This experiment gives the user time to practice creating a library and loading the library into the flow cell without wasting DNA from the experimental source of interest. This control experiment is great for initial hiccups in the protocol, but can also be a large consumer of reagents, time, and flow cells. The issues that may arise in the actual experiment, which is the main aim, will not be the same issues that arise in the control experiments. These issues could range from low-quality DNA, not enough extracted DNA used in library prep, or even too much DNA used. Problems that may arise in the main experiment due to these issues may be irreversible and can cost the experimenter time and money.

The subject of interest for obtaining WGS was a 6-month-old Akaushi (Red/Brown Wagyu) bull. Whole blood was collected from this calf and DNA was extracted using the

QIAamp blood kit (*QIAamp DNA Blood Kits*, n.d.). The LSK-109 protocol was laid out as an easy-to-use process, which included three steps: 1) DNA repair and end-prep, 2) adapter ligation and clean-up and 3) priming and loading the SpotON flow cell. The first step was done with the reagents included in the NEBNext Companion Module for ONT ligation sequencing, which specifically includes the DNA repair buffer and mix and the end-prep reaction buffer and enzyme mix. This step was the most difficult, as most of these reagents used were mixed at low quantities (1-3 μ l) and without pipetting experience could be difficult to be precise. The most important part of this step is the DNA sample itself, which the protocol only calls for 1 μ l if the sample is of high-density (100-200 ng). The DNA used for this study was measured in the range of 130-140 fmol with the Qubit fluorometer. Each run consisted of creating a library, loading it onto a flow cell and running the flow cell for about 72hrs. After each run was finished, the flow cell was washed to remove the library and put back into refrigeration.

The first runs were all done with 1 μ l of DNA, but due to low sequencing output, the amount of DNA used was increased to 2 μ l on the last 2 runs. Increasing the amount of extracted DNA used had a positive outcome to the amount of DNA sequenced, as the final runs produced the most sequence. This could also be due to the experience of the user increasing. Increasing the amount of DNA used is a risk though, as it could either increase DNA sequence output due to the abundance of DNA primed in the sample or it can clog the nanopores in the flow cell array (Kubota et al., 2019) with an overabundance of DNA and result in low output.

The library preparation is designed to take around 60 minutes, but without proper training, this preparation can take up to 2 hours. The library preparation via the protocol provided by ONT contains most reagents needed after the first step of DNA end-repair, except for the AMPpure beads, which are molecular beads that help clean the sample throughout the protocol. Most steps in the protocol are easy to follow, but some can be time sensitive, and many steps require a steady hand. These are all hurdles that a beginner will overcome with practice, but slight missteps will result in having to restart the preparation. Some troubleshooting steps can be found in Table 3.1.

Troubleshooting Table			
Step	Problem	Possible Reason	Possible Solution
Ethanol Wash	Low amount of sequenced DNA	Evidence of ethanol still in sample	Use higher ethanol % in wash (75-80%) to dry quicker
Loading flow cell	Bubble formation	Incorrect pipetting, introduction of air in nanopores	More stable/advanced pipetting practice
Pipetting DNA into sample	Low amount of sequenced DNA	Not enough DNA used in library preparation	Increased amount of DNA used (2 µl)
Thermal Cycler	Low amount of sequenced DNA	Not enough time in thermal cycler Not hot enough in thermal cycler	Increased time at higher temperature in first end-prep step which includes use of thermal cycler
Use of Ampure beads	Low amount of sequenced DNA	Need longer amount of time binding and un-binding Need more beads for clean-up when increased DNA sample	Incubated for 15 min at 37 to allow for binding (or unbinding) Increased beads introduced by 1.5x when using larger initial DNA sample

Table 3.1 Problems that arose in sequencing analysis with the ONT MinION and solutions that were explored to troubleshoot

There are two stopping points within the library preparation, one after the DNA repair and end prep and one after last step in preparation before flow cell loading. Each stopping point states that the library can be stored and should be used within the next 12 to 24 hours to get the best results. This was not done in this study but could be vital to other experiments.

After library preparation the DNA library is loaded into the flow cell within the MinION through the SpotON port on the flow cell. Loading the SpotON point in the flow cell is the most detrimental step within the protocol if done incorrectly. The library must be loaded along with a priming solution without the introduction of air bubbles. These bubbles have the capability to permanently damage the sequencing array which contain the “nanopores” in a lipid bilayer which do all the genome sequencing.

Basecalling

Sequencing is done by the nanopores by recording the difference in electrical signals from the baseline as each base passes through the pores. These signals are lovingly named “squiggles” which are the raw signal output from the MinION (Rang et al., 2018). Changes in electrical current can be determined as different base pairs by a basecaller, the most common one being from Nanopore itself, “guppy”. The guppy software is fast and efficient, but is not free for anyone to use, as it is only available to Nanopore customers. Other 3rd party software is available for download and has proven to produce high-quality and accurate data that is alike to the output from “guppy”. Basecalling converts raw squiggles into a fastq format, which is then used in the alignment procedure.

Bioinformatics

Post-basecalling procedures process these reads into aligned sequences and are dependent on bioinformatic programs with a variety of steps. These steps include an initial quality control, an adapter trimming, alignment of the sequence, sorting all files, converting files into bam format, calling the variants, and finally obtaining alignment statistics such as depth of coverage, percent of genome covered, and total number of variants called. Alignment of this data consisted of established bioinformatic software which includes porechop (Wick et al., 2019), samtools and bcftools (Li et al., 2009; Li & Barrett, 2011), longshot (Edge & Bansal, 2019) and minimap2 (Li, 2018). Many other programs exist that do a myriad of analyses for alignment. Here, we outline a proposed pipeline for analysis of nanopore reads with common bioinformatic tools.

The first program used for these reads is the R-script based program MinIONQC (Lanfear et al., 2019), which outputs plots of diagnostics of each run on a flow cell. Such diagnostics include the q-score of the reads, the length of all the reads, how many reads were

produced, the number of reads produced over time, and output diagnostics per nanopore in each flow cell. These are all important outputs to understand each run on the MinION itself, as many tools focus on the output fast5 files, it is still crucial to look at how each run compared to improve future runs on the MinION.

The bioinformatic pipeline continues with adapter trimming done by porechop, where adapters from reads were removed. The alignment was done with minimap2 using the cattle reference assembly. Sorting the reads, indexing the reference genome and reads (creating a blueprint to align reads to the genome more efficiently) and merging all reads together was done by many operators in samtools. Variant calling was done with Longshot, which is a variant caller designed for long-read sequences. Alignment statistics were produced by using operators of bcftools that manipulate the final vcf format.

This pipeline aims at aligning sequence from multiple runs with the MinION with a reference genome while combining all runs together to get one whole genome sequence from the animal of interest. The pipeline is not unlike many proposed bioinformatic pipelines that exist for other sequence information (Zhou et al., 2019) but is specifically tailored to reads from the ONT MinION. This technology was used for sequencing of the whole genome, which may be the incorrect approach to utilize this sequencer best. The MinION is best suited for obtaining a small amount of genomic information very quickly while also being small enough to be mobile.

The bioinformatic pipeline to align and call variants for the Illumina output was done using IVDP (<https://github.com/rodrigopsav/IVDP>). All of these programs are commonly used for Illumina outputs, as these genotypes are the most commonly used within the genomic space.

Results

Application of Protocol

The final whole genome sequence obtained from this method had 5.2x coverage depth and 2,329,110 variants called. These variants called by longshot only include one sample, which may explain the low number of variants called. Of these variants called, there were 856,924 homozygous (1/1) calls and 926,780 heterozygous (1/0) calls. Output from each run can be found in Table 3.2.

<i>Run Number</i>	<i>Total Gb</i>	<i>Total Reads</i>	<i>Mean Length</i>	<i>Max Length</i>	<i>Ultra- long reads</i>
1	0.13	12,000	11,011	66,867	0
2	1.08	136,000	7,993	117,208	4
3	2.61	262,516	9,945	126,941	6
4	1.79	244,000	7,340	107,872	4
5	3.36	372,000	9,035	112,098	10
6	0.93	112,000	8,334	99,962	0
7	5.53	644,000	8,597	129,442	11

Table 3.2 Output from each run at 72 hours from the MinION.

The first runs had the worst outcome due to the learning curve of the protocol techniques that required some troubleshooting. By the final run, the protocol was working more efficiently, and the researcher had more practice with the laboratory techniques. This is still a low amount of coverage for producing a full genome sequence, with a more acceptable coverage hovering from 10-30x. Utilizing low-pass imputation techniques may be able to bypass this issue and fill in the areas of lower coverage with reference genomes (Snelling et al., 2020).

The sequencing protocol that was generated through all runs in this experiment differed from the original protocol taken from the Nanopore website. The general path taken to acquire WGS can be found in Figure 3.2. This was expected, as not all experiments are created equally, and the experience level of the author was low to none at the start. Most of the protocol differences were taken from the Nanopore community page, which filled in the gaps of knowledge where the protocol was lacking. The protocol outlined in the Methods section is aligned with many of the suggestions shared online, as well as trial and error from the runs done.

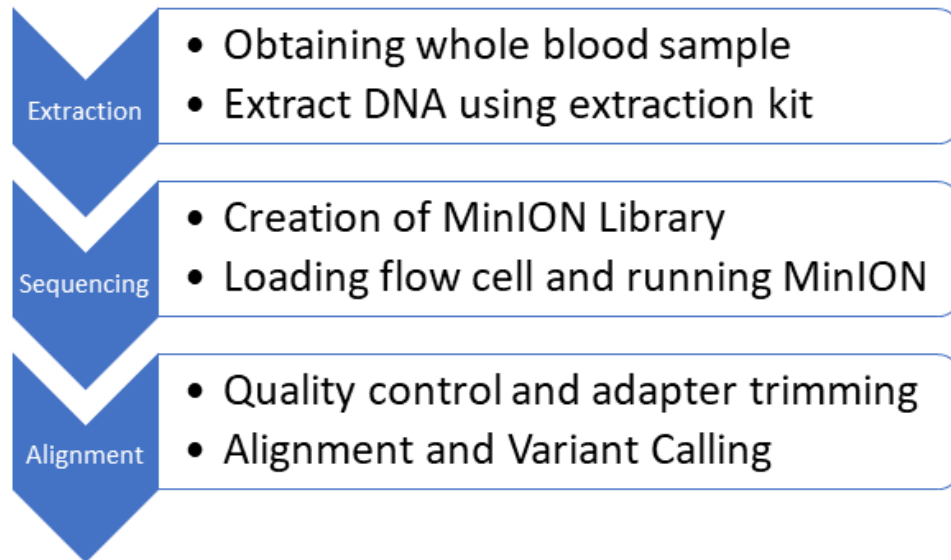


Figure 3.2 Workflow for sequencing with an ONT MinION in this study.

The bioinformatic pipeline produced for alignment of these reads was the easiest to diagnose, as data can be manipulated much easier than the physical DNA sample. This pipeline, like most, consists of 3rd party software that is free to use except the basecalling guppy software. This could lead to issues for future users who may be using ONT products but do not have access to download guppy. The alignment of all sequence data would take the user a couple of days to do, as some of these programs are doing complex operations on huge data files. The user must also be aware of memory and storage issues, as these files are not known to be small, although the speed of alignment does correlate with the size of the genome being used. The longest step is the basecalling process, which can take many days to complete just one run that produces over 4 GB without GPU capabilities. The success and time spent in this process depends on the experience of the user, as this study aimed to use this technology through the eyes of a beginner.

Comparison of Sequencing Platforms

Multiple methods of genomic sequencing were considered for this study: the Illumina HiSeq and Oxford Nanopore's MinION with flow cells. These technologies come at different price points; the MinION costing around \$5,000 for the technology, 3rd party reagents and flow cells to run experiments, and the HiSeq, which was around \$2,500 for one animal at WGS. The difference in price is large when considering the "start-up" one-time costs with the MinION, but future runs will only cost the user the amount of the consumables needed.

Coverage output of the Illumina HiSeq was at 40x with 98% of the genome covered, comparative to the Nanopore WGS, which had a much lower coverage depth, but did have 97% of the genome covered (Table 3.3).

	Coverage Depth (bam files)		
	Coverage Depth	Breadth	Depth of Covered Positions
Illumina HiSeq	39.16x	98.62%	39.71x
Nanopore Flow Cells	8.21x	97.18%	8.44x

Table 3.3. Coverage depth and breadth of the sequencing platforms used for WGS on one Akaushi bull.

Comparison of each method yielded comparable results in breadth but had a significant difference in coverage depth. This is partly due to protocols that do not need amplification of the DNA sample, which makes the library preparation process much quicker. DNA amplification does have its benefits, most clearly, the increase in coverage depth as seen in the Illumina output which utilizes amplification. Due to the time amplification can take (up to many hours), removing amplification from the protocol can bring sequence to the user much faster.

Conclusions

The new frontier in animal sequencing is to get it out of the lab, speed up the results collected, and to make it available to all types of researchers and producers. The MinION has the possibility to do all these things, as the size and protocol are easily picked up and easy to adjust to any type of experiment with an experimenter of any skill level. The protocol provided by ONT for producing sequence could be more in-depth, as many steps were simply glazed over, and inexperienced lab users could easily be lead astray. Producing good results of each experiment run with the MinION depends on the user tweaking the amount of reagent used, the time spent incubating or even the time/speed of mixing. Understanding ONT protocols for a specific experiment for a beginner requires reliance on problem solving with the Nanopore community, which can make the sequencing process more labor intensive, with a constant “debugging” of sorts of the MinION and library preparation. The bioinformatic pipeline produced for alignment of these

reads consists of 3rd party software that is free to use except the basecalling guppy software and the MinKNOW sequencing software. This could lead to issues for future users who may be using ONT products but do not have access to download this software.

Future technologies from ONT could help remove the human error and produce more accurate sequences. The VolTRAX (Oxford Nanopore Technologies) is a recent launch from ONT that takes the human out of the library preparation all together, as it is a device that does the library preparation itself. Other library kits also exist, such as the Rapid kit, which can take a high-quality DNA sample from raw DNA to library in ~20 minutes. Depending on the objectives of the research, the technologies and chemistry can be “mixed-and-matched” to create a user-friendly protocol.

The MinION has a long laundry list of pros and cons, but the novelty of this technology is important for use in the animal industry. This is because the idea of mobile sequencing, which can be used by any person of any skill level, is unheard of in the sequencing space. This is the only sequencer on the market that can be taken out of the laboratory, which opens the door to the opportunity of on-farm sequencing, at the right price. The practical application of this aspect allows for on-site diagnosis of disease, quick genomic sequencing of animals for quick parentage testing and even traceability of animal products.

LITERATURE CITED

- Bull, R. A., Adikari, T. N., Ferguson, J. M., Hammond, J. M., Stevanovski, I., Beukers, A. G., Naing, Z., Yeang, M., Verich, A., Gamaarachchi, H., Kim, K. W., Luciani, F., Stelzer-Braid, S., Eden, J. S., Rawlinson, W. D., van Hal, S. J., & Deveson, I. W. (2020). Analytical validity of nanopore sequencing for rapid SARS-CoV-2 genome analysis. *Nature Communications* 2020 11:1, 11(1), 1–8.
- Dijk, E. van, Jaszczyszyn, Y., Naquin, D., & Thermes, C. (2018). The third revolution in sequencing technology. *Trends in Genetics*, 34(9), 666–681.
- Edge, P., & Bansal, V. (2019). Longshot enables accurate variant calling in diploid genomes from single-molecule long read sequencing. *Nature Communications* 2019 10:1, 10(1), 1–10.
- Faria, N. R., Sabino, E. C., Nunes, M. R. T., Alcantara, L. C. J., Loman, N. J., & Pybus, O. G. (2016). Mobile real-time surveillance of Zika virus in Brazil. *Genome Medicine*, 8(1), 97.
- Hayes, B. J., & Daetwyler, H. D. (2019). 1000 Bull Genomes Project to Map Simple and Complex Genetic Traits in Cattle: Applications and Outcomes. *Annual Review of Animal Biosciences*, 7, 89–102.
- Jain, M., Olsen, H. E., Paten, B., & Akeson, M. (2016). The Oxford Nanopore MinION: Delivery of nanopore sequencing to the genomics community. *Genome Biology*, 17(1).
- Kubota, T., Lloyd, K., Sakashita, N., Minato, S., Ishida, K., & Mitsui, T. (2019). Clog and Release, and Reverse Motions of DNA in a Nanopore. *Polymers*, 11(1).
- Lamb, H. J., Hayes, B. J., Nguyen, L. T., & Ross, E. M. (2020). The Future of Livestock Management: A Review of Real-Time Portable Sequencing Applied to Livestock. *Genes*, 11(12), 1478.
- Lamb, H. J., Hayes, B. J., Randhawa, I. A. S., Nguyen, L. T., & Ross, E. M. (2021). Genomic prediction using low-coverage portable Nanopore sequencing. *PLOS ONE*, 16(12), e0261274.
- Lanfear, R., Schalamun, M., Kainer, D., Wang, W., & Schwessinger, B. (2019). MinIONQC: fast and simple quality control for MinION sequencing data. *Bioinformatics*, 35(3), 523–525.
- Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 34(18), 3094–3100.
- Li, H., & Barrett, J. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, 27(21), 2987–2993.

- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & Subgroup, 1000 Genome Project Data Processing. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079.
- Lu, H., Giordano, F., & Ning, Z. (2016). Oxford Nanopore MinION Sequencing and Genome Assembly. In *Genomics, Proteomics and Bioinformatics* (Vol. 14, Issue 5).
- NEBNext® Companion Module for Oxford Nanopore Technologies® Ligation Sequencing / NEB. (n.d.). Retrieved January 9, 2022, from <https://www.neb.com/products/e7180-nebnext-companion-module-for-oxford-nanopore-technologies-ligation-sequencing>
- Oxford Nanopore Technologies. (n.d.). VolTRAX. Retrieved February 6, 2021, from <https://nanoporetech.com/products/voltrax>
- QIAamp DNA Blood Kits. (n.d.). Retrieved January 9, 2022, from <https://www.qiagen.com/us/products/discovery-and-translational-research/dna-rna-purification/dna-purification/genomic-dna/qiaamp-dna-blood-kits/>
- Quick, J., Loman, N. J., Duraffour, S., Simpson, J. T., Severi, E., Cowley, L., Bore, J. A., Koundouno, R., Dudas, G., Mikhail, A., Ouédraogo, N., Afrough, B., Bah, A., Baum, J. H. J., Becker-Ziaja, B., Boettcher, J. P., Cabeza-Cabrerizo, M., Camino-Sánchez, Á., Carter, L. L., ... Carroll, M. W. (2016). Real-time, portable genome sequencing for Ebola surveillance. *Nature*, 530(7589), 228–232.
- Rang, F. J., Kloosterman, W. P., & de Ridder, J. (2018). From squiggle to basepair: Computational approaches for improving nanopore sequencing read accuracy. In *Genome Biology* (Vol. 19, Issue 1, pp. 1–11). BioMed Central Ltd.
- Snelling, W. M., Hoff, J. L., Li, J. H., Kuehn, L. A., Keel, B. N., Lindholm-Perry, A. K., & Pickrell, J. K. (2020). Assessment of Imputation from Low-Pass Sequencing to Predict Merit of Beef Steers. *Genes*, 11(11), 1312.
- Wick, R. R., Judd, L. M., & Holt, K. E. (2019). Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biology*, 20(1), 1–10.
- Zhou, A., Lin, T., & Xing, J. (2019). Evaluating nanopore sequencing data processing pipelines for structural variation identification. *Genome Biology*, 20(1), 237.

CHAPTER 4: Mobile, Rapid Beef Product Identification through 3rd generation Sequencing Methods

Hanna Ostrovski, Yasir Nawaz, Rodrigo P Savegnago, Wen Huang, and Cedric Gondro

Abstract

The highly sought-after Wagyu cattle breed, celebrated for its exceptional quality, originated in Japan, yet access to these animals has been globally restricted since the 1990s, with no live animal Wagyu genetics becoming available. The early slow adoption of Wagyu in the US and their high-quality beef products has seen a recent rise. The surge in the availability of Wagyu in the United States has coincided with a notable increase in consumer demand for high-marbling beef, in which Wagyu are prized for. This surge has brought concerns about product authenticity, as the label "Wagyu" has been liberally applied to products without stringent verification processes, raising doubts about the true origin. Quick verification of product is needed and can be achieved through genotype. Recent sequencing technologies have made rapid, on-site genomic traceability a thing of the present. This study aims to create a protocol to quickly identify product breed composition through mobile sequencing products from Oxford Nanopore's MinION. This mobile protocol was aimed at the everyday user with a low cost that has wide accessibility. Wagyu samples were genotyped with the MinION and 100K Bovine SNP chip, then compared via principal component analysis for initial identification. Low output from the MinION found PCA as a poor identifier and more precise breed classification was needed. This was accomplished through haplotype correlation and concordance rate against a large reference of many breeds at whole-genome sequence. The MinION output identified these animals through haplotype matching with high correlation (0.55) and concordance rates (0.94). Product identification and certification through genotype for breed claims on Wagyu was accomplished through this rapid sequencing kit. Further exploration into Nanopore products will pave a path of putting the power of high-quality beef product verification in the hands of a consumer.

Introduction

Identifying products from producer to consumer has shown to be difficult in the modern era of high production and consumption of millions of pounds of animal goods. The term traceability in research rhetoric has become commonplace and many studies have proposed approaches to tackle this problem (Aung & Chang, 2014; Bosona & Gebresenbet, 2013; Schroeder & Tonsor, 2012). One of the most varied products in the United States is Wagyu beef. The labeling of Wagyu products is not heavily protected and encompasses all animals from 50% to 100% Wagyu breed composition.

Many tagging or electronic identification practices are not a permanent identifier and could fall off, not be read correctly by scanners, or may be installed incorrectly and result in a human error. The most permanent identifier of animal-based products is DNA. Identifying animal products from the beginning of its life to the product via sequencing could be the answer to product identification without relying on physical or electronic tagging. Although it is known that every animal has their own genotype, it is not always easy to sequence and connect that sequence to a product. The process of DNA sequencing relies on molecular biologists, bioinformatic scientists, geneticists, animal scientists and even computer scientists. The demand for genomic information obtained out of the lab has grown, as technologies have opened doors for scientists, researchers, and students to obtain genotypic information by themselves for their applications (Gupta & Gupta, 2014).

Third-generation sequencing technologies have paved the way for the possibility of mobile sequencing, with the MinION being introduced in 2014 (Jain et al., 2016). Other technologies for DNA extraction in an easy and quick method have also been introduced (*QIAamp DNA Blood Kits*). Mobile DNA sequencing kits have not been established by a specific technology company, but other studies have introduced ways to sequence DNA out of the lab (Lamb et al., 2020). Nanopore's MinION has paved the way for taking the complexity out of wet lab protocols through a protocol that requires less time and lab materials than traditional sequencing methods. The outputs of these protocols are genomic sequences that are obtained through long-read technology. These long reads can identify structural variation that is not detectable with traditional short-read sequencers (Nguyen et al., 2023; Zhou et al., 2019).

Comparison of multiple sequencing techniques using Nanopore technologies has been documented (Bowden et al., 2019; Stefan et al., 2022; Tyler et al., 2018) but an exploration into the use of the “flongle” has still to be done. This flongle (R9.4.1) is some of the most cost-effective technology for sequencing using the MinION, with the cost of a single flongle starting at \$90 (Oxford Nanopore Technology, n.d.). Previous studies with the MinION have outlined obtaining WGS with smaller genomes with ease and sufficient coverage (King et al., 2021; McNaughton et al., 2019; Taylor et al., 2019), most notably, obtaining significant and accurate sequence of pathogens and bacteria. As of this date, there have been no studies that have tried to achieve sequencing for breed composition in an animal population. Utilization of rapid, mobile sequencing using this flongle technology could pave the way for animal identification in an out-of-lab setting at a lower price.

This study aims to create an out of lab sequencing protocol that is now available to the everyday scientist for sample identification. Initial exploration into feasibility of obtaining sequence from the MinION flongle will create the baseline for creation of a low-cost mobile sequencing kit to identify breed composition, especially in Wagyu labeled products. The “gold standard” of sequencing technologies in the animal sciences, the Illumina 100k Bovine chip, was utilized in this study as a baseline of breed identification. Further placement of animals in their respective breeds will be done using PCA. Previous studies have outlined the impact that PCA has on breed identification (Destefanis et al., 2000) to identify a sample’s positive match for breed identification. Rapid identification of sample breed composition, especially in the Wagyu breed, can give consumers peace of mind when purchasing these high-end beef products. Ease of use of these sequencing products can open the door to verification by any beef enthusiast.

Material and Methods

Sample Collection

Initial blood collection was done on 14 Wagyu animals and 15 Akaushi animals at the Michigan State UPREC research center in Chatham, MI. The blood was sent off to NeoGEN in Lincoln, NE and genotyped at Illumina 100k Bovine Chip and used for sequencing with the out of lab protocol.

DNA extraction for Flongle Methods

DNA extraction was done with the QIAamp mini blood kit (*QIAamp DNA Blood Kits*), but the protocol was changed for adoption to an out-of-lab protocol. The use of a mini centrifuge at 6,000 RPM (*Mini Centrifuge, 6,000 RPM, White / Southern Labware*) and a mobile heating block (*One-Block Digital Dry Bath 115V / Southern Labware*), were crucial for creating a protocol that can be mobile while also keeping costs low. Increasing the concentration of DNA for use in the flongles for each sample was integral to the out-of-lab protocol, so the amount of buffer in the final elution step was halved in the DNA extraction. Testing the concentration of the samples was done with the NanoDrop ND-1000 in a lab setting and was only used to identify if this protocol was viable outside of the lab. The samples ranged from a concentration of 72-89 ng/ul, which is within the needs of the MinION flongle flow cells (Lu et al., 2016). All samples were extracted using these tested methods and labware that can be used in an out-of-lab setting.

Nanopore Flongle Methods and Mobile Sequencing Kit

Not all animals were considered in the flongle analysis due to time and cost limitations. In total, 9 animals per breed were chosen and run separately on individual flongles. The animals chosen from each breed were at random with the “random” function in R. The initial total of 18 animals were used to establish the flongle pipeline for out-of-lab sequencing.

The technology used to obtain genomic information was ONT’s MinION, along with the flongle flow cell and flongle adapter. The kit used to make the DNA library for the flongles was the Rapid Sequencing Kit (RAP-004), which boasts a very quick library preparation time. After initial practice, the protocol can be done within 30 minutes. Third party materials were needed for this protocol (AMPpure beads), which does largely increase the cost of the budget.

The protocol was designed to be a mobile lab that any level of scientist could do at a reasonable cost. The goal for this study was to create a whole mobile kit that could follow the sample extraction, DNA extraction, flongle library preparation and final bioinformatic pipeline that could be completed out of the wet lab. This was done by purchasing mobile lab elements, such as a heating block, mini centrifuge and pipettes that were specifically for out-of-lab usage. Other consumables were needed such as pipette tips, nuclease free water, ethanol and tubes. All these elements and their approximate price can be found in Table 4.1.

Mobile Lab Component	Price (USD)
MinION Device	\$1,000
Library Preparation	\$1,460
Flongles (x12) and Flongle Adapter	
Pipettes	\$477
Pipette Tips	\$150 per box
Nuclease Free Water	\$40 for 500mL
Omega XP Beads	\$107 for 5mL
Ethanol	\$38 for 500mL
Heating Block + Adapters	\$381 + \$39.51 & \$80
Mini Centrifuge	\$150
Magnetic Rack	\$59
Computer (Dell g15 Gaming Laptop)	\$1,200

Table 4.1 Price breakdown of mobile lab for MinION using flongles.

The initial runs were done following the exact protocol laid out by Nanopore. After the first runs, some editing was done to the protocol to best fit the needs of the input sample and the wants of the goals of this study. Increasing the amount of DNA used in the preparation steps and increasing the Rapid Adapter (RAP) had positive outcomes by increasing total output.

Output from some flongle runs was very low, while others were quite successful. Troubleshooting at the beginning of analysis was necessary, as this was a new technology introduced into the lab. Most problems arose due to flongle quality control, which is run through the MinKNOW program before library preparation is done. Many flongles were received as defective, which is classified as a flongle with less than 60 pores as functional out of the 126 total pores (Rang et al., 2018). In this case, Nanopore will send a new flongle to replace the defective one. Time between quality control of the flongle and receiving the replacement may take up to 2-3 weeks, so timing issues can arise if a batch of flongles was found to be defective.

Bioinformatic Analysis

The pipeline for the flongle samples from fast5 output files from the MinION to final vcf files is as follows: guppy (Wick et al., 2019) was used for basecalling and alignment with the UCD ARS 2.1 cattle genome, samtools (Li et al., 2009) was used for merging, sorting and indexing bam files, bcftools (Danecek et al., 2021) was used for final vcf calling. Bcftools was used because the output from the flongles were minimal compared to the initial exploration between Illumina and Nanopore at whole genome level.

Since output from flongles was low, utilization of vcf outputs was compared at different number of calls per position as well as quality of the position call. This was done through manipulation of the vcf columns of DP and QUAL. DP standing for the depth of coverage at a certain call and QUAL standing for the quality of that call. Low quality levels and lower depth included more variants, but then sacrificed the accuracy of these calls. Common depth for high accuracy is around 40x with quality above 20 (De La Cerda et al., 2023; Delahaye & Nicolas, 2021).

Breed Identification through PCA

Originally, 14 animals from the Akaushi and Wagyu herds were sampled, but not all animals were used in the protocol (Table 4.2). This was due to time and material availability, as the flongles received did not all pass the quality number of pores available. Some samples were also run multiple times (H002, 732, J101 and H006), as the protocol was not well established at the beginning of this study and is highly prone to human error. Those re-runs are marked by a “redo” in the sample name.

BREED			
AKAUSHI (RED WAGYU)		BLACK WAGYU	
H002redo	H002	19	20
H006	H006redo	J101	J101redo
H007	G902	J103	J102
H009	H004	732	408
H010	A101	199	D808
		732redo	

Table 4.2 Breed identification of all samples used. Samples with "redo" were sequenced with more than 1 flongle due to low initial output.

Initial exploration into breed identification utilized principal component analysis. Since each of these samples had different outputs, with different calls, one large matrix containing all animals could not have been created. Each sample was run through a bioinformatic analysis that included subsetting the calls the sample had against calls available in the 1000 bull's directory. This subset of calls per sample yielded a PCA per sample. To obtain principal components, a genetic relationship matrix was created using this (VanRaden, 2008) method:

$$G = \frac{Z'Z}{2 \sum p_i(1 - p_i)}$$

Where Z is a matrix of centered allele effects and p_i is the allele frequency at locus i . The resulting G matrix was utilized to obtain the eigen values and eigenvectors for PCA. This analysis was done in R using the eigen package and plotting the first 2 principal components. The dispersal of these values per animal around the mean 0 was then plotted by breed, with percent of variation explained calculated by taking the variance of the eigenvalues (McVean, 2009; Patterson et al., 2006).

Breed Identification Method through haplotype Blocks

The method of obtaining low pass sequencing with the MinION yielded results that are not fit for imputation due to the long-read structure of the reads obtained. Solving this issue through imputation of the dataset is not feasible. Long-read sequencing has this downfall, as it is not a pinpointed sequencing method in this case, which then outputs large chunks of the genome

which are usually not evenly distributed throughout the genome. A more effective method of identification of these animals was created without the need for imputation.

Identifying haplotype blocks from long-read sequences from the flongle output proved to be straightforward approach for breed identification with the flongle. The initial haplotype from the flongle was used without the need for a variant caller. The haplotype blocks were then compared to haplotypes from all breeds in the 1000 bulls directory as well as Wagyu and Akaushi samples that had been collected at whole genome sequence.

Samtools mpileup was used to get all the nucleotides and positions for reads with a mapping quality threshold of 60. Insertions were deleted from the data and positions were matched to 1000 bulls genome data to subset the reads to keep variants only. The number of total variants sequenced, number of reads obtained with the flongle and number of those reads aligned as well as passed quality control can be found in Table 4.3. Those samples that had low output of reads subsequently had a low number of variants and aligned reads that were able to be used in haplotype correlation. The sample nucleotides were considered as a haplotype and converted into 0 and 1 based on counts of the reference allele.

Sample ID	Breed	Number of Variants	Total Reads	Aligned Reads	QC Reads
19	Wagyu	107,726	10,816	6,895	5,219
199	Wagyu	254,558	10,538	6,831	4,834
20	Wagyu	761,962	28,902	18,840	14,823
408	Wagyu	765,020	39,206	25,653	20,506
732	Wagyu	348,650	29,789	21,386	16,815
732redo	Wagyu	4,014,254	206,208	135,512	108,950
A101	Akaushi	741,301	29,402	19,209	14,304
D808	Wagyu	2,884,050	120,262	76,594	61,300
G902	Akaushi	1,444,057	49,817	31,815	24,950
H002	Akaushi	39,616	1,973	1,349	1,025

Table 4.3 Number of variants obtained for haplotype matching procedure. This includes total number of variants obtained, total number of DNA strands read, number of those strands that were aligned and number of those strands that passed quality control.

Table 4.3 (cont'd)

<i>H002redo</i>	Akaushi	1,727,485	111,330	70,094	53,510
<i>H004</i>	Akaushi	2,598,474	116,639	71,686	56,608
<i>H006</i>	Akaushi	31,082	1,810	1,261	1,040
<i>H006redo</i>	Akaushi	3,207,971	160,989	99,679	78,601
<i>H007</i>	Akaushi	3,707,136	146,562	89,832	70,287
<i>H009</i>	Akaushi	997,026	50,891	28,674	21,176
<i>H010</i>	Akaushi	1,196,141	66,197	41,036	30,719
<i>J101</i>	Wagyu	157,110	13,583	8,684	6,713
<i>J101redo</i>	Wagyu	1,305,754	59,603	38,517	31,426
<i>J102</i>	Wagyu	699,505	21,716	12,464	9,611
<i>J103</i>	Wagyu	4,711,443	163,471	98,634	76,461

The reference data, which consisted of many cattle breeds, was subset down based on positions of haplotypes obtained from the flongle output and converted into haplotype format. The correlations of sample haplotypes with reference haplotypes were calculated using Pearson's correlation.

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

Concordance percentages were also calculated for every pairwise comparison between sample and reference haplotypes. This was done across the genome for each animal in the reference population for each breed. Correlations and concordances were used against each breed in the reference population to identify the breed with highest match with the sample. The largest concordance and correlation of haplotypes from sample to reference breed animal were then grouped as that reference animal's breed.

Results

Flongle Output

The output from each flongle run can be found in table 4. Three scenarios were considered, the position had a depth of 1x, 2x, or 3x or greater, with all quality filtering set to 15. Differences in number of variants used from Table 3 are due to differences in quality threshold, as the quality threshold used for PCA analysis was set very low to obtain as many variants as

possible, without consideration of accuracy. Output of the flongles from the mobile sequencing kit showed promising results, as some sample runs had a high number of variants sequenced (see 732 redo and J103 in Table 4.4). The coverage depth and breadth of the flongles may not be enough for an accurate whole genome sequence but can give some insight into the sequence of the animal. The issue lies in the reliability of the genome, as low depth and breadth may introduce inaccurate calls in the genome.

<i>Sample ID</i>	<i>Position Coverage Depth</i>		
	<i>Depth at 1x</i>	<i>Depth at 2x</i>	<i>Depth at 3x</i>
<i>19</i>	4,930,124	34,765	3,625
<i>20</i>	32,767,833	825,457	40,237
<i>199</i>	10,730,608	61,254	6,788
<i>408</i>	33,942,193	883,729	55,992
<i>732</i>	15,907,514	281,956	8,347
<i>732 redo</i>	174,803,776	16,473,840	1,138,012
<i>A101</i>	32,014,043	973,671	58,633
<i>D808</i>	126,877,137	8,928,299	605,251
<i>G902</i>	618,447,212	2,332,279	163,267
<i>H002</i>	1,704,607	19,969	4,143
<i>H002 REDO</i>	76,524,848	3,538,080	223,932
<i>H006</i>	1,347,878	8,897	0
<i>H006 REDO</i>	139,912,708	10,554,241	693,717
<i>H007</i>	162,263,656	17,100,294	1,385,234
<i>H009</i>	43,039,064	1,308,600	95,175
<i>H010</i>	52,202,351	2,005,568	158,992
<i>J101</i>	7,102,697	84,323	2,022
<i>J101 REDO</i>	57,663,228	1,845,338	97,239
<i>J102</i>	29,885,656	828,256	28,632
<i>J103</i>	202,218,709	22,203,105	1,713,119

Table 4.4 Output of each sample at quality filtering of 15 at different depth

There were many flongles that did not run correctly due to low pore capacity or low amount of sequence in the library preparation (see H002 and H006 samples in Table 4.4). The minimum number of flongle pores that pass quality control to run is 50 pores (Delahaye & Nicolas, 2021) out of the 126 pores available per flongle. Anything under that threshold can be replaced by a new flongle if under the 4-week window of warranty that the flongles are allowed. Due to the variability in flongle quality received and initial protocol troubleshooting, only 7 Akaushi and 9 Wagyu were used, as multiple flongles were run for a single animal if sequencing had failed.

Due to low coverage and the nature of long-read sequencing, traditional imputation methods of the flongle outputs were difficult. This is due to the long-read sequence itself, as it is not conducive to imputation in that it outputs large chunks of the genome that may not have adequate coverage of the genome. In short-read sequencing, it is more likely to get many small segments of the genome that will cover a larger range of the genome (Whiteford et al., n.d.). Imputation programs are written for the latter sequencing scenario, in which the program can take those smaller sections in masse and imply the larger sections. When given the case of large sections with large missing sections in-between, the imputation software can falter.

Breed Identification with 100k

As a baseline for the greater sample size of animals used, the blood samples were not only sequenced with the flongle mobile set up but were genotyped at 100k as well. The animals were genotyped to understand what the PCA should look like with the flongle output, and if the sampled animals would group correctly within their respective breeds. In theory, the PCA plots should look identical if enough data was collected on the flongles. The PCA of animals used can be found in Figure 4.1. All samples of the two breeds, Wagyu and Akaushi, can be found via the legend in salmon (akaushi_test) and in red (wagyu_test). The samples that were collected for this study followed the pattern that was expected and were grouped into the correct breeds. This baseline confirms that these animals are in fact either Wagyu or Akaushi and group within what breed that they were assigned.

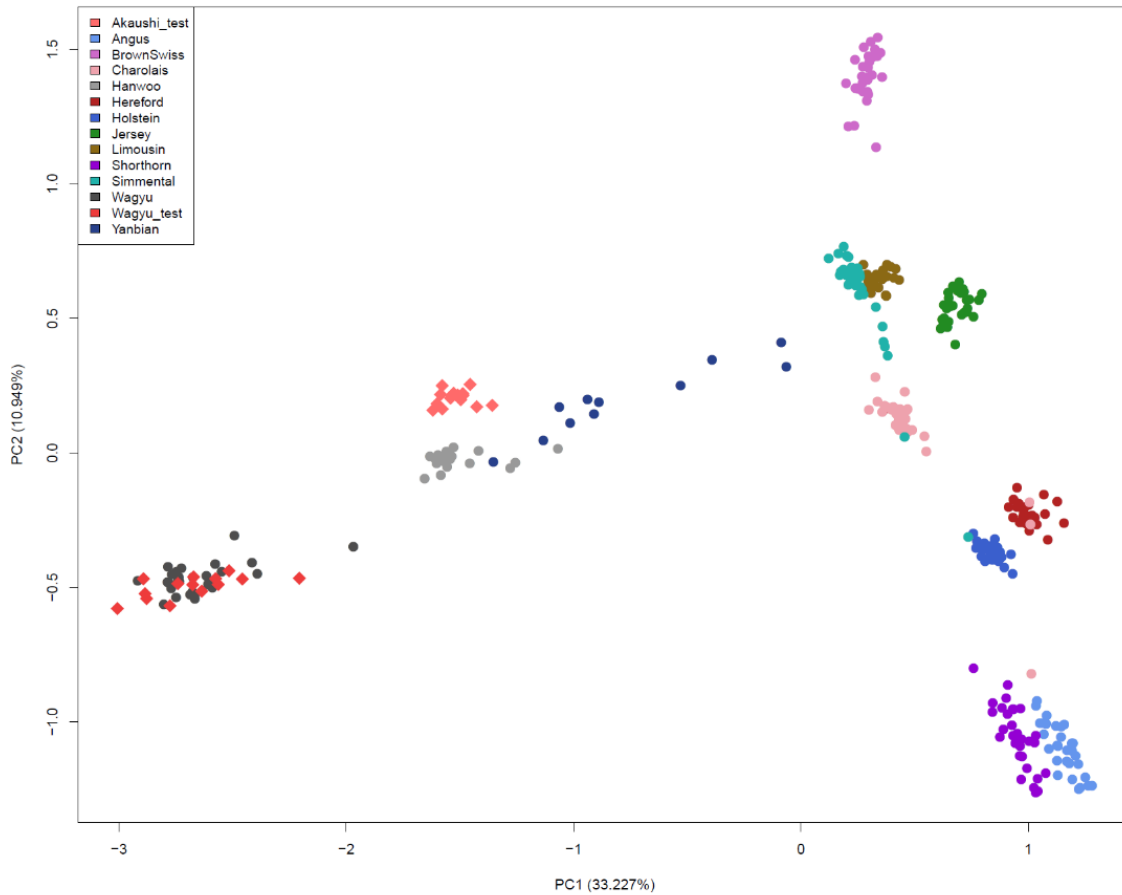


Figure 4.1 Principal Component Analysis with 100k genotypes of Wagyu and Akaushi samples with other cattle breeds to identify breed composition.

Breed Identification using Flongle Mobile Kit via PCA

The goal for the flongle output is to mimic the grouping in Figure 4.1. To do this, each VCF output for 9 Wagyu and 7 Akaushi were filtered at low depth and quality at each position to make up for low output. Each DNA strand in the library preparation is read into Nanopores as one long read (Clamer et al., 2014) unlike the 100k genotype that consists of small reads that cover each position at a large depth. The nature of long-read sequencing may not acquire considerable depth at each position but can span large swaths of the genome in single sequence strands. Figures 4.2, 4.3 and 4.4 show each scenario when filtering for depth with 3 examples chosen based on output: a high amount of output (J103), an average amount of output (H002redo), and a low amount of output (19).

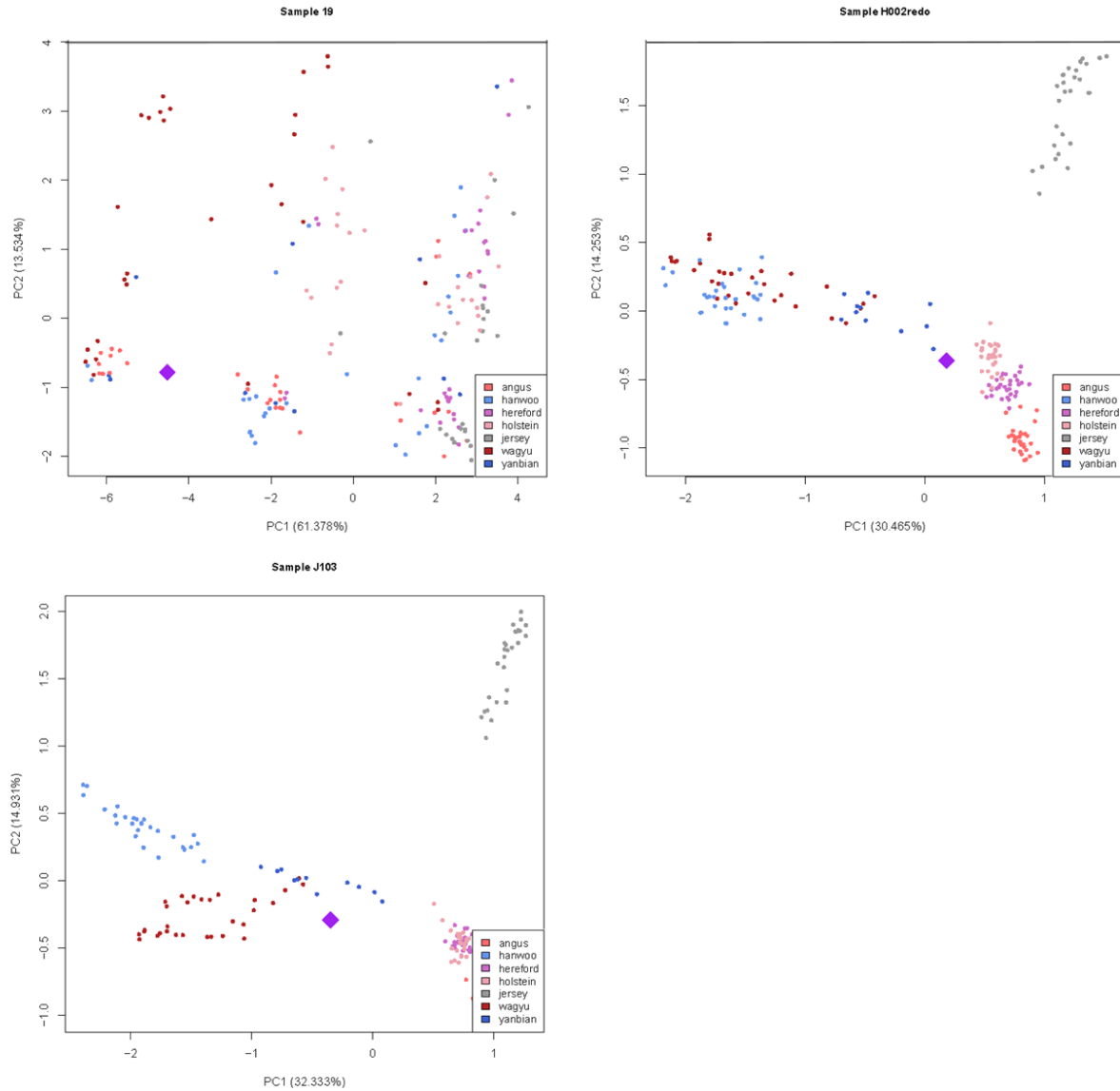


Figure 4.2 Principal component analysis with coverage depth filtering set to 1 and quality of 15. Sample 19 had low sequence output, sample H002 had an average sequence output and sample J103 had a high sequence output. Each sample can be identified as a purple diamond and should have been grouped in the Wagyu breed.

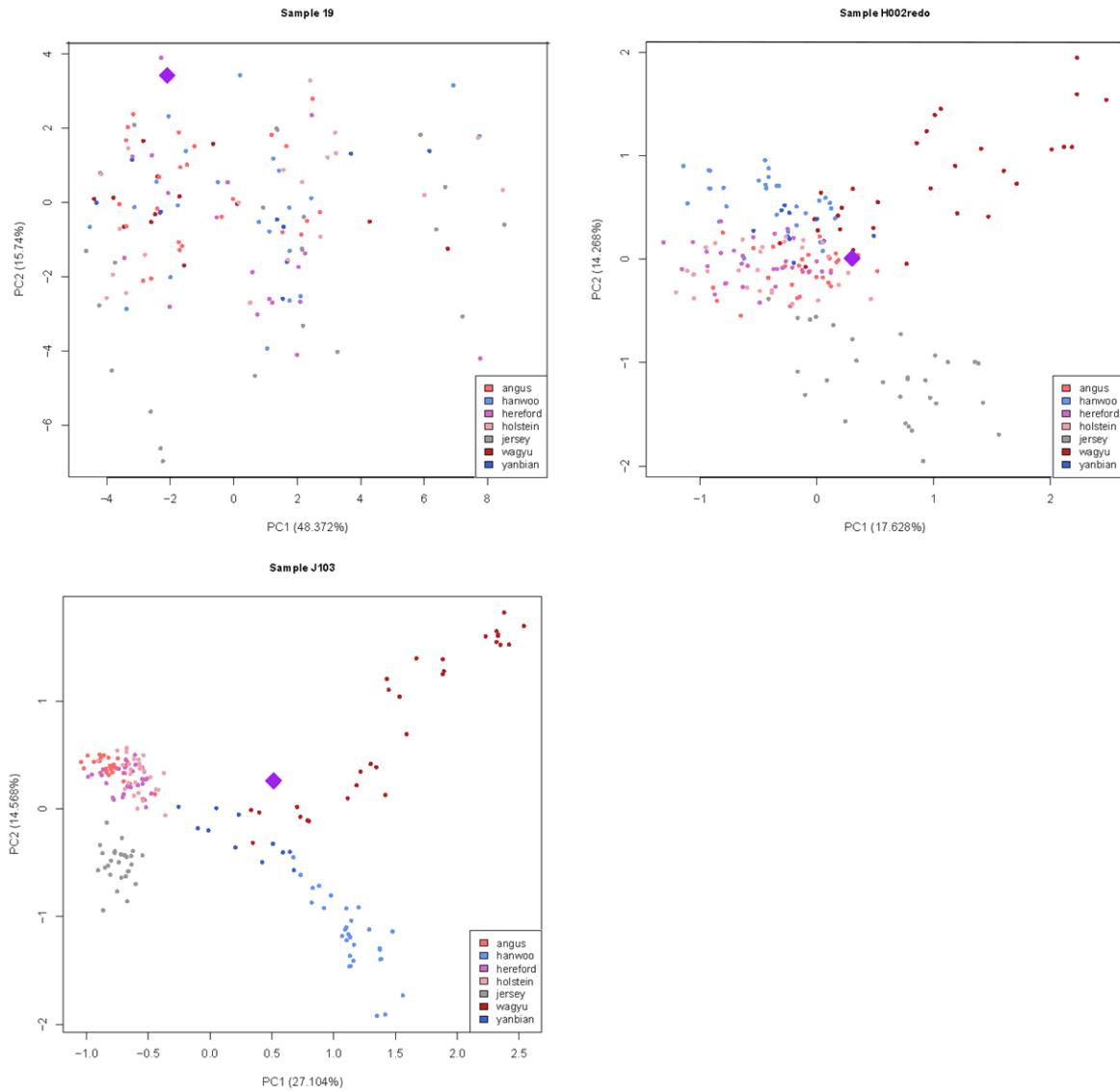


Figure 4.3 Principal component analysis with coverage depth filtering set to 2 and quality of 15. Sample 19 had low sequence output, sample H002 had an average sequence output and sample J103 had a high sequence output. Each sample can be identified as a purple diamond and should have been grouped in the Wagyu breed..

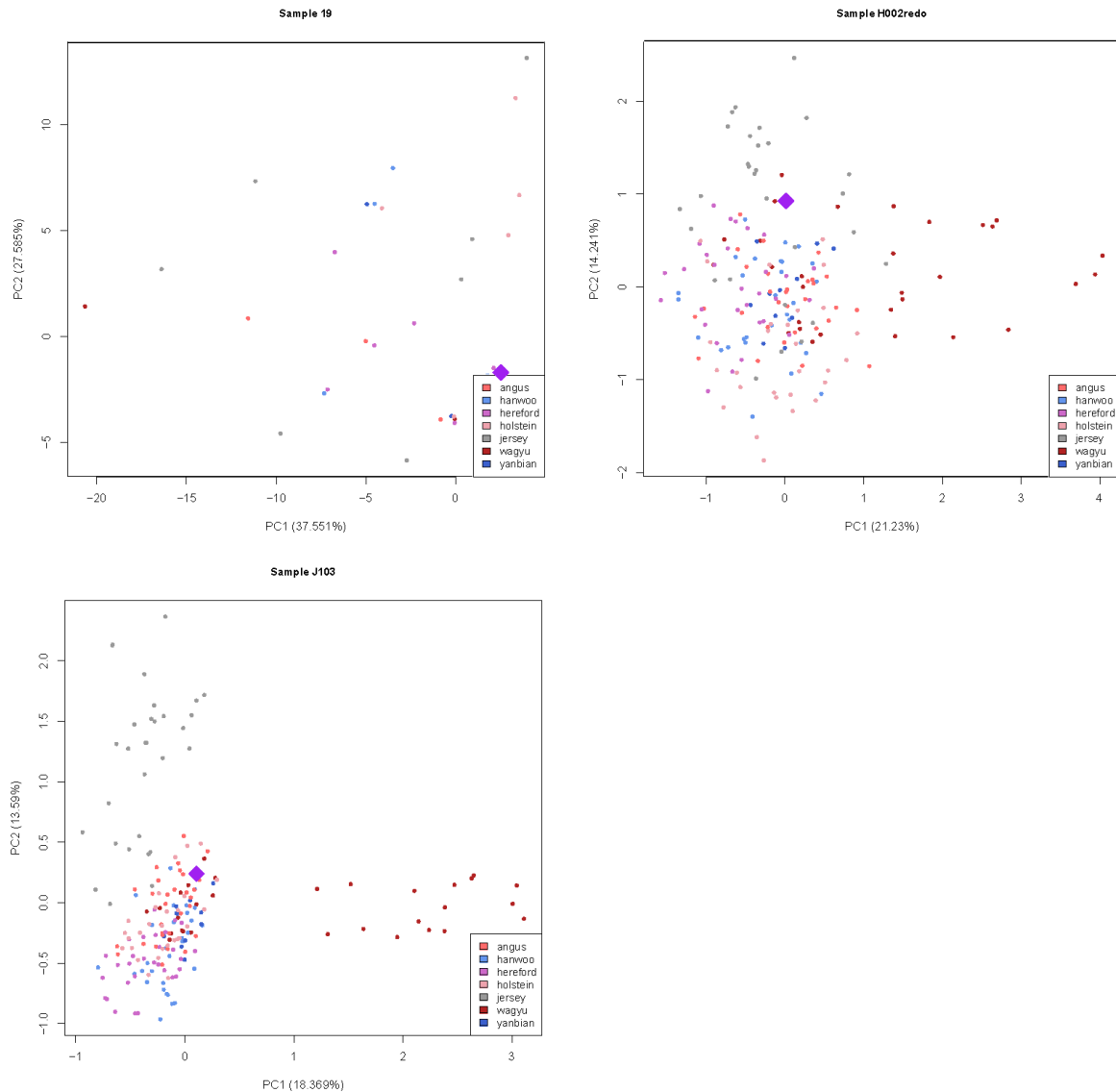


Figure 4.4 Principal component analysis with coverage depth filtering set to 3 and quality of 15. Sample 19 had low sequence output, sample H002 had an average sequence output and sample J103 had a high sequence output. Each sample can be identified as a purple diamond and should have been grouped in the Wagyu breed..

The overall trend shows that the more positions you sequence, the better you may be at predicting the breed of the animal sequenced with the flongle. Yet, when you have the most positions when filter for low quality and depth, you are risking those positions being incorrect. The lower the accuracy of the call, the possibility of these animals' correct breed identification is also lower or cannot even be realized with low amount of sequence available, such as in Sample 19. The possibility of inaccurate placement of an animal in a breed becomes very high especially

in animals who belong to closely related groups, such as the Akaushi (Red Wagyu) and Black Wagyu.

Breed Identification through Haplotype Correlation

The number of reads obtained from this procedure was low, so traditional imputation strategies could not be employed without possibility of imputation “toward the mean” of a reference population. As seen in PCA results of figures 4.2, 4.3, and 4.4, it was not possible to subset the positions obtained against a reference to run principal component analyses for placement in a breed group. Another method of breed identification was employed through matching haplotypes with the reference population. Results of haplotype correlation indicated that the correlations ranged between 0.18 to 0.55 while the concordance percentages ranged between 0.79 to 0.94 for all animals in the reference genome.

A plot of haplotype correlations per breed with each sample group (Wagyu or Akaushi) can be seen in Figure 4.5 and concordance of samples with reference haplotypes can be seen in Figure 4.6.

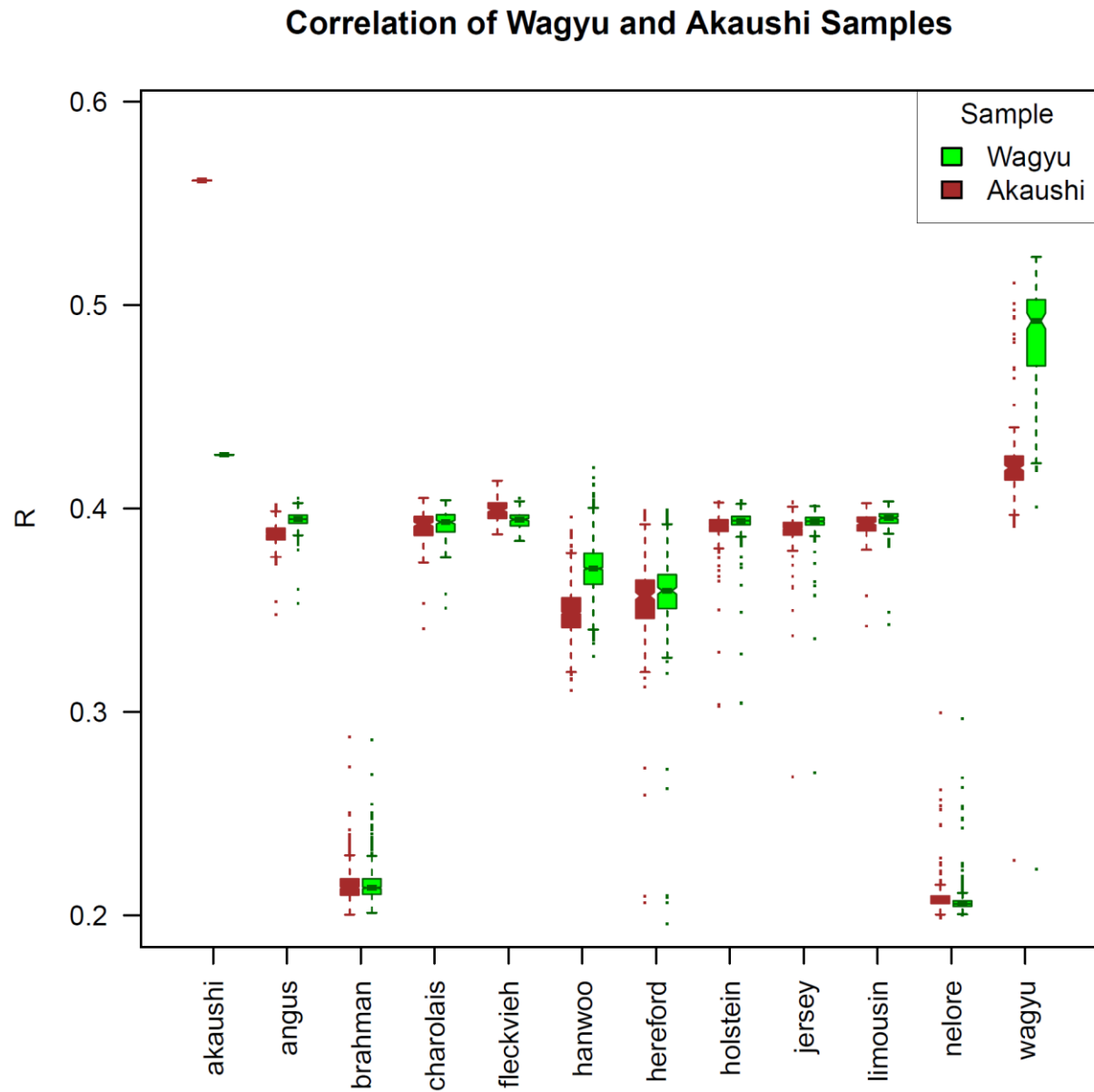


Figure 4.5 Correlation of Wagyu and Akaushi haplotypes with reference population haplotypes. Each breed considered can be seen on the X-axis, and the correlation of haplotype blocks in red (Akaushi) or green (Wagyu). Highest overall correlation can be seen between the sample's realized breed and their respective reference breed group.

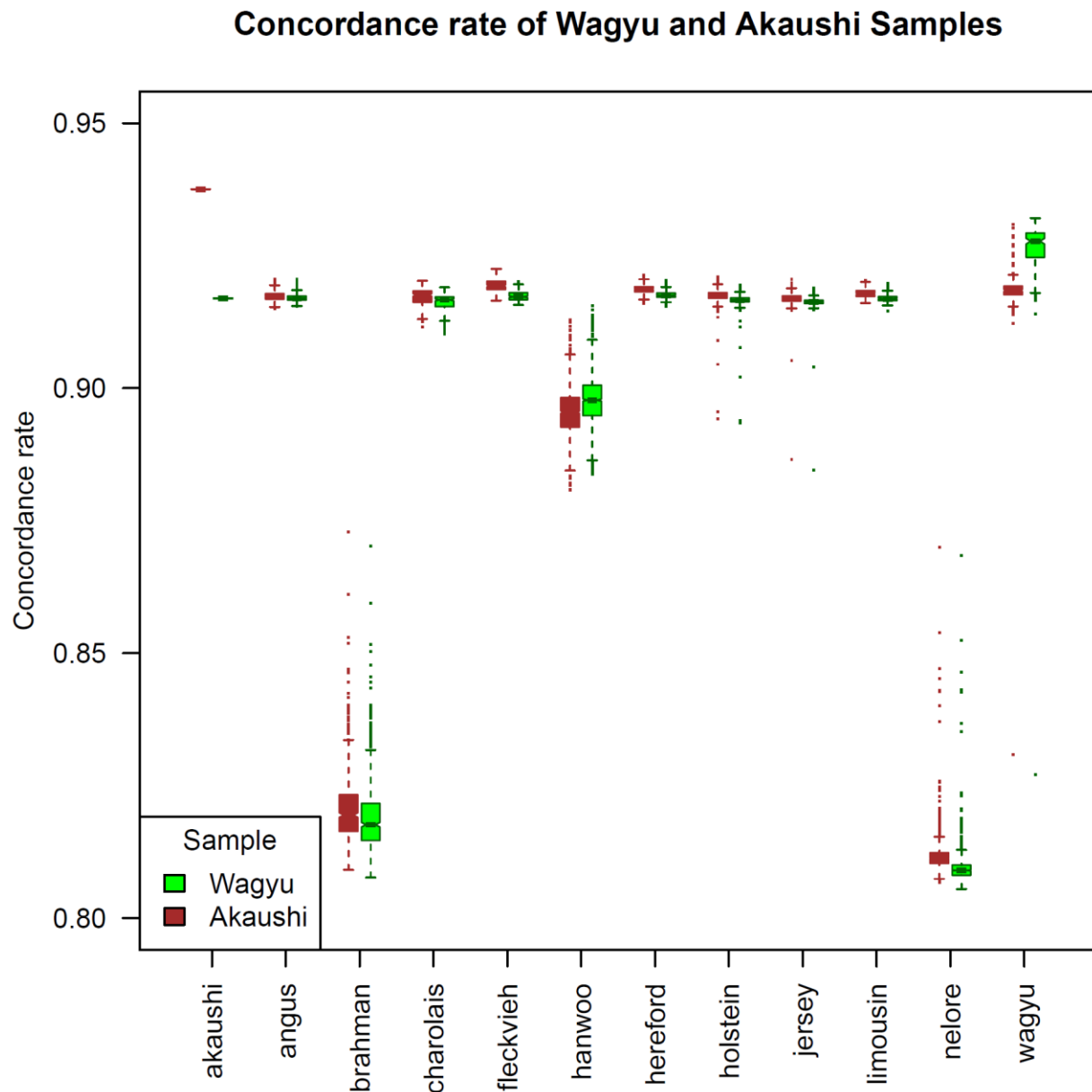


Figure 4.6 Concordance rates for Wagyu and Akaushi haplotype blocks with reference haplotypes. Highest overall correlation can be seen between the sample's realized breed and their respective reference breed group.

The mean concordance and correlations of sample haplotypes were highest for their respective breeds (i.e., Wagyu and Akaushi). The lowest correlations and concordances were observed for distantly related breeds like Brahman and Nelore which are indicine cattle.

Conclusions

Overall, the mobile kit proved to be a viable option for an out-of-lab sequencing protocol. The initial setup is costly, but most consumables, excluding the nanopore flongles, are a cheap replacement. The output from the flongle sequencing via mobile kit may not prove to be effective in principal component analysis, but enough sequence was obtained to identify animals through another method using haplotypes. Breed identification using a PCA sacrifices specificity and accuracy, as the need for dense variant arrays that span the genome is crucial for building a genetic relationship matrix that can identify breed composition. Many variants may be sequenced, but only a few may contribute into breed identification (O'Brien et al., 2020). This process, which includes all breeds that are to be considered, demands some sort of phasing or imputation, as the breed database is built on genotyped animals with a standard SNP map. To solve this issue, a new approach was established to better estimate breed composition without the need to phase to a certain SNP density.

Recent imputation programs have been created specifically for long read nanopore data. The QUILT (Davies et al., 2021) program utilizes these long-reads in a haplotype imputation via Gibbs sampling. Utilizing this program for low-coverage, whole-genome data will be crucial in identifying breeds from a group of samples. Accurate imputation from low density genotype samples opens the door for more intensive breed identification, from principal component analysis to STRUCTURE (Porrás-Hurtado et al., 2013) analysis. The setback to this method is the reliance on good databases. Breed identification cannot be known without first having the many breeds to compare to.

Bringing efficient and accurate genomic sequencing out of the lab has proven to be a difficult task. It is even harder to bring sequencing to the masses at a low cost with an easy-to-follow protocol. More up-front costly setups can achieve greater output (Lamb et al., 2021), which can be a solution if up-front cost is not issue. Many logistical issues such as sample collection, processing time, faulty materials, and insufficient data for bioinformatic analysis have been in the way of real-time sequencing success by any scientist. These are all hurdles that were faced while creating this protocol and sequencing kit. This technology is still in its infancy, with new flongle chemistry and engineering being produced annually. The cost of this technology and the difficulty of use reflects its novelty, and the protocol is still out of reach for an everyday

farmer or scientist. Identification of these samples was achieved, but through a novelty approach, which would not have been possible without the knowledge to do so, which is not commonplace. Even with all the setbacks, this new frontier of sequencing has proved to be able to identify breed composition through genotype in an out-of-lab setting. This achievement in technology was not even feasible 25 years ago, and new technologies upcoming from Nanopore aim to make sequencing more efficient and user friendly.

The successful identification and certification of Wagyu beef through genotype analysis using rapid sequencing kits marks a significant stride in ensuring the authenticity of this high-quality breed. These advancements not only offer a reliable means of verifying breed claims but also emphasize the potential for wider consumer trust in US Wagyu. Nanopore products have the potential to enhance transparency and trust within the food industry, ensuring that premium products like Wagyu maintain their integrity and value throughout the supply chain.

LITERATURE CITED

- Aung, M. M., & Chang, Y. S. (2014). Traceability in a food supply chain: Safety and quality perspectives. In *Food Control* (Vol. 39, Issue 1, pp. 172–184). Elsevier BV.
- Bosona, T., & Gebresenbet, G. (2013). Food traceability as an integral part of logistics management in food and agricultural supply chain. In *Food Control* (Vol. 33, Issue 1, pp. 32–48). Elsevier.
- Bowden, R., Davies, R. W., Heger, A., Pagnamenta, A. T., de Cesare, M., Oikkonen, L. E., Parkes, D., Freeman, C., Dhalla, F., Patel, S. Y., Popitsch, N., Ip, C. L. C., Roberts, H. E., Salatino, S., Lockstone, H., Lunter, G., Taylor, J. C., Buck, D., Simpson, M. A., & Donnelly, P. (2019). Sequencing of human genomes with nanopore technology. *Nature Communications* 2019 10:1, 10(1), 1–9.
- Clamer, M., Höfler, L., Mikhailova, E., Viero, G., & Bayley, H. (2014). Detection of 3'-end RNA uridylation with a protein nanopore. *ACS Nano*, 8(2), 1364–1374.
- Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., & Davies, R. M. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, 10(2), 1–4.
- Davies, R. W., Kucka, M., Su, D., Shi, S., Flanagan, M., Cunniff, C. M., Chan, Y. F., & Myers, S. (2021). Rapid genotype imputation from sequence with reference panels. *Nature Genetics*, 53(7), 1104.
- De La Cerda, G. Y., Landis, J. B., Eifler, E., Hernandez, A. I., Li, F. W., Zhang, J., Tribble, C. M., Karimi, N., Chan, P., Givnish, T., Strickler, S. R., & Specht, C. D. (2023). Balancing read length and sequencing depth: Optimizing Nanopore long-read sequencing for monocots with an emphasis on the Liliales. *Applications in Plant Sciences*, 11(3).
- Delahaye, C., & Nicolas, J. (2021). Sequencing DNA with nanopores: Troubles and biases. *PLoS ONE*, 16(10).
- Destefanis, G., Barge, M. T., Brugiapaglia, A., & Tassone, S. (2000). The use of principal component analysis (PCA) to characterize beef. *Meat Science*, 56(3), 255–259.
- Gupta, A. K., & Gupta, U. D. (2014). Next Generation Sequencing and Its Applications. *Animal Biotechnology: Models in Discovery and Translation*, 345–367.
- Jain, M., Olsen, H. E., Paten, B., & Akeson, M. (2016). The Oxford Nanopore MinION: Delivery of nanopore sequencing to the genomics community. *Genome Biology*, 17(1).
- King, J., Pohlmann, A., Dziadek, K., Beer, M., & Wernike, K. (2021). Cattle connection: molecular epidemiology of BVDV outbreaks via rapid nanopore whole-genome sequencing of clinical samples. *BMC Veterinary Research*, 17(1).

- Lamb, H. J., Hayes, B. J., Nguyen, L. T., & Ross, E. M. (2020). The Future of Livestock Management: A Review of Real-Time Portable Sequencing Applied to Livestock. *Genes*, *11*(12), 1478.
- Lamb, H. J., Hayes, B. J., Randhawa, I. A. S., Nguyen, L. T., & Ross, E. M. (2021). Genomic prediction using low-coverage portable Nanopore sequencing. *PLOS ONE*, *16*(12), e0261274.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, *25*(16), 2078–2079.
- Lu, H., Giordano, F., & Ning, Z. (2016). Oxford Nanopore MinION Sequencing and Genome Assembly. In *Genomics, Proteomics and Bioinformatics* (Vol. 14, Issue 5).
- McNaughton, A. L., Roberts, H. E., Bonsall, D., de Cesare, M., Mokaya, J., Lumley, S. F., Golubchik, T., Piazza, P., Martin, J. B., de Lara, C., Brown, A., Ansari, M. A., Bowden, R., Barnes, E., & Matthews, P. C. (2019). Illumina and Nanopore methods for whole genome sequencing of hepatitis B virus (HBV). *Scientific Reports*, *9*(1).
- McVean, G. (2009). A genealogical interpretation of principal components analysis. *PLoS Genetics*, *5*(10).
- Mini Centrifuge, 6,000 RPM, White / Southern Labware.* (n.d.). Retrieved September 7, 2023, from <https://www.southernlabware.com/mini-centrifuge>
- Nguyen, T. V., Vander Jagt, C. J., Wang, J., Daetwyler, H. D., Xiang, R., Goddard, M. E., Nguyen, L. T., Ross, E. M., Hayes, B. J., Chamberlain, A. J., & MacLeod, I. M. (2023). In it for the long run: perspectives on exploiting long-read sequencing in livestock for population scale studies of structural variants. *Genetics Selection Evolution* *2023* *55*:1, *55*(1), 1–15.
- O'Brien, A. C., Purfield, D. C., Judge, M. M., Long, C., Fair, S., & Berry, D. P. (2020). Population structure and breed composition prediction in a multi-breed sheep population using genome-wide single nucleotide polymorphism genotypes. *Animal*, *14*(3), 464–474.
- One-Block Digital Dry Bath 115V / Southern Labware.* (n.d.). Retrieved September 7, 2023, from <https://www.southernlabware.com/one-block-digital-dry-bath-115v>
- Oxford Nanopore Technology. (n.d.). *Flongle*. <https://Nanoporetech.Com/Products/Flongle>.
- Patterson, N., Price, A. L., & Reich, D. (2006). Population Structure and Eigen analysis. *PLoS Genetics*, *2*(12), 2074–2093.

- Porras-Hurtado, L., Ruiz, Y., Santos, C., Phillips, C., Carracedo, Á., & Lareu, M. V. (2013). An overview of STRUCTURE: Applications, parameter settings, and supporting software. *Frontiers in Genetics*, 4(MAY), 48396.
- QIAamp DNA Blood Kits*. (n.d.). Retrieved January 9, 2022, from <https://www.qiagen.com/us/products/discovery-and-translational-research/dna-rna-purification/dna-purification/genomic-dna/qiaamp-dna-blood-kits/>
- Rang, F. J., Kloosterman, W. P., & de Ridder, J. (2018). From squiggle to basepair: Computational approaches for improving nanopore sequencing read accuracy. In *Genome Biology* (Vol. 19, Issue 1, pp. 1–11). BioMed Central Ltd.
- Schroeder, T. C., & Tonsor, G. T. (2012). International cattle ID and traceability: Competitive implications for the US. *Food Policy*, 37(1), 31–40.
- Stefan, C. P., Hall, A. T., Graham, A. S., & Minogue, T. D. (2022). Comparison of Illumina and Oxford Nanopore Sequencing Technologies for Pathogen Detection from Clinical Matrices Using Molecular Inversion Probes. *Journal of Molecular Diagnostics*, 24(4), 395–405.
- Taylor, T. L., Volkening, J. D., DeJesus, E., Simmons, M., Dimitrov, K. M., Tillman, G. E., Suarez, D. L., & Afonso, C. L. (2019). Rapid, multiplexed, whole genome and plasmid sequencing of foodborne pathogens using long-read nanopore technology. *Scientific Reports*, 9(1).
- Tyler, A. D., Mataseje, L., Urfano, C. J., Schmidt, L., Antonation, K. S., Mulvey, M. R., & Corbett, C. R. (2018). Evaluation of Oxford Nanopore's MinION Sequencing Device for Microbial Whole Genome Sequencing Applications. *Scientific Reports*, 8(1), 10931.
- VanRaden, P. M. (2008). Efficient Methods to Compute Genomic Predictions. *Journal of Dairy Science*, 91(11), 4414–4423.
- Whiteford, N., Haslam, N., Weber, G., Prü Gel-Bennett, A., Essex, J. W., Roach, P. L., Bradley, M., & Neylon, C. (n.d.). *An analysis of the feasibility of short read sequencing*.
- Wick, R. R., Judd, L. M., & Holt, K. E. (2019). Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biology*, 20(1), 1–10.
- Zhou, A., Lin, T., & Xing, J. (2019). Evaluating nanopore sequencing data processing pipelines for structural variation identification. *Genome Biology*, 20(1), 237.

CHAPTER 5: Genetic Characterization of the Akaushi Breed in the United States

Hanna Ostrovski and Cedric Gondro

Abstract

Understanding the relationship between cattle breeds is important when considering crossbreeding in an industry setting or knowing the feasibility of a multi-breed genomic prediction. This is especially important when considering Japanese Wagyu breeds, which are not well characterized in a population analysis. Today, in the United States (US), Akaushi (also known as Japanese Brown or Red) and Japanese Black cattle are rising to prevalence due to their high-quality beef products. The American Akaushi population structure has never been explored since the arrival of these animals in the 1990's. This study aims to better understand the Akaushi population by analyzing such population parameters as inbreeding, heterozygosity, opposing homozygotes, linkage disequilibrium (LD), and effective population size. Comparison with other Asian and European cattle breeds using principal component analysis was performed to identify the relatedness of the Akaushi breed in the US (United States) to other cattle breeds. Our study found that out of the 43 animals studied, the genetic variation within the Akaushi population was comparable to other Asian breeds but showed large amount of LD within the genome. Analyzing relatedness to other cattle breeds observed that the Akaushi are most closely related to the Korean Hanwoo and Chosun. This study outlines that within the American Akaushi population, large LD blocks are apparent and that this group of cattle are most closely related to Hanwoo cattle along with other Japanese Breeds. Further exploration into the American Japanese breeds is needed to fully understand the selection pressures that have occurred since the 1990's.

Introduction

The first cattle domestication event of *Bos taurus taurus* occurred around 8,500 B.C. in the Fertile Crescent (Bruford et al., 2003). Soon after this first domestication, a separate domestication event of *Bos taurus indicus* occurred in the Indus Valley (Loftus et al., 1994) which led to the two specific subspecies that are now present today. Pinpointing the dissemination of these two species throughout the world is important to understand the genetic

makeup of our current cattle breeds, specifically those breeds of high-quality, for this study, high-marbling animals originating from the Asian continent. The evidence of *Bos taurus taurus* in the Asian continent placed their arrival around 2,500-1,900 B.C. (Cai et al., 2014) and these domesticated cattle then spread to the outreaches of Japan from the Asian continent via Korea (Sasazaki et al., 2006b).

All modern cattle in Japan can be genetically classified as a cross between an imported European cattle breed, such as Simmental, and the imported *Bos taurus taurus* from Korea (Gotoh et al., 2018). The animals of focus for this study are a group of American Akaushi cattle that have been reproducing in the United States for ~30 years. The Akaushi population in America is known to have originated from the Japanese Brown (also sometimes known as Red) Wagyu cattle that were imported into the United States from Japan in 1994. Before this small nucleus, there existed two Wagyu bulls in the United States which were brought in 1976 and had already been used to cross with other breeds. This population has grown from around a dozen animals to a large herd of Fullblood and Purebred American Akaushi and Wagyu in the United States (Beeman, 2019. Interview).

The Akaushi breed is known to be classified as a Japanese breed, but this study aims to identify relatedness between American Akaushi and other Asian breeds through genomic analysis using population parameters. A wide variety of cattle breeds were analyzed and compared to the Akaushi to understand the breed in relation to cattle used around the world. The methods used to define the American Akaushi population included calculating the level of inbreeding, the observed heterozygosity, the number of opposing homozygotes, the measure of linkage disequilibrium (LD), the effective population size (N_e), and the genetic relationship between the Akaushi population and other traditionally Eastern and Western cattle breeds. Identification of origin of this breed could lead to a better understanding of breed composition to make better management decisions based on evaluations of other comparable breeds.

Materials and Methods

Population measures were calculated to better understand the Akaushi population in the United States. A principal component analysis (PCA) was used firstly to understand the breeds that were most like the Akaushi based on genotype. After PCA, the breeds that most resembled

the Akaushi genetically were included with the Akaushi in population parameter analysis. The measures used in this study included the level of inbreeding, the observed heterozygosity, the number of opposing homozygotes, the measure of linkage disequilibrium, and the effective population size.

Data

Original animal counts for each breed can be found in Table 5.1. The sample size from each breed ranged from 12 to ~1500 animals.

Breed	Number of Animals Analyzed
Angus	112
Charolais	12
Hereford	79
Holstein	978
Limousin	14
Murray Grey	36
Shorthorn	77
Jersey	688
Hanwoo	1492
Wagyu	119
Yeonbyun	63
Akaushi	43
Akaushi Crossbred	56
Brindle	20
Jeju Black	20
Chosun	20
Hanwoo Population 2	239

Table 5.1 Original animal counts per breed considered in population analysis.

Imputation

All genotyped animals were imputed up to a subset 700k chip that contained 507,261 SNP. Subsetting the 700k SNP panel was necessary to remove those SNP calls that purely

contained NAs across most breeds. The ~507k SNP chip was chosen because it represents only a maximum of 10% of missing calls within the whole 700k genotyped animals.

The imputation from 50k to 507k SNP chip was done using the phasing program EAGLE (Loh et al., 2016) and the imputation program MINIMAC (Das et al., 2016). The phasing and imputation protocol that was followed can be found in the paper published by Al-Mamun et. Al. (Al-Mamun et al., 2017).

Genetic relationship to other breeds

To understand the genetic background of the Akaushi cattle and how related it was to other Asian cattle breeds, principal components were calculated using the Akaushi genetic data as well as data from Angus, Charolais, Hanwoo, Hereford, Holstein, Jersey, Limousin, Murray Grey, Shorthorn, Wagyu, Yeonbyun, Akaushi crossbreed, Brindle, Chosun, and Jeju Black cattle.

All animals from all breeds mentioned above were imputed up to the subset 700k chip which contained 507k SNP. After imputation, a G matrix was constructed with all animals and was centered around the mean of the matrix. A singular value decomposition was employed to calculate the principal components of the centered G. The first two principal components were plot against each other to understand the genetic variability between all breeds.

After analysis using PCA for all breeds, the East Asian cattle populations were chosen to compare population structure to the American Akaushi breed. These breeds include the Hanwoo, Chosun, Jeju Black, Yeonbyun, an Akaushi cross, and Wagyu. These breeds would be the most interesting to analyze, as the Akaushi in America has not been thoroughly studied while other similar cattle breeds have existed for many years and studied extensively.

Similarities in population parameters between breeds would result in the conclusion of origins for the Akaushi breed. This was done using common population parameters such as measure of inbreeding, level of heterozygosity, analysis of opposing homozygotes, level of linkage disequilibrium (LD) and estimation of the effective population size.

Level of Inbreeding

The level of inbreeding in a population can give insight into the population structure itself. The inbreeding coefficient explains how related the parents of an animal are. The increase

in inbreeding can cause inbreeding depression which can cause fitness problems in the population (Falconer, 1960).

To calculate inbreeding in the population, a G matrix was constructed by use of the R package “snpReady” (Granato & Fritsche-Neto, 2018). The VanRaden method (VanRaden, 2008) for calculating a G matrix was implemented in this study and is as follows:

$$G = \frac{ZZ'}{2 \sum p_i(1 - p_i)}$$

Where G is the genotype relationship matrix, Z is a design matrix containing centered allele effects and p_i are allele frequencies.

This matrix is the relationship covariance matrix using the genotypes from the SNP chip. The diagonal of this matrix contains the inbreeding coefficient for each individual.

The observed heterozygosity

Heterozygosity can indicate the amount of genetic variability found in a population. The heterozygosity per animal was estimated by:

$$HE = f_{AB}$$

Where f_{AB} is the frequency of the heterozygous loci and HE is the measure of heterozygosity as a count. Heterozygosity was measured per animal per breed.

The number of opposing homozygotes

The number of opposing homozygotes between animals checks for mendelian inconsistencies and aids in understanding relationships in the genetic information. If animals have opposing homozygotes at a certain locus, then they cannot be related. On the other hand, if animals share an allele at a locus, then there is a chance that they might be related (Calus et al., 2011).

Opposing homozygotes were calculated using methods published by Calus, Mulder, & Bastiaansen, 2011.

$$Op = M_0 * M_2'$$

Where Op is the matrix of opposing homozygotes, M_0 is a square matrix of 0 and 1 which correspond to those SNP which are coded as “0”, and M_2' is another square matrix of 0 and 1 corresponding to those SNP which are coded “2”.

Linkage Disequilibrium

Linkage disequilibrium (LD) was measured by rate of LD decay to account for the physical distance between SNPs (Single Nucleotide Polymorphisms) that are included in the dataset. The decay of LD over time is important to understand how linked SNPs are in the population which is an important parameter to understand the history of populations and for association studies (Vos et al., 2017).

To measure LD decay, the average LD was estimated between a certain distance between the SNP. This was done using the “snpStats” package in R, which estimates the r^2 value of LD. The LD was then plot on distances on the genome of 1000 base pairs to 4000000 base pairs.

Effective Population Size

The effective population size parameter explains the rate of change in a population due to e.g., genetic drift, selection, bottlenecks and other evolutionary factors. Effective population size explains how many animals would be needed in an idealized population to create the same amount of genetic variation (Charlesworth, 2009). The larger the effective population size, the more variable the population was and vice versa.

The measure of Ne_T over T generations ago, was calculated by sampling 30 random animals from each breed except for the Brindle, Chosun, and Jeju Black breeds due to small number of total animals in the dataset. After correcting for population size, r^2 was obtained to get the pairwise LD across the genome. The r^2 was measured between each SNP in each animal. The Ne was then calculated by using the average r^2 over a specific distance based on the measure of Ne_T at a certain time in the past. The equation used to calculate Ne_T was from de Roos, 2008:

$$Ne_T = (1/4c)(1/(\bar{r}^2 - 1))$$

where c was the marker distance in morgans related to the population size T generations ago (de Roos et al., 2008).

Results

The principal component analysis plot between the Akaushi population and other previously mentioned cattle breeds can be seen in Figure 5.1.

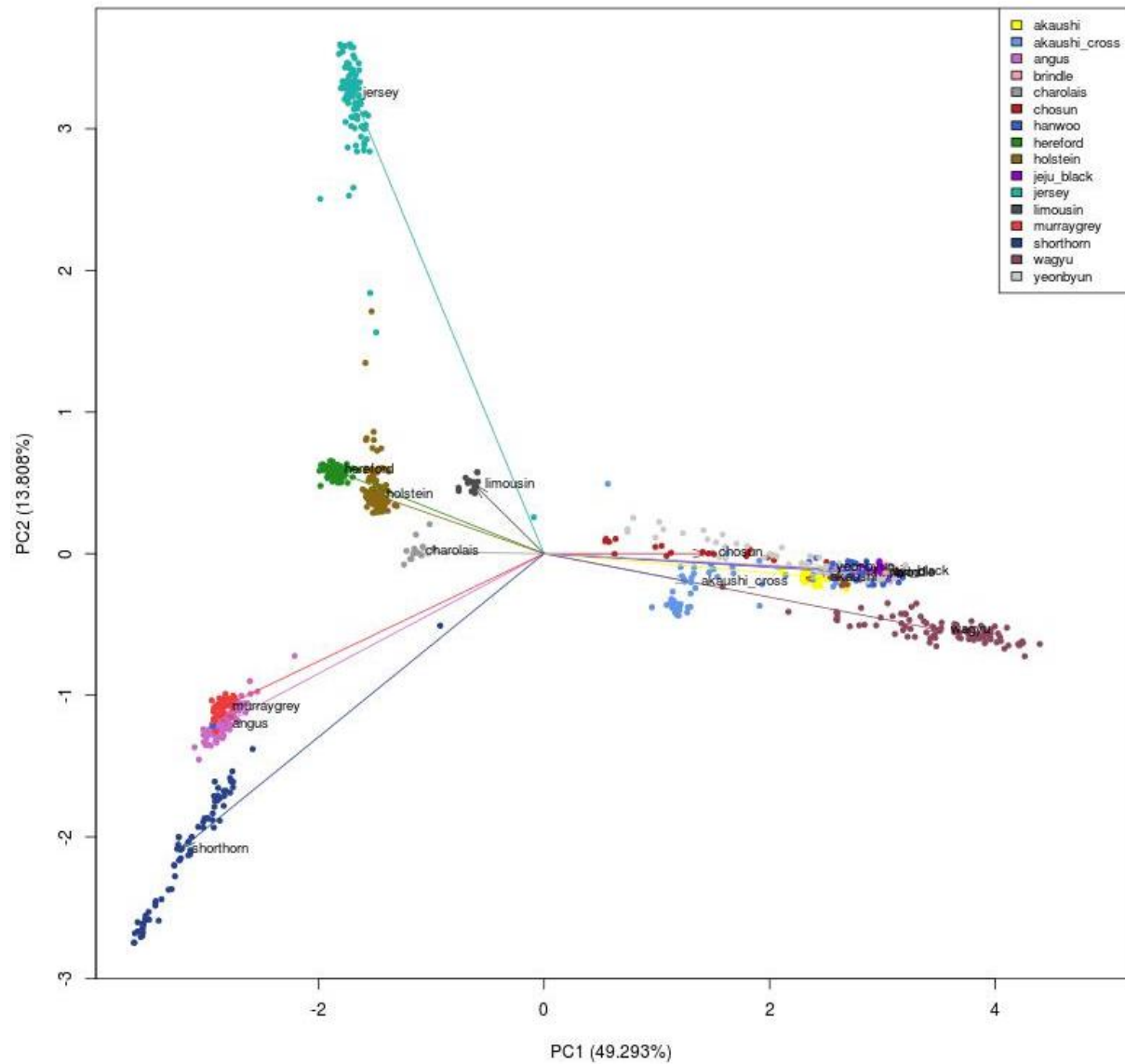


Figure 5.1 Principal component analysis of all breeds considered in the Akaushi population analysis.

This plot considers all the European and Asian breeds that were considered in this study. The principal components which account for variation between each of the breeds are on either the x- or y- axes. The largest variation was found between the Asian and European breeds, around 50% of the total variation. There was also a separation of animals within the continental breeds as well, with the most variation on the y-axis occurring between the Shorthorn and Jersey cattle at about 10%.

The Akaushi population within this Figure show little to no separation from many of the other Asian cattle breeds. A distinct separation can be seen within the Asian cattle breeds, with one cluster containing Akaushi cattle and the other, Wagyu cattle. This bolsters the claim that these cattle are genetically more like non-Japanese cattle.

The inbreeding within the Akaushi (Fig. 5.2) wavers around 0 (when using a centered G) with one outlier around 0.4. The other breeds analyzed also showed the same trend, with some outliers in the Hanwoo population (Fig. 5.3) as well.

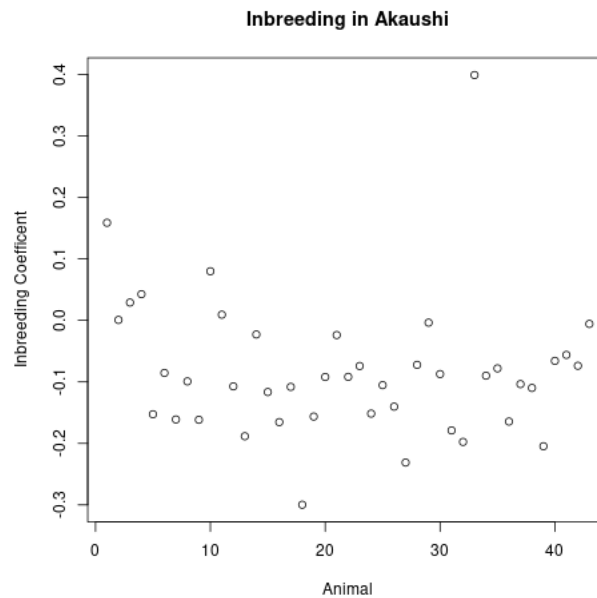


Figure 5.2 Inbreeding coefficient of the Akaushi population from the genotypic relationship matrix.

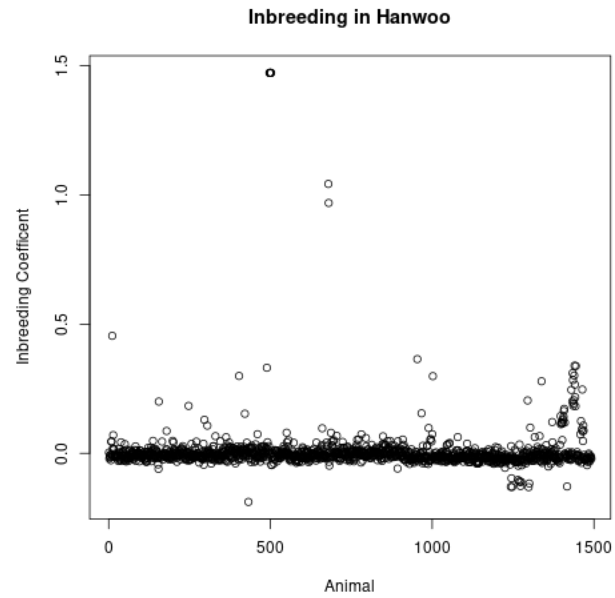


Figure 5.3 Inbreeding coefficient of the Hanwoo population from the genotypic relationship matrix

The level of inbreeding within these breeds could vary depending on the animals that are contained within the sample. Larger sample sizes of those breeds that were sampled with a small number of animals within this study could shed light on the greater population inbreeding trends.

The observed heterozygosity within all breeds wavers around 0.2-0.5. This is also seen in the Akaushi (Fig. 5.4), which can show that the Akaushi has some variation within the population sampled.

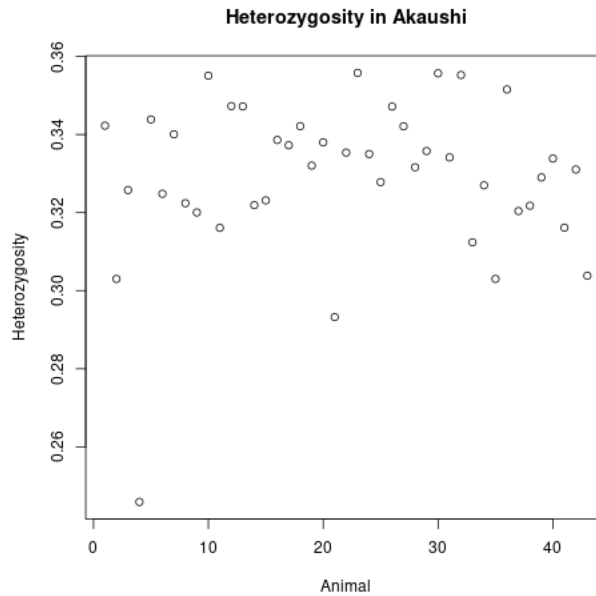


Figure 5.4 Observed heterozygosity in the Akaushi population.

As in most populations, variation in each breed does occur, where some outliers exist either extremely negative or positive. The Akaushi population shows one animal has a lower observed heterozygosity than the average, which can lead to decreased genetic variation.

The opposing homozygote estimation is employed to find animals that have similar genetics within a population as to identify possible related pairs, such as siblings or a parent-offspring. This analysis may bring those pairs to light and could solve pedigree discrepancies. Within the Akaushi (Fig. 5.5), there is a definitive observation of genetic relatedness between a few pairs of animals. Comparatively to other breeds analyzed, the only other distinct separation can be found in the Yeonbyun population (Fig. 5.6).

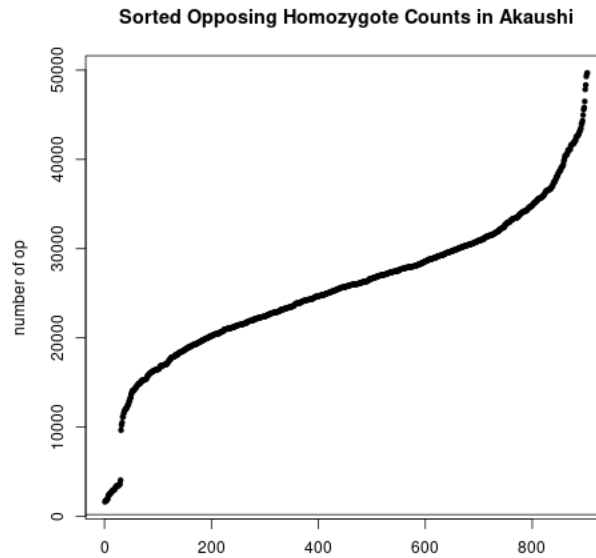


Figure 5.5 Opposing homozygotes in the Akaushi population.

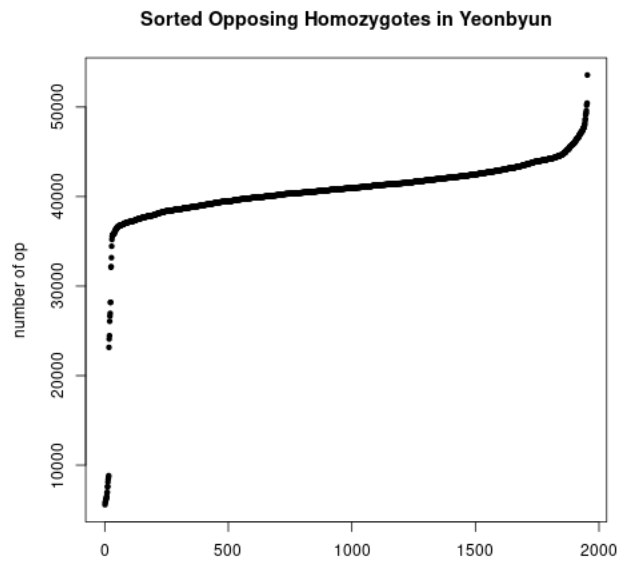


Figure 5.6 Opposing homozygotes in the Yeonbyun population.

LD decay is the measure of LD (on the x-axis) over the distance between SNPs in base pairs (y-axis). The LD structure in the Akaushi population (Fig. 5.7) falls into the same trend as other breeds with a tight clustering of points as the distance is closer together such as the Hanwoo (Fig. 5.8).

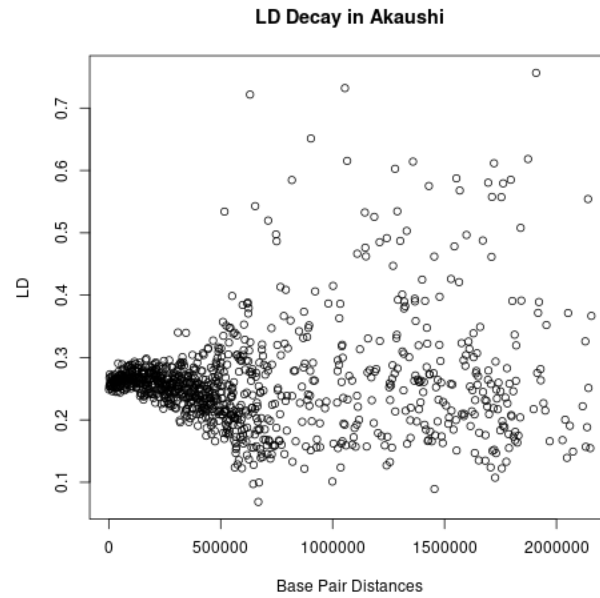


Figure 5.7 Linkage disequilibrium decay in the Akaushi population.

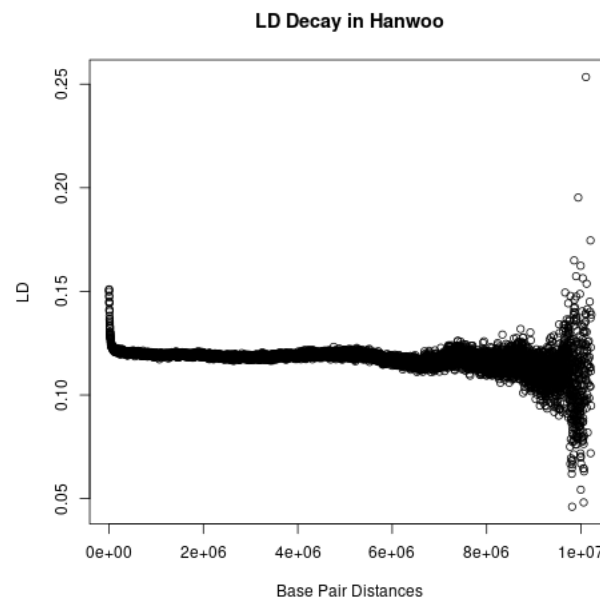


Figure 5.8 Linkage disequilibrium decay in the Hanwoo population.

As the distance between the SNPs grows, the more sporadic the LD becomes. The one notable difference between the Akaushi population and other populations is that LD measure is large. The beginning cluster starts around 0.3 and many LD points land above 0.7, while the highest values from other breeds only reach just above 0.6.

The effective population size at 1, 5, 10 and 20 generations ago of each breed can be found in Table 5.2 and the N_e across multiple generations in all breeds can be visualized in Figure 5.9.

Breeds	Previous Number of Generations			
	1	5	10	20
Angus	3.50	10.87	13.92	17.45
Hanwoo	7.47	22.32	25.05	28.44
Hereford	3.24	9.98	12.63	16.47
Holstein	5.07	15.08	18.16	21.41
Jersey	2.77	10.44	14.00	17.72
Murray Grey	2.29	8.54	11.62	15.14
Shorthorn	2.15	9.17	12.76	16.65
Wagyu	4.29	16.16	21.59	27.58
Yeonbyun	11.18	24.33	26.66	29.59
Akaushi	2.88	9.53	13.57	17.65
Akaushi Cross	6.41	18.13	22.49	26.59
Brindle	5.92	16.69	21.14	25.73
Chosun	6.36	20.76	25.23	28.78
Hanwoo Pop2	7.34	18.97	22.70	26.70
Jeju Black	6.27	19.08	24.07	28.43

Table 5.2 Effective population size at 1, 5, 10 and 20 generations ago of each breed considered. Most inbred breeds show very low N_e across generations.

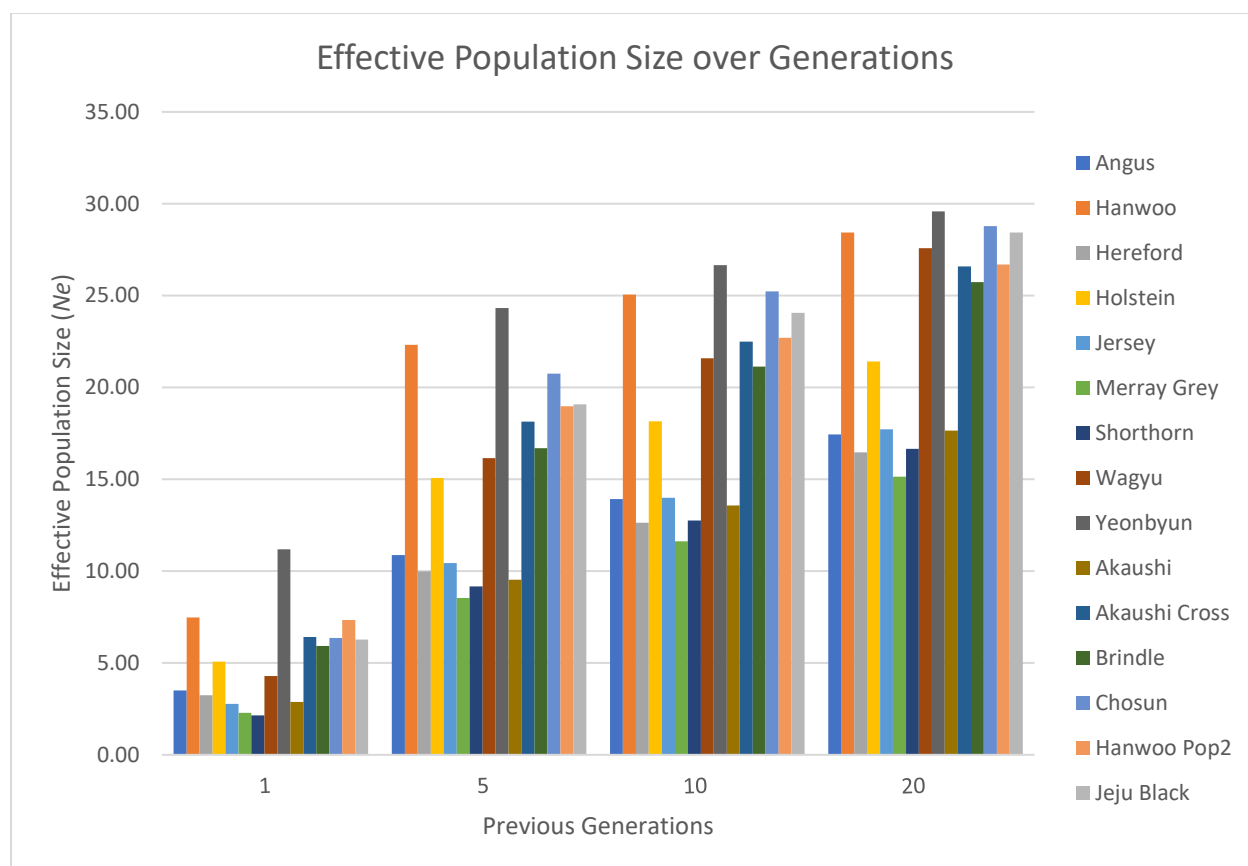


Figure 5.9 Effective population size per 1, 5, 10, and 20 generations ago per breed considered.

The Akaushi sample was found to have an N_e of 14 10 generations ago, which is one of the lower sizes compared to the other breeds analyzed. At 10 generations, the Yeonbyun breed was seen to have the highest N_e and the Shorthorn and Jersey breeds were seen to have the lowest with values ranging from 11 to 27. Low N_e values are common within breeds which are genetically similar and may have smaller population sizes. The Akaushi sample comes from an exceedingly small population that does not have a large amount of variation due to the limited number of animals, which leads to small N_e estimations.

Conclusions

The obvious separation in Figure 5.1 of the Akaushi (and other Asian breeds) from European breeds was to be expected. The relationship of the Akaushi to the other Asian breeds was confirming of the breed background, as it was genetically most like the Hanwoo. This analysis confirms the relatedness to the Hanwoo, as the Akaushi and Hanwoo seem to have a

large overlap in the PC analysis. This was also to be expected as breed lineage of the Akaushi population, as well as many Japanese breeds, was said to have been imported from Korea (Sasazaki et al., 2006a). Previous studies have also outlined the similarities between the Japanese breeds and Hanwoo (Kawaguchi et al., 2022) and showed overlap between Japanese cattle, which were raised in Japan, and Hanwoo through PCA analysis. This study upholds these mirrored findings between the American Akaushi and Hanwoo populations and helps to visualize the association of underlying population structure within the PC analysis.

The results of analysis of population parameters within all Asian breeds shows that the Akaushi population was very closely related, but not unlike other Asian breeds. The inbreeding within the Akaushi showed that some animals in the population do have a large inbreeding coefficient, which could be due to the small amount of Akaushi currently present in the United States. The similarities within the Akaushi population can also be seen in the measure of opposing homozygotes which show a very explicit separation of unrelated to related pairs. This separation shows that some animals are very closely related and includes a parent-offspring pair.

The measure of LD decay specifically shows that the Akaushi population has a high amount of LD, even in respect to other Asian breeds. With a high LD in this population, we can assume large chunks of the genome are linked together, which gives way to inheriting these large blocks (or haplotype blocks) together (Slatkin, 2008). This contributes to the measure of homozygosity in the population. This increased homozygosity can lead to recessive homozygous traits becoming common within the population. The sample of Akaushi cattle used here is shown to be very related within itself, but there was still some genetic variation found. The heterozygosity per animal shows some variation within this group of animals, even if it was still quite low. Identifying variation in this population and capitalizing on such genetic variation can curb the possible negative effects of breeding animals that are genetically similar (Curik, Ferenčaković and Sölkner, 2014), most common of these effects being inbreeding depression.

The effective population size of the Akaushi compared to other breeds show that this population was similar because of the estimated small N_e . This was to be expected because of the background of the Akaushi breed in the United States as well as the low size of the number of animals analyzed. Other studies in Korean Hanwoo cattle also show a decline in the effective population size in recent years (Li & Kim, 2015).

The analysis of the American Akaushi breed in comparison to Eastern and Western breeds shows that the relatedness between the Akaushi and the Hanwoo is strong, which is to be expected due to the breed origin background. When American Akaushi population parameters were compared to the population parameters of other Asian breeds, many similarities arose, such as the similar amount of heterozygosity and inbreeding. The main difference when comparing the Akaushi with other Asian breeds was the measure of LD decay within the populations. The Akaushi showed to have more LD within the genome, leading way to assume that the Akaushi population has not varied much through generations and becoming more genetically related. These population parameters have shed light on the American Akaushi population, most importantly, this population is most like Korean breeds and may lack genetic diversity due to the breeding structure.

LITERATURE CITED

- Al-Mamun, H. A., Bernardes, P. A., Lim, D., Park, B., & Gondro, C. (2017). A guide to imputation of low density single nucleotide polymorphism data up to sequence level. *Journal of Animal Breeding and Genomics*.
- Bruford, M. W., Bradley, D. G., & Luikart, G. (2003). DNA markers reveal the complexity of livestock domestication. *Nature Reviews Genetics*, 4(11), 900–910.
- Cai, D., Sun, Y., Tang, Z., Hu, S., Li, W., Zhao, X., Xiang, H., & Zhou, H. (2014). The origins of Chinese domestic cattle as revealed by ancient DNA analysis. *Journal of Archaeological Science*, 41, 423–434.
- Calus, M. P., Mulder, H. A., & Bastiaansen, J. W. (2011). Identification of Mendelian inconsistencies between SNP and pedigree information of sibs. *Genetics Selection Evolution*, 43(1), 34.
- Charlesworth, B. (2009). *Effective population size and patterns of molecular evolution and variation*.
- Curik, I., Ferencaković, M., & Sölkner, J. (2014). Inbreeding and runs of homozygosity: A possible solution to an old problem. *Livestock Science*.
- Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A. E., Kwong, A., Vrieze, S. I., Chew, E. Y., Levy, S., McGue, M., Schlessinger, D., Stambolian, D., Loh, P.-R., Iacono, W. G., Swaroop, A., Scott, L. J., Cucca, F., Kronenberg, F., Boehnke, M., ... Fuchsberger, C. (2016). Next-generation genotype imputation service and methods. *Nature Genetics*, 48(10), 1284–1287.
- de Roos, A. P. W., Hayes, B. J., Spelman, R. J., & Goddard, M. E. (2008). Linkage Disequilibrium and Persistence of Phase in Holstein–Friesian, Jersey and Angus Cattle. *Genetics*, 179(3), 1503–1512.
- Falconer, D. (1960). *Introduction to quantitative genetics*.
- Gotoh, T., Nishimura, T., Kuchida, K., & Mannen, H. (2018). The Japanese Wagyu beef industry: current situation and future prospects - A review. *Asian-Australasian Journal of Animal Sciences*, 31(7), 933–950.
- Granato, I., & Fritsche-Neto, R. (2018). *snpReady: Preparing Genotypic Datasets in Order to Run Genomic Analysis. R package version 0.9.6*.
- Kawaguchi, F., Nakamura, M., Kobayashi, E., Yonezawa, T., Sasazaki, S., & Mannen, H. (2022). Comprehensive assessment of genetic diversity, structure, and relationship in four Japanese cattle breeds by Illumina 50 K SNP array analysis. *Animal Science Journal*, 93(1), e13770.

- Li, Y., & Kim, J. J. (2015). Effective population size and signatures of selection using bovine 50K SNP chips in Korean native cattle (Hanwoo). *Evolutionary Bioinformatics*.
- Loftus, R. T., MacHugh, D. E., Bradley, D. G., Sharp, P. M., & Cunningham, P. (1994). Evidence for two independent domestications of cattle. *Proceedings of the National Academy of Sciences of the United States of America*, 91(7), 2757–2761.
- Loh, P.-R., Danecek, P., Palamara, P. F., Fuchsberger, C., A Reshef, Y., K Finucane, H., Schoenherr, S., Forer, L., McCarthy, S., Abecasis, G. R., Durbin, R., & L Price, A. (2016). Reference-based phasing using the Haplotype Reference Consortium panel. *Nature Genetics*, 48(11), 1443–1448.
- Sasazaki, S., Odahara, S., Hiura, C., Mukai, F., & Mannen, H. (2006). Mitochondrial DNA Variation and Genetic Relationships in Japanese and Korean Cattle. *Asian-Australasian Journal of Animal Sciences*, 19(10), 1394–1398.
- Slatkin, M. (2008). Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. *Nature Reviews. Genetics*, 9(6), 477–485.
- VanRaden, P. M. (2008). Efficient Methods to Compute Genomic Predictions. *Journal of Dairy Science*, 91(11), 4414–4423.
- Vos, P. G., Paulo, M. J., Voorrips, R. E., Visser, R. G. F., van Eck, H. J., & van Eeuwijk, F. A. (2017). Evaluation of LD decay and various LD-decay estimators in simulated and SNP-array data of tetraploid potato. *Theoretical and Applied Genetics*, 130(1), 123–135.

CHAPTER 6: Estimation of Within and Across Breed Prediction Accuracies in the Wagyu Population in the United States and the Korean Hanwoo

Hanna Ostrovski, Daniela Lourenco, Andra Nelson, Cedric Gondro

Abstract

Understanding the population structure among the various Wagyu subtypes outside of Japan, specifically the Red and Black varieties, is essential due to their high value and unique marbling characteristics. The comparison between these groups aims to uncover the extent of genomic relatedness which can directly impact the accuracy of genomic prediction of breeding values. Further investigation into the relationship between the Korean Hanwoo and US Wagyu will be explored due to historical accounts that suggest that Japanese animals may have originated from Korea. Relationship status between all breeds was done through Principal Component Analysis (PCA) and through estimating genomic prediction accuracies between breeds. Further investigation into the shared genetic elements between all cattle breeds was done through genome-wide association (GWA). Genomic prediction accuracies obtained through training and testing groups utilizing GBLUP using weaning weight as a phenotype for growth showed low accuracy between Red and Black Wagyu, around 0.10. A RedBlack population of crossed Red and Black Wagyu and the Hanwoo were able to predict other breeds with moderate accuracy, ranging from 0.23 to 0.27. To address unbalanced breed group sizes (~150 Black Wagyu versus ~5000 Red Wagyu), the total population was divided into 10 balanced groups based on animal relatedness via the first principal component. Testing prediction accuracies within these splits revealed higher accuracies between closely related splits, up to 0.45. Notably, the split involving Red Wagyu (1st PC split) and Korean Hanwoo (10th PC split) demonstrated the highest accuracy, reinforcing the close genetic relationship between these breeds. The GWA identified new genomic regions on chromosomes 6, 10, and 14 associated with growth. These findings signify the early stages of unraveling the intricate relationships between different subsets of US Wagyu. Utilizing this knowledge in estimating breeding values within Wagyu will impact breeding practices, enhancing the selection of desirable traits.

Introduction

Genomic prediction between breeds has not yielded very promising results due to the nature of breed structure as they have different genomic architectures. This is especially important in the Wagyu breed in the United States, as Red and Black Wagyu are registered in the same Association and combined in the genomic evaluation. Low accuracy of prediction may occur between the two groups due to differences in origin as both Red and Black American Wagyu originate from Japan but resided in different prefectures in Japan (Namikawa, 1980). Further investigation into the population structure between these two subtypes in the US has not been thoroughly explored, but initial papers suggest both groups, and the population, are highly inbred (Heffernan, 2022; Scraggs et al., 2014). This is to be expected, as this group of animals originated from a couple of bulls that were exported to the US in the 1970's and a larger herd of animals brought over in the early 1990s. No other outside genetics have been utilized, so current genomic variability relies on the initial structure of animals that were imported. The relationships between these animals from the original group of Wagyu imported to the US is now used as the base generation, as further genomic information of previous Japanese generations is not available for analysis on current US Wagyu. This leaves the American population with low genomic variability, as well as a tight genomic population due to low effective population size.

Genomic prediction within breed has been seen to result in high accuracy (Hayes et al., 2009; Karaman et al., 2016) due to the underlying genomic architecture of related animals between testing and training groups. This predictive power has given those in the cattle industry an edge on selection, as animals can be culled from programs before needing to be proved by progeny. The accuracy of prediction has been proven to degrade when predicting the performance of an animal from a training population that is not related to the animal of interest. This usually occurs in prediction between breeds that may not be tightly related. Low accuracy of prediction has been reported, with bumps in accuracy attributed to inclusion of crossbred animals (Lund et al., 2014; Misztal et al., 2022; Olson et al., 2012), different modeling methods (Khansefid et al., 2020) or utilization of whole-genome sequence (Nawaz et al., 2022; Raymond et al., 2018).

Utilization of whole genome sequence data can be advantageous when predicting between breeds or groups that are not closely related (Druet et al., 2013; Meuwissen & Goddard,

2010). This slight bump in accuracy can have rippling effects when considering how heavily breeding values are utilized in the cattle industry. Whole genome sequence can uncover more similar areas in the genome than a standard SNP chip, as more variants are available to create genomic relationships between animals. While increasing the amount of sequence data may show improvement in accuracies, other additions in the training group, inclusion of related animals, and utilization of other prediction models has also shown to increase accuracy (Meuwissen et al., 2021). The use of whole genome sequence should be explored as an avenue to increase accuracy between breeds when all other methods are exhausted, as the cost of obtaining WGS has decreased significantly over time.

Degradation of accuracy also happens per generation, as linkages between the testing and training populations can break down if the generations between the two grow farther apart (Castro Dias Cuyabano et al., 2019). This is due to linkage disequilibrium breaking down through generations through recombination. Differences in breed composition between the training and testing populations can decrease accuracy as well (van den Berg et al., 2020), especially if one breed is overrepresented in a training group. All these potential breaks in connection between the testing and training groups can result in lower prediction accuracies.

Yet, crossing between the Red and the Black Wagyu in the US population occurs and can link these two breed groups together. Across-breed prediction accuracy has been seen to increase when links between the two breeds are included, such as these crossbred animals in the US Wagyu population. Increases in accuracy have been shown to be directly related to an increase in profit (Thomasen et al., 2014), as more accurate breeding values can be assigned to breeding animals.

The selection and origins of many Asian breeds are similar, as they are high marbling with a selection pattern of being used as work animals before being selected for beef consumption (Gotoh et al., 2018; Motoyama et al., 2016). Further selection within breed to create modern breeds, such as the Japanese Black Wagyu or the Korean Hanwoo, would create larger genomic differences in modern cattle populations. This study aims at exploring the structure of the American Wagyu and Korean Hanwoo population today through genome-wide association and estimating the accuracy of prediction between the Red and Black populations, including the crossed Red/Black animals.

Materials and Methods

Data

Wagyu data from the American Wagyu Association was utilized. This data includes animals that are categorized Black Wagyu, Red Wagyu (also known as Brown or Akaushi animals), and Red/Black crosses. Hanwoo data originated from Korea and includes animals that are 100% Hanwoo. All animals have been genotyped with a bovine SNP chip, the Wagyu animals being genotyped with either the 50K Illumina Bovine chip or the 100k Illumina Bovine chip. All Hanwoo animals were initially genotyped at a 50K Illumina Bovine chip. All animals were phased and subset to 70k after quality control using eagle and impute5 software (Rubinacci et al., 2019; Loh et al., 2016). Each animal was then imputed up to whole genome sequence using the 1000 bulls as a reference population utilizing impute5 (Rubinacci et al., 2019) including Hanwoo and Wagyu animals that were previously genotyped at WGS. After quality control, the 70K chip contained 70, 343 SNP and animals with WGS contained 14,454,093 SNP.

The phenotype used to test the predictive ability between these populations was the phenotype of weaning weight. All animals considered within this study had to be genotyped and had to include a phenotype of weaning weight in pounds. To scale these two datasets, a linear model was employed, with the phenotype of all animals as the dependent variable and the variable of breed fit as the independent variable. Residuals from this model were then utilized as the adjusted phenotypes for genomic prediction.

Principal Component Analysis

Principal component analysis was firstly used to understand the population structure that exists between the Red Wagyu, Black Wagyu, and the Hanwoo. Obtaining eigen values and eigen vectors was done on the genetic relationship matrix of all animals, which was created via VanRaden, 2008:

$$G = \frac{Z'Z}{2 \sum p_i(1 - p_i)}$$

The package ‘eigen’ in R was utilized to obtain the eigenvectors and eigenvalues to plot each animal by the principal components that explained most of the variation within the dataset.

The eigen decomposition of the genomic relationship matrix uncovers population structures that are only available through genotype (McVean, 2009; Patterson et al., 2006). Eigenvectors are used to understand the grouping of animals per breed, the principal components, while eigenvalues explain the variance between the principal components (Karamizadeh et al., 2013). The top principal components that explained the most variation was then plot against each other to visualize how these animals grouped together.

Obtaining Accuracy of Prediction

Training and testing groups were utilized to understand the accuracy of prediction in certain prediction scenarios. The number of animals in each breed used for within and across breed prediction scenarios are in Table 6.1:

	<i>Breed Type</i>			
	Wagyu Red/Black			
	Wagyu Red	Wagyu Black	Cross	Hanwoo
<i>Number of Animals</i>	4766	147	598	3096

Table 6.1. Number of animals considered for each breed group; Wagyu and Hanwoo.

Prediction scenarios were done within each breed with an 80/20 split for training and testing groups. Across breed scenarios were run, with each breed combination considered for training and testing scenarios. Additional scenarios were considered due to unbalanced breed groups; principal components were employed to create balanced training and testing groups based off genotypic relationships. Creating these groups was done through sorting all animal's scores based off the first and second principal components then the animals were split up across PC1. See Fig. 6.1 for better visualization of splits using the PCA scores:

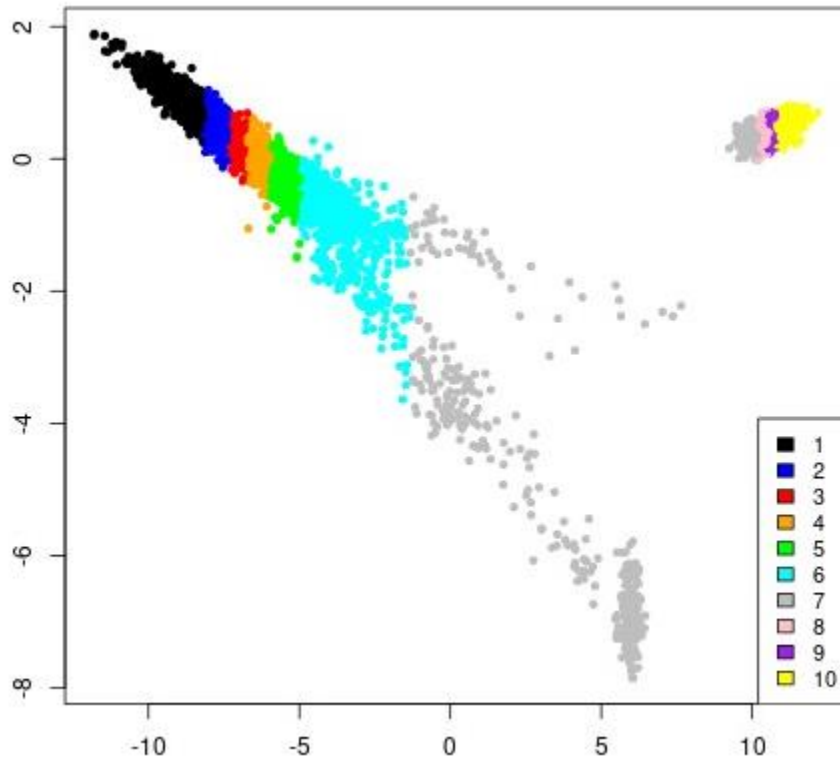


Figure 6.1. Principal Component splits over all animals using the first principal component scores as a baseline. Equal number of animals were represented in each split.

All scenarios were run at 70k density and WGS density. Both Hanwoo and Wagyu populations utilized pre-adjusted phenotypes utilizing farm, sex, and date of birth as fixed effects. Further centering methods between Hanwoo and Wagyu populations were done by fitting breed as a fixed effect to the adjusted phenotype. All predicted breeding values were done through GBLUP:

$$y = Xa + e$$

Where y is the adjusted phenotype, X is the design matrix connecting phenotypes to genetic values, a is the vector of random animal effects $a \sim N(0, G\sigma_a^2)$ and e is the vector of residual effects, where we assume $e \sim N(0, I\sigma_e^2)$. Accuracy of prediction was done through Pearson correlation of predicted EBV and adjusted phenotype within the training group.

Genome-Wide Association Study

A final step in understanding the significant areas on the genome was done through genome-wide association study. This was done by backsolving the SNP effects, then obtaining

the p-value of each SNP effect. This model estimates the genomic EBVs ($\hat{a} = Z\hat{g}$) through a snpGBLUP:

$$y = Mg + e$$

Where y is the adjusted phenotype, M matrix of $n \times m$, where n is the number of animals and m are the number of markers, g is the vector of random SNP effects evenly distributed, $g \sim N(0, I\sigma_g^2)$ and e is the vector of residual effects, where we assume $e \sim N(0, I\sigma_e^2)$.

Backsolving of SNP effects was done through this equation:

$$EFF = \frac{1}{d} * Z * G * \hat{a}$$

Where $d = 2 * \sum p_i(1 - p_i)$. To obtain p-values of each SNP effect we can use :

$$pval_i = 2 \left(1 - \Phi \left(\left| \frac{\widehat{EFF}_i}{sd(\widehat{EFF}_i)} \right| \right) \right)$$

Which identifies the significance of the SNP effect \widehat{EFF}_i as $pval_i$ through the density distribution (t-distribution). All p-values were corrected using the Bonferroni correction to account for multiple comparison, dividing by number of SNPs. Visual identification of significant SNP throughout a genome was by way of Manhattan plots. The highest p-values of significance that cross the p-value threshold of $-\log_{10}(5 \times 10^{-8})$ were considered significant. Locations of significant SNP were evaluated within ensembl database for candidate genes.

Results

A breakdown of population structure between all groups can be seen in the principal component analysis in Figure 6.2.

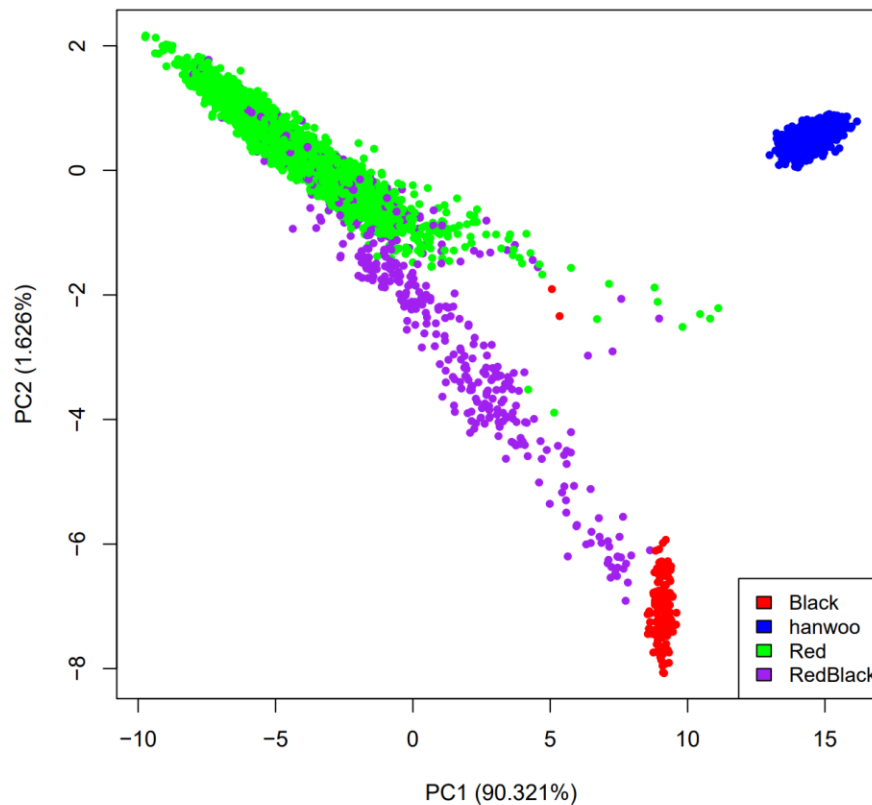


Figure 6.2. Principal Component Analysis of all animals considered. The largest PC1 accounted for 90% of the variation between the Hanwoo and Wagyu populations.

It is apparent that each breed type is singular enough to separate into groups, while the Red/Black animals are rightfully connecting the Red and Black Wagyu groups. There are some animals within that group that may be misidentified, as some crosses are closer to Red Wagyu and some Black animals are gravitating more towards the mean. Crossing and backcrossing of animals is common within the breed, so some crosses may be more genetically like the Fullblood Red Wagyu than a Black Wagyu.

The prediction accuracies per each scenario can be found in Table 6.2 and Table 6.3. The accuracies increased when utilizing whole genome sequence, especially in the across breed prediction accuracies. Prediction accuracies did not increase in the RedBlack population when using the WGS. This could be due to the nature of imputation to WGS, as no animals in the reference population for imputation were RedBlack Wagyu crosses.

70K	Red	RedBlack	Black	Hanwoo	All
<i>Red</i>	0.374594	0.335652	-0.11648	0.085771	0.154744
<i>RedBlack</i>	0.217795	0.383666	0.198072	0.086485	0.235117
<i>Black</i>	0.12328	0.302299	0.11055	0.077759	0.122974
<i>Hanwoo</i>	0.148088	0.273783	0.054902	0.19351	0.223874
<i>All</i>	0.229882	0.413217	0.133042	0.101837	0.38774

Table. 6.2. Prediction accuracies using GBLUP analysis between all breed types at 70K density chip. The training groups are in the rows, and testing groups in columns. Diagonal, within-breed accuracies were split 80/20.

WGS	Red	RedBlack	Black	Hanwoo	All
<i>Red</i>	0.371206	0.341941	-0.03513	0.083345	0.160455
<i>RedBlack</i>	0.214048	0.385192	0.186228	0.064573	0.224814
<i>Black</i>	0.133806	0.300887	0.142109	0.072744	0.126242
<i>Hanwoo</i>	0.182419	0.260344	0.128508	0.192171	0.23864
<i>All</i>	0.233802	0.408325	0.167007	0.095279	0.389136

Table 6.3 Prediction accuracies using GBLUP analysis between all breed types at whole genome sequence. The training groups are in the rows, and testing groups in columns. Diagonal, within-breed accuracies were split 80/20.

These accuracies reflect the structure seen within the PCA, with random sampling of animals from all breed groups having the highest accuracy. The lowest accuracies occurred between the Red and Black Wagyu.

Creating training and testing groups with the first principal component yielded results found in Table 6.4.

<i>PC Split</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>	<i>10</i>
<i>1</i>	0.741538	0.22561	0.274736	0.274841	0.211956	0.148241	-0.0684	0.092799	0.080853	0.124792
<i>2</i>	0.232632	0.757092	0.310467	0.313882	0.232305	0.254334	0.140827	0.075368	0.067799	0.058163
<i>3</i>	0.246209	0.287265	0.771826	0.296464	0.255255	0.288262	-0.01909	0.052816	0.056025	0.096405
<i>4</i>	0.253963	0.285115	0.292298	0.760715	0.257011	0.256342	0.153722	0.049149	0.045374	0.077921
<i>5</i>	0.182446	0.204958	0.260903	0.256986	0.786462	0.297649	0.328171	0.096377	0.067579	0.043934
<i>6</i>	0.162094	0.254469	0.276273	0.257071	0.28735	0.766839	0.457719	0.051433	0.078018	0.059165
<i>7</i>	0.120543	0.115265	0.108513	0.020819	0.100567	0.161845	0.762403	0.118151	0.086604	0.060939
<i>8</i>	0.090184	0.098663	0.098948	0.083739	0.130673	0.060615	0.180494	0.891996	0.164283	0.09978
<i>9</i>	0.087728	0.070019	0.056425	0.050179	0.108749	0.111891	0.242234	0.194645	0.884068	0.140601
<i>10</i>	0.140437	0.058859	0.114395	0.108622	0.052981	0.076539	0.271762	0.062811	0.108292	0.832933

Table 6.4. Prediction accuracies using balanced principal component splits, utilizing the first PC split 10 ways. Most related groups showed the highest prediction accuracies. The training groups are in the rows, and testing groups in columns.

The accuracies for the within PC split accuracy on the diagonal was larger than the within breed accuracies found in Table 6.2 and 6.3. This can be attributed to the balance of genomic relatedness within these PC splits, as well as the balance of samples per PC split. On average, the farther the PC split groups are apart, the lower the prediction accuracy. In many scenarios using split 10 (which included all Hanwoo animals) for the training group yielded higher results than other splits.

The GWAS for the 70k and WGS of all animals can be found in Fig 6.3 and Fig 6.4 respectively.

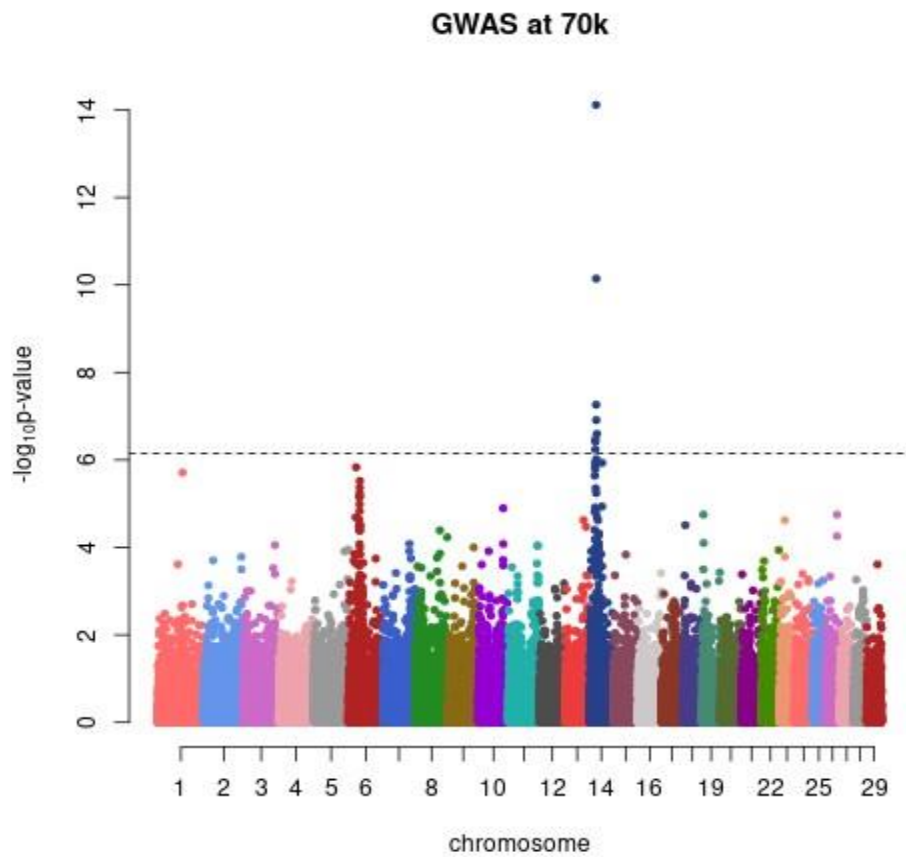


Figure 6.3. Genome Wide Association Study for all animals at 70K for the phenotype of weaning weight.

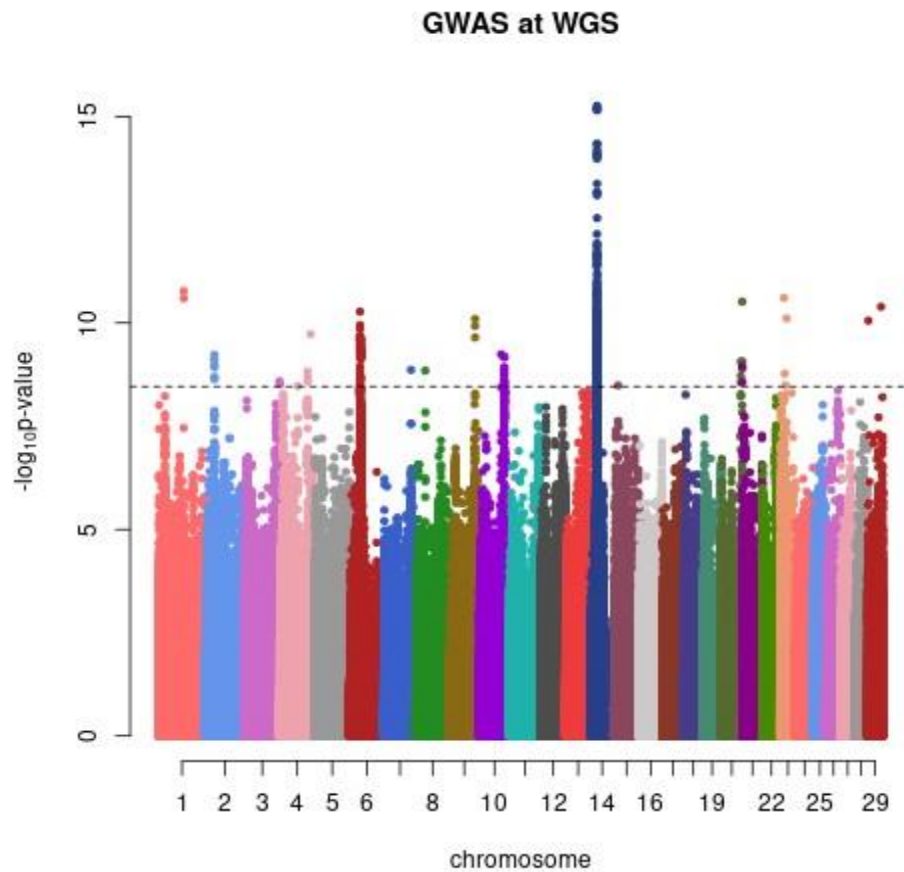


Figure 6.4. Genome Wide Association Study for all animals at whole genome sequence for the phenotype of weaning weight.

In Fig 6.3, a very apparent peak can be seen in chromosome 14. When run against previous significant QTL observed in cattle, the PLAG1 locus was detected and has been previously identified as a growth locus on the genome (see all significant SNP in Table 6.5).

In Fig 6.4, whole-genome analysis identified more areas on the genome, specifically areas on chromosomes 6, 10 and 14. All significant SNP identified with weaning weight (cattle growth), can be seen in supplemental information. Those areas identified included the PLAG1 locus on chromosome 14, PTGR2 on chromosome 10 (which is a part of the fatty-acid metabolism pathway), and NPBWR1 on chromosome 14 (which is involved in food regulation pathway). Significant runs identified on chromosome 6 did not include any significant SNP that are available within the ensembl browser for Bos Taurus. Previous studies have identified QTL on chromosome 6 associated with growth in Hanwoo and Piedmontese cattle (Bongiorni et al.,

2012; Naserkheil et al., 2021): LCORL, NCAPG and LAP3. The location of these QTL previously identified were between the 2 significant peaks that were identified in this study, located on chromosome 6 35550018-35560448 and 38559100-38569258.

SNP Name	CHR	POSITION	QTL
14:20640612_G_A	14	20640612	PLAG1
14:20642540_A_G	14	20642540	PLAG1
14:20646499_G_A	14	20646499	PLAG1
14:23343150_A_G	14	23343150	PLAG1
14:23446328_A_C	14	23446328	PLAG1
14:23630896_T_C	14	23630896	PLAG1
14:24026168_A_G	14	24026168	PLAG1
14:25839968_G_A	14	25839968	PLAG1

Table 6.5. All significant SNP in the genome wide association study at 70K with weaning weight as the phenotype.

Conclusions

The US Wagyu breed sub-types showed to specifically separate which was illustrated through Principal Component Analysis (PCA) and offered intriguing insights to the Wagyu relationship to the Hanwoo. Distinct separations among the Wagyu and Hanwoo populations can be seen, underlining their genetic divergence. Yet, when the between breed genomic prediction was done between the Red Wagyu and the Hanwoo, it showed a moderate prediction accuracy. This linkage further proves the ancestral links between the Hanwoo and Red (Akaushi) Wagyu and points to possible inconsistencies within the PCA in this study. Inclusion of other cattle breeds may aid in uncovering the true relationships in the PCA, as the prediction accuracy showed the Hanwoo and Red Wagyu should group closer together.

Interestingly, a significant separation was evident between Red and Black Wagyu, this distinct divide was further emphasized in prediction accuracies between the two breeds. resulting in the lowest prediction accuracy in this scenario. In practice, all Wagyu are considered with estimating breeding values. It is necessary to include the RedBlack cross animals to increase

accuracy within the whole US Wagyu population, as this group of crossbred animals showed high prediction accuracy when predicting both the Red and Black Wagyu populations. Whole-genome sequence did increase prediction accuracy, but its implementation in industry settings might be hindered by cost implications and the computational demands for handling extensive data.

Genome-Wide Association Studies (GWAS) uncovered notable regions on the genome within this population for growth. While some regions corresponded to previously identified Quantitative Trait Loci (QTL), specifically *PLAG1*, others unveiled unreported areas in Asian populations. These newfound genomic peaks could uncover QTL or represent novel regions in growth traits in Asian cattle breeds.

Moving forward, deeper analyses within these groups should prioritize carcass-based phenotypes, aligning with the breeds' renowned high-marbling qualities. With prediction accuracy observed to be low between Wagyu sub-breeds, the inclusion of RedBlack animals will be pivotal in keeping and improving accuracy levels in a whole population analysis of the US Wagyu. Balancing the differing breed structure of the US Wagyu population will be crucial in guiding future research and estimating breeding values within the industry.

LITERATURE CITED

- Bongiorni, S., Mancini, G., Chillemi, G., Pariset, L., & Valentini, A. (2012). Identification of a short region in chromosome 6 affecting direct calving ease in Piedmontese cattle breed-Supplementary material. *PLoS One*, 7(12).
- Castro Dias Cuyabano, B. C. D., Wackel, H., Shin, D., & Gondro, C. (2019). A study of Genomic Prediction across Generations of Two Korean Pig Populations. *Animals*, 9(9), 672.
- Druet, T., Macleod, I. M., & Hayes, B. J. (2013). Toward genomic prediction from whole-genome sequence data: impact of sequencing design on genotype imputation and accuracy of predictions. *Heredity* 2014 112:1, 112(1), 39–47.
- Gotoh, T., Nishimura, T., Kuchida, K., & Mannen, H. (2018). The Japanese Wagyu beef industry: Current situation and future prospects - A review. In *Asian-Australasian Journal of Animal Sciences* (Vol. 31, Issue 7, pp. 933–950). Asian-Australasian Association of Animal Production Societies.
- Hayes, B. J., Visscher, P. M., & Goddard, M. E. (2009). Increased accuracy of artificial selection by using the realized relationship matrix. *Genetics Research*, 91(1), 47–60.
- Heffernan, K. (2022). Evaluating the Genetic Architecture of the Japanese Wagyu Breed Within the United States.
- Karaman, E., Cheng, H., Firat, M. Z., Garrick, D. J., & Fernando, R. L. (2016). An Upper Bound for Accuracy of Prediction Using GBLUP. *PLOS ONE*, 11(8), e0161054.
- Karamizadeh, S., Abdullah, S. M., Manaf, A. A., Zamani, M., & Hooman, A. (2013). An Overview of Principal Component Analysis. *Journal of Signal and Information Processing*, 04(03), 173–175.
- Khansefid, M., Goddard, M. E., Haile-Mariam, M., Konstantinov, K. V., Schrooten, C., de Jong, G., Jewell, E. G., O'Connor, E., Pryce, J. E., Daetwyler, H. D., & MacLeod, I. M. (2020). Improving Genomic Prediction of Crossbred and Purebred Dairy Cattle. *Frontiers in Genetics*, 11.
- Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 34(18), 3094–3100.
- Loh, P.-R., Danecek, P., Palamara, P. F., Fuchsberger, C., A Reshef, Y., K Finucane, H., Schoenherr, S., Forer, L., McCarthy, S., Abecasis, G. R., Durbin, R., & L Price, A. (2016). Reference-based phasing using the Haplotype Reference Consortium panel. *Nature Genetics*, 48(11), 1443–1448.

- Lund, M. S., Su, G., Janss, L., Guldbrandtsen, B., & Brøndum, R. F. (2014). Genomic evaluation of cattle in a multi-breed context. *Livestock Science*, 166(1), 101–110.
- McVean, G. (2009). A genealogical interpretation of principal components analysis. *PLoS Genetics*, 5(10).
- Meuwissen, T., & Goddard, M. (2010). Accurate Prediction of Genetic Values for Complex Traits by Whole-Genome Resequencing. *Genetics*, 185(2), 623–631.
- Meuwissen, T., van den Berg, I., & Goddard, M. (2021). On the use of whole-genome sequence data for across-breed genomic prediction and fine-scale mapping of QTL. *Genetics Selection Evolution*, 53(1).
- Misztal, I., Steyn, Y., & Lourenco, D. A. L. (2022). Genomic evaluation with multibreed and crossbred data. *JDS Communications*, 3(2), 156–159.
- Motoyama, M., Sasaki, K., & Watanabe, A. (2016). Wagyu and the factors contributing to its beef quality: A Japanese industry overview. *Meat Science*, 120, 10–18.
- Namikawa, K. (1980). *Breeding History Of Japanese Beef Cattle And Preservation Of Genetic Resources As Economic Farm Animals*.
- Naserkheil, M., Mehrban, H., Lee, D., & Park, M. N. (2021). Genome-wide Association Study for Carcass Primal Cut Yields Using Single-step Bayesian Approach in Hanwoo Cattle. *Frontiers in Genetics*, 12, 752424.
- Nawaz, M. Y., Bernardes, P. A., Savegnago, R. P., Lim, D., Lee, S. H., & Gondro, C. (2022). Evaluation of Whole-Genome Sequence Imputation Strategies in Korean Hanwoo Cattle. *Animals*, 12(17), 2265.
- Olson, K. M., VanRaden, P. M., & Tooker, M. E. (2012). Multibreed genomic evaluations using purebred Holsteins, Jerseys, and Brown Swiss. *Journal of Dairy Science*, 95(9), 5378–5383.
- Patterson, N., Price, A. L., & Reich, D. (2006). Population Structure and Eigenanalysis. *PLoS Genetics*, 2(12), 2074–2093.
- Raymond, B., Bouwman, A. C., Schrooten, C., Houwing-Duistermaat, J., & Veerkamp, R. F. (2018). Utility of whole-genome sequence data for across-breed genomic prediction. *Genetics Selection Evolution*, 50(1), 1–12.
- Scraggs, E., Zanella, R., Wojtowicz, A., Taylor, J. F., Gaskins, C. T., Reeves, J. J., de Avila, J. M., & Neibergs, H. L. (2014). Estimation of inbreeding and effective population size of full-blood wagyu cattle registered with the American Wagyu Cattle Association. *Journal of Animal Breeding and Genetics*, 131(1), 3–10.

- Thomasen, J. R., Egger-Danner, C., Willam, A., Guldbrandtsen, B., Lund, M. S., & Sørensen, A. C. (2014). Genomic selection strategies in a small dairy cattle population evaluated for genetic gain and profit. *Journal of Dairy Science*, 97(1), 458–470.
- van den Berg, I., MacLeod, I. M., Reich, C. M., Breen, E. J., & Pryce, J. E. (2020). Optimizing genomic prediction for Australian Red dairy cattle. *Journal of Dairy Science*, 103(7), 6276–6298.
- VanRaden, P. M. (2008). Efficient Methods to Compute Genomic Predictions. *Journal of Dairy Science*, 91(11), 4414–4423.

Supplemental Tables

SNP Name	CHR	POSITION
1:85082264_C_T	1	85082264
1:85082469_C_T	1	85082469
2:32148926_G_A	2	32148926
2:32152141_T_A	2	32152141
2:32153462_G_A	2	32153462
2:32157167_G_A	2	32157167
2:32160234_A_G	2	32160234
2:32163031_A_G	2	32163031
3:119087112_T_C	3	119087112
3:120616386_T_G	3	120616386
4:61336290_T_C	4	61336290
4:95309987_G_A	4	95309987
4:95319999_T_G	4	95319999
4:95339993_C_T	4	95339993
4:105609727_T_A	4	105609727
6:35550018_T_A	6	35550018
6:35550066_T_A	6	35550066
6:35550067_T_A	6	35550067
6:35550287_C_G	6	35550287
6:35556613_T_C	6	35556613
6:35557555_C_T	6	35557555
6:35557917_G_A	6	35557917
6:35557983_T_G	6	35557983
6:35558042_C_T	6	35558042
6:35558523_T_C	6	35558523
6:35558737_A_G	6	35558737
6:35558908_T_C	6	35558908
6:35559215_G_A	6	35559215
6:35559492_A_C	6	35559492
6:35559558_G_A	6	35559558
6:35559635_C_T	6	35559635
6:35559640_A_T	6	35559640

6:35559668_G_A	6	35559668
6:35559795_G_A	6	35559795
6:35559801_G_A	6	35559801
6:35559804_T_C	6	35559804
6:35559823_A_G	6	35559823
6:35559874_C_G	6	35559874
6:35559930_A_G	6	35559930
6:35559958_C_T	6	35559958
6:35559987_G_A	6	35559987
6:35559989_C_A	6	35559989
6:35560103_G_T	6	35560103
6:35560136_G_A	6	35560136
6:35560262_A_T	6	35560262
6:35560448_G_T	6	35560448
6:38559100_T_A	6	38559100
6:38559119_A_G	6	38559119
6:38559838_C_T	6	38559838
6:38559871_A_G	6	38559871
6:38559901_G_A	6	38559901
6:38559927_G_A	6	38559927
6:38560042_C_T	6	38560042
6:38560045_T_C	6	38560045
6:38560268_A_G	6	38560268
6:38560977_T_C	6	38560977
6:38560980_T_C	6	38560980
6:38561113_T_C	6	38561113
6:38561154_C_G	6	38561154
6:38561156_G_A	6	38561156
6:38561261_A_G	6	38561261
6:38561900_G_A	6	38561900
6:38562229_A_G	6	38562229
6:38562559_C_T	6	38562559
6:38562675_A_G	6	38562675
6:38563338_T_G	6	38563338
6:38563658_C_G	6	38563658
6:38563833_A_T	6	38563833
6:38564037_T_G	6	38564037
6:38564080_T_C	6	38564080

Supplemental Table 6.6: Name, chromosome, and position of each significant SNP related to growth in Asian cattle identified through genome-wide association with whole-genome sequence.

Supplemental Table 6.6 (cont'd)

6:38564419_T_C	6	38564419
6:38565084_T_C	6	38565084
6:38565365_C_G	6	38565365
6:38566242_A_G	6	38566242
6:38566447_C_T	6	38566447
6:38568231_C_T	6	38568231
6:38569258_T_C	6	38569258
7:93070914_G_A	7	93070914
8:31787017_G_T	8	31787017
9:89474378_C_T	9	89474378
9:89480926_C_T	9	89480926
9:89480968_G_A	9	89480968
10:74881199_A_G	10	74881199
10:85206757_G_A	10	85206757
10:85207181_A_G	10	85207181
10:85212383_C_T	10	85212383
10:85216887_T_C	10	85216887
10:85221410_T_G	10	85221410
10:85228665_C_T	10	85228665
10:85230003_C_T	10	85230003
10:85231678_C_T	10	85231678
10:85244139_A_C	10	85244139
10:85336209_A_G	10	85336209
10:85336554_C_T	10	85336554
10:85338651_C_A	10	85338651
14:20089453_T_C	14	20089453
14:20222988_C_A	14	20222988
14:20223163_C_T	14	20223163
14:20269421_C_T	14	20269421
14:21609324_C_T	14	21609324
14:21609334_C_T	14	21609334
14:21609399_G_C	14	21609399
14:21609456_C_T	14	21609456
14:21609557_C_T	14	21609557
14:21609685_G_T	14	21609685
14:21609731_C_T	14	21609731
14:21609861_G_A	14	21609861
14:21610015_G_A	14	21610015
14:21610069_G_A	14	21610069
14:21610423_A_T	14	21610423
14:21610510_C_T	14	21610510
14:21610668_T_C	14	21610668

14:21610669_T_G	14	21610669
14:21610861_G_A	14	21610861
14:21610871_C_T	14	21610871
14:21610876_G_T	14	21610876
14:21610895_A_C	14	21610895
14:21610975_T_C	14	21610975
14:21611515_A_G	14	21611515
14:21611539_A_G	14	21611539
14:21612303_A_G	14	21612303
14:21612536_G_A	14	21612536
14:21612729_T_A	14	21612729
14:21612745_G_T	14	21612745
14:21612768_C_T	14	21612768
14:21612786_C_A	14	21612786
14:21612829_C_A	14	21612829
14:21612930_T_C	14	21612930
14:21612980_A_C	14	21612980
14:21613086_T_C	14	21613086
14:21613099_T_G	14	21613099
14:21613232_T_C	14	21613232
14:21613332_G_A	14	21613332
14:21613344_G_A	14	21613344
14:21613362_G_A	14	21613362
14:21613382_T_G	14	21613382
14:21613383_T_G	14	21613383
14:21613409_A_C	14	21613409
14:21613413_A_C	14	21613413
14:21613427_C_T	14	21613427
14:21613441_T_C	14	21613441
14:21613480_A_G	14	21613480
14:21613585_G_A	14	21613585
14:21613587_T_C	14	21613587
14:21613637_T_C	14	21613637
14:21613664_A_T	14	21613664
14:21613690_T_C	14	21613690
14:21613761_G_A	14	21613761
14:21613818_C_T	14	21613818
14:21613953_C_T	14	21613953
14:21613960_C_T	14	21613960
14:21613962_T_C	14	21613962
14:21613964_T_G	14	21613964
14:21614346_C_T	14	21614346

Supplemental Table 6.6 (cont'd)

14:21614387_T_A	14	21614387
14:21614411_A_T	14	21614411
14:21614491_C_T	14	21614491
14:21614565_A_G	14	21614565
14:21614587_A_G	14	21614587
14:21614646_A_C	14	21614646
14:21614907_T_C	14	21614907
14:21614946_A_G	14	21614946
14:21615021_G_A	14	21615021
14:21615254_G_A	14	21615254
14:21615342_C_T	14	21615342
14:21615348_A_G	14	21615348
14:21615363_C_T	14	21615363
14:21615367_G_A	14	21615367
14:21615406_C_G	14	21615406
14:21615508_G_A	14	21615508
14:21615676_T_A	14	21615676
14:21616096_A_G	14	21616096
14:21616168_G_A	14	21616168
14:21616723_T_A	14	21616723
14:21618368_A_G	14	21618368
14:21618606_C_A	14	21618606
14:21619114_A_G	14	21619114
14:21619120_G_A	14	21619120
14:21619285_G_A	14	21619285
14:21619764_C_T	14	21619764
14:21620605_G_A	14	21620605
14:21621311_A_C	14	21621311
14:21621477_C_A	14	21621477
14:21623073_G_C	14	21623073
14:21623991_G_C	14	21623991
14:21625776_A_G	14	21625776
14:22670309_G_A	14	22670309
14:22805657_G_A	14	22805657
14:22809167_G_A	14	22809167
14:22813456_C_T	14	22813456
14:22814269_T_C	14	22814269
14:22814595_C_T	14	22814595
14:22815875_G_C	14	22815875
14:22817795_C_T	14	22817795
14:22840845_G_A	14	22840845
14:22867326_T_G	14	22867326

14:22977521_G_A	14	22977521
14:22982222_T_C	14	22982222
14:22983440_C_T	14	22983440
14:22989917_G_A	14	22989917
14:22993121_C_T	14	22993121
14:22995861_C_G	14	22995861
14:22997098_A_T	14	22997098
14:22997286_C_A	14	22997286
14:22997290_G_A	14	22997290
14:22999212_T_G	14	22999212
14:22999280_G_A	14	22999280
14:23001109_G_T	14	23001109
14:23003295_T_C	14	23003295
14:23004135_A_C	14	23004135
14:23004758_T_C	14	23004758
14:23009266_C_T	14	23009266
14:23029146_A_G	14	23029146
14:23128784_T_C	14	23128784
14:23155902_T_G	14	23155902
14:23157224_C_T	14	23157224
14:23165825_C_T	14	23165825
14:23166431_A_T	14	23166431
14:23168386_G_A	14	23168386
14:23168918_C_G	14	23168918
14:23170297_G_A	14	23170297
14:23172658_T_C	14	23172658
14:23172995_A_G	14	23172995
14:23173588_C_G	14	23173588
14:23173674_T_C	14	23173674
14:23175099_T_C	14	23175099
14:23176131_A_C	14	23176131
14:23178091_A_G	14	23178091
14:23179478_C_T	14	23179478
14:23179846_G_C	14	23179846
14:23182232_G_C	14	23182232
14:23183709_C_A	14	23183709
14:23183729_G_A	14	23183729
14:23184342_A_C	14	23184342
14:23184460_T_C	14	23184460
14:23184735_T_A	14	23184735
14:23185438_C_T	14	23185438
14:23188778_G_A	14	23188778

Supplemental Table 6.6 (cont'd)

14:23189327_T_C	14	23189327
14:23189446_C_T	14	23189446
14:23189453_C_A	14	23189453
14:23189633_G_T	14	23189633
14:23189695_G_A	14	23189695
14:23189757_A_G	14	23189757
14:23189783_G_C	14	23189783
14:23189891_T_C	14	23189891
14:23190227_T_G	14	23190227
14:23190282_A_G	14	23190282
14:23190369_A_T	14	23190369
14:23190575_C_A	14	23190575
14:23190736_C_A	14	23190736
14:23190805_G_A	14	23190805
14:23190818_G_A	14	23190818
14:23190885_A_G	14	23190885
14:23190918_A_G	14	23190918
14:23190948_A_G	14	23190948
14:23191049_A_C	14	23191049
14:23191219_C_T	14	23191219
14:23191296_G_A	14	23191296
14:23191494_T_C	14	23191494
14:23191567_A_G	14	23191567
14:23191628_A_G	14	23191628
14:23191765_T_C	14	23191765
14:23191781_A_G	14	23191781
14:23191800_A_G	14	23191800
14:23191895_C_T	14	23191895
14:23192069_C_G	14	23192069
14:23192247_T_C	14	23192247
14:23192277_G_A	14	23192277
14:23192311_G_A	14	23192311
14:23192569_G_C	14	23192569
14:23192583_C_T	14	23192583
14:23192592_G_A	14	23192592
14:23192602_A_C	14	23192602
14:23192794_A_G	14	23192794
14:23192837_G_A	14	23192837
14:23192858_T_G	14	23192858
14:23192959_T_C	14	23192959
14:23192973_A_G	14	23192973
14:23193080_A_G	14	23193080

14:23193261_T_C	14	23193261
14:23193270_T_C	14	23193270
14:23194095_T_C	14	23194095
14:23194344_T_C	14	23194344
14:23194350_T_C	14	23194350
14:23194604_G_A	14	23194604
14:23194808_A_T	14	23194808
14:23194809_A_C	14	23194809
14:23196362_C_G	14	23196362
14:23196417_T_G	14	23196417
14:23196725_C_T	14	23196725
14:23197858_C_T	14	23197858
14:23199580_A_G	14	23199580
14:23199714_G_T	14	23199714
14:23201028_G_A	14	23201028
14:23201789_C_T	14	23201789
14:23212612_T_C	14	23212612
14:23214574_G_C	14	23214574
14:23215630_T_G	14	23215630
14:23253786_A_G	14	23253786
14:23264575_A_G	14	23264575
14:23296927_G_C	14	23296927
14:23297204_T_C	14	23297204
14:23297472_A_G	14	23297472
14:23298062_A_G	14	23298062
14:23300304_G_C	14	23300304
14:23314460_T_C	14	23314460
14:23314730_C_T	14	23314730
14:23314761_T_C	14	23314761
14:23326588_C_G	14	23326588
14:23329375_T_C	14	23329375
14:23338890_G_T	14	23338890
14:23343150_A_G	14	23343150
14:23346065_C_G	14	23346065
14:23354569_A_G	14	23354569
14:23383008_A_G	14	23383008
14:23384445_T_C	14	23384445
14:23392938_T_G	14	23392938
14:23414524_A_G	14	23414524
14:23416592_C_G	14	23416592
14:23418801_T_C	14	23418801
14:23423823_T_C	14	23423823

Supplemental Table 6.6 (cont'd)

14:23427548_T_C	14	23427548
14:23434525_A_G	14	23434525
14:23438738_C_T	14	23438738
14:23440914_A_G	14	23440914
14:23443706_C_T	14	23443706
14:23471725_C_T	14	23471725
14:23473697_C_T	14	23473697
14:23478900_A_G	14	23478900
14:23479427_C_T	14	23479427
14:23495771_A_T	14	23495771
14:23529103_C_G	14	23529103
14:23552239_G_A	14	23552239
14:23552679_C_T	14	23552679
14:23553197_A_G	14	23553197
14:23554044_C_G	14	23554044
14:23554490_A_C	14	23554490
14:23555313_T_C	14	23555313
14:23555320_A_C	14	23555320
14:23555430_A_G	14	23555430
14:23555460_G_A	14	23555460
14:23555549_A_G	14	23555549
14:23555550_A_G	14	23555550
14:23555551_C_A	14	23555551
14:23555570_A_T	14	23555570
14:23559882_G_C	14	23559882
14:23564369_C_A	14	23564369
14:23565143_T_C	14	23565143
14:23566885_C_T	14	23566885
14:23578196_G_A	14	23578196
14:23583466_A_G	14	23583466
14:23583576_A_G	14	23583576
14:23583640_T_C	14	23583640
14:23585785_G_A	14	23585785
14:23587335_C_T	14	23587335
14:23587652_G_A	14	23587652
14:23587653_A_G	14	23587653
14:23587737_T_G	14	23587737
14:23590382_A_G	14	23590382
14:23590403_C_T	14	23590403
14:23590438_C_G	14	23590438
14:23590493_T_A	14	23590493
14:23590931_C_T	14	23590931

14:23590939_G_A	14	23590939
14:23590941_T_C	14	23590941
14:23591033_A_G	14	23591033
14:23591827_T_C	14	23591827
14:23592917_A_G	14	23592917
14:23593940_C_T	14	23593940
14:23594517_C_T	14	23594517
14:23596025_C_T	14	23596025
14:23596251_G_A	14	23596251
14:23596422_A_G	14	23596422
14:23596429_A_T	14	23596429
14:23600267_G_A	14	23600267
14:23601668_G_T	14	23601668
14:23605155_T_G	14	23605155
14:23607944_A_C	14	23607944
14:23608832_A_G	14	23608832
14:23612458_C_T	14	23612458
14:23612481_A_G	14	23612481
14:23616675_A_G	14	23616675
14:23617053_G_A	14	23617053
14:23618229_T_G	14	23618229
14:23618495_A_G	14	23618495
14:23618614_A_G	14	23618614
14:23618734_G_T	14	23618734
14:23618798_G_T	14	23618798
14:23618800_T_A	14	23618800
14:23618814_T_G	14	23618814
14:23619269_G_A	14	23619269
14:23619297_T_C	14	23619297
14:23620105_C_T	14	23620105
14:23620227_A_T	14	23620227
14:23620640_G_C	14	23620640
14:23621082_T_C	14	23621082
14:23621083_G_A	14	23621083
14:23621256_G_A	14	23621256
14:23630842_T_C	14	23630842
14:23630896_T_C	14	23630896
14:23633052_G_A	14	23633052
14:23634452_A_T	14	23634452
14:23637052_G_T	14	23637052
14:23638330_C_G	14	23638330
14:23638895_A_G	14	23638895

Supplemental Table 6.6 (cont'd)

14:23639458_G_A	14	23639458
14:23639633_T_C	14	23639633
14:23639701_C_A	14	23639701
14:23639932_T_A	14	23639932
14:23640016_A_G	14	23640016
14:23644290_A_C	14	23644290
14:23645199_G_A	14	23645199
14:23646689_G_A	14	23646689
14:23647559_C_T	14	23647559
14:23652025_C_T	14	23652025
14:23652804_G_A	14	23652804
14:23653071_T_A	14	23653071
14:23864771_A_G	14	23864771
14:23888150_C_T	14	23888150
14:23888591_A_T	14	23888591
14:23888612_C_T	14	23888612
14:23890981_G_T	14	23890981
14:23892930_C_T	14	23892930
14:23896609_G_A	14	23896609
14:23903530_C_T	14	23903530
14:23908621_T_A	14	23908621
14:23912123_C_T	14	23912123
14:23912998_G_T	14	23912998
14:23915419_C_T	14	23915419
14:23916478_G_T	14	23916478
14:25446205_G_T	14	25446205
15:13337618_G_A	15	13337618
20:67275290_C_G	20	67275290
20:67276662_G_A	20	67276662
20:67276917_A_G	20	67276917
20:71900962_G_A	20	71900962
20:71904894_A_G	20	71904894
20:71905923_C_A	20	71905923
21:285066_T_G	21	285066
21:285235_G_A	21	285235
23:12495116_C_T	23	12495116
23:15370628_A_T	23	15370628
23:18775299_A_G	23	18775299
23:21751972_A_G	23	21751972
29:2392400_G_C	29	2392400
29:45964941_G_A	29	45964941

CHAPTER 7: Conclusions

The high-quality beef revolution in the US has already begun, as consumers are demanding high-marbling beef products, with validation of the breed composition of these high-cost products. Rapid identification of US Wagyu products is needed, and this can be accomplished through genomic sequencing. New equipment by Oxford Nanopore Technologies specifically designed for out-of-lab sequencing was shown to be able to tackle this through matching the genotypic output per sample to reference haplotypes. This out-of-lab protocol was also done at a low-cost and was proven to have high correlation and concordance rates of sample to reference, even at low depth and coverage. This technique showed that even with low genomic output, and relatively no knowledge about laboratory procedures, a sample from an animal can be traced back to the breed of origin. The out-of-lab possibility of these technologies could be endless, as genomic information can be utilized for breed composition, disease testing, or even quick identification of parentage with the correct bioinformatic tools.

Although product verification is crucial for consistency of Wagyu product in the market, understanding the population structure of the Wagyu breed in the United States is of equal importance. This breed originates from a small number of animals, imported under strange circumstances, and specific selection pressures are singular to this group of animals outside of Japan. Genomic architectures uncovered a population that is inbred and shows signals of low genomic variability. This is to be expected and is good confirmation of the suspected population status in the US. Keeping a beat on the inbreeding and variation available to Wagyu producers ensures a healthy breeding stock for years to come.

The Wagyu breed in the US is considered as a single large population, yet it comprises three primary subgroups: Red/Brown Wagyu (Akaushi), Black Wagyu, and the RedBlack crossbreed. A Principal Component Analysis (PCA) revealed distinct connections among these breeds, with Black and Red Wagyu groups segregating and the RedBlack animals bridging these groups. When the Korean Hanwoo was included in this analysis, it stood apart from the Wagyu animals, forming a separate cluster. However, predictive accuracy highlighted a moderate relationship between Red Wagyu and Korean Hanwoo. Notably, prediction accuracies were lowest when attempting to predict Black Wagyu solely from Red Wagyu, possibly affected by the substantial imbalance in breed population sizes (~150 Black Wagyu compared to ~5000 Red). Further investigations warrant a larger representation of Black Wagyu and other breed

groups to unravel additional genomic structures not encompassed in this study. This exploration of connections between US Wagyu sub-breeds and their Asian relatives holds promise for understanding the evolutionary trajectories of this understudied breed and potentially leveraging breeding practices across various Asian breeds. Specifically, this study illuminated the relationships between the subtypes within the US Wagyu population and with the Korean Hanwoo, revealing that Akaushi (Red Wagyu) shares a closer genetic link with Hanwoo than with Black Wagyu. Further exploration into the population structure of US Wagyu for the estimation of breeding values within the whole population will be explored to increase accuracy. Inclusion of crossbred RedBlack animals will be a crucial piece in linking the two Red and Black populations together

Inclusion of all Asian breeds studied in the GWAS uncovered areas of the genome that were not previously researched for growth traits in Asian cattle. This discovery is a new look into the selection patterns of Asian breeds for growth and may uncover areas of the genome that emerged to be highly associated with phenotypes that are previously unknown. Although this new discovery is an exciting development in understanding genomic architecture, the growth traits are not a huge target for Wagyu cattle. Further investigation through genome-wide association on the marbling phenotypes, such as intramuscular fat percentage or marbling fineness, would be the most beneficial for Wagyu producers and researchers. New genomic QTL controlling these sought after qualities would further open the curtain into the selection pressures on the most desired traits in Wagyu.

As Wagyu emerges as a prominent breed in the United States, regulation and verification of animals and product through genotype can be done on-site, rapid sequencing technologies. Outlook of the Wagyu and Akaushi populations in the United States is positive if inbreeding is kept at bay and genomic variation is targeted. Connections between the Asian breeds in the United States and abroad should be further explored, as differing selection pressures may drift these groups farther apart. Improving the accuracy of breeding values in Wagyu will start with understanding the population currently available in the United States and working towards bettering those accuracies for all industry professionals through inclusion of all Wagyu-type animals in the reference population. The impact of this breed on the beef market to increase quality is an avenue for producers to best utilize these cattle, and selection for those high-quality animals then becomes crucial for the everyday rancher.

APPENDIX A: Sale report for Wagyu in March 2023.

Triangle B Ranch – 15th Annual Production Sale March 18th, 2023 Stigler, Oklahoma

Sale Manager: James Danekas & Associates Inc.

Live Broadcast & Bidding: LiveAuctions.tv

Averages:

20 Fullblood Females	\$26,525.00
12 Fullblood Bulls	\$33,471.00
3 Pregnancies	\$8,667.00
1 Flush	\$8,500.00
6 Embryos	\$1,025.00/embryo
60 Units of Semen	\$167/unit

TOPS:

Females:

Lot 1: TBR HIKOKURA 035 7702K, 8/23/2022 sired by Arubial United; **\$400,000** to **Eden Valley Wagyu**, Eden Valley, MN.

Lot 5: TBR HIKOHIME B438 2-1 7637, 3/13/2021 sired by TBR MITSUITOFUKU 2149Y; **\$13,500** to **G. W. Wagyu Farm**, Jay, OK.

Lot 7: TBR HIKOFUKU 3-9-3 7148H, 5/06/2020 sired by TBR KIKUTNAMI 4051A; **\$10,000** to **Perpetua Wagyu**, Tulsa, OK.

Lot 3: TBR CHIYOTAKE B487 7040G, 12/2/2019 sired by BLACKMORE YASUCHIYO C058; **\$9,000** to **G. W. Wagyu Farm**, Jay, OK.

Bulls:

Lot 32: TBR SHIGEFUKUNAMI 7704K, 8/26/2022 sired by Arubial United; **\$325,000** to **Flying A Wagyu**, La Salle, CO.

Lot 39: TBR KIKUTNAMI 7569K, 5/01/2022 sired by TBR KIKUTNAMI 4051A; **\$9,000** to **Philip Parish**, Eddyville, KY.

Lot 44: TBR ITOZURUNAMI 7626K, 3/19/2022 sired by TBR SHIGENAMINAMI 3024Z; **\$8,250** to **Philip Parish**, Eddyville, KY.

Pregnancy:

Lot 46: ARUBIAL BOND Q007 X TBR HIKOKURA 035 5 7232J; **\$15,000** to **Wilders Wagyu**, Turkey, NC.

With a shot of rain the day before the sale, the pastures started to pop with fresh, bright green grass on sale day. With the sun shining and a large crowd gathering, the 15th Annual Triangle B Ranch sale was a huge success. The guests enjoyed a fullblood Wagyu lunch and a very exciting sale. Bidding was very active and to start the auction off, a record high seller was sold. The 80+

people online and the full loft at the sale witnessed history. It didn't stop there as half way through the sale yet another record was shattered to make this sale hold the top selling fullblood Wagyu bull and female ever sold in North America. To top it off, in the crowd to witness all this go down was "Bubbles", the cutest little girl pet monkey. The day couldn't have gone any better.

APPENDIX B: Sale report for Angus sale in April 2023.

Trask Angus Bull Sale

Herd City: Wall, SD US
 Sale City: Philip Livestock, Philip, SD US
 Sale Date: 04/18/2023

Auctioneer: Lynn Weishaar
 Angus Media Rep: Rod Geppert

Lots	Gross	Average
10 Older Bulls	\$40,750.00	\$4,075.00
67 Yrlg.Bulls	\$312,250.00	\$4,660.00
77 Total Registered Bulls	\$353,000.00	\$4,584.00
77 Reported Sale Total	\$353,000.00	\$4,584.00

Top Bull		
Lot 5	Trask Full House 2008	\$9,500.00
Buyer: Jim Wright - Parade, SD Sire Name: WR Full House Seller: Trask Angus Bull Sale Wall SD More Info...		
Lot 56	Trask Justified 2048	\$8,000.00
Buyer: Jim Wright - Parade, SD Sire Name: D U Defender 032 Seller: Trask Angus Bull Sale Wall SD More Info...		
Lot 20	Trask Full House 2032	\$7,750.00
Buyer: Jim Witcher - Scenic, SD Sire Name: WR Full House Seller: Trask Angus Bull Sale Wall SD More Info...		
Lot 4	Trask Full House 2090	\$7,500.00
Buyer: HoosCow Angus - Rushville, NE Sire Name: WR Full House Seller: Trask Angus Bull Sale Wall SD More Info...		
Lot 29	Trask Full House 2185	\$7,000.00
Buyer: Pat Trask - Wasta, SD Sire Name: WR Full House Seller: Trask Angus Bull Sale Wall SD More Info...		

Comments: 1 Gelding Horse - \$8,750

APPENDIX C: Live animal specifications for Wagyu in the United States from the USDA.



Agricultural
Marketing
Service

Quality Assessment Division
1400 Independence Avenue SW, Stop 0258
Washington, DC 20250

AWA Specification
May 5, 2021
Page 1 of 2

LIVE ANIMAL SPECIFICATION AMERICAN WAGYU ASSOCIATION'S SPECIFICATION FOR CHARACTERISTICS OF CATTLE ELIGIBLE FOR APPROVED PROGRAMS CLAIMING WAGYU INFLUENCE

1 SCOPE

- 1.1** This specification sets forth the requirements for live cattle from which carcasses can be derived for certification in approved branded beef programs claiming Wagyu influence. Cattle may qualify for the breed influence specification of these branded beef programs by meeting the Genotype requirements specified below.

2 REQUIREMENTS

- 2.1** Cattle eligible for Wagyu Fullblood (100% Wagyu) certified beef programs must be registered with the American Wagyu Association and DNA parent verified through the American Wagyu Association to both sire and dam. Both parents will be registered by the American Wagyu Association.
- 2.2** Cattle eligible for Wagyu Purebred (93.75% Wagyu minimum) certified beef programs must be registered with the American Wagyu Association and DNA parent verified through the American Wagyu Association to both sire and dam. Both parents will be registered by the American Wagyu Association.
- 2.3** Cattle eligible for Wagyu influence certified beef programs must have positive identification (e.g. ear tags, tattoos, brands) and be traceable to provable (e.g. registration papers) Wagyu parentage. Qualifying cattle must be traceable to at least one registered parent (Fullblood or Purebred). The registration categories can be viewed at www.wagyu.org. For an animal to qualify as Wagyu, subsequently allowing the meat from the animal to be eligible to claim Wagyu influence, the animal must be an offspring of:
- 2.3.1 A registered sire of at least 15/16 or 93.75% Wagyu blood (i.e. Fullblood or Purebred) or;
 - 2.3.2 A registered dam of at least 15/16 or 93.75% Wagyu blood (i.e. Fullblood or Purebred);

3. VERIFICATION

- 3.1 Genotype.** Compliance with genotype requirements shall be determined and controlled through a quality management program approved and monitored by the Livestock and Poultry Program.

USDA is an equal opportunity provider, employer, and lender.



**United States
Department of
Agriculture**

Agricultural
Marketing
Service

Quality Assessment Division
1400 Independence Avenue SW, Stop 0258
Washington, DC 20250

AWA Specification
May 5, 2021
Page 2 of 2

Chad Nelson, National Meat Supervisor
Standards and Specifications Division
Livestock & Poultry Program