A SURVEY OF RELATIONSHIPS AMONG RARE BREEDS OF SWINE

A Thesis

presented to

the Faculty of the Graduate School

at the University of Missouri-Columbia

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

---

by

KIZZI ROBERTS

Dr. William Lamberson, Thesis Supervisor

December 2014

The undersigned, appointed by the dean of the Graduate School, have examined the thesis entitled

A SURVEY OF RELATIONSHIPS

AMONG RARE BREEDS OF SWINE

presented by Kizzi Roberts,

a candidate for the degree of master of science

and hereby certify that, in their opinion, it is worthy of acceptance.

Professor William Lamberson

Professor Timothy Safranski

Professor Matthew Gompper

Thanks Mom and Dad.

ACKNOWLEDGEMENTS

I would like to acknowledge and thank my advisor Dr. Bill Lamberson for all of his help and support over the years as I worked toward completing my degree. I could not have asked for a better mentor and professor.

A big thank you to my committee members Dr. Tim Safranski and Dr. Matthew Gompper for their support and patience as I worked toward finishing my thesis. I appreciate their willingness to work with me and their contributions toward helping me complete this process.

I would also like to thank Tasia Taxis, my fellow graduate student, for all of her support and guidance during this process.

Thank you to Cinda Hudlow for being an outstanding resource for all things related to graduate school and life.

TABLE OF CONTENTS

LIST OF FIGURES AND EQUATIONS

Figure Page

Equation

LIST OF TABLES AND PLOTS

Table Page

Plot

LIST OF ABBREVIATIONS

TLC - The Livestock Conservancy

F - Inbreeding coefficient

GRM - Genomic Relationship Matrix

MAF - Minor Allele Frequency

MDS - Multidimensional Scaling

R - Relationship between two individuals

SNP - Single Nucleotide Polymorphism

A SURVEY OF RELATIONSHIPS
AMONG RARE BREEDS OF SWINE

Kizzi Roberts

Dr. William Lamberson, Thesis Supervisor

ABSTRACT

Genetic diversity allows adaptation to environmental changes and varied disease resistance. Without such diversity, a population could be decimated by disease or environmental fluctuations. Swine breeds facing extinction share characteristics such as small size, slow growth rate and high fat percentage which eliminate them from the high-input high-output business of commercial production. Small populations and lack of genetic information increases the chance that producers are breeding closely related individuals, which ultimately eliminates genetic diversity by increasing levels of homozygosity in subsequent generations. By making genetic data available, producers can make more educated breeding decisions to preserve genetic diversity in future generations.

Hair samples were collected from Guinea, Ossabaw Island, Red Wattle, American Saddleback, and Mulefoot pigs and genotyped with the Porcine 60k SNP chip. Publicly available genotyping data were obtained for British Saddleback, Duroc, Landrace, Large White, Pietrain and Tamworth pigs. PLINK was used to construct a genomic relationship matrix and to calculate inbreeding coefficients. American Saddleback and British

Saddleback showed relatedness across the two breeds, so they were combined. Breed averages for relatedness (R) and inbreeding coefficient (F) were compared using SAS. The model was significant (P-value <0.0001). PLINK was also used to determine SNPs in linkage disequilibrium and to create a multidimensional scaling diagram.

Popular breeds (Landrace, Large White, and Duroc) exhibit lower levels of R among individuals, on average, as compared to R between individuals of endangered breeds, especially Ossabaw Island, Red Wattle, Mulefoot and Tamworth. Following a similar pattern, F is high for Ossabaw Island, Tamworth, and Mulefoot, and low for Large White and Landrace. While less common in the United States, Pietrain is a popular breed in Europe which likely accounts for low R and F values.

Having complete pedigrees and large populations allows commercial breeds to maintain low levels of R and F within a population. For heritage type breeds, lack of popularity means fewer individuals to select among, and, within a viable populous, even fewer have known pedigrees. This research is the first step toward preserving genetic diversity by providing producers with accurate genetic information.

CHAPTER 1: OBJECTIVES AND LITERATURE REVIEW

Introduction and Objectives

In recent years, heritage breed producers have become interested in the genetics of their animals and in ways to preserve diversity for future generations. The focus on swine arises from a greater threat of extinction as compared to other endangered livestock breeds. Raising pigs is far more management intensive than other livestock production systems, therefore it is rare for farmers to keep pigs as a hobby. This puts rare breeds of pigs at a greater risk of being lost, because overall there are fewer independent pig farmers than other types of livestock farmers.

For this project, breeds of interest were selected among those listed as critical by The Livestock Conservancy (TLC), and by ease of access to samples. Endangered breeds include the Guinea, Ossabaw Island, Mulefoot, Red Wattle and Saddleback.

For further comparison, genotyping data were obtained for common industry breeds. By comparing heritage breeds to industry breeds, differences and similarities may be identified. Understanding such traits is paramount to educating producers and the public about the value endangered breeds offer through genetic diversity.

Single nucleotide polymorphisms, SNPs, will be used to identify alleles specific to a breed, allowing animals of unknown origin to be identified. For heritage producers this is important because they are working toward preserving the integrity of the breed. Animals may sometimes share phenotypic characteristics of the breed of interest but on a genetic level, they are something different. Classifying such animals appropriately is important if rare breeds are to be successfully conserved.

Hair samples were genotyped using the porcine 60k SNP chip at GeneSeek in Lincoln, Nebraska. Results were uploaded to a Sharefile account. Files could then be further analyzed using genome wide association software such as PLINK.

The objectives of this research project are to better understand relationships among rare breeds of pigs and identify breed differences through the use of SNP data. Such relationships include those between individuals as well as between breeds.

Literature Review

Mankind entered the 20th century with a global population of roughly 1.6 billion. By the year 2000, the population surpassed six billion people (Population Reference Bureau, 2013). Today, world population exceeds seven billion, and in the United States alone there are over 315 million people (United States Census Bureau, 2010). With populations expected to continue growing over the next few decades, a major concern in every country is how to provide stable food sources for their people.

In the United States, agriculture has gone through significant changes over the course of the last century. In 1900, 40 percent of the American workforce was employed in agriculture. By 2000, this number had dropped to 1.9 percent of the workforce (Dimitri et al., 2005). As the workforce changed, so did the structure of the farms themselves.

In the 20th century, farming practices, used the world over and unchanged for hundreds of years, would experience a radical revolution in the United States. Farmers using horses for power in the early 1900's would begin using tractors with horsepower by 1930 (Dimitri et.al, 2005). A family farm previously able to pick 500 pounds of cotton in a day, could pick 40,000 pounds in a day using new technology (Paarlburg, 2000). In 1900, the average farm produced five different commodities, but by 2000, farming had become more specialized with each farm producing, on average, one commodity (Dimitri et al., 2005). By the end of the 20th century, the modern farm held little resemblance to the farm of 1900.

With new technologies and a changing economy, jobs off the farm became more available and increasingly appealing. In 1930, only 30 percent of farmers also held off farm jobs, but by 2000, 93 percent of farmers earned additional income off the farm (Dimitri et al., 2005). To get many of the new industry jobs, families moved from rural to

urban settings (Paarlburg, 2000), and the number of farms in the U.S. fell by 63 percent. Concurrently, average farm size increased by 67 percent (Dimitri et al., 2005). This new, larger farm paved the way for modern agriculture as farmers relied heavily on new technologies to feed a growing population.

New production systems required more capital and made use of artificial insemination and antibiotics, as well as advances in knowledge of breeding, genetics, disease control, and nutrition (National Research Council, 1993). With the growing U.S. population demanding convenience and consistency from agriculture products, especially meat (Windig, 2010; Pinstrup-Anderson, 1999), larger operations specializing in one species were better equipped to meet these demands.

The emergence of confined animal feeding operations, CAFO, allowed farmers to control growing conditions in an environment capable of supporting greater numbers of livestock. Livestock such as poultry and swine, once raised outdoors, were increasingly housed in these operations. With nearly absolute control over indoor environments, farmers rigorously selected animals able to thrive in such conditions. By manipulating both the growing environment and the animals themselves, the ultimate goal became producing a uniform end product.

Uniform systems result in uniform animals (Hall, 2004; Hassebrook, 1999). Uniformity is best illustrated by the poultry and swine industries, in which diversity is limited to small-scale or hobby farms with the commercial industry favoring a select few breeds (Cardellino, 2004). By the 1990's, four major processing firms controlled more than 50% of the pork industry, with each firm producing their own uniform pork product

(Heffernan, 1999). Today, five firms control nearly 75% of the industry (National Pork Board, 2009). In agriculture the needs of many are being provided by a handful of breeds.

The shift to large-scale operations and the perceived loss of the family farm stirs nostalgia in many, but the efficiency of modern agriculture is hard to dispute. In 1910, a farmer could expect to earn one dollar of income for just over an hour of labor. By 1980, a farmer could earn the same amount in about four minutes of labor (Paarlburg, 2000). For the American farmer, it's difficult to feel nostalgia for former techniques when modern technology has proven exceptionally profitable. Additionally, the farmer was not the only beneficiary of improved efficiency.

In 1900 the average household spent 25 percent of its income on food, by 2000 this number would decrease to 14 percent (Paarlburg, 2000), and today, citizens of the United States spend less than 7 percent of their income on food. This number is lower than any other country in the world (USDA, 2013; Washington State University, 2011).

For all the benefits of modern agriculture, there are also drawbacks. Concerns often focus on the environmental impact of concentrated operations or the welfare of the animals. While both issues are important to the sustainability of agriculture, another essential component is often overlooked, biodiversity.

The importance of biodiversity in nature is often easier to see than its importance on the farm (Taberlet, 2008; Hall, 2004).The public sees value in conservation of wild species such as the rhinoceros or giant panda. Preservation of these unique species is given much attention and funding (Taberlet, 2008). Individual breeds represent biodiversity in livestock, but their preservation makes a less compelling argument to the public. The perception being that since the species as a whole is thriving there is no threat

of extinction. Unbeknownst to many, modern practices have diminished the number of viable breeds and breed extinction is a real possibility.

Each year many breeds face the threat of extinction as populations dwindle to nonviable numbers. According to the Food and Agriculture Organization, FAO, approximately 20 percent of documented livestock breeds are at risk of extinction (Animal Genetics Resources Group, 2006), with 300 breeds having become extinct in the past 15 years (Cardellino, 2004). With each breed lost, a certain amount of genomic information is lost forever.

By definition, a breed consists of a group of animals sharing resemblance to one another and which produce offspring resembling them as well. Animals of the same breed also exhibit similar behavior or abilities (Dohner, 2001). These behaviors and abilities determined which breeds farmers valued for their operation, and as the agriculture world changed, so did the perceived value of a breed's qualities.

The push toward high-output operations forces many farmers to abandon heritage or local breeds in favor of industry breeds in order to remain competitive (Windig, 2010; National Research Council, 1993). Consumers support a market that places value on an animal only for how much meat, wool, milk, etc. it can produce (Taberlet, 2008). Breeds unable to keep up with industry standards are seen as less valuable (Windig, 2010), even though they may offer unique genetic traits not seen in industry breeds.

Intensive breeding systems have successively increased production traits while genetic variation and fitness have decreased (Villanueva et al., 2004). Breeds adapted to survive droughts, extreme temperatures and other environmental hardships have been

replaced with breeds adapted to ambient temperatures, constant attention and ad libitum access to food and water (NRC, 1993).

Many heritage breeds of pigs are able to forage for themselves, thrive on limited resources and care for their own young (Dohner, 2001). The modern pig is housed indoors, receives high quality feed, and has little maternal instinct. But the commercial pig serves as a fast-maturing source of cheap, abundant protein (National Pork Board, 2009), which limits the heritage breed to small operations or hobby farms where they are often raised to be sold into a niche market or for purely personal fulfillment. The lack of popularity at the large scale production level reduces numbers and hurts the chances of many heritage breeds surviving for future generations.

Breeds of Interest

In an attempt to understand how these breeds relate to one another and fit into the pork industry, it is important to know how they were developed and for what purpose.

Saddlebacks

Figure 1. British Saddleback.
Source. http://www.britishpigs.org.uk/breed_bs.htm

An example of challenges facing the conservation of breeds is lack of background information including pedigrees but not limited to basic breed origin. Referring to Mason's World Dictionary of Livestock Breeds, Types and Varieties, American

Saddleback is not listed as a breed; however American Essex, Essex, and several varieties of other Saddleback are recorded.

American Essex was developed in the USA from Black Essex which were imported in 1820. The American Essex was supposedly extinct until revived in 1967 in Texas as an experimental breed and renamed Guinea-Essex. The Black Essex was a variety of the Small Black breed, of England, which was the origin of the Essex as well as the American Essex. The Essex originated in England around 1918 and in 1967 was combined with the Wessex Saddleback to form the British Saddleback (Mason, 2002). A breed society was formed for the British Saddleback in 1996, and today it is considered a rare breed in the United States (American Livestock Breed Conservancy, 2011).

Terminology for breeds is sometimes loosely defined and the designation of "American" may simply differentiate American-born Saddlebacks as opposed to Saddlebacks born and raised in England.

Red Wattle

Figure 2. Red Wattle
Source. American Livestock Conservancy

The Red Wattle has a more obscure background, and its precise breed origins are unknown. It was developed around 1970 in Texas, from feral pigs residing in wooded areas in the Eastern part of the state (Mason, 2002). It is classified as rare in the United States by Mason's Dictionary and listed as critical by The Livestock Conservancy (TLC).

An animal is considered critical by TLC if there are fewer than 200 animals registered with their breed association each year and the global population is estimated at less than 2000.

Ossabaw Island

Figure 3. Ossabaw Island
Source. American Livestock Conservancy

Ossabaw Island hogs are named for the island from which they originate, Ossabaw Island. They exist on this island off the coast of Georgia in isolation as feral pigs. Though they have thrived on the island for many years, it is believed the pigs have a Spanish heritage. In 1986 a herd book was established for pigs removed from the island (Mason, 2002).

Though the Ossabaw is still raised for meat on small farms, pigs removed from the island are most likely to be found in research settings where they are used as models for non-insulin dependent diabetes. Ossabaw Island pigs are listed as critical by TLC.

Mulefoot

Figure 4. Mulefoot
Source. www.maveric9.com/recent/american-mulefoot-hogs

The Mulefoot is named for its unique hoof. Unlike most pigs which have a cloven hoof, the Mulefoot has a solid hoof, like a horse or mule. Other breeds possessing a solid hoof include the Choctaw and Casco de Mula.

Developed in the early 1900's in Missouri, it is believed to originate from a cross of Berkshire and feral Razorback (Mason, 2002). The breed was originally developed for its lard (Dohner, 2001). In a market that no longer places high value on lard, the Mulefoot has seen its numbers dwindle and is listed as nearly extinct by Mason's dictionary and critical by TLC.

Guinea Hog

Figure 5. Guinea Hog
Source. American Livestock Conservancy

The Guinea Hog is a small, black breed developed in the United States in Minnesota and Alabama. The breed originated from pigs of West African descent. Other names for the breed include: African Guinea, Guinea Forest Hog, and Gulf pig (Mason, 2002). Classified as critical by ALC, the American Guinea Hog association is exceptionally active in their attempts to raise awareness and learn more about their breed.

Duroc

Figure 6. Duroc
Source. www.ansi.okstate.edu

The Duroc was developed around 1872 by combining old Duroc of New York with Jersey Red of New Jersey (Mason, 2002). In the years following, it was established in other countries around the world. Though its original purpose was for lard (Dohner, 2001), it is one of only a few breeds used fairly extensively in current commercial pig operations.

Landrace

Figure 7. Landrace
Source. www.ansi.okstate.edu

Popular worldwide, the Landrace is utilized extensively in commercial operations. It was developed in Northwest Europe from native breeds in that area (Mason, 2002). Landrace pigs are characterized by their white skin, floppy ears and good maternal ability.

Large White

Figure 8. Large White
Source. www.ansi.okstate.edu

Like the Landrace, the Large White is quite popular with commercial producers and can be found throughout the world. Developed in England in the early 18th century, the Large White originated from a cross of Yorkshire and Chinese pigs (Mason, 2002). Large White pigs are characterized by their white skin, erect ears, large litter size and good maternal ability. Today, Yorkshire is synonymous with Large White.

Pietrain

Figure 9. Pietrain
Source. www.ansi.okstate.edu

Pietrain originated in Belgium around 1920 by crossing Bayeux with other local breeds. The Bayeux originated by combining Tamworth with another breed, possibly Normand. Developed in France, Normand was crossed extensively with Large White and

went extinct in 1937 (Mason, 2002). Today, Pietrain is found in many countries including the United States, but is more popular throughout Europe.

Tamworth

Figure 10. Tamworth
Source. www.ansi.okstate.edu

The Tamworth originates from England circa 1850. The Breed Society was established in the United States in 1923 (Mason, 2002). Tamworth is also known as Staffordshire. Little else is known about the origination of the Tamworth breed.

Today it is nearly extinct in Great Britain and listed as threatened in the United States, with a global population estimated at less than 5000.

Population size

In breeds with dwindling populations, effective population size becomes increasingly important (Kliman, 2008). Effective population size is essentially the number of individuals capable of contributing unique genetic material at an allele frequency equal to that of an ideal population, and predicts the rate of genetic drift within a population (Kliman, 2008). A reduced effective population size can lead to loss of neutral genetic variation, fixation of certain alleles, and increased rate of inbreeding (England, 2006). The effective population size is often much smaller than the census population of a breed (Kliman, 2008; Taberlet, 2008; England, 2006).

To understand why a discrepancy exists between census population (N) and effective population (N_e), it's important to understand factors affecting N_e and that smaller populations are more affected by these factors. Forces acting on N_e include: fluctuating population size, mutation, breeding sex ratio, and overlapping generations (Kliman, 2008). These factors are tied into the characteristics of an "ideal" population. An ideal population, where $N = N_e$, is characterized by equal numbers of males and females, all individuals being equally likely to reproduce, random mating, and number of breeding individuals remaining constant from generation to generation (Ollivier, 2004).

The simple equation used for calculation of N_e accounts for the male (N_m) and female (N_f) census populations within a group.

$$N_e = \frac{4 \times N_m \times N_f}{N_m + N_f}$$

Equation 1. Effective Population Size

It is immediately apparent how a domestic population would violate the characteristics of an ideal population. In livestock operations, a single male breeds many females, and if artificial insemination is used the ratio of males to females can become even more skewed. Due to selection by the producer, not all animals are likely to reproduce, and mating is generally not random. Since the producer is operating with the goal of meeting various production traits, breeding is carefully planned to maximize genetic potential in animals displaying desired traits.

Maintaining a high N_e is important for endangered breeds, but it is also important for more populous breeds. With industry breeds the population can number in the millions, and N_e is assumed to be high, but this is not always the case. Since these breeds are selected for uniformity, to thrive in specific environmental conditions, and often

artificially inseminated, genetic diversity can be severely reduced. For example, in 1999 the Holstein cattle population numbered 8.5 million in the United States, but the effective population was calculated at only 39 (Weigel, 2001). Conservation of genetic diversity to maintain N_e is important at all levels of production, whether the population numbers in the hundreds or the millions.

Farmers who attempt preservation of rare breeds are often challenged by both small populations and a lack of pedigree data. Without pedigree data, inbreeding, which increases homozygosity (Alderson, 1992, Wright, 2008) and decreases N_e (Kliman, 2008), is a real issue (Garcia-Gamez, 2012). Inbreeding and increased homozygosity can concentrate undesirable or deadly recessive abnormalities which reduce a breed's viability (Wright, 2008). Concentration of certain alleles and increased homozygosity leading to the reduced performance of an individual or a population is referred to as inbreeding depression (Silio, 2013). Knowing relationships enables farmers to plan matings to maintain genetic integrity within a population and reduce the instances of inbreeding depression. By reducing inbreeding, producers decrease the risk of concentrating undesirable homozygous alleles in a population.

Traditionally, relationships between animals are calculated by referring to each animal's pedigree. The coefficient of relationship, or R, estimates the percentage of alleles between two individuals which are identical by descent, IBD (Wright, 2008). The following formula is generally used to calculate relationships:

Equation 2. Coefficient of Relationship

Where is the coefficient of relationship, and *n* is the number of paths separating two individuals (Falconer). When two individuals are connected by more than one path, the coefficients are computed separately and then summed. Summing the values from separate paths accounts for a shared ancestor and increases relatedness due to inbreeding.

The inbreeding coefficient also gives the probability of two alleles at any given locus being identical by descent, but within an individual rather than between two individuals. Having alleles IBD creates a situation where an animal has increased levels of homozygosity as compared to levels expected in a non-inbred base population. Previously, inbreeding was discussed on a population basis, but in terms of an individual's inbreeding coefficient the following equation will be used.

Equation 3. Inbreeding Coefficient

Where F_x is the inbreeding coefficient, *n* is the number of paths between the common ancestor and the parents of an individual, and F_a is the inbreeding coefficient of the common ancestor (Falconer, 1996).

Producers use pedigrees to minimize matings between related individuals, which maximizes genetic diversity (Windig, 2010), and helps maintain a healthy effective population size. Until recently, producers relied solely on pedigrees to document relationships. But for rare and endangered breeds, pedigrees are often incomplete or missing. (Garcia-Gamez, 2012; Taberlet, 2008).

Through advances in genomics, it is now possible to genotype animals and estimate relatedness among a population of genotyped individuals (Garcia-Gamez, 2012; VanRaden, 2008; Windig, 2010). Genotyping enables farmers to select animals with

greater heterozygosity and use them for breeding. By maintaining heterozygosity, farmers are preserving genetic diversity and increasing the viability of the breed for future generations (Gill, 1992). Genotyping also recognizes individual alleles, making it possible to identify individuals possessing exceptionally rare or valuable alleles within a population (Lamberson, 1998). Such individuals are especially important to maintain genetic variation in small populations.

Single nucleotide polymorphisms, or SNPs, are variations in the DNA where one nucleotide base is replaced with another nucleotide. SNPs are the most common type of genetic marker, occurring once about every 300 base pairs, making them useful for characterizing many genetic traits (Zhu, 2004). Most SNPs are found in non-coding DNA where they have no effect. SNPs found within genes, however, can cause genotypic and phenotypic differences in an individual (Stachowiak, 2007). Recent studies have shown just how significant these differences can be.

In humans, especially, several SNPs have been identified as being closely related to certain traits or diseases. The range of diseases is vast and include cancer, heart disease, diabetes, and Alzheimer's (Altshuler, 2008). In livestock, researchers most often try to identify SNPs related to production (Goddard, 2009). As interest in SNPs grew, so did the ease of access to such information.

The development of SNP chips for livestock species has advanced the use of genetic testing on farms across the United States. When the process of collecting DNA can be as simple as plucking a few hairs, and the cost less than one hundred dollars, producers are seeing the advantages to getting their animals genotyped. Questions concerning an offspring’s sire can now be answered with genotyping, or an animal of

unknown origins can be genotyped and identified as a specific breed (McKay et al., 2008).

For swine, the Porcine 60k SNP chip is the popular choice for extensive genetic testing. Other chips are available which cover only a few hundred SNPs, but the 60k chip covers over 60,000 SNPs in a pig's DNA. Genotyping can then offer insight into an animal's heritage, genetic diversity and genetic merit. In terms of effective population size, such information is especially useful in a small population when other data is unavailable, such as pedigree records.

Endangered livestock breeds are classified by The Livestock Conservancy as critical when global populations fall below 2000 individuals (American Livestock Breed Conservancy, 2011). These breeds have fallen out of favor with either the public or producers for a variety of reasons, most related to the change in farming operations. As producers and organizations such as TLC attempt to save these critically endangered breeds, genetic information becomes increasingly important. In order for breeds to sufficiently recover and thrive, not merely survive, genetic diversity must be preserved, and new technologies such as genotyping can help save these breeds.

DNA testing can provide the history of an animal when paper records have been lost or were never even established. Most breeds facing extinction are old and did not have an established breed organization for many years, therefore accurate records are hard to come by. SNPs can be used in place of missing paperwork. More than just showing which copy of allele an animal received, SNPs provide insight into the individual's as well as the breed's history when appropriately analyzed.

Since SNPs occur at fairly regular intervals, they can be used to determine whether or not alleles at different loci are in linkage disequilibrium. Linkage disequilibrium, LD, occurs when alleles at different loci associate at a rate which differs from what is expected (Gaut, 2003, Slatkin, 2008). Linkage disequilibrium can reveal much about the history of a breed (Amaral, 2008) considering it can be influenced by such events as selection, genetic drift, inbreeding, changes in population size, and bottlenecks (Gaut, 2003; Slatkin, 2008).

LD can be calculated using the following equation:

Equation 4. Linkage Disequilibrium

Where D_{12} represents the LD between alleles 1 and 2, and is the frequency of sequences containing allele A at the one site and allele G at the second site. Marginal frequency of each allele at each site is represented by and (Gaut, 2003). If D is equal to zero, then no LD is present, and the alleles are assumed to be assorting randomly.

Another measure of LD is r^2 which may be calculated in PLINK and is a multiple correlation coefficient (Purcell, 2007). As is relates to D, r^2 is equal to D^2 divided by the product of the allele frequencies at the two loci (Gaut, 2003). Since r^2 represents a correlation, it may take any value between 0 and 1, with 0 being no correlation and 1 being completely correlated.

For small populations of interest, understanding a breed's history and genetics can prove invaluable to saving the breed for future generations. Whether a breed provides a function through milk, wool, or meat, or is simply aesthetically pleasing, these breeds make unique contributions to the diversity of production agriculture.

CHAPTER 2: RELATIONSHIPS AND INBREEDING COEFFICIENTS

Introduction

Farmers who attempt preservation of rare breeds are challenged by both small populations and a lack of pedigree data. Without pedigree data, inbreeding, which increases homozygosity (Wright, 2008; Alderson, 1992), can become a problem (Garcia-Gamez, 2012). Inbreeding and increased homozygosity can increase the frequency of undesirable or lethal homozygous recessive genotypes which reduce a breed's viability (Wright, 2008). Knowing relationships enables farmers to plan matings to maintain genetic integrity within a population by reducing inbreeding, thereby decreasing the prevalence of undesirable homozygous traits.

Producers use pedigrees to minimize matings between related individuals, which maximizes genetic diversity (Windig, 2010) and helps maintain a healthy effective population size, but for rare and endangered breeds, pedigrees are often incomplete or missing.

Through advances in genetics it is now possible to genotype animals and estimate relatedness among a population of genotyped individuals (Garcia-Gamez; Windig 2010; VanRaden). Genotyping could enable farmers to select less related or unrelated individuals for breeding, thereby preserving heterozygosity in the process. By maintaining heterozygosity, farmers are preserving genetic diversity and increasing the viability of the breed for future generations (Gill, 1992).

Materials and Methods

A selection of both heritage breeds and commercial breeds were utilized in this study. Breeds were chosen based on level of endangerment and ease of access. Ultimately samples or data were obtained for seven heritage breeds and four commercial breeds.

Producers were contacted via email to evaluate their ability and willingness to participate in this research project by providing hair samples from their pigs. Hair collection cards, provided by GeneSeek (Neogen Corporation, Lincoln, NE), were mailed to producers along with instructions for sample collection.

Hair samples were collected from Guinea, Ossabaw Island, Red Wattle, American Saddleback, and Mulefoot pigs and genotyped with the Porcine 60k SNP chip at GeneSeek in Nebraska. Hair samples could be plucked from any location on the animal. The follicle needed to be intact and at least fifteen hairs were collected from each animal to ensure enough tissue was available for genotyping.

Sample size varied from breed to breed but when possible, samples were obtained from individuals at various locations in an effort to get a more representative sample of the breed as a whole. The final count of individuals per breed was 13 Guinea, 10 Ossabaw Island, 5 Red Wattle, 2 American Saddleback, and 4 Mulefoot pigs.

Publicly available genotyping data were obtained for British Saddleback, Duroc, Landrace, Large White, Pietrain and Tamworth pigs (Goedbloed, 2012). The article, "Genome-wide single nucleotide polymorphism analysis reveals recent genetic introgression from domestic pigs into Northwest European wild boar populations" (Goedbloed, 2012), included data from 120 individuals representing six breeds. (Goedbloed, 2012).

Overall, information was collected for eleven breeds of pigs, consisting of seven heritage breeds and four commercial breeds. Sample size for each breed ranged from two to twenty individuals. After initial analysis, American Saddleback and British Saddleback were combined as one breed for a total of twenty-two Saddleback individuals.

For samples genotyped by Geneseek, resulting SNP data were converted to PED and MAP files for analysis in PLINK. Data obtained online required no conversion. It was available in PED and MAP files for use with PLINK.

PLINK (Purcell, 2013) is a program available for download online. It features a wide range of tools allowing whole genome association analysis (Purcell, 2007). By converting files to PED and MAP formats, they can be easily analyzed with PLINK. Source codes for analysis completed in PLINK are provided in Appendix 1.

PLINK was used to construct a genomic relationship matrix (GRM) for each breed. A GRM calculates the percent of alleles identical by descent between two individuals. A value of zero indicates no relationship and a value of one would indicate an identical twin or clone. For two non-inbred individuals, the highest expected value is 0.5 which would represent a parent-offspring or full-sib relationship. Individuals or SNPs not meeting the following criteria were removed: Minor allele frequency greater than 0.05 or call rate for individual greater than 0.90. After selection 29,887 SNPs remained. PLINK compares SNPs in order to calculate identity-by-descent for all pairs of individuals, which can be used to estimate relationships when pedigrees are unavailable.

Results

The following section contains results from GRM analysis.

Red Wattle

Table 1. Genomic Relationship Matrix for Red Wattle Individuals

	3-5	15-5	1-3	3-2	16-1
3-5	.	.	.	.	.
15-5	0.41	.	.	.	.
1-3	0.67	0.48	.	.	.
3-2	0.64	0.36	0.67	.	.
16-1	0.42	0.45	0.50	0.33	.

Table 1 summarizes the GRM for Red Wattle individuals. The value for relationship between individuals ranged from 0.36 to 0.67. Though samples were collected from individuals belonging to two separate farms, all individuals shared some level of relationship.

Ossabaw Island

Table 2. Genomic Relationship Matrix for Ossabaw Island Individuals

	OS-0153	Betty	Barney	Witchling	OS-1240	10-2sow	Wilma	OS-1207	Fred	Rip
OS-0153	.	.	.	.	.	.	.	.	.	.
Betty	0.52	.	.	.	.	.	.	.	.	.
Barney	0	0	.	.	.	.	.	.	.	.
Witchling	0.34	0.33	0.33	.	.	.	.	.	.	.
OS-1240	0.61	0.62	0.29	0.35	.	.	.	.	.	.
10-2sow	0.32	0.30	0.27	0.31	0.27	.	.	.	.	.
Wilma	0.68	0.66	0	0.33	0.67	0.33	.	.	.	.
OS-1207	0.66	0.59	0	0	0.65	0.29	0.66	.	.	.
Fred	0.54	0.63	0	0	0.63	0.29	0.61	0.66	.	.
Rip	0.33	0.30	0.26	0.35	0.33	0.49	0.34	0	0.29	.

Table 2 summarizes the GRM for Ossabaw Island individuals. The value for relationship between individuals for Ossabaw Island, ranged from 0.0 to 0.68. Samples

were collected from at least three different farms. Exact farm of origin information was not available for Ossabaw Island samples submitted by the American Guinea Hog Association.

Mule Foot

Table 3. Genomic Relationship Matrix Mulefoot Individuals

	Sow-25	Sow-21	Sow-22	Tubby-boar
Sow-25	.	.	.	.
Sow-21	0.66	.	.	.
Sow-22	0.65	0.71	.	.
Tubby-boar	0.75	0.69	0.71	.

Table 3 summarizes the GRM results for Mulefoot individuals. Values for relationship among Mulefoot individuals ranged from 0.65 to 0.75. Since Mulefoot samples were difficult to locate, all individuals sampled for this study resided on the same property. However, these values are still higher than would be expected even for a parent-offspring or full-sib calculation, making it likely that the parents of these were related at some level.

Guinea Hog

Table 4 summarizes the GRM for Guinea Hog individuals. Values for relationship in Guinea hogs ranged from 0 to 0.64. Two individuals were removed as they were duplicates of DNC A and AllBlack. Each sample had a unique identifier, but shared a 0.99 relationship with either DNC A or AllBlack. All samples were submitted by the American Guinea Hog Association. Farm information or location of animals was not provided. However, it is known that the samples provided for this research were collected from various farms across the country.

Table 4. Genomic Relationship Matrix Guinea Hog Individuals

	Blue	Wart	AllBlack	WartSow	BigBoar	Houdini	DNC A	RoseDNC	DNCJunior	SettyLilly	Samson	Killian	Barnett
Blue	.	.	.	.	.	.	.	.	.	.	.	.	.
Wart	0.42	.	.	.	.	.	.	.	.	.	.	.	.
AllBlack	0	0	.	.	.	.	.	.	.	.	.	.	.
WartSow	0	0.29	0	.	.	.	.	.	.	.	.	.	.
BigBoar	0	0.25	0	0.64	.	.	.	.	.	.	.	.	.
Houdini	0.16	0	0	0	0	.	.	.	.	.	.	.	.
DNC A	0	0	0	0	0	0	.	.	.	.	.	.	.
RoseDNC	0.20	0.11	0	0	0.22	0.51	0	.	.	.	.	.	.
DNCJunior	0.19	0.17	0	0.32	0.28	0	0.54	0.17	.	.	.	.	.
SettyLilly	0	0.32	0	0.56	0.59	0.26	0	0.25	0.31	.	.	.	.
Samson	0.14	0	0.26	0	0	0	0	0.09	0.17	0	.	.	.
Killian	0.16	0.18	0	0.35	0.39	0.32	0.37	0.35	0.43	0.38	0.12	.	.
Barnett	0.16	0.22	0	0.50	0.42	0	0.29	0.27	0.32	0.47	0	0.32	.

Table 5. Genomic Relationship Matrix Duroc Individuals

	DU20 M08	DU20 M01	DU20 M09	DU20 M11	DU20 M12	DU20 M02	DU20 M03	DU21 M04	DU21 M01	DU21 M02	DU21 M03	DU22 M01	DU22 M05	DU22 M06	DU22 M11	DU23 M03	DU23 M01	DU23 M02	DU23 M04	DU23 M05
DU20 M08	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
DU20 M01	0.37	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
DU20 M09	0.30	0.36	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
DU20 M11	0.33	0.34	0.29	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
DU20 M12	0.31	0.29	0.27	0.34	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
DU20 M02	0.32	0.67	0.37	0.29	0.26	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
DU20 M03	0.31	0.36	0.35	0.27	0.28	0.34	.	.	.	.	.	.	.	.	.	.	.	.	.	.
DU21 M04	0	0	0	0	0.21	0	0	.	.	.	.	.	.	.	.	.	.	.	.	.
DU21 M01	0	0.23	0	0.22	0.19	0	0	0.27	.	.	.	.	.	.	.	.	.	.	.	.
DU21 M02	0	0.21	0.22	0.18	0.24	0	0.21	0.24	0.22	.	.	.	.	.	.	.	.	.	.	.
DU21 M03	0.22	0.23	0.21	0	0.23	0	0	0.29	0.27	0.23	.	.	.	.	.	.	.	.	.	.
DU22 M01	0	0	0	0	0.20	0	0	0.24	0.24	0.21	0.24	.	.	.	.	.	.	.	.	.
DU22 M05	0	0	0	0.25	0.22	0	0	0.19	0.25	0.23	0.22	0.29	.	.	.	.	.	.	.	.
DU22 M06	0.23	0.21	0	0.22	0.22	0	0.22	0.26	0.29	0.25	0.29	0.25	0.27	.	.	.	.	.	.	.
DU22 M11	0	0.20	0	0.21	0.21	0	0	0.24	0.20	0.21	0.27	0.25	0.24	0.27	.	.	.	.	.	.
DU23 M03	0.34	0.31	0.28	0.28	0.29	0.27	0.26	0.19	0.17	0.19	0.22	0.21	0.21	0.20	0.21	.	.	.	.	.
DU23 M01	0.29	0.28	0.23	0.27	0.36	0.26	0.27	0.22	0.21	0.21	0.23	0.23	0.22	0.21	0.21	0.40	.	.	.	.
DU23 M02	0.22	0.22	0.23	0.22	0.25	0.22	0.20	0.19	0.20	0.19	0.20	0.18	0	0.24	0.20	0.22	0.31	.	.	.
DU23 M04	0.22	0.28	0.29	0.27	0.25	0.28	0.28	0.21	0.22	0.19	0.23	0.22	0.23	0.21	0.22	0.29	0.33	0.26	.	.
DU23 M05	0.21	0.22	0.21	0.22	0.25	0	0.21	0.21	0.19	0.25	0.20	0.22	0.20	0.23	0.21	0.26	0.29	0.31	0.25	.

Table 5 summarizes the GRM for Duroc individuals. Relationship values for Duroc individuals ranged from 0 to 0.67. All data for Duroc individuals were obtained from an online source (Goedbloed, 2012).

Table 6 summarizes the GRM for Landrace individuals. Relationship values for landrace ranged from 0 to 0.44. Again, these values were obtained from an online source (Goedbloed, 2012).

Table 7 summarizes the GRM for Pietrain individuals. Relationship values for Pietrain individuals ranged from 0 to 0.37. Data for these individuals were obtained from an online source (Goedbloed, 2012).

Table 8 summarizes the GRM for Saddleback individuals. Relationship values for Saddleback individuals ranged from 0 to 0.73. Data for these individuals were obtained from an online source (Goedbloed, 2012) and from the ALC.

Table 9 summarizes the GRM for Tamworth individuals. Relationship values for Tamworth individuals ranged from 0.35 to 0.64. Data for these individuals were obtained from an online source (Goedbloed, 2012).

Table 10 summarizes the GRM for Large White individuals. Relationship values for Large White individuals ranged from 0 to 0.46. Data for these individuals were obtained from an online source (Goedbloed, 2012).

Table 6. Genomic Relationship Matrix Landrace Individuals

	LR22M16	LR22M17	LR22M18	LR22M19	LR20M02	LR20M03	LR21M04	LR21M02	LR21M03	LR23M01	LR23M02	LR23M03	LR30F08	LR30F09	LR30F10	LR30F11	LW27F03	LW27F02	LW27M11	LW27F12
LR22M16	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
LR22M17	0.18	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
LR22M18	0.24	0.20	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
LR22M19	0.24	0.19	0.30	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
LR20M02	0	0	0	0	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
LR20M03	0	0	0	0	0.44	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
LR21M04	0	0	0	0	0	0	.	.	.	.	.	.	.	.	.	.	.	.	.	.
LR21M02	0	0	0	0	0	0	0.08	.	.	.	.	.	.	.	.	.	.	.	.	.
LR21M03	0	0	0	0	0	0	0.07	0.06	.	.	.	.	.	.	.	.	.	.	.	.
LR23M01	0.16	0	0	0	0	0	0	0	0	.	.	.	.	.	.	.	.	.	.	.
LR23M02	0	0	0	0	0	0	0	0	0	0.18	.	.	.	.	.	.	.	.	.	.
LR23M03	0.11	0	0	0	0	0	0	0	0	0.18	0.15	.	.	.	.	.	.	.	.	.
LR30F08	0	0	0	0	0	0	0	0	0	0	0	0	.	.	.	.	.	.	.	.
LR30F09	0	0	0	0	0	0	0	0	0	0	0	0	0.15	.	.	.	.	.	.	.
LR30F10	0	0	0	0	0	0	0	0	0	0	0	0	0.18	0.16	.	.	.	.	.	.
LR30F11	0	0	0	0	0	0	0	0	0	0	0	0	0.21	0.14	0.27	.	.	.	.	.
LW27F03	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.	.	.	.
LW27F02	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.14	.	.	.
LW27M11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.31	0.20	.	.
LW27F12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.20	0.14	0.23	

Table 7. Genomic Relationship Matrix Pietrain Individuals

	PI03F03	PI03F04	PI03F06	PI03F07	PI03F08	PI20U02	PI20U01	PI21M04	PI21M13	PI21M06	PI21M15	PI21M01	PI22M01	PI22M02	PI22M03	PI23F13	PI23F07	PI23F03	PI23M08	PI23M07
PI03F03	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
PI03F04	0.13	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
PI03F06	0.34	0	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
PI03F07	0.22	0.28	0.18	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
PI03F08	0.37	0.13	0.23	0.16	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
PI20U02	0.09	0.14	0.12	0	0.16	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
PI20U01	0.12	0	0.14	0.13	0	0.23	.	.	.	.	.	.	.	.	.	.	.	.	.	.
PI21M04	0.16	0	0.14	0.13	0.12	0.11	0.14	.	.	.	.	.	.	.	.	.	.	.	.	.
PI21M13	0.12	0	0	0	0.15	0	0	0.15	.	.	.	.	.	.	.	.	.	.	.	.
PI21M06	0.15	0.08	0.12	0.14	0.13	0.13	0.10	0.13	0.12	.	.	.	.	.	.	.	.	.	.	.
PI21M15	0.16	0.14	0.14	0.16	0.12	0	0	0	0.14	0.09	.	.	.	.	.	.	.	.	.	.
PI21M01	0.11	0.12	0.19	0.11	0.09	0	0.07	0.15	0.15	0.19	0.21	.	.	.	.	.	.	.	.	.
PI22M01	0.15	0.13	0.15	0.11	0.12	0.11	0.11	0.11	0.13	0.15	0	0.10	.	.	.	.	.	.	.	.
PI22M02	0.10	0	0.10	0.11	0.11	0	0	0	0	0.10	0	0	0.18	.	.	.	.	.	.	.
PI22M03	0.07	0	0.09	0	0.14	0.10	0.10	0.09	0	0	0.09	0	0.16	0.57	.	.	.	.	.	.
PI23F13	0.06	0	0.10	0	0	0	0	0	0	0.09	0	0	0	0	0	.	.	.	.	.
PI23F07	0.07	0	0	0	0.10	0	0	0	0	0	0	0	0	0	0	0.20	.	.	.	.
PI23F03	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.22	0.28	.	.	.
PI23M08	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.21	0.15	0.18	.	.
PI23M07	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.19	0.23	0.18	0.19	.

Table 8. Genomic Relationship Matrix Saddleback Individuals

	Honeybun	Dinah	BS01_M18	BS01_M33	BS01_M36	BS01_M37	BS01_M44	BS01_F17	BS01_M01	BS01_M04	BS01_F29	BS01F02	BS01F03	BS01F06	BS01F10	BS01F11	BS01F13	BS01F15	BS01F25	BS01F28	BS01F31	BS01F34
Honeybun	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Dinah	0.16	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
BS01_M18	0	0	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
BS01_M33	0.18	0	0	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
BS01_M36	0.16	0.13	0	0	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
BS01_M37	0.15	0.13	0	0.15	0.30	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
BS01_M44	0.11	0.15	0	0	0	0.17	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
BS01_F17	0.18	0.13	0.10	0.27	0.14	0.17	0.23	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
BS01_M01	0.13	0	0	0.21	0.14	0.12	0	0.23	.	.	.	.	.	.	.	.	.	.	.	.	.	.
BS01_M04	0.18	0.17	0	0.20	0.16	0.16	0	0.23	0.42	.	.	.	.	.	.	.	.	.	.	.	.	.
BS01_F29	0.15	0	0	0.18	0.16	0.15	0	0.13	0.28	0.34	.	.	.	.	.	.	.	.	.	.	.	.
BS01F02	0.14	0	0	0.22	0.15	0.12	0	0.21	0.67	0.43	0.30	.	.	.	.	.	.	.	.	.	.	.
BS01F03	0.18	0.12	0.12	0.20	0.14	0.21	0	0.29	0.31	0.22	0.14	0.22	.	.	.	.	.	.	.	.	.	.
BS01F06	0.18	0.12	0	0.20	0.15	0.14	0	0.27	0.55	0.41	0.31	0.48	0.29	.	.	.	.	.	.	.	.	.
BS01F10	0.12	0.12	0.15	0	0	0.15	0	0.13	0	0.12	0.11	0.11	0.12	0.11	.	.	.	.	.	.	.	.
BS01F11	0.12	0.14	0.12	0	0	0.11	0	0.08	0	0.10	0.09	0	0.10	0.10	0.44	.	.	.	.	.	.	.
BS01F13	0.15	0.14	0.12	0	0.12	0.12	0	0.08	0	0.11	0	0	0.10	0.10	0.44	0.73	.	.	.	.	.	.
BS01F15	0.13	0.11	0	0	0	0	0	0.13	0	0.14	0	0	0.13	0.13	0.21	0.20	0.25	.	.	.	.	.
BS01F25	0.13	0.11	0	0.32	0.14	0.15	0.17	0.24	0.33	0.29	0.41	0.26	0.23	0.30	0.10	0.08	0.08	0	.	.	.	.
BS01F28	0.16	0	0	0.36	0	0.18	0	0.29	0.19	0.17	0.18	0.17	0.21	0.18	0.14	0	0	0	0.36	.	.	.
BS01F31	0	0	0	0.35	0	0	0	0.25	0	0.18	0	0	0.18	0	0	0	0	0	0.32	0.44	.	.
BS01F34	0.16	0	0	0.49	0	0.11	0	0.20	0.26	0.19	0.19	0.27	0.18	0.26	0	0	0	0	0.33	0.26	0.26	.

Table 9. Genomic Relationship Matrix Tamworth Individuals

	TA01F02	TA01F04	TA01F07	TA01F08	TA01F09	TA01F10	TA01F13	TA01F14	TA01F16	TA01F18	TA01M03	TA01M06	TA01M15	TA01M17	TA01M20	TA01M23	TA01M26	TA01M27	TA01M32	TA01M34
TA01F02	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
TA01F04	0.39	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
TA01F07	0.38	0.40	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
TA01F08	0.43	0.39	0.63	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
TA01F09	0.45	0.46	0.60	0.61	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
TA01F10	0.38	0.38	0.51	0.43	0.49	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
TA01F13	0.42	0.40	0.43	0.52	0.47	0.41	.	.	.	.	.	.	.	.	.	.	.	.	.	.
TA01F14	0.35	0.40	0.44	0.42	0.45	0.45	0.51	.	.	.	.	.	.	.	.	.	.	.	.	.
TA01F16	0.37	0.43	0.49	0.48	0.59	0.45	0.44	0.43	.	.	.	.	.	.	.	.	.	.	.	.
TA01F18	0.38	0.43	0.55	0.50	0.60	0.54	0.43	0.47	0.64	.	.	.	.	.	.	.	.	.	.	.
TA01M03	0.44	0.69	0.39	0.38	0.45	0.37	0.40	0.39	0.46	0.41	.	.	.	.	.	.	.	.	.	.
TA01M06	0.46	0.45	0.58	0.76	0.64	0.45	0.53	0.45	0.52	0.54	0.42	.	.	.	.	.	.	.	.	.
TA01M15	0.40	0.39	0.40	0.44	0.49	0.36	0.37	0.46	0.43	0.42	0.35	0.49	.	.	.	.	.	.	.	.
TA01M17	0.39	0.39	0.48	0.43	0.59	0.42	0.46	0.40	0.72	0.61	0.42	0.50	0.37	.	.	.	.	.	.	.
TA01M20	0.41	0.38	0.40	0.42	0.45	0.36	0.42	0.40	0.44	0.42	0.40	0.44	0.44	0.43	.	.	.	.	.	.
TA01M23	0.37	0.45	0.41	0.41	0.49	0.40	0.51	0.45	0.45	0.46	0.48	0.43	0.41	0.45	0.49	.	.	.	.	.
TA01M26	0.42	0.48	0.49	0.46	0.47	0.48	0.47	0.48	0.45	0.47	0.49	0.47	0.47	0.42	0.48	0.56	.	.	.	.
TA01M27	0.40	0.43	0.42	0.44	0.45	0.42	0.43	0.44	0.44	0.43	0.43	0.44	0.45	0.38	0.49	0.67	0.56	.	.	.
TA01M32	0.38	0.46	0.40	0.44	0.53	0.43	0.43	0.42	0.40	0.43	0.48	0.47	0.38	0.44	0.39	0.46	0.41	0.40	.	.
TA01M34	0.45	0.48	0.44	0.44	0.52	0.49	0.43	0.42	0.47	0.50	0.48	0.45	0.47	0.45	0.43	0.51	0.52	0.50	0.47	.

Table 10. Genomic Relationship Matrix Large White Individuals

	LW20M02	LW20M01	LW21M01	LW21M02	LW22M01	LW22M11	LW22M19	LW22M04	LW22M12	LW22M05	LR29F11	LR29M01	LR29F10	LR29F01	LR29M13	LR29M14	LR29F03	LR29F14	LW31_B3	LW31_CQ
LW20M02	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
LW20M01	0	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
LW21M01	0	0	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
LW21M02	0	0	0.21	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
LW22M01	0	0	0	0	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
LW22M11	0	0	0	0	0.15	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
LW22M19	0	0	0	0	0.13	0.23	.	.	.	.	.	.	.	.	.	.	.	.	.	.
LW22M04	0	0	0	0	0.09	0.10	0.09	.	.	.	.	.	.	.	.	.	.	.	.	.
LW22M12	0	0	0	0	0.19	0.11	0.14	0.11	.	.	.	.	.	.	.	.	.	.	.	.
LW22M05	0	0	0	0	0.14	0.20	0.19	0.12	0.13	.	.	.	.	.	.	.	.	.	.	.
LR29F11	0	0	0	0	0	0	0	0	0	0	.	.	.	.	.	.	.	.	.	.
LR29M01	0	0	0	0	0	0	0	0	0	0	0.23	.	.	.	.	.	.	.	.	.
LR29F10	0	0	0	0	0	0	0	0	0	0	0.27	0.21	.	.	.	.	.	.	.	.
LR29F01	0	0	0	0	0	0	0	0	0	0	0.20	0.27	0.21	.	.	.	.	.	.	.
LR29M13	0	0	0	0	0	0	0	0	0	0	0.21	0.22	0.19	0.35	.	.	.	.	.	.
LR29M14	0	0	0	0	0	0	0	0	0	0	0.28	0.27	0.16	0.29	0.28	.	.	.	.	.
LR29F03	0	0	0	0	0	0	0	0	0	0	0.27	0.28	0.24	0.25	0.22	0.28	.	.	.	.
LR29F14	0	0	0	0	0	0	0	0	0	0	0.27	0.23	0.23	0.27	0.21	0.23	0.46	.	.	.
LW31_B3	0	0	0.18	0.14	0.10	0	0	0	0	0	0	0	0	0	0	0	0	0.08	.	.
LW31_CQ	0	0	0	0	0.09	0	0.08	0	0	0	0	0	0	0	0	0	0	0	0	.

PLINK also contains a method for calculating inbreeding coefficients. The data set was first pruned to only include SNPs in approximate linkage equilibrium. By excluding SNPs in linkage disequilibrium, bias due to selection is eliminated. After pruning, 9130 SNPs were removed. The remaining data were analyzed in Plink. The resulting F value is the inbreeding coefficient estimate (Purcell, 2013).

After estimating relationships between individuals and inbreeding coefficients, an average level of relatedness and inbreeding coefficient was calculated for each breed. Using an average for each breed allowed comparison across breeds.

Breed averages and standard errors for relationships between individuals and inbreeding coefficients were obtained by fitting relationships between individuals within breed and inbreeding coefficients to a model including breed as a categorical fixed effect (SAS® Version 9.2, SAS Institute Inc., Cary, NC). Code used to complete analysis in SAS can be found in Appendix 2.

Table 11 summarizes average relationships (R) between individuals (n) within a breed, and average inbreeding coefficient (F) of individuals within a breed. American Saddleback and British Saddleback showed relatedness across the two breeds, so they were combined. The model was significant (P-value <0.0001) and significant differences across breeds are indicated by superscripts ($\alpha \leq 0.05$).

Table 11. Breed Relationship (R) and Inbreeding Coefficient (F) Averages

Breed	n	R	F
Guinea	13	0.17^{a}	0.26^{a}
Ossabaw Island	10	0.37^{b}	0.47^{b}
Red Wattle	5	0.49^{bc}	0.28^{a}
Saddleback	22	0.14^{a}	0.15^{d}
Mulefoot	4	0.69^{d}	0.39^{ab}
Duroc	20	0.21^{a}	0.25^{a}
Landrace	20	0.03^{e}	0.15^{c}
Large White	20	0.05^{ef}	0.15^{c}
Pietrain	20	0.07^{f}	0.12^{c}
Tamworth	20	0.46^{c}	0.41^{b}
Values in columns with no superscripts in common are significantly different ($\alpha \leq 0.05$)			

Mulefoot individuals showed the highest level of relatedness, while Landrace individuals had the lowest level of relatedness. These numbers indicate that, on average, two individuals selected at random from the Mulefoot breed are more likely to have some level of relatedness than two individuals selected from Landrace.

In the table, F is an average of inbreeding coefficients for all individuals within a breed. F is highest for heritage breeds and lower in commercial breeds. Ossabaw Island has the highest value (0.47) which is expected considering all Ossabaw Island hogs originate from the isolated herd on Ossabaw Island.

As expected, popular breeds (Landrace, Large White, and Duroc) exhibit lower levels of R between individuals, on average as compared to R between individuals of endangered breeds, especially Ossabaw Island, Red Wattle, Mulefoot and Tamworth. Following a similar pattern, F is high for Ossabaw Island, Tamworth, and Mulefoot, and low for Large White and Landrace. While less common in the United States, Pietrain is a popular breed in Europe which likely accounts for low R and F values.

Having complete pedigrees and large populations allows commercial breeds to maintain low levels of R and F within a population. Producers are able to draw from a

wide selection of viable animals when making breeding decisions for their herd. While some levels of relatedness are present, they are not maintained at the levels seen in the heritage breeds. Extensive records for commercial animals allow producers to make educated decisions and mate individuals to emphasize genetic merit, while avoiding matings between close relatives.

For heritage type breeds, lack of popularity means fewer individuals to select among, and, within a viable populous, even fewer have known pedigrees. Even if animals are purchased from another farm, without complete records, there is no guarantee that the animals will be unrelated to a producer's current stock. As seen with the "American" Saddlebacks, even British Saddlebacks on a different continent were related to them. Distance between farms is not necessarily a factor in predicting whether or not two individuals will be related.

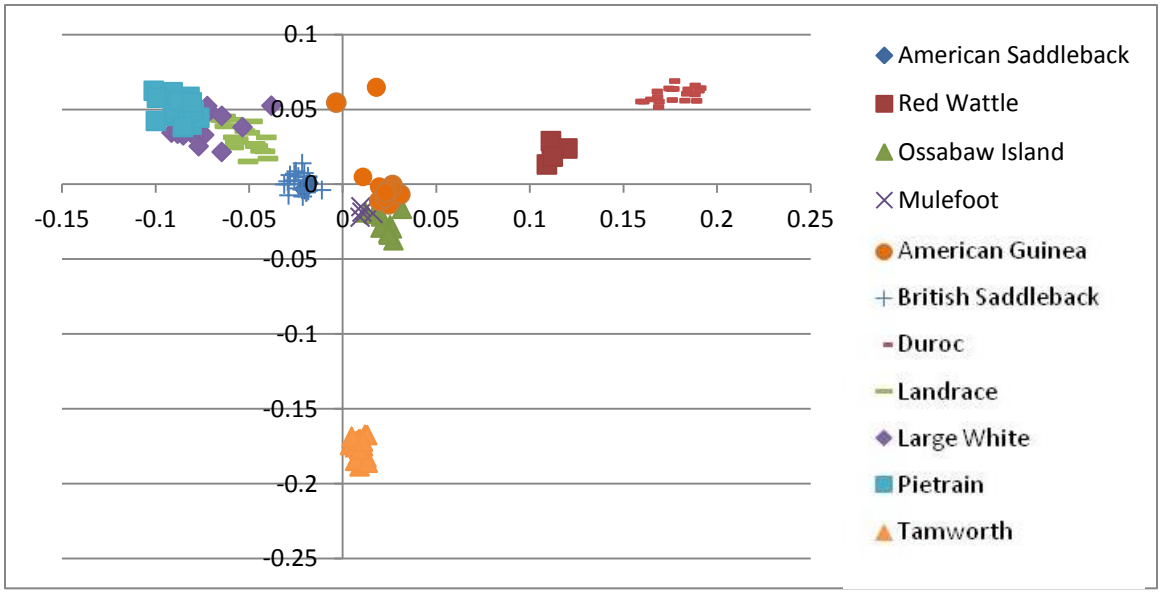
CHAPTER 3: MULTIDIMENSIONAL SCALING

Multidimensional scaling, MDS, allows for comparisons between things that are difficult to compare. Multidimensional simply indicates the data points are plotted in more than one dimension (i.e. two-dimensional) and scaling refers to the adjustment of the distances between points based on how well they correlate to one another (Borg et al., 2013; Cox, 2001). Instead of having a chart with a series of numerical correlation values, the MDS plot provides a way to visualize data and see patterns (Borg et al., 2013; Cox, 2001). Since the MDS plot is constructed from correlations, all comparisons are made relative to one another with points closer together being more highly correlated than points farther apart (Borg et al., 2013; Cox, 2001).

The MDS plot was prepared in PLINK by using whole genome SNP data to produce pairwise identity by state distance. (Borg et al., 2013; Cox, 2001). This plot shows relative relationships between breeds. Breeds clustered more closely together (i.e. Landrace, Pietrain, and Large White) share a more similar genetic makeup than breeds plotted farther apart.

Such a plot can be used to evaluate which breeds might possess unique traits as compared to commercial breeds. For example, outlying breeds such as Tamworth, Red Wattle and Duroc may carry unique genes not seen in breeds sharing a similar genetic makeup.

Plot 1. Multidimensional scaling plot of pig breeds



Two Guinea hog outliers are evident from Plot 1. One of these hogs, Samson, was of older bloodlines and found to be less related to the other Guinea samples, as was the other outlier. It may be that these two hogs share similar bloodlines from the founders of the Guinea hogs. The other Guineas were from more recent generations and they all group closely together.

From this plot, there appear to be six distinct groups, or clusters of individuals. The first group consists of Pietrain, Large White and Landrace. These breeds may share origin. The second group includes British Saddleback and American Saddleback. These plotted on top of one another and were ultimately combined for this study into the single breed of Saddleback. These are located near groups 1 and 3 but are obviously their own unique group, having no overlap with other breeds. The third group consists of three breeds, Mulefoot, Guinea and Ossabaw Island. Individuals from these breeds are plotted

exceptionally close together, and in some cases are overlapping. As with group one, these three breeds share some physical characteristics such as color and general conformation. The fourth and fifth groups consist of Red Wattle and Duroc respectively. These two breeds appear to be plotted closer to each other than to the rest of the breeds. While there is no overlap between these two, they do share similar characteristics and may have had common origin. The sixth group consists of the Tamworth breed. The Tamworth individuals are located toward the bottom of the plot in a tight grouping. They are plotted well away from all other breeds in this diagram.

In reference to this particular MDS plot, Red Wattle, Duroc and Tamworth appear to be the most different from all other breeds. Since the Red Wattle and Duroc fall close to one another, it is likely they share a similar genetic makeup. Tamworth draws the most interest as it is located in an isolated portion of the plot and does not appear to fall in line with any other breeds. Tamworth individuals could possess exceptionally unique alleles not carried by individuals in the other breeds. Ultimately, the MDS plot provides insight into the similarities of these breeds without explaining what those similarities, or differences, might be.

CHAPTER 4: EFFECTIVE POPULATION SIZE AND LINKAGE DISEQUILIBRIUM

Using estimates of census population size from The Livestock Conservancy's website, effective population size can be roughly estimated for the heritage breeds of pigs in this study. In addition to census numbers, the average breeding rate of one boar to twenty sows will be used. While this does not account for artificial insemination, which would skew the ratio, it is a reasonable estimate given that many heritage operations are small and may not be utilizing as many genetic tools as a large, commercial operation.

The following equation will be used to calculate effective population size.

$$N_e = \frac{4 \times N_m \times N_f}{N_m + N_f}$$

The following table summarizes population information for all heritage breeds of pigs, which assumes the census population is actually a count of breeding individuals. The 1:20 ratio is used for number of boars to sows.

Table 12. Effective Population Size

Breed	**N_C**	**N_m**	**N_f**	**N_e**
Saddleback	600	30	570	114
Guinea	4500	225	4275	855
Red Wattle	4000	200	3800	760
Tamworth	5000	250	4750	950
Mulefoot	1500	75	1425	285
Ossabaw Island	2000	100	1900	380

The second table assumes the census population is a count of all animals within the breed, and not just those given the opportunity to reproduce. The population is assumed to be fifty percent female and fifty percent male, and the ratio of 1 boar to 20 sows is again used. Now a certain portion of the census population is eliminated completely since those individuals would not be contributing their genetic material through reproduction.

Table 13. Effective Population Size Adjusted

Breed	**N_C**	**Males**	**N_m**	**N_f**	**N_e**
Saddleback	600	300	15	300	57
Guinea	4500	2250	113	2250	431
Red Wattle	4000	2000	100	2000	381
Tamworth	5000	2500	125	2500	477
Mulefoot	1500	750	38	750	145
Ossabaw Island	2000	1000	50	1000	191

Linkage Disequilibrium

Linkage disequilibrium, LD, occurs when alleles at different loci associate at a rate which differs from what is expected (Gaut, 2003; Slatkin). Linkage disequilibrium can reveal much about the history of a breed (Amaral) considering it can be influenced by such events as selection, genetic drift, inbreeding, changes in population size, and bottlenecks (Gaut, 2003; Slatkin).

A breed having a higher number of SNPs in LD might correspond to a lower effective population size since LD is affected by selection and inbreeding. Both of these

factors can have negative effects on N_e. The r^2 value is the multiple correlation coefficient (Purcell, 2007). This value is an indicator of the level of association between two SNPs and ranges from 0 to 1, with 0 being no correlation and 1 representing complete linkage.

The following table summarizes the number of SNPs by breed found to be in complete LD ($r^2 = 1$) and the average r^2 for all SNPs within a breed.

Table 14. Count of Single Nucleotide Polymorphisms in linkage disequilibrium and r^2

Breed	SNPs in LD	r^2
Red Wattle	34,693	0.72
Guinea	16,988	0.60
Mulefoot	35,009	0.77
Ossabaw	24,514	0.66
Tamworth	19,562	0.71
Pietrain	6,615	0.57
Large White	6,024	0.55
Landrace	6,102	0.54
Duroc	11,737	0.62
Saddleback	8,571	0.55

Red Wattle, Mulefoot and Ossabaw have the highest number of SNPs in complete LD. Tamworth, Guinea and Duroc have the next highest number, and Saddleback, Pietrain, Landrace and Large White have the lowest numbers of SNPs in LD.

The values for Saddleback are surprising given the breed's low effective population size and status as an endangered heritage breed. Higher levels of LD were expected given these circumstances.

CHAPTER 5: UNIQUE SINGLE NUCLEOTIDE POLYMORPHISMS

Single nucleotide polymorphisms, or SNPs, are variations in the DNA where one nucleotide base is replaced with another. Single nucleotide polymorphisms are the most common type of genetic marker, occurring once about every 300 base pairs, making them useful for characterizing many genetic traits (Zhu). Most SNPs are found in non-coding DNA where they have no phenotypic effect. SNPs found within genes, however, can cause genotypic and phenotypic differences in an individual (Stachowiak). Recent studies have shown just how significant these differences can be.

In humans, several SNPs have been identified as being closely related to certain traits or diseases. These diseases are vast and include cancer, heart disease, diabetes, and Alzheimer's (Altshuler, 2008). In livestock, researchers most often try to identify SNPs related to production (Goddard, 2009). As interest in SNPs grew, so did the ease of access to such information.

The development of SNP chips for livestock species has advanced the use of genetic testing on farms across the United States. When the process of collecting DNA can be as simple as plucking a few hairs, and the cost less than one hundred dollars, producers are seeing the advantages to getting their animals genotyped. Questions concerning an offspring's sire can now be answered with genotyping or an animal of unknown origins can be genotyped and identified as a specific breed (McKay et al., 2008).

The following table contains the number of SNPs with a minor allele frequency, MAF, of 1 for a given breed. For each of these SNPs, MAF equaled 1 for only one breed,

so there is no overlap of SNPs across breeds. Additionally, overall MAF was reported for each breed. This is the frequency of the minor allele across all SNPs for each breed.

Table 15. Minor allele frequency

Breed	Count of SNPs with MAF=1	Average MAF
Duroc	152	0.20
Red Wattle	2406	0.24
Ossabaw	1294	0.23
Mulefoot	4830	0.23
Guinea	49	0.25
Saddleback	605	0.24
Landrace	2	0.20
Large White	2	0.21
Pietrain	34	0.21
Tamworth	488	0.16

Tamworth had the lowest average MAF and Guinea had the highest. Landrace and Large White only had two SNPs where the MAF was equal to 1. Mulefoot had the highest number of homozygous minor alleles with 4,830. Since the Mulefoot individuals were highly inbred, that could be a contributing factor to the high level of homozygous minor alleles.

These results say something about the design of the Porcine 60k SNP chip, considering it was created based on SNPs known to be polymorphic in commercial

breeds, i.e. Landrace and Large White. Therefore it is expected that few SNPs for commercial breeds would be monomorphic.

Taking this into consideration, it is surprising that Guinea hogs had only 49 SNPs with a MAF of 1. Considering Guinea hogs are an endangered population and the individuals sampled for this project had varying levels of inbreeding, more SNPs were expected to have a MAF of 1. This could indicate that Guinea hogs are more genetically diverse than other heritage breeds, and their chances of preservation remain high.

CHAPTER 6: DISCUSSION

Coefficients of Relationship and Inbreeding

The research presented here shows SNP information is a valuable aspect of the conservation of endangered breeds of swine. In recent years, many projects have illustrated the potential SNP data holds for unlocking the history of a species as well as its current state. Some have used SNPs to determine inbreeding and relationships among individuals with the focus often on heritage or rare breeds of pigs (Lopes, 2013; Saura, 2013; Silio, 2013; Herrero-Medrano, 2012).

Saura et al. (2013), estimated inbreeding in ancient Iberian pigs. The endangered Iberian pig plays an important role in the Mediterranean area. Researchers calculated inbreeding coefficients from genealogical data as well as from molecular data in the form of SNPs. They found inbreeding coefficients to be about twice as high when calculated from SNPs versus pedigree.

Average molecular inbreeding coefficient was 0.81 while average genealogical inbreeding coefficient was 0.39 for Iberian pigs. The researchers performed a regression comparing molecular to genealogical coancestry, or kinship, values and determined molecular coancestry was a good indicator of genealogical coancestry. So if the researchers had not had access to extensive pedigrees, as is the case for most rare breeds, the molecular calculations alone would be good indicators of relatedness between individuals.

Lopes et al. (2013) analyzed SNPs for 1,565 individuals from three commercial pig populations. As with Saura et al. (2013), this study also had access to complete pedigrees, allowing genealogical and molecular comparisons.

Researchers looked at three lines of pigs, and found lines 2 and 3 to have higher molecular inbreeding coefficients as compared to pedigree inbreeding coefficient. For line 1, molecular coefficients were lower than pedigree inbreeding coefficients. For lines 1, 2 and 3, average molecular inbreeding coefficient across all SNPs were calculated at -0.01, 0.09 and 0.06 respectively. As compared to the present study, results are similar in that the commercial breeds had lower mean inbreeding coefficients as compared to heritage breeds. Lopes et al. (2013) does not indicate which breeds were used just that they are commercial breeds. Even though this is a European paper, it is likely some Landrace or Large White lines were included, as these breeds are also popular commercially in Europe.

Herrero-Medrano et al. (2012) researched the Chato Murciano pig breed which is found in an isolated region of Spain. Samples were taken from eight different farms and results focused on the differences among farms. The research reports average heterozygosity as compared to expected heterozygosity across all SNPs. Overall, the average heterozygosity, or genetic diversity, was found to be in range with other European breeds. Since the researchers reported average heterozygosity rather than inbreeding coefficients, it is not possible to make direct numerical comparisons to this thesis research.

The researchers found differences among farms indicating management and breeding practices could influence the levels of genetic diversity maintained within an isolated population. This is interesting in that it gives hope to preserving endangered breeds in the U.S. through changes is management or breeding programs. For being an

isolated and endangered population, the Chato Murciano appears to have maintained genetic diversity and survive as a viable population.

Silio et al. (2013) compared pedigree calculated inbreeding coefficients to runs of homozygous SNPs for 64 Iberian pigs. The individuals were sampled from experimental animals known to have extensive inbreeding variation. Researchers found high correlation between pedigree inbreeding coefficient and runs of homozygous SNPs.

As all these studies indicate, SNPs can be used to accurately predict whether two individuals are related without the aid of pedigree data. Using SNP data can provide more detailed information about the genetics of an individual as compared to pedigree data alone. By using SNPs, it is possible to determine which alleles were inherited, whereas calculations based on pedigrees make many assumptions as to the percent of alleles inherited from a particular source. For non-inbred parent-offspring relationships, this is not as much of a concern since the offspring inherits half its DNA from each parent. For inbred individuals or other types of relationships, such as siblings, SNPs can provide valuable insight into which individuals could produce the most genetically diverse offspring.

Multidimensional Scaling

MDS plots provide a visual representation of breed differences and similarities (Herrero-Medrano, 2012; Kijas, 2012). These plots can provide a starting point for looking into particular breeds more closely. Since they only indicate that a similarity or difference exists and provide no information as to what that similarity might be, further

analysis is often required. These plots are still useful, however, and provide a 'big picture' view of the research.

Similar research (Herrero-Medrano, 2012), compared Meishan, Large White, Landrace, Chato Murciano, Berkshire, Tamworth, Iberian Pig and Duroc in an MDS plot. The researchers found Large White and Landrace clustered closely together when compared with all pigs and when compared with only European breeds.

Differences between the findings of Herrero-Medrano et al. (2012), and this research, include the clustering of Tamworth with other breeds. Tamworth appears closely clustered with Berkshire and Chato Murciano when all breeds are compared. When only European breeds are compared, Tamworth loosely clusters with Berkshire and Iberian Pig. In the findings of this research, Tamworth did not cluster with any other breeds, however Berkshire and Iberian Pig were not part of this study.

Similar findings include the isolation of Duroc as compared to all breeds. In the research presented here, Duroc clustered alone, but it was near Red Wattle. In Herrero-Medrano's findings, Duroc clusters well away from all other breeds.

Creating an MDS plot with additional breeds could provide more information about how these breeds are related. Comparing European versus United States sourced Large White and Landrace could reveal differences in breeds due to selection.

Effective Population Size and Linkage Disequilibrium

Much research has been done regarding effective population size of livestock breeds (Uimari 2011; MacEachern 2009; Ollivier 2004; Weigel 2001). Research

analyzing LD in addition to effective population size is abundant, but most focus on commercial breeds of swine (Badke, Uimari).

Badke, et al., estimated LD in four breeds of swine, Duroc, Hampshire, Landrace and Yorkshire. Average r^2 for Landrace was 0.15 and Duroc was 0.46. These numbers are lower than r^2 values found by this research. For Duroc, r^2 was 0.62 and for Landrace, r^2 was 0.54.

Uimari et al., estimated LD for Finnish Landrace and Yorkshire. For Finnish Landrace, the r^2 value was 0.43. This number is closer to what was found in the present research. In the present study, Landrace SNP data were obtained from Goedbloed, et al., and individuals for that study were sourced from Europe. The geographical source of individuals could impact LD.

Little research has been done to determine LD present in heritage breeds. Additionally, little is known about the effective population size. It is assumed the effective population size is small for these breeds (American Livestock Breed Conservancy, 2011) based off of census data alone. More research to determine current census populations of these breeds would provide a more detailed picture on the outlook of the breed as a whole.

This research indicates heritage breeds possess high levels of LD and low effective population sizes. This would be expected for herds existing in isolation or on small farms where new bloodlines are seldom introduced. The Ossabaw Island pigs existed in isolation on an island, where natural selection created an animal capable of surviving in the island's unique environment. Mulefoot individuals possessed the highest

value of r^2 and were also sampled from a single location. Additionally, they were found to be highly inbred.

Unique Single Nucleotide Polymorphisms

Cooper et al. identified SNPs unique to different breeds of cattle. Ayrshire, Brown Swiss, Holstein and Jersey individuals were identified by SNP frequencies. To determine if a SNP was unique to one of the breeds, a selected allele had to occur at a frequency of ≥0.90 monomorphic, i.e. homozygous, for one breed, and be ≤0.30 monomorphic for all other breeds. For Ayrshire cattle, twelve SNPs met these criteria and could be used in identifying Ayrshire individuals.

In this thesis, only the counts of SNPs with a MAF of 1.0 were reported for each breed. Further analysis could reveal SNPs meeting the criteria put forth by Cooper et al., allowing SNPs unique to a particular breed to be identified. Identifying animals of a particular breed by use of SNPs has been explored in cattle (Cooper, 2013; MacKay, 2008) but little if any research for pigs is available.

CHAPTER 7: GENERAL DISCUSSION

This study indicates heritage breeds have qualities, such as hardiness, and self-sufficiency, worth preserving that are not seen in industry breeds and conservation efforts should continue. The results also show that loss of genetic diversity from high levels of inbreeding is a real threat to the conservation of heritage breeds.

Additional research could be done by gathering genetic information from more breeds or more individuals within the breeds in this study. Creating a detailed map of all swine breeds could provide information benefiting both the small farmer and the industry. Alleles found only in heritage breeds could be of value to large scale operations, but without further research, such alleles could be lost forever. Genes for disease resistance are of particular interest to farmers everywhere, but especially in large operations where swine live close together sharing food and water. In such operations, disease can spread quickly, and with increasing pressure from outside organizations to reduce or eliminate antibiotic use, having animals that are naturally disease resistant could prove vital to continued success of confined feeding operations.

Breeds of particular interest are the Red Wattle, Duroc and Tamworth, due to their apparent genetic diversity as compared to other breeds in this study. It would be interesting to obtain SNP data for additional breeds not seen in this study. This research focused on breeds endangered in the United States, but there are many breeds all over the world facing extinction, and a global database could provide more information about the origin of many breeds.

Additionally, resources for producers could be developed to aid them in creating breeding programs to maximize the preservation of genetic diversity. Such resources could include the ability to send DNA samples to a university and have the DNA analyzed to provide the farmer genetic information. This information could include the number of hetero- and homozygous alleles as well as the individual's relatedness to other individuals within a breed. If a database is maintained over time for SNP data of individuals from various breeds, this could prove to be an invaluable resource to the preservation of rare breeds.

Currently, a program to assist in preserving genetic diversity through breeding plans is available for cattle producers. Matesel, developed by Brian Kinghorn, allows producers to submit information about individuals in their herd and the program puts out a breeding plan to maximize genetic gain. The program does not use SNP data, instead it relies on pedigree and best linear unbiased prediction, BLUP, information to produce Estimated Breeding Values, EBVs, for cattle (Breedplan, 2013). Producers using it have seen rapid genetic improvements within their herds (Beef Central, 2013).

As long as producers and researchers remain interested in rare and endangered breeds of pigs, there is hope that these breeds might survive. Through research and education, producers will gain access to tools and knowledge enabling them to raise animals that provide the best future for their breed.

Appendix 1. Plink Program Codes

plink --file mydata --allow-no-sex --maf 0.05 --mind 0.1 --make-bed --out nameoffile

This command filters SNPs based on a minor allele frequency of 0.05 and filters individuals missing 0.1 of their snps. Then a new file is generated (.bed) which contains SNPs and individuals meeting these filters. Also notice, the command 'allow-no-sex' is used. This tells plink that there is no information for sex. Without this statement, plink will set ambiguously sexed individuals as missing and no results will be produced. When sex information is not available, using this command is easier than attempting to modify the original file.

plink --file mydata --allow-no-sex --merge-list allfiles.txt --make-bed --out Porcineall

This command merges files together and outputs '.bed' files. Bed files are easier to work with since they run faster in Plink.

plink --bfile mydata --allow-no-sex --maf 0.05 --mind 0.1 --genome --out filename

This code creates the GRM. Notice the change from 'file' to 'bfile'. When using '.bed' files, the command 'bfile' must be used to call up the correct file. The command does not actually construct the matrix, it simply outputs the data in column form. The matrix will need to be created by hand.

Since the 'genome' command can produce a large dataset if many individuals are used. The results can be narrowed by applying constraints to the estimated relationship value. By adding a minimum, only animals exceeding that level of relatedness will be reported.

plink --bfile mydata --allow-no-sex --maf 0.05 --mind 0.1 --genome --min 0.05 --out filename

By keeping the minimum small, animals sharing even a small level of relatedness are still reported, but the output isn't overwhelmed by animals with zero relation.

plink --bfile mydata --allow-no-sex --indep 50 5 2

This command prunes the SNPs prior to calculating inbreeding coefficients. The file 'plink.prune.out' will contain a list of SNPs in linkage disequilibrium. This list can then be used to tell plink which SNPs to exclude when calculating inbreeding coefficient. By eliminating SNPs in linkage disequilibrium, bias due to selection is reduced.

plink --bfile mydata --allow-no-sex --exclude plink.prune.out --het --out filename

The 'het' command produces a file containing a calculation for inbreeding coefficient (F). The 'exclude' command followed by the file 'plink.prune.out' will eliminate SNPs in linkage disequilibrium from the calculations.

To perform multidimensional scaling, the 'genome' command must first be applied to the data. Then the MDS plot can be created with the following code.

plink --bfile mydata --allow-no-sex --read-genome myfile.genome --cluster --mds-plot 4 --out mymdsfile

From the results, plot C1 values against C2 values to create a scatter plot. The plot can be created in a program such as Microsoft Office Excel.

To calculate allele freqencies, including minor allele frequency (MAF), use the following code.

plink --file mydata --allow-no-sex --freq --out newfilename

The 'freq' command produces allele frequencies. In order to cluster the results by a given criteria (i.e. breed), first create a cluster file assigning each individual to a cluster. This can easily be done by accessing the 'fam' file. In the 'fam' file only keep the first and second columns then add a third column. Use the third column to designate the clusters. Save this file as a text document (.txt) so it can be used in the following code.

plink --file mydata --allow-no-sex --within cluster.txt --missing --freq --out newfilename

The following code was used to remove SNPs in LD and also to calculate r^2

plink --file mydata --allow-no-sex --keep mylist.txt --indep-pairwise --50 5 1 --out filename

Since all pigs were combined into one file for analysis, each breed had to be isolated to count LD. The command "keep" followed by the text file containing individuals to keep allows the program to only run these individuals. The command "indep-pairwise" performs the SNP by SNP comparison looking for LD, with the following criteria: 50 indicates a window size of 50 SNPs, 5 is the number of SNPs to shift the window, and 1 is the r^2 value. SNPs in complete LD were desired so r^2 was set to 1, meaning only SNPs completely associated with another SNP would be removed.

The following code was used to calculate r^2 for all SNPs within a breed.

plink --file mydata --allow-no-sex --keep mylist.txt --r2 --out filename

Again, individuals within a breed were isolated using the "keep" command. The "r2" command calculates the multiple correlation coefficient for all SNPs. The ".ld" output file contains the r^2 data.

Appendix 2: SAS Code

```
Data pigs;
Input breed$ relationship;
datalines;
SB      0.1555
SB      0
SB      0.1827
SB      0.156
SB      0.1482
SB      0.1114
SB      0.1771
SB      0.1311
SB      0.1778
SB      0.154
....

TA01    0.3527
TA01    0.4277
TA01    0.3852
TA01    0.3141
;
ods rtf style=journal file= 'outputfilename.rtf';
proc glm;
class breed;
model F=breed;
means breed/Tukey;
lsmeans breed;
run;
ods rtf close;
```

The only changes in the code when running SAS for inbreeding instead of relationship,

included changing the input and datalines. The proc glm statement did not change.

Appendix 3: Producers Involved with Research

Kevin Fall of the American Guinea Hog Association provided genotyping results for ten Guinea Hogs and three Ossabaw Island hogs, sourced from various farms across the United States. Robert Long of Flintrock Bison Ranch in Ash Grove, Missouri, and Cheryl Fanning of Dogwood Hill Farm, LLC in Lamar, Missouri submitted hair from four and three Ossabaw Island hogs respectively.

Samples from Red Wattle hogs were collected from Garret Caryl and Clyde Grover at the Small Farmer's Forum in Columbia, Missouri. Mulefoot samples were collected by the researcher from animals owned by Art and Vera Gelder at Walk-About Acres in Columbia, Missouri. Jeanette Beranger of the American Livestock Conservancy submitted two Saddleback samples from the farm of Matt and Michele Whalen of Green Mountain Heritage farm in Vermont.

Literature Cited

Alderson, G.L.H. 1992. A System to Maximize the Maintenance of Genetic Variability in Small Populations. In: L. Alderson and I. Bodo, editors, Genetic Conservation of Domestic Livestock. C.A.B. International, UK. p. 18-29.

Altshuler, D., M.J. Daly, and E.S. Lander. 2008. Genetic Mapping in Human Disease. Science. 322:5903:881-888.

Amaral, A.J., H.J. Megens, R.P.M.A. Croojmans, H.C.M. Heuven and M.A.M. Groenen. 2008. Linkage disequilibrium decay and haplotype block structure in the pig. Genetics Society of America.

American Livestock Breed Conservancy, The (ALBC). "Breed information – ALBC Conservation Priority List." Retrieved from: http://www.albc-usa.org/cpl/wtchlist.html. Retrieved on: January 8, 2011.

Animal Genetics Resources Group (AGRG). 2006. Protecting animal genetic diversity for food and agriculture: Time for action. Food and Agriculture Organization of the United Nations. Rome, Italy. http://www.fao.org/ag/magazine/pdf/angr.pdf.

Badke, Y.M., R.O. Bates, C.W. Ernst, C. Schwab and J.P. Steibel. 2012. Estimation of linkage disequilibrium in four United States pig breeds. BMC Genomics. 13:24.

Beef Central. 2013. New software tools helps optimise Breedplan-driven breeding decisions. Available: http://www.beefcentral.com/production/new-software-tools-helps-optimise-breedplan-driven-breeding-decisions. Accessed Aug. 4, 2014.

Borg, I., P.J.F. Groenen, and P. Mair. 2013. Applied Multidimensional Scaling. 1st ed. Springer Heidleberg. New York, NY.

Breedplan. 2013. International Beef Recording Scheme: Breedplan- Overview. Available: http://breedplan.une.edu.au/index. Accessed Aug. 4, 2014.

Cardellino, R.A. 2004. Conservation of farm animal genetic resources – a global view. In: G. Simm, B. Villanueva, K.D. Sinclair and S. Townsend, editors, Farm Animal Genetic Resources. Nottingham University Press, Nottingham, UK. p. 1-14.

Cooper, T.A., G.R. Wiggans, D.J. Null, J.L. Hutchison, and J.B. Cole. 2013. Genomic evaluation, breed identification, and discovery of a haplotype affecting fertility for Ayrshire dairy cattle. J. Dairy Sci. 97:3878-3882.

Cox, T.F. and M.A.A. Cox. 2001. Multidimensional Scaling, Second Edition. 2nd ed. Chapman & Hall/CRC. Boca Raton, FL.

Dimitri, C. A. Effland, and N. Conklin. 2005. The 20th Century Transformation of U.S. Agriculture and Farm Policy. Economic Information Bulletin. EIB 3:17.

Dohner, J.V. 2001. The Encyclopedia of Historic and Endangered Livestock Breeds. 1st ed. Yale University Press, New Haven, CT.

Engelsma, K.A., R.F. Veerkamp, M.P.L. Calus, P. Bijma and J.J. Windig. 2012. Pedigree and marker-based methods in the estimation of genetic diversity in small groups of Holstein cattle. J.Anim.Breed.Genet. 129:195-205.

England, P.R., J.M. Cornuet, P. Berthier, D.A. Tallmon, and G. Luikart. 2006. Estimating effective population size from linkage disequilibrium: severe bias in small samples.” Consv.Gen. 7:303-308.

Falconer, D.S. and T.F.C. Mackay. 1996. Introduction to Quantitative Genetics. 4th ed. Longman Group Ltd. Essex, England.

Garcia-Gamez, E., G. Sahana, B. Gutierrez-Gil, J.J. Arranz. 2012. Linkage disequilibrium and inbreeding estimation in Spanish Churra sheep. BMC Genetics. 13:43.

Gaut, B. S., and A. D. Long. 2003. The lowdown on linkage disequilibrium. The Plant cell.15:7:1502-1506.

Gill, J.J.B. and M. Harland. 1992. Maximal Maintenance of Genetic Variation in Small Populations. In: L. Alderson and I. Bodo, editors, Genetic Conservation of Domestic Livestock. C.A.B. International, UK. 3-17.

Goddard M., and B. Hayes. 2009. Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nat. Rev.10:381–391.

Goedbloed, D.J., H. Megens, P. van Hooft, J.M. Herrero-Medrano, W. Lutz, P. Alexandri, R.P.M.A. Crooijmans, M. Groenen, S.E. van Wieren, R.C. Ydenberg, and H.H.T. Prins. 2012. Genome-wide single nucleotide polymorphism analysis reveals recent genetic introgression from domestic pigs into Northwest European wild boar populations. Molecular Ecology 22:3:856–866.

Goedbloed, D.J., H. Megens, P. van Hooft, J.M. Herrero-Medrano, W. Lutz, P. Alexandri, R.P.M.A. Crooijmans, S.E. van Wieren, R.C. Ydenberg, H.H.T. Prins, and M. Groenen. 2012. Data from: Genome-wide single nucleotide polymorphism analysis reveals recent genetic introgression from domestic pigs into Northwest European wild boar populations. Dryad Digital Repository.

Hall, S. J. G. 2004. Livestock Biodiversity: Genetic Resources for the Farming of the Future. Blackwell Science Ltd., Oxford, UK.

Hassebrook, C. 1999. Meeting Food Needs through Sustainable Production Systems and Family Farms. In: D. P. Weeks, J. B. Segelken and R. W. F. Hardy, editors, NABC Report 11: World Food Security and Sustainability: The Impacts of Biotechnology and Industrial Consolidation. National Agricultural Biotechnology Council, Ithaca, NY. p. 111-119.

Heffernan, W.D. 1999. Biotechnology and Mature Capitalism. In: D. P. Weeks, J. B. Segelken, and R. W. F. Hardy, editors, NABC Report 11: World Food Security and Sustainability: The Impacts of Biotechnology and Industrial Consolidation. National Agricultural Biotechnology Council, Ithaca, NY. p. 121-135.

Herrero-Medrano, J. M., H. J. Megens, R. P. Crooijmans, J. M. Abellaneda, and G. Ramis. 2012. Farm-by-farm analysis of microsatellite, mtDNA and SNP genotype data reveals inbreeding and crossbreeding as threats to the survival of a native Spanish pig breed. An. Genet. 44:259-266.

Kijas, J.W., et al. 2012. Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. PLOS Bio. 10:2.

Kliman, R., B. Sheehy, and J. Schultz. 2008. Genetic Drift and Effective Population Size. Nature Education. 1:3.

Lamberson, W.R. 1998. Marker allele frequency index selection as a method for reducing loss of genetic variation in small populations. World Congr. Genet. Appl. Livestock Prod. 28:107-110.

Lopes, M. S., F. F. Silva, B. Harlizius, N. Duijvesteijn, P. S. Lopes, S. EF. Guimaraes, and E. F. Knol. 2013. Improved estimation of inbreeding and kinship in pigs using optimized SNP panels. BMC Genetics. 14:92.

MacEachern, S., B. Hayes, J. McEwan, and M. Goddard. 2009. An examination of positive selection and changing effective population size in Angus and Holstein cattle populations (*Bos taurus*) using a high density SNP genotyping platform and the contribution of ancient polymorphism to genomic diversity in Domestic cattle. BMC Genomics. 10:181

Mason, I.L. and V. Porter. 2002. Mason's World Dictionary of Livestock Breeds, Types and Varieties. 5th ed. CABI, New York, NY.

McKay, S. D., R. D. Schnabel, B. M. Murdoch, L. K. Matukumalli, J. Aerts, W. Coppieters, D. Crews, E. D. Neto, C. A. Gill, C. Gao, H. Mannen, Z. Wang, C. P. Van Tassell, J. L. Williams, J. F. Taylor and S. S. Moore. 2008. An assessment of population structure in eight breeds of cattle using a whole genome SNP panel. BMC Genetics. 9:37.

National Pork Board. 2009. Quick Facts: The Pork Industry at a Glance. America's Pork Checkoff Program. Des Moines, IA.

National Research Council, NRC. 1993. Managing Global Genetic Resources: Livestock. 1st ed. National Academy Press, Washington, D.C.

Ollivier, L., and J. W. James. 2004. Predicting the annual effective size of livestock populations. Genetical Research. 84:1:41-46.

Paarlberg, D. and P. Paarlberg. 2000. The Agricultural Revolution of the 20th Century. 1st ed. Iowa State University Press, Ames, IA.

Pinstrup-Anderson, A., and R. Pandy-Lorch. 1999. Securing and Sustaining Adequate Food Production for the Third Millenium. In: D. P. Weeks, J. B. Segelken, and R. W. F. Hardy, editors, NABC Report 11: World Food Security and Sustainability: The Impacts of Biotechnology and Industrial Consolidation. National Agricultural Biotechnology Council, Ithaca, NY. p. 121-135.

Population Reference Bureau, PRB. 2013. Population Growth. http://www.prb.org/Educators/TeachersGuides/HumanPopulation/PopulationGrowth.aspx. (Accessed 22 April 2013.)

Purcell, S. 2013. Package: PLINK 1.07. http://pngu.mgh.harvard.edu/purcell/PLINK/. (Accessed 5 January 2013.)

Purcell S., B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, and P. C. Sham. 2007. PLINK: a toolset for whole-genome association and population-based linkage analysis. American Journal of Human Genetics. 81.

Saura, M., A. Fernandez, M. C. Rodriguez, M. A. Toro, C. Barragan, A. I. Fernandez, and B. Villanueva. 2013. Genome-wide estimates of coancestry and inbreeding in a closed herd of ancient Iberian pigs. PLOS One. 8:10.

Silio, L., M. C. Rodriguez, A. Fernandez, C. Barragan, R. Benitez, C. Ovilo, and A.I. Fernandez. 2013. Measuring inbreeding and inbreeding depression on pig growth from pedigree or SNP-derived metrics. J. Anim. Breed. Genet. 130:349-360.

Slatkin, M. 2008. Linkage Disequilibrium - Understanding the evolutionary past and mapping the medical future. Nature Reviews Genetics. 9:477-485.

Stachowiak, M., M. Szydlowski, J. Cieslak, and M. Switonski. 2007. SNPs in the Porcine PPARGC1A gene: Interbreed differences and their phenotypic effects. Cellular and Molecular Biology Letters. 12:231-239.

Taberlet, P, A. Valentini, H. R. Rezaei, S. Naderi, F. Pompanon, R. Negrini, P. Ajmone-Marsan. 2008. Are cattle, sheep, and goats endangered species?. Molecular Ecology. 17:275–284.

Tenesa, A., P. Navarro, B. J. Hayes, D. L. Duffy, G.M. Clarke, M. E. Goddard, and P. M. Visscher. 2007. Recent human effective population size estimated from linkage disequilibrium. Genome. 17:520-526.

Uimari, P. and M. Tapio. 2011. Extent of linkage disequilibrium and effective population size in Finnish Landrace and Finnish Yorkshire pig breeds. Journal of An. Sci. 89:3:609-614.

United States Census Bureau. 2010. U.S. and World Population Clock. http://www.census.gov/popclock. (Accessed 20 April 2013.)

USDA. 2013. Economic Research Service: Food Expenditures. http://www.ers.usda.gov/data-products/food-expenditures.aspx. (Accessed 20 April 2013.)

VanRaden, P.M. 2008. Efficient Methods to Compute Genomic Predictions. J. of Dairy Sci. 91:4414-4423.

Villanueva, B., R. Pong-Wong, J. A. Wooliams, and S. Avendano. 2004. Managing genetic resources in selected conserved populations. In: G. Simm, B. Villanueva, K.D. Sinclair and S. Townsend, editors, Farm Animal Genetic Resources. Nottingham University Press, Nottingham, UK. p. 113-131.

Washington State University, WSU. 2011. Annual income spent on food. Washington State Magazine. Available: https://wsm.wsu.edu/researcher/WSMaug11_billions.pdf. Accessed Apr. 20, 2013.

Weigel, K. A. 2001. Controlling inbreeding in modern breeding programs. Journal of Dairy Science. E. Suppl. 84:E177–E184.

Windig, J.J., and K.A. Engelsma. 2010. Perspectives of genomics for genetic conservation of livestock. Consv.Gen. 11:635-641.

Wright, L. I., T. Tregenza, and D. J. Hosken. 2008. Inbreeding, inbreeding depression and extinction. Conservation Genetics. 9:4, 833-843.

Zhu, X., S. Zhang, D. Kan, and R. Cooper. 2004. Haplotype block definition and its application. Pacific Symposium on Biocomputing 9:152-163.