THESIS

CROSS COUNTRY COMPARISON OF GENETIC DIVERSITY BY MERGING MICROSATELLITE DATA FROM PIGS

Submitted by

Cyril O. Ibeka

Department of Animal Sciences

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Spring 2014

Master's Committee:

Advisor: Jack Whittier
Co-Advisor: Harvey Blackburn

Phillip Chapman

ABSTRACT

CROSS COUNTRY COMPARISON OF GENETIC DIVERSITY BY MERGING MICROSATELLITE DATA FROM PIGS

The need to develop a comprehensive method to extend genetic diversity to a global level has never being more important given the increase in the amount of genetic information being generated within various countries using different genotyping procedures, hence the need to merge datasets. Merging also enables a country to compare the genetic diversity of its animals to other countries. Many genetic diversity studies have been limited either due to the use of manual merging techniques, which ignores the effect of different laboratories with different laboratory protocols and are obviously time consuming or due to allele size inconsistencies in the case of automated merging processes. Our objective was to develop a method that extends genetic diversity globally using inexpensive methods based on Bayesian approach. Thirty microsatellite markers were originally genotyped from 220 Brazilian pigs and thirty-five microsatellite markers were genotyped from 179 pigs from the United States. Fourteen microsatellite markers were common between pigs in both countries. However, twelve microsatellite markers with posterior probability greater than 0.550 were successfully merged using a Bayesian cluster method. GENALEX and FSTAT results showed that the Brazilian pigs were significantly different from the Chinese ($P< 0.01667$) and US ($P < 0.01667$) pigs but were genetically closer to the US breeds (0.25) than the Chinese breeds (0.42). Population structure results obtained from STRUCTURE software showed peaks at $K = 2, 4, 13$ and 15. STRUCTURE result with $K = 4$ showed evidence of geographic differentiation of breeds into Brazil, China and the United States.

STRUCTURE result at K = 9 revealed evidence of admixture within countries. The Chinese pigs showed evidence of genetic differentiation within the breeds. Many Brazilian pigs are not unique pigs but are admixtures developed by crossing local pigs with commercial pigs. The genetic diversity of the US swine population may need to be increased to prevent loss of biodiversity in the event of disease outbreaks or natural disaster.

KEYWORDS

Animal genetic resources

Bayesian methods

genetic diversity

pigs

ACKNOWLEDGEMENTS

I would like to thank my graduate advisor Dr. Jack Whittier, who through his effort ensured that my final semester in the department of Animal Science at Colorado State University was a huge success. I reserve the highest respect for Dr. Harvey Blackburn, my co-advisor who was directly responsible for all my research success at Colorado State University. To Dr. Phillip Chapman from the department of Statistics at Colorado State University, who later joined my graduate committee, I would like to thank you immensely for all your contributions; I learnt a lot from you.

I would especially like to acknowledge Samuel Paiva from the Brazilian Agricultural Research Corporation (EMBRAPA). You were of immense help throughout this research.

I like to thank Dr. Green Gerwyn from the department of Statistics; your vast knowledge of the field of Statistics gave me a lot to think about. I can say with confidence that I am thinking Statistics, Programming and Genetics.

I would like to thank Dr. Angela Presson whose communication was helpful during the merging process using MICROMERGE. Not forgetting Jerome Goudet who discussed aspects of the FSTAT software program with me.

Finally, I remember the words of Mohandas Gandhis, “First they ignore you, then they laugh at you, then they fight you, then you win”. I thank God I finally succeeded.

DEDICATION

This thesis is dedicated to my family especially my sister Janet Ibeka who not only provided the finances to complete my graduate work, but also supported me during my difficult period at Colorado State University. To my mom of blessed memory who knew I was going to succeed no matter the challenge. March 13 makes it 14 years from the day you left for a better place; you are gone now but you are never forgotten. Not forgetting my dad, this work is dedicated to you all.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

1.0 INTRODUCTION

Species that have low genetic diversity at several or all of its genes are most likely to be affected from the impact of human-induced changes and pressures of selection, genetic drift and mutation (Herrero-Medrano *et al*., 2013). Lose of variation may lead to reduced response to changing environments, of which genetic susceptibility to novel infectious diseases is a major concern (Herrero-Medrano *et al*., 2013).

Livestock breeds are excellent biological entities with a profoundly important, and growing, role in maintaining food safety (FAO, 2007). They serve not only as food and sources of income to local producers but are also at the center of strategies by the United Nations towards achieving the Millennium Development goals (FAO, 2007).

The term biodiversity or biological diversity refers to the variability among living organisms from all sources which include diversity within species, between species and of the ecosystems (Tandon and Bhattacharjee, 2010). Different alleles may exist for a particular trait; however, the DNA sequence may or not be identical. When alleles that govern the expression of that trait are identical within a population, the population is said to have a low genetic diversity for the trait. Likewise, when the alleles that contribute to the expression of a trait within a population are different, the population is said to have high genetic diversity for that trait.

Numerous genetic diversity studies have been conducted within countries using microsatellite markers. What is often lacking in such studies is the ability to extend results to a regional or global level. Furthermore, the consequence of domestication and post-domestication events has made it even more imperative to extend genetic diversity studies to regional and

global level. For example, it is possible to find similar breeds separated by geographical location but having different names. Also, it is possible to find animals within a breed that may have accumulated different alleles due to genetic drift.

Attempts have been made in the past to merge microsatellite markers datasets for cross country genetic diversity studies. Merging data sets manually is difficult, time consuming, and error prone (Weeks *et al.,* 2002; Presson *et al*., 2006; Presson *et al*., 2008) due to several factors notably: differences in genotyping process and hardware; binning methods during the end after genotyping processes; molecular weight standards and finally due to curve fitting algorithms (Jianlin, 2011).

Re-genotyping of genetic materials is a choice since it is becoming faster and cheaper to genotype samples. However, it is impractical to re-genotype all samples at one center (Presson *et al*., 2006). It is easier to exchange allele frequency than samples across international borders.

COMBI.PL merges data sets by assigning allele sizes across studies using maximum likelihood estimation (Taubert and Bradley, 2008). The algorithm is based on the assumption that assumption that allele frequencies are not expected to be equal. However, size discrepancies between genotyping facilities for microsatellite markers; inaccuracies and common differences in grouping and naming alleles make the use of allelic size very inefficient (Presson *et al*., 2006).

It would be impossible to identify the extent of similarities or the lack there of between breeds without extending genetic diversity to a global level. By using Bayesian approaches (Presson *et al*., 2008; Paiva *et al*. 2011) these investments can be extended to provide greater insight into the status of genetic diversity among countries without additional sampling and genotyping. In an effort to better understand genetic diversity among western hemisphere sheep

breeds Paiva, *et al*. (2011) showed how such an approach was useful in comparing Brazilian and US sheep breeds.

In order to conserve national genetic diversity, countries such as Brazil (BR) and the United States (US), have embarked on specific national conservation strategies to protect local breeds as well as increase animal genetic diversity. Although conserving local genetic diversity is extremely important, any attempt to increase species specific diversity is dependent on the genetic diversity available within a gene pool.

Brazil and United States pig breeds originated from different European countries or China (CH). However, once imported, breeding decisions were made to mold the populations to meet the needs within each country and in some instances to develop new breeds (e.g., Duroc). In this study Brazil and US microsatellite data for pigs was merged using a Bayesian based method developed by Presson *et al*. (2008).

1.1 Objective

The objective of the study was to provide a comprehensive method to study genetic diversity of United States and Brazil pigs using inexpensive yet advanced bioinformatics tools.

CHAPTER TWO

2.0 LITERATURE REVIEW

2.1 Domestication and Movement of Pigs

Genetic and archaeological findings suggest that pig domestication began at 6 locations namely: Europe, Near and Middle East, China, Indian Subcontinent and South East Asia (FAO, 2007). Western and Far Eastern pigs have strong genetic differentiation and are phenotypically different. The Far Eastern pigs are fatter with shorter legs, dropping belly, prick ears and wrinkled skin (Amills *et al*., 2010). Recent reports suggest that China might have been one of the most important pig domestication sites in the Far East (Yuan *et al*., 2008).

There appears to be contradictory reports about the origin of European domestic pigs. Evidence from mitochondrial markers suggests an absence of Near Eastern haplotypes in modern European swine breeds (Larson *et al*., 2005; Ramirez *et al*., 2009; Amills *et al*., 2010). However, autosomal and Y-chromosome markers provide evidence of a close relationship between the wild boar, Sus scrofa from Europe and that from Near East (Ramirez *et al*., 2009; Amills *et al*., 2010). The use of Y-chromosome markers and mitochondrial DNA in the different studies make both reports credible since lineage markers do not recombine during meiosis and are inherited unchanged from generation to generation. Larson *et al*. (2005) and Amills *et al*. (2010) suggested that the contradictory results could be the consequence of independent domestication events in Europe.

Breeds are usually created by admixing diverse population types with different production and genetic backgrounds (Amills *et al*., 2010). Today's pig breeds have been developed from sustained breed development. The United States, China and England are

historically the 3 main centers of pig development (Jones, 1998) with China, the European Union, United States, Brazil and Canada are the largest producers of pigs as a source of meat (40%, 93 million metric tons; Orr and Shen, 2006).

Since domestication, pigs have been moved across the world. For instance, Iberian pigs became prominent in South America after colonization; European alleles are predominant in creole pigs (Ramirez *et al*., 2009; Souza *et al*., 2009); Far Eastern, European and Chinese alleles have been observed at significant frequencies in Africa pigs (Ramirez *et al*., 2009).

2.2 Swine Breeds from the United States and Brazil

Table 1. Present Location, Origin, Composition, Type, Morphology, Genetics and Uses of Some Swine Breeds in the United States

Breed	Present location	Origin	Composition	Type	Morphology	Genetics	Uses
Meishan	USA	North Shanghai, China	Taihu	Local	Small to medium size; black wrinkled coat and face.	Prolific	Meat and Lard
Fengjing	USA	North Shanghai, China	Taihu	Local	Black and wrinkled face and skin	Prolific	Meat and Lard
Minzhu	USA	Liaoning, Jilin and Heilongjiang, NE China	North China	Local	Black color with large lop ears	Prolific	Meat and lard
Yorkshire	USA	York, England	Yorkshire with Cantonese	Commercial	White in color with erect ears	Maternal	Meat
Berkshire	USA	Reading, England	Cantonese and Siamese x Local Old English before or Neapolitan	Commercial	Black color with white points on legs, face and tail	Terminal	Meat
Duroc	USA	New York; New Jersey; England	Jersey Red of New Jersey and Old Duroc of New York	Commercial	Red color but range from light golden, almost yellow to very dark red	Terminal	Meat and formerly lard

Sources: Mason (2002)

Table 2. Present Location, Origin, Composition, Type, Morphology, Genetics and Uses of Some Swine Breeds in Brazil

Breed	Present location	Origin	Composition	Type	Morphology	Genetics	Uses
Landrace	Brazil	Denmark		Commercial	White with black spots	Maternal	Meat
Monteiro	Brazil	Brazil		Naturalized			Meat
MS60	Brazil	Brazil	Duroc x Large White x Pietrain	Composite			Meat
Piau	Brazil	Paranaiba River basin of S.W. Brazil	Canastra and/or Canastrao x Poland China x Duroc	Naturalized	White/cream with black or red spots		Meat and lard
Caruncho	Brazil	Brazil	Piau x Tatu	Naturalized	White/sandy with black spots, sometimes red -and-white/black		Lard
Casco de Burro	Brazil	Brazil		Naturalized			Lard
Pirapetinga	Brazil	Minas Gerais, Brazil	Nilo x Tatu.	Naturalized	Black or Violet		Meat
Large White	Brazil	England		Commercial			Meat
Nilo	Brazil	Brazil	Canastra x Tatu	Naturalized	Black with white spots		Lard
Pietrain	Brazil	Brabant in Belgium	Bayeux (unknown male and Tamworth) x local	Naturalized	Dirty white with black or reddish spots with red hairs	Terminal	Meat
Moura	Brazil	S. Brazil	Duroc x Canastra x Canastrao	Naturalized	Blue or with red roan		Meat
Mamelado	Brazil	Uruguay		NaturalIzed			Meat

2.3 Sustainability of Genetic Diversity

Maintaining the survival of local breeds is crucial in the fight to sustain global diversity. As exotic breeds are imported, there is the danger of diluting and eventually disintegrating the local breeds by indiscriminate cross-breeding with imported stock (FAO, 2007). The Food and Agricultural organization strongly has encouraged countries to develop strategies to conserve their local breeds. In Brazil, the Animal Genetics Resource Conservation Program coordinated by the National Research Centre for Genetic Resources and Biotechnology (Cenargen) of the Brazilian Agricultural Research Corporation (EMPRAPA) was set up and charged with this responsibility (Mariante *et al*., 1999). The National Center for Genetic Resources Preservation (NCGRP) through the Agricultural Research Service (ARS) of the United States Department of Agriculture (USDA) conserves crops and animal genetic resources that are important to the US (USDA... http://www.ars-grin.gov/ncgrp/index.htm).

When mutation, selection and migration are absent, the genetic properties such as heterozygosity and genotype frequency, of an infinitely large population will remain constant over time (Wang, 2005). However, with a finite population size; the genetic properties will change from generation and the rate of change depends on the effective population size (Crow and Kimura, 1970; Wang, 2005).

Genetic drift is an extremely important force in evolution, although the strength of its effect is dependent on the total population size (or the effective population size if some individuals do not breed). Small population size can lead to loss of neutral genetic variation, fixation of mildly deleterious alleles, and thereby reduces population fitness (Kalinowski and Waples, 2002). The rate of this process depends on the effective size of a population (Kalinowski and Waples, 2002).

Decrease in heterozygosity for a given locus within a population can also result from inbreeding. One consequence of the loss of genetic diversity through inbreeding is the increase in homozygotes in a population which is the main cause of inbreeding depression. Bichard (2002) warned that animals could be susceptible to diseases and parasites. A more alarming consequence of decrease in heterozygosity is the extinction of whole animal populations.

2.4 Genetic Markers for Genetic Diversity Studies

Genetic markers are specific DNA sequences in the genome of an organism that can be used to track inheritance and distribution of parental genotypes among a segregating progeny, and to determine the extent of genetic variation in a population (Schlo¨tterer, 2004). Due to the vast size of the genome, genomic analysis involving random gene sequencing between individuals may not lead to detectable differences; however, there are certain highly variable regions in the genome of organisms where differences in the sequence of individuals are highly prevalent (Fredua-Agyeman *et al*., 2008). These regions could provide useful information to study genetic diversity within and between populations.

Recent advances in molecular biology provide novel tools for addressing research questions that involve centers of domestication, breed development and genetic diversity across geographic regions (Duran *et al*., 2009; Paiva *et al*., 2011). The availability of sound genetic information ensures that decisions made on conservation will be better informed and will result in improved germplasm management (Vicente *et al*., 2005). Well informed sampling strategies for germplasm material destined for ex situ conservation and designation of priority sites (i.e. identifying specific areas with desirable genetic diversity) for in situ conservation are both crucial for successful conservation efforts. The choice of type of material to be conserved and the sampling strategies depend among other things on the objectives of cryopreservation program

(ERFP, 2003). Although the cost of embryo and semen collection for ex-situ cryopreservation is relatively high, it is useful in preserving gametes or offspring of rare or extinct breeds in the long term especially when used alongside other conservation strategies (FAO, 2006; FAO, 2009).

It is therefore important to carefully describe the conservation objectives and the nature of the conservation since there is no clear boundary between in situ and ex situ in vitro conservation (FAO, 2007). In turn, defining strategies is dependent on knowledge of location, distribution and extent of genetic diversity (de Vicente *et al*., 2005). Paiva *et al*. (2011) reported that in addition to location of samples, laboratory method and suitable genetic markers that allow for a straightforward combination of genetic resources are also important. In order to improve the conservation of animal genetic resources, the global community approved and published a Global Plan of Action for Animal genetic Resources through FAO (2007). The Global plan of Action articulated strategic priority for the characterization of genetic resources in a way to optimize the cross-national comparability of data in order to monitor trends and risks to animal genetic resources at national, regional and global levels (FAO, 2007).

Molecular characterization, by itself or in conjunction with other data (phenotypic traits or geo-referenced data), provides useful information for assessing, among other factors, the amount of genetic diversity (Perera *et al*., 2000), the structure of diversity in samples and populations (Shim and Jorgensen, 2000; Figliuolo and Perrino, 2004), rates of genetic divergence among populations (Maestri *et al*., 2002) and the distribution of diversity in populations found in different locations (Perera *et al*., 2000; Ferguson *et al*., 2004).

Measures of molecular genetic diversity unlike trait diversity are more attractive as a basis for conservation decisions because they yield quantitative measures of relatedness which can, in turn, be used to assess genetic diversity within a set of breeds (FAO, 2007). Furthermore,

molecular genetic information is not affected by environmental influences which affect trait or quantitative data to variable extents.

Several types of genetic markers have been developed for use in genetic studies. Such markers include microsatellites, single nucleotide polymorphisms (SNPS), amplified fragment length polymorphism (AFLPs), variable number of tandem repeats (VNTRs), random amplified polymorphism DNA (RAPD), single strand conformation polymorphisms (SSCPs), restriction fragment length polymorphisms (RFLPs), short tandem repeats, single feature polymorphism, diversity arrays technology (DArT), restriction site associated DNA (RAD) markers, mtDNA and Y chromosome. However, for diversity studies, the use of formerly used category genetic markers such as major histocompatibility complex (MHC), or other protein polymorphisms or less-specific markers such as RAPD and AFLP are not recommended for genetic diversity studies unless they are needed to answer specific questions FAO (2011).

2.5 Use of Microsatellite Markers for Genetic Diversity Studies

Microsatellite markers are tandem repeat short DNA sequences, and are often highly polymorphic (Nagarajan *et al*., 2009). They consist of motifs of one to six nucleotides repeated several times that have a characteristic mutational behavior (Kelkar *et al*., 2010). The high mutation rates and simple Mendelian mode of inheritance of microsatellite markers make them particularly suitable for the study of fine population structure, mating systems and pedigree (Abdelkrim *et al*., 2009). They are abundant, fairly evenly distributed throughout the Euchromatic part of the genome and most microsatellite loci can be amplified by standard PCR (Schlo¨tterer, 2004).

The variability of the number of repeat units at a particular locus and the conservation of the sequences flanking the tandem repeat make microsatellites valuable, codominant genetic

markers [Tautz, 1989]. Different individuals exhibit variation manifested as repeat number differences (Guichoux *et al*., 2011). When microsatellite markers for a particular species are developed, their capacity to be amplified by PCR allows large-scale genotyping on automated DNA analyzers for the construction of genetic linkage maps, and facilitates studies of population genetics and reproduction ecology (Tani *et al*., 2004).

The advantages of microsatellite markers over single-nucleotide polymorphism (SNPs) include high allelic diversity and relative ease of transfer between closely related species; however, its cost and length of the development phase are some drawbacks to its use compared to SNPs (Guichoux *et. al*., 2011). Recent progress made in developing microsatellite marker and genotyping has ensured that there has been a growing interest in the use of microsatellite markers in several countries for genetic diversity of various livestock species (Guichoux *et al*., 2011).

2.6 The Role of Bioinformatics in Computing Genetic Diversity

Bioinformatics is needed to make sense of the huge volume of information constantly being generated from available techniques in molecular genetics; as these techniques evolve and become more specialized due to new information or approach, so will the requirement for more advanced tools to handle the research objectives associated with the new data sets.

Various measures exist for computing the similarities and differences within individuals and between populations based on genetic distance or genetic differentiation. Wright's fixation indices or F-statistics are the most widely used parameters to describe population differentiation due to population structure (Bird *et al*., 2011). Wright (1965) defined the fixation index as the correlation between uniting gametes. He developed the fixation index, F_{ST} as part of a set of hierarchical system of parameters in terms of total populations T, subdivisions S and individual I

to describe the properties of hierarchically subdivided of natural populations (Wright 1943, 1946 and 1951).

Wright (1965) defined F_{IT} as the correlation between gametes that unit to produce the individual, relative to the gametes of the total population. F_{IS} which is also referred to as the inbreeding coefficient was defined as the average over all subdivisions of the correlation between uniting gametes relative to those of their own subdivisions (Wright, 1965). A high F_{IS} value indicates a considerably high level of inbreeding. F_{ST} was defined as the correlation between random gametes within subpopulations (Wright, 1965). F_{ST} usually ranges from 0 to 1, although a slight negative value may be observed; it is the most used metric for measuring subpopulation differentiation with a value of 0 indicating a panmictic population or a population in Hardy-Weinberg equilibrium. This is a population where there is no subdivision, random mating exists and no genetic differentiation is observable within the population. In terms of allele frequencies, it refers to identical allele frequencies in a pair of populations; a value of 0.05 shows a negligible differentiation; a value of 0.25 indicates a very high level of differentiation; while a value of 1 indicates a complete isolation and extreme subdivision.

Wright's parameters were originally designed for bi-allelic locus and were found to be inapplicable to multiple alleles except for special cases of random differentiation with no selection. In order words, Wright's Parameters decrease with increasing allelic diversity (Hedrick 1999, 2005; Jost, 2008; Bird *et al*., 2011; Meirmans and Hedrick, 2011), potentially affecting the conclusions of a number of published studies (Neigel, 2002; Heller and Siegismund, 2009; Bird *et al*., 2011).

Nei (1973, 1987) showed that the genetic frequency variation in a sub-structured population could be analyzed directly for any hierarchical structure of populations. He obtained the gene differentiation relative to the total population as:

$$G_{ST} = \frac{D_{ST}}{H_T} \quad (1)$$

Where D_{ST} is the average gene diversity between subpopulations, including the comparisons of subpopulations with themselves; H_T is the gene diversity in the total/pooled population (total population heterozygosity) and is given by:

$$H_T = 1 - J_T \quad (2)$$

J_T is the gene identity in the total population which is given by:

$$J_T = J_S - D_{ST} \quad (3)$$

By substituting equation 3 into equation 2, then:

$$H_T = 1 - J_S + D_{ST} \quad (4)$$

But the mean diversity within each population (average subpopulation heterozygosity) is given by:

$$H_S = 1 - J_S \quad (5)$$

From equations 4 and 5:

$$H_T = H_S + D_{ST} \quad (6)$$

Nei (1973) therefore showed that the gene diversity in the pooled population could be partitioned into the gene diversity within and between subpopulations.

Furthermore, if we rearrange equation 6:

$$D_{ST} = (H_T - H_S) \quad (7)$$

Substituting equation 7 into the equation 1:

$$G_{ST} = \frac{\left(H_T - H_S\right)}{H_T} \quad (8)$$

Equation 8 could further be broken down to give:

$$G_{ST} = 1 - \frac{H_S}{H_T} \quad (9)$$

Nei (1973) noted that if two alleles are present in a locus, the G_{ST} is identical to the Wright's F_{ST}. He further provided evidence that this could be applied to the case of multiple alleles. However, G_{ST} has been known to be affected by mutation and heterozygosity, in a way that makes comparison of loci that have high genetic mutation difficult (Hedrick, 1999). Hedrick (2005) reported that G_{ST} is dependent on the amount of genetic variation for highly variable genes like microsatellites where H_S and H_T can approach unity. Consequently G_{ST} may not range between 0 and 1 and could be very small even if the subpopulations have no overlapping sets of alleles (Hedrick, 1999).

Hedrick (2005), obtained a standardized G_{ST} by scaling the observed value of G_{ST} by the maximum level of G_{ST} obtainable ($G_{ST(max)}$), assuming the subpopulations are of equal sizes, k and the alleles within subpopulations are unique. Therefore,

$$G'_{ST} = \frac{G_{ST}}{G_{ST(\max)}} \quad (10)$$

$$\text{Where } G_{ST(\max)} = \frac{H_{T(\max)} - H_S}{H_{T(max)}} \quad (11)$$

$H_{T(\max)}$, is the maximum heterozygosity possible in the total population, given the observed heterozygosity within the subpopulations.

$$H_{T(\max)} = 1 - \frac{1}{k^2}\sum_i \sum_j p^2_{ij} \quad (12)$$

Where p_{ij} is the frequency of allele A_i in subpopulation j.

Combining equation 10 and 11,

$$G_{ST}^{'} = \frac{G_{ST}\left(k - 1 + H_S\right)}{\left(k - 1\right)\left(1 - H_S\right)} \tag{13}$$

Hedrick (2005) computed the average subpopulation heterozygosity as:

$$H_S = 1 - \frac{1}{k}\sum_i \sum_j p^2_{ij} \tag{14}$$

He computed the average within subpopulation homozygosity as:

$$1 - H_S = \frac{1}{k}\sum_i \sum_j p^2_{ij} \tag{15}$$

He then re-computed $H_{T(max)}$:

$$H_{T(max)} = \frac{\left(k - 1 + H_S\right)}{k} \tag{16}$$

Hedrick (2005) also recomputed G_{ST} as:

$$G_{ST} = \frac{\left(k - 1\right)\left(1 - H_S\right)}{k - 1 + H_S} \tag{17}$$

Reynold *et al*. (1983) utilized the coancestry coefficient (θ) initially proposed by Cavalli-Sforza and Bodmer (1971) and Cockerham (1973) as a measure of genetic distance. He defined θ as the probability that a random pair of genes at the same locus within a random chosen population is identical by descent. There are two general ways individuals are identical: animals could be identical by descent or identical by state. Reynold *et al*. (1983) also reported the genetic distance D as:

$$D = -\ln(1 - \theta) \tag{18}$$

Reynold *et al*. (1983) computed the coancestry, θ as:

$$\theta_l = \frac{a_l}{a_l + b_l} \tag{19}$$

Where,

$$a_l = \frac{[2\sum_{i=1}^{r} n_i \sum_{u=1}^{u_l} (\tilde{p}_{ilu} - \tilde{p}_{lu})^2 - (r-1)b_l]}{2(r-1)n_l} \tag{20}$$

$$b_l = 2\sum_{i=1}^{r} \frac{n_i \tilde{\alpha}_{il}}{r(2\bar{n}-1)} \tag{21}$$

For a sample of n_i individuals from the i-th replicate population (i = 1, 2,…., r), $\tilde{p}_{ilu}$ is the frequency of the u-th allele (u = 1, 2, …..,u_l) at the locus (l = 1, 2,…..,m). The expectation of $E[\tilde{p}_{ilu}] = p_{lu}$, then,

$$\bar{n} = \sum_{i=1}^{r} n_i / r \tag{22}$$

$$n_l = \frac{(r\bar{n} - \sum_{i=1}^{r} n_i^2 / r\bar{n})}{(r-1)} \tag{23}$$

$$\tilde{p}_{il} = \frac{\sum_{i=1}^{r} n_i \tilde{p}_{ilu}}{r\bar{n}} \tag{24}$$

$$\tilde{\alpha}_{il} = 1 - \sum_{u=1}^{u_l} \tilde{p}_{ilu}^2 \tag{25}$$

The methods for computing D and θ are based on assumptions and are more difficult to implement compared to Nei's or Wright's genetic distances. One assumption of Reynold *et al*. (1983) was that the divergence between populations with a common ancestry was based only on genetic drift.

Excoffier *et al*. (1992) developed the ϕ statistic to compute DNA haplotype divergence at different levels of hierarchical subdivisions using a linear model of the form:

$$p_{jig} = p + a_g + b_{ig} + c_{jig} \quad (26)$$

Where p_{jig} is the jth chromosome which is equivalent to the jth individual ($j = 1,\ldots..,\ I_g$), in the gth group ($g = 1,\ldots..,\ G$), and p is the overall true mean which is the unknown expectation of p_{jig} averaged over the whole study.

Excoffier *et al*. (1992) obtained the correlation of random haplotypes within populations relative to that of random pairs of haplotypes drawn from the whole species as:

$$\Phi_{ST} = \frac{\sigma_b^2 + \sigma_b^2}{\sigma^2} \quad (27)$$

The correlation of random haplotypes within a group of populations, relative to that of random pairs of haplotypes drawn from the whole species was computed as:

$$\Phi_{CT} = \frac{\sigma_a^2}{\sigma^2} \quad (28)$$

The correlation of the molecular diversity of random haplotypes within populations, relative to that of random pairs of haplotypes drawn from the region was computed as:

$$\left(\Phi_{SC}\right) = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_c^2} \quad (29)$$

σ_a^2, σ_b^2 and σ_c^2 are the variance components for the group effect (a), the population effect (b) and the individuals within population effect (c) and the sum of the variances gives σ^2.

Jost (2008) suggested another method to eliminate the effect of heterozygosity on G_{ST}. He developed a measure of genetic differentiation, D which quantifies genetic diversity in terms of effective number of alleles rather than heterozygosity. However, Ryman and Leimar (2009) showed that this measure of differentiation is affected by mutation and heterozygosity in a way

that is even more pronounced than G_{ST}. Bird *et al*. (2011), pointed out that overall, there is no single metric that best quantifies population genetic differentiation; recommending that researchers report both a fixation index and an index of genetic differentiation because they represent different properties of population partitioning.

2.7 Use of Genetic Programs in Animal Conservation Studies

To make computation of the genetic distances and other population parameters easier, several software programs have been developed. The STRUCTURE software was developed for multi-loci data to investigate population structure. The software is based on the Bayesian clustering approach and can be applied to various types of markers such as SNPs, RFLP and microsatellites (Pritchard *et al*., 2000). The method attempts to assign individuals to populations based on their genotypes while simultaneously estimating population allele frequencies (Pritchard *et al*., 2000; Falush *et al*., 2003 and 2007).

GENALEX is an Excel add-on which can be run on both PC and Macintosh (Peakall and Smouse 2006, 2012). The software accepts a number of data formats not limited to codominant genotypic data with 2 columns per locus; codominant and haploid raw allele frequency data; dominant (binary), haploid (including haplotypes) or sequence data coded numerically with 1 column per locus/base. The template and create options are useful in creating a suitable format for analysis in GENALEX. The software allows for various computations such as frequency based statistical procedures including Fst, Fit, Fis, private allele list, expected heterozygosity, Nei's genetic distances and Nei's genetic identity. The software plots several graphs such as allelic patterns and frequency graphs by locus or by population for each locus. It also outputs G-statistics and its analogues combined over all populations along with standard errors and confidence intervals. This option is useful for testing population differences. Other useful options

in GENALEX compute AMOVA as well as the Principal Coordinate Analysis (PCOA). These options provide a method to decompose total variations in a population.

The software program, FSTAT is a computer package for PCs which estimates and tests gene diversities and differentiation statistics including both Nei, and Weir and Cockerham estimators from codominant genetic markers (Goudet, 2002). FSTAT computes similar estimates to the GENALEX, but in addition, computes allelic richness using the method of rarefraction. Rarefraction is a method to standardize samples with varying sample sizes collected from different populations so as to make comparisons across the populations (Kalinowski, 2004; 2005).

GENEPOP is another program used in computing genetic diversity estimates (Rousset, 2008). GENEPOP computes exact tests for Hardy-Weinberg equilibrium, for population differentiation and for genotypic disequilibrium among pairs of loci (Rousset, 2008). It performs analyses of isolation by distance from pairwise comparison of individuals or population samples (Rousset, 2008). The F-statistics and private allele list from GENEPOP are similar to GENALEX. It also computes null alleles; a null allele is an allele that is present in an allele yet it is not amplified. Chapuis and Estoup (2006) reported that null alleles are commonly encountered in populations with large effective size, with an unusually high mutation rate in flanking regions, and that have differentiated from the population from which the cloned allele state was drawn and the primer designed. They further suggested that when populations are significantly differentiated, F_{ST} and genetic distances were overestimated in the presence of null alleles.

Some other programs developed for genetic diversity studies include: ARLEQUIN, POPGENE, DnaSP, PHYLIP and many more have been discussed by Labate (2000). Although

several software programs are constantly being developed or updated, they essentially compute similar estimates, however, with few differences.

To tackle the issue arising from data obtained from different countries, different laboratories or using different protocols, various merging programs have been developed (Paiva *et al*., 2011). The methods used in the past have been difficult to implement with large data sets. An obvious advantage of equating alleles between data sets into one large dataset is the increased accuracy that would be obtained thereby increasing power to detect linkages and association (Presson *et al*., 2006).

The use of MICROMERGE software program is most likely a better method. This program automates merging of Microsatellite data sets using the Bayesian statistical model that matches allele frequencies between datasets while the model selection is accomplished with the Markov Chain Monte Carlo (MCMC) algorithm that samples alignments according to their posterior probabilities (Presson *et al*., 2006). The observed bin (observed data set alleles) is aligned to a proposed set of theoretical alleles and each Micro-merge alignment is characterized by 3 parameters: the number of theoretical alleles; the theoretical frequencies and the partition (Presson *et al*., 2008). During each iteration, a change is proposed to one of the three alignment parameters of the MCMC chain, and the proposal is accepted or rejected by comparing the likelihoods of the proposed and current states (Presson *et al*., 2008). MICROMERGE is both quick and accurate (Presson *et al*., 2006) and aligns data sets marker by marker, matching each marker's allele frequencies while preserving size order (Presson *et al*., 2008).

The program requires a minimum of 5 input files: a micro-merge control file; 2 or more Mendel locus files; and 2 or more Mendel pedigree files for each data set to be merged (Presson, 2007). In addition to the required files, 1 micro-merge inclusion status file and 1 micro-merge

samples in common file may be included as optional files (Presson, 2007). In order for MICROMERGE to work efficiently, the datasets must consist of samples from the same population since allele frequencies vary across populations (Presson *et al*., 2008). Furthermore, a large sample size is needed to obtain accurate frequency estimates (Presson *et al*., 2008). The default sample size is equal to the size of the other data sets (with a minimum value of 200 samples, but the user can specify an alternative sample size) (Presson *et al*., 2008).

Some special features included in MICROMERGE version 2.0 such as one-to-one alignment and re-merging markers with low posterior probabilities have been developed to provide a more suitable alignment format for use by most other programs and improve alignment of markers that have low posterior probabilities and rare bins by zeroing these bins and re-merging the data (Presson *et al*., 2008). MICROMERGE was designed to merge multi-allelic marker datasets and may not be applicable to SNP datasets for two reasons. First, SNPs are designed to detect a specific nucleotide variation within a DNA sequence so the correspondence between the datasets is already known. Secondly, the major and minor alleles are ordered in the same way.

CHAPTER THREE

3.0 MATERIALS AND METHODS

3.1 Animal Care

Animal care and use committee approval was not necessary for this project because the data sets were obtained from preexisting databases.

3.2 Sample Collection, Genotyping and Description of Data

Pigs used for the analyses were chosen to be a representative sample of each breed. DNA samples were obtained from either the hair, semen, blood or the muscle. Genotyping using microsatellite markers of the Brazilian samples was done by Brazilian Agricultural Research Corporation Embrapa, while the genotyping using microsatellite markers of the US samples was done by the Agriculture and Agri-Foods Canada, Saskatoon. The US and Brazilian genotyped datasets were then stored in databases at the Genebank for the United States Department of Agriculture and the DNA Bank of the laboratory of Animal Genetics located at the Embrapa Genetic Resources and Biotechnology, Brasilia respectively. The genotyped datasets were retrieved from these databases and used for the study.

Table 3 and 4 show the summary of the BR and the US swine data sets used for study. Two hundred and twenty animals were sampled from eleven Brazilian breed while one hundred and seventy nine animals were sampled from seven US breeds. The sampled animals were chosen so that they were unrelated or at least separated by three generations. All the samples for the Brazilian data set were done in Brazil with the exception of the Uruguayan Mamelado. The Brazilian breeds consisted of eight naturalized breeds (MS60, Moura, Piau, Monteiro, Nilo, Mixed, Pietrain and Mamelado) and three commercial breeds (Large White, Landrace and

Duroc). The MS60 is an industry composite developed by crossing Duroc, Large White and Pietrain. The Mixed group consisted of locally adapted pig breeds: Caruncho, Casco de Burro and Pirapetinga. The US data contained three Chinese breeds (Meishan, Fengjing and Minzhu), Two English breeds (Yorkshire and Berkshire) and one US breed (Duroc). The Meishan breed (n =64) was subdivided into two groups: Meishan-China (n=22) originally imported from China in the 1980's and Meishan-US (n=42) from samples taken during the last decade from randomly mated populations. Thirty microsatellite markers from Brazil as well thirty five microsatellite markers from US were used to genotype the populations in their respective countries. There were no common samples analyzed in either laboratory.

The data sets were pooled after merging into two regions consisting of ten Brazilian breeds (MS60, Nilo, Moura, Monteiro, Duroc, Pietrain, Large White, Landrace, Mixed and Piau) in the Brazilian region and seven US breeds (Meishan-US, Meishan-BR, Fengjing, Minzhu, Yorkshire, Berkshire and Duroc) in the US region. The merged data set was also pooled into three regions consisting of ten Brazilian swine breeds as the Brazilian region; four Chinese breeds (Meishan-US, Meishan-BR, Fengjing, Minzhu) as the Chinese region; and three breeds (Yorkshire, Berkshire and Duroc) as the US-English region.

Table 3. Genetic Group and Sample Size of 30 Complete Microsatellite Markers Genotyped from 220 Brazilian Pigs with 14 Microsatellite Markers in Common with Pigs Sampled from the United States

Country Sampled	Genetic Group	Sample Size	Microsatellite Markers	
			Complete	Common
Brazil	Landrace	31	30	14
Brazil	Monteiro	37	30	14
Brazil	Moura	35	30	14
Brazil	Mamelado	16	30	14
Brazil	Piau	31	30	14
Brazil	MS60	48	30	14
Brazil	Duroc	4	30	14
Brazil	Large White	3	30	14
Brazil	Pietrain	4	30	14
Brazil	Nilo	5	30	14
Brazil	Mixed	6	30	14

Table 4. Genetic Group and Sample Size of 35 Complete Microsatellite Markers Genotyped from 179 United States Pigs with 14 Microsatellite Markers in Common with Pigs Sampled from Brazil

Country Sampled	Genetic Group	Sample Size	Microsatellite Markers	
			Complete	Common
United States	USA Meishan	42	35	14
United States	China Meishan	22	35	14
United States	Fengjing	22	35	14
United States	Minzhu	20	35	14
United States	Yorkshire	21	35	14
United States	Berkshire	26	35	14
United States	Duroc	26	35	14

3. 3 Genetic Population Analyses

All genetic diversity parameters with the exception of allelic richness were computed using GENALEX version 6.501 (Peakall and Smouse 2006, 2012). The number of alleles, Na was obtained by direct count and with the mean computed as the arithmetic mean across loci.

The number of private alleles was computed as the number of alleles unique to a single population in the dataset.

The effective number of alleles was obtained as:

$$Ne = \frac{1}{1 - H_E} \tag{30}$$

Where H_E is the expected heterozygosity or genetic diversity within a population and was computed as one minus the sum of squared allele frequency. The expected heterozygosity was computed per locus as shown below:

$$H_E = 1 - \sum p_i^2 \tag{31}$$

The mean heterozygosity or genetic diversity per population, $\bar{H}_E$ was computed as the expected heterozygosity averaged across populations. The average within population heterozygosity, H_S was computed the same way as the mean heterozygosity computed and is given as:

$$H_S = \bar{H}_E = \frac{\sum H_{Es}}{k} \tag{32}$$

Where H_{Es} is the expected heterozygosity in the s-th population and k is the number of populations.

Allele frequency was computed as:

$$F_x = \frac{2Nxx + Nxy}{2N} \tag{33}$$

Where N_{xx} is the number of XX homozygous individuals; N_{xy} is the number of heterozygous individuals, where Y can be any other allele and N is the number of samples.

F_{IS}, the inbreeding coefficient within individuals relative to the subpopulation was computed as:

$$F_{IS} = 1 - \frac{H_O}{H_S} \tag{34}$$

F_{IT}, the inbreeding coefficient within individuals relative to the total population was computed as:

$$F_{IT} = \frac{H_T - \overline{H}_O}{H_T} \tag{35}$$

Fst, the inbreeding coefficient within subpopulations, relative to the total was computed as:

$$F_{ST} = \frac{H_T - \overline{H}_E}{H_T} \tag{36}$$

H_T is the total expected heterozygosity if all populations were pooled (no subdivision) and is defined as:

$$H_T = 1 - \sum_{i=1}^{h} \overline{p}_i^{\,2} \tag{37}$$

The observed heterozygosity, H_o for a single locus within a population was computed as:

$$H_o = \frac{Number\ of\ Heterozygotes}{N} \tag{38}$$

Where the number of heterozygotes was determined by direct count and N is the sample size.

The mean observed heterozygosity, $\overline{H}_o$ was computed as:

$$\overline{H}_o = \frac{\sum H_{Os}}{k} \tag{39}$$

Nei's genetic distance D, was computed as:

$$D = -\ln(I) \tag{40}$$

Where.

$$I = \frac{J_{XY}}{\sqrt{(J_X J_Y)}} \tag{41}$$

$$J_{XY} = \sum_{i=1}^{k} P_{ix} P_{iy} \tag{42}$$

$$J_X = \sum_{i=1}^{k} p_{ix}^2 \tag{43}$$

$$J_Y = \sum_{i=1}^{k} p_{iy}^2 \tag{44}$$

J_{XY}, J_X and J_Y were summed over all loci and alleles and divided by the number of loci to get the average values which were then used to compute the Nei's genetic identity, I. Also, p_{ix} and p_{iy} are the frequencies of the i-th allele in populations x and y

The number of migrants was computed as:

$$Nm = \frac{\left(\frac{1}{F_{ST}}\right) - 1}{4} \tag{44}$$

The Analysis of molecular variance (AMOVA) procedure was computed following Excoffier *et al*. (1992); Huff *et al*. (1993); Excoffier and Lischer (2010).

The principal coordinate analysis was as described by Orloci (1978). Genetic distance matrices were obtained separately for the common and the complete marker datasets as well as

the merged data set. Variance-covariance matrices were obtained from the genetic distances and then standardized by multiplying each element in the genetic distance matrix by the square root of n-1. The principal components were computed from the standardized covariances. The first and second principal components were plotted for each data set. Principal coordinate analysis was also performed for breed averages for the common marker dataset, the complete marker dataset for both the US and the Brazilian breeds and merged data set.

To compute allelic richness, the genotype frequency was first converted to GENEPOP format using GENALEX and then converted to FSTAT format. Allelic richness was computed using rarefraction method in FSTAT (Goudet, 2002). The p-values obtained after 2720 permutations with 5% adjusted nominal level for multiple comparisons of the genetic differentiation was 0.000368 for the merged data set; The p-values obtained after 60 permutations with 5% adjusted nominal level for multiple comparisons of the genetic differentiation was set at 0.016667 when the merged populations was pooled into three regions; The p-values obtained after 20 permutations with 5% adjusted nominal level for multiple comparisons of the genetic differentiation was 0.05 when the merged populations was pooled into two regions.

3.4 Population Structure Analysis

The datasets were converted from GENALEX format into structure format using CONVERT version 1.31 (Glaubitz, 2005). Structure analysis was done using STRUCTURE software version 2.3 developed by Pritchard *et al*. (2010). Missing data were represented as -9. The data contained individual labels, population identifiers and row of marker names. The program was started with a random configuration before iterating through steps dependent only on the parameter values at the previous step. The correlations induced between the state of the

Markov chain at different points during the run was eliminated or made negligible by running several simulations.

Table 5 shows a summary of MCMC parameters set used to run structure for the unmerged data sets. Analysis was done separately for the US and the Brazilian datasets using all the Microsatellite markers and then using only the common set of markers. MCMC selection criteria were based on the consistency of the MCMC results as shown by the line scatterplots.

An MCMC combination of burn-in length of 50000, MCMC number of replications of 100000 with 5 replications was chosen as parameters to run STRUCTURE software for the merged data set since the STRUCTURE runs for the common markers showed very similar variances. The number of Bayesian clusters for the merged population was set at k = 18.

Structure runs were also done for the pooled data sets for k =2 for the 2 regions and k = 3 for the three regions using burn-in length of 50000, MCMC replication of 100000 and 5 iterations.

3.5 Merging Procedure

Merging of the data sets was done using MICROMERGE software program. MICROMERGE software program merges data sets based on allele frequencies. The allele frequencies computed with GENALEX for the Brazilian and US datasets were used as MICROMERGE inputs. The locus, pedigree and the inclusion status files were coordinated so that the markers appeared in the same order. There were fourteen microsatellite markers common to both datasets. Merging was done using the allele frequencies of the fourteen common markers and the genotype information both from the US and the Brazilian datasets. Two locus and 2 pedigree files were created one for the Brazilian pig data and the other for the US pig data. The locus files were formatted in comma-separated Mendel format. The alleles and their

corresponding frequencies for each marker were ordered in ascending allele size order in both locus files.

The software program requires a pedigree file to be available or simulated. Since no pedigree data was available, the pedigree file was simulated by generating a list of sequential numbers for the individual and its parents. The pedigree files were formatted in comma-separated Mendel pedigree format. The individual genotypes were formed by separating allele pair for each locus with a dash. The genotype data started from column 7 in the pedigree file. The Missing genotypes were left blank. The control file was named as control.txt.

The minimum posterior probability for aligning each marker using a one-to-one alignment was set at 0.575. For the first round of the merging procedure, the burn-in length was set at 5000 with a MCMC number of iterations set at 5000000 and a genotype error rate of 0.05. The burn-in length for the second round of the merging procedure was increased to 10000and a rare allele threshold of 0.05 was included. A rare allele threshold was included because the presence of rare alleles may lead to low posterior probability, rejection of the alignment, and the loss of data for an entire marker (Presson *et al*., 2008).

MICROMERGE carries out model selection by an MCMC algorithm that samples alignments according to their posterior probabilities. Each iteration of the MCMC algorithm proposes a change to an alignment characteristic, which is then accepted or rejected by a Metropolis-Hastings step (Metropolis *et al*., 1953; Hastings, 1970). Since the dimensionality of the parameter space changes whenever the number of allele changes, micro-merge uses a reversible jump Markov chain (Green, 1995; Presson *et al*., 2006).

Once the microsatellites are merged we explore genetic diversity parameters within and among breeds for the two countries.

3.6 Positive Control

The GENALEX results from the unmerged data sets was used as positive control to compare GENALEX results obtained after merging the common marker sets using MICROMERGE. Since the merged data was obtained from the common markers from both the US and BR data sets, the success of the merging procedure was accessed by comparing the genetic analysis for the common markers and the complete markers with the merged data set.

Table 5. Burn-in Length, Markov Chain Monte Carlo (MCMC) Number of Replications and Iterations for a Fixed Bayesian Population, K, Analyzed Using Complete Microsatellite Markers Genotyped from Pigs in Brazil and United States with 14 Microsatellite Markers Common between the Countries

Burn-in	MCMC	Numberof Iterations	K
50000	100000	1	5
50000	100000	5	5
50000	300000	1	5
50000	300000	5	5
100000	100000	1	5
100000	100000	5	5
100000	300000	1	5
100000	300000	5	5
200000	600000	1	5
200000	600000	5	5

Number of complete microsatellite markers for BR breeds = 30
Number of complete microsatellite markers for US breeds = 35

CHAPTER FOUR

3.0 RESULTS AND DISCUSSION

The results of the population structure analysis as well as the analysis of molecular variance for the unmerged datasets, that is, the complete markers as well as the common markers will be used as evidence to support the merging procedure. The results of the merging procedure as well as results of post merging analyses would be compared to the pre merged datasets to determine the success of the merging procedure. Post merging analyses including genetic analysis and population structure analysis would then be presented for the merged data as well as the pooled data set.

The estimated log likelihood for the complete and common microsatellite markers genotyped from pigs in Brazil and the United States are shown in tables 18 and 19. Several runs from K = 1 to K = 5 were done to determine the appropriate burn-in length and the MCMC replication after burn-in that would be used to run structure on the merged data set. Each K-th replicate has its corresponding estimated log likelihood. The estimated log likelihood was similar from K=1 to K=5 for burn-in periods of 50000, 100000 and 200000 and MCMC replication after burn-in period of 100000,300000 and 600000 for the complete microsatellite markers in both the Brazilian and US datasets. The same trend was observed in the common microsatellite markers.

This is an important result because instead of using much higher burn-in periods and MCMC replication after burn-in, for example, a burn-in periods of 200000 and MCMC replication after burn-in of 600000, we would obtain similar population structures from our target data (merged data set) but with the added advantage of reducing the computational time if we used a burn-in period of 50000 and MCMC replication after burn-in period of 10000. The

common marker sets for both the US and Brazilian data sets had the largest estimated log likelihood. The estimated log likelihood was lowest for the US complete genotyped data set for K=1 to K=5.

A slight decrease in estimated log likelihood was observed at K = 3 for all 4 major combinations with the exception of the Brazilian data set when only 14 marker sets were used. This could indicate the number of distinct populations present in the data set. The Brazilian data set with 14 markers showed little deviations for K = 1 to K = 5 for all 5 MCMC combinations. This is most likely the reason why the deviation at K = 3 was not noticeable. As the number of Bayesian clusters (K) increased from K =3 to K = 5, the estimated log likelihood became more uniform for all 4 major clusters. The model choice criterion was based on determining a constant MCMC combination across all the data sets. Pritchard *et al*. (2010) suggested that determining the appropriate burn-in length is important to minimize the effect of the starting configuration while determining the number of MCMC replication after burn-in is important to get accurate parameter estimates.

Several approaches have been suggested to determine the best K; one approach is based on the log likelihood for each K and is represented as Ln P(D), where D is the data (Pritchard *et al*., 2000). The second method computes delta K (ΔK) based on the second order rate of change of the likelihood (ΔK) (Evanno *et al*., 2005). Evanno *et al*. (2005) reported that as K approached the true value, ΔK shows a clear peak. However, since we set our K to 5, this approach was inappropriate for our analysis. The L(K) values are given as the average of the Ln P(D) values.

Table 6 shows the AMOVA from microsatellite markers from pigs in Brazil and US after 999 permutations. Twenty eight percent of the variation in the US pig population was observed among the breeds for both the common marker set and for the complete marker set. Within

individual and within population differences for this dataset decreased from seventy percent to sixty seven percent when the number of loci genotyped increased from 14 to 35 in the US pig dataset. This decrease corresponded with a percent increase in among individual between breed differences. However, these represent minor changes in how genetic variations are distributed.

The within individual and within population differences for the Brazilian swine population did not change when the number of loci genotyped increased from 14 to 30. However, the among population difference decreased by one percent while among individual between breed difference increased by one percent.

The result shows that majority of the genetic variation occurs at the individual level which is consistent with other reports in the literature (Blackburn *et al*., 2006; Swart *et al*., 2010; Paiva *et al*., 2010). It is possible to combine the breeds for each country as one population and still not lose a lot of significant variation between breeds or subpopulations.

The results of the MCROMERGE merging procedure are shown in tables 7 to 11. The posterior probabilities of the merged and re-merged markers of the first (A1) and second (A2) top alignments are shown in tables 7 and 8. The difference between the top two alignments is given in column four. The one-to-one alignment default setting chooses the highest likelihood to represent the lumped alignment. The likelihood is computed from the product of the genotype frequencies of the data sets and the bin frequencies were obtained from the theoretical allele frequencies (Presson *et al*., 2008). Of the fourteen markers common to both the Brazilian and the US datasets, three markers (SW240, SW951 and SW857) had top alignments with posterior probability greater than 0.575 which is the minimum posterior probability that must be attained for a marker to 'pass' and not require re-merging if the one-to-one alignment option is chosen (Presson, 2007). The other eleven markers with posterior probability less than 0.575 were

remerged. After remerging the markers, four markers still had posterior probabilities below 0.575. Of the four markers with low posterior probabilities, aligning marker SW632 was the most problematic due to very high amount of missing data especially in the Brazilian data set and was consequently removed. The breed Mamelado was also deleted due to missing data across all markers. Although the posterior probability of marker SW830 was below the minimum threshold, Micro-merge did not remerge this marker; however, it was still included in subsequent runs.

The posterior probabilities of the merged and re-merged markers of the first (A1) and second (A2) top alignments for second round of iterations are shown from tables 9 and 10. Markers SW951, SW857 and S0225 had posterior probabilities greater than 0.575. Ten markers were re-merged because their posterior probabilities were below the minimum posterior probability threshold. All markers with the exception of markers SW24 and S0226 had posterior probabilities above 0.575 after re-merging. The minimum posterior probability threshold for one-to-one alignment was reduced from 0.575 to 0.550 to include marker SW24 but marker S0226 was discarded.

4.1 Positive Control

Merging results for the final 12 microsatellite markers are shown in table 11. The posterior probabilities for the final 12 markers ranged from 0.554 to 0.924. Based on the minimum posterior probability for inclusion of alleles in the merged data (0.550), the merging process was successful. Presson *et al*. (2008) reported similar low posterior probability for lumped alignments while the one-to-one alignment increased the posterior probability of the markers. Presson *et al*. (2008) also reported that the inclusion of a rare allele threshold increased

the posterior probability of microsatellite markers. The merging result reported in this study was therefore comparable to Presson *et al*. (2008).

The final merged data contained 383 pigs; 10 Brazilian pig breeds and 7 US pig breeds. Mean allelic pattern for the unmerged Brazilian and US swine breeds as well as the merged data set are shown in figures 1 and 2. In the Brazilian data, Piau had the largest mean number of alleles of 6.53 when all 30 markers were analyzed and 2.53 for the common markers. Duroc had the smallest mean number of alleles of 2.53 for the complete marker set while Large White had the least number of alleles of 2.43 for the common set of markers. The mean effective number of alleles was highest for Piau for both the common set of markers (6.53) and the complete marker set (2.53). Duroc had the smallest mean effective number of alleles of 1.99 for the complete markers while Large White had the smallest mean effective number of alleles for the common marker set.

MS60, Moura, large White, Nilo, Landrace and Pietrian all had slightly lower mean expected heterozygosity compared to mean observed heterozygosity in the common marker set as well as the complete marker set. The mean expected heterozygosity is approximately the same as the mean observed heterozygosity for MS60. Any difference may have been due to computational error. This would therefore result in negative fixation indexes. Negative indexes indicate excess heterozygosity due to negative assortative mating or heterotic selection (Peakall and Smouse, 2012). Negative assortative mating occurs when two phenotypically dissimilar breeds are mated to one another. The difference between the expected and observed heterozygosity was highest for the mixed breed, otherwise the difference was essentially zero or very close to zero in Duroc, Piau, Monteiro and Mamelado. The number of private alleles for all the Brazilian breeds was close to zero, with Piau and Moura just above 0.14. The mean allelic

richness appeared to be generally similar to the mean effective number of alleles, although values for the mean allelic richness were smaller.

The USA dataset showed that the highest mean number of alleles was observed in the Ming breed for both the common set of markers and the complete marker set. The smallest mean number of alleles and mean effective number of alleles was observed in Berkshire in the common marker set. However, the US Meishan breed showed the least mean number of alleles. The expected heterozygosity was smaller than the observed heterozygosity for the Fengjing breed. The average number of private alleles for the US data was relatively high in Fengjing, Ming, Yorkshire, Duroc and Berkshire.

Private alleles have been implicated as an indicator of gene flow. Slatkin (1985) and Barton and Slatkin (1986) showed that private alleles was related to the mean number of migrants exchanged per generation between population.

The genetic distances between the Brazilian breeds using the common and complete markers are shown in tables 12. As expected, the genetic distances for the complete marker set was the same or slightly higher the genetic distances for the common marker set. The exceptions were genetic distances that involved the Mixed group as well as the Moura breed. This may likely be the consequence of much higher number of missing alleles resulting from pooling 2 breeds which individually have a high number of missing data. For the common microsatellite markers, lowest genetic distance was observed between MS60 and Large White while the highest genetic distance was observed between the Brazilian Duroc and the Mixed group (0.70). The genetic distances between the Mixed group and other Brazilian breeds were relatively high. For the complete marker set, the lowest genetic distance was observed between the Brazilian Duroc

and Piau (0.21) while the highest genetic distance was observed between the Mixed group and Pietrain.

The genetic distances between the US breeds using the common and complete markers are shown in tables 13. The result also showed that the genetic distances from the complete marker set were slightly higher than the genetic distances obtained using the common marker set. The noticeable exceptions were observed with the Berkshire and Duroc breeds. The percentage of missing values was 7% each for both Duroc and Berkshire at locus S0068. Although slightly lower distances were observed for the China Meishan and Minzhu using the complete marker set, this most likely was due to computational errors rather than missing values. The smallest genetic distance was observed between the Meishan sub-populations for the common marker set and for the complete marker set. The highest genetic distance was observed between Mesishan-CH and Berkshire for both the common marker set (1.73) and the complete marker set (1.55).

The estimates of the number of alleles, effective number of alleles, number of private alleles, allelic richness, and the observed and expected heterozygosity for the complete and common marker sets were similar. The mean allelic pattern as well as the genetic distances between breeds for merged data set appeared to be similar with the estimates for the common marker set.

Table 6. Analysis of Molecular Variance (%) of Complete Microsatellite Markers Genotyped from Pigs in Brazil and the United States with 14 Microsatellite markers in Common to both Countries

	Complete Markers		Common Markers	
	BR	US	BR	US
Among Population	12	28	11	28
Among Individual	17	5	18	2
Within Individual	71	67	71	70

Table 7. First Round of Merging Showing Posterior and Overlapping Probabilities with Difference between Posterior Probability of Top Two Alignments, A1 and A2 Using 14 Microsatellite Markers Common between Pigs in Brazil and the United States

	Posterior Probabilities			Overlapping Probabilities	
Locus	A1	A2	Difference	A1	A2
SW632	0.409	0.328	0.081	0.905	0.899
SW24	0.403	0.384	0.019	0.903	0.898
S0226	0.337	0.077	0.260	0.849	0.778
SW240	0.657	0.127	0.530	0.939	0.847
S0227	0.564	0.272	0.292	0.928	0.876
SW951	0.889	0.070	0.819	0.986	0.852
SW2406	0.462	0.403	0.059	0.913	0.905
SW830	0.414	0.142	0.272	0.881	0.790
SW2008	0.496	0.099	0.397	0.868	0.777
SW857	0.906	0.049	0.857	0.981	0.829
S0026	0.376	0.201	0.175	0.916	0.916
SW2410	0.149	0.142	0.007	0.837	0.817
S0225	0.567	0.153	0.414	0.873	0.727
SW72	0.455	0.276	0.179	0.850	0.872

Table 8. First Round of Re-merging Showing Posterior and Overlapping Probabilities with Difference between Posterior Probability of Top Two Alignments, A1 and A2 Using 14 Microsatellite Markers Common to Pigs in Brazil and the United States

	Posterior Probabilities			Overlapping Probabilities	
Locus	A1	A2	Difference	A1	A2
SW632	0.301	0.267	0.034	0.839	0.800
SW24	0.918	0.049	0.869	0.976	0.707
S0226	0.324	0.220	0.104	0.855	0.847
S0227	0.855	0.058	0.797	0.973	0.831
SW2406	0.869	0.108	0.761	0.976	0.850
SW2008	0.628	0.124	0.504	0.923	0.827
S0026	0.652	0.209	0.443	0.972	0.972
SW2410	0.329	0.313	0.016	0.888	0.885
S0225	0.550	0.436	0.114	0.930	0.913
SW72	0.775	0.208	0.567	0.937	0.774

Table 9. Second Round of Merging Showing Posterior and Overlapping Probabilities with Difference between Posterior Probability of Top Two Alignments, A1 and A2 Using 14 Microsatellite Markers Common to Pigs in Brazil and the United States

	Posterior Probabilities			Overlapping Probabilities	
Locus	A1	A2	Difference	A1	A2
SW24	0.244	0.194	0.050	0.716	0.711
S0226	0.275	0.153	0.122	0.845	0.804
SW240	0.512	0.174	0.338	0.895	0.772
S0227	0.537	0.302	0.235	0.923	0.878
SW951	0.813	0.120	0.693	0.975	0.855
SW2406	0.468	0.354	0.114	0.906	0.900
SW830	0.400	0.119	0.281	0.880	0.807
SW2008	0.500	0.146	0.354	0.884	0.816
SW857	0.924	0.049	0.875	0.986	0.833
S0026	0.454	0.224	0.230	0.923	0.859
SW2410	0.304	0.210	0.094	0.825	0.810
S0225	0.906	0.041	0.865	0.985	0.847
SW72	0.501	0.111	0.390	0.880	0.752

Table 10. Second Round of Re-merging Showing Posterior and Overlapping Probabilities with Difference between Posterior Probability of Top Two Alignments, A1 and A2 Using 14 Microsatellite Markers Common to Pigs in Brazil and the United States

	Posterior Probabilities			Overlapping Probabilities	
Locus	A1	A2	Difference	A1	A2
SW24	0.554	0.385	0.169	0.727	0.579
S0226	0.287	0.168	0.119	0.764	0.721
SW240	0.764	0.175	0.589	0.948	0.809
S0227	0.870	0.044	0.826	0.978	0.822
SW2406	0.863	0.107	0.756	0.975	0.849
SW830	0.966	0.014	0.952	0.992	0.752
SW2008	0.572	0.216	0.356	0.887	0.792
S0026	0.811	0.128	0.683	0.985	0.985
SW24110	0.744	0.136	0.608	0.952	0.868
SW72	0.843	0.114	0.729	0.955	0.712

Table 11. Comparing the Number of Alleles per Locus between the Final Merged Microsatellite Markers and the Microsatellite Markers in Common Prior to Merging for Pigs in Brazil and the United States Including the Posterior Probability during Merging

	Allele			Posterior
Loci	US	BR	Final	Probabilities
SW24	12	10	15	0.554
SW240	13	10	14	0.764
S0227	9	10	12	0.870
SW951	5	10	12	0.813
SW2406	11	10	14	0.863
SW830	9	9	13	0.966
SW2008	8	9	12	0.572
S0026	9	7	11	0.811
SW2410	13	7	17	0.744
SW72	8	5	9	0.843
SW857	9	8	11	0.924
S0225	11	6	12	0.906

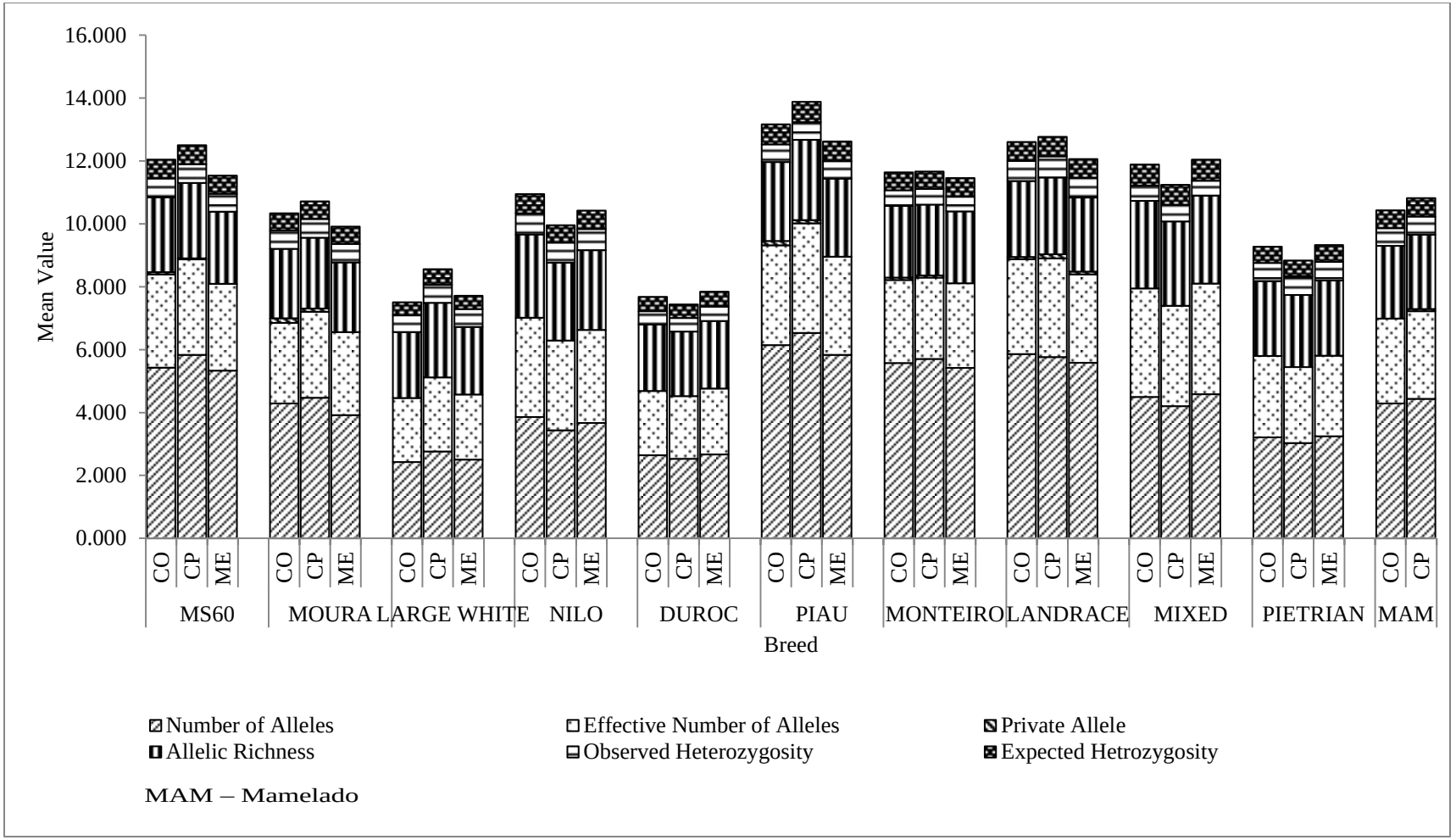


Figure 1. Number of Alleles, Effective Number of Alleles, Number of Private Alleles, Allelic Richness and Observed and Expected Heterozygosity for Complete (CP) and Merged (ME) Microsatellite Markers genotyped from Brazilian Pigs with Microsatellite Markers in Common (CO) with Pigs from the United States

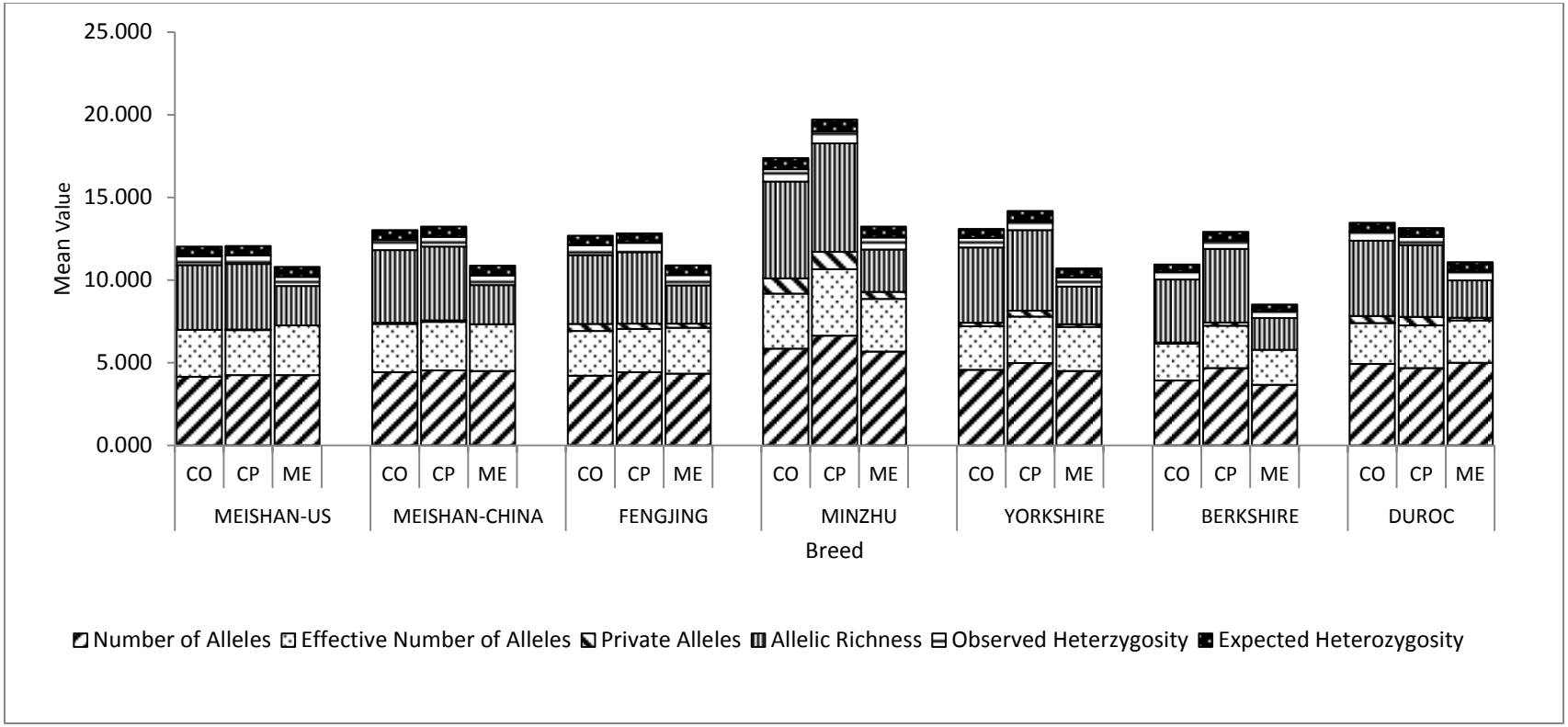


Figure 2. Number of Alleles, Effective Number of Alleles, Number of Private Alleles, Allelic Richness and Observed and Expected Heterozygosity for Complete (CP) and Merged (ME) Microsatellite Markers genotyped from United States Pigs with Microsatellite Markers in Common (CO) with Pigs from Brazil

Table 12. Nei's Genetic Distances Using 30 Complete Microsatellite Markers (Upper Diagonal) Genotyped from 11 Brazilian Swine Breeds and Using 14 Microsatellite Markers in Common (Lower Diagonal) with the United States Pigs

				Breed						
1	2	3	4	5	6	7	8	9	10	11
0.00	0.28	0.37	0.38	0.32	0.25	0.41	0.21	0.43	0.22	0.31
0.30	0.00	0.53	0.42	0.27	0.29	0.44	0.41	0.40	0.56	0.47
0.29	0.36	0.00	0.65	0.57	0.43	0.53	0.43	0.58	0.54	0.44
0.31	0.47	0.52	0.00	0.48	0.21	0.22	0.39	0.36	0.53	0.53
0.31	0.28	0.33	0.41	0.00	0.32	0.56	0.40	0.61	0.66	0.48
0.18	0.29	0.30	0.21	0.29	0.00	0.20	0.26	0.28	0.44	0.36
0.39	0.51	0.44	0.33	0.51	0.20	0.00	0.44	0.42	0.51	0.45
0.15	0.45	0.40	0.38	0.37	0.19	0.42	0.00	0.50	0.38	0.39
0.43	0.50	0.65	0.46	0.70	0.32	0.56	0.50	0.00	0.62	0.49
0.21	0.50	0.47	0.49	0.50	0.27	0.40	0.29	0.62	0.00	0.39
0.23	0.48	0.38	0.45	0.36	0.26	0.34	0.19	0.67	0.29	0.00

1 – MS60, 2 – Moura, 3 – Large White, 4 – Nilo, 5 Brazilian Duroc, 6 – Piau, 7 – Monteiro, 8 – Landrace, 9 – Mixed, 10 – Pietrain, 11 – Mamelado

Table 13. Nei's Genetic Distances Using 35 Complete Microsatellite Markers (Upper Diagonal) Genotyped from 7 United States Swine Breeds and Using 14 Microsatellite Markers in Common (Lower Diagonal) with Brazilian Pigs

		Breed				
1	2	3	4	5	6	7
0.00	0.11	0.81	0.90	1.37	1.25	1.33
0.15	0.00	0.84	0.85	1.53	1.55	1.41
0.87	0.78	0.00	0.75	1.23	1.28	1.36
0.75	0.66	0.65	0.00	0.63	0.62	0.71
1.21	1.35	0.91	0.64	0.00	0.36	0.50
1.33	1.73	1.01	0.60	0.26	0.00	0.48
1.04	1.10	1.44	0.74	0.49	0.54	0.00

1 – US-Meishan, 2 – China-Meishan, 3 – Fengjing, 4 – Minzhu, 5 – Yorkshire, 6 – Berkshire, 7 – US Duroc

4.2 Within Breed Diversity for Merged Data Set

This result from the merging procedure showed that in general, for similar number of microsatellite markers and number of breeds, the merged data correctly represented the unmerged Brazilian and US data.

Figure 3 shows the observed, expected and total heterozygosity with fixation indices (F_{IS}, F_{IT} and F_{ST}) per locus from 12 merged microsatellite markers common to pigs in the United States and Brazil. Fourteen microsatellite loci were genotyped from two hundred and twenty Brazilian and US swine breeds. The average within population heterozygosity ranged from (S0225) 0.38 to (SW24) 0.74 with a mean of 0.60. Five microsatellite loci had average within population heterozygosity below the within population heterozygosity averaged over populations and loci. These microsatellite markers were S0227 (0.43), SW830 (0.566), SW2008 (0.55), SW2410 (0.56) and S0225 (0.38); and of these loci, two were below one standard deviation from the mean (S0227 and S0225). The mean observed heterozygosity ranged from S0225 (0.28) to SW240 (0.74) with a mean of 0.5628. Like the average within population heterozygosity and mean observed heterozygosity, S0225 (2.31) had the lowest mean allelic richness. The highest mean allelic richness was observed at microsatellite loci SW857 (3.23). High inbreeding coefficient (positive F_{IS}) was observed within the populations in each locus except at loci SW240 (-0.07), SW2008 (-0.04), SW857 (-0.01), SW2410 (-0.22) and SW72 (-0.11), where negative inbreeding coefficient were observed.

Figure 4 shows the mean allele pattern across seventeen breeds genotyped at 12 microsatellite loci. The highest mean number of alleles was observed for Piau (5.83) followed by Ming (5.67) while the smallest mean number of alleles (2.43) was observed for Large White. The mean number of alleles for the Duroc from the United States was closer to MS60 (5.33) and

Monteiro (5.42), than the Brazilian Duroc (2.67). The mean effective number of alleles was 2.560 for the US Duroc and 2.100 for the Brazilian Duroc. The mean effective number of alleles for Large White was the same as the mean effective number of alleles for the Brazilian Duroc. The highest mean effective number of alleles was observed in the mixed breed while the smallest mean effective number of alleles (2.07) was observed for Large White. The highest and lowest observed and expected heterozygosity were observed in the US breeds compared. Among all Brazilian breeds, only the Landrace had private alleles. Private alleles were observed in Fengjing (0.25), Ming (0.42), Yorkshire (0.17) and Duroc (0.17).

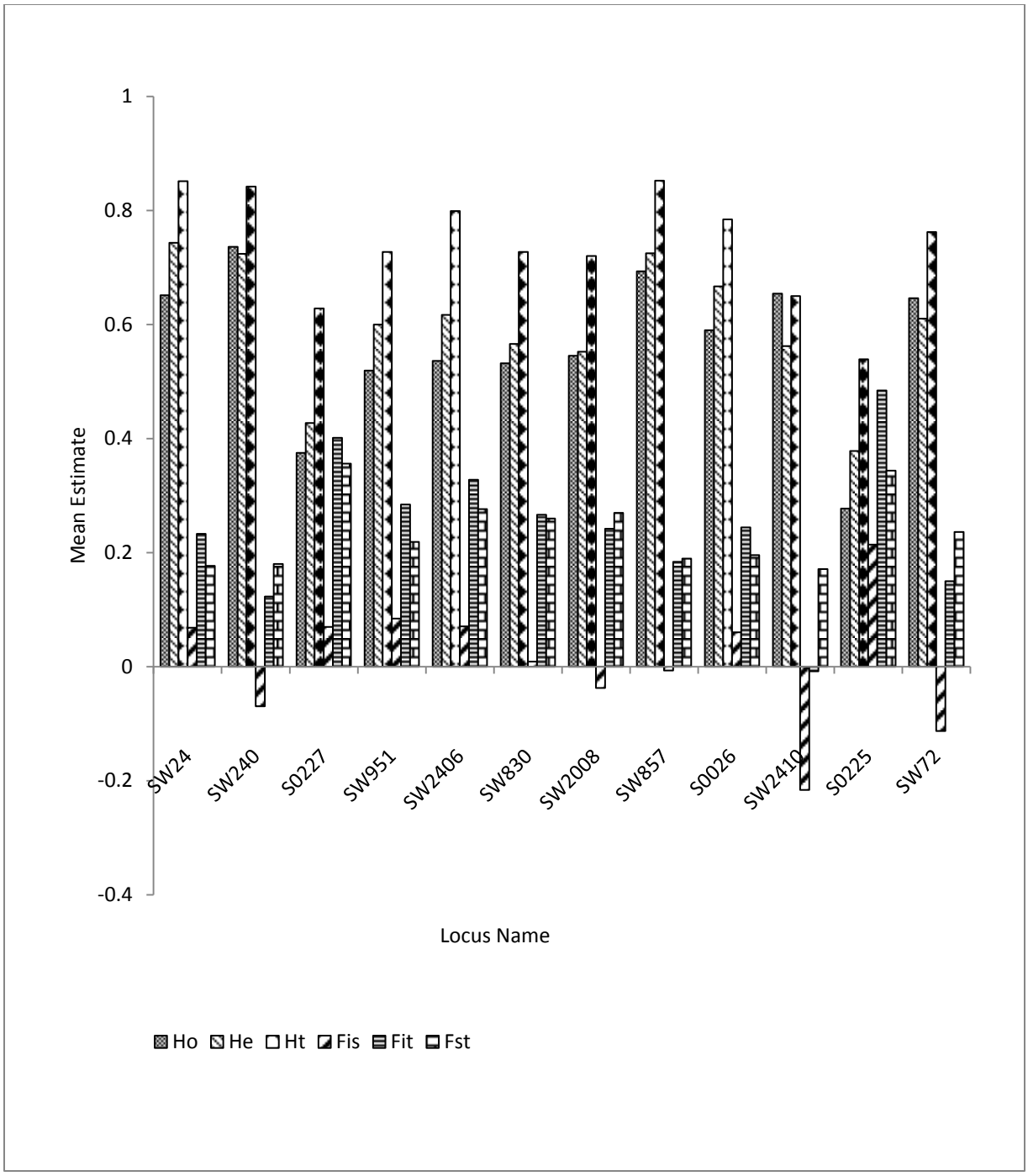


Figure 3. Observed, Expected and Total Heterozygosity with Fixation Indices (F_{IS}, F_{IT} and F_{ST}) Per Locus from 12 Merged Microsatellite Markers Common to Pigs in the United States and Brazil

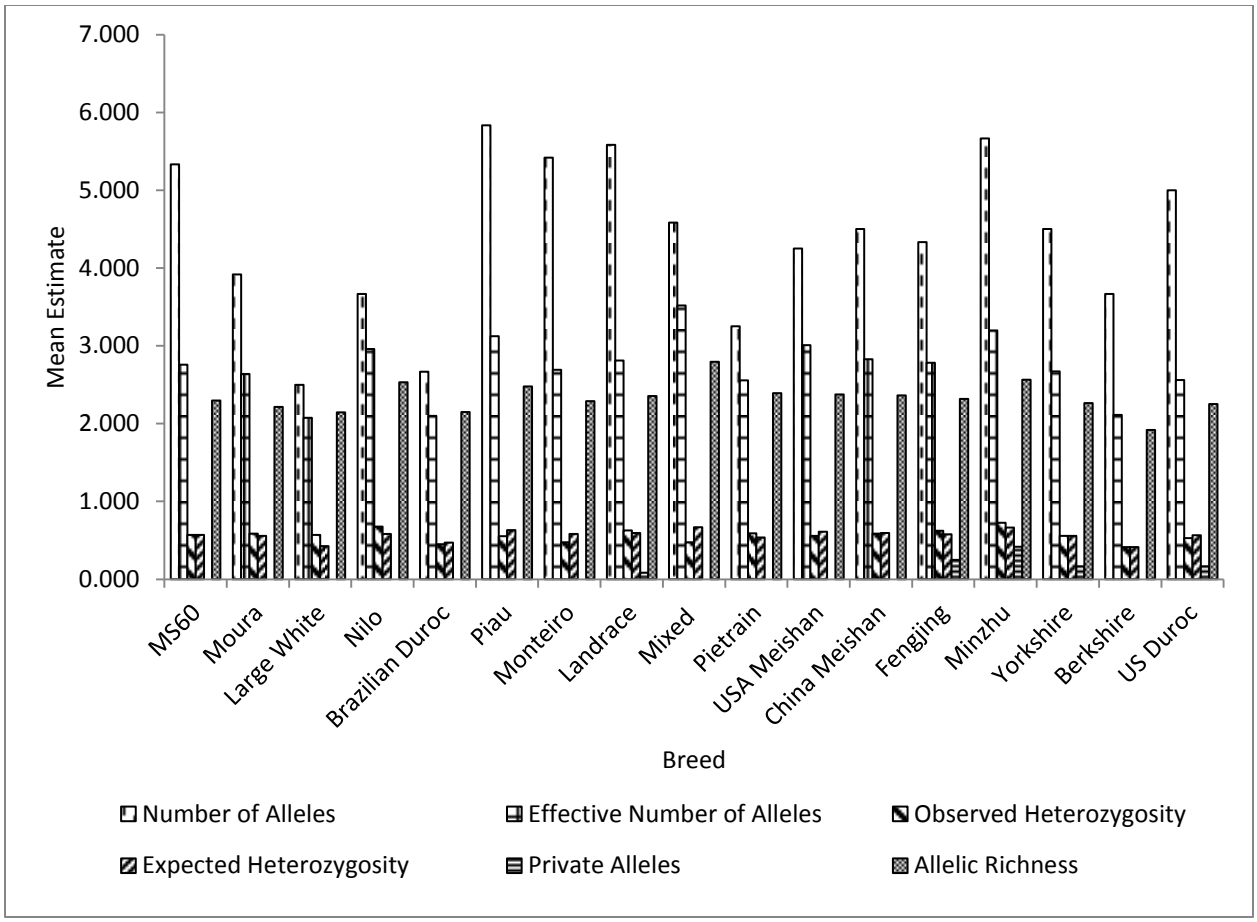


Figure 4. Number of Alleles, Effective Number of Alleles, Observed and Expected Heterozygosity, Number of Private Alleles and Allelic Richness for 12 Merged Microsatellite Markers Genotyped from 17 Pigs in Brazil and the United States

4.3 Between Breed Diversity for Merged Data set

F_{ST} values ranged from 0.03 to 0.44 with a mean of 0.19 (table 14). F_{ST} values were higher when the Brazilian breeds were compared to the US breeds. The highest F_{ST} value was highest between Meishan and Berkshire. F_{ST} value between the US Meishan and Berkshire (F_{ST} = 0.40, $p < 0.001$) was slightly lower than the F_{ST} value between the Chinese Meishan and Berkshire. The smallest F_{ST} was observed between Nilo and Piau. F_{ST} values between the Duroc from Brazil and that from the US was 0.20. Genetic distance ranged from 0.12 to 1.87 with a mean of 0.59 (table 14). Like the F_{ST} values, genetic distance was highest between Berkshire and the Meishan breeds. Genetic distance was also high between Berkshire and the Meishan subpopulations ($p < 0.00037$). The smallest genetic distance was between the two Meishan subpopulations (0.12, $p < 0.00037$)). Genetic distance between the Duroc breeds was 0.50 ($p > 0.00257$) and this was below the mean genetic distance. The genetic distances between MS60 and Pietrain, Large White and the Brazilian Duroc were 0.19 ($p > 0.01654$), 0.25 and 0.25 ($p > 0.00294$) respectively. The p-value for the genetic distance between MS60 and Large White could not be computed because of the small sample size for Large White (n=3). The genetic differentiation between MS60 and the three breeds corresponded to F_{ST} of 0.05, 0.09 and 0.09 respectively. The genetic distances between each Chinese breed (the US and the Chinese Meishan, Fengjing and Ming) and Berkshire ($p < 0.00037$) were quite large. The highest mean effective number of alleles was observed between Nilo and Piau. Higher mean effective number of alleles was observed among the Brazilian swine.

Table14. Nei's Genetic Distances (Lower Diagonal) and Fixation index (Upper Diagonal) Using 12 Merged Microsatellite Markers Common to Pigs in Brazil and the United States

Breed																
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
0.00	0.08	0.10	0.07	0.09	0.05	0.10	0.05	0.09	0.06	0.17	0.17	0.15	0.11	0.10	0.13	0.12
0.26	0.00	0.15	0.10	0.10	0.06	0.12	0.11	0.10	0.11	0.17	0.18	0.16	0.14	0.12	0.16	0.18
0.25	0.42	0.00	0.13	0.14	0.10	0.14	0.13	0.16	0.14	0.25	0.24	0.22	0.16	0.10	0.17	0.22
0.28	0.38	0.41	0.00	0.09	0.05	0.08	0.10	0.11	0.12	0.19	0.19	0.17	0.14	0.09	0.14	0.14
0.25	0.25	0.33	0.28	0.00	0.08	0.13	0.09	0.15	0.13	0.21	0.19	0.20	0.14	0.11	0.15	0.15
0.16	0.23	0.25	0.18	0.20	0.00	0.06	0.06	0.07	0.08	0.15	0.15	0.12	0.10	0.07	0.13	0.14
0.38	0.46	0.41	0.29	0.42	0.22	0.00	0.11	0.11	0.12	0.19	0.21	0.13	0.13	0.11	0.15	0.17
0.13	0.38	0.33	0.36	0.26	0.20	0.46	0.00	0.10	0.10	0.15	0.15	0.14	0.11	0.12	0.15	0.11
0.37	0.37	0.52	0.49	0.60	0.29	0.51	0.49	0.00	0.14	0.14	0.15	0.12	0.10	0.13	0.19	0.17
0.18	0.35	0.35	0.44	0.36	0.28	0.44	0.31	0.60	0.00	0.21	0.22	0.20	0.15	0.12	0.13	0.18
0.84	0.80	1.12	1.04	0.88	0.74	1.10	0.80	0.81	1.16	0.00	0.04	0.20	0.14	0.21	0.28	0.16
0.86	0.86	1.06	1.06	0.77	0.74	1.22	0.71	0.90	1.32	0.12	0.00	0.18	0.12	0.22	0.31	0.18
0.62	0.66	0.79	0.80	0.76	0.50	0.49	0.59	0.60	0.91	1.10	0.84	0.00	0.12	0.17	0.23	0.22
0.48	0.64	0.56	0.72	0.53	0.50	0.65	0.52	0.63	0.73	0.80	0.65	0.56	0.00	0.12	0.17	0.13
0.37	0.46	0.25	0.35	0.33	0.30	0.42	0.48	0.63	0.44	1.15	1.33	0.78	0.57	0.00	0.09	0.13
0.38	0.50	0.45	0.43	0.40	0.40	0.48	0.45	0.74	0.35	1.41	1.87	0.91	0.61	0.23	0.00	0.17
0.48	0.77	0.76	0.61	0.50	0.66	0.85	0.45	1.03	0.79	0.88	0.98	1.27	0.64	0.48	0.51	0.00

1 – MS60, 2 – Moura, 3 – Large White, 4 – Nilo, 5 Brazilian Duroc, 6 – Piau, 7 – Monteiro, 8 – Landrace, 9 – Mixed, 10 – Pietrain, 11 – US-Meishan, 12 – China-Meishan, 13 – Fengjing, 14 – Minzhu, 15 – Yorkshire, 16 – Berkshire, 17 – US Duroc.

Silva *et al*. (2011) reported higher mean number of alleles and effective number of alleles of 3.61 and 3.01 respectively for Brazilian Duroc, however with slightly lower observed and expected heterozygosity (0.57, 0.58). In the merged data set, Landrace had 1 private allele; Fengjing had 3; Ming had 5; Yorkshire had 2 and the US Duroc also had 2. Sollero *et al*. (2009) observed 8 private alleles in Landrace, 7 private alleles in Monteiro, 3 in Moura, 6 in MS60 and 13 in Piau. The mean effective number of alleles, observed and expected heterozygosity obtained from our study for the Brazilian Duroc was similar to estimates for the Portuguese Duroc obtained by Vicente *et al*. (2008), although, they obtained a much higher mean number of alleles (3.77). Vicente *et al*. (2008) reported a slightly higher observed heterozygosity (0.56) and a slightly expected heterozygosity (0.59) for the Pietrain. Rodrigannez *et al*. (2008) reported a total of 16 private alleles with a frequency of 1.11 when estimated by rarefaction (n = 64) for the Spanish Duroc.

Allelic richness or diversity is an important alternative measure of genetic diversity and has been suggested to provide some useful information to aid conservation decisions (Simianer, 2005; Foulley and Ollivier, 2006). It was important to eliminate the effect of samples size on allelic richness especially within the Brazilian swine data set because the observed number of alleles and the frequency distribution are highly dependent on the sample size and the marker system used (Kalinowski 2004; Bashalkhanov *et al*., 2009). Allelic richness observed for Nilo was 2.53, however, the sample size was small (n=5). Allelic richness for the mixed group was really not a surprise given its composition. Allelic richness when the data sets were merged for the Brazilian Duroc was 2.15 (n = 4) while for the US Duroc it was 2.25 (n = 26). On the basis of allelic richness, this suggests that the two Duroc breeds are similar. It would be expected that Duroc breeds from any part of the globe should be more closely related. Rodriganez *et al*. (2008)

observed a much higher allelic richness in Spanish wild boar (7.72, n = 68); in Spanish Duroc (6.97, n = 64); and in Iberian breeds (7.80, n = 107) using the rarefraction method. However, Swart *et al*. (2010) observed allelic richness in Durocs (3.65, n = 22) which was similar to the value obtained in this study. Allelic richness in a population may indicate a decrease in population size or past bottlenecks (Nei *et al*., 1975).

MS60 is a composite breed developed by crossbreeding Duroc, Large White and Pietrain (Sollero *et al*., 2009). Nei's genetic distance estimates between MS60 and Piau (0.16) and between MS60 and Landrace (0.13) showed that Piau and Landrace appeared to be genetically similar to MS60 than the three breeds from which MS60 was formed. One explanation for this is that the Old Duroc and subsequent improved Duroc varieties have been implicated in the formation of both MS60 and Piau. Another explanation for this may suggest a recent conservation direction towards increasing the genetic diversity of MS60 by crossing the breed with both commercial and local breeds. This is supported by the observed heterozygosity for MS60 (0.57) which is higher than the mean observed heterozygosity within the Brazilian breeds (0.56) and the overall mean observed heterozygosity (0.31) for both the US and the Brazilian data set. Sollero *et al*. (2009) reported a higher observed heterozygosity for MS60 (0.61).

The low genetic distance between the Meishan subpopulations suggest a close relationship between the two breeds. The significant p-value observed does not suggest that the two groups are genetically distinct breeds; rather it provides evidence about the conservation approach implemented by the United States since the importation of Meishan semen and embryos from China in 1989. One rather interesting observation from the result was the genetic distance between the Brazilian Duroc and the US Duroc (0.50). Although, as expected; this was not significantly different, however, the moderately high genetic distance between the Duroc

from the US and that from Brazil may indeed suggest changes due to the utilization of the breed in the two countries. It is also more likely that the differences observed could be as a result of small sample size of the Duroc breed from Brazil. Observed heterozygosity was 0.45 for the Brazilian Duroc and 0.53 for the US Duroc.

Figure 5 shows the mean allelic patterns across when the merged populations were pooled into three major regions. Breeds from the US had the smallest mean allelic patterns compared to the other two regions. The Brazilian breeds had a higher mean number of alleles but had a smaller mean effective number of alleles compared to the breeds from China. The Chinese breeds had higher observed and expected heterozygosity but smaller number of private alleles compared to the Brazilian breeds. Also, the Chinese breeds had the highest allelic richness across all three regions.

When the merged populations were pooled into two major regions (figure 6), the mean number of alleles, the mean effective number of alleles, the expected heterozygosity, the mean number of private alleles and the mean allelic richness for the US breeds were higher than those of the Brazilian breeds. However, the observed heterozygosity in the US breeds was lower than the observed heterozygosity in the Brazilian breeds.

Table 15 shows the genetic distance, fixation index and number of migrants when the merged populations were pooled into 2 and 3 regions. Among the 3 regions, the F_{ST} and genetic distance was highest between the Chinese and US breeds (0.19, $p < 0.001$; and 0.70, $p < 0.01667$ respectively) but with the smallest Nm. Between the 2 pooled regions, F_{ST} and genetic distance between the Brazilian breeds were 0.06 ($p < 0.001$) and 0.20 ($p < 0.05$) respectively.

The Chinese breeds were furthest away from the US/English breeds. The result was consistent with reports that Asian native pigs are closely related but different from the European

pigs with the Exception of Berkshire (Kim *et al*., 2002; Chen *et al*., 2011). However, our study showed that both Berkshire and Large White were also genetically distant from the Chinese breeds unlike the study reported by Kim *et al*. (2002) and Chen *et al*. (2011). It is important to point out that in our study; only three Large White pigs were sampled. Sample size was also very small in the study conducted by Kim *et al*. (2002). Several authors have reported that studies of Chinese porcine diversity in most cases were either limited to one or a small number of indigenous breeds or based on either restriction fragment length polymorphism analysis of mtDNA (Lan and Shi, 1993; Huang and Qui, 1998) or focused on a relatively small region of the mtDNA control region (Kim *et al*., 2002; Okumura *et al*., 2001). Yu *et al*. (2013) reported that Large White did not cluster together with the Chinese breeds. Yu *et al*. (2013) reported that while the Berkshire was clustered together with the Chinese breeds and had been implicated in its formation (Mason, 2004); it was still a European breed.

Another interesting observation was the smaller genetic distance between the US/European and the Brazilian breeds compared to the genetic distance between the Chinese breeds and the US/European breeds. This is not surprising given that several authors have reported that the Brazilian local breeds originated from the activity of Spanish and Portuguese settlers in the 16^{th} century (Souza *et al*., 2009; McManus *et al*., 2010). Furthermore, several US/European breeds contributed in the formation of several Brazilian breeds such as MS60 and Moura (Mason, 2004; Sollero *et al*., 2009).

The observed heterozygosity for the Chinese breeds shows that the breeds have a vast amount of diversity. This was consistent with the findings reported by Yang *et al*. (2003). The Brazilian breeds appeared to also have a higher genetic diversity compared to the US breeds. The smaller F_{ST} and genetic distance among the Brazilian breeds compared to breeds within the US is

an indication of the high level of gene flow among different herds from the same geographical region (Santos-Silva *et al*., 2009; Silva *et al*., 2011).

Figure 7 shows breed relationships for the merged data based on two principal coordinates. The first two principal coordinates account for 52.09% of the genetic variation in the merged data set. Fengjing and the US Duroc were placed on extreme ends. Yorkshire and Berkshire were grouped together. MS60, Moura, Large White, Nilo, the Brazilian Duroc, Piau, Landrace and Pietrain were all grouped together in a cloud of points. Minzhu was grouped together with the Brazilian pigs. Fengjing was grouped together with the Mixed group. This is not surprising since both Caruncho and Pirapetinga were developed by crossing Piau with a Chinese breed. Obviously, there is a connection between Brazilian and Chinese breeds. The US and Chinese Meishan were grouped together while the Brazilian and US Duroc subpopulations were kept away from each other.

Figure 8 shows the breed relationships for three regions: Brazil, China and the US. The first principal coordinate accounts for 78.52% of the total genetic variation between the breeds from the three regions. Breeds from Brazil appear to lie somewhere at the line joining the US breeds and the Chinese breeds.

The principal coordinate analysis was consistent with the genetic distances reported in this study; however, the principal coordinate plot (figure 8) provides more evidence that the Brazilian and the US Durocs are not genetically distinct breeds. The two Duroc breeds appeared to be grouped together. A better description of the breeds would be improved Duroc varieties or subpopulations of the Duroc breed. This description would explain the impacts of artificial selection and geographical isolation which may result in breed differentiation/speciation. It is important to point out that the small sample size of the Brazilian Duroc may have affected the

genetic distance. The F_{ST} value for the two Duroc varieties and the principal coordinate analysis suggest that both of them are not well differentiated and still account for a considerable amount of the original genetic diversity (4.9% for the US Duroc and 4.2% for the Brazilian Duroc).

The Principal component analysis for the three regions suggests that the Brazilian breeds were also influenced by breeds from both China and US/England. This is consistent with reports from several authors (Mason, 2004; Sollero *et al*., 2009).

Figure 9 shows delta k values for the merged data set. Delta k peaks were observed at clusters 2, 4, 13 and 15 with the highest delta at cluster 2. Figures 11 shows the genetic structure of the merged Brazilian and US breeds when the Bayesian cluster equaled 4 (figure 10a) and 9 (figure 10b). When k equaled 4 (figure 10a), the breeds were separated into respective countries. All the Brazilian breeds were grouped into one cluster. The US and English breeds were grouped into one cluster. However, the pig breeds from China were differentiated into two clusters with the Meishan breeds in one cluster and Minzhu and Fengjing in another cluster. When k equaled 9 (figures 10b), the population structure within each country was differentiated further into their respective breeds.

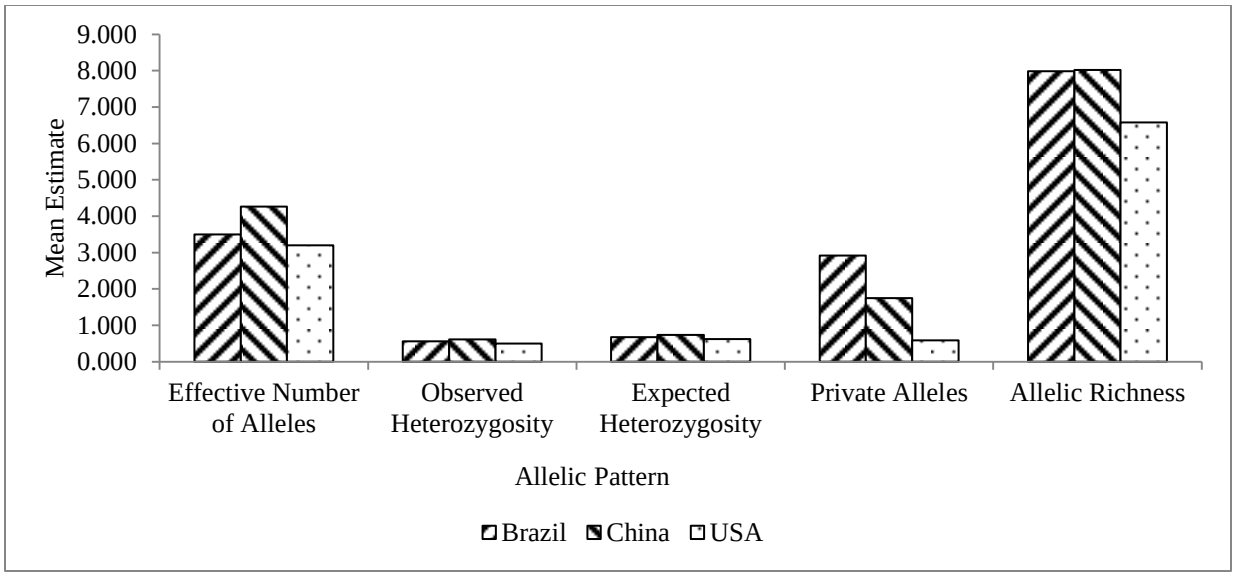


Figure 5. Effective Number of Alleles, Observed and Expected Heterozygosity, Number of Private Alleles and Allelic Richness Using 12 Microsatellite Markers Genotyped from Pigs in Brazil, China and the United States

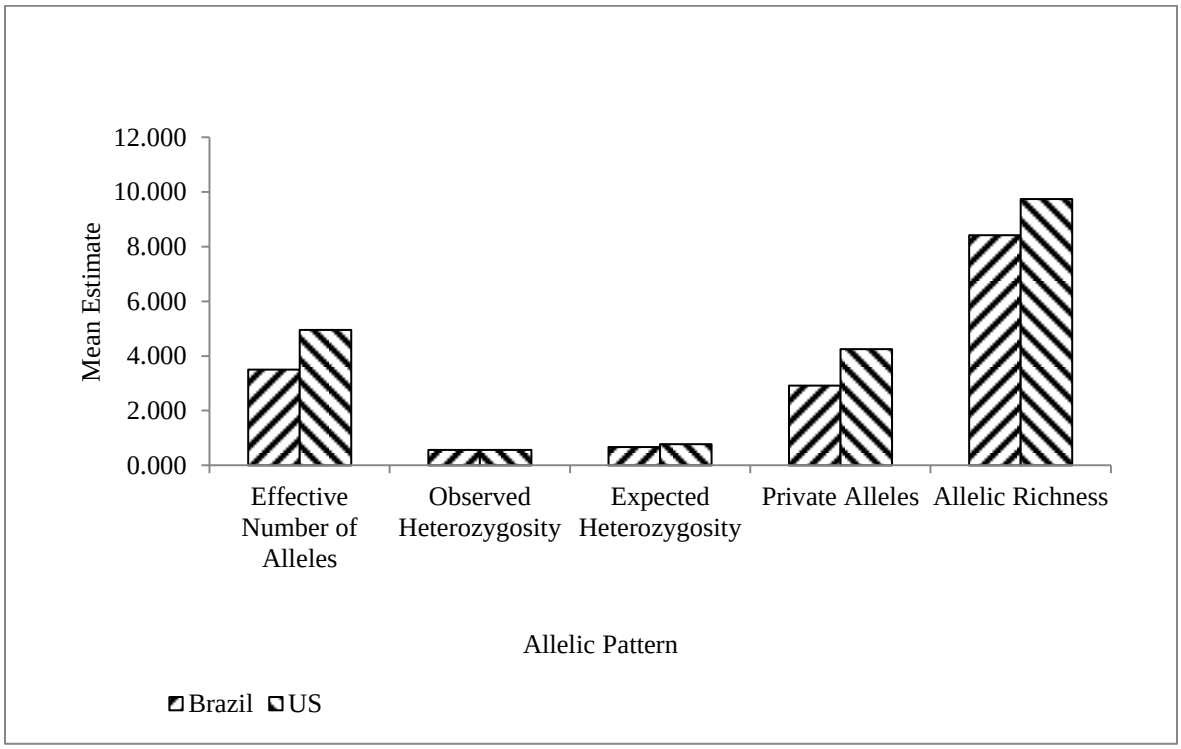


Figure 6. Effective Number of Alleles, Observed and Expected Heterozygosity, Number of Private Alleles and Allelic Richness Using 12 Microsatellite Markers Genotyped from Pigs in Brazil and the United States.

Table 15. Nei's Genetic Distance, Fixation Index (F_{ST}) and Number of Migrants (Nm) Using Microsatellite Markers from Pigs Pooled into 3 Regions: Brazil (BR), China (CH) and the United States; and two regions: China and the United States

	Pooled - 3 Regions			Pooled - 2 Regions
	BR-CH	BR-US	CH-US	BR-US
Nei D	0.42	0.25	0.70	0.20
F_{ST}	0.11	0.10	0.19	0.06
Nm	1.97	2.35	1.10	4.01

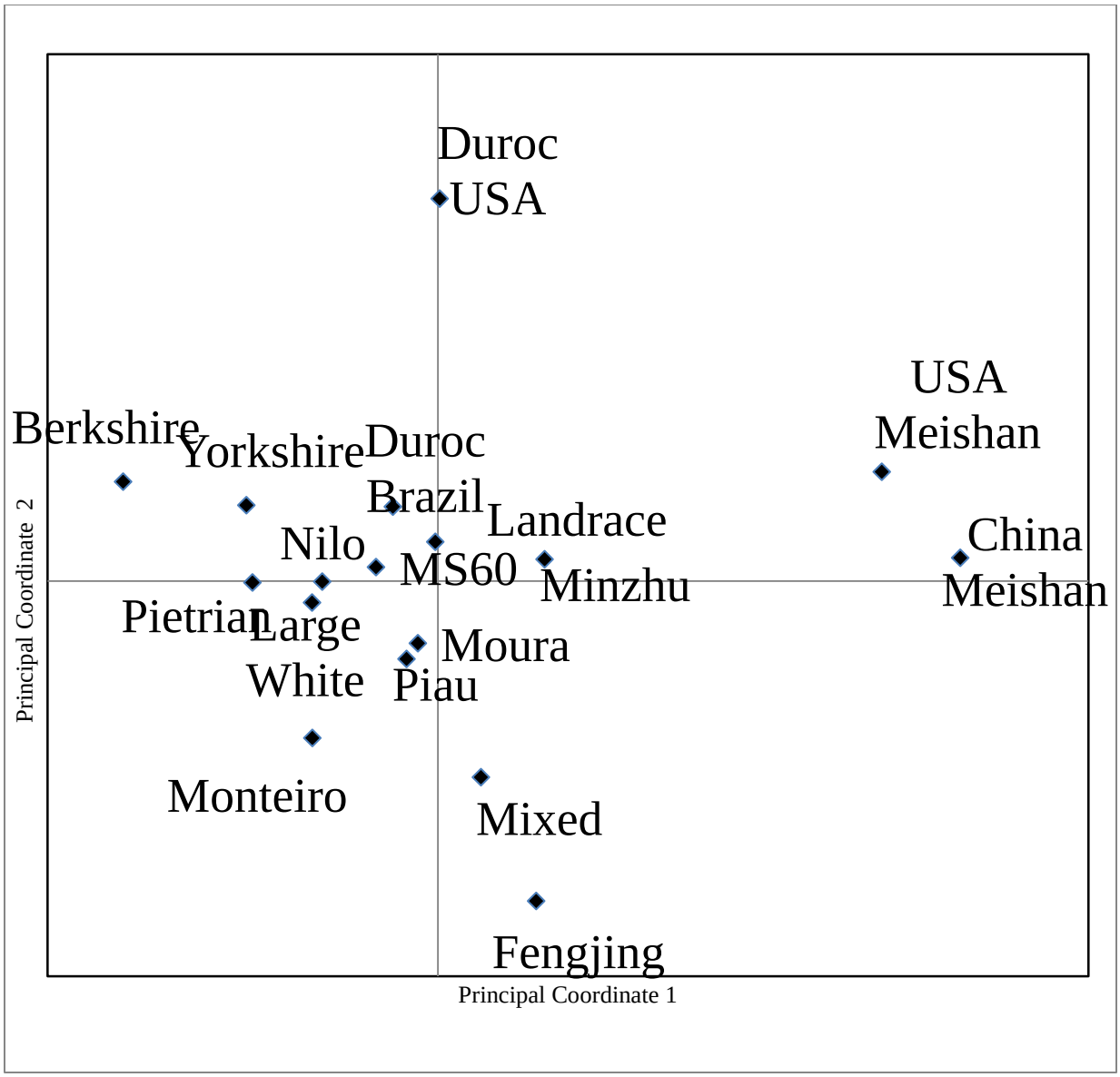


Figure 7. Breed Relationship between Pigs from the United States and Brazil Based on Two Principal Coordinates

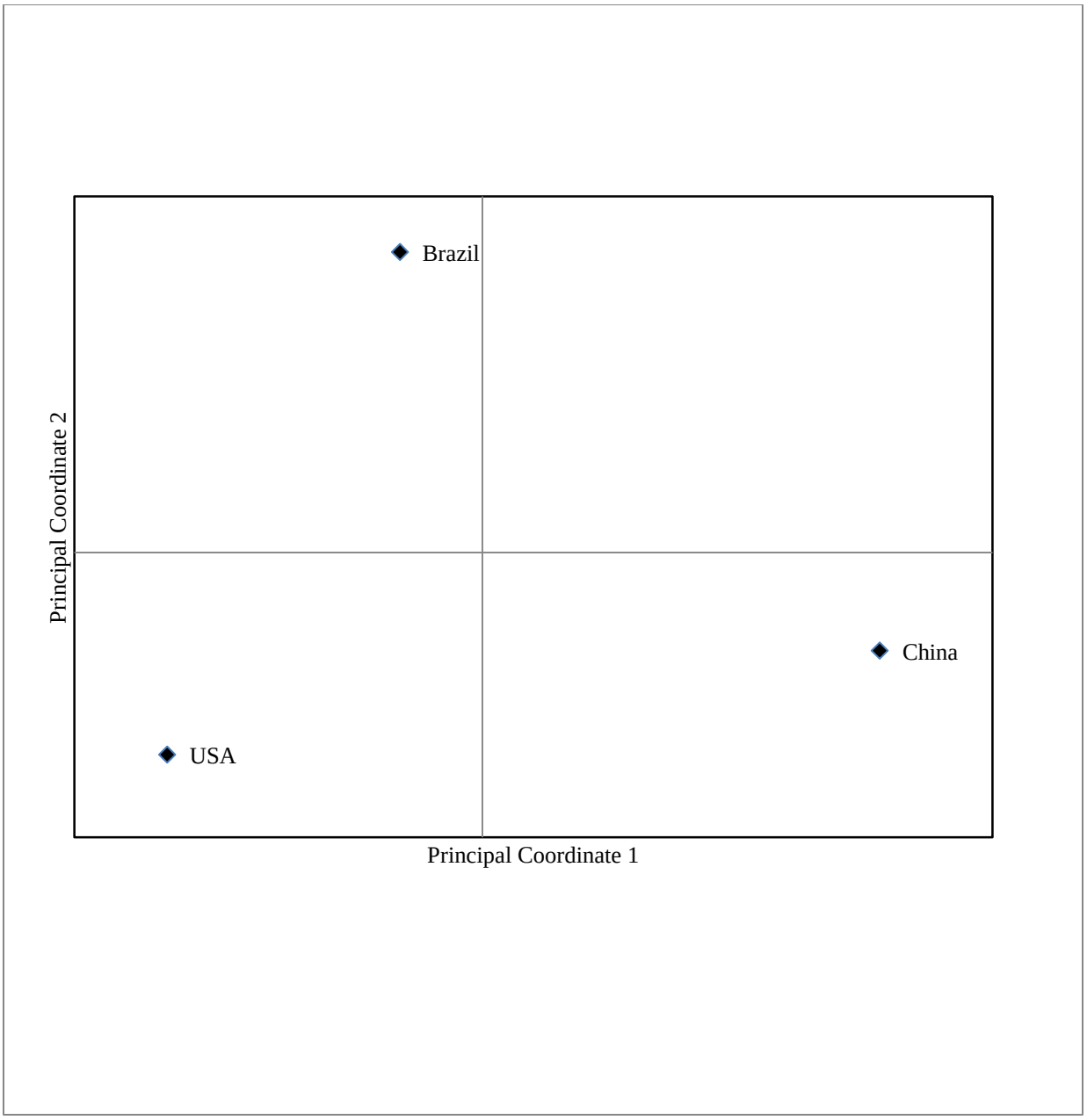


Figure 8. Regional Pig Relationship between Brazil, China and United States Based on Two Principal Coordinates

The population structure showed a lot of genetic admixtures which may indicate crossbreeding. MS60 was placed in the same cluster with Large White, Brazilian Duroc and Pietrain. The Brazilian Duroc showed similarities with Piau and Monteiro. The similarity between Duroc and Monteiro was consistent with the study conducted by Grossi *et al*. (2006). Mason (2002) reported that Duroc was important in the formation of Piau. There was no clear similarity between the US Duroc and the Brazilian Duroc. This was most likely as a result of the small sample size of the Brazilian Duroc (n=4). The two Meishan subpopulations were grouped together; Fengjing and Ming were also grouped together so were Yorkshire and Berkshire. Yorkshire and Berkshire are both English breeds; Yorkshire originated in York, England while Berkshire originated from Reading, England (Mason, 2004). Due to the small sample sizes of Casco de Burro, Caruncho and Pirapetinga, the data for these breeds were combined together and as expected, this group showed a lot of admixtures. The mixed group appeared to share some genetic similarity with Piau.

Figure 11 shows the genetic structure of the pooled populations. The US breeds showed genetic similarity with the Brazilian breeds when the breeds were pooled into two regions (figure 11a). The breeds from the US that appeared to share similar genetic material with Brazilian breeds were the Duroc, the English breeds (Yorkshire and Berkshire), the Chinese breeds (Fengjing and Ming). When the breeds were pooled into three regions (figure 11b), the Brazilian breed population appeared to be made up of a mixture of breeds from China and the United States. This was not surprising given that swine breeds from Brazil and US were developed using breeds from China and Europe.

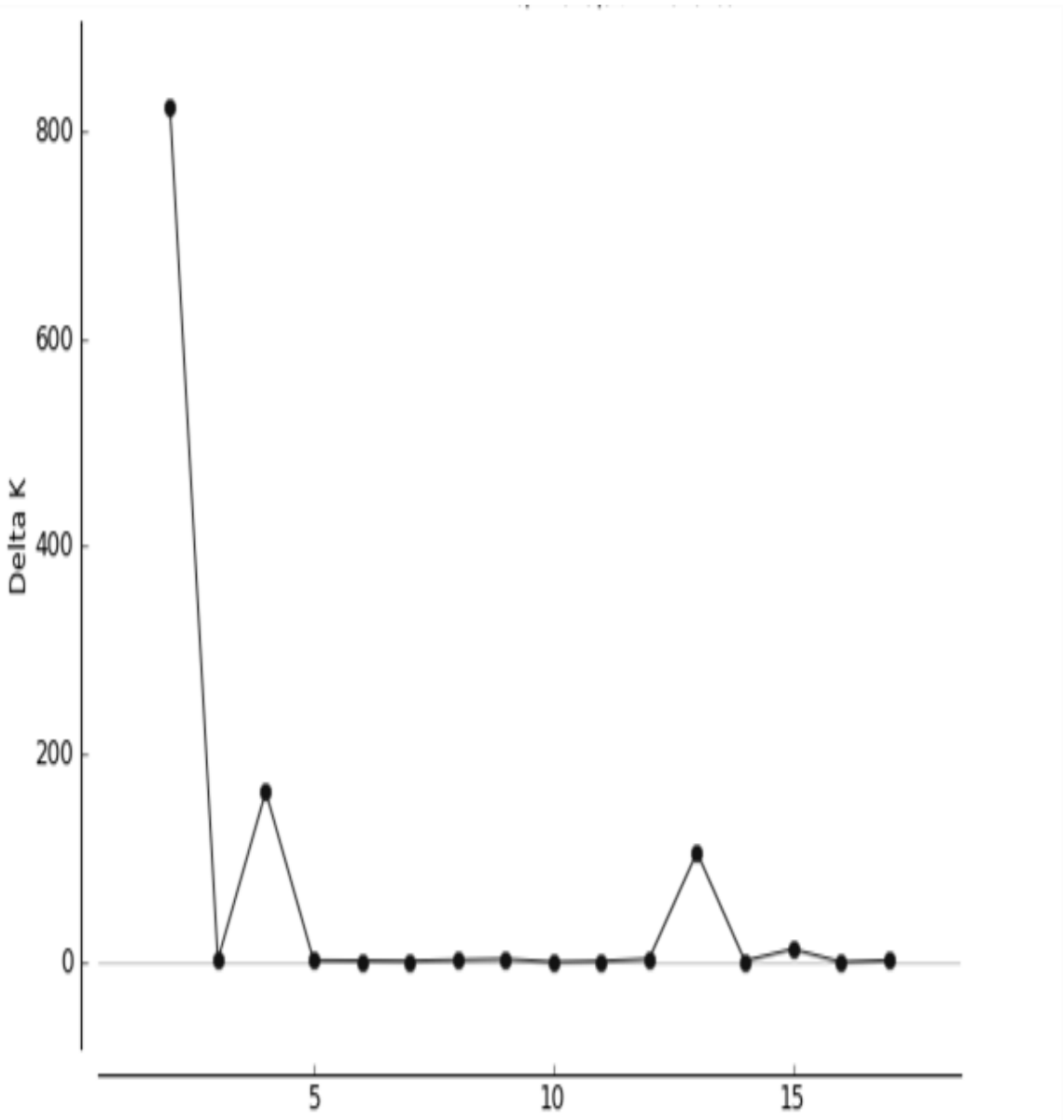
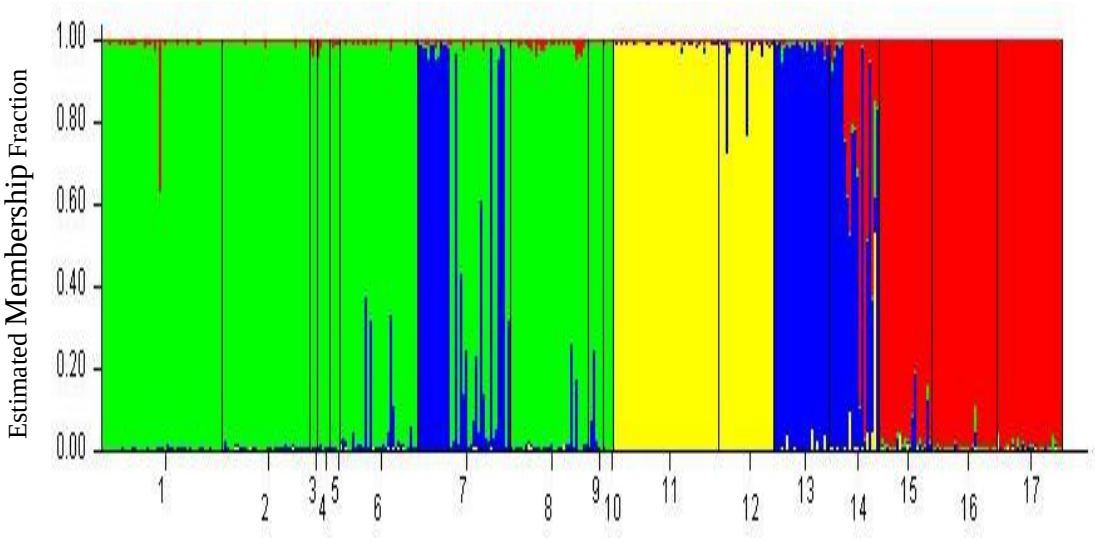
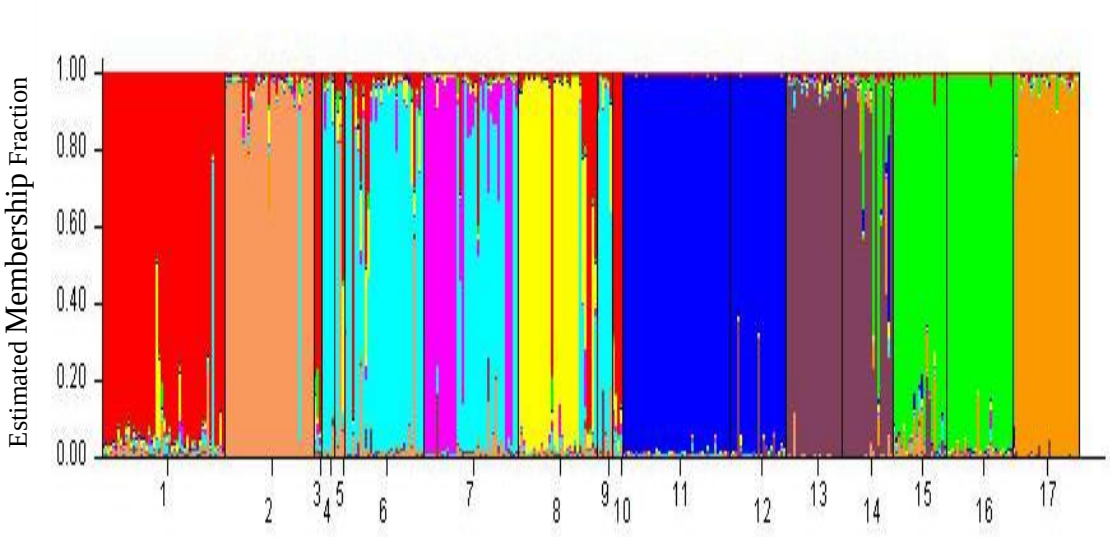


Figure 9. Bayesian Cluster and the Corresponding Delta K Using 12 Merged Microsatellite Markers from Pigs in Common between Brazil and the United States



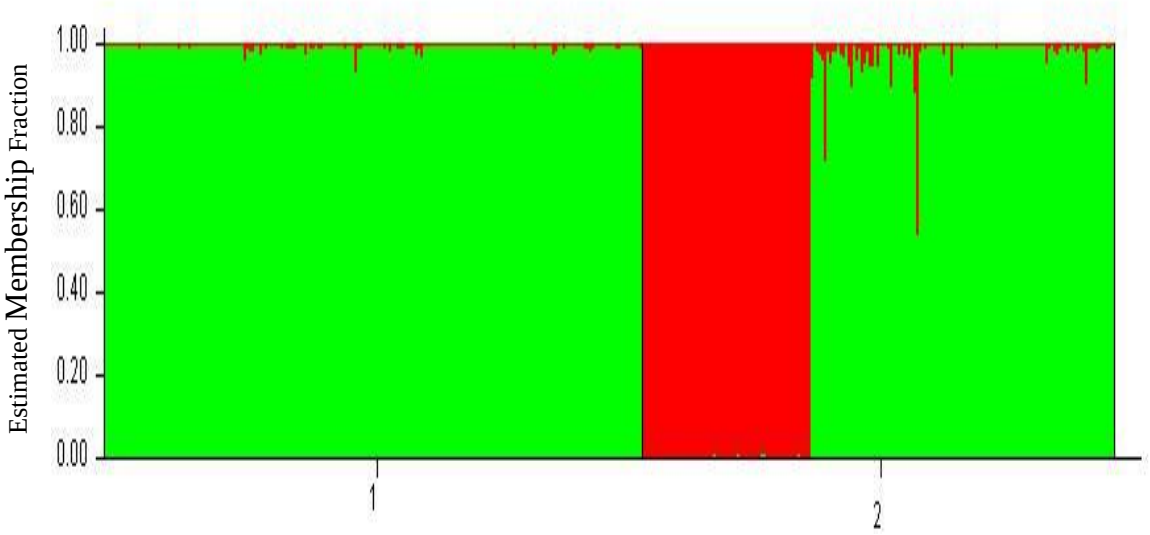
10a



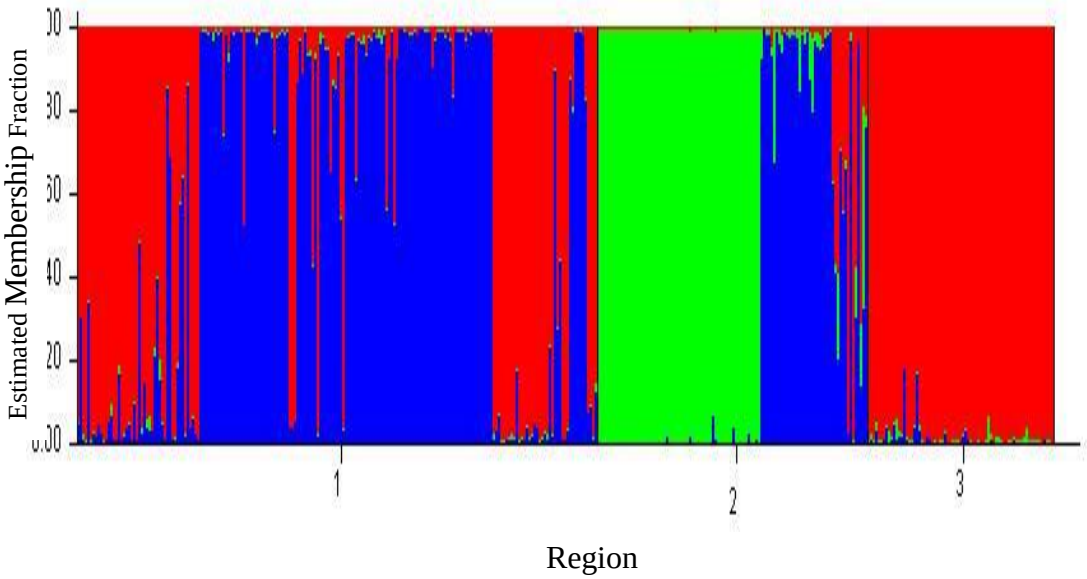
10b.

1 – MS60, 2 – Moura, 3 – Large White, 4 – Nilo, 5 Brazilian Duroc, 6 – Piau, 7 – Monteiro, 8 – Landrace, 9 – Mixed, 10 – Pietrain, 11 – US-Meishan, 12 – China-Meishan, 13 – Fengjing, 14 – Minzhu, 15 – Yorkshire, 16 – Berkshire, 17 – US Duroc

Figure 10. Estimated Membership of Individual Pigs from Brazil and the United States when Bayesian Cluster Number equaled 4 (10a) and 9 (10b)



11a.



11b.

Figure 11. Estimated Membership of Pigs Pooled into the Brazilian (1) and the United States (2) Regions (11a); and Pigs Pooled into the Brazilian (1), the Chinese (2) and the United States (3) Regions (11b)

CHAPTER FIVE

5.0 CONCLUSION

The study provides a detailed approach to extend genetic conservation study to a regional or global level. Results from previously used merging methods have not only been challenging to reproduce, but also too difficult to compare with studies conducted in other regions. There are obvious advantages to merging microsatellite data. Merging increases power and accuracy to detect association prior to analysis. The use of MICROMERGE provides a relatively faster and easier method to implement since it is based only on allele frequency rather than allele size. Once merging has been successfully implemented, comparing genetic diversity/breeds between different countries can easily be done. This increases opportunities for countries to engage in collaborative efforts at low additional cost.

Like all computations, small errors are inevitable. However, attempts have been made to carefully reduce any error as a result of computation during the merging process. The top two alignments for each merging process had posterior probabilities less than 100%. Two markers were also included in the final merged data with posterior probability less than the minimum recommended criterion. These computational adjustments reduced the accuracy of the merging process.

The result showed that although the Brazilian swine breeds have different names, some of the breeds are actually admixtures, derived from crossing local Brazilian breeds with Chinese and/ or US-English breeds. There is an indication of differentiation between the Brazilian Duroc and the US Duroc which suggests the result of introgression. However, the differentiation did not result in major differences between the two subpopulations.

The Chinese breeds had the highest genetic diversity compared to both the United States-English breeds and the Brazilian breeds. The amount of genetic diversity observed among the Brazilian breeds may be an indication of the efforts towards increasing local breed diversity; however, there appears to be evidence of crossbreeding with commercial/exotic breeds. The lower genetic diversity among the US-English breeds compared to the Brazilian and Chinese breeds as shown by the mean within population heterozygosity and the allelic richness, suggest that conservation efforts may have decreased over time and action to conserve the animal genetic resources should be undertaken.

The result shows that the pigs all share some level of genetic similarity irrespective of the country they are currently found. The genetic similarities stem from pre-domestication events. Some breed differences may be due to the effect of within country breed utilization and introgression. However, due to the small sample size of some breeds like the Brazilian Duroc and the Large White, the magnitudes of the genetic distances that involve these breeds are not conclusive.

Small sample size could be the consequence of a larger problem of small effective population size for some breeds of pigs whether within a country or globally. Genetic drift and inbreeding are problem most commonly associated with small population size. Genetic drift and inbreeding could potentially increase the emergence of deleterious alleles within a population which further reduces the number of breeding males and females in that population. The effect of such bottlenecks can remain in a population for several generations even when the population increases. From this study, similar pig breeds could be crossed together to preserve pigs with small populations. More intense global germplasm collection strategy is another approach to identify the same breeds, collect and store germplasm to preserve the existence of these pigs.

The Chinese breeds may not be as similar as was previously reported by several authors. Evidence of genetic differentiation was observed within the Chinese region. The Meishan breeds were shown to be different from the other Chinese breeds. This was an interesting finding; however, it is beyond the scope of this study.

5.1 Future Study

This study was intended to provide a comprehensive approach to conduct cross country genetic diversity by merging microsatellite markers from pigs and provide opportunities for countries to compare their local animal genetic resources with other countries. There are obvious limitations observed in this study especially with the small sample size of some pigs. Refining and improving this method would be an obvious opportunity for future studies.

CITATIONS

Abdelkrim, J., B.C. Robertson., J.L. Staton and N.J. Gemmell. 2009. Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing. Bio Techniques 46:185-192.

Amills, M., A. Clop., O. Ramirez and, M. Perez-Enciso. 2010. Origin and Genetic Diversity of Pigs. In: Encyclopedia of Life Sciences (ELS). John Wiley & Sons, Ltd: Chichester. Doi: 10.1002/9780470015902.a0022884.

Barton, N.H., and M. Slatkin. 1986. A Quasi-Equilibrium Theory of the Distribution of Rare Alleles in a Subdivided Population. Heredity 56:409-415.

Bashalkhanov, S., M. Pandey, and P. Rajora. 2009. A simple Method for Estimating Genetic Diversity in Large Populations from Finite Sizes. BMC Genetics 10:84.

Bichard, M. 2002. Genetic Improvement in Dairy Cattle - An Outsider's Perspective. Livestock Production Science 75: 1-10.

Bird, C.E., S.A. Karl., P.E. Smouse, and R.J. Toonen. 2011. Detecting and Measuring Genetic Differentiation. CrustIssues19.10(27):31-39. Retrieved August 20, 2013 from www.epscor.hawaii.edu/system/files/er_book/2012/05/Birdetal_Detecting_measu ring.pdf

Blackburn, H.D., Y. Toishibekov., M. Toishibekov., C.S. Welsh., S.F. Spiller., M. Brown, and S.R. Paiva. 2006. Genetic Diversity of *Ovis aries* Populations Near Domestication Centers and in the New World. Genetica 139:1169-1178.

Castro, S.T.R., M.S.M. Albuquerque, and J.L. Germano. 2002. Census of Brazilian Naturalized Swine Breeds. Arch. Zootec. 51:235-239

Cavalli-Sforza, L .L, and W. F. Bodmer. 1971. The Genetics of Human Populations. W. H. Freeman and Company, San Francisco.

Chapuis, M.P., and A. Estoup. 2006. Microsatellite Null Alleles and Estimation of Population Diffrentiation. Mol. Biol. Evol. 24(3)621.

Chen, C.H., H.L. Huang., H.Y. Yang., S.H. Lai., N.T. Yen., M.C. Wu, and M.C. Huang. 2011. Mitochondial Genome of Taiwan Pig (Sus scrofa). Afr. J. Biotechnol. 10:2556-2561.

Cockerham, C.C . 1973. Analyses of gene frequencies. Genetics 74:679-700.

Crow, J. F, and M. Kimura. 1970. An introduction to population genetics theory. In Harper and Row New York: Harper and Row.

de Alba, J. 1987. Animal Genetic Resources. Strategies for Improved use and conservation. FAO Animal Production and health 66:336. URL: http://www.fao.org/docrep /010/ah806e/ah806e00.htm.

de Vicente, M.C., F.A. Guzman., J. Engels, and V. Ramanatha Rao. 2005. Genetic Characterization and its Use in Decision Making for the Conservation of Crop Germplasm. The Role of Biotechnology. Villa Gualino, Turin, Italy.

Development and characteristics of microsatellite markers for sugi (Cryptomeria japonica D. Don) derived from microsatellite-enriched libraries. Ann. For. Sci. 61 (2004) 569–575.

distruct.html

Duran, C., N. Appleby., D. Edwards, and J. Batley. 2009. Molecular Genetic Markers: Discovery, Applications, Data Storage and Vitualization. Current Bioinformatics 4:16-27.

Ellis, J.S ., J. Gilbey., A. Armstrong., T. Balstad., E. Cauwelier ., C. Cherbonnel., S. Consuegra., J. Coughlan., T. F. Cross ., W. Crozier., E. Dillane., D. Ensing., C. Garcı´a de Lea´niz., E. Garcı´a-Va´zquez., A. M. Griffiths., K. Hindar., S. Hjorleifsdottir., D. Knox., G. Machado-Schiaffino., P. McGinnity., D. Meldrup., E. E. Nielsen., K. Olafsson., C. R. Primmer ., P. Prodohl., L. Stradmeyer., J.-P. Va¨ha¨., E. Verspoor., V. Wennevik, and J. R. Stevens. 2011. Microsatellite standardization and evaluation of genotyping error in a large multi-partner research programme for conservation of Atlantic salmon (Salmo salar L.).

ERFP. 2003. Guidelines for the constitution of national cryopreservation programmes for farm animals. *In* S.J. Hiemstra, ed. *European Regional Focal Point on Animal Genetic Resources,* Publication.No.1.

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the Number of Clusters of Individuals Using the Software STRUCTURE: a Simulation Study. Mol. Ecol., 14:2611-2620.

Excoffier, L., and H.E.L. Lischer. 2010. Arlequin Suite ver 3.5: a New Series of Programs to Perform Population Genetics Analyses under Linux and Windows. Molecular Ecology Resources 10: 564-567.

Excoffier, L., P.E. Smouse, and J.M. Quattro. 1992. Analysis of Molecular Variance Inferred from Metric Distances among DNA Haplotypes: Application to Human Mitochondrial DNA Restriction Data. Genetics 131:479-491.

Falush, D., M. Stephens, and J.K. Pritchard.2003. Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies. Genetics Society of America. 164: 1567-1587.

Falush, D., M. Stephens, and J.K. Pritchard.2007. Inference of Population Structure Using Multilocus Genotype Data: Domimant Markers and Null Alleles. Molecular Ecology notes.doi:10.1111/j.1471-8286.2007.01758.x.

FAO, 2009. Learning from the past: Successes and failures with agricultural biotechnologies in developing countries over the last 20 years. Background Document to Conference 16 of the FAO Biotechnology Forum (8 June to 5 July 2009). Accessed on March 30, 2014 from http://www.fao.org/biotech/C16doc.htm

FAO. 2006. The Role of Biotechnology in Exploring and Protecting Agricultural Genetic Resources. pg 190. Retrieved on May 15, 2014 from ftp://ftp.fao.org/docrep /fao/009/a0399e/a0399e00.pdf.

FAO. 2007: Rischkowsky B, Pilling D (eds). The State of the World's Animal Genetic Resources for Food and Agriculture. FAO, Rome. Multiple chapters.

FAO. 2011. Molecular characterization of animal genetic resources. FAO Animal Production and Health Health Guidelines. No.9.Rome. Retrieved August 17, 2013 from www.fao.org/docrep/014/i2413e/i2413e00.pdf.

Ferguson, M.E., P.J Bramel and S. Chandra. 2004. Gene diversity among botanical varieties in peanut (*Arachis hypogaea L.*). Crop Sci., Sci 44:1847-1854.

Figliuolo, G. and P. Perrino. 2004. Genetic diversity and intra-specific phylogeny of Triticum turgidum L. subsp. Dicoccon (Schrank) Thell. Revealed by RFLPs and SSRs. Genet. Resour. Crop Ev., 52:519-527.

Flicek, P., M. R. Amode., D. Barrell., K. Beal., K. Billis., S. Brent., D. Carvalho-Silva., P. Clapham., G. Coates., S. Fitzgerald., L. Gil., C. G. Girón., L. Gordon., T. Hourlier., S. Hunt., N. Johnson., T. Juettemann., A. K. Kähäri., S. Keenan., E. Kulesha., F. J. Martin., T. Maurel., W. M. McLaren., D. N. Murphy., R. Nag., B. Overduin., M. Pignatelli., B. Pritchard., E. Pritchard., H. S. Riat., M. Ruffier., D. Sheppard., K. Taylor., A. Thormann., S. J. Trevanion., A. Vullo., S. P. Wilder., M. Wilson., A. Zadissa., B. L. Aken., E. Birney., F. Cunningham., J. Harrow., J. Herrero., T.J.P. Hubbard., R. Kinsella., M. Muffato., A. Parker., G. Spudich., A. Yates., D. R. Zerbino., and S.M.J. Searle. Ensembl. 2014. Nucleic Acids Research 2014 42 Database issue: D749-D755. doi: 10.1093/nar/gkt1196

Foulley, J.L., and L. Ollivier. 2006. Estimating Allelic Richness and its Diversity. Liv Sci. 101:150-158.

Fredua-Agyeman, R., D. Adamski., R.J. Liao, and D. Borthakur. 2008. Development and characterization of microsatellite markers for analysis of population differentiation in the tree legume Acacia koa (Fabaceae : Mimosoideae) in the Hawaiian Islands. Genome 51:1001-1015.

Gama, L. T., A. M. Martinez., I. Carolino., V. Landi., J. V. Delgado., A. A. Vicente., J. L. Vega-Pla., O. Cortés., C. O. Sousa, and BIOPIG Consortium. 2013. Genetic Structure, Relationships and Admixture with Wild Relatives in Native Pig Breeds from Iberia and its Islands. *Genetics Selection Evolution* 45:18.

Glaubitz, J.C. 2005. CONVERT: A User-Friendly Program to Reformat Diploid Genotypic Data for Commonly Used Population Genetic Software Packages. *Molecular Ecology Notes* 4: 309–310.

Goudet, J. 2002. FSTAT. A Program to Estimate and Test Gene Diversities and Fixation Indices (Version 2.9.3.2). Retrieved on October 31, 2013 at www.unil.ch/popgen/softwares/fstat.htm.

Green, P.J. 1995. Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination. Biometrika 82: 711-732.

Guichoux, E., L. Lagache., S. Wagner., P. Chaumeil., P. Le´ Ger.., O. Lepais., C. Lepoittevin., T. Malausa., E. Revardel., F. Salin, And R. J . Petit. 2011. Current Trends in Microsatellite Genotyping. Molecular Ecology Resources. 11, 591–611.

Hastings, W.K. 1970. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. Biometrika 57:97-109.

Hedrick, P.W. 1999. Highly variable loci and their interpretation in evolution and conservation. Evolution 53: 313–318.

Hedrick, P.W. 2005. A standardized genetic differentiation measure. Evolution 59: 1633– 1638.

Heller, R. and, H.R. Siegismund. 2009. Relationship between three measures of genetic differentiation GST , Dest , and G0ST : how wrong have we been? Mol. Ecol. 18:

2080– 2083.

Hoffmann, I., P.A. Marsan., J.S.F. Barker., E.G. Cothran., O. Hanotte., J.A. Lenstra., D. Milan., S. Weigend, and H. Simianer. 2004. New MoDAD marker sets to be used in diversity studies for the major farm animals species: recommendations of a joint ISAG/FAO working group. Proceedings of the 29th International Conference on Animal Genetics, 2004 Sep 11-16; Tokyo, Japan 123 [Abstract].

Huang, Y.F., and X.P. Qui. 1998. The Relationship between Mitochondrial DNA Polymorphism and Origin and Differentiation of Native Pig Breeds in China. Yi Chuan Xue Bao 25:322-329.

Huff, D.R., R. Peakall, and P.E. Smouse. 1993. RAPD Variation within and among Natural Populations of Outcrossing Buffalograss *Buchloe dactyloides* (Nutt) Engelm. Theoretical and Applied Genetics 86:927-934.

Jianlin, H. 2011. Meta-analysis of genetic diversity in goats with microsatellites. International Livestock Research Instiute (ILIR) Nairobi, Kenya and CAAS-ILRI Joint Laboratory on Livestock and Forage Genetic Resources, Chinese Academy of Agricultural Sciences (CAAS), Beijing, China). Globaldiv Final Workshop EPFL Campus, Lausanne, Switzerland.

Jones, G.F. 1998. Genetic Aspects of Domestication, Common Breeds and Origins. In: Rothschild, M.F, and A. Ruvinsky. (eds). The Genetics of Pigs. Wallingford: CABI Publishing.

Jost, L. 2008. G_{ST} and its relatives do not measure differentiation. Mol. Ecol. 17: 4015–4026.

Juliano, R.S., M.C da Silva and E. Bianchini . n.d. Pantaneiro cattle. Retrieved August 17, 2013 from http://www.buiatrics.com/fileok/file/Pantaneiro_Buiatria_TRAD_ok-M.pdf

Kalinowski S.T, and R.S. Waples. 2002. Relationship of Effective to Census Size in Fluctuating Populations. Conservation Biology. 16: 129-136.

Kalinowski, S.T. 2004. Counting Alleles with Rarefaction: Private Alleles and Hierarchical Sampling Designs. Conservation Genetics 5:539-543.

Kalinowski, S.T. 2005. HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. Mol. Ecol. Notes. 5:187–189.

Kelkar YD, Strubczewski N, Hile SE et al. (2010) What is a microsatellite: a computational and experimental definition based upon repeat mutational behavior at A/T and GT/AC repeats. Genome Biology and Evolution, **2**, 620–635.

Kim, K.I., J.H. Lee., K. Li., Y.P. Zhng., S.S. Lee., J. Gongora, and C. Moran. 2002. Phylogenetic Relationships of Asian and European Pig Breeds Determined by Mitochondrial DNA D-loop Sequence Polymorphism. Anim. Genet. 33:19-25.

Labate, J.A. 2000. Software for population genetic analyses of molecular marker data. Crop Sci. 40:1521-1528.

Lan, H., and L.M. Shi. 1993. The Origin and Genetic Differentiation of Native Breeds of Pigs in Southwest China: an Approach from Mitochondrial DNA Polymorphism. Biochem. Genet. 31:51-60.

Larson, G., K. Dobney., U. Albarella., M. Fang., E. Matisoo-Smith., J. Robins., S. Lowden., H. Finlayson., T. Brand., E. Willerslev., P. Rowley-Conwy., L. Andersson and A. Cooper. 2005. Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. Science 307(5715): 1618-1621.

Maestri, E., A. Malcevschi., A. Massari and N Marmiroli. 2002. Genomic analysis of cultivated barley (Hordeum vulgare) using sequence-tagged molecular markers. Estimates of divergence based on RFLP and PCR maekers derived from stress-responsive genes, and simple-sequence repeats (SSRs). *Mol. Gen. Genomics* 267(2):186-201.

Mariante A. da S. Mariante., M. do S.M. Albuquerque., A.A. do Egito, and C. McManus. 1999. Advances in the Brazilian animal genetic resources conservation programme. AGRI:109-123.

Mason, I.L. 2002. Mason's World Dictionary of Livestock Breeds, Types and Varieties. 5th rev. ed. CAB INTERNATIONAL. Wallingford Oxon OX10 8DE, UK.

McManus. C., S.R. Paiva., A.V.R. Silva., L.S. Murata., H. Louvandini., G.P.B. Cubillos., G. Castro., R. A. Martinez., M.S.L. Dellacasa and J.E.Perez.2010. Phenotypic Characterization of Naturalized Swine Breeds in Brazil, Uruguay and Columbia. Braz.arch. boil. Technol. 53:3

McTavish, E.J., J.E. Decker., R.D. Schabel., F.T. Jeremy, and D.M Hillis. 2013. New World cattle show ancestry from multiple independent domestication events. Proceeding of the National Academy of Sciences of the United States of America.

Meirmans, P.G and, P.W Hedrick. 2011. Assessing population structure: F_{ST} and related measures. Mol. Ecol. Res. 11: 5–18.

Metropolis, N., A.W. Rosenbluth., M.N. Rosenbluth., A.H. Teller, and E. Teller. 1953. Equations of State Calculations by Fast Computing Machines. J. Chem. Phys. 21:1087-1092.

Nagarajan, M., N. Kumar., G. Nishanth., R. Haribaskar., K. Paranthaman., J. Gupta., M. Mishra., R. Vaidhegi., S. Kumar., A. K. Ranjan, and S. Kumar. 2009. Microsatellite Markers of Water Buffalo, *Bubalus bubalis* - Development, Characterisation and Linkage Disequilibrium Studies. BMC Genetics. 10:68.

Nei, M. 1973. The theory and estimation of genetic distance, pp. 45-51. In: Genetic Structure of Populations, Edited by N. E. MORTON. University Press of Hawaii, Honolulu.

Nei, M. 1987. Molecular Evolutionary Genetics. New York: Columbia university Press.

Neigel, J.E. 2002. Is FST obsolete? Conserv. Genetics 3: 167–173.

Okumura, N., Y. Kurosawa., E. Kobayashi., T. Watanobe., N. Ishiguro., H. Yasue, and T. Mitsuhashi. 2001. Genetic Relationship Amongst the Major Non-coding Regions of Mitochondrial DNAs in Wild Boars and Several Breeds of Domesticated Pigs. Anim. Genet. 32:139-147.

Orloci, L. 1978. Multivariate analysis in vegetation research. 2nd ed. Dr W. Junk, The Hague.

Orr, D.E, and Y.S. Shen. 2006. World Pig Production, Opportunity or Threat? Proceedings of the 2006 Midwest Swine Nutrition Conference, Indianapolis, Indiana, USA.

Paiva, S.R., A.S. Mariante, and H.D. Blackburn. 2011. Combining US and Brazilian Microsatellite Data for a Meta-Analysis of Sheep (Ovis aries) Breed Diversity: Facilitating the FAO Global Plan of Action for Conserving Animal Genetic Resources. J. Hered. 102(6):697-704.

Peakall, R., and P.E. Smouse. 2006. GenAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol. Ecol. Notes. 6:288-295.

Peakall, R., and P.E. Smouse. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update.Bioinformatics 28:2537-2539.

Perera, L., J.R. Russell., J. Provan and W. Powell. 2000. Use of microsatellite DNA markers to investigate the level of genetic diversity and population genetic structure of coconut (*Cocos nucifera L.*). Genome 43: 15-21.

Presson, A.P. 2007. MicroMerge Input File Specifications. Retrieved on August 10, 2013 at https://www.genetics.ucla.edu/software/MicroMerge_Input-File_Specifications _2007.pdf

Presson, A.P., E. Sobel., K. Lange, and J.C Papp. 2006. Merging Microsatellite Data. Journal of Computational Biology.13 (6):1131-1147.

Presson, A.P., E.M. Sobel., P. Pajukanta., C. Plaiser., D.E. Weeks., K. Aberg, and J.C Papp. 2008. Merging microsatellite data: enhanced methodology and software to combine genotype data for linkage and association analysis. BMC Bioinformatics.9:317

Pritchard, J.K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics Society of Anerica.155:945-959.

Pritchard, J.P., X. Wen, and D. Falush. 2010. Documentation for structure software: Version 2.3. Retrieved on August 20, 2013 at http://pritch.bsd.uchicago.edu /structure.html.

Ramı´rez, O., A. Ojeda., A. Toma`s., D. Gallardo., L.S. Huang.,J.M. Folch., A. Clop., A. Sa´nchez., B. Badaoui., O. Hanotte., O. Galman-Omitogun., S.M. Makuza., H. Soto., J. Cadillo., L. Kelly.,I.C. Cho.,S. Yeghoyan., M. Pe´rez-Enciso and M. Amills. 2009. Integrating Y- Chromosome, Mitochondrial, and Autosomal Data to Analyze the Origin of Pig Breeds. Molecular Biology and Evolution 26(9):2061-2072.

Reynolds, J., B.S Weir, and C.C Cockerham. 1983. Estimation of the Coancestry Coefficient: Basis for a Short term genetic Distance. Genetics Society of America.

Rodriganez, J., C. Barragan., E. Alves., C. Gortazar., M.A. Toro, and L. Silio. 2008. Genetic Diversity and Allelic Richness in Spanish Wild and Domestic Pig Population Estimated from Microsatellite Markers. Spanish J. Agricultural Research 6:107-115.

Rosenberg, N.A. 2007. DISTRUCT. A Program for the Graphical Display of Population Structure. Retrieved on January 15, 2014 at http://www.stanford.edu/group/rosenberglab/

Rosenberg, N.A., T. Burke., K. Elo., M.W. Feldman., P.J. Freidlin., M.A Groenen., J. Hillel., A. Maki-Tanila., M. Tixier-Boichard., A. Vignal., K. Wimmers, and S. Weigend. 2001.

Empirical Evaluation of Genetic Clustering Methods Using Multilocus Genotypes from 20 Chicken Breeds. Genetics 159:699-713.

Rousset, F. 2008. GENEPOP'007. A complete re-implementation of the genepop software for Windows and Linux. Mol. Ecol. Res. 8: 103-106.

Ryman, N., and O. Leimar. 2009. Comment: G_{ST} is still a useful measure of genetic differentiation – a comment on Jost's D. Molecular Ecology 18:2084-2087.

Santos-Silva, F., R. Ivo., M.C Sousa., A Vicente., I. Carolino., N. Carolino and L.T. Gama. 2009. Analise da estrutura genetica de populacoes ovinas churros portuguesas. Archvos de Zootecnia. 58:493-496.

Schlo¨tterer, C. 2004. The Evolution of Molecular Markers – Just a matter of Fashion? Nat. Rev. Genet. 5: 63-69. Doi:10.1038/nrg1249. PMID:14666112.

Shim, S.I. and R.B Jorgensen. 2000. Genetic structure in cultivated and wild carrots (Daucus carota L.) revealed by AFLP analysis. *Theor. Appl. Genet*., 101:227-233.

Silva, E. C., W.M.D. Junior., P. Ianella., M.A.G Filho., C.J.P. de Oliveira., D.N.M Ferreira., A.R. Caetano and S.R. Paiva. 2011. Patterns of Genetic Diversity of Local Pig Populations in the State of Pernambuco, Brazil. 2011. R. Bras. Zootec 8:1691-1699.

Simianer, H. 2005. Using Expected Allele Number as Objective Function to Design between and within Breed Conservation of Farm Animal Biodiversity. J. Anim. Breed. Genet. 122:177-187.

Slatkin, M. 1985. Rare Alleles as Indicators of Gene Flow. Evolution 39:53-63.

Sollero B.P., S.R. Paiva., D.A. Faria., S.E.F. Guimaraes., S.T.R. Castro., A.A. Egito., M.S.M. Albuquerque., U. Piovezan., G.R. Bertani and A. da S. Mariante. 2009. Genetic Diversity of Brazilian Pig Breeds Evidenced by Microsatellite Markers. Livestock Science 123:8-15.

Souza, C.A., S.R. Paiva., R.W. Pereira., S.E.F. Guimaraes., W.M. Dutra., L.S. Murata and A.S. Mariante. 2009. Iberian origin of Brazilian local pigs breeds based on Cytochrosome b (MT-CYB) sequence. Animal Genetics 40(5): 759-762.

Swart, H., A. Kotze., P.A.S. Olivier, and J.P. Grobler. 2010. Microsatellite-based Characterization of Sourthern African Domestic Pigs (Sus Scrofa domestica). South African Jornal of Animal Science.40:2.

Tandon, P., and P. Bhattacharjee. 2010. Bioinformatics in Biodiversity Management. ISBN: 81-88237-81-7.Retrieved on August 17,2013 from http://dspace.nehu.ac.in/handle/1/2373

Tani, N., T. Takahashi., T. Ujino-Ihara., H. Iwata., K. Yoshimura, and Y. Tsumura. 2004.

Taubert, H., and D.G. Bradley. 2008. COMBI.PL. A computer program to combine data with microsatellite marker allele size information. Mol Ecol Resour. 8:572-574.

Tautz D., Hypervariability of simple sequences as a general source for polymorphic DNA markers, Nucleic Acids Res. 17 (1989) 6463–6471.

Vicente, A.A., M.I, Carolino., M.C.O. Sousa., C. Ginja., F.S. Silva., A.M. Martinez., J.L. Vega-Pla., N.Carolino, and L.T Gama. 2008. Genetic Diversity in Native and Commercial Breeds of Pigs in Portugal Assessed by Microsatellites. J.Anim.Sci. 86:2496-2507.

Wang, J. 2005. Estimation of effective population sizes from data on genetic markers. Phil. Trans. R. Soc. B. 360 no. 1459 1395-1409.

Weeks, D.E., Y.P. Conley., R. E. Ferrell., T. S. Mah, and M.B. Gorin. 2002. A tale of two genotypes: Consistency between two high-throughput genotyping centers. Genome Res. 12, 430–435.

Wright, S. 1943. Isolation by distance. Genetics 28: 114-138.

Wright, S. 1946. Isolation by distance under diverse systems of matings. Genetics 31:39.

Wright, S. 1951. The genetical structure of populations. Annals of Eugenetics 15: 323-354.

Wright, S. 1965. The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution, 19, 395–420.

Yang, J., J. Wang., J. Kijas., B. Liu., H. Han., M. Yu., H. Yang., S. Zhao, and K. Li. 2003. Genetic Diversity Present within the Near-Complete mtDNA Genome of 17 Breeds of Indigenous Chinese Pigs. Journal of Heredity 94(5):381-385.

Yu, G., H. Xiang., J. Wang, and X. Zhao. 2013. The Phylogenetic Status of Typical Chinese Native Pigs: Analyzed by Asian and European Pig Mitochondrial Genome Sequences. Journal of Animal Science and Biotechnology 4:9.

Yuan, J., H. Jianlin, and R. Blench. 2008. Livestock in ancient China: an archaeozoological persective. In: Sanchez-Mazas A, Blench R.M, and Ross M (eds) Human Migrations in Continental East Asia and Taiwan. Genetic, Linguistic and Archaeological Evidence. Abingdon: Routledge.

APPENDICES

Table 16. Chromosome Position and Sequence of 14 Microsatellite Markers from Pigs Common to Brazil and the United States

Marker Name	Chromosome	Position	Sequence
SW24	17	23.3	5’-3’: CTTTGGGTGGAGTGTGTGC 3’-5’: ATCCAAATGCTGCAAGCG
SW240	2	42.0	5’-3’: AGAAATTAGTGCCTCAAATTGG 3’-5’: AAACCATTAAGTCCCTAGCAAA
S0227	4	4.1	5’-3’: GATCCATTTATAATTTTAGCACAAAGT 3’-5’: GCATGGTGTGATGCTATGTCAAGC
SW951	10	101.0	5’-3’: TTTCACAACTCTGGCACCAG 3’-5’: GATCGTGCCCAAATGGAC
SW2406	6	21.4	5’-3’: AATGTCACCTTTAAGACGTGGG 3’-5’: AATGCGAAACTCCTGAATTAGC
SW830	10	0.0	5’-3’: AAGTACCATGGAGAGGGAAATG 3’-5’: ACATGGTTCCAAAGACCTGTG
SW2008	11	14.1	5’-3’: CAGGCCAGAGTAGCGTGC 3’-5’: CAGTCCTCCCAAAAATAACATG
S0026	16	46.9	5’-3’: AACCTTCCCTTCCCAATCAC 3’-5’: CACAGACTGCTTTTTACTCC
SW2410	8	-1.3	5’-3’: ATTTGCCCCCAAGGTATTTC 3’-5’: CAGGGTGTGGAGGGTAGAAG
SW72	3	17.8	5’-3’: ATCAGAACAGTGCGCCGT 3’-5’: TTTGAAAATGGGGTGTTTCC
SW857	14	7.4	5’-3’: TGAGAGGTCAGTTACAGAAGACC 3’-5’: GATCCTCCTCCAAATCCCAT
S0225	8	82.8	5’-3’: GCTAATGCCAGAGAAATGCAGA 3’-5’: CAGGTGGAAAGAATGGAATGAA
SW632	7	104.4	5’-3’: TGGGTTGAAAGATTTCCCAA 3’-5’: GGAGTCAGTACTTTGGCTTGA
S0226	4, 2		5’-3’: ACACACACATACACACACATGC 3’-5’: ATGAGAGCTGGGATTTGGG

Source: Flicek *et al*. (2014)

Table 17. Number of Alleles per Locus and Percentage of Missing Data from Merged Microsatellite Markers from Pigs Common to Brazil and the United States

Locus	Number of Alleles	Percentage of Missing Data
SW24	15	5.2
SW240	14	4.7
S0227	12	6.0
SW951	12	6.3
SW2406	14	6.3
SW830	13	0.0
SW2008	12	1.6
SW857	11	1.6
S0026	11	5.0
SW2410	17	2.6
S0225	12	5.0
SW72	9	2.6

Table 18. Estimated Log Likelihood Using Complete Microsatellite Markers from Brazilian Pigs with 14 Microsatellite Markers in Common with Pigs from the United States

Number of Run	K	Estimated log likelihood									
		Common Marker Set					Complete Marker Set				
		1	2	3	4	5	1	2	3	4	5
1	1	-8634.5	-8634.6	-8635.2	-8634.5	-8634.9	-18874.3	-18873.7	-18874.0	-18874.6	-18873.7
2	1	-8634.4	-8634.8	-8635.0	-8635.0	-8634.9	-18874.4	-18873.4	-18874.1	-18873.3	-18873.9
3	1	-8636.4	-8634.9	-8634.9	-8634.6	-8634.6	-18874.7	-18873.9	-18873.2	-18873.8	-18874.4
4	1	-8635.0	-8634.9	-8634.3	-8634.8	-8634.7	-18875.7	-18874.3	-18875.1	-18873.8	-18873.6
5	1	-8634.1	-8634.8	-8633.1	-8635.0	-8634.7	-18873.4	-18873.8	-18874.4	-18873.8	-18873.1
6	2	-8140.3	-8096.2	-8098.9	-8095.9	-8105.1	-17564.0	-17565.0	-17565.1	-17549.4	-17549.4
7	2	-8094.8	-8096.9	-8098.5	-8104.6	-8141.2	-17550.9	-17565.7	-17548.3	-17565.1	-17548.7
8	2	-8145.9	-8137.9	-8140.4	-8097.0	-8104.4	-17548.5	-17565.0	-17548.8	-17565.7	-17566.1
9	2	-8097.4	-8108.0	-8097.9	-8142.7	-8105.1	-17547.7	-17564.8	-17550.3	-17548.7	-17565.7
10	2	-8138.7	-8105.5	-8138.5	-8140.4	-8097.3	-17548.5	-17565.3	-17549.5	-17564.5	-17549.9
11	3	-7655.7	-7655.5	-7656.4	-7657.2	-7655.9	-16563.5	-16562.2	-16563.0	-16564.7	-16563.0
12	3	-7656.5	-7655.9	-7655.0	-7659.4	-7658.1	-17015.9	-16561.3	-16561.1	-16563.6	-16563.8
13	3	-7657.2	-7655.2	-7656.8	-7659.4	-7656.1	-16560.2	-16562.0	-16560.9	-16564.3	-16563.0
14	3	-7656.6	-7657.0	-7657.5	-7656.8	-7658.1	-16566.6	-16562.9	-16564.8	-16563.7	-16561.1
15	3	-7657.1	-7657.4	-7658.5	-7655.7	-7655.6	-16560.9	-16561.5	-16559.1	-16562.5	-16561.7
16	4	-7429.7	-7450.3	-7439.9	-7430.4	-7445.3	-16185.2	-16186.4	-16043.9	-16190.3	-16031.6
17	4	-7444.6	-7444.6	-7428.0	-7451.1	-7444.7	-16041.5	-16046.5	-16045.5	-16186.2	-16186.5
18	4	-7432.9	-7452.6	-7427.8	-7449.0	-7453.8	-16162.3	-16185.0	-16030.6	-16043.7	-16184.9
19	4	-7431.2	-7442.2	-7440.8	-7428.2	-7444.8	-16189.0	-16191.5	-16043.6	-16032.3	-16188.5
20	4	-7439.1	-7449.1	-7431.5	-7430.6	-7443.8	-16047.4	-16191.2	-16046.2	-16045.9	-16190.8
21	5	-7233.0	-7235.5	-7234.1	-7232.7	-7232.5	-15665.4	-15668.0	-15665.7	-15669.4	-15666.4
22	5	-7234.7	-7234.7	-7234.1	-7232.4	-7234.0	-15665.1	-15666.9	-15785.1	-15664.7	-15649.4
23	5	-7233.4	-7237.8	-7231.2	-7233.7	-7235.6	-15663.4	-15784.8	-15664.9	-15666.1	-15645.3
24	5	-7233.0	-7236.1	-7234.4	-7235.9	-7234.4	-15646.5	-15667.3	-15645.3	-15781.3	-15650.9
25	5	-7234.3	-7233.1	-7236.3	-7232.7	-7232.4	-15676.0	-15666.9	-15779.5	-15671.1	-15669.9

1 – 50000 x 10000, 2 – 50000 x 300000, 3 – 100000 x 100000, 4 – 100000 x 300000, 5 – 200000 x 600000

The first of each combination is the burn-in period while the second combination is the Markov Chain Monte Carlo replications

Table 19. Estimated Log Likelihood Using Complete Microsatellite Markers from Pigs in the United States with 14 Microsatellite Markers in Common with Pigs from Brazil

Number of Runs	K	Estimated log likelihood									
		Common Marker Set					Complete Marker Set				
		1	2	3	4	5	1	2	3	4	5
1	1	-8822.8	-8822.6	-8823.4	-8822.0	-8822.4	-23163.1	-23161.0	-23162.5	-23163.1	-23162.5
2	1	-8822.5	-8821.8	-8821.6	-8822.2	-8822.1	-23162.5	-23162.4	-23163.6	-23161.7	-23162.6
3	1	-8822.7	-8822.5	-8822.1	-8822.9	-8822.5	-23164.1	-23163.0	-23163.3	-23161.8	-23163.0
4	1	-8822.6	-8822.4	-8823.1	-8822.5	-8822.8	-23162.8	-23163.2	-23164.0	-23162.6	-23162.3
5	1	-8822.5	-8822.6	-8822.5	-8822.4	-8822.8	-23161.8	-23163.5	-23164.3	-23162.9	-23161.9
6	2	-7389.7	-7389.6	-7393.9	-7392.3	-7390.8	-19198.1	-19196.9	-19197.2	-19195.7	-19196.5
7	2	-7389.3	-7393.1	-7390.8	-7391.6	-7390.9	-19201.9	-19197.6	-19198.5	-19196.8	-19196.5
8	2	-7390.1	-7388.6	-7391.4	-7390.4	-7390.8	-19196.2	-19198.8	-19202.0	-19197.6	-19195.0
9	2	-7388.0	-7390.9	-7392.0	-7390.4	-7391.9	-19199.9	-19196.1	-19193.6	-19197.9	-19199.0
10	2	-7389.5	-7392.2	-7390.9	-7389.9	-7391.5	-19194.1	-19196.9	-19205.8	-19197.6	-19196.4
11	3	-6679.9	-6679.3	-6679.0	-6678.3	-6678.9	-17567.6	-17574.3	-17575.4	-17571.8	-17575.2
12	3	-6995.4	-6679.8	-6677.0	-6678.8	-6678.1	-17565.4	-17575.6	-17563.1	-18208.8	-17571.7
13	3	-6677.8	-6678.0	-6678.6	-6678.1	-6678.3	-17575.6	-17567.2	-17575.6	-17566.9	-17571.4
14	3	-6679.5	-6679.4	-6676.8	-6678.1	-6679.1	-17580.5	-17567.2	-17567.1	-17568.7	-17571.1
15	3	-6678.7	-6678.8	-6677.6	-6678.6	-6678.9	-17574.3	-17568.4	-17570.1	-17569.3	-17571.3
16	4	-6399.1	-6399.7	-6317.0	-6399.5	-6318.7	-16816.7	-16625.0	-16806.9	-16808.7	-16806.7
17	4	-6319.6	-6398.9	-6319.5	-6318.5	-6319.6	-16630.3	-16808.8	-16802.0	-16622.7	-16810.7
18	4	-6400.3	-6401.2	-6401.0	-6398.6	-6399.8	-16627.7	-16811.7	-16813.1	-16810.1	-16797.7
19	4	-6400.9	-6393.6	-6318.9	-6398.9	-6319.4	-16621.6	-16906.7	-16827.9	-16796.7	-16808.1
20	4	-6318.0	-6400.0	-6398.1	-6320.4	-6318.9	-16807.2	-16813.5	-16821.2	-16802.7	-16806.8
21	5	-6040.2	-6225.5	-6040.2	-6039.4	-6039.2	-16104.1	-16106.3	-15844.7	-16104.0	-16109.8
22	5	-6038.1	-6038.0	-6041.5	-6040.4	-6039.0	-16103.5	-16004.5	-16370.9	-16110.8	-16109.3
23	5	-6038.8	-6040.0	-6038.6	-6039.7	-6038.4	-16113.5	-16105.9	-15843.0	-15862.5	-16110.4
24	5	-6037.4	-6038.7	-6039.8	-6039.6	-6040.0	-16683.9	-15858.1	-16102.1	-15853.7	-16042.3
25	5	-6039.6	-6038.6	-6040.3	-6039.2	-6039.2	-15857.6	-16109.7	-16102.3	-15858.6	-16681.2

1 – 50000 x 10000, 2 – 50000 x 300000, 3 – 100000 x 100000, 4 – 100000 x 300000, 5 – 200000 x 600000

The first of each combination is the burn-in period while the second combination is the Markov Chain Monte Carlo replications

Table 20. Number of Alleles (Na), Effective Number of Alleles (Ne), Observed and Expected Heterozygosity (Ho and H_E), Number of Private Alleles (PA) and Allelic Richness (AR) Using 12 Merged Microsatellite Markers Common to Pig in Brazil and the United States

	Breed																
Para-meter	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Na	5.3	3.9	2.5	3.7	2.7	5.8	5.4	5.6	4.6	3.3	4.3	4.5	4.3	5.7	4.5	3.7	5.0
Ne	2.8	2.6	2.1	3.0	2.1	3.1	2.7	2.8	3.5	2.6	3.1	2.8	2.8	3.2	2.7	2.1	2.6
Ho	0.6	0.6	0.6	0.7	0.5	0.6	0.5	0.6	0.5	0.6	0.6	0.6	0.6	0.7	0.6	0.4	0.5
H_E	0.6	0.6	0.4	0.6	0.5	0.6	0.6	0.6	0.7	0.6	0.6	0.6	0.6	0.7	0.6	0.4	0.6
PA	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.3	0.4	0.2	0.0	0.2
AR	2.3	2.2	2.1	2.5	2.1	2.5	2.3	2.4	2.8	2.4	2.4	2.4	2.4	2.6	2.3	1.9	2.3

1 – MS60, 2 – Moura, 3 – Large White, 4 – Nilo, 5 Brazilian Duroc, 6 – Piau, 7 – Monteiro, 8 – Landrace, 9 – Mixed, 10 – Pietrain, 11 – US-Meishan, 12 – China-Meishan, 13 – Fengjing, 14 – Minzhu, 15 – Yorkshire, 16 – Berkshire, 17 – US Duroc