Genetic and Geographic Diversity of Gyr (Bos Indicus) Cattle in Brazil

Yu Wang
Registration Number: 1441626
06 / 2015

Division of Livestock Sciences (NUWI)
University of Natural Resources and Life Sciences, Vienna, Austria

SUPERVISORS:
Univ.Prof. Dr. Johann Sölkner
Dr. Gábor Mészáros
Dr. Marcos Vinicius Gualberto Barbosa da Silva

BOKU

Education and Culture
Erasmus Mundus

Contents

Abstract

Das Gyr (Bos indicus) ist eine der wichtigsten Rinderrassen zur Milchproduktion in tropischen Gebieten. Die Rasse hat sich in Brasilien aus relativ wenigen aus Indien importierten Gründertieren entwickelt, verbreitete sich jedoch kürzlich sehr weit und wurde gleichzeitig strenger Selektion unterzogen. Dieses Vorgehen könnte zu erhöhten Inzuchtwerten und niedriger genetischer Diversität innerhalb der Gyr Population geführt haben. Information über die genetische Diversität der Gyr Population und deren Einflussfaktoren sind deshalb unverzichtbar um nachhaltig Zuchtfortschritt gewährleisten sowie die genetische Vielfalt dieser Rasse zu erhalten. Diese Studie untersuchte die genetische Diversität der Rasse und den Zusammenhang zwischen genetischer und geografischer Distanz. Es wurden 588 Kühe von sieben Herden mit unterschiedlicher geografischer Lage aus zwei Brasilianischen Staaten für 45797 Einzelnukletid-Polimorphismus (SNP) Marker genotypisiert. Eine Teilmenge von 9176 SNP wurden zur Bestimmung der genetischen Diversität verwendet. Die insgesamte Heterozygotie der Gesamtpopulation lag bei 0.264 ± 0.158, wobei keine signifikanten Unterschiede zwischen den einzelnen Herden festgestellt werden konnten. Der Durchschnitt der genetischen Differenz zwischen allen Herden wurde mit dem F_{ST} Wert gemessen und betrug 0.050 ± 0.041. Ein F_{IT} Wert für die gesamte Population von 0.019 ± 0.058 sowie ein FIS Wert von -0.031 ± 0.047 zeigen einen Überschuss an Heterozygoten. Der Inzuchtgrad wurde als jener Anteil des Genoms der sich in „Runs of Homozygosity“ befindet gemessen und betrug 0.0537. Die Berechnung der effektiven Anzahl an Migranten pro Generation (Nm) ergab 3.49. Mit 532 km befand sich der größte geografische Abstand zwischen Herde_265 und Herde_551. Ebenso waren der F_{ST} Wert und Nei's genetische Distanz zwischen Herde_265 und Herde_551 am größten mit den entsprechenden Werten von 0.077 und 0.02. In der Hauptkomponentenanalyse (PCA) wies die erste Hauptkomponente 4.895% der gesamten Varianz aus und die zweite Hauptkomponente stand für 2.526% der Variation. Die Anwendung von Rousset’s Methode der Isolation durch Distanz zeigte einen linearen Zusammenhang zwischen genetischer und geografischer Distanz, wobei dieses Modell 29.89% der gesamten Varianz erklärt. Der Mantel Test, welcher Matrizen mit geografischer Distanz und genetischer Distanz vergleicht, zeigte beinahe signifikant, dass diese beiden Arten von Distanzen positiv miteinander korrelieren (r=0.624, P=0.068).

Keywords: Gyr; genetische Diversität; SNPs; genetische Distanz; geografische Distanz

Abstract

The Gyr cattle breed (*Bos indicus*) has become an important breed for milk production throughout the tropical areas. The breed was developed from a comparatively small number of founder animals imported to Brazil from India, with rapid recent expansion of the population and intensification of selection procedures. This strategy indicates a potential increase of inbreeding levels and reduction of genetic diversity in the Gyr population. Information about genetic diversity within Gyr cattle is therefore essential for genetic improvement, understanding of environmental adaptation as well as genetic conservation. This study investigated the genetic diversity and the relationship between genetic distance and geographic distance. In total, 588 Gyr cows of seven herds from different geographical locations of two states in Brazil were genotyped for 45797 single nucleotide polymorphism (SNP) markers and a subset of 9176 SNP were used to assess genetic diversity. The overall expected heterozygosity of total population was 0.264 ± 0.158, and there was no significant difference between herds. The average of genetic differentiation among all herds measured as F_{ST} value was 0.050 ± 0.041. F_{IT} value for the whole population was 0.019 ± 0.058 while the F_{IS} value for the whole population was -0.031 ± 0.047, showing an excess of heterozygotes. Genome wide level of inbreeding based on proportions of the genome being in runs of homozygosity was 0.0537. The assessment of the effective number of migrants by generation (*Nm*) was 3.49. The largest geographic distance between subpopulations in this study was between Herd_265 and Herd_551 with a spatial distance of 532 km. Both F_{ST} and Nei's genetic distance were also greatest between Herd_265 and Herd_551 with values of 0.077 and 0.022 respectively. In the principle component analysis (PCA), the first principle component）accounted for 4.895% of the total variance and the second principle component condensed 2.526% of the variation. The application of Rousset's isolation by distance method provided a linear relationship between genetic distance and geographical distance: $\frac{F_{ST}}{1-F_{ST}} = -0.035 + 0.014 \ln(d)$; with 29.89% of the variance explained by this model. The Mantel test, comparing matrices of geographical and genetic distances indicated a positive correlation between those two types of distance, with a trend toward significance (r=0.624, P=0.068).

Keywords: Gyr cattle; biodiversity; SNPs; genetic distance; geographical distance

Introduction

The Gyr breed (*Bos indicus*) is one of the principal Zebu breeds originated from India. The native origin of this breed is around the Gyr hills and forests of Kathiawar in India, which lies between 20 ° and 22 ° North latitude and 70 ° and 72 ° East longitude (Gaur, Kaushik et al. 2003). The Gyr animals are famous for their high milk yield producing ability (Madalena 1988), tolerance to heat stress (Torres-Júnior, de FA Pires et al. 2008) and resistance to various tropical diseases. Thus in the last century Brazil, Mexico and USA imported these animals where they have been bred successfully.

The Gyr breed arrived in Brazil via importation, started from 1870 with a total number of 700 Gyr animals. Currently the population size has already reached 400,000 all over Brazil (Santana, Pereira et al. 2014). At the beginning Gyr was imported as beef cattle, but some breeders started to use them for milk production afterwards (Queiroz and Lôbo 1993). Gyr breed is also mainly utilized as the basis for crosses with European dairy breeds in order to maintain rusticity, adaptability and resistance to parasites, especially in grazing systems. The breeding program plays a role not only for genetic improvement but also for maintaining the genetic diversity of a population. The last major export of the breed to Brazil took place in 1960, after which laws made the importation and exportation of animals rather difficult. Due to the rapid growth and dissemination of Gyr breed, along with the limited introduced animals, the demand of genetically superior proven sires has increased rapidly in Brazil. The current situation is that a small number of proven bulls with high breeding value are frequently used. Therefore the risk of reducing genetic diversity and increasing inbreeding level becomes much higher, which may lead to reproductive, productive and economic loss in the future. Thus it has already becoming an urgent issue of maintaining genetic variability in the Gyr cattle in Brazil.

There are many influential factors that may contribute to the change of genetic diversity such as artificial selection (Charlesworth, Nordborg et al. 1997), migration (Goossens, Chikhi et al. 2005), mutation (Lacy 1987) etc. Since artificial selection and migration may lead to the

spatial separation of animals, the geographical dispersion may also become an indirect influential factor to genetic differentiation. It has already been suggested by two theories that an increase of geographic distance will lead to the increase of genetic diversity: the theory of isolation by distance (Wright 1943) and by the stepping-stone model (Kimura and Weiss 1964). The theory of isolation by distance was first developed by Sewall Wright indicates that populations in remote locations may become differentiated simply by isolation by distance. This isolation by distance can create genetic differentiation among subpopulations. Individuals within a subpopulation are neighbors in the sense that their gametes may come together and inbreeding within the subpopulation reduces heterozygosity. Both models indicate that gene flow can be limited to shorter distances with greater genetic differentiation with increasing geographic distance.

There are several studies related to the relationship between geographical distance and genetic distance in different species. In human population, it has been confirmed that a strong positive linear relationship existed between genetic differentiation and geographic distance in African human population (Ramachandran, Deshpande et al. 2005). Afterwards a similar genetic structure study of European human population supported the former study result that the small genetic differentiation present between subpopulations was characterized by a significant correlation between genetic and geographic distance (Lao, Lu et al. 2008). However this significant positive correlation cannot always be found. In chicken, relevant results showed that although the potential linear correlation between geographical distance with genetic distance may exist, the correlation was non-significant (Bao, Shu et al. 2009). In a similar study of European sheep breeds, PCA analysis on SNP data showed that differentiated breeds were with good correspondence to geographical locations. However the genetic diversity is a consequence mainly brought by selection operated by local sheep farmers, different flock management and breed admixture. There is no solid evidence to prove that geographical isolation is a potential reason for genetic differentiation (Pariset, Mariotti et al. 2011).

In several similar studies conducted in cattle, the results showed that genetically close related cattle breeds were more likely to be clustered in close original geographical locations (Felius 1995, Gautier, Faraut et al. 2007, Kugonza, Jianlin et al. 2011, Pham, Do et al. 2013). Yet, a significant relationship between spatial distances and genetic distance could not be detected. Relevant research in beef cattle showed that neither isolation by distance nor hierarchical structure associated with geography were detected in 18 local breeds from southwest Europe (Jordana, Alexandrino et al. 2003). The main influential factor of genetic differentiation was genetic drift.

To the best of our knowledge, no similar research has been done for Gyr cattle in Brazil, connecting genetic differentiation and geographic isolation. The aim of this study was to evaluate genetic diversity of Gyr cattle breed in Brazil and analyze the relationships between all subpopulation pairs of geographical distance and genetic distance. The results aim to provide suggestions and related information for decisions concerning breeding in the future.

Materials and Methods

Data Resource

To evaluate the genetic diversity and investigate whether there is a significant relationship between geographic and genetic distance in Brazilian Gyr cattle, we applied both GPS data and genotype data managed by EMBRAPA Dairy Cattle, Brazil. The data set consisted of 1613 cows from 81 herds for this study. The original GPS dataset consisted of GPS coordinates and altitude information of 277 herds spread in 8 states in Brazil. The original genotype dataset consisted of 445 bulls and 1,663 cows genotyped with the Illumina BovineSNP50 v2 (Illumina Inc., San Diego, CA, US). For the current study only the genotype and GPS information for cows were used.

Several editing steps were conducted prior to analysis: 596 cows from 12 herds matched with the corresponding GPS coordinates were used for the next step analyses; 4 herds (Herd_1054, Herd_1198, Herd_1309 and Herd_31681) were removed because for each herd there was only a single cow record; The geographical distance between herd_1304 and herd_24426 was 0.000134259 km after calculation. So the two animals in herd_1304 were merged into herd_24426; 4 targeted cows (CAL4235, FBGA5186, EFC376, RRP4927) without genotype information were removed. After editing, a total of 588 records from 7 Gyr cattle herds representing 7 subpopulations were retained for this study. Population size, farm name and relevant geographical information can be seen in Table 1. The corresponding locations of the target herds were presented via Google Earth (Figure 1).

Table 1 Description of 7 targeted herds of Gyr cattle in Brazil

Herd ID	Population Size	Farm Name	City	State	Latitude	Longitude	Altitude (m)
265	267	Fazenda Brasília	SÃO PEDRO DOS FERROS	Minas Gerais	-19.93	-42.57	268
266	72	Santana da Serra	CAJURU	São Paulo	-21.36	-47.21	817
270	38	Estância Silvânia	CAÇAPAVA	São Paulo	-23.20	-45.73	585
277	77	Terra Vermelha/Campo Alegre	ITOBI	São Paulo	-21.89	-46.92	690
551	81	Nova Estiva	BURITIZAL	São Paulo	-20.25	-47.65	916
750	5	NA	GUAPÉ	Minas Gerais	-20.71	-45.95	817
24426	46	NA	MOCOCA	São Paulo	-21.33	-47.10	648

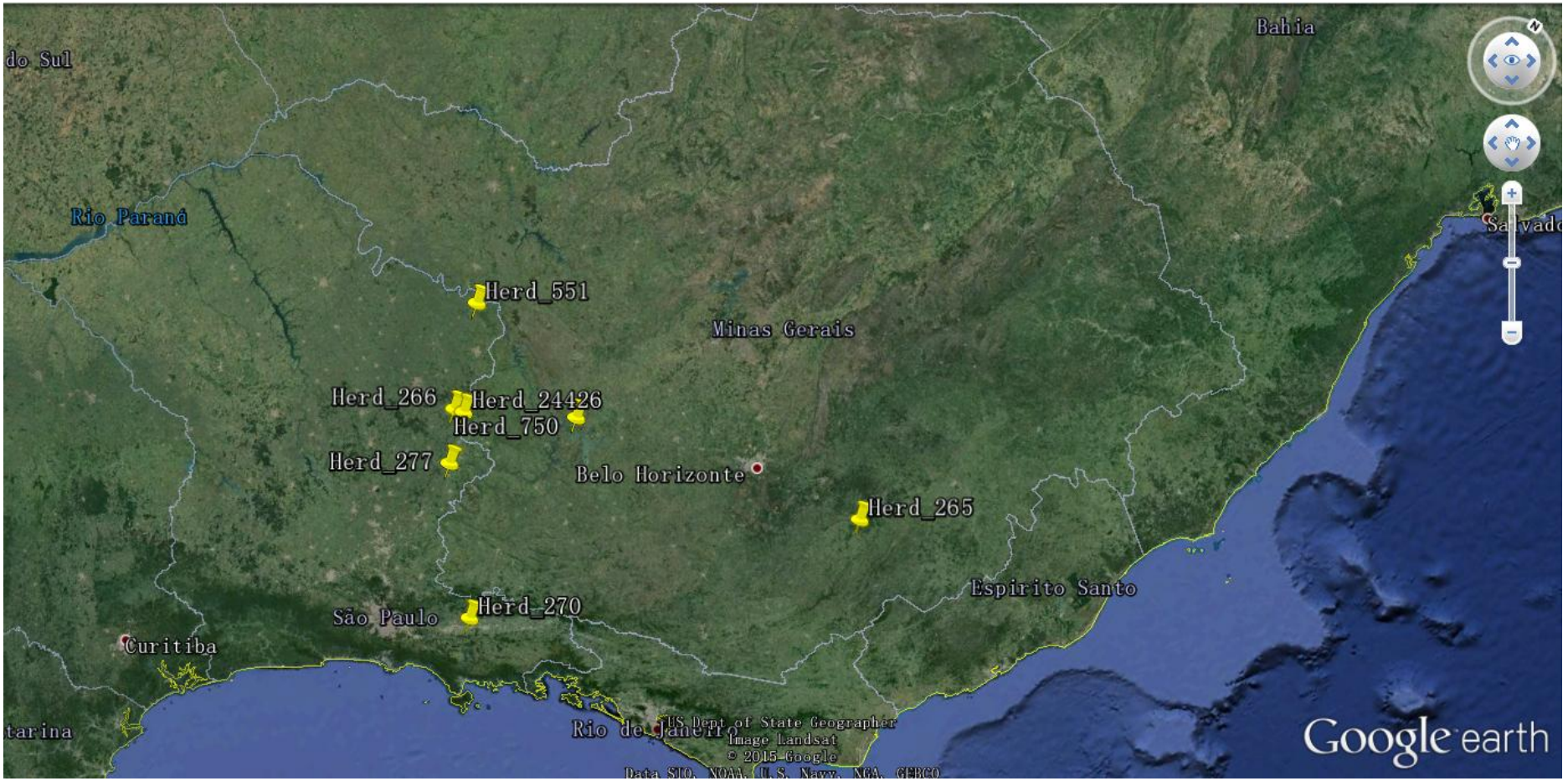

Figure 1 Farm locations in Brazil where the cattle populations were sampled indicated by yellow pins.

Quality control process and marker selection

Original genotype dataset consisted of 45797 single nucleotide polymorphic markers. Genotype quality checks were accomplished via using PLINK v1.9 (Chang, Chow et al. 2014): Markers selected for diversity analysis were required to be located on autosomal chromosomes; Only animals with call rates ≥95% and SNP call rate ≥95% were kept in the final dataset (no animal and SNPs eliminated); The SNP markers with minor allele frequency (MAF) <0.01 were discarded (19323 SNPs were eliminated). SNP markers with value for the Hardy-Weinberg equilibrium test $<10^{-6}$ were also discarded (33 SNPs were eliminated). After qualify control process, 26441 SNPs were retained. Due to the SNP number limitation (10000 SNPs maximum) in the SPAGeDi v1.4c (Hardy and Vekemans 2002) software employed for analysis, 35% SNPs were randomly selected. A total number of 9176 SNPs were kept as the final genotype dataset for most analyses. For the search of segments of the genome that are autozygous, the full set of 26441 SNPs was used.

Statistical analyses

Analyses of within population genetic diversity

Observed heterozygosity (H_O), expected heterozygosity (H_E) and allelic richness (A_R) were applied to evaluate genetic variability and compare the levels of heterogeneity within population. Observed heterozygosity, the frequency of heterozygous individuals in a population, and expected heterozygosity, the probability that two gametes randomly chosen from the gene pool, are the measures most commonly used by papers that present a genetic summary of populations. Both measures are very sensitive to the allele frequencies in the population rather than just to the number of alleles. Allelic richness (also referred to as allelic diversity), calculated as the average number of alleles per locus, is another commonly reported measure of genetic variation (Leberg 2002). Observed heterozygosity, expected heterozygosity and allelic richness were calculated via using SPAGeDi v1.4c. Besides, except

for the above three parameters, the inbreeding coefficient (Fi) for each herd was also calculated via SPAGeDi v1.4c. The statistical significance between parameter values of subpopulations and the values of overall population were tested via t-test in R (Team 2014) respectively. Level of autozygosity, which has been confirmed to be a realistic indicator of level of inbreeding (Curik, Ferenčaković et al. 2014), was calculated using the SNP & Variation suite v7.6.8 Win64 from Golden Helix (SVS) (http://www.goldenhelix.com). Runs of homozygosity (ROH) segments longer than 4 Mb were searched for, under the following criteria among all chromosomal segments: 15 or more coherent homozygous SNPs, a density of at least 1 SNP every 150 kb, gaps of no more than 1000 kb between SNPs and less than 2 missing genotypes across all individuals. In order to take genotyping errors into account and avoid underestimation of long ROH (Ferenčaković, Hamzić et al. 2013), 1 heterozygous genotype call per segment was allowed. According to McQuillan et al. (McQuillan, Leutenegger et al. 2008) the genomic inbreeding coefficient (F_{ROH}) was estimated by the sum of ROH lengths of an individual divided by the total length of the autosomes as follows:

$$F_{ROH} = \frac{\sum_{j=1}^{n} L_{ROH_j}}{L_{TOTAL}} \quad [1]$$

which L_{TOTAL} was the total size of the genome covered by markers, calculated from the sum of inter marker distances in the UMD v3.1 assembly as 2,506,343,112 bp. Results obtained from SNP & Variation Suite were analyzed using R and package "psych" (Revelle 2014). Inbreeding levels were calculated for the whole population and for each herd.

For allele frequencies, both minor allele and major allele for each herd were calculated via SPAGeDi v1.4c. Deviation from Hardy-Weinberg equilibrium (HWE; heterozygote deficiency) was performed in PLINK v1.9 for each SNP and each population.

Genetic distance and geographical distance among different populations

A hierarchical analysis of molecular variance (AMOVA) was applied using the software ARLEQUIN 3.5 (Excoffier and Lischer 2010) to quantify the degree of differentiation among different subpopulations and the differences were tested for significance using 1023 bootstrap

permutations. Genetic differentiation between subpopulations were estimated using F_{ST} coefficient (Weir and Cockerham 1984). F_{ST} is a measure of allele frequency divergence among demes or subpopulations which can be described as the amount of allele frequency variance in a subpopulation relative to the maximum variance which can be defined as follows:

$$F_{ST} = \frac{\sigma_S^2}{\sigma_T^2} = \frac{\sigma_S^2}{\bar{p}(1-\bar{p})} \quad [2]$$

where σ_S^2 is the variance in the frequency of the allele between different subpopulations, weighted by the sizes of the subpopulations and σ_T^2 is the variance of the allelic state in the total population, $\bar{p}$ is the average frequency of an allele in the total population. And from population level, relevant F-statistics can be calculated as below:

$$F_{ST} = \frac{F_{it}-F_{is}}{1-F_{is}} \quad [3]$$

where F_{it} is a measure of the overall departure from Hardy-Weinberg (HW) proportions in the entire population due to both nonrandom mating within local subpopulations (F_{is}), and allele frequency divergence among subpopulations (F_{ST}); F_{is} is a measure of departure from Hardy-Weinberg proportions within local demes or subpopulations due to nonrandom mating. Pairwise genetic distances, D_S (Nei 1972), between subpopulations were estimated. Nei's standard genetic distance (D_S) is a measure of genetic differentiation between two populations often used in phylogenetic reconstruction, which is calculated as below:

$$D_S = -\ln\left(\frac{J_{XY}}{J_{XX}J_{YY}}\right) \quad [4]$$

where $J_{XY} = \sum_{i=1}^{m}\sum_{j=1}^{r} x_{ij}y_{ij}/r$; $J_{XX} = \sum_{i=1}^{m}\sum_{j=1}^{r} x_{ij}^2/r$; $J_{YY} = \sum_{i=1}^{m}\sum_{j=1}^{r} y_{ij}^2/r$. x_{ij} is the frequency of the i-th allele at the j-th locus in population X, and y_{ij} is the frequency of the i-th allele at the j-th locus in population Y.

The effective number of migrants (*Nm*) was estimated, assuming the n-island model of population structure, on the basis of the relationship:

$$F_{ST} = \frac{1}{1+4Nm\alpha}, where\ \alpha = \left(\frac{n}{n-1}\right)^2 \quad [5]$$

Isolation by distance was investigated as the correlation between pairwise F_{ST} / (1 - F_{ST})

and ln(km) (Rousset 1997). A linear regression was used to estimate the coefficients:

$$\frac{F_{ST}}{1-F_{ST}} = \alpha + \beta \ln(km) \quad [6]$$

Pairwise geographic distances were calculated based on great circle distances using the package "geosphere" (Hijmans, Williams et al. 2014) in R, according to which the distance D between two points specified by (latitude, longitude) coordinates (α_1, δ_1) and (α_2, δ_2), with a central angle of θ between the two point is

$$D = 2R \arctan\left(\frac{\sqrt{sin^2\left(\frac{\delta_1-\delta_2}{2}\right)+cos\delta_1 cos\delta_2 sin^2\left(\frac{\alpha_1-\alpha_2}{2}\right)}}{\sqrt{1-\left[sin^2\left(\frac{\delta_1-\delta_2}{2}\right)+cos\delta_1 cos\delta_2 sin^2\left(\frac{\alpha_1-\alpha_2}{2}\right)\right]}}\right) \quad [7]$$

and R is the radius of the Earth, which we assume to be 6378.187km (Shumaker and Sinnott 1984).

Pearson correlation coefficients were computed for the genetic divergence (F_{ST}) between subpopulations and the great-circle geographical distance between subpopulations. The statistical significance of both sets of correlation coefficients was assessed by means of a Mantel test (Smouse, Long et al. 1986) with 999 permutations via the package "vegan" (Oksanen, Blanchet et al. 2013) in R. Based on individual level, the genomic relationship matrix was constructed using the software "Gmatrix" (Su and Madsen 2011). The geographic data were converted into a spatial distance matrix, where all the individuals from same herd shared the same average spatial location. Similar Mantel test was also performed to test the correlation coefficient between genomic relationship matrix and spatial distance matrix.

To investigate relationships between subpopulations, neighbor-joining (N-J) (Saitou and Nei 1987) dendrograms were constructed from Nei's D_S genetic distances using Poptree Version 2 (Takezaki, Nei et al. 2010). Bootstrap values were obtained with 1,000 replicates over loci. A principal component analysis (PCA) was carried out to illustrate the relationship among the subpopulations using PLINK v1.9 with default settings. PCA was carried out based on the variance-standardized relationship matrix and top principal components were generally used as covariates in association analysis regressions to help correct for population stratification. Top 20 principal components of the variance-standardized relationship matrix were extracted and eigenvectors were calculated. Based on the result obtained from PLINK

v1.9, a PCA plot was obtained from the software "Genesis" (Buchmann and Hazelhurst 2014) .

Results

Genetic variability within Gyr cattle sub-populations

A total of 9176 SNPs in 588 samples from seven Gyr cattle subpopulations were studied. Expected heterozygosity, observed heterozygosity, allelic richness, individual inbreeding coefficient, genomic inbreeding level, major allele frequency, minor allele frequency and percentage of SNPs which were not in Hardy Weinberg Equilibrium for each subpopulation are given in Table 2.

The expected heterozygosity ranged from 0.239 ± 0.183 (Herd_750) to 0.262 ± 0.160 (Herd_24426) and the value for the full population was 0.264 ± 0.158, which was not significantly different from the values of each of the subpopulations (P=0.265). For average observed heterozygosity, the lowest value appeared in Herd_277 (0.260 ± 0.171) and highest in Herd_270 (0.268 ± 0.189) and the value for the overall population was 0.263 ± 0.158, which was not significantly different (P= 0.281). The allelic richness among all subpopulations ranged from 1.65 (Herd_750) to 1.71 (Herd_265 and Herd_24426). For inbreeding coefficient (F_{IS}) in each subpopulation, the lowest inbreeding coefficient was shown in Herd_270 with a value of -0.045, while the highest was shown in Herd_750 with a value of 0.003. Based on calculation of F_{IS}, the inbreeding coefficient of overall population was -0.031, which has no significant difference with all subpopulations with a P-value 0.472. When inbreeding level was calculated based on homozygous segments >4Mb, F_{ROH} for the whole population was 0.0537. Herd averages of F_{ROH} were from 0.0373 (Herd_270) to 0.0663 (Herd_265), with no significant difference between herds (P = 0.0797).

There was no significant difference in minor allele frequencies among subpopulations. Minor allele frequency ranged from 0.442 ± 0.351 in Herd_551 to 0.445 ± 0.347 in Herd_270 and Herd_750. The P-value for the difference in both minor and major allele frequencies was 0.638.

Table 2 Genetic variability within Gyr cattle sub-populations*

Population	A_R	He (SD)	Ho (SD)	F_{IS}	F_{ROH}	Major Allele Frequency (SD)	Minor Allele Frequency (SD)
Herd_265	1.71	0.251 (0.171)	0.261 (0.180)	-0.037	0.0663	0.557 (0.348)	0.443 (0.348)
Herd_266	1.66	0.259 (0.163)	0.266 (0.172)	-0.018	0.0415	0.556 (0.342)	0.444 (0.342)
Herd_270	1.70	0.254 (0.171)	0.268 (0.189)	-0.045	0.0373	0.555 (0.347)	0.445 (0.347)
Herd_277	1.68	0.256 (0.165)	0.260 (0.171)	-0.011	0.0505	0.556 (0.345)	0.444 (0.345)
Herd_551	1.69	0.247 (0.174)	0.264 (0.190)	-0.059	0.0428	0.558 (0.351)	0.442 (0.351)
Herd_750	1.65	0.239 (0.183)	0.265 (0.241)	0.003	0.0423	0.555 (0.357)	0.445 (0.357)
Herd_24426	1.71	0.262 (0.160)	0.266 (0.169)	-0.004	0.0392	0.557 (0.340)	0.443 (0.340)
Overall population	1.71	0.264 (0.158)	0.263 (0.158)	-0.031	0.0537	0.556 (0.347)	0.444 (0.347)

*: A_R (k=10); Allelic richness (expected number of alleles among 10 gene copies); He: gene diversity corrected for sample size; Ho: observed heterozygosity; F_{IS}: individual inbreeding coefficient based on information on Ho and He; F_{ROH}: genomic inbreeding level, level of autozygosity

Genetic distance and geographic distance between subpopulations

Population differentiation was examined by fixation indices F_{ST}, F_{IT} and F_{IS} across all loci. The average of genetic differentiation among all herds measured as F_{ST} value was 0.050 ± 0.041. F_{IT} value for the whole population was 0.019 ± 0.058 while the F_{IS} value for the whole population was -0.031 ± 0.047, showing an excess of heterozygotes (negative value). Analysis of molecular variance illustrated that within subpopulation genetic variation accounted for 94.98% among 7 subpopulations and between subpopulation genetic variation accounted for 5.02% of the variance. For each pair of subpopulations, the differentiation index (F_{ST}) and its significance are presented in Table 4 above diagonal, with the significant values marked bold. F_{ST} values ranged from 0.004 (pairwise between Herd_277 and Herd_750) to 0.077 (pairwise between Herd_265 and Herd_551). All the pairwise F_{ST} values showed significant difference except pairwise F_{ST} between Herd_277 and Herd_750 (F_{ST}=0.005) and between Herd_750 and Herd_24426 (F_{ST}=0.006). Nei's genetic distances between subpopulations are shown in Table 4 below diagonal. The greatest Nei's genetic distance was found between Herd_265 and Herd_551 with the value of 0.022 while the smallest Nei's genetic distance was found between Herd_277 and Herd_750 with the value of 0.001. The assessment of the effective number of migrants by generation (*Nm*) in Gyr populations was 3.49.

The geographical distance between seven herds constituted a symmetrical matrix shown in lower triangle of Table 5. The greatest geographical distance appeared between Herd_265 and Herd_551 with a spatial distance of 532 km while the shortest geographical distance appeared between Herd_266 and Herd_277 with a spatial distance of 66 km. The application of Rousset's isolation by distance method, as implemented in SPAGeDi v1.4c, allowed the computation of parameters α and β in Equation 6. The values obtained were -0.035 and 0.014 for both α and β, respectively, with 29.89% of the variance explained by the regression model (R^2).

$$\frac{F_{ST}}{1-F_{ST}} = -0.035 + 0.014 \ln(d) \quad [8]$$

The Mantel's test indicated that a potential positive correlation between genetic distance and geographic distance existed at a considerable trend toward significance (r=0.624, P=0.068).

The phylogenetic relationships based on a neighbor-joining tree constructed based on Nei's genetic distance and PCA were in agreement. Figure 3 shows an unrooted phylogenetic neighbor-joining tree obtained from Poptree Version 2. The tree branch sizes are proportional to the genetic distances among the Brazilian Gyr cattle subpopulation in this study. The numbers shown on the nodes are the calculated bootstrapping values after 1000 replicates. The neighbor-joining tree displayed two main clusters: Herd_270 and Herd-551 (Cluster 1), Herd_265, Herd_24426, Herd_277, Herd_750 and Herd_266 (Cluster 2). Herd_265 and Herd_551 had the longest genetic distance which has also been confirmed in the PCA plot (Figure 4). The first principle component（PC） accounts for 4.895% of the total variance and clearly distinguishes Herd_265 and Herd_750. The second principle component condenses 2.526% of the variation.

Table 3 Analysis of Molecular Variance among seven Gyr cattle subpopulations in Brazil

Source of Variation	Sum of Squares	Variance Components	Percentage variation
Among Populations	60174.139	61.779	5.023
Within Populations	1365574.877	1168.156	94.977
Total	1425749.015	1229.935	

Table 4 Pair-wise genetic differentiation (F_{ST}) values (above diagonal) and Nei's standard genetic distance (Ds) value (below diagonal) between the seven cattle populations. Significant pairwise F_{ST} values were marked bold

Population	Herd_265	Herd_266	Herd_270	Herd_277	Herd_551	Herd_750	Herd_24426
Herd_265	-	**0.053**	**0.060**	**0.051**	**0.077**	**0.052**	**0.032**
Herd_266	0.019	-	**0.036**	**0.029**	**0.056**	**0.019**	**0.015**
Herd_270	0.022	0.013	-	**0.034**	**0.058**	**0.029**	**0.022**
Herd_277	0.018	0.010	0.012	-	**0.057**	0.005	**0.007**
Herd_551	0.028	0.020	0.021	0.021	-	**0.051**	**0.047**
Herd_750	0.017	0.006	0.009	0.001	0.016	-	0.006
Herd_24426	0.011	0.005	0.008	0.003	0.017	0.002	-

Table 5 Pairwise spatial distances (km) between seven cattle populations

	Herd_265	Herd_266	Herd_270	Herd_277	Herd_551	Herd_750	Herd_24426
Herd_265							
Herd_266	508.373						
Herd_270	488.928	254.935					
Herd_277	501.304	65.980	190.076				
Herd_551	531.671	131.152	382.745	196.822			
Herd_750	363.098	149.329	277.638	164.742	184.081		
Herd_24426	496.376	12.008	250.793	64.491	132.846	137.695	

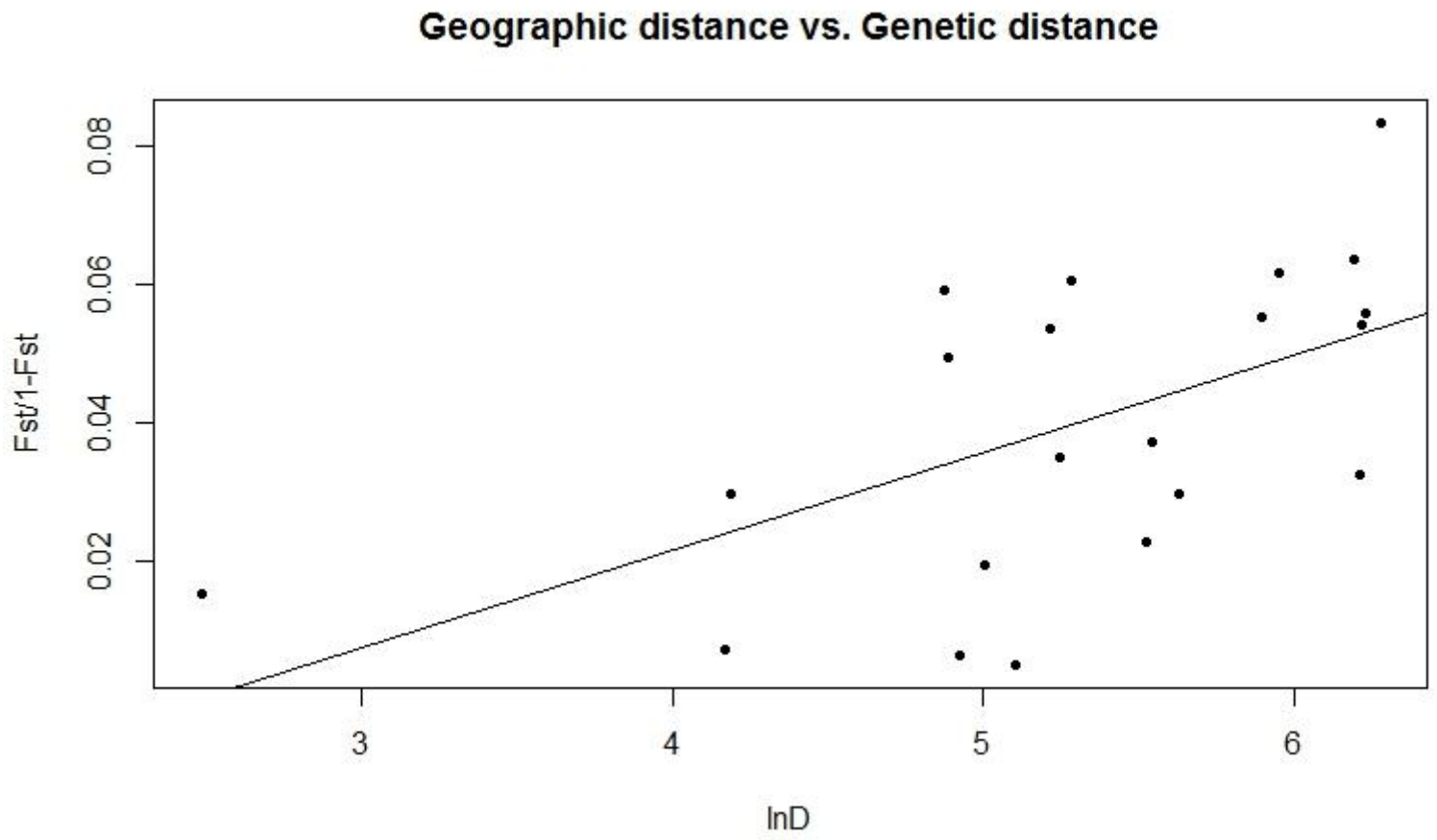

Figure 2 Plot of relationship between geographical distance ln(d) and pairwise $F_{ST}/(1-F_{ST})$ for all subpopulation pairs

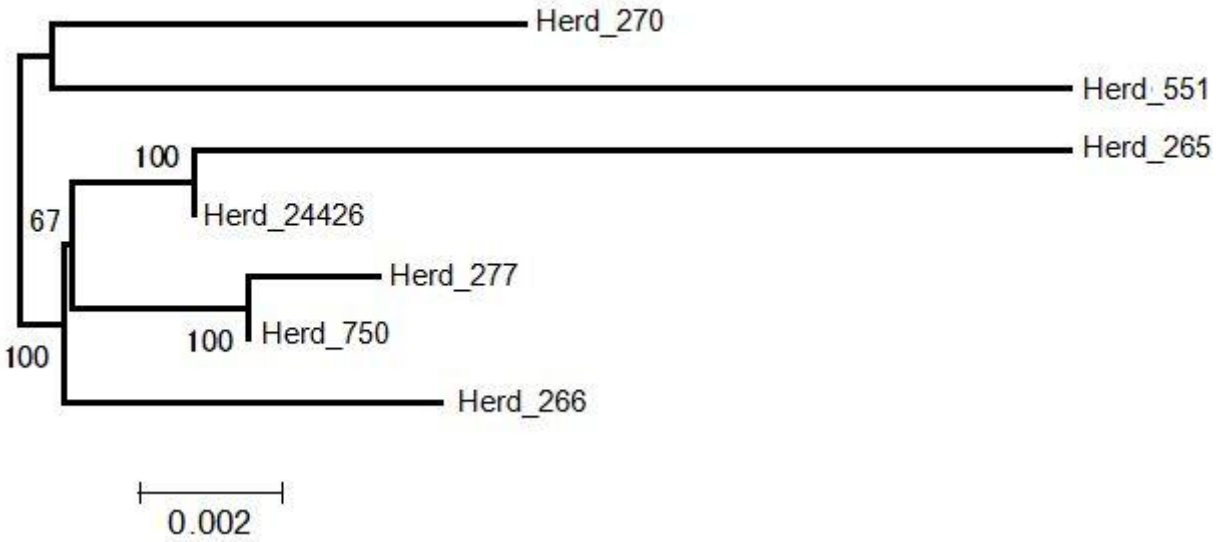

Figure 3 Dendrogram showing genetic similarities among 7 Brazilian Gyr cattle subpopulations. This tree is constructed by the neighbor-joining method from Nei's genetic distance. Numbers at the nodes represent the percentage of a group's occurrence in 1000 bootstrap replicates. This is an unrooted tree. The scale bar represents a length of 0.002.

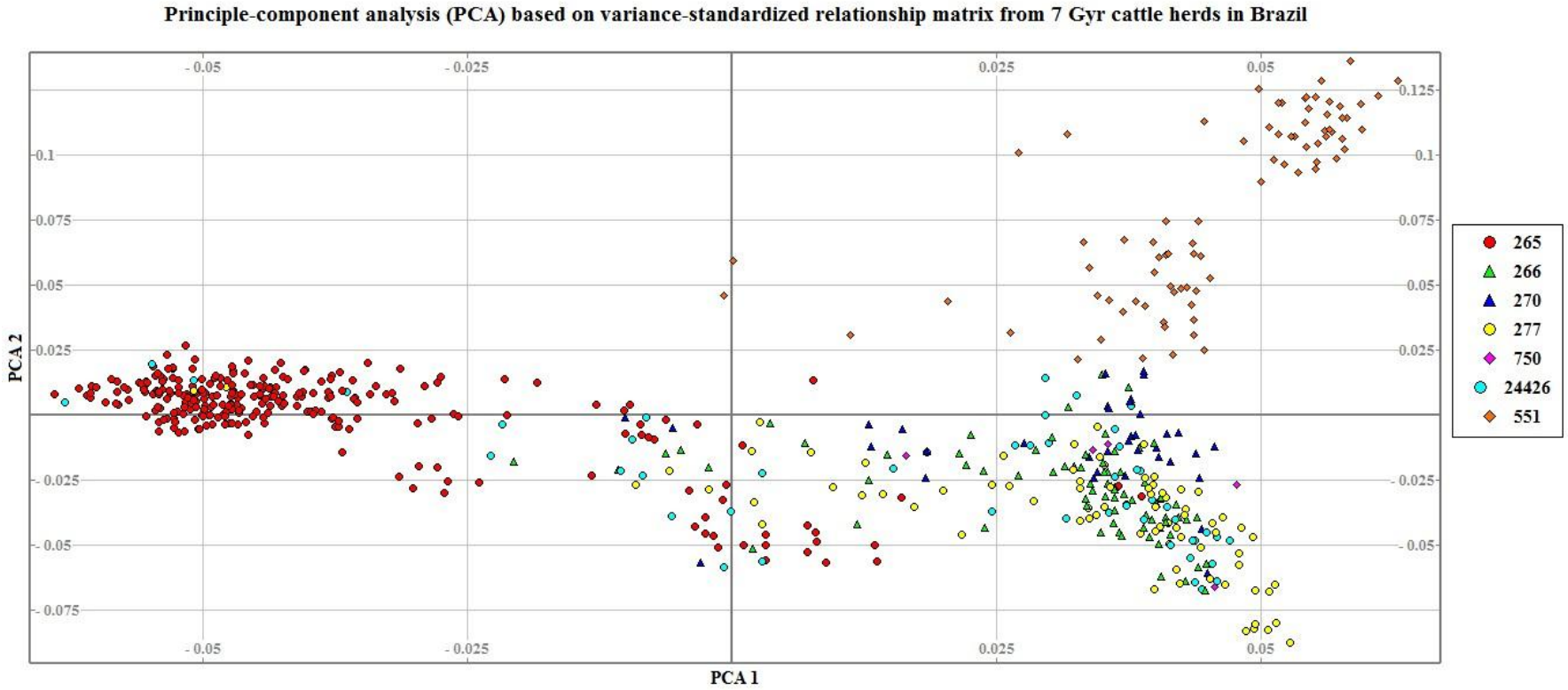

Figure 4 SNP-based Principal Component Analysis (PCA) of 588 Gyr cattle from 7 subpopulations.

Discussion

Analyses of within-breed genetic diversity determination

Genetic diversity is essential for genetic improvement and to promote rapid adaptation to changing environments and breeding goals. In this study, 588 cows from seven different geographical locations of two states in Brazil were genotyped for 45797 single nucleotide polymorphism (SNP) markers and 9176 of them were applied to assess genetic diversity. For the allele frequency there was no significant difference between the values obtained from different herds. The nonsignificant values between subpopulations do not prove the loss of genetic variation between subpopulations. Among all the herds, the average heterozygosity was found to be 0.263, which is obviously higher than the result obtained from a similar study based on SNP data of French Gyr cattle (0.156) (Gautier, Laloë et al. 2010). However the current value of heterozygosity is much lower compare to other former diversity analysis of Gyr cattle based on microsatellite dataset (Machado, Schuster et al. 2003, Azevedo 2010). It was concluded that the average direct count heterozygosity for the nine loci was 0.306 and the expected Hardy-Weinberg heterozygosity was 0.550 in 2003. The research afterwards confirmed that the mean observed and expected heterozygosities for Gyr cattle across loci were 0.73 and 0.75 respectively. The results of both observed and expected heterozygosity values obtained from SNP markers were much lower than those estimated via using microsatellite markers. It has become a debate of how to compare diversity estimates among markers, with much focus on the effect of different mutation rates and levels of heterozygosity between highly polymorphic markers, such as microsatellites, and less variable markers such as SNPs (Waples and Gaggiotti 2006). It has been proved that the mutation rates are relatively lower for SNPs than for microsatellite markers. For a microsatellite marker, the expected locus-specific heterozygosity may reach more than 0.95, while for SNP marker the maximum expected heterozygosity can only reach to 0.5. Thus this could be a reflection of the multi-allelic nature of microsatellite markers to achieve a higher

heterozygosity level than SNP markers. (Vignal, Milan et al. 2002).

The level of heterozygosity of Gyr cattle may also varied because of different spatial locations. A recent study revealed that based on microsatellite information, the observed and expected heterozygosity of Gyr cattle can reach 0.600 and 0.663 respectively in US (Villalobos-Cortés, Martínez et al. 2015). Similar study showed that average heterozygosity of Gyr cattle was observed to be 0.679 ± 0.09 among all the loci using microsatellite markers in India (Kale, Rank et al. 2010). This could be explained by differences in population history and management.

The inbreeding level based on F_{IS} values of current Gyr population in Brazil is lower compared to previous results. F_{IS} was -0.031 ± 0.047 and except for Herd_750, all the values of inbreeding coefficients were negative, which indicated an excess of observed over expected heterozygosity. This negative inbreeding value was also confirmed recently where F_{IS} was evaluated as -0.0097 in Gyr cattle in Brazil (O'Brien, Höller et al. 2015). Similar result was obtained that based on allele frequency, F_{IS} value of Gyr was 0.0054 (Porto-Neto, Sonstegard et al. 2013). With traditional methods using pedigree information, in 2007 an average inbreeding coefficient of 1.96% was obtained in a commercial cattle breeds in Brazil (Egito, Paiva et al. 2007). Similar average inbreeding coefficients of 2.82% (Reis Filho, Lopes et al. 2010) and 2.14% (Santana, Pereira et al. 2014) were obtained on the population structure analysis in Gyr cattle. However compared the other two alternative approaches, the result obtained from F_{ROH} showed a higher level of inbreeding. Currently, ROH is the most reliable approach to estimate the levels of inbreeding in cattle. Compared to the estimation based on pedigree information, the inbreeding coefficient based on ROH can capture both recent and distant inbreeding. Besides, ROH is also sensitive to stochastic nature of recombination (Ferenčaković, Hamzić et al. 2013). Yet, the levels of autozygosity captured by runs of homozygosity are substantially affected by the minimum length of a ROH allowed. There, 4Mb was chosen as a threshold based on results of Ferenčaković et al.(Ferenčaković, Sölkner et al. 2013), indicating that shorter runs include some wrong calls with the density of the 50k SNP chip used in this study.

Relationships between seven herds

Wright's F-statistics is an important tool to provide insights into evolutionary process that influence the genetic variation between subpopulations. In this study a relatively high density of markers was applied. The genetic differentiation analysis showed that the Gyr cattle populations were genetically distinct. On average, the genetic differentiation (F_{ST}) among different subpopulations was 0.050, a highly significant ($p < 0.001$) value, which indicates that there was a great differentiation and a relatively low gene flow among the 7 herds included. It is also clear that most of the genetic variation (95%) is inter-individual and 5% of the total variation is due to isolated subpopulation differences. Estimates of pairwise genetic differentiation based on the infinitesimal model (F_{ST}) were all significant after Bonferroni corrections ($p <0.05$) except for the pair of Herd_277-Herd_750 and Herd_750-Herd_24426, indicating that most of the subpopulations can be considered as genetically separate entities. The estimated number of migrants per generation (*Nm*) between populations was generally low (3.49) compared to the value among other breeds (Mart ńez, Gama et al. 2012). The estimated number of migrants, i.e., the number of individuals exchanged between populations per generation that would balance the diversifying effect of genetic drift. The low value *Nm* indicates that the level of genetic material exchange is low, although this level of migration is expected to maintain the genetic differentiation observed between the subpopulations. It is also possible that the bulls are acquired from specific subpopulations.

Similar magnitude of genetic differentiation has been reported in Zebu cattle between different breeds. The result is in accordance with relevant study in diversity analysis of Brazilian cattle breeds which confirmed that the average F_{ST} value was 0.0496 between zebu breeds in Brazil (Egito, Paiva et al. 2007). However the value of overall genetic differentiation (F_{ST}) among Gyr cattle subpopulations is higher than the former population structure analyses of Brazilian Gyr cattle ($F_{ST} = 0.023$) (Reis Filho, Lopes et al. 2010). It was predicted that the F_{ST} value between Gyr breed and Nelore breed was 0.0475 (O'Brien,

Höller et al. 2015). The result we obtained was even higher compared to the value obtained between different cattle breeds. Currently there are four types of farms in Gyr cattle breeding system in Brazil : 1. multiplier herds which use both their own sires and sires from outside. In the mean time, they also sell animals; 2. commercial herds which use both their both their own sires and sires from outside. However all the animals are kept in the farm instead of selling them; 3. nucleus farm which don't use the sires from outside but sell animals; 4. isolated herds which have no connection with any other farms (Reis Filho, Lopes et al. 2010). In this study, the information about type of farms of our targeted herds was not available. Thus structuring of populations may occur through subpopulations with limited migration and gene flow, resulting in completely isolated subpopulations that aggravate the problems of conservation of genetic groups.

The highest pairwise F_{ST} value appeared between Herd_265 and Herd_551 with a value of 0.077. For this pair of herds the geographic distance is also the greatest in this study. Although the result obtained from the Mantel test indicated that the correlation between genetic distance and geographic distance is not significant, this positive correlation was also predicted via the phylogeny tree and principle variance component analysis.

However there are still three limitations in this current study. In current study, the degree of spatial dispersion is still low. Although number of targeted animals is large (1613 cows), only 36% of the animals (588) have access to complete GPS information. The cows considered in this study were only spread in seven herds located in two states of Brazil. Compared to the spatial distribution of Gyr cattle, it is hard to predict whether the result will be representative to the current situation of the whole country.

Another limitation for this study is the restriction of sample size per herd. For example in this study, Herd_750 only has five animals. It is hard to predict whether the genetic differentiation of these five animals can represent the genetic differentiation of the whole herd.

Ascertainment bias is a further issue that needs to be considered when using SNPs for population genetic analyses as it may introduce a systematic bias in estimates of variation

within and between populations. Ascertainment bias is the systematic deviation from the expected allele frequency distribution that occurs because of the sampling processes used to select marker loci (Nielsen 2000). Because of the small size of the ascertainment panel, the possibility that this SNP is identified is a function of its minor allele frequency in this panel. Thus it may influence the statistical measures which rely on allele frequency, such as estimating nucleotide diversity, population structure etc (Helyar, Hemmer‐Hansen et al. 2011). Applying a more robust method such as genotype by sequencing technique can provide a potential approach for correcting ascertainment bias (De Donato, Peters et al. 2013).

Conclusions and Suggestions for Future Work

This is the first study related to genetic and geographic diversity of Gyr cattle in Brazil. Judging from all the parameters for evaluating genetic diversity, the current study results showed a higher genetic diversity than former research results in Gyr cattle. The overall expected heterozygosity of total population was 0.264 ± 0.158. The average of genetic differentiation among all herds measured as F_{ST} value was 0.050 ± 0.041. F_{IT} value for the whole population was 0.019 ± 0.058 while the F_{IS} value for the whole population was -0.031 ± 0.047, showing an excess of heterozygotes (negative value). However the genome wide level of inbreeding based on proportions of the genome being in runs of homozygosity was 0.0537, which was higher than the values obtained based on allele frequency or pedigree information, has confirmed to be a more realistic reflection of inbreeding level. The assessment of the effective number of migrants by generation (*Nm*) in Gyr populations was 3.49. The largest geographic distance appeared between Herd_265 and Herd_551 with a spatial distance of 532 km. As expected, F_{ST} and Nei's genetic distance were also greatest between Herd_265 and Herd_551 with average value of 0.077and 0.022 respectively. The application of Rousset's isolation by distance method provided the linear regression of genetic distance on geographical distance: $\frac{F_{ST}}{1-F_{ST}} = -0.035 + 0.014 \ln(d)$ with 29.89% variance explained by the model. The Mantel test indicated that a potential positive correlation between genetic distance and geographic distance with a considerable trend toward significance (r=0.624, P=0.068). In the PCA analysis, the first principle component （PC）accounted for 4.895% of the total variance and the second principle component condensed 2.526% of the variation.

The genetic differentiation analysis showed that the Gyr cattle populations considered in this study were genetically distinct. However in order to achieve a more representative results, an increased number of animals per herd with complete GPS information is needed, preferably from geographically more distinct areas. Besides, applying a more robust method

such as applying genotype by sequencing technique could be a potential approach for correcting ascertainment bias.

Acknowledgement

I would like to thank for EMBRAPA, Brazil for offering relevant information and data for this study. My deepest gratitude goes first and foremost to my supervisors, Prof. Johann, Sölkner, Dr. Gábor Mészáros and Dr. Marcos Vinicius Gualberto Barbosa da Silva for their patience, guidance and unconditional support. From them I learnt how to perform real research and how to write in a scientific way. I would never accomplish my master thesis without their help. I would like to thank them very much for their support and understanding for my situation.

Most importantly, I would like to express my thanks to my family and friends. For my parents, who give me unconditional support, encourage and love, I am forever grateful. When I meet difficulties, they always stand by me and give me maximum support. I would also like to express my appreciation to all my colleagues of Animal Breeding and Genetics group, who offers help every time when I ran into all kinds of problems.

It is my honor to study in University of Natural Resources and Life Sciences, Vienna and to do my master thesis in the Animal Breeding and Genetics Group. All the experience here will stay with me in my whole life.

References

1. Azevedo, A. (2010). "Molecular characterization of Brazilian Dairy Gyr using microsatellite markers."
2. Bao, W., J. Shu, X. Wu, H. Musa, C. Ji and G. Chen (2009). "Genetic diversity and relationship between genetic distance and geographical distance in 14 Chinese indigenous chicken breeds and red jungle fowl." Czech Journal of Animal Science **54**(2): 74-83.
3. Buchmann, R. and S. Hazelhurst (2014). "The 'Genesis' Manual."
4. Chang, C. C., C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell and J. J. Lee (2014). "Second-generation PLINK: rising to the challenge of larger and richer datasets." arXiv preprint arXiv:1410.4803.
5. Charlesworth, B., M. Nordborg and D. Charlesworth (1997). "The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations." Genetical research **70**(02): 155-174.
6. Curik, I., M. Ferenčaković and J. Sölkner (2014). "Inbreeding and runs of homozygosity: a possible solution to an old problem." Livestock Science **166**: 26-34.
7. De Donato, M., S. O. Peters, S. E. Mitchell, T. Hussain and I. G. Imumorin (2013). "Genotyping-by-sequencing (GBS): a novel, efficient and cost-effective genotyping method for cattle using next-generation sequencing."
8. Egito, A. A., S. R. Paiva, M. Maria do Socorro, A. S. Mariante, L. D. Almeida, S. R. Castro and D. Grattapaglia (2007). "Microsatellite based genetic diversity and relationships among ten Creole and commercial cattle breeds raised in Brazil." BMC genetics **8**(1): 83.
9. Excoffier, L. and H. E. 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**(3): 564-567.
10. Felius, M. (1995). Cattle breeds: an encyclopedia, C Misset bv.
11. Ferenčaković, M., E. Hamzić, B. Gredler, T. R. Solberg, G. Klemetsdal, I. Curik and J. Sölkner (2013). "Estimates of autozygosity derived from runs of homozygosity: empirical evidence from selected cattle populations." Journal of Animal Breeding and Genetics **130**(4): 286-293.

12. Ferenčaković, M., J. Sölkner and I. Curik (2013). "Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors." Genetics, selection, evolution: GSE **45**(1): 42.

13. Gaur, G., S. Kaushik and R. Garg (2003). "The Gir cattle breed of India-characteristics and present status." Animal Genetic Resources Information **33**: 21-29.

14. Gautier, M., T. Faraut, K. Moazami-Goudarzi, V. Navratil, M. Foglio, C. Grohs, A. Boland, J.-G. Garnier, D. Boichard and G. M. Lathrop (2007). "Genetic and haplotypic structure in 14 European and African cattle breeds." Genetics **177**(2): 1059-1070.

15. Gautier, M., D. Laloë and K. Moazami-Goudarzi (2010). "Insights into the genetic history of French cattle from dense SNP data on 47 worldwide breeds." PLoS One **5**(9): e13038.

16. Goossens, B., L. Chikhi, M. Jalil, M. Ancrenaz, I. LACKMAN-ANCRENAZ, M. Mohamed, P. Andau and M. W. Bruford (2005). "Patterns of genetic diversity and migration in increasingly fragmented and declining orang-utan (Pongo pygmaeus) populations from Sabah, Malaysia." Molecular Ecology **14**(2): 441-456.

17. Hardy, O. J. and X. Vekemans (2002). "SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels." Molecular ecology notes **2**(4): 618-620.

18. Helyar, S. J., J. Hemmer-Hansen, D. Bekkevold, M. Taylor, R. Ogden, M. Limborg, A. Cariani, G. Maes, E. Diopere and G. Carvalho (2011). "Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges." Molecular Ecology Resources **11**(s1): 123-136.

19. Hijmans, R. J., E. Williams, C. Vennes and M. R. J. Hijmans (2014). "Package 'geosphere'."

20. Jordana, J., P. Alexandrino, A. Beja-Pereira, I. Bessa, J. Cañon, Y. Carretero, S. Dunner, D. Laloë, K. Moazami-Goudarzi and A. Sanchez (2003). "Genetic structure of eighteen local south European beef cattle breeds by comparative F-statistics analysis." Journal of Animal Breeding and Genetics **120**(2): 73-87.

21. Kale, D., D. Rank, C. Joshi, B. Yadav, P. Koringa, K. Thakkar, T. Tolenkhomba, J. Solanki, S. George and P. Ponnachan (2010). "Genetic diversity among Indian Gir, Deoni and Kankrej cattle breeds based on microsatellite markers." Indian Journal of Biotechnology **9**: 126-130.

22. Kimura, M. and G. H. Weiss (1964). "The stepping stone model of population structure and the decrease of genetic correlation with distance." Genetics **49**(4): 561.

23. Kugonza, D. R., H. Jianlin, M. Nabasirye, D. Mpairwe, G. H. Kiwuwa, A. M. Okeyo and O. Hanotte (2011). "Genetic diversity and differentiation of Ankole cattle populations in Uganda inferred from microsatellite data." Livestock Science **135**(2): 140-147.

24. Lacy, R. C. (1987). "Loss of genetic diversity from managed populations: interacting effects of drift, mutation, immigration, selection, and population subdivision." Conservation Biology **1**(2): 143-158.

25. Lao, O., T. T. Lu, M. Nothnagel, O. Junge, S. Freitag-Wolf, A. Caliebe, M. Balascakova, J. Bertranpetit, L. A. Bindoff and D. Comas (2008). "Correlation between genetic and geographic structure in Europe." Current Biology **18**(16): 1241-1248.

26. Leberg, P. (2002). "Estimating allelic richness: effects of sample size and bottlenecks." Molecular ecology **11**(11): 2445-2449.

27. Machado, M. A., I. Schuster, M. L. Martinez and A. L. Campos (2003). "Genetic diversity of four cattle breeds using microsatellite markers." Revista Brasileira de Zootecnia **32**(1): 93-98.

28. Madalena, F. (1988). "A note on the effect of variation of lactation length on the efficiency of tropical cattle selection for milk yield." Theoretical and applied genetics **76**(6): 830-834.

29. Mart ínez, A. M., L. T. Gama, J. Ca ñón, C. Ginja, J. V. Delgado, S. Dunner, V. Landi, I. Mart ín-Burriel, M. C. T. Penedo and C. Rodellar (2012). "Genetic footprints of Iberian cattle in America 500 years after the arrival of Columbus."

30. McQuillan, R., A.-L. Leutenegger, R. Abdel-Rahman, C. S. Franklin, M. Pericic, L. Barac-Lauc, N. Smolej-Narancic, B. Janicijevic, O. Polasek and A. Tenesa (2008). "Runs of homozygosity in European populations." The American Journal of Human Genetics **83**(3): 359-372.

31. Nei, M. (1972). "Genetic distance between populations." American naturalist: 283-292.

32. Nielsen, R. (2000). "Estimation of population parameters and recombination rates from single nucleotide polymorphisms." Genetics **154**(2): 931-942.

33. O'Brien, A. M. P., D. Höller, S. A. Boison, M. Milanesi, L. Bomba, Y. T. Utsunomiya, R. Carvalheiro, H. H. Neves, M. V. da Silva and C. P. VanTassell (2015). "Low levels of taurine introgression in the current Brazilian Nelore and Gir indicine cattle populations."

Genetics Selection Evolution **47**(1): 1-7.

34. Oksanen, J., F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens and H. Wagner (2013). "Package 'vegan'." Community ecology package, version **2**(9).

35. Pariset, L., M. Mariotti, M. Gargani, S. Joost, R. Negrini, T. Perez, M. Bruford, P. Ajmone Marsan and A. Valentini (2011). "Genetic diversity of sheep breeds from Albania, Greece, and Italy assessed by mitochondrial DNA and nuclear polymorphisms (SNPs)." The Scientific World Journal **11**: 1641-1659.

36. Pham, L. D., D. N. Do, N. T. Binh, N. Van Ba, T. T. T. Thuy, T. X. Hoan, V. C. Cuong and H. N. Kadarmideen (2013). "Assessment of genetic diversity and population structure of Vietnamese indigenous cattle populations by microsatellites." Livestock Science **155**(1): 17-22.

37. Porto-Neto, L. R., T. S. Sonstegard, G. E. Liu, D. M. Bickhart, M. V. Da Silva, M. A. Machado, Y. T. Utsunomiya, J. F. Garcia, C. Gondro and C. P. Van Tassell (2013). "Genomic divergence of zebu and taurine cattle identified through high-density SNP genotyping." BMC genomics **14**(1): 876.

38. Queiroz, S. and R. Lôbo (1993). "Genetic relationship, inbreeding and generation interval in registered Gir cattle in Brazil." Journal of Animal Breeding and Genetics **110**(1-6): 228-233.

39. Ramachandran, S., O. Deshpande, C. C. Roseman, N. A. Rosenberg, M. W. Feldman and L. L. Cavalli-Sforza (2005). "Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa." Proceedings of the National Academy of Sciences of the United States of America **102**(44): 15942-15947.

40. Reis Filho, J. C., P. S. Lopes, R. d. S. Verneque, R. d. A. Torres, R. L. Teodoro and P. L. S. Carneiro (2010). "Population structure of Brazilian Gyr dairy cattle." Revista Brasileira de Zootecnia **39**(12): 2640-2645.

41. Revelle, W. (2014). "psych: Procedures for personality and psychological research." Northwestern University, Evanston. R package version **1**(1).

42. Rousset, F. (1997). "Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance." Genetics **145**(4): 1219-1228.

43. Saitou, N. and M. Nei (1987). "The neighbor-joining method: a new method for

reconstructing phylogenetic trees." Molecular biology and evolution **4**(4): 406-425.

44. Santana, M., R. Pereira, A. Bignardi, L. El Faro, H. Tonhati and L. Albuquerque (2014). "History, structure, and genetic diversity of Brazilian Gir cattle." Livestock Science **163**: 26-33.

45. Shumaker, B. and R. Sinnott (1984). "Astronomical computing: 1. Computing under the open sky. 2. Virtues of the haversine." Sky and telescope **68**: 158-159.

46. Smouse, P. E., J. C. Long and R. R. Sokal (1986). "Multiple regression and correlation extensions of the Mantel test of matrix correspondence." Systematic zoology: 627-632.

47. Su, G. and P. Madsen (2011). "User's Guide for Gmatrix. A program for computing Genomic relationship matrix." University of Aarhus, Tjele, Denmark.

48. Takezaki, N., M. Nei and K. Tamura (2010). "POPTREE2: Software for constructing population trees from allele frequency data and computing other population statistics with Windows interface." Molecular biology and evolution **27**(4): 747-752.

49. Team, R. C. (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2012, ISBN 3-900051-07-0.

50. Torres-Júnior, J. d. S., M. de FA Pires, W. De Sa, A. d. M. Ferreira, J. H. M. Viana, L. S. d. A. Camargo, A. d. A. Ramos, I. Folhadella, J. Polisseni and C. De Freitas (2008). "Effect of maternal heat-stress on follicular growth and oocyte competence in Bos indicus cattle." Theriogenology **69**(2): 155-166.

51. Vignal, A., D. Milan, M. SanCristobal and A. Eggen (2002). "A review on SNP and other types of molecular markers and their use in animal genetics." Genetics Selection Evolution **34**(3): 275-306.

52. Villalobos-Cortés, A., A. Martínez, J. L. Vega-Pla, V. Landi, J. Quiroz, J. R. Marques and J. V. Delgado (2015). "Genetic relationships among five zebu breeds naturalized in America accessed with molecular markers." Italian Journal of Animal Science **14**(2).

53. Waples, R. S. and O. Gaggiotti (2006). "INVITED REVIEW: What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity." Molecular ecology **15**(6): 1419-1439.

54. Weir, B. S. and C. C. Cockerham (1984). "Estimating F-statistics for the analysis of population structure." evolution: 1358-1370.

55. Wright, S. (1943). "Isolation by distance." Genetics **28**(2): 114.