Population structure of Brazilian Gyr dairy cattle

João Cruz Reis Filho, Paulo Sávio Lopes, Rui da Silva Verneque, Robledo de Almeida Torres, Roberto Luiz Teodoro, Paulo Luiz Souza Carneiro

1 Ministério da Agricultura, Pecuária e Abastecimento, Brasília, DF.
2 Universidade Federal de Viçosa (UFV), Departamento de Zootecnia, Viçosa, MG, bolsista do CNPq.
3 Embrapa Gado de Leite, Juiz de Fora, MG.
4 Universidade Estadual do Sudoeste da Bahia (UESB), Departamento de Ciências Biológicas, Jequié, BA.

ABSTRACT - The objective of the present study was to evaluate the genetic structure of Gyr cattle selected for milk production. Files of pedigree and production were composed of 27,610 animals. The ENDOG program was used for the calculation of individual inbreeding coefficient (F) and coefficient of average relatedness (AR), effective number of animals (Ne), effective number of founders (fe) and ancestors (fa), and generation interval (GI). Individual inbreeding coefficients and average relatedness in the population were 2.82% and 2.10%, respectively. It was observed a reduction in the effective number of animals, especially after publication of the results of the first progeny test. The estimated effective number of founders was 146 and 75 for the ancestrals. Out of those, only 28 ancestors accounted for the origin of 50% of the population genes. The average generation interval was 8.41 years and it was longer for males than for females. For maintaining genetic variability in future generations, it should be invested mating strategies that reduce inbreeding and which do not use massively only some high breeding value sires.

Key Words: effective size, generation interval, inbreeding, progeny test

Introduction

Considering that pedigree data continue to be important, numerous studies evaluating inbreeding, effective population size, genetic diversity and other important population parameters are available (Cleveland et al., 2005; Parland et al., 2007; Carolino & Gama, 2008; Malhado et al., 2008; Martinez et al., 2008). Some simple population parameters that depend mainly on management and mating systems have an important impact on the genetic variability of a population. Therefore, studies on population structure may elucidate important factors that affect the genetic evolution of populations (Valera et al., 2005).

Gyr dairy cattle is a subpopulation of Gyr breed selected for milk production. Gyr Dairy breed is found in more than 80% of the Brazilian dairy herds (Santos, 1994), either as pure animals or as crosses with Holstein animals and it is used mainly because of its rusticity, adaptability and resistance to endo- and ectoparasites, especially in grazing systems. The Programa Nacional de Melhoramento do Gir Leiteiro was implemented in Brazil in 1985 and...
since that bull progeny tests have been performed (Martinez et al., 2006; Neves et al., 2009).

Concerns regarding inbreeding, genetic structure and diversity of breeds and populations under selection have increased. (Nomura et al., 2001; Weigel & Lin, 2002; Gutiérrez et al., 2003; Chikhi et al., 2004; Hagger, 2005). Unfavorable effects of inbreeding on milk yield and lactation length (Queiroz et al., 1993), and on weaning and yearling weights (Queiroz et al., 2000) were reported for Gyr cattle. Also, it was observed, from 1979 to 1988 (Faria, 2002), an effective size decrease of registered Gyr population and an increase in the total inbreeding rate.

The massive use of a small number of proven bulls with high breeding value may affect long-term genetic gain by reducing genetic variability. Because of this, the objective of the present study was to analyze the genetic structure of the Brazilian Gyr Dairy cattle and to provide information for decisions concerned to breeding.

**Material and Methods**

The historical database of the Gyr breed, which is consisted of data from the progeny test of the Programa Nacional de Melhoramento do Gir Leiteiro and the Arquivo Zootécnico Nacional, maintained by the Associação Brasileira dos Criadores de Zebu, was used in present study.

Production, reproduction and pedigree data were available; however, records from crossbred animals were not used. Unregistered Gyr animals data were included to complement the pedigree information. Therefore, the complete file comprised 27,610 animals, in which data from one or both parents were unknown for 3,709 animals and the reference population comprehended 23,901 animals (both parents known). Pedigree data from animals born from 1966 to 2003 were used.

The complete pedigree file was used for analysis of the genetic structure of the population using the ENDOG v. 4.5 program (Gutiérrez & Goyache, 2005). The ENDOG program uses the algorithm proposed by Meuwissen & Luo (1992) for computation of the inbreeding coefficient (F). The relative increase in inbreeding per generation ($\Delta F$) was calculated by using the formula (Gutiérrez & Goyache, 2005):

$$\Delta F = \frac{F_t - F_{t-1}}{1 - F_{t-1}}$$

where, $F_t$ = average inbreeding coefficient estimated for the current generation; $F_{t-1}$ = average inbreeding coefficient estimated for the previous generation. The effective population size ($N_e$) was calculated as $N_e = \frac{1}{2\Delta F}$.

The ENDOG program also provides three additional values of $N_e$ by computing the regression coefficient (b) of the individual inbreeding coefficient over: i) the number of full generations traced; ii) the maximum number of generations traced and iii) the equivalent complete generations. Considering the corresponding regression coefficient as the increase in inbreeding between two generations ($F_t - F_{t-1} = b$), it is used:

$$N_e = \frac{1}{2b}$$

The number of traced full generations (g) is defined as the furthest generation in which all the ancestors are known, i.e., the generation that separates the progeny from the most distant generation in which 2$^g$ ancestors of the individual are known. The maximal number of traced generations is the number of generations separating the individual from its furthest ancestor. The equivalent complete generation is computed as the sum over all known ancestors of the terms computed as the sum of $(1/2)^t$, where $t$ is the number separating the individual from each known ancestor (Maingel et al., 1996).

The average relatedness (AR) was computed using an algorithm to obtain $c'$. Each element of $c'$ is the AR of the respective animal, defined as:

$$c' = \left(\frac{1}{n}\right)A$$

in which $A$ = the numerator relationship matrix (Henderson, 1976; Quaas, 1976), $1$ = vector of ones (1 x n), and $n$ = the number of animals. The average relatedness (AR) coefficient of each individual represents the probability of an allele randomly chosen from the whole population in the pedigree that belongs to that individual. The AR of a founder indicates the probability of this founder really be an animal from the base population. Therefore, AR is an alternative or complement to F, and it can be used to predict the long-term inbreeding of a population because it takes into account the percentage of the complete pedigree originated from a given animal (Gutiérrez & Goyache, 2005).

Animals imported from India and from unknown parents were considered to be the founders. The effective number of founders ($f_e$) was calculated using the equation:

$$f_e = \frac{1}{\sum_{k} q_k^2}$$

in which $q_k$ = the probability of a random allele in the population be inherited from founder $k$, $q_k$ = the AR of founder $k$ (Gutiérrez & Goyache, 2005), $f_e$ = the number of animals that is expected to produce the same genetic diversity as observed in the studied population (Lacy, 1989)
and it is also an estimate on how genetic diversity was preserved over time.

The effective number of ancestors was computed as:

$$f_a = \frac{1}{\sum q_j}$$

in which $q_j$ = the marginal contribution of an ancestor $j$ (not necessarily a founder), i.e., the genetic contribution of an ancestor that is not explained by an ancestor chosen before; $f_a$ = the minimal number of ancestors, not necessarily founders, explaining the complete genetic diversity of a population (Boichard et al., 1997). This parameter complements the information provided by the effective number of founders by accounting for the losses of genetic variability produced by the intense use of few individual reproducers (Gutiérrez & Goyache, 2005).

The generation interval, defined as the average age of parents on the birth of their offspring, was also calculated using the ENDOG program. This parameter was computed for the four pathways (sire-son, sire-daughter, dam-son and dam-daughter), including only animals with known birth dates after 1966. In addition to the generation intervals estimated with the ENDOG program, it was included two generation intervals, the average between sire and dam to son or to daughter, which are more useful to calculate genetic gains.

Wright’s F coefficients (Wright, 1978), Fit, Fst and Fis, were calculated (Gutiérrez & Goyache, 2005). Fit is the average inbreeding coefficient of the population, Fst corresponds to the expected average inbreeding coefficient if sires from each generation were randomly mated and it was estimated for hypothetical populations produced by random mating, and Fis expresses the deviation from causality obtained in current matings.

Results and Discussion

The complete relationship matrix included data from 27,610 animals; an average inbreeding coefficient of 2.82% and AR coefficient of 2.10% were observed (Figure 1). Although inbreeding could be considered to be low in the studied population, inbred animals accounted for more than 60% of all animals (16,687 animals), with an average inbreeding coefficient of 4.66%, indicating that concern regarding to inbreeding in the next generations is necessary. The percentage of inbred animals increased over time but the average inbreeding coefficient did not follow this trend and it remained low. The results of the bull progeny test performed within the Programa Nacional de Melhoramento do Gir Leiteiro provide important information that helps breeders to plan matings aiming at genetic gain without neglecting the relatedness among animals (Martinez et al., 2006).

The largest number of complete generations found in the present study was five, corresponding to the complete knowledge of its 32 closest ancestors ($2^5$). The average inbreeding coefficient, AR coefficient and percentage of inbred animals increased consistently over the generations (Table 1). These results are partially due to the knowledge that more complete generations of one individual increases the possibility that an important ancestor will appear several times in the pedigree.

According to Gutiérrez & Goyache (2005), the advantages of using the AR coefficient are that: i) the computational cost to calculate AR coefficients is similar to the computational cost of the numerator relationship matrix because both procedures use common algorithms; ii) AR coefficients can also be used as a measure of inbreeding in the whole population since they take into account both inbreeding and coancestry coefficients, and iii) AR can be used as an index to maintain the initial genetic stock, selecting animals with the lowest AR value as breeding animals. In addition, AR can be used to compute the effective size ($f_a$) of the founder population as the inverse of the sum of the square AR coefficients across founder animals.

The differences in the observed effective population size ($N_e$) were probably caused by the variation in the average inbreeding coefficient across generations because $N_e$ was calculated as a function of the increase in inbreeding ($\Delta F$, Table 1). Similar results have been reported in the study of Faria (2002), in which $N_e$ ranged from 41 to 99 for the registered Gyr breed using only data from animals for which a genealogical record could be obtained from the Associação Brasileira dos Criadores de Zebu.

Figure 1 - Percentage of inbred cows among the total number of cows and respective average inbreeding coefficient according to year of the birth of the cow.
Ne was calculated by regression of the individual inbreeding coefficient over the number of different types of traced generations (Table 2). When the available information is scarce, these estimates can be useful to approximate the upper, lower and the real limits of Ne in the studied population (Gutiérrez & Goyache, 2005). The maximal number of generations separating one individual from its most distant ancestor was 22.

Ne increased from 1985 to 1994-2002 (Figure 2) when it reached its highest value (about 200). This trend reflects the history of the breed. Before the implementation of the Programa Nacional de Melhoramento do Gir Leiteiro in 1985, selection was performed exclusively using information within herds and the number of animals was very small. After the establishment of the program, exchange of genetic material markedly increased since the design of the progeny test itself encourages the simultaneous use of different sires in various herds to permit a more accurate prediction of breeding values. After publication of the progeny test results in 1992, the best sires were used intensively which decreased Ne.

The reference population for which both parents are known contained only 23,901 animals, with one or both parents being unknown for 3,709 animals. The effective number of founders (fe) was 146 and the expected increase in inbreeding, caused by the unbalanced contribution of founders, was 0.34%.

The number of ancestors contributing to the reference population was 2,872, with only 28 ancestors accounting for 50% of the population genes. The effective number of ancestors (fa) was 75. The fa parameter is always equal to or lower than fe since it takes into account the "bottleneck effect" in the pedigree which, in turn, is the main cause of the loss of alleles in populations such as dairy cattle. Studying the same breed, Faria (2002) observed a decrease in fe and fa over time (284 and 211, respectively), but the obtained values were still higher than the estimates observed in the present study. Several ancestors and founders of the Gyr breed did not contribute to the dairy population.

The fe/fa ratio of 1.95, which expresses the "bottleneck effect", was intermediate between the ratios 1.08 to 1.34 obtained by Faria (2002) for the Gyr breed and the ratio of 3.0 was reported by Boichard et al. (1997) for Normande cattle. Overestimation of the effective number of founders (fe) is particularly serious for intensive selection programs when the germplasm of a limited number of animals is widely distributed, for example, by artificial insemination (Boichard et al., 1997).

The average age of the parents on the birth of their offspring was high and similar to average estimates reported for other Zebu herds in Brazil (Table 3): 6.56 years of age for Gyr Mocho (Faria et al., 2001); 7.09 years of age for Nelore Mocho (Faria et al., 2002); 7.17 years of age for Tabapuá (Vercesi Filho et al., 2002); 8.02 years of age for the Gyr breed (Queiroz & Lobo, 1993), and 8.3 years of age for Nelore cattle in the State of Bahia (Malhado et al., 2008). These high average ages reflect the selection process applied in these breeds.}

### Table 1 - Inbreeding, relatedness coefficient and effective population size for full generations traced

<table>
<thead>
<tr>
<th>Generation</th>
<th>Number of individuals</th>
<th>Average inbreeding coefficient (%)</th>
<th>% Inbred</th>
<th>F (_{\text{inbred}}) (%)</th>
<th>Average relatedness coefficient (%)</th>
<th>Effective population size</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>3709</td>
<td>0.00</td>
<td>0.05</td>
<td>0.00</td>
<td>0.17</td>
<td>100.80</td>
</tr>
<tr>
<td>1</td>
<td>5297</td>
<td>0.50</td>
<td>8.38</td>
<td>5.91</td>
<td>1.06</td>
<td>16.10</td>
</tr>
<tr>
<td>2</td>
<td>7102</td>
<td>3.57</td>
<td>73.01</td>
<td>4.89</td>
<td>2.22</td>
<td>125.30</td>
</tr>
<tr>
<td>3</td>
<td>8176</td>
<td>3.95</td>
<td>96.04</td>
<td>4.12</td>
<td>2.89</td>
<td>36.60</td>
</tr>
<tr>
<td>4</td>
<td>3234</td>
<td>5.26</td>
<td>99.88</td>
<td>5.27</td>
<td>3.67</td>
<td>80.20</td>
</tr>
<tr>
<td>5</td>
<td>92</td>
<td>5.86</td>
<td>100.00</td>
<td>5.86</td>
<td>3.91</td>
<td>80.20</td>
</tr>
</tbody>
</table>

% Inbred = percentage of inbred individuals, $F_{\text{inbred}}$ = average inbreeding coefficient of inbred individuals.

### Table 2 - Average number of generations, inbreeding increase ($\Delta F$) and effective size (Ne) for different types of generations

<table>
<thead>
<tr>
<th>Type of generation</th>
<th>Average</th>
<th>$\Delta F$</th>
<th>Ne</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete(^1)</td>
<td>2.08</td>
<td>1.14%</td>
<td>35.42</td>
</tr>
<tr>
<td>Maximum(^2)</td>
<td>8.89</td>
<td>0.16%</td>
<td>308.30</td>
</tr>
<tr>
<td>Equivalent(^3)</td>
<td>3.86</td>
<td>0.74%</td>
<td>67.38</td>
</tr>
</tbody>
</table>

\(^1\) Furthest generation in which all ancestors are known.
\(^2\) Number of generations separating the individual from its furthest ancestor.
\(^3\) Sum over all known ancestors of the terms computed as the sum of (1/2)\(^t\), where \(t\) is the number of generations separating the individual from each known ancestor.
estimates can be partially attributed to the longevity of the Gyr breed, with frequent cases of animals that are maintained for reproduction until 20 years of age.

Using information on cows and bulls from Gyr dairy cattle, Queiroz et al. (1991) estimated generation intervals and obtained values similar to the ones reported here, indicating that few changes were observed in the generation interval of this breed in the last decades (dam-daughter = 8.20; dam-son = 8.86; sire-daughter = 7.36; sire-son = 8.78; general average = 8.02). The authors suggested that bulls should be used for a smaller period of time to reduce the generation interval.

Long generation intervals reduce annual genetic gains for selected traits, with a consequent lower economical return for the breeding program. Razook et al., (1993) reported short generation intervals for Nelore and Guzerá experimental herds in Brazil (sire-son interval of 3.66 and 3.74 years, respectively). The age of the sire on the birth of its the son was higher than that observed for the other paths (sire-daughter, dam-son and dam-daughter). This can be explained by the continuous use of certain bulls without replacement, or by the old age of sires when their first test is made public. Gyr selection programs have been encouraging the use of young bulls in progeny tests. Madalena et al. (1996) alerted to the fact that the time interval between the communication of the breeding value of the proven bulls submitted to the Gyr Dairy cattle progeny test and their use may increase the generation interval.

Estimates of Wright’s F statistics obtained for the Gyr population were $F_{it} = 0.0188$, $F_{st} = 0.023$ and $F_{is} = -0.0057$. These coefficients measure causality deviation in matings, and values which are negative and close to zero indicate a slight predominance of matings between subpopulations. When $F_{is}$ is lower than zero, mating across subpopulations should be avoided. The negative value of $F_{is}$ observed in the present study can be attributed to the use of within-herd sires because 61.7% of the 26 analyzed herds used sires from their stock.

The low value of $F_{st}$ indicates the lack of structure of the population (Table 4). All herds used outside bulls and/or artificial insemination and none of them was classified as a nucleus (herds that do not use outside sires but sell animals) neither as a herd without connection (isolated). Most herds (26, corresponding to 96.3%) were classified as multiplier herds (herds that use outside or their own sires and sell animals) and only one (3.7%) was classified as commercial (herds that use outside or their own sires and do not sell animals for reproduction). The percentage of outside bulls used in the herds classified as multiplier was 38.8% and the only herd classified as commercial used 100% outside bulls. The use of outside sires, especially through artificial insemination, explains the lack of subdivision or structure of the population. 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. Structured subpopulations are not common in dairy or beef cattle.

According to Carneiro et al. (2007), genetic drift is one of the main factors that affect genetic gains in populations under selection. The small effective population size may cause variations in genetic gain, especially using BLUP as predicted breeding values. To maximize genetic gain in Gyr dairy cattle, it is recommended to increase the population size, to control inbreeding through guided matings, to reduce generation interval, and to avoid intensive use of few bulls.

### Table 4 - Classification of Gyr Dairy herds according to origin and use of bulls

<table>
<thead>
<tr>
<th>Type of herd</th>
<th>Use outside bulls</th>
<th>Use their own bulls</th>
<th>Sell bulls</th>
<th>Number of herds</th>
<th>Percentage of outside bulls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nucleus</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Multiplier</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>26</td>
<td>38.30</td>
</tr>
<tr>
<td>Multiplier</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>Commercial</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>Commercial</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>1</td>
<td>100</td>
</tr>
<tr>
<td>Isolated</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

These coefficients measure causality deviation in matings, and values which are negative and close to zero indicate a slight predominance of matings between subpopulations.
Conclusions

Inbreeding coefficients breed and effective population size of Gyr Dairy Cattle should be considered in mating decisions. Breeding strategies that minimize inbreeding and do not rely on the massive use of few sires with high breeding values are suggested to maintain genetic variability in future generations. In addition, long generation intervals should be reduced to maximize genetic progress.

Acknowledgements

The authors wish to thank Embrapa Gado de Leite and Associação Brasileira dos Criadores de Gir Leiteiro (ABCGIL), for providing the data for this study. The authors also wish to thank Dr. Mário Luiz Martinez (in memoriam) for his contribution.

References


