

Amostragem de indivíduos representativos em uma população de animais da raça Canchim aparentados e genotipados

Ismael Urbinati¹

Roberto Hiroshi Higa²

Fabiana Barichello Mokry³

Amostragem consiste em estudar um pequeno grupo, designado amostra, retirado de uma população, de modo que a amostra represente toda a população. Quando a população possui características que permitem a criação de subconjuntos, as amostras extraídas por amostragem simples são menos representativas. Nesse caso, é utilizada a amostragem estratificada (BARBETTA, 2007).

Os estratos podem ser delimitados pela análise de agrupamento, processo pelo qual se agrupa um conjunto de objetos físicos ou abstratos em classes de objetos similares (EVERITT et al., 2001).

O objetivo neste estudo foi avaliar diferentes métodos para selecionar uma amostra de uma população contendo animais aparentados, representando estratos de famílias, de forma que os animais selecionados representem toda a população.

Utilizaram-se duas fontes de informação para o estudo. A primeira foi o arquivo de pedigree (animal, pai e mãe) constituído de 6.801 animais da raça Canchim, sendo 400 animais genotipados. A segunda foi a matriz de parentesco constituída pelos coeficientes de parentesco dos 400 animais genotipados, organizados na forma de matriz, com os animais nas linhas e colunas e os correspondentes coeficientes de parentesco dentro da matriz (BOURDON, 2000).

¹ Universidade Estadual Paulista “Júlio de Mesquita Filho”,
ismaelurbinati@zootecnista.com.br

² Embrapa Informática Agropecuária, roberto@cnptia.embrapa.br

³ Universidade Federal de São Carlos, bolsista Capes, fbmokry@gmail.com

A seleção de animais representativos dessa população foi realizada utilizando três diferentes métodos. Todas as análises foram realizadas utilizando o software estatístico R (R DEVELOPMENT CORE TEAM, 2009).

Método 1: O arquivo de pedigree foi utilizado para selecionar os animais mais representativos. Com o auxílio do software estatístico R foram calculadas as distâncias entre cada linha do arquivo, pelo método Manhattan (EVERITT, 2001), e em seguida foi construído um dendrograma com os animais. Foram calculados os coeficientes de parentesco médio para cada animal utilizando a matriz de parentesco. O dendrograma foi particionado em 50 grupos distintos, e dentro de cada grupo foi amostrado o animal que possuía o maior parentesco médio.

Método 2: A própria matriz de parentesco foi utilizada para calcular uma matriz de dissimilaridade dada por $D = \max(D) - D$. Com essa matriz de dissimilaridade foi construído um dendrograma, posteriormente particionado em 50 grupos, dentro dos quais se amostrou o animal com o maior parentesco médio.

Método 3: Os animais foram amostrados de acordo com a seguinte rotina:

- 1º Calculou-se o parentesco médio de cada animal com todos os animais da matriz de parentesco (medida de representatividade);
- 2º Selecionou-se o animal de maior parentesco médio com a população;
- 3º Calculou-se o parentesco médio de cada um dos animais ainda não selecionados com os animais já selecionados;
- 4º Dentre aqueles animais de menor parentesco médio com os animais já selecionados (divergência genética), escolheu-se o animal de maior parentesco médio com os animais ainda não selecionados (representatividade);
- 5º Repetiu-se a rotina até completar o número de animais necessários (50 amostras).

Para comparar os diferentes métodos de amostragem, foi desenhado um diagrama de Venn (Figura 1) (OLIVEROS, 2007), com as quantidades de animais selecionados por cada método dentro de cada círculo, e nas intersecções a quantidade de animais que foram amostrados por mais de um método. Os números em destaque representam a numeração dos animais amostrados.

Os métodos 1 e 2 são mais parecidos entre si pois ambos utilizaram a divisão dos dendrogramas das distâncias entre os animais para a amostra-

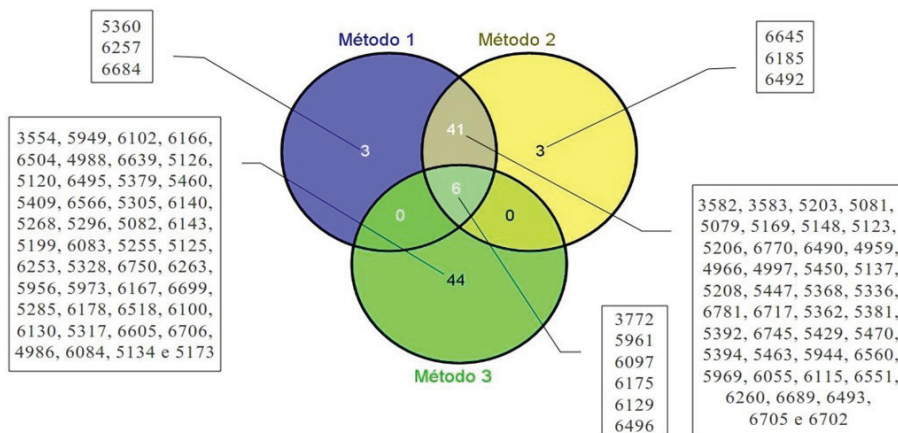


Figura 1. Comparação entre os conjuntos de animais selecionados por cada um dos 3 métodos implementados.

gem, e o método 3 utilizou diretamente as médias de parentesco. Para trabalhos futuros, sugere-se a comparação dos três métodos utilizando, como fonte de informações, os dados moleculares (genótipos) dos animais.

Referências

- BARBETTA, P. A. **Estatísticas aplicadas às Ciências Sociais**. 7. ed. Florianópolis: Ed. da UFSC, 2007. 340 p.
- BOURDON, R. M. **Understanding animal breeding**. 2nd ed. Upper Saddle River, NJ : Prentice Hall, 2000. 538 p. ill.
- EVERITT, B. S.; LANDAU, S.; LEESE, M. **Cluster analysis**. 4th ed. LoLondon: Arnold; New York: Oxford University, 2001. 237 p.
- OLIVEROS, J. C. VENNY. **An interactive tool for comparing lists with Venn Diagrams**. Disponível em: <<http://bioinfogp.cnb.csic.es/tools/venny/index.html>>. Acesso em: 1 out. 2011.
- R DEVELOPMENT CORE TEAM. **R: A language and environment for statistical computing**. Vienna, Austria: R Foundation for Statistical Computing, 2009. 409 p. Disponível em: <<http://www.R-project.org>>. Acesso em: 1 out. 2011.