



## VII Simpósio Brasileiro de Melhoramento Animal São Carlos, SP, 10 e 11 de julho de 2008

### **Análise de Componentes Principais como ferramenta na avaliação da diversidade genética de suínos**

Bruna Pena Sollero<sup>1</sup>, Marcos Yamaki<sup>1</sup>, Simone Eliza Facioni Guimarães<sup>2</sup>, Paulo Luiz Souza Carneiro<sup>3</sup>, Samuel Rezende Paiva<sup>4</sup>, Arthur da Silva Mariante<sup>4</sup>

<sup>1</sup>Estudante de doutorado da Universidade Federal de Viçosa, email: [brunasollero@yahoo.com.br](mailto:brunasollero@yahoo.com.br), email: [marcos.yamaki@sadia.com.br](mailto:marcos.yamaki@sadia.com.br)

<sup>2</sup>Professora adjunta da Universidade Federal de Viçosa, email: [sfacioni@ufv.br](mailto:sfacioni@ufv.br)

<sup>3</sup>Professor adjunto da Universidade Estadual do Sudoeste da Bahia, email: [plscarneiro@gmail.com](mailto:plscarneiro@gmail.com)

<sup>4</sup>Pesquisador da Embrapa Recursos Genéticos e Biotecnologia, e-mail: [samuel@cenargen.embrapa.br](mailto:samuel@cenargen.embrapa.br), [mariante@cenargen.embrapa.br](mailto:mariante@cenargen.embrapa.br)

**Resumo** - A análise de Componentes Principais (PCA) é uma técnica de estatística multivariada de grande aceitação na análise de dados genéticos. Aplicada no presente trabalho para avaliar cinco grupos genéticos de suínos brasileiros (n=182), genotipados por meio de 28 marcadores microssatélites, a PCA foi capaz de, essencialmente, identificar a relação genética atual existente entre estes grupos, de forma a representar grande parte da variação total existente (74,15%) por apenas dois componentes principais. Comparada com o método baseado em distâncias genéticas (UPGMA), a hipótese de que os grupos genéticos naturalizados e comerciais se mostraram geneticamente diferenciados não foi refutada. O não agrupamento da raça naturalizada Moura com os outros grupos genéticos enfatiza, ainda mais, a relevância destes resultados para programas de conservação genética.

**Palavras-chave:** conservação, marcadores microssatélites, métodos estatísticos

### **Principal Component Analysis as a tool in assessing genetic diversity of pigs**

**Abstract** - Principal Component Analysis (PCA) is a multivariate statistical technique of great acceptance in genetics data analysis. Applied in this study to evaluate five genetic groups of pigs (n=182), genotyped through 28 microsatellite markers, the PCA was able to identify the current genetic relationship among these groups, in order to represent most part of the total variance (74.15%) by only two principal components. Compared with the method based on genetic distances (UPGMA), the hypothesis that the naturalized and commercial genetic groups were genetically differentiated was not refuted. The absence of clustering of the Moura naturalized pig breed with the others emphasizes, even more, the relevance of these results for conservation genetic programs.

**Keywords:** conservation, microsatellite markers, statistical methods

## Introdução

A espécie suína, entre outras de animais domésticos, se destaca, como importante recurso complementar da diversidade genética mundial. Por isso, várias são as metodologias estatísticas testadas e aplicadas na investigação da variabilidade genética, estrutura populacional e relação genética atual destes grupos genéticos.

A Análise de Componentes Principais (PCA) é capaz de reduzir linearmente a dimensão de um conjunto de dados, de modo que este novo conjunto, em ordem de estimação, retenha o máximo de informações em termos de variação total. Esta técnica tem se mostrado uma poderosa ferramenta na identificação da estrutura e relações genéticas entre populações de acordo com suas distribuições geográficas (Rout *et al.*, 2008). Desta forma, a PCA é eficientemente aplicável em painéis de dados em múltiplas dimensões, que se estendem desde experimentos com marcadores genéticos, como os microssatélites, até amplos experimentos para análise de expressão gênica.

Este trabalho objetiva identificar a relação genética existente entre cinco grupos genéticos de suínos por meio da técnica de PCA, a partir de um painel de 28 marcadores microssatélites utilizados na avaliação da diversidade genética de suínos.

## Material e Métodos

Foram utilizados neste trabalho, dados de genotipagem de 182 animais a partir de 28 marcadores microssatélites. Estes animais representaram três raças naturalizadas de suínos (Piau, n=31; Monteiro, n=37 e Moura, n=35), uma raça comercial (Landrace, n=31) e um composto comercial (MS60, n=48). Com a perspectiva de abranger a maior parte do genoma suíno, os marcadores microssatélites propostos para a análise foram: S0313, SW240, SW72, SW445, S0227, S0226, SW2425, SW995, SW2406, S0025, SW632, OPN, SW2410, S0225, SW539, SW951, SW830, S0230, SW2008, SW957, SW1962, SWR1008, SW769, S0007, SW857, S0088, S0026 e SW24.

A fim de visualizar a relação entre os grupos genéticos, foi realizada Análise de Componente Principal por meio do programa PCA-GEN (Goudet, 1999) e os grupos genéticos foram representados num gráfico de dispersão. Valores de *bootstrapping* foram obtidos a partir de 1000 repetições para testar a consistência dos componentes principais e gerar os resultados. Por meio desta técnica, as frequências de todos os alelos de um loco foram consideradas variáveis independentes. Consequentemente, matrizes de correlação foram computadas a partir das frequências alélicas de todos os locos, bem como os autovalores de todos os componentes principais, a proporção de autovalores individuais na variância total (taxas de contribuição dos componentes), e os scores de cada indivíduo para cada componente principal.

Por meio do programa Dispan (Ota, 1993), um dendrograma foi gerado (1000 repetições) baseando-se no método de UPGMA a partir da matriz de distância de Nei  $D_A$  (Nei *et al.*, 1983) a fim de comparar o agrupamento proposto via PCA.

## Resultados e Discussão

A Análise de Componentes Principais incluiu todos os cinco grupos genéticos e os 28 locos para sumarizar graficamente a relação genética existente entre estes (Figura 1). Desta forma, o primeiro e segundo componentes apresentaram valores de 41,83% e 32,32%, respectivamente, de maneira que o resultado percentual acumulativo correspondeu a 74,15% do total da variação observada. Comparativamente, Fan *et al.* (2002) observaram que apenas 41% da variação total existente entre raças suínos chinesas foram representadas

pelos dois primeiros componentes principais. Já Rout *et al.* (2008), observaram que 65% do total da variação existente entre sete raças de cabras indígenas, por meio de 17 marcadores microssatélites, foi explicada pelos três primeiros componentes principais inferidos.

De acordo com a distribuição dos grupos genéticos pelos quadrantes, a presente análise permitiu discriminar claramente os grupos genéticos comerciais daqueles naturalizados de suínos. Mais especificamente, a raça Moura mostrou-se mais distinta em relação a todas as outras analisadas, provavelmente em razão do baixo fluxo gênico existente. A proximidade da raça naturalizada Piau principalmente com a raça Monteiro, sugere a ocorrência de compartilhamento de alelos entre estes. Por outro lado, a raça Piau tem sido também utilizada em cruzamentos com raças comerciais, o que pode explicar a sua proximidade com este grupo. O valor de  $F_{ST}$  global de 0,12556 detectado, pode sugerir certo grau de estruturação de populações entre estes grupos genéticos de suínos.

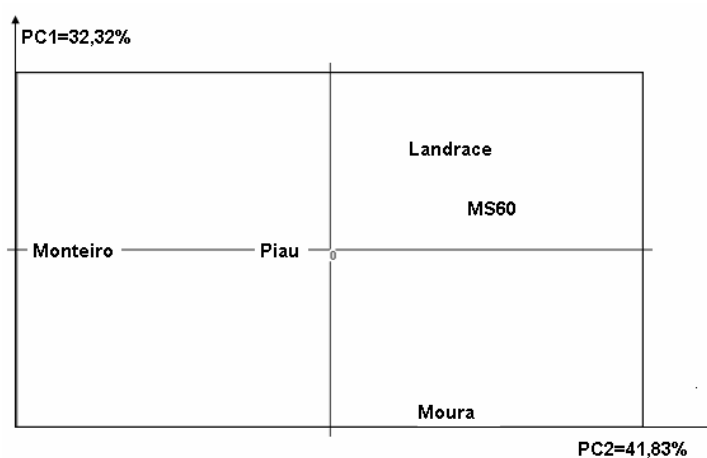


Figura 1- Representação gráfica bidimensional dos dois primeiros componentes principais que explicaram 74% da variação total observada entre os cinco grupos genéticos de suínos, a partir de 28 locos de microssatélites. PC1 e PC2= primeiro e segundo Componentes Principais, respectivamente.



Figura 2- Cinco grupos genéticos representados pelo dendrograma obtido pelo método UPGMA, com os respectivos valores de *bootstrap* nas raízes dos ramos.

Alternativamente, métodos baseados em distâncias genéticas são também utilizados para explicar a relação genética entre populações, mas a eficiência da análise de componentes principais pode se sobressair, como constatado por Saitbekova *et al.* (1999), em que se mostrou mais explicativa do que a árvore inferida pelo método algoritmo de *Neighbor joining* para diferenciar claramente raças nativas de cabras. A presente análise

comparativa, com o método de UPGMA, sugeriu similarmente ao PCA, o padrão de agrupamento entre os cinco grupos genéticos estudados, com valores de *bootstrapping* que demonstraram confiabilidade nos resultados (99, 86 e 83) (Figura 2).

Os resultados aqui expostos podem ser considerados relevantes em programas de conservação destas raças naturalizadas, uma vez que grupos genéticos geneticamente mais distantes, como a raça Moura, podem ser considerados de maior grau de identidade genética. Posteriormente, propõe-se um estudo baseando-se na mesma técnica de PCA, porém, para interpretar o peso de cada uma das variáveis originais (locos) na diversidade entre os grupos genéticos.

### Conclusões

A Análise de Componentes Principais foi capaz de distinguir os grupos genéticos de suínos naturalizados daqueles comerciais de forma a visualizar o conjunto da amostra apenas pelo gráfico dos dois primeiros componentes principais, os quais explicaram a maior parte da variação original entre os grupos. A eficiência desta técnica, utilizando dados moleculares, foi reforçada pelo dendrograma obtido pelo método UPGMA, cujos resultados corroboraram.

### Literatura Citada

- FAN, B.; WANG, Z.G.; LI, Y.J. et al. Genetic variation analysis within and among Chinese indigenous swine populations using microsatellite markers. **Animal Genetics**, v.33, p.422-427, 2002.
- GOUDET, J. **PCA-GEN, Version 1.2**. <http://www.unil.ch/izea/research.html>. Institute of Ecology, Biology Building, UNIL, Lausanne, Switzerland, 1999.
- NEI, M.; TAJIMA, F.; TATENO, Y. Accuracy of estimated phylogenetic trees from molecular data. **Journal of Molecular Evolution**, v.19, p.153-170, 1983.
- OTA, T. **Dispan, Genetic distance and phylogenetic analysis**. Institute of Molecular Evolutionary Genetics, The Pennsylvania State University, 1993.
- ROUT, P.K.; JOSHI, M.B; MANDAL, A. et al. Microsatellite-based phylogeny of Indian domestic goats. **BMC Genetics**, v.9, p.1-11, 2008.
- SAITBEKOVA, N.; GAILLARD, C.; OBEXER-RUFF. et al. Genetic diversity in Swiss goats breeds based on microsatellite analysis. **Animal Genetics**, v.30, p.36-41, 1999.