

Impacto do desequilíbrio de ligação em análises de *admixture*

Primeiro autor: Paula Adas Pereira Suniga

Demais autores: Suniga, P. A. P.¹; Silva, M. V. G. B.²; Menezes, G. R. O.³; Egito, A. A.³

Resumo

O desequilíbrio de ligação (LD) é resultado da história evolutiva que reflete a herança compartilhada em qualquer população. É uma medida da associação não aleatória de alelos em diferentes loci e pode ajudar a identificar regiões cromossômicas que expliquem variações fenotípicas. O LD varia de acordo com os fenômenos que ocasionalmente ocorrem nas populações como recombinação, deriva genética, mutação, seleção natural ou artificial. O objetivo do trabalho foi evidenciar as implicações do LD em análises de *admixture*, que visam estimar a ancestralidade das populações. Foram realizadas análises em um conjunto de dados de genótipos composto por 777.000 SNPs de 16 raças obtidos em painéis de alta densidade. O controle de qualidade dos dados foi realizado duas vezes utilizando os mesmos parâmetros, no entanto, em uma das vezes foram retirados os marcadores que apresentaram alto LD. As análises foram realizadas utilizando o programa *ADMIXTURE* e os resultados foram plotados em gráficos para melhor visualização. Além de realizar análises não supervisionadas, também foi calculado o erro de validação cruzada. A primeira observação realizada é que, ao excluir os marcadores, o banco de dados fica menor e a análise

(1) Mestranda da Universidade Federal de Mato Grosso do Sul - UFMS, paula_adas@hotmail.com. (2) Pesquisador da Embrapa Gado de Leite. (3) Pesquisadores da Embrapa Gado de Corte. * Autor correspondente.

torna-se mais rápida devido à diminuição do número de SNPs. Altera também a formação dos conjuntos populacionais. Além disso, pode-se observar uma alteração nos erros de validação cruzada, tanto no padrão da distribuição quanto na discrepância dos valores, o que influencia na escolha do melhor número de grupos (*cluster* - K) a ser escolhido. Conclui-se que a retirada dos marcadores em LD torna a análise mais rápida e robusta e, pela maior variação dos erros, sendo possível optar pelo melhor K a ser estudado com maior segurança.

Parceria / Apoio financeiro

Embrapa Gado de Corte, UFMS e Capes.