Efeito dos critérios e limiares do software PLINK na identificação de corridas de homozigose (ROH) em populações simuladas de bovinos de leite

Francisco Rosa[1]; Rodrigo de Almeida Teixeira[2]; Laila Talarico Dias[3]; Gisele Ferreira da Silva[4].

[1]Universidade Federal do Paraná – UFPR. E-mail: ffrancisco.rosa@gmail.com; [2]Universidade Federal do Paraná – UFPR. E-mail: rteixeira@ufpr.br; [3]Universidade Federal do Paraná – UFPR. E-mail: lailatalarico@ufpr.br; [4]Universidade Federal do Paraná – UFPR. E-mail: gisele.zootecnia@gmail.com.

Resumo

Com o advento do genoma bovino e a genotipagem com milhares de marcadores SNP, a corrida de homozigose surge como uma nova metodologia para estimar o grau de parentesco entre animais de produção. O objetivo desse trabalho foi de avaliar como os critérios e limiares para identificação das corridas de homozigose estão relacionados ao seu tamanho e número. Por meio do software QMSim foi simulado o genoma de uma população bovinos de leite imitanto inseminação artificial. A partir dos marcadores, foram estimados os seguimentos em homozigose utilizando o software PLINK. No programa os critérios e limiares considerados na avaliação foram pelo menos 15 ou 50 SNPs em homozigose na janela, permitido apenas 0, 1, 2 ou 4 SNPs heterozigotos no trecho. O número de ROH é entre 1 e 6% menor para ROHs com no mínimo 50 SNPs comparado a 15 SNPs em homozigose. A presença de 4 heterozigotos comparada à sua ausência como critério para a determinação de corridas de homozigose superestimou em 17 e 21% o número médio de trechos em janelas com 15 e 50 SNPs respectivamente. Diferentes critérios e limiares definidos para a identificação de ROH no programa PLINK proporcionam uma variação expressiva no número de trechos de DNA em homozigose. O aumento no número de heterozigotos permitidos superestima o número de ROHs.

Palavras-chave: Endogamia; genoma; marcadores SNP; painéis SNP.

1. Introdução

A maior parte dos animais de produção possui algum grau de parentesco (ROBERTSON, 2008), no entanto, com o avanço dos painéis de marcadores SNP, novas metodologias surgiram para melhor estimar e controlar a endogamia em rebanhos comerciais. Dentre elas podemos citar a estimativa de endogamia (F_{ROH}) a partir da corrida de homozigose (ROH) (CURIK; FERENČAKOVIĆ; SÖLKNER, 2014). O software PLINK (PURCELL et al., 2007) é uma ferramenta comumente utilizada para identificar ROH, onde é aplicado uma metodologia que determina previamente um número mínimo de SNPs em homozigose com distância específica entre eles. Com grande número de marcadores é possível identificar longos trechos em homozigose (ROH), porém o tamanho e número desses trechos no genoma pode variar de acordo com o número de marcadores e a metodologia utilizada para sua estimativa (FERENČAKOVIĆ; SÖLKNER; CURIK, 2013).

A genotipagem com marcadores SNP já está sendo amplamente utilizada, no entanto, ainda é muito cara. Uma alternativa para o estudo de fatores que podem influenciar a identificação das corridas de homozigose é a utilização de softwares de simulação. Dentre os descritos na literatura o QMSim se destaca porque cria diferentes cenários envolvendo grande número de marcadores. Além disso, está muito bem documentado e possibilita imitar situações muito próximas ao que encontramos em animais de produção (SARGOLZAEI; SCHENKEL, 2009).

O objetivo desse trabalho é de avaliar como os critérios e limiares para identificação de ROH está relacionada ao seu tamanho e número em populações simuladas de bovinos da raça holandesa.

2. Material e Métodos

O pedigree e genótipo dos animais foram simulados com o software QMSim (SARGOLZAEI; SCHENKEL, 2009). A população histórica foi simulada partindo de 30 até 20600 animais (20000 fêmeas e 600 machos) em 100 gerações para estabelecer o desequilíbrio de ligação (LD) desejado. Foi simulado um "afunilamento" na população histórica para criar um alto nível de LD, isso foi possível pela queda brusca no número de animais na geração 70 a 78 e posterior aumento até a geração 100. Para criar um cenário com inseminação artifical, a partir da população histórica foi simulado uma população 0,1% de machos em relação ao número de fêmeas. Nas 10 gerações de seleção o acasalamento foi assortativo positivo, com 1 filhote por acasalamente e 50% de probabilidade de ser macho ou fêmea. Em cada geração foi considerada uma taxa de reposição de 25% para fêmeas e 95% para machos, e utilizou-se a idade como critério de descarte.

O genoma foi simulado com 29 pares de cromossomos autossomos, com o tamanho variando entre 60 e 154 cM, e número de QTLs relacionados à produção de leite, variando entre 0 e 41, distribuidos aleatoriamente entre os cromossomos. A taxa de mutação aplicada na população histórica foi de $2{,}5x10^{-5}$ para marcadores e QTLs. Os genótipos foram simulados com 50558 marcadores SNP, com a densidade de aproximadamente 17 marcadores a cada cM imitando um painel de 50 kb. Sobre o genoma foram distribuídos 93 QTLs com número de alelos variando aleatoriamente entre 2, 3 e 4 (DEHNAVI et al., 2018).

A identificação das corridas de homozigose foi feita com o software PLINK. Os critérios e limiares considerados na análise foram (i) pelo menos 15 ou 50 SNPs em homozigose; (ii) ROH com no mínimo 1 Mb de comprimento; (iii) densidade de pelo menos 1 SNP a cada 50 kb; (iv) máximo de 4 SNPs perdidos no trecho; (v) permitido apenas 0, 1, 2 ou 4 SNPs heterozigotos no trecho; (vi) intervalo entre SNPs no trecho de 500 kb; (vii) sobreposição de trechos em homozigose de 0,05.

3. Resultados e Discussão

O nível de endogamia de um indivíduo pode ser subestimado de acordo com os critérios e limiares utilizados para a identificação das corridas de homozigose (FERENČAKOVIĆ; SÖLKNER; CURIK, 2013). Na Tabela 1 estão descritos os resultados da estimativa de ROH utilizando diferentes critérios para o número mínimo de SNPs em homozigose e o número marcadores heterozigotos permitidos no trecho.

Table 1- Número médio de ROHs (nROH) e comprimento médio do ROH e seus desvios padrão (σ); comprimento mínimo (Mín) e máximo (Máx) de ROH em análises realizadas com diferentes critérios e limiares para número mínimo de SNPs em homozigose (jSNP) e número de heterozigotos permitidos no ROH

jSNP	Heterozigotos	nROH	σ	ROH	σ(Mb)	Mín	Máx
15	0	128,37	8,45	11,25	10,75	1,01	153,86
	1	142,48	9,43	10,89	10,03	1,01	153,86
	2	149,26	9,86	10,74	9,70	1,03	153,86
	4	154,52	10,25	11,00	9,56	1,08	153,86
50	0	119,93	8,05	11,86	11,25	1,84	153,86
	1	135,99	8,97	11,28	10,35	1,74	153,86
	2	144,14	9,36	11,03	9,94	1,75	153,86
	4	152,88	10,12	11,08	9,63	1,93	153,86

O número médio de ROH variou de 128,37 ± 4,45 a 154,52 ± 10,25, quando foi aceito pelo menos 15 SNPs em homozigose no trecho. Para 50 SNPs na janela, o número médio de ROH variou

entre 119,93 ± 8,05 a 154,88 ± 10,12. Valores médios de 128,37 e 119,93 ROHs foram estimados quando a análise desconsiderou a presença de SNPs heterozigotos na janela com 15 e 50 SNPs respectivamente. Considerando 1, 2 ou 4 heterozigotos o número médio de ROH foi de 142,48; 149,26 e 154,52 respectivamente, para janela com 15 SNPs. Médias de 135,99; 144,14 e 152,88 foram encontradas respectivamente para 1, 2 ou 4 heterozigotos em janelas com no mínimo 50 SNPs. Portanto, o número de ROH é entre 1 e 6 % menor para ROHs com no mínimo 50 SNPs comparado a 15 SNPs em homozigose. A presença de 4 heterozigotos comparada à sua ausência como critério para a determinação de corridas de homozigose superestimou em 17 e 21% o número médio de trechos em janelas com 15 e 50 SNPs respectivamente.

O comprimento médio de ROH variou de 10,74 a 11,86 Mb com valores altos de desvio padrão. Para cada análise com critérios e limiares diferentes os comprimentos variaram entre 1 Mb e 153 Mb aproximadamente, o que justifica a grande dispersão descrita.

4. Conclusão

Diferentes critérios e limiares definidos para a identificação de ROH no programa PLINK proporcionam uma variação expressiva no número de trechos de DNA em homozigose. O aumento no número de heterozigotos permitidos superestima o número de ROHs. A genotipagem com painéis mais densos é uma alternativa para diminuir os vieses na identificação de trechos em homozigose.

5. Referências

CURIK, I.; FERENČAKOVIĆ, M.; SÖLKNER, J. Inbreeding and runs of homozygosity: A possible solution to an old problem. **Livestock Science**, v. 166, n. 7, p. 26–34, 6 ago. 2014.

DEHNAVI, E. et al. The effect of using cow genomic information on accuracy and bias of genomic breeding values in a simulated Holstein dairy cattle population. **Journal of Dairy Science**, p. 5166–5176, 2018.

FERENČAKOVIĆ, M.; SÖLKNER, J.; CURIK, I. Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors. **Genetics Selection Evolution**, v. 45, n. 1, p. 42, 2013.

PURCELL, S. et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. **The American Journal of Human Genetics**, v. 81, n. 3, p. 559–575, 2007.

ROBERTSON, A. Inbreeding in artificial selection programmes. **Genetics Research**, v. 89, n. 5–6, p. 275–280, 2008.

SARGOLZAEI, M.; SCHENKEL, F. S. QMSim: A large-scale genome simulator for livestock. **Bioinformatics**, v. 25, n. 5, p. 680–681, 2009.