

# **SIMULAÇÃO DE MARCADORES GENÉTICOS EM CARACTERÍSTICAS QUANTITATIVAS<sup>1</sup>.**

Adam Taiti Harth Utsunomiya<sup>2</sup>, Ricardo da Fonseca<sup>3</sup>, Michele Porto Pires<sup>2</sup>, Rafael Keith Ono<sup>2</sup>. - Zootecnia – Faculdade de Zootecnia – Campus de Dracena.

<sup>1</sup> Apoio Fapesp – SP e CNPq – Brasil.

<sup>2</sup> Aluno de graduação da Faculdade de Zootecnia – Unesp/Dracena – Laboratório de Computação Científica Aplicada à Zootecnia (LuCCA-Z).

<sup>3</sup> Professor Assistente da Faculdade de Zootecnia – Unesp/Dracena – Laboratório de Computação Científica Aplicada à Zootecnia (LuCCA-Z).

Recentemente, a genética molecular tem produzido grande quantidade de informações sobre marcadores e genes. Desde então, diversos estudos foram realizados em áreas relacionadas tentando encontrar maneiras de incorporar estas informações em suas aplicações ou para otimizar seu uso em tarefas diárias. Particularmente no melhoramento animal, os estudos visam compreender os efeitos de incorporar informação molecular, principalmente informações de marcadores, nos programas de seleção e cruzamento. Importantes resultados foram encontrados por Fernando e Gianola (1986), entre outros. Em muitos deles a simulação provou-se uma importante ferramenta.

Também no ensino, a simulação tem sido importante, por exemplo, para gerar cromossomos, genes e marcadores ou para mostrar os efeitos de uma alteração na frequência alélica em uma população.

Contudo, em ambas as áreas, ciência e educação, uma discussão detalhada de algoritmos usados para gerar dados moleculares não foi encontrada. Por exemplo, esta informação seria valiosa para melhoristas para estudar como incorporar informações moleculares e medir seus efeitos em programas de melhoramento. A demanda da comunidade científica em melhoramento animal por estes algoritmos foram observadas também em grupos de discussões via e-mail como: AGDG, AngenMap e Quan\_gen.

Portanto, esse trabalho tem por objetivo iniciar discussões e publicar algoritmos para gerar marcadores, que possam ser usados em ambas as áreas, ciência e educação.

Dois algoritmos foram desenvolvidos, o primeiro para aplicar em ciência e o segundo em educação.

O problema básico é: em uma população de indivíduos, simular um genoma de tamanho  $s$  com  $g$  genes junto com  $m$  marcadores espalhados entre  $c$  cromossomos.

Criar marcadores, essencialmente, gerando endereços nos cromossomos. Se apenas um cromossomo existe, o processo é razoavelmente simples. O cromossomo pode ser visto como uma linha onde o primeiro endereço é zero e o último é o tamanho do genoma,  $s$ .

O algoritmo para aplicação científica é baseado na teoria de genética de populações, onde o número de cromossomos não é relevante, então pode-se trabalhar com um par de cromossomos homólogos somente. Todo o tamanho do genoma estará concentrado em um cromossomo, onde  $g$  genes e  $m$  marcadores serão endereçados.

O algoritmo é descrito como segue:

1. criar uma lista vazia,  $b$ ;
  2. para cada gene no cromossomo:
    1. gere um número aleatório ( $v$ ) de uma distribuição uniforme no intervalo  $[0, \lfloor s/g \rfloor]$ ;  
 $\lfloor s/g \rfloor$  está representado o tamanho do cromossomo.
    2. some o valor de  $v$  com o valor previamente armazenado;  
o valor  $v$  representa o endereço do gene no cromossomo atual. No primeiro, o valor armazenado seria zero;
    3. armazene  $v$  em  $b$ ;  
 $b$  seria de tamanho  $g$ ;
- Procedimento para marcadores.
3. para cada marcador no cromossomo:
  4. crie uma lista vazia,  $d$ ;

1. gere um número aleatório ( $q$ ) de uma distribuição uniforme no intervalo  $[0, \mathbb{L}/m]$ ;
2. some o valor de  $q$  com o valor previamente armazenado;  
o valor de  $q$  representa o endereço do marcador no cromossomo atual. No primeiro, o valor armazenado seria zero;
3. armazene  $q$  em  $d$ ;  
 $d$  seria de tamanho  $m$ .

Deve-se observar que  $b$  e  $d$  são, respectivamente, mapas dos endereços de genes e marcadores no cromossomo original.

Já para o algoritmo de aplicação educacional o número de cromossomos é importante, tanto quanto a variação de seus tamanhos. Os estudantes demandam cenários com menos abstração, principalmente nos seus primeiros cursos de genética.

Um possível algoritmo segue:

1. crie uma lista ( $\mathbb{L}$ ) com 2 elementos, onde o primeiro é zero e o segundo é  $s$ ;
2. para cada indivíduo na população repita os passos seguintes  $c - 1$  vezes:
  1. gere um número aleatório ( $r$ ), de uma distribuição uniforme, dentro do intervalo  $[1, (s - 1)]$ ;
  2. verifique se  $r$  está inserido em  $\mathbb{L}$ ;
  3. se FALSO, armazene  $r$  em  $\mathbb{L}$ ;
  4. se VERDADEIRO, retorne ao passo 2.1.;
  5. organize  $r$  em ordem ascendente em  $\mathbb{L}$ ;
3. para cada cromossomo a ser criado:
  1. crie uma variável para armazenar o tamanho do cromossomo,  $w$ ;
  2. calcule  $l_{i+1} - l_i$ ;  
 $l_i$  representa o  $i^{\text{th}}$  elemento da lista  $\mathbb{L}$ ;
  3. armazene o resultado em  $w$ ;

Agora, procedimento para gerar endereços de genes e marcadores nos cromossomos.

4. crie uma lista  $a$ ;
5. para cada cromossomo ou quando sobrar genes ou marcadores para atribuir um endereço:
  1. gere um número aleatório,  $r$ , no intervalo  $[1, c]$ ;  
este passo garante que o cromossomo será escolhido aleatoriamente;
  2. verifique se  $r$  está inserido em  $a$ ;
  3. se FALSO, armazene  $r$  em  $a$ ;
  4. se VERDADEIRO, retorne para o passo 5.1.;
  5. gere um outro número aleatório ( $u$ ) de uma distribuição uniforme no intervalo  $[1, g]$ ;  
o valor de  $u$  representa o número de genes no cromossomo  $r$ ;
  6. faça  $g = g - u$ ;
  7. crie outra lista,  $b$ ;
  8. para cada gene no cromossomo  $r$ :
    1. gere um número aleatório ( $v$ ) de uma distribuição uniforme no intervalo  $[0, \mathbb{L}/u]$ ;
    2. some o valor de  $v$  com os valores previamente armazenados;
    3. armazene  $v$  em  $b$ ;

Procedimento para marcadores.

9. gere um número aleatório ( $t$ ) de uma distribuição uniforme no intervalo  $[1, m]$ ;  
o valor de  $t$  representa o número de marcadores no cromossomo  $r$ ;
10. faça  $m = m - t$ ;
11. para cada marcador no cromossomo  $r$ :
12. crie uma lista  $d$ ;
  1. gere um número aleatório ( $q$ ) de uma distribuição uniforme no intervalo  $[0, \mathbb{L}/t]$ ;
  2. some o valor de  $q$  com os valores previamente armazenados;
  3. armazene  $q$  em  $d$ .

Se o software está visando ensino, muitos dos detalhes e ilustrações, usualmente não necessários em aplicações científicas, devem ser incluídos. Frequentemente, estas inclusões têm efeito negativo na eficiência dos algoritmos mas sua execução é imperativa para a aceitação do software.

## **Referências Bibliográficas**

Fernando, R. L. and Gianola, D. (1986). Optimal properties of the conditional mean as a selection criterion. *Theor. Appl. Genet.*, 72:822–825.

**Bolsa:** FAPESP