

**UNIVERSIDADE FEDERAL DOS VALES DO
JEQUITINHONHA E MUCURI**

MICHEL MARQUES FARAH

**SIMULAÇÃO EM NÍVEL DE GENE E DE INDIVÍDUO APLICADA AO
MELHORAMENTO ANIMAL**

**DIAMANTINA - MG
2010**

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

MICHEL MARQUES FARAH

**SIMULAÇÃO EM NÍVEL DE GENE E DE INDIVÍDUO APLICADA AO
MELHORAMENTO ANIMAL**

Dissertação apresentada à Universidade Federal dos Vales do Jequitinhonha e Mucuri, como parte das exigências do Programa de Pós-Graduação em Zootecnia, para obtenção do título de *Magister Scientiae*.

Orientador: Prof. Aldrin Vieira Pires

DIAMANTINA - MG
2010

Ficha Catalográfica - Serviço de Bibliotecas/UFVJM
Bibliotecária Viviane Pedrosa de Melo CRB6 2641

**F219s
2010**

Farah, Michel Marques

Simulação em nível de gene e de indivíduo aplicada ao melhoramento animal./Michel Marques Farah– Diamantina: UFVJM, 2010.

54p.

Dissertação (Dissertação apresentada ao Curso de Pós- Graduação Stricto Sensu em Produção Animal) - Faculdade de Ciências Agrárias, Universidade Federal dos Vales do Jequitinhonha e Mucuri.

Orientador: Prof. Aldrin Vieira Pires

1. Algoritmo 2. LZ5 3. Marcadores moleculares 4. QTL 5. Simulação estocástica 6. Simulação determinística I. Título

CDD 636.0821

MICHEL MARQUES FARAH

**SIMULAÇÃO EM NÍVEL DE GENE E DE INDIVÍDUO APLICADA AO
MELHORAMENTO ANIMAL**

Dissertação apresentada à Universidade Federal dos Vales do Jequitinhonha e Mucuri, como parte das exigências do Programa de Pós-Graduação em Zootecnia, para obtenção do título de *Magister Scientiae*.

APROVADA em 15/07/2010

Prof. Aldrin Vieira Pires – UFVJM
Orientador

Prof. Idalmo Garcia Pereira – UFVJM
Co-orientador

Prof. Ricardo da Fonseca – UNESP
Co-orientador

Prof. Frederico de Castro Figueiredo - IFES

DIAMANTINA - MG
2010

*“Há tantos quadros na parede
Há tantas formas de se ver o mesmo
quadro”
- Humberto Gessinger -*

DEDICO

Aos meus pais.

À Camila.

Aos meus irmãos.

À Yasmin, Gabriela e Jehad.

AGRADECIMENTO

Ao professor Aldrin Vieira Pires, pela orientação e pela paciência durante a execução deste trabalho.

À Camila, meu amor, por toda a ajuda, carinho, dedicação e respeito durante todos esses anos.

Ao professor Idalmo pela amizade e por tudo que já foi feito.

Ao professor Ricardo, pela orientação, dedicação e grande ajuda na execução deste trabalho.

Aos meus pais, minha inspiração de vida, respeito e admiração, por tudo que já fizeram para mim e para meus estudos.

Aos meus irmãos pela força e incentivo.

À Yasmin, Gabriela e Jehad por alegrarem minha vida.

À Sônia Tângari, exemplo de vida, e família por tudo que já me ajudaram durante todo período.

Ao DTI – UFVJM por toda ajuda principalmente na parte inicial do trabalho.

Às Repúblicas Fenda do Biquini e Eternament.

Aos meus amigos: Tio Chico, Miudo, Dinho, Raphael Linear, Pipeta, Felipe Nilvan, Hudson, Renan, Diogo, Adam, Rodrigo, Mclean, Rafael e todos os amigos de Diamantina e Dracena.

À CAPES, pela concessão da bolsa de estudos.

Ao departamento de Zootecnia e pós-graduação e à Adriana.

E a todos os demais, muito obrigado!

RESUMO

FARAH, Michel Marques. Universidade Federal dos Vales do Jequitinhonha e Mucuri, julho de 2010. 54p. **Simulação em nível de gene e de indivíduo aplicada ao melhoramento animal**. Orientador: Aldrin Vieira Pires. Co-orientador: Idalmo Garcia Pereira. Co-orientador: Ricardo da Fonseca. Dissertação (Mestrado em Zootecnia).

A simulação de dados apresenta diversas vantagens, como proporcionar a obtenção de respostas à seleção e diminuir o tempo necessário para a avaliação das metodologias estudadas no melhoramento genético animal. Porém, os trabalhos que utilizam simulação empregam vários termos como simulação estocástica, simulação determinística, simulação de Monte Carlo, simulação em nível de gene e simulação em nível de indivíduo e, muitas vezes, estes termos são utilizados de maneiras diferentes ou em outras condições, causando uma divergência nos termos utilizados. Assim, os objetivos deste trabalho foram agrupar, definir e diferenciar os termos técnicos utilizados nos trabalhos de simulação em melhoramento genético animal e comparar e definir as propriedades dos procedimentos de simulação em nível de indivíduo e em nível de gene. Foram desenvolvidos três cenários de simulação, em nível de indivíduo, em nível de gene com e sem marcador utilizando o software LZ5. Foram simuladas três populações de suínos para cada cenário e com diferentes herdabilidades (0,12, 0,27 e 0,47). A população-base foi constituída de 1500 animais, sendo 750 machos e 750 fêmeas e para as duas simulações em nível de gene foi considerado um genoma de 2800 cM e 18 cromossomos de tamanhos aleatórios, as características foram governadas por 500 locos poligênicos dialélicos, com frequências alélicas iguais e taxa de recombinação de 0,01. Para a simulação em nível de gene com marcadores, ainda foram distribuídos marcadores distanciados igualmente a 50 cM e distribuídos aleatoriamente 5 QTLs por todo o genoma. Os valores amostrados apresentaram bem semelhantes para os três tipos de simulação, apresentando um aumento das variâncias aditiva e fenotípica e da herdabilidade nas primeiras gerações e depois decrescendo ao longo das gerações. Já para a média fenotípica, houve um ganho genético por geração, indicando que todos os métodos utilizados são eficientes para a obtenção de dados simulados. Assim, a vantagem da simulação em nível de gene é que é possível simular marcadores moleculares e QTLs, enquanto a simulação em nível de indivíduo é muito eficiente para obtenção de dados como o valor genético do indivíduo e da média fenotípica da população em um período de tempo muito menor, pois demanda menos recursos computacionais e de algoritmos estruturados para desenvolver quando comparado com a simulação em nível de gene. Portanto, define-se simulação em nível de indivíduo como uma metodologia de simulação que consiste em gerar valores genéticos (G) a partir de uma distribuição normal com média e variância previamente definidas; enquanto para a simulação em nível de gene a metodologia consiste em gerar os valores dos efeitos de cada loco poligênico e seus QTLs, a partir de uma distribuição normal com média e variância previamente definidas para cada componente, e pela soma destes, obtém-se o G de cada indivíduo da população. Para a geração do efeito residual (E) as duas metodologias de simulação são feitas da mesma forma, gerando-se um efeito aleatório amostrado, também, de uma distribuição normal e assim obtém-se os valores fenotípicos (P) de cada indivíduo pela soma destes dois componentes (G+E).

Palavras-chave: algoritmo, LZ5, marcadores moleculares, QTL, simulação estocástica, simulação determinística

ABSTRACT

FARAH, Michel Marques. Federal University of the Jequitinhonha and Mucuri Valleys, July 2010. 54p. **Simulation of individual and gene level applied to animal breeding.** Adviser: Aldrin Vieira Pires. Committee members: Idalmo Garcia Pereira, Ricardo da Fonseca. Dissertation (Master's degree in Animal Science).

The simulation data has several advantages, such as providing the obtaining responses to selection and reduce the time required for evaluation methodologies studied in animal breeding. However, simulation studies employ various terms such as simulation stochastic, deterministic simulation, Monte Carlo simulation, simulation level of gene and simulation at the individual level and often these terms are used in different ways or in other conditions, causing a divergence in the terms used. Thus, the objectives were cluster, define and differentiate the technical terms used in the work of simulation in animal breeding and compare and define the properties procedures for simulation-level and individual-level gene. There had been developed three scenarios for simulation at the individual-level and level gene, with and without marker, using the software LZ5. There had been simulated three pig populations for each scenario, with different heritabilities (0.12, 0.27 and 0.47). The base population consisted of 1500 animals, 750 males and 750 females and for both simulations at the level of the gene was considered a genome of 2800 cM, and 18 chromosomes in random sizes, the characteristics were governed by 500 loci diallelic polygenic, with equal allele frequencies and recombination rate of 0.01. For the simulation Level with gene markers, were also distributed bookmarks equally spaced at 50 cM and five QTL distributed randomly across the genome. The sampled values were very similar for the three types of simulation, an increase of additive variance and phenotype and heritability in the first generations and then decreasing to over the generations. As for the average phenotype was a genetic per generation, indicating that all methods used are efficient for obtain simulated data. Thus, the advantage of gene-level simulation is that it can simulate molecular markers and QTLs, while the simulation at individual level is very efficient for obtaining data as the individual's genetic value and phenotypic average of the population over a period of much less time, since it requires less computational resources and algorithms structured to develop, when compared with the simulation-level gene. Therefore, it is defined as the individual level simulation a methodology simulation that generates breeding values (G) from a normal distribution with mean and variance as previously defined; and the gene level simulation is defined as a methodology that generates the values of effects of each locus and their polygenic QTLs from a normal distribution with mean and variance previously defined for each component, and the sum of these gives the G of each individual in the population. For the generation of residual effect (E) the two simulation methodologies are made in the same way, generating a random effects sampled also a normal distribution and so it was obtained the phenotypic values (P) of each individual by summing these two components (G+E).

Keywords: algorithm, deterministic simulation, LZ5, molecular markers, QTL, stochastic simulation

Sumário

| | |
|---|----|
| 1. INTRODUÇÃO | 11 |
| 2. REVISÃO DE LITERATURA | 12 |
| 2.1. Simulação de dados no Melhoramento Genético Animal | 12 |
| 2.2. Simulações determinística, estocástica e de Monte Carlo..... | 12 |
| 2.3. Simulação em nível de gene e em nível de indivíduo | 14 |
| 3. MATERIAL E MÉTODOS | 18 |
| 3.1. Simulação de dados..... | 18 |
| 3.2. Geração da população-base (G_0) | 19 |
| 3.3. Definição dos candidatos à seleção..... | 27 |
| 3.4. Realização da seleção..... | 29 |
| 3.5. Formação da próxima geração (G_{n+1})..... | 31 |
| 3.6. Formação e armazenamento de uma população..... | 42 |
| 3.7. Formação dos gametas e de um novo indivíduo | 43 |
| 3.8. Simulação de QTLs e Marcadores | 43 |
| 4. RESULTADOS E DISCUSSÃO | 43 |
| 5. CONCLUSÃO | 51 |
| 6. REFERÊNCIAS BIBLIOGRÁFICAS | 52 |

1. INTRODUÇÃO

Com o avanço da zootecnia e dos programas de melhoramento genético animal, cada vez mais são exigidas respostas rápidas sobre os programas de melhoramento e quais as suas implicações futuras no desenvolvimento de uma determinada população sob condições de ambientes adversos.

Uma das formas de se obter respostas em um curto período de tempo está ligada a informática pelo uso de softwares que obtém os dados necessários de maneira mais rápida e eficiente, de alto valor, para ajudar nas tomadas de decisões, tornando o melhoramento genético animal mais eficiente.

Portanto, o desenvolvimento de programas de simulação genética se torna, cada vez mais, uma ferramenta muito útil para avaliação de características e cenários diversos, permitindo obter predições sobre as possíveis respostas de uma determinada população sob a influência de diversos fatores auxiliando os profissionais nas tomadas de decisões e na avaliação da eficiência de metodologias aplicadas ao melhoramento genético.

A simulação de dados apresenta diversas vantagens, como, proporcionar a obtenção de respostas à seleção e diminuir o tempo necessário para a avaliação das metodologias estudadas no melhoramento genético animal. Porém, os trabalhos que utilizam simulação empregam vários termos como simulação estocástica, simulação determinística, simulação de Monte Carlo, simulação em nível de gene e simulação em nível de indivíduo e, muitas vezes estes termos são utilizados de maneiras diferentes ou em outras condições, causando uma divergência nos termos utilizados.

A simulação em nível de indivíduo e em nível de gene, muitas vezes, são utilizadas em um mesmo contexto, causando uma divergência entre os pesquisadores sobre qual seria definição adequada para cada tipo de simulação, sendo que a principal diferença entre elas está na metodologia de simulação dos valores genéticos aditivos, e quais os cenários são mais adequados para determinada metodologia.

Assim, os objetivos deste trabalho foram agrupar, definir e diferenciar os termos técnicos utilizados nos trabalhos de simulação em melhoramento genético animal e comparar e definir as propriedades dos procedimentos de simulação em nível de indivíduo e em nível de gene.

2. REVISÃO DE LITERATURA

2.1. Simulação de dados no Melhoramento Genético Animal

A simulação de dados está sendo cada vez mais utilizada no melhoramento animal, tornando-se uma importante ferramenta, não só para a diminuição dos custos de implantação de um programa de melhoramento, mas também para a diminuição do período de tempo necessário para avaliar o comportamento de diferentes metodologias sob diferentes condições, e, assim, aplicar os resultados em populações reais. Diversos pesquisadores utilizaram a simulação para avaliação de programas de melhoramento genético (POLLAK et al., 1984; JOHNSON & NOTTER, 1987; SORENSEN et al., 2001; CARVALHEIRO et al., 2002; HAO et al., 2004; KUEHN, 2005; CUNHA et al., 2006; ASSIS et al., 2007; SOUZA, 2008).

A simulação genética permite gerar informações de indivíduos por várias gerações e em um período de tempo muito menor, quando comparado com os procedimentos de campo. Segundo Pereira et al. (2007), através da simulação é possível conhecer os verdadeiros parâmetros das distribuições dos dados, o que é de extrema importância para avaliação de metodologias utilizadas na estimação dos componentes de variância e predição dos valores genéticos. Assim, é possível gerar dados fenotípicos a partir do valor genético hipotético dos indivíduos, possibilitando condições para a comparação das metodologias. Apesar de ser computacionalmente oneroso, apresenta a vantagem de ser possível para situações complexas em que a aplicação de outro método para avaliação seria inviável (CARNEIRO JR. et al., 2004).

Como destacado por Cunha et al. (2006), o uso de dados simulados permite conduzir a seleção por várias gerações sucessivas, necessárias para avaliar o comportamento dos diferentes parâmetros e, dessa forma, extrapolar os resultados para as populações reais. Assim, para se obter os dados simulados, diversas metodologias são utilizadas no melhoramento animal, como a simulação determinística, simulação estocástica, simulação de Monte Carlo, simulação em nível de indivíduo e em nível de gene.

2.2. Simulações determinística, estocástica e de Monte Carlo

Segundo Pereira (2000), as simulações são determinísticas quando as variáveis de entrada assumem valores exatos sendo os resultados de saída sempre os mesmos, independente do número de repetições. Já a simulação estocástica permite que as variáveis de

entrada assumam diversos valores dentro de uma distribuição de probabilidades previamente definida. Os resultados gerados pela simulação estocástica podem ser diferentes a cada replicação, em função da natureza aleatória das variáveis que dão entrada no modelo.

Seguindo o modelo de simulação determinística, Formigoni et al. (2004) desenvolveram um modelo bioeconômico para avaliar a relação de custos e receitas das características, probabilidade a prenhez aos 14 meses (PP14) e a *stayability*, e sua sensibilidade a diferentes circunstâncias de produção para um tradicional sistema de cria. Estes autores concluíram que a *stayability*, em valores absolutos, mostrou-se economicamente mais relevante que a PP14 de acordo com as condições econômicas e de produção simulados, enquanto ambas as características foram sensíveis as diferentes circunstâncias de produção avaliada.

Já Neves et al. (2009) desenvolveram um estudo de simulação estocástica com o objetivo de verificar as consequências do uso combinado de acasalamento dirigido e sêmen sexado em uma população de bovinos de corte, simulando seis gerações de seleção para três cenários de acasalamento e uso de sêmen sexado. O primeiro cenário o usou-se a acasalamento aleatório e o uso exclusivo de sêmen convencional, o segundo cenário caracterizou-se pelo uso de acasalamento associativo positivo nas 40% melhores vacas e acasalamento associativo negativo nas demais, sem uso de sêmen sexado. O terceiro cenário seguiu o mesmo procedimento de acasalamento do segundo, combinando-o com uso de sêmen sexado nas vacas submetidas a acasalamento associativo positivo. Os autores concluíram que o uso de acasalamento associativo negativo não foi eficiente para reduzir a variabilidade dos animais destinados ao abate, ao contrário do uso combinado de acasalamento associativo positivo e sêmen sexado que aumentou a produção de animais geneticamente superiores.

Outro método de simulação é o de Monte Carlo, que é um método estatístico utilizado em simulações estocásticas com diversas aplicações. Este método é muito utilizado para obter aproximações numéricas de funções complexas e envolve a geração de observações de alguma distribuição de probabilidade pré-definida e o uso das amostras obtidas. O amostrador de Gibbs é utilizado para aproximar a função de interesse e estimar distribuições marginais *a posteriori* a partir de densidades condicionais complexas (CARNEIRO JR. et al., 2005).

Diversos pesquisadores utilizam este método para estimação de componentes de variância utilizando a amostragem de Gibbs, como Reis et al. (2009) que utilizaram o software R para gerar dados de populações selecionadas e não selecionadas, utilizando o método de Monte Carlo, a fim de se compararem os dados simulados com os dados reais. Para

isso, simularam populações de 40 indivíduos (20 machos e 20 fêmeas) sob seleção e acasalamento aleatório, durante seis gerações sucessivas. Os resultados simulados mostraram-se próximos aos reais.

Este programa de simulação descrito por Reis et al. (2009), além de utilizarem a simulação de Monte Carlo, utilizaram também propriedades da simulação em nível de indivíduo para realizarem a evolução da população. A seguir serão descritas as metodologias de simulação em nível de indivíduo e em nível de gene, que também são exemplos de procedimentos de simulação estocástica.

2.3. Simulação em nível de gene e em nível de indivíduo

Segundo Carneiro Jr. et al. (2004), basicamente, existem dois procedimentos para a simulação de dados: em nível de indivíduo e em nível de gene.

Na simulação em nível de indivíduo utilizam-se desvios normais, ou seja, simulam-se vetores de números com uma média e uma variância obtendo-se o valor genético dos indivíduos e o efeito do ambiente sobre as características, e, pela soma desses gera-se o valor fenotípico de cada indivíduo (CARNEIRO JR. et al., 2004).

Na a simulação em nível de gene utiliza-se estrutura semelhante à do material genético, em que a determinação da característica é feita pela contribuição de uma série de genes, os quais podem apresentar várias formas alélicas e alterar a expressão genotípica, de acordo com um código de expressão genética determinado pelo pesquisador, em que podem ocorrer, à semelhança do processo biológico, dominância, epistasia, pleiotropia e ligação gênica (CARNEIRO JR. et al., 2004). Os valores genéticos aditivos são simulados pela soma da contribuição de cada loco envolvido na determinação da característica, às quais são somados os efeitos de ambiente, amostrados de uma distribuição normal, para formação dos valores fenotípicos.

Tanto a simulação em indivíduo quanto em nível de gene são baseadas na pressuposição do modelo infinitesimal, proposto por Fisher (1918), no qual uma característica é determinada por um número infinito de locos, sem ligação e com efeito aditivo somente. Cada um dos locos apresenta um efeito infinitesimal e as grandes mudanças na média ocorrem pela soma dos efeitos de um grande número de locos com efeitos muito pequenos.

O modelo infinitesimal assume que uma característica é afetada por um número muito grande de locos, cada um, com efeito muito pequeno, satisfazendo tanto a normalidade quanto a linearidade. Assim, a maioria dos modelos de seleção são baseados no pressuposto de que

uma população sob seleção pode ser adequadamente descrita pelo modelo infinitesimal (LYNCH & WALSH, 1998).

Na literatura, encontram-se vários trabalhos que utilizam a simulação em nível de indivíduo em diversos ambientes. Carvalheiro et al. (2002) utilizaram simulação para avaliar os efeitos da heterogeneidade de variância residual entre grupos de contemporâneos na avaliação genética de bovinos de corte. Para isso, utilizaram uma população inicial real a fim de evitar que estruturas artificiais ou distantes da realidade fossem simuladas e, em seguida, geraram os ganhos de peso pós-desmame, utilizando a função RANNOR do SAS, que gera um conjunto de valores com distribuição normal. Os autores concluíram que o uso de avaliações ponderadas produz resultados não inferiores às avaliações que assumiram homogeneidade de variância.

Sorensen & Waagepetersen (2002) simularam duas populações durante 25 gerações em 10 repetições, utilizando, para cada população, os métodos de reprodução natural e reprodução por seleção com uso de informações dos progenitores. As simulações utilizadas permitiram obter resultados reais para as populações. Concluiu-se que os esquemas de simulações são ferramentas que podem ser implantadas nos programas de melhoramento genético animal.

A fim de se compararem as estimativas de componentes de variância e os valores genéticos preditos a partir de modelos lineares mistos Gaussianos e Robustos, Pereira et al. (2007) simularam 50 populações de 1000 animais cada, com seus valores fenotípicos para uma característica produtiva hipotética. Cada população foi composta de cinco gerações com um núcleo de dez pais e 50 mães, e, nas gerações seguintes, outros dez machos e 50 fêmeas foram selecionados aleatoriamente, a partir da geração anterior, utilizando o modelo linear normal contaminado (com formato leptocúrtico, ou seja, cauda longa), sugerindo que este modelo oferece uma alternativa para a estimação dos componentes de variância.

Também são encontrados diversos trabalhos no melhoramento animal que utilizam a metodologia de simulação em nível de gene. Fonseca et al. (2001) simularam o genoma de uma única característica, governada por 500 locos, para duas populações, com 500 animais cada e com diferentes herdabilidades para cada população, com o objetivo de verificar o efeito das violações de pressuposições sobre os ganhos genéticos obtidos durante dez gerações de seleção, não verificando efeito das intensidades de erro, aplicadas ao componente de variância genética aditiva, sobre o ganho genético em diferentes distribuições.

Já Carneiro et al. (2001) a fim de compararem o efeito da conexidade de dados sobre o valor fenotípico médio e variância genética aditiva simularam um genoma constituído de uma

característica quantitativa com 15 pares de cromossomos e 500 locos com efeito aditivo distribuídos ao acaso nestes cromossomos. Assim, foram simulados os indivíduos da população inicial que deram origem às próximas gerações, simulando todo o processo reprodutivo, desde a gametogênese até a formação dos novos indivíduos. Concluindo que a conexidade de dados reduz o valor fenotípico médio para características de baixa herdabilidade e tamanho reduzido de progênie.

Utilizando a metodologia de simulação em nível de gene com o objetivo de recuperar dados perdidos pela seleção e assim diminuir o efeito da seleção na avaliação genética, Fonseca (2003) simulou um rebanho bovino com peso a desmama e peso aos 550 dias, sendo estas características controladas por 20 e 40 genes, respectivamente. Na população base foram simulados 600 vacas e 30 touros, que depois de uma geração de acasalamento ao acaso, foram gerados 300 fêmeas e 300 machos e, em cada geração, foram selecionados 20 machos e 100 fêmeas, acasalados ao acaso, produzindo a geração subsequente. Este processo foi repetido por 30 gerações e cada análise foi conduzida na presença ou ausência de pré-seleção no peso à desmama, sendo os dados analisados via BLUP. O autor concluiu que a inclusão das informações de genes identificados não contribuiu para a redução do viés devido à seleção.

Cunha et al. (2006) simularam um genoma contendo 29 pares de cromossomos autossômicos distribuídos por um comprimento total de 3000 cM e por todo o genoma foram alocados, ao acaso, 200 locos quantitativos dialélicos, seguindo distribuição normal uniforme. Os efeitos do ambiente também seguiram distribuição normal. A população-base foi formada por 1000 indivíduos, 500 machos e 500 fêmeas, não aparentados e, pelo acasalamento destes, gerou-se a população inicial com 1500 animais, obtendo-se, assim, o parentesco dos indivíduos. Em seguida, formaram-se os rebanhos que foram submetidos à seleção por cinco gerações consecutivas. Estes autores concluíram que é importante considerar, simultaneamente, o método de seleção bem como o tamanho efetivo da população, no processo seletivo. O método de seleção utilizado foi mais influente sobre os níveis da endogamia do que o próprio tamanho efetivo da população.

Carneiro et al. (2007) avaliaram o coeficiente de endogamia, a fixação de alelos e o limite de seleção em populações selecionadas durante 20 gerações, simulando uma característica quantitativa com herdabilidade igual a 0,10 e um genoma constituído por 200 locos quantitativos polialélicos (oito alelos), com frequências alélicas iniciais baseadas na distribuição uniforme e estes locos foram distribuídos aleatoriamente em 15 pares de cromossomos autossômicos. A resposta à seleção reduziu em longo prazo, devido ao incremento da endogamia, às perdas pela fixação dos alelos e à redução do limite de seleção.

Embora a simulação de dados seja bastante utilizada, parece haver a adoção de um mesmo termo para a classificação de dois procedimentos diferentes de simulação. Neuenschwander et al. (2008) desenvolveram um software de simulação denominada por eles em base no indivíduo, em que valor genético do indivíduo depende da quantidade de locos que irão afetar a característica. Entretanto, essa é a definição e o procedimento de simulação em nível de gene, como denominada por Carneiro Jr. et al. (2004), que consiste em simular os efeitos de cada loco e assim obter o valor genético aditivo do indivíduo, enquanto na simulação em nível de indivíduo consiste em obter apenas o valor genético do indivíduo.

Portanto, percebe-se que não existe um consenso entre os autores sobre a definição dos tipos de simulação, apesar de haver uma definição objetiva dos procedimentos de simulação, dados por Carneiro Jr. et al. (2004). Porém, alguns pesquisadores usam o termo de simulação em nível de indivíduo para descrever a simulação em nível de gene. Além disso, não foi encontrado na literatura recomendações de cenários em que se deva utilizar a simulação em nível de indivíduo ou de gene.

De maneira similar, também não foram encontrados trabalhos descrevendo as diferenças, vantagens e desvantagens sobre estes dois tipos de simulação. Uma análise rigorosa entre os tipos de simulação permite realizar comparações e definições conceituais entre a simulação em nível de indivíduo e de gene. Estas informações são de interesse ao melhorista, pois permitem que se definam cenários em que um tipo de simulação deve ter desempenho superior ao outro, ou ainda em que ambos sejam equivalentes.

3. MATERIAL E MÉTODOS

3.1. Simulação de dados

Os cenários de simulação utilizados foram desenvolvidos por meio do software LZ5 (FONSECA et al., 2010) que é uma ferramenta para a simulação de dados e avaliação genética no melhoramento animal. Este software é composto por 5 módulos, como representado na Figura 1:

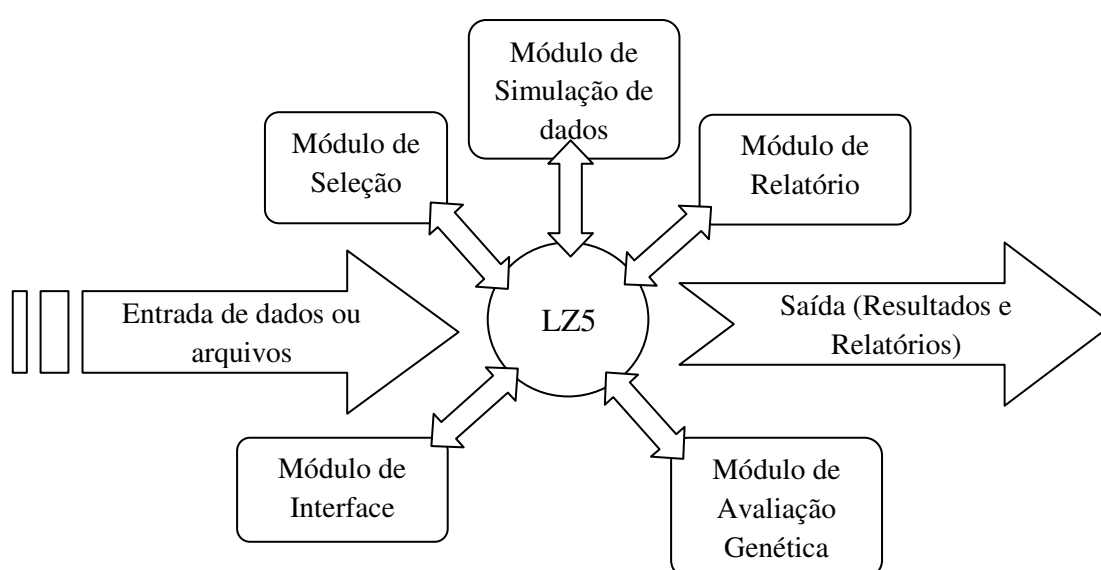


Figura 1 – Representação dos módulos que compõem o simulador LZ5.

O LZ5 foi desenvolvido utilizando a linguagem de programação C++, e o compilador gcc (GNU *Compiler Collection*).

A rotina de simulação, tanto para indivíduo quanto de gene, foi implementada seguindo três passos básicos: 1) geração da população base; 2) processo de seleção e 3) constituição da próxima geração (processo de reprodução), resumidamente descritos a seguir:

Passo 1 - Geração da população base (G_0): os valores genéticos e ambientais foram simulados aleatoriamente, através de uma distribuição normal com média e variância pré-definidas. A soma destes valores resultou nos valores fenotípicos de cada indivíduo.

Passo 2 - Processo de Seleção: inicialmente definiram-se os candidatos à seleção, formando dois vetores, de machos e de fêmeas, com os valores fenotípicos e, ou genéticos. Os candidatos foram selecionados com sobreposição de geração, considerando a geração atual

mais duas gerações anteriores e assim foram selecionados os indivíduos aptos à reprodução, para serem pais da próxima geração.

Passo 3 - Formação da próxima geração (G_{n+1}): este passo iniciou-se com a definição do tipo de acasalamento, preferencial positivo ou negativo, assim, o primeiro macho selecionado se acasalou com duas fêmeas, aleatoriamente, gerando as progênes de acordo com as informações dos pais.

Os passos 2 e 3 foram repetidos até se atingir 15 gerações, incluindo as informações das características de interesse. O mesmo processo de simulação foi repetido por 500 vezes.

3.2. Geração da população-base (G_0)

Para comparação das metodologias de simulação, tanto em nível de indivíduo quanto de gene, foram simuladas três populações, geradas a partir de uma população-base constituída de 1500 indivíduos (750 machos e 750 fêmeas), utilizando características e parâmetros da literatura, de acordo com a Tabela 1.

Tabela 1. Parâmetros genéticos utilizados para simulação de três populações de suínos

| Características | Referência | Herdabilidade (h^2) | Var. Genética Aditiva (σ_a^2) | Média Fenotípica (μ) |
|-----------------------|--------------------|-------------------------|--|----------------------------|
| Peso da leitegada | PIRES et al., 2000 | 0,12 | 0,6 | 15,38 |
| Peso aos 70 dias | GINÉ et al. 2004 | 0,27 | 0,04 | 1,65 |
| Espessura de Toucinho | GINÉ et al. 2004 | 0,47 | 7 | 30 |

Para a simulação em nível de gene, foi considerado o genoma de 2800 cM e 18 cromossomos de tamanhos aleatórios, as características foram governadas por 500 locos poligênicos dialélicos, as frequências alélicas foram iguais e a taxa de recombinação foi de 0,01. Nesta simulação foram ilustrados dois cenários, sem marcadores e com marcadores distanciados igualmente (50cM), sendo gerados 5 QTLs por característica, distanciados aleatoriamente no genoma.

a) Simulação em nível de indivíduo

Depois de definidos os parâmetros iniciais para a geração da população-base, foram amostrados os valores genéticos aditivos de cada indivíduo, seguindo uma distribuição normal com média e variância previamente definidas (Tabela 1). Em seguida foram amostrados os desvios causados pelo ambiente, seguindo também uma distribuição normal, com média zero e variância de acordo com a fórmula a seguir:

$$\sigma_e^2 = \frac{\sigma_a^2}{h^2} - \sigma_a^2$$

em que:

σ_e^2 = variância residual;

σ_a^2 = variância aditiva;

h^2 = herdabilidade.

Considerando o modelo infinitesimal proposto por Fisher (1918), em que os locos têm efeito muito pequeno e a soma desses define o valor genético de um indivíduo, e ainda que os efeitos dos locos têm distribuição desconhecida, podemos afirmar, pelo teorema do limite central, que os valores genéticos tem distribuição normal com média zero e variância conhecida.

Espera-se que a média dos valores genéticos seja zero porque os efeitos de cada loco são, por sua vez, constituídos da soma dos efeitos médios dos alelos, os quais são definidos como desvios da população.

De acordo com Falconer (1996), os efeitos de ambiente são definidos como desvios, considerando a definição de que:

$$P = G + E$$

em que,

P = valor fenotípico simulado;

G = valor genotípico aditivo do animal;

E = desvios causados pelo ambiente.

Pressupondo ausência de dominância e epistasia pode-se reescrever o modelo como:

$$P = A + E$$

em que,

P = valor fenotípico simulado;

A = valor genético aditivo do animal;

E = desvios causados pelo ambiente.

Então,

$$E = P - A$$

E, portanto, os desvios causados pelo ambiente devem ter média zero e variância conhecida. De modo geral, pressupõe-se que os desvios de ambiente também têm distribuição normal. Sendo assim, os valores genéticos aditivos (a) e desvios do ambiente (e) são vetores aleatórios normalmente distribuídos $a \sim N(0, I\sigma_a^2)$ e $e \sim N(0, I\sigma_e^2)$.

O número de valores amostrados sempre foi proporcional ao número de indivíduos a ser gerado para a população-base. Assim, considerando n indivíduos:

$$a = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix} \quad e = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix}$$

Os valores fenotípicos, representados como desvios da média foram obtidos por meio da soma dos vetores a e e . Assim, considerando $p \sim N(0, I\sigma_p^2)$, $p = a + e$ ou:

$$p = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ \vdots \\ p_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ x_3 + y_3 \\ \vdots \\ x_n + y_n \end{bmatrix}$$

Assim, p tem distribuição normal com média 0 e variância σ_p^2 pelo teorema da combinação linear. Este teorema garante que, uma variável aleatória obtida pela soma de duas outras variáveis aleatórias independentes com distribuição normal, também apresenta distribuição normal com sua média e variância obtida da soma das médias e variâncias das variáveis originais.

$$E(p) = E(a) + E(e)$$

$$V(p) = E(a) + E(e)$$

assim,

$$E(p) = 0 + 0 = 0$$

$$V(p) = \sigma_a^2 + \sigma_e^2$$

Considerando, também, que a estrutura de variâncias é $I\sigma_a^2$, pois os indivíduos gerados nessa população são independentes, significando, do ponto de vista genético, que os indivíduos não compartilham alelos e, portanto, não são aparentados e consequentemente não-endogâmicos.

Portanto, para a geração da população-base na simulação em nível de indivíduo, utilizou-se o seguinte algoritmo básico:

1. Amostre um vetor a de números aleatórios a partir de uma distribuição normal com média zero e variância σ_a^2 . O número de amostragens deve ser igual ao número de indivíduos na população;
2. Amostre o vetor e de números aleatórios a partir de uma distribuição normal com média zero e variância σ_e^2 . O número de amostragens deve ser igual ao número de indivíduos na população;
3. Some os dois vetores aleatórios e armazene o resultado em um vetor p .

De modo geral, o número total de amostragens nesse algoritmo é igual a $2n$, em que n é o tamanho da população-base. Considerando cada amostragem e cada operação aritmética básica como um elemento de cálculo (ou de esforço de processador), representado pela constante x , o algoritmo acima demanda $n+n+n$ elementos (operações) de cálculo para ser executado, assim, são necessários $3n$ operações, n amostragens para constituir o vetor a , mais n amostragens para constituir o vetor e e, mais n adições para constituir o vetor p .

Por outro lado, considere que cada variável aleatória represente uma unidade de armazenamento, β , assim o algoritmo demanda $3n$ unidade de armazenamento, sendo parte devido a geração dos valores genéticos aditivos, outra devido a geração dos desvios de ambiente e a terceira devido aos valores fenotípicos.

b) Simulação em nível de gene

Para este tipo de simulação, o valor genético aditivo de cada indivíduo foi calculado pela soma dos efeitos de cada loco poligênico mais os efeitos dos QTLs, quando for o caso. Foram simulados os valores genéticos dos QTLs seguindo uma distribuição normal com média e variância aditiva segundo a Tabela 1. O efeito médio de cada loco poligênico foi obtido a partir da divisão da variância genética aditiva pelo número de locos poligênicos

seguindo o modelo infinitesimal proposto por Fisher (1918). O cálculo dos desvios causados pelo ambiente foi determinado da mesma forma que na simulação em nível de indivíduo.

Portanto, assim como na simulação em nível de indivíduo, para a simulação dos animais na simulação em nível de gene geraram-se os valores fenotípicos (P) das características:

$$P = A + E$$

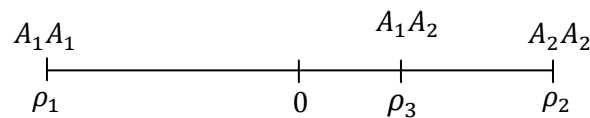
Sendo o valor genético do indivíduo (A) calculado como:

$$A = \rho_1 + \rho_2 + \rho_3 + \cdots + \rho_l = \sum_{i=1}^l \rho_i$$

em que,

ρ_i = soma dos efeitos médios dos dois alelos no loco i , ou a contribuição do loco i para o valor genético aditivo do indivíduo para a característica estudada.

A contribuição de cada loco para o valor genético da característica depende da configuração dos alelos no loco, uma vez que cada alelo contribui, aditivamente, com seu efeito médio para o valor de ρ_i . Assim, consideraram-se dois alelos (A_1 e A_2), com os respectivos efeitos médios associados (m_1 e m_2), então, segundo Falconer (1996):



em que,

$$\rho_1 = m_1 + m_1$$

$$\rho_2 = m_2 + m_2$$

$$\rho_3 = m_1 + m_2$$

Considerou-se ainda efeito de dominância nulo, $\rho_3 = m_1 + m_2 = 0$, existindo, portanto, três contribuições possíveis para cada loco, representadas pelo vetor:

$$c_k = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \end{bmatrix}$$

em que,

c_k = vetor das possíveis contribuições do loco k para o valor genético de um indivíduo.

Para a determinação dos elementos do vetor c_k assumiu-se o modelo infinitesimal. Uma vez que os locos têm contribuições muito pequenas para o valor genético aditivo do indivíduo, assumiu-se que todos os locos contribuíram igualmente à mesma configuração de alelos. Essa simplificação permitiu definir o vetor c_k como a seguir:

$$\sigma_a^2 = 2pq\rho_1^2\lambda$$

em que,

p = frequência do alelo A1;

q = frequência do alelo A2;

λ = coeficiente de endogamia.

E assim, o efeito médio para o valor de ρ_i :

$$\rho_1 = \frac{\sigma_a^2}{\sqrt{2pq\lambda}}$$

Como $\rho_2 = -\rho_1$ então,

$$\rho_2 = -\frac{\sigma_a^2}{\sqrt{2pq\lambda}}$$

$$c = \begin{bmatrix} \frac{\sigma_a^2}{\sqrt{2pq\lambda}} \\ -\frac{\sigma_a^2}{\sqrt{2pq\lambda}} \\ 0 \end{bmatrix}$$

Assim, para a formação de uma população-base de n indivíduos, considerou-se o vetor n dimensional de valores genéticos aditivos a :

$$a = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_n \end{bmatrix}$$

em que,

a_d = soma das contribuições de cada loco que determina a característica.

Assim,

$$a = \begin{bmatrix} \rho_{11} + \rho_{21} + \rho_{31} + \cdots + \rho_{l1} \\ \rho_{12} + \rho_{22} + \rho_{32} + \cdots + \rho_{l2} \\ \rho_{13} + \rho_{23} + \rho_{33} + \cdots + \rho_{l3} \\ \vdots \\ \rho_{1n} + \rho_{2n} + \rho_{3n} + \cdots + \rho_{ln} \end{bmatrix}$$

A contribuição ρ_{ij} para cada indivíduo foi definida a partir do vetor c e depende da configuração dos alelos. Assim, se para um indivíduo qualquer a configuração de um loco é A_1A_1 , a contribuição atribuída a esse loco foi ρ_1 ; se A_1A_2 , a contribuição foi ρ_2 ; e, se A_2A_2 , a contribuição foi ρ_3 .

Uma vez que os efeitos médios dos locos são, por definição, desvios em relação a um grupo específico de indivíduos (Falconer, 1996), também são contribuições ρ_i e, portanto, os valores genéticos aditivos também são representados como desvios e assim $E(a) = 0$. Como o elemento do vetor a é a soma de contribuições dos locos, pelo teorema do limite central, espera-se que a seja normalmente distribuído e, portanto $a \sim N(0, I\sigma_a^2)$. Os desvios de ambiente, assim como na simulação em nível de indivíduo, foram amostrados de acordo com $e \sim N(0, I\sigma_e^2)$. Então, o vetor de valores fenotípicos (p) foi calculado da mesma forma:

$$p = a + e$$

e

$$p = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ \vdots \\ p_n \end{bmatrix} = \begin{bmatrix} a_1 + e_1 \\ a_2 + e_2 \\ a_3 + e_3 \\ \vdots \\ a_n + e_n \end{bmatrix} = \begin{bmatrix} (\rho_{11} + \rho_{21} + \rho_{31} + \cdots + \rho_{l1}) + e_1 \\ (\rho_{12} + \rho_{22} + \rho_{32} + \cdots + \rho_{l2}) + e_2 \\ (\rho_{13} + \rho_{23} + \rho_{33} + \cdots + \rho_{l3}) + e_3 \\ \vdots \\ (\rho_{1n} + \rho_{2n} + \rho_{3n} + \cdots + \rho_{ln}) + e_n \end{bmatrix}$$

O algoritmo para geração da população-base de n indivíduos foi definido como:

1. Amostre n desvios de uma distribuição normal com média zero e σ_e^2 e armazene-os no vetor e ;
2. Para cada loco de um indivíduo amostre dois números entre os inteiros 0 e 1, um que representa o alelo A_1 e outro o alelo A_2 e armazene-os em um vetor g . Se o resultado da amostragem for $g=[0 \ 0]$ atribua ρ_1 ao loco; se $g=[1 \ 1]$, atribua ρ_3 ; e ρ_2 , caso contrário;
3. Para cada indivíduo, some as contribuições de todos os locos e armazene no vetor a ;
4. Obtenha p a partir da soma de $a+e$.

Esse algoritmo requer n amostragens para o vetor e , e mais n operações para o vetor p . Portanto, para simulação dos valores genéticos para cada loco (l) foram necessárias duas amostragens para definir a configuração dos alelos e, portanto, o número de amostragens para cada indivíduo foi $2l$.

Após definições de configurações, as operações de associação do efeito com a configuração requerem l operações e mais l operações de adição foram necessárias para se gerar o valor genético aditivo.

Assim, para um indivíduo, $4l$ operações são necessárias para gerar o valor genético. Em uma população de n indivíduos $4ln$ operações são necessárias.

Portanto, a complexidade do algoritmo é linear, significando que quanto maior o tamanho da população e o número de locos o esforço de processamento aumenta na proporção de $4ln$.

Nesta simulação em nível de gene, necessita-se armazenar os vetores g , a , e e p , e também os valores de ρ_i , e, considerando que cada constante ocupa uma unidade de memória, então para uma população de n indivíduos e l locos determinando uma característica precisa-se de $n+n+n+nl+2nl$ unidades de memória, que equivale a $3n(l+1)$. Portanto, o consumo de memória para armazenar uma simulação é, aproximadamente, três vezes o tamanho da população vezes o número de locos.

A população-base não possui indivíduos aparentados ou endogâmicos e, portanto, não corresponde a uma situação real, onde as populações já evoluíram por muitas gerações e possuem pedigrees complexos e diferentes graus de endogamia entre os indivíduos.

Assim, a população-base evolui por algumas gerações. Os próximos passos foram:

- a) Definir os candidatos à seleção;
- b) Realizar seleção;
- c) Definir os tipos de acasalamento;
- d) Gerar novos indivíduos.

3.3. Definição dos candidatos à seleção

Foi implementado o método de seleção individual com base no fenótipo para ambos os tipos de simulação. Selecionaram-se 50 machos e 300 fêmeas a partir dos vetores contendo os candidatos à seleção com seus respectivos fenótipos.

Os candidatos à seleção foram definidos pelo número de gerações consideradas (atual e as duas anteriores a ela). Assim, para cada categoria de sexo, considerou-se uma série de vetores, g_i de dimensões n_i , em que $i = 0, 1, 2, \dots, t$; t é o índice que representa a geração atual. Desta forma, cada geração pode ser representada por dois vetores, um para cada sexo:

$$G_{i_m} = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_m \end{bmatrix} \quad G_{i_f} = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_f \end{bmatrix}$$

em que,

m e f são respectivamente o número de machos e fêmeas da geração;

G_{i_m}, G_{i_f} são os vetores contendo as informações (por exemplo, valor fenotípico, valor genético aditivo, índices etc.) de machos e fêmeas respectivamente.

Os candidatos a seleção foram escolhidos através de gerações sobrepostas, significando que foram escolhidos da geração atual e de k gerações anteriores. Assim, se a escolha dos candidatos acontece em gerações não-sobrepostas, somente os vetores G_T (para machos e fêmeas) devem ser considerados, se os candidatos a serem selecionados pertencem a i gerações sobrepostas, então o seguinte conjunto de vetores deve ser selecionado para cada um dos sexos:

$$T = \{G_i / i = 1, 2, \dots, t\}$$

O conjunto T também pode ser representado por um vetor $C_{(gxn),1}$ em que:

$$C = \begin{bmatrix} G_0 \\ G_1 \\ \vdots \\ G_t \end{bmatrix}$$

Expandindo o vetor C , para um dos sexos:

$$C = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_{n_1} \\ \hline \bar{s}_1 \\ s_2 \\ \vdots \\ s_{n_2} \\ \hline \vdots \\ \hline s_1 \\ s_2 \\ \vdots \\ s_{n_t} \end{bmatrix}$$

O algoritmo para definição dos candidatos é o mesmo para os procedimentos em nível de indivíduo ou gene e pode ser descrito como:

1. Defina o conjunto T por meio do número de gerações das quais os indivíduos são candidatos a seleção;
2. Concatene por colunas os elementos de T em ordem crescente dos valores de i ;

O vetor C resultante deve ter dimensão $(\sum r_i)$, em que:

r_i = o número de linhas do vetor G_i .

Considerando uma única geração, o vetor G_i ocupará uma quantidade de memória que é proporcional ao número de elementos, n . Por sua vez, os elementos s_i são conjuntos de informações sobre os indivíduos e ocupará mais memória quanto maior for a quantidade de informações que eles armazenam, digamos z .

Assim, considerando uma geração qualquer, a quantidade de memória ocupada, M , é proporcional ao número de indivíduos e quantidade de informações sobre eles:

$$M \propto nz$$

Para g gerações a quantidade de memória usada é:

$$M \propto gnz$$

E uma vez que os dois vetores C precisam ser gerados, um para cada sexo:

$$M \propto 2gnz$$

Uma vez que nova informação deve ser gerada nesse processo, todo o esforço computacional é dirigido para a execução de cópias para a formação dos vetores C . Assim, se uma operação de cópia é considerada um elemento de esforço, a quantidade de cópias a serem realizadas é definida da mesma forma que para o cálculo da quantidade de memória, ou seja:

$$E \propto gnz$$

em que,

E = total de operações de cópia realizadas pelo processador.

Assim, o algoritmo para geração de candidatos à seleção tem complexidade linear.

3.4. Realização da seleção

O procedimento para a execução da seleção foi o mesmo para os dois tipos de simulação. Assim, considerou-se que foram selecionados m machos e f fêmeas para formação da geração i .

Considerou-se também o critério de seleção cs , em que cs pode ser, por exemplo, o valor fenotípico, ou o valor genético aditivo ou ainda um índice qualquer associado aos indivíduos.

Para a execução da seleção, os candidatos foram ordenados de acordo com o critério cs . Consideraram-se então os vetores ordenados dos candidatos à seleção:

$$C_m = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_{n_m} \end{bmatrix} \quad C_f = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_{n_f} \end{bmatrix}$$

em que,

C_m = vetor dos candidatos à seleção machos;

C_f = vetor dos candidatos à seleção fêmeas.

Se o desejável é que o critério cs seja o maior possível, então os primeiros indivíduos dos vetores C_m e C_f , representados pelo conjunto s_l são os que possuem o maior valor para o cs .

A seleção de m machos e f fêmeas foi executada pela escolha dos m primeiros elementos do vetor C_m e dos f primeiros elementos do vetor C_f , os quais podem ser transferidos para dois vetores de selecionados, diga-se, S_m e S_f :

$$s_m = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_m \end{bmatrix} \quad s_f = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_f \end{bmatrix}$$

O processo de seleção demandou memória adicional para armazenamento dos indivíduos selecionados, vetores s_m e s_f . A memória requerida M é diretamente proporcional ao número de conjuntos s_i selecionados:

$$M = 2n_s z$$

em que,

n_s = número total de indivíduos selecionados.

O esforço total de processamento E nesse algoritmo foi em função das demandas de ordenação dos vetores de candidatos e de cópias dos selecionados para as respectivas estruturas de armazenamento.

O número de operações de cópia, considerando um elemento de esforço, foi proporcional ao número de indivíduos copiados, portanto:

$$E_c = 2n_s z$$

O cálculo de esforço devido ao ordenamento dos indivíduos é complexo e depende do tipo de algoritmo de ordenação utilizado. Por questões de simplificação foi considerada uma relação linear entre o número de indivíduos a ser ordenados e a quantidade de ordenações de fato realizadas. Assim:

$$E_0 \approx 2n$$

Portanto, o esforço total realizado é:

$$E \approx 2n_s z + 2n$$

A quantidade de memória requerida foi relativamente pequena, uma vez que dependeu, principalmente, do número de selecionados. Por outro lado, o esforço computacional pode ser significativo se considerarmos grande número de candidatos a serem ordenados para fins de seleção, pois existe dependência direta do número de candidatos à seleção n .

3.5. Formação da próxima geração (G_{n+1})

Assim que foram selecionados os reprodutores, foi realizado o acasalamento entre os indivíduos, de forma aleatória, seguindo os métodos reprodutivos descritos para cada tipo de simulação, em nível de indivíduo ou de gene.

a) Simulação em nível de indivíduo

Para este tipo de simulação, o valor genético aditivo de uma progênie foi determinado pela soma da metade dos valores genéticos dos pais, e somado a um valor aleatório determinado como segregação amostral mendeliana, que consiste em uma amostragem aleatória dos genes dos pais causada pela segregação independente dos genes durante a formação das células germinativas e pela seleção aleatória de gametas na formação do embrião (BOURDON, 2000). Como nesta metodologia têm-se apenas os valores genéticos aditivos dos pais, para simular possíveis recombinações, interações epistáticas e outros fatores genéticos que podem influenciar na expressão do valor fenotípico, foram realizadas amostragens dos efeitos da segregação dos alelos.

Assim, a geração de novos indivíduos envolveu a criação de gametas pelos pais e a amostragem de dois deles, um do pai e outro da mãe, que se juntaram para restabelecer o número diplóide e formando a progênie. Durante o processo de gametogênese pode ocorrer o evento de recombinação e mudar a configuração original dos cromossomos paternos e maternos das células germinativas, produzindo gametas recombinantes. Além disso, os cromossomos foram amostrados de maneira aleatória durante o processo de meiose, sendo que cada gameta possui uma amostra de metade do material genético do genitor (amostragem mendeliana).

No novo indivíduo, diferentes configurações de alelos no genoma, estão presentes devido aos processos de recombinação e amostragem ocorridos durante a gametogênese nos pais. A consequência desse fato é que a progênie, de modo geral, tem seu valor genético aditivo divergente da média do valor genético aditivo de seus pais, ou seja:

$$a_i \neq \frac{1}{2}(a_s + a_d)$$

em que,

a_i = valor genético aditivo da progênie;

a_s = valor genético aditivo do pai;

a_d = valor genético aditivo da mãe.

Dessa forma, o modelo que explica a desigualdade apresentada pode ser escrito como:

$$a_i = \frac{1}{2}(a_s + a_d) + \Phi$$

em que:

Φ = efeito da amostragem mendeliana.

Portanto, Φ pode ser escrito como uma diferença entre a média dos valores genéticos aditivos dos pais e o valor genético aditivo do indivíduo:

$$\Phi = a_i - \frac{1}{2}(a_s + a_d)$$

Assim, pode-se escrever que:

$$\begin{aligned} E(\Phi) &= E\left[a_i - \frac{1}{2}(a_s + a_d)\right] \\ &= E(a_i) - E\left[\frac{1}{2}(a_s + a_d)\right] \\ &= 0 - \frac{1}{2}E[(a_s + a_d)] \\ &= 0 - \frac{1}{2}[E(a_s) + E(a_d)] \\ &= 0 - \frac{1}{2}[0 + 0] \\ E(\Phi) &= 0 \end{aligned}$$

da mesma forma, pode mostrar que:

$$\begin{aligned} V(\Phi) &= V\left[a_i - \frac{1}{2}(a_s + a_d)\right] \\ &= V(a_i) - \frac{1}{4}V(a_s + a_d) + 2cov\left[a_i, \frac{1}{2}(a_s + a_d)\right] \\ &= V(a_i) - \frac{1}{4}[V(a_s) + V(a_d) + 2cov(a_s, a_d)] + cov[a_i, (a_s + a_d)] \\ &= V(a_i) - \frac{1}{4}V(a_s) - \frac{1}{4}V(a_d) - \frac{1}{2}cov(a_s, a_d) + cov(a_i, a_s) + cov(a_i, a_d) \end{aligned}$$

Considerando que $V(a_i) = V(a_s) = V(a_d)$,

$$V(\Phi) = V(a_i) - \frac{1}{4}V(a_i) - \frac{1}{4}V(a_i) - \frac{1}{2}cov(a_s, a_d) + cov(a_i, a_s) + cov(a_i, a_d)$$

Após devidas manipulações pode-se mostrar que:

$$V(\Phi) = \frac{1}{2}V(a_i)(1 - \bar{F})$$

em que,

\bar{F} = coeficiente médio de endogamia dos pais do indivíduo i , ou seja, $\frac{1}{2}(F_d + F_s)$.

Uma vez que na simulação em nível de indivíduo, genomas não estão disponíveis, os efeitos de amostragem de cromossomos e recombinações para a formação dos gametas devem ser simulados como um efeito conjunto Φ .

De acordo com a equação que descreve o valor genético aditivo do indivíduo, dependente da segregação amostral mendeliana, Φ é uma combinação linear dos valores genéticos aditivos do indivíduo e de seus pais. Assumindo que os valores genéticos possuem distribuição normal, então pelo teorema da combinação linear:

$$\Phi \sim N\left[0, \frac{1}{2}\sigma_a^2(1 - \bar{F})\right]$$

Para cada indivíduo a ser simulado, um efeito de amostragem mendeliana deve ser amostrado da distribuição de probabilidade a cima. Portanto o número de amostragens é proporcional ao número de indivíduos (n).

Assim, se n_i novos indivíduos devem ser gerados, o conjunto M , deve ser formado:

$$M = \{m_i: i = 1, 2, \dots, n_i\}$$

Considerando agora que existe também um conjunto A_m , de médias de valores genéticos aditivos dos pais dos indivíduos:

$$A_m = \left\{\frac{1}{2}(a_{si} + a_{di}): i = 1, 2, \dots, n_i\right\}$$

Os valores genéticos aditivos dos novos indivíduos foram obtidos por:

$$A = A_m + M$$

Se representarmos os conjuntos como vetores:

$$A = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_{n_1} \end{bmatrix} = \begin{bmatrix} 1/2 (a_{s1} + a_{d1}) \\ 1/2 (a_{s2} + a_{d2}) \\ 1/2 (a_{s3} + a_{d3}) \\ \vdots \\ 1/2 (a_{sn_1} + a_{dn_1}) \end{bmatrix} + \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ \vdots \\ m_{n_1} \end{bmatrix}$$

Portanto, na simulação em nível de indivíduo o seguinte algoritmo foi utilizado:

Para cada novo indivíduo a ser gerado:

1. Calcule a quantidade: $1/2 (a_{s_i} + a_{d_i})$;
2. Calcule a quantidade: $\bar{F} = 1/2 (F_{s_i} + F_{d_i})$;
3. Amostre m de (S);
4. Some as quantidade dos passos 1 e 3 e armazene no vetor a .

De modo geral, não há interesse no armazenamento dos conjuntos A_m e M mas somente na formação de A .

O cálculo de \bar{F} , no passo 2, requer informação dos coeficientes de endogamia dos pais. Esses devem ser obtidos previamente por algum algoritmo (metodologia) disponível para este fim.

Agora, se cada operação de amostragem e soma puder ser considerada como um esforço de processamento, x , então o algoritmo, desconsiderando os esforços demandados para o cálculo de F_{s_i} e F_{d_i} , demanda $n_1x + n_1x + n_1x + n_1x$, ou seja, $4(n_1x)$ esforços de processamento para os passos 1, 2, 3 e 4 respectivamente.

Portanto pode-se dizer que $Tp \propto n_1$, uma vez que $4x$ é constante para esse algoritmo e Tp = tempo de processamento demandado pelo algoritmo.

Se os esforços demandados pelo cálculo de F_{s_i} e F_{d_i} forem considerados, então, o passo 2 demanda n_x unidades de esforço de processament devido a soma e mais U unidades de esforço devido ao algoritmo para o cálculo de F_{s_i} e F_{d_i} .

Normalmente, como pode ser visto por Quaas (1976) a quantidade de processamento depende do tamanho da população e portanto $U \propto n$, em que n é o tamanho da população antes da formação de n_1 .

Assim a equação do tempo de processamento pode ser reescrita como:

$$Tp \propto n_1 + n$$

Com relação a quantidade de memória o algoritmo mostrado demanda somente o armazenamento do conjunto A para cada geração de novos indivíduos, e portanto a quantidade de memória RAM é proporcional a quantidade de indivíduos novos gerados:

$$\beta_1 = n_1 b$$

em que,

β_1 = quantidade de memória demandada na geração 1;

n_1 = número de indivíduos novos formados;

b = unidade de memória.

Assim, o possível algoritmo para o acasalamento é:

1. Defina, de acordo com algum critério os pares de conjuntos de machos e fêmeas a serem acasalados;
2. Embaralhe os pares dos vetores (m_i e f_j);
3. Escolha o primeiro macho do m_i e acasale com as f primeiras fêmeas de f_j . Retire-os do conjunto;
4. Repita o procedimento até que todos os pares (m_i e f_j) tenham sido utilizados, ou seja, sejam conjuntos vazios.

Uma vez que esses procedimentos podem ser executados nos próprios vetores selecionados, nenhuma memória adicional é demandada, entretanto, o caso é diferente para processamento, pois são exigidos muitos esforços computacionais no momento do embaralhamento dos indivíduos. O esforço é maior quanto maior o tamanho do “objeto” a ser embaralhado. Portanto, na simulação em nível de gene demanda mais esforço quando comparado com a simulação em nível de indivíduo.

b) Simulação em nível de gene

Uma vez que os cromossomos e os locos são criados, é possível simular a gametogênese considerando a amostragem de cromossomos e eventos de recombinação e portanto, não é necessário a simulação do efeito da amostragem mendeliana, Φ , como mostrado na simulação em nível de indivíduo.

Suponha então que um cromossomo qualquer possa ser representado como um conjunto de locos, diga-se, H_i . Assim,

$$H_i = \{L_j: j = 1, 2, \dots, l_i\}$$

em que,

H_i = i-ésimo cromossomo do genoma;

L_j = j-ésimo loco do cromossomo;

l_i = número de locos do cromossomo i.

Os locos dentro de cada cromossomo são também conjuntos de informações:

$$L_j = \{x_i: i = 1, 2, \dots, q\}$$

em que,

x_i = i-ésima informação do loco

q = número de informações do conjunto L_j .

Para a descrição da simulação da gametogênese, basta considerar um loco como um conjunto de alelos do loco:

$$L_j = \{x_1\}$$

em que,

x_1 = conjunto de dois elementos, cada um representando o estado de um alelo, ou seja,

$$x_1 = \{A_i, A_j: i = 1, 2, \dots, m; j = 1, 2, \dots, m\}$$

em que,

m = número de alelos.

Portanto, H_i , pode ser representado como:

$$H_i = \{(A_i, A_j), (B_i, B_j), (C_i, C_j), \dots, (Z_i, Z_j)\}$$

Percebe-se então que, de fato, H_i , representa o par de cromossomos homólogos, onde o primeiro elemento de x_i é o alelo do primeiro homólogo e o segundo elemento é a representação do alelo do segundo homólogo. Essa idéia é melhor ilustrada na Figura 2.

$$H_i = \{(A_1, A_2), (B_1, B_2), (C_1, C_2)\}$$

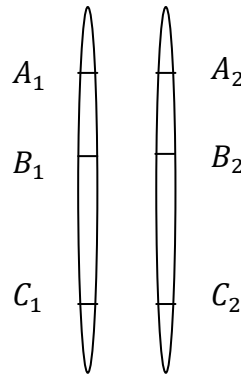


Figura 2 – Representação do conjunto de cromossomos homólogos.

No momento da meiose, os homólogos se pareiam e recombinações podem ocorrer em qualquer posição nos cromossomos. Neste trabalho a probabilidade de ocorrência de um evento de recombinação foi a mesma para todo o genoma (0,01):

$$P(recomb) = k$$

Um teste foi executado a cada loco do cromossomo por meio da amostragem de um número aleatório, entre 0 e 1, de uma distribuição uniforme (r). Assim, se:

$r > k$ não ocorre recombinação;

$r < k$ ocorre recombinação.

Sempre que um evento de recombinação ocorrer, os fragmentos dos homólogos daquela posição até seus finais devem ser trocados:

$$H_i = \{(A_i, A_j), (B_j, B_i), (C_j, C_i), \dots, (Z_j, Z_i)\}$$

↑
recombinação

Se um novo evento de recombinação ocorrer então:

$$H_i = \{(A_i, A_j), (B_j, B_i), (C_i, C_j), \dots, (Z_i, Z_j)\}$$

↑
recombinação

O efeito da amostragem dos alelos nos gametas é devido a separação independente dos cromossomos homólogos durante a meiose. Esse efeito pode ser simulado pela escolha aleatória de um dos homólogos para cada par de cromossomos do genoma, o que pode ser feito pela amostragem de um número aleatório, r , uniformemente distribuído entre 0 e 1, para cada cromossomo. Se:

$r > 0,5$, é escolhido o homólogo paterno;

$r \leq 0,5$, é escolhido o homólogo materno.

Se cada homólogo puder ser representado como H_{io} , em que $o = s$ ou d , referindo-se aos homólogos de origem paterna e materna, respectivamente, então um gameta pode ser descrito como:

$$G = \{h_{io} : i = 1, 2, \dots, c \text{ e } o = s \text{ ou } d\}$$

em que,

c = o número de pares de cromossomos no genoma.

Considerou-se que cada indivíduo possui dois cromossomos em seu genoma, H_1 e H_2 , e que o gameta que dará origem a um novo indivíduo deve ser informado, então, por exemplo:

$$H_1 = \{(A_1, A_2), (B_1, B_2), (C_1, C_2)\}$$

$$H_2 = \{(D_1, D_2), (E_1, E_2)\}$$

A Figura 3 representa a formação de um gameta considerando a amostragem dos cromossomos.

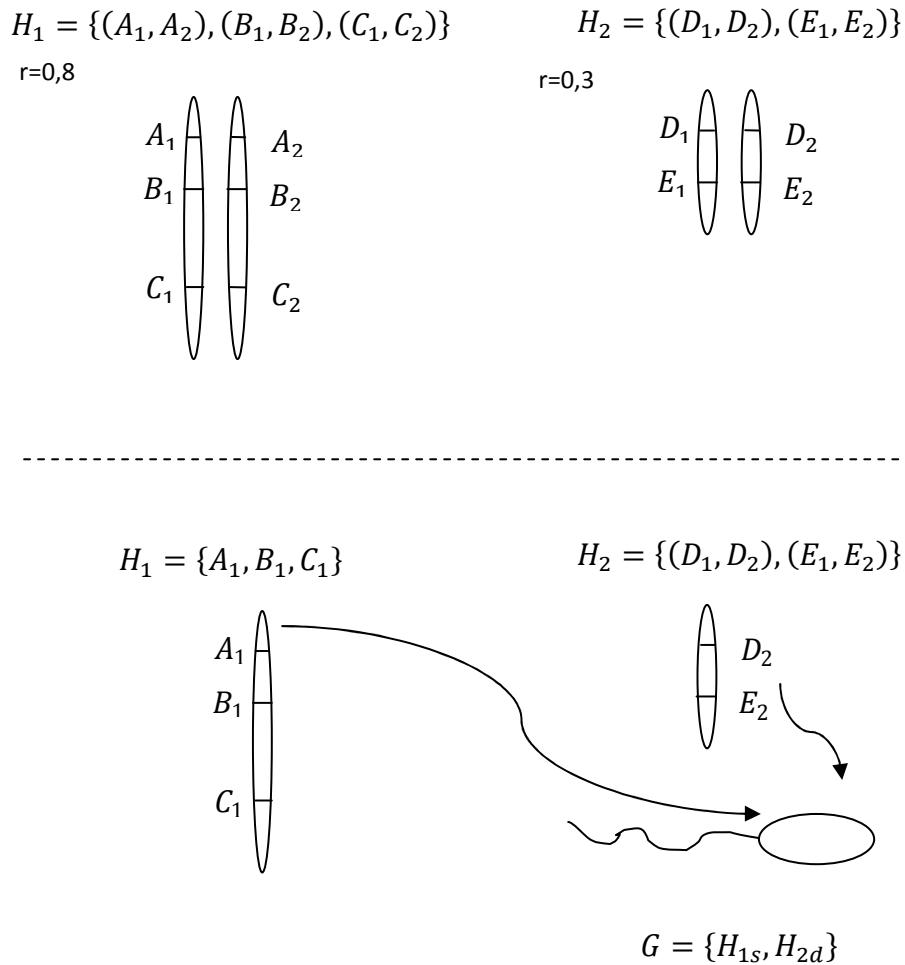


Figura 3 – Representação da gametogênese em que H_1 , ocorreu a escolha da cromátide paterna ($r=0,8$) e H_2 , ocorreu a escolha da cromátide materna ($r=0,3$).

O processo de gametogênese descrito ocorre em ambos os pais e os gametas formados foram reunidos em um novo indivíduo. Assim, pode-se representar o genoma do novo indivíduo como:

$$M = \{H_i: i = 1, 2, \dots, c\}$$

em que,

M = conjunto de pares de cromossomos de um indivíduo;

H_i = i -ésimo par de cromossomo, onde:

$$H_i = \{(a, b): a \in H_{is}, b \in H_{id}\}$$

O par (a, b) é, de fato, a informação de configuração de alelos do loco, x_i .

Finalmente, para a simulação do valor genético de um indivíduo, foi considerado que um loco qualquer, como definido na equação da configuração dos locos dentro de cada

cromossomo (h_j), foi um conjunto de duas informações, x_1 e x_2 , em que x_1 é a informação de configuração dos alelos e x_2 representa o valor genético aditivo do loco, ou seja, a soma dos efeitos médios dos alelos representados em x_1 .

Assim, considerando as equações da configuração dos locos dentro de cada cromossomo (h_j) e o genoma do indivíduo (M), o valor genético aditivo foi calculado como:

$$a = \sum_{i=1}^c \sum_{j=1}^{l_i} x_{ij}$$

em que,

a = valor genético aditivo do indivíduo;

x_{ij} = valor genético aditivo do j -ésimo loco do i -ésimo cromossomo.

O algoritmo para simulação de um novo indivíduo e definição de seu valor genético é:

Para ambos os pais do novo indivíduo a ser gerado:

1. Amostre um número aleatório entre 0 e 1, de uma distribuição uniforme (r);
 - 1.1 Se $r \geq 0,5$, copie o elemento a_1 do primeiro loco de H_{io} do pai para o conjunto H_{io} do indivíduo;
 - 1.2 Se $r < 0,5$, copie o elemento b_1 do primeiro loco de H_{io} do pai para o conjunto H_{io} do indivíduo.
2. Para o próximo loco (ou elemento de H_{io}) amostr um número aleatório, entre 0 e 1, de uma distribuição uniforme (r_1);
 - 2.1. Se $r_1 > k$, então copie para H_{io} , do indivíduo, o elemento (a_j ou b_j) correspondente à mesma posição em x_i do loco anterior;
 - 2.2. Se $r_1 < k$, então copie para H_{io} , do indivíduo, o elemento (a_j ou b_j) da outra posição em x_i em relação à escolha feita no loco anterior;
3. Repita o passo 2, 2.1 e 2.2 até o final dos cromossomos dos pais.

Os esforços de processamento (p) envolvidos em cada passo do algoritmo são:

1. Amostragens de números aleatórios para cada par de cromossomo do genoma do indivíduo.
2. Para cada loco a ser formado no novo indivíduo, uma cópia do alelo de ambos os pais deve ser feito nos conjuntos H_{io} do indivíduo para a formação dos homólogos de origem paterna e materna.

3. Para cada loco um número aleatório deve ser gerado para o teste de recombinação e o teste deve ser realizado.

Assim, considerando que os esforços de cópia, geração de número aleatório e testes demandam os mesmos esforços de processamento, então podemos dizer que o total de esforços de processamento (p) é proporcional ao tempo de processamento (T_p).

Portanto, se pretendemos criar n_i novos indivíduos, onde o genoma é constituído por c pares de homólogos e l locos, os esforços de processamento por etapas são:

1. $P_1 \propto n_1 c$, ou seja, os esforços de processamento do passo 1 do algoritmo (P_1) é proporcional ao número de indivíduos a ser gerado e ao número de pares de cromossomos no genoma.
2. $P_2 \propto 2n_1 l$, por indivíduo, duas cópias de configuração de alelos deve ser feita por loco.
3. $P_3 \propto n_1 (l - 1)$, o número de amostragens e testes é proporcional ao número de loco de cada indivíduo, exceto para a primeira, por isso o número de esforços é proporcional a $(l-1)$.

Assim, o total de esforços de processamento é:

$$\begin{aligned}
 P &= P_1 + P_2 + P_3 \\
 &= n_1 c + 2n_1 l + 2n_1 (l - 1) \\
 &= n_1 (c + 2l + 2(l - 1)) \\
 &= n_1 (c + 2(l + l - 1)) \\
 &= n_1 (c + 2(2l - 1))
 \end{aligned}$$

A expressão anterior pode ser simplificada se notar que em relação ao número de locos, as quantidades c e -1 são relativamente menos significativas, ou seja:

$$\begin{aligned}
 \frac{c}{l} &\rightarrow 0 \\
 \frac{-1}{l} &\rightarrow 0
 \end{aligned}$$

E, portanto, a expressão pode ser reescrita como:

$$\begin{aligned}
 P &= n_1 \times 4l \\
 &= 4n_1 l
 \end{aligned}$$

Percebe-se que todo o esforço computacional vem dos passos que trabalham sobre os locos, ou seja, os passos 2 e 3, e que a quantidade de operações é 4 vezes o número de loco.

O consumo de memória RAM é exclusivamente devido a formação do novo genoma do novo indivíduo e, portanto, depende do número de locos simulados. Assim, a quantidade de memória RAM demandada (R) pode ser representada como:

$$R \propto n_1 lb_i$$

em que,

b_i = unidade de memória.

3.6. Formação e armazenamento de uma população

A estrutura básica de armazenamento da simulação em nível de indivíduo seguiu o esquema representado na Figura 4, onde a classe *população* armazenou os indivíduos de cada geração e cada *indivíduo* armazenou suas informações genéticas.

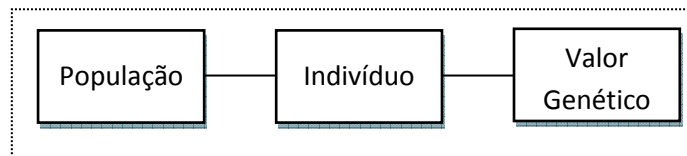


Figura 4 – Representação das classes de estrutura da metodologia de simulação em nível de indivíduo.

Já a simulação em nível de gene a estrutura de armazenamento é bem mais complexa (Figura 5). Criou-se um vetor da população, esta população armazena os indivíduos de cada geração, estes indivíduos foram constituídos de um objeto genoma, que por sua vez armazena informações dos cromossomos, que armazenava informações dos locos e por fim os alelos que compõe cada loco.

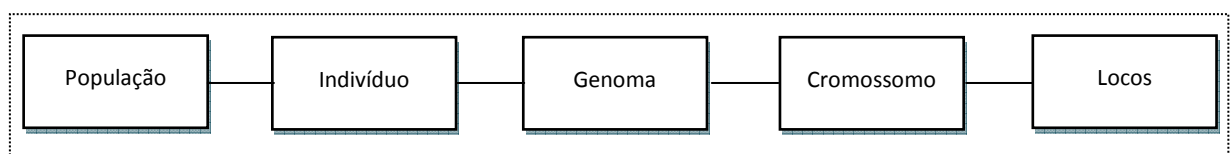


Figura 5 – Representação das classes de estrutura da metodologia de simulação em nível de gene.

3.7. Formação dos gametas e de um novo indivíduo

A programação do algoritmo para formação dos gametas, na simulação em nível de indivíduo, consistiu em apenas dividir pela metade o valor genético dos progenitores selecionados e em seguida realizou-se uma amostragem da segregação amostral Mendeliana. Portanto, para esta metodologia, necessitou-se apenas realizar quatro cálculos básicos e uma amostragem da segregação e o cálculo do coeficiente de parentesco dos pais.

Para a simulação em nível de gene, necessitou-se percorrer todos os cromossomos de cada progenitor e a cada alelo realizou-se uma amostragem que verifica a ocorrência de recombinação na posição. Depois de formados o gameta dos pais, realizou-se a zigotogênese (união dos gametas). Depois de formado os indivíduos percorrem-se todos os alelos para calcular o valor genético aditivo do indivíduo.

3.8. Simulação de QTLs e Marcadores

Além da simulação em nível de gene sem marcadores, foi considerado também cenário considerando os marcadores dialélicos. Estes marcadores foram distribuídos em 18 cromossomos autossômicos, seguindo a distância média de 50 cM partindo da posição 0 cM. Cinco QTLs, também dialélicos, foram distribuídos aleatoriamente por todo o genoma de 2800 cM.

Não foi realizada a simulação de marcadores e QTLs na simulação em nível de indivíduo, pois seguindo a descrição desta metodologia, este tipo de simulação amostra apenas o valor da soma dos efeitos de cada loco, não sendo possível gerar posições ou efeitos de locos individuais. Se for amostrado o efeito médio do QTL a simulação estaria dividindo efeitos do genoma do indivíduo, iniciando uma simulação em nível de gene, o que não se enquadra neste processo de simulação.

4. RESULTADOS E DISCUSSÃO

Os resultados das simulações (indivíduo, gene sem marcador e gene com marcador) para h^2 de 0,12, são apresentados na Tabela 2 e destacados na Figura 6. Os valores amostrados apresentaram um comportamento bem semelhante para os três tipos de simulação, apresentando um aumento da variância aditiva nas primeiras gerações, seguindo de um decréscimo. Para a simulação em nível de indivíduo, a máxima variância aditiva ocorreu na

terceira geração, sendo que houve uma diminuição mais rápida quando comparado com as simulações em nível de gene com ou sem marcador. Sugere-se que este comportamento das variâncias são explicadas pela baixa relação de parentesco entre os indivíduos das primeiras gerações, portanto existe uma baixa relação de alelos idênticos por ascendência, mesmo tendo a possibilidade de se ter alelos idênticos por posição, assim as variâncias apresentam um comportamento inesperado nas primeiras gerações e em seguida apresentam o comportamento adequado para populações em que está praticando seleção.

O mesmo comportamento das variâncias aditivas foi encontrado para o parâmetro herdabilidade da característica, porém a queda ao longo das gerações foi menor, como mostra a Figura 6b. Este comportamento ocorreu devido a herdabilidade ser um coeficiente genético que expressa a relação entre a variância genética aditiva e a variância fenotípica, ou seja, mede o nível de correspondência entre o valor fenotípico e o valor genético aditivo.

Tabela 2. Variâncias aditiva (σ_a^2) e fenotípica (σ_p^2), média fenotípica da população (μ) e herdabilidades (h^2) de uma população utilizando as metodologias de simulação em nível de indivíduo e de gene (com e sem marcador) para uma característica com h^2 de 0,12

| Ger* | Simulação em nível de Indivíduo | | | | Simulação em nível de Gene | | | | | | | |
|------|---------------------------------|--------------|-------|-------|----------------------------|--------------|-------|-------|--------------------|--------------|-------|-------|
| | | | | | Sem Marcador e QTL | | | | Com Marcador e QTL | | | |
| | σ_a^2 | σ_p^2 | μ | h^2 | σ_a^2 | σ_p^2 | μ | h^2 | σ_a^2 | σ_p^2 | μ | h^2 |
| 0 | 0,62 | 5,37 | 15,31 | 0,12 | 0,57 | 4,82 | 15,28 | 0,12 | 0,57 | 5,31 | 15,44 | 0,11 |
| 1 | 0,90 | 5,25 | 15,75 | 0,17 | 0,84 | 5,06 | 15,82 | 0,17 | 0,95 | 5,48 | 15,81 | 0,17 |
| 2 | 0,84 | 5,46 | 16,38 | 0,15 | 0,76 | 5,04 | 16,26 | 0,15 | 0,85 | 5,40 | 16,34 | 0,16 |
| 3 | 1,01 | 5,54 | 16,80 | 0,18 | 0,72 | 5,27 | 16,47 | 0,14 | 0,83 | 5,34 | 16,87 | 0,16 |
| 4 | 0,95 | 5,32 | 17,33 | 0,18 | 0,82 | 4,85 | 16,75 | 0,17 | 0,78 | 5,33 | 17,18 | 0,15 |
| 5 | 0,86 | 5,37 | 17,69 | 0,16 | 0,83 | 5,19 | 17,29 | 0,16 | 0,89 | 4,87 | 17,67 | 0,18 |
| 6 | 0,85 | 5,24 | 18,22 | 0,16 | 0,89 | 5,23 | 17,80 | 0,17 | 0,82 | 5,18 | 18,15 | 0,16 |
| 7 | 0,73 | 5,16 | 18,72 | 0,14 | 0,86 | 5,15 | 18,33 | 0,17 | 0,87 | 5,62 | 18,64 | 0,15 |
| 8 | 0,76 | 5,19 | 19,02 | 0,15 | 0,86 | 5,37 | 18,66 | 0,16 | 0,93 | 5,16 | 18,96 | 0,18 |
| 9 | 0,72 | 5,26 | 19,62 | 0,14 | 0,74 | 5,30 | 19,10 | 0,14 | 0,68 | 4,88 | 19,32 | 0,14 |
| 10 | 0,75 | 5,41 | 19,93 | 0,14 | 0,72 | 4,89 | 19,53 | 0,15 | 0,67 | 4,92 | 19,79 | 0,14 |
| 11 | 0,71 | 5,34 | 20,34 | 0,13 | 0,69 | 5,01 | 19,85 | 0,14 | 0,74 | 5,03 | 20,03 | 0,15 |
| 12 | 0,72 | 4,97 | 20,74 | 0,15 | 0,68 | 5,34 | 20,15 | 0,13 | 0,70 | 5,21 | 20,32 | 0,14 |
| 13 | 0,66 | 5,22 | 20,99 | 0,13 | 0,68 | 5,18 | 20,53 | 0,13 | 0,70 | 5,14 | 20,64 | 0,14 |
| 14 | 0,66 | 5,00 | 21,47 | 0,13 | 0,58 | 4,93 | 21,08 | 0,12 | 0,70 | 4,68 | 21,01 | 0,15 |
| 15 | 0,74 | 5,07 | 21,72 | 0,15 | 0,58 | 5,15 | 21,35 | 0,11 | 0,62 | 5,11 | 21,41 | 0,12 |

*Ger = geração

Já na Figura 4c observa-se um comportamento crescente da média fenotípica devido ao processo de seleção aplicado na população simulada, indicando que o processo de simulação dos dados para ambas as metodologias foi eficiente mesmo para característica com baixa herdabilidade.

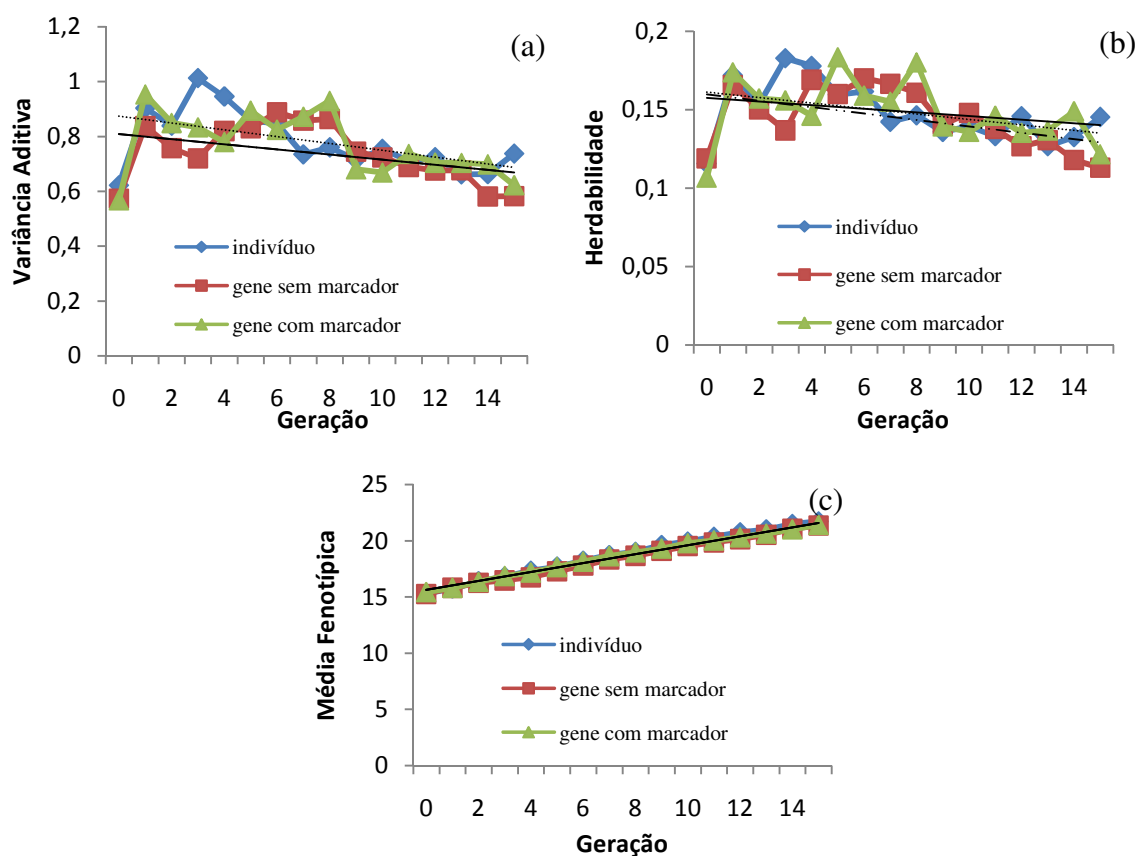


Figura 6 - Comportamento da variância aditiva (a) herdabilidade (b) e média fenotípica (c) de uma população utilizando as metodologias de simulação em nível de indivíduo e gene (com e sem marcador) para uma característica com h^2 de 0,12.

Os resultados das simulações para a característica com herdabilidade igual a 0,27 estão apresentados na Tabela 3 e destacadas na Figura 7. Os mesmos comportamentos dos resultados das simulações anteriores ($h^2=0,12$) foram verificados para esta característica.

Na Figura 7a é apresentado o comportamento da variância aditiva ao longo das gerações. Para a simulação em nível de gene com marcador, ocorreram variações quando comparado com as outras simulações, isto pode ser explicado devido a presença dos QTLs no genoma dos indivíduos, o que causa uma variação até a fixação destes QTLs na população.

Tabela 3. Resultados das variâncias aditivas (σ_a^2), variâncias fenotípicas (σ_p^2), média fenotípica da população (μ) e herdabilidades (h^2) de uma população simulada por 10 gerações utilizando as metodologias de simulação em nível de indivíduo e de gene para uma característica com h^2 de 0,27

| Ger* | Simulação em nível de Indivíduo | | | | Simulação em nível de Gene | | | | | | | |
|------|---------------------------------|--------------|-------|-------|----------------------------|--------------|-------|-------|--------------------|--------------|-------|-------|
| | | | | | Sem Marcador e QTL | | | | Com Marcador e QTL | | | |
| | σ_a^2 | σ_p^2 | μ | h^2 | σ_a^2 | σ_p^2 | μ | h^2 | σ_a^2 | σ_p^2 | μ | h^2 |
| 0 | 0,04 | 0,16 | 1,65 | 0,27 | 0,04 | 0,15 | 1,65 | 0,29 | 0,04 | 0,15 | 1,66 | 0,25 |
| 1 | 0,05 | 0,16 | 1,79 | 0,32 | 0,06 | 0,17 | 1,83 | 0,34 | 0,06 | 0,17 | 1,81 | 0,36 |
| 2 | 0,05 | 0,17 | 1,97 | 0,30 | 0,05 | 0,15 | 1,98 | 0,36 | 0,05 | 0,15 | 1,98 | 0,32 |
| 3 | 0,05 | 0,16 | 2,16 | 0,30 | 0,06 | 0,17 | 2,09 | 0,34 | 0,05 | 0,16 | 2,13 | 0,30 |
| 4 | 0,05 | 0,16 | 2,26 | 0,33 | 0,05 | 0,15 | 2,25 | 0,33 | 0,06 | 0,17 | 2,31 | 0,36 |
| 5 | 0,04 | 0,15 | 2,44 | 0,30 | 0,05 | 0,16 | 2,39 | 0,33 | 0,06 | 0,16 | 2,50 | 0,36 |
| 6 | 0,04 | 0,15 | 2,56 | 0,29 | 0,05 | 0,16 | 2,50 | 0,32 | 0,05 | 0,15 | 2,69 | 0,33 |
| 7 | 0,04 | 0,15 | 2,69 | 0,27 | 0,05 | 0,15 | 2,65 | 0,32 | 0,05 | 0,15 | 2,81 | 0,34 |
| 8 | 0,04 | 0,14 | 2,84 | 0,26 | 0,05 | 0,15 | 2,82 | 0,31 | 0,05 | 0,17 | 2,95 | 0,30 |
| 9 | 0,04 | 0,15 | 2,95 | 0,28 | 0,04 | 0,15 | 2,91 | 0,29 | 0,05 | 0,16 | 3,11 | 0,34 |
| 10 | 0,04 | 0,15 | 3,06 | 0,27 | 0,04 | 0,14 | 3,04 | 0,28 | 0,04 | 0,15 | 3,29 | 0,29 |
| 11 | 0,04 | 0,15 | 3,17 | 0,27 | 0,04 | 0,16 | 3,17 | 0,28 | 0,04 | 0,15 | 3,39 | 0,29 |
| 12 | 0,04 | 0,16 | 3,33 | 0,27 | 0,04 | 0,14 | 3,28 | 0,29 | 0,04 | 0,14 | 3,54 | 0,29 |
| 13 | 0,04 | 0,15 | 3,48 | 0,29 | 0,04 | 0,14 | 3,38 | 0,28 | 0,04 | 0,14 | 3,68 | 0,27 |
| 14 | 0,04 | 0,15 | 3,61 | 0,25 | 0,04 | 0,14 | 3,54 | 0,28 | 0,04 | 0,15 | 3,77 | 0,25 |
| 15 | 0,04 | 0,16 | 3,71 | 0,24 | 0,03 | 0,15 | 3,69 | 0,23 | 0,03 | 0,14 | 3,87 | 0,24 |

*Ger = Geração

Já o parâmetro herdabilidade (Figura 7b) apresentou valores mais constantes ao longo das gerações para a metodologia de simulação em nível de indivíduo quando comparada com as simulações em nível de gene com e sem marcadores. Estes valores podem ser explicados devido ao processo de simulação, já que a variância aditiva varia de acordo com a amostragem da segregação amostral mendeliana, que por sua vez é dependente da variância genética aditiva e dos coeficientes de endogamia dos pais do indivíduo. Este coeficiente de endogamia é calculado utilizando o método tabular e é uma estimativa da quantidade de alelos idênticos por ascendência, e pode sofrer alteração, como por exemplo, a variação entre 0 e 0,5 para indivíduos meio-irmãos.

A Figura 7c representa a média fenotípica crescente através das gerações, com um valor maior para a simulação em nível de gene com marcador, devido à contribuição dos efeitos dos QTLs. Isto não foi verificado para a característica de h^2 igual a 0,12, por apresentar um valor muito baixo e por ter sua expressão muito influenciada pela variância residual.

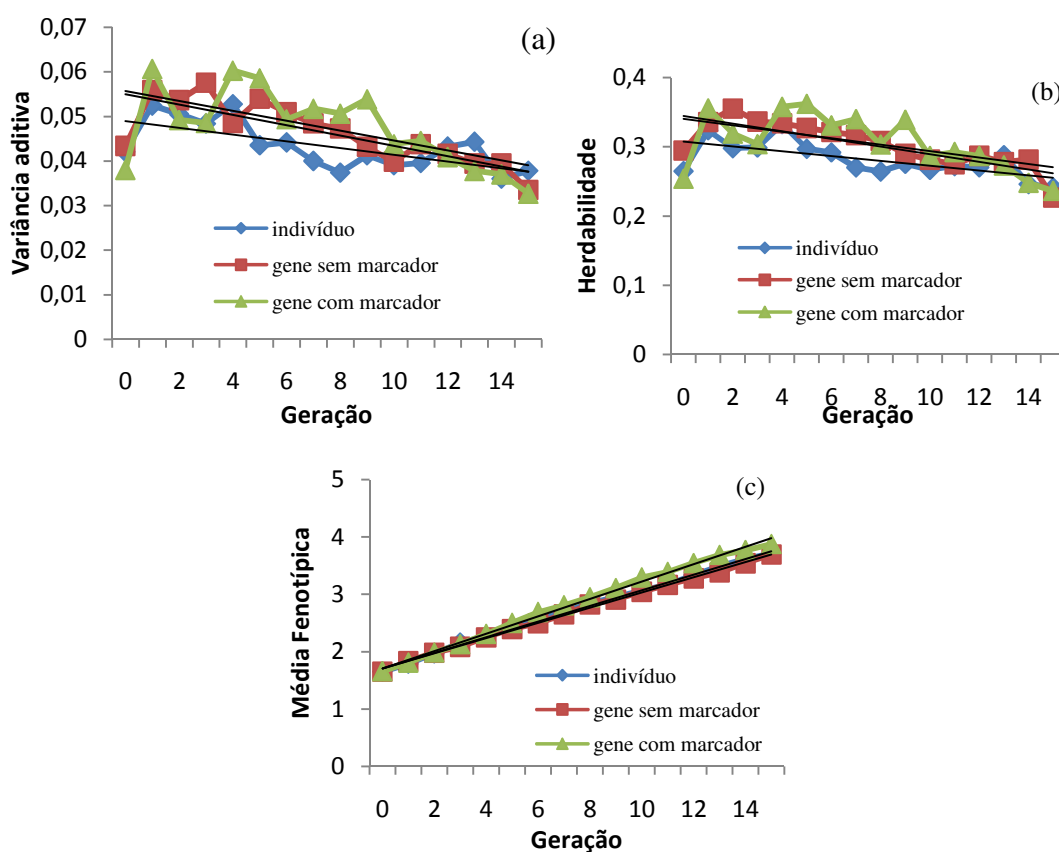


Figura 7 – Comportamento da variância aditiva (a), herdabilidade (b) e média fenotípica (c) de uma população utilizando as metodologias de simulação em nível de indivíduo e gene (com e sem marcador) para uma característica com h^2 de 0,27.

A Tabela 4 apresenta os resultados das simulações para a característica com herdabilidade de 0,47 e destacadas na Figura 8. O comportamento dos valores também foi semelhante às simulações anteriores, exceto para a média fenotípica que foi decrescente, pois o tipo de acasalamento adotado foi o preferencial negativo, ou seja, o vetor dos indivíduos selecionados pelo algoritmo foi ordenado inversamente, do menor para o maior valor fenotípico, selecionando os menores valores.

Tabela 4. Resultados das variâncias aditivas (σ_a^2), variâncias fenotípicas (σ_p^2), média fenotípica da população (μ) e herdabilidades (h^2) de uma população simulada por 10 gerações utilizando as metodologias de simulação em nível de indivíduo e de gene para uma característica com h^2 de 0,47

| Ger* | Simulação em nível de Indivíduo | | | | Simulação em nível de Gene | | | | | | | |
|------|---------------------------------|--------------|-------|-------|----------------------------|--------------|-------|-------|--------------------|--------------|-------|-------|
| | | | | | Sem Marcador e QTL | | | | Com Marcador e QTL | | | |
| | σ_a^2 | σ_p^2 | μ | h^2 | σ_a^2 | σ_p^2 | μ | h^2 | σ_a^2 | σ_p^2 | μ | h^2 |
| 0 | 7,03 | 15,21 | 30,08 | 0,46 | 7,18 | 14,74 | 30,03 | 0,49 | 7,26 | 15,58 | 30,09 | 0,47 |
| 1 | 9,89 | 17,33 | 27,00 | 0,57 | 8,93 | 17,26 | 27,11 | 0,52 | 9,74 | 17,84 | 26,81 | 0,55 |
| 2 | 8,63 | 16,73 | 24,07 | 0,52 | 8,56 | 16,99 | 24,50 | 0,50 | 10,04 | 18,58 | 23,63 | 0,54 |
| 3 | 8,64 | 16,59 | 22,15 | 0,52 | 8,32 | 16,33 | 22,29 | 0,51 | 10,37 | 18,44 | 20,42 | 0,56 |
| 4 | 8,61 | 16,03 | 19,69 | 0,54 | 7,79 | 15,65 | 19,85 | 0,50 | 8,76 | 17,36 | 17,66 | 0,50 |
| 5 | 9,80 | 18,46 | 17,34 | 0,53 | 7,56 | 15,12 | 17,81 | 0,50 | 8,80 | 17,09 | 15,12 | 0,52 |
| 6 | 8,83 | 16,99 | 14,05 | 0,52 | 7,16 | 15,33 | 16,31 | 0,47 | 8,02 | 15,91 | 12,72 | 0,50 |
| 7 | 7,57 | 15,10 | 11,64 | 0,50 | 7,48 | 14,81 | 14,45 | 0,50 | 7,16 | 15,02 | 10,28 | 0,48 |
| 8 | 7,75 | 15,72 | 9,66 | 0,49 | 7,39 | 15,51 | 12,40 | 0,48 | 6,47 | 14,18 | 8,56 | 0,46 |
| 9 | 7,14 | 14,92 | 6,91 | 0,48 | 6,38 | 13,88 | 10,64 | 0,46 | 6,80 | 15,13 | 6,93 | 0,45 |
| 10 | 6,67 | 14,14 | 4,68 | 0,47 | 6,47 | 14,82 | 8,73 | 0,44 | 5,68 | 14,40 | 4,94 | 0,39 |
| 11 | 5,70 | 13,41 | 2,54 | 0,43 | 5,90 | 13,74 | 6,62 | 0,43 | 5,94 | 13,71 | 3,34 | 0,43 |
| 12 | 5,15 | 13,15 | 0,64 | 0,39 | 5,74 | 14,02 | 5,02 | 0,41 | 5,48 | 12,98 | 1,50 | 0,42 |
| 13 | 5,32 | 13,06 | -0,89 | 0,41 | 5,33 | 13,04 | 3,53 | 0,41 | 5,43 | 14,20 | -0,10 | 0,38 |
| 14 | 4,20 | 11,62 | -2,36 | 0,36 | 5,43 | 13,44 | 2,07 | 0,40 | 5,01 | 12,08 | -1,96 | 0,41 |
| 15 | 3,86 | 12,15 | -3,57 | 0,32 | 5,20 | 12,88 | 0,50 | 0,40 | 5,36 | 13,02 | -3,49 | 0,41 |

*Ger = Geração

Este pico das variâncias aditiva e fenotípica e da herdabilidade, nas primeiras gerações e para todas as simulações, pode ser explicado devido ao fato de que as metodologias de simulação adotaram a seleção com base no fenótipo dos indivíduos, e por serem essas características de moderadas à baixas herdabilidades, o efeito do ambiente aleatório pode ter interferido na seleção de animais com alelos indesejáveis, o que aumenta a variância genética aditiva da população (Figuras 6a e 6b).

Já a Figura 6c mostra uma tendência à queda da média fenotípica da população, por ser uma característica selecionada utilizando o acasalamento preferencial negativo, e portanto no algoritmo para a definição dos candidatos a seleção, os vetores dos indivíduos são ordenados em ordem decrescente e são selecionados os menores valores fenotípicos.

Assim, os três processos de simulação foram adequados para amostrar a evolução de um rebanho, para as herdabilidades estudadas, quando aplicado metodologias de seleção devido ao progresso genético apresentados nos gráficos (Figuras 4c, 5c e 6c).

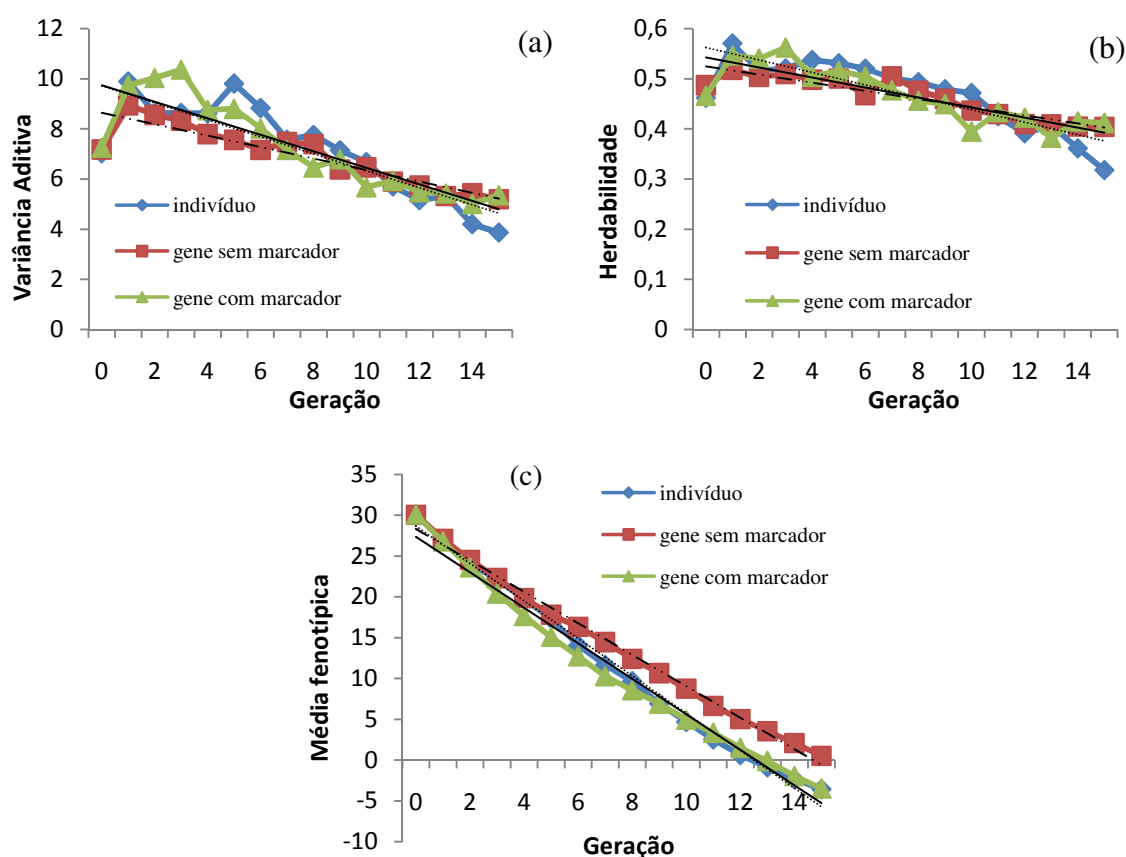


Figura 8 – Comportamento da variância aditiva (a), herdabilidade (b) e média fenotípica (c) de uma população utilizando as metodologias de simulação em nível de indivíduo e gene (com e sem marcador) para uma característica com h^2 de 0,47.

A vantagem da simulação em nível de gene é que é possível simular marcadores moleculares e QTLs e esta metodologia também permite, utilizando equipamentos mais modernos e algoritmos estruturados próprios, a implementação de metodologias para a avaliação de Seleção Assistida por Marcadores ou a Seleção Genômica, que utilizam das informações de marcadores para prever os valores genéticos com maiores precisões, levando a um avanço nesta metodologia. Enquanto a simulação em nível de indivíduo o processo de simulação é eficiente apenas para a determinação dos valores genéticos e os comportamentos dos componentes de variância ao longo das gerações.

Sendo assim, a definição mais adequada para a metodologia de simulação em nível de indivíduo é a de uma metodologia que consiste em apenas amostrar os valores genéticos de cada indivíduo de uma população e para seu desenvolvimento é necessária a determinação do grau de parentesco entre os indivíduos e de seus respectivos coeficientes de endogamia, obtidos pelo cálculo das probabilidades de indivíduos parentes apresentarem alelos idênticos por ascendência. Já a a simulação em nível de gene consiste em simular o efeito da

contribuição de cada alelo presente no genoma do indivíduo permitindo obter o coeficiente de endogamia real de cada indivíduo através da análise do genoma e dos alelos idênticos por ascendência.

É notável também a diferença de quantidade de memória RAM utilizada e tempo de processamento (Tabela 5), já que os algoritmos da simulação em nível de indivíduo são de fácil implementação e utilizam apenas de valores amostrais e cálculos simples, enquanto os algoritmos da metodologia de simulação em nível de gene são mais complexos e necessitam de um número maior de informações e alocações de memória, e se torna cada vez mais oneroso quanto maior o número de informações adicionadas nos cromossomos, isso faz com que o algoritmo percorra todos os alelos de cada cromossomo do indivíduo para se obter os dados necessários, aumentando o tempo de processamento e o consumo de recursos computacionais.

Tabela 5. Tempo total de processamento dos três ambientes de simulação, consumo média de memória RAM por repetição da simulação e total de memória RAM consumida no fim de cada repetição da simulação

| | Simulação em nível de Indivíduo | Simulação em nível de Gene | |
|------------------|---------------------------------|----------------------------|--------------------|
| | | Sem Marcador e QTL | Com Marcador e QTL |
| Tempo (hh:mm:ss) | 00:01:21 | 21:40:13 | 25:12:36 |
| Total consumido* | 80 Mb | 2,0 Gb | 3,2 Gb |

*Total do consumo de memória RAM em um computador de 4Gb de memória e processador Intel® Core™2Duo (2,10GHz)

Portanto, de acordo com as descrições anteriores, para a simulação em nível de gene, grande parte dos esforços de processamento para a formação de um novo indivíduo são originárias das amostragens de números aleatórios, comparações entre quantidades requeridas pelo teste de recombinação e cópias dos alelos dos homólogos dos pais. Na simulação em nível de indivíduo os esforços maiores são devido ao cálculo dos coeficientes de endogamia dos pais do indivíduo a ser formado.

A simulação em nível de gene requer memória para armazenar o genoma do novo indivíduo, enquanto a simulação em nível de indivíduo necessita armazenar as informações necessárias ao cálculo dos coeficientes de endogamia dos pais. A quantidade de memória necessária para o cálculo dos coeficientes de endogamia não é grande e geralmente é menor que a necessária para armazenar os genomas de uma população de indivíduos de mesmo tamanho. Por exemplo, se considerar que os coeficientes de endogamia devem ser obtidos em

uma população de tamanho 1000, pelo método tabular, a quantidade de unidades de memória necessária para a obtenção desses é $(1000 \times 1000)b$.

Para essa mesma população, considerando também que cada loco no genoma ocupa 10 unidades de memória e que a característica a ser simulada é determinada por 500 locos, então a quantidade total de memória é dada por $(1000 \times 500 \times 10)b$, o que demanda 5 vezes mais memória que a simulação em nível de indivíduo.

A simulação em nível de gene somente consome menos memória necessária se o número de locos for reduzido para menos do que 1000, o que é uma situação que não se aproxima da realidade para características quantitativas.

5. CONCLUSÃO

As duas metodologias de simulação são semelhantes e adequadas para estudos da evolução de rebanhos no melhoramento genético animal. A simulação em nível de indivíduo é recomendada para simulações simples devido a fácil implementação do algoritmo e baixa exigência de desempenho computacional. Já a simulação em nível de gene é recomendada para simulações complexas do genoma dos animais e para obtenção dos coeficientes de endogamia reais de cada indivíduo.

A geração do valor genético para novos indivíduos deixa claro uma das principais diferenças entre as simulações em nível de indivíduo e em nível de gene. A simulação em nível de indivíduo por não ter disponível as informações do genoma, simula o efeito da amostragem mendeliana e da recombinação de uma distribuição específica, com parâmetros calculados a cada nova geração.

Na simulação em nível de gene tanto os cromossomos e locos quanto a amostragem mendeliana e recombinação podem ser simulados, não sendo necessária a utilização do efeito conjunto de amostragem mendeliana como na simulação em nível de indivíduo.

Portanto, a definição mais exata para a simulação em nível de indivíduo é que esta consiste em gerar valores genéticos aditivos a partir de uma distribuição normal com média e variância previamente definidas. A simulação em nível de gene consiste em gerar os valores dos efeitos de cada loco poligênico e dos QTLs que influenciam na característica, a partir de uma distribuição normal com média e variância previamente definidas para cada componente e pela soma do efeito de cada loco e QTL, obtém-se o valor genotípico de cada indivíduo da população. Assim, qualquer fragmentação para simulação do valor genotípico dos indivíduos, pode ser caracterizada como simulação em nível de gene.

6. REFERÊNCIAS BIBLIOGRÁFICAS

ASSIS, G.M.L.; CARNEIRO JR., J.M.; EUCLYDES, R.F. *et al.* Estimação de componentes de variância sob influência de genes de efeito principal, comparando-se metodologias Bayesiana e clássica sob diferentes cenários. **Revista Brasileira de Zootecnia**, v.36, n.5, p.1266-1274, 2007.

BOURDON, G.E.P. **Understanding animal breeding**. Upper Saddle River: Prendice-Hall, 2000. 538p.

CARNEIRO, A.P.S.; TORRES, R.A.; EUCLYDES, R.F. *et al.* Efeito da Conexidade de dados sobre o valor fenotípico médio e a variância genética aditiva. **Revista Brasileira de Zootecnia**, v.20, n.2, p. 336-341, 2001.

CARNEIRO, P.L.S.; MALHADO, C.H.M.; EUCLYDES, R.F. *et al.* Endogamia, fixação de alelos e limite de seleção em populações selecionadas por métodos tradicionais e associados a marcadores moleculares. **Revista Brasileira de Zootecnia**, v.36, n.2, p.369-375, 2007.

CARNEIRO JR., J.M.; EUCLYDES, R.F.; LOPES, P.S. *et al.* Avaliação de métodos de estimação de componentes de variância utilizando dados simulados. **Revista Brasileira de Zootecnia**, v.33, n.2, p.328-336, 2004.

CARNEIRO JR., J.M.; ASSIS, G.M.L.; EUCLYDES, R.F. *et al.* Influência da informação *a priori* na avaliação genética animal utilizando dados simulados. **Revista Brasileira de Zootecnia**, v.34, n.6, p.1905-1913, 2005.

CARVALHEIRO, R.; FRIES, L.A.; SCHENKEL, F.S. *et al.* Efeitos da heterogeneidade de variância residual entre grupos de contemporâneos na avaliação genética de bovinos de corte. **Revista Brasileira de Zootecnia**, v.31, n.4, p.1680-1688, 2002.

CUNHA, E.E.; EUCLYDES, R.F.; TORRES, R.A. *et al.* Simulação de dados para avaliação genética de rebanhos de gado de corte. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v.58, n.3, p.381-387, 2006.

FALCONER, D.S.; MACKAY, T.F.C. **Introduction to Quantitative Genetics**. 4^a ed. England : Longman Scientific and Technical, 1996. 464p.

FISHER, R.A. The correlation between relatives on the supposition of Mendelian Inheritance. **Royal Society of Edinburgh**, v.52, p.399-433, 1918. Disponível em: <http://digital.library.adelaide.edu.au/coll/special/fisher/9.pdf> acessado em 21/04/2009.

FONSECA, R., EUCLYDES, R.F., TORRES, R.A. *et al.* Efeito da violação de pressuposições da metodologia de modelos mistos na avaliação genética animal. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v.53, n.1, 2001.

FONSECA, R. **Use of information on identified genes to reduce the selection bias on genetic evaluation**. Tese (Doutorado em Genética e Melhoramento) – Universidade Federal de Viçosa, Minas Gerais, 2003.

FONSECA, R.; VERNIZ, V.; UTSUNOMIYA, A.T.H.; *et al.* LZ5 a free-software for data simulation. In: 32nd Conference for the International Society for Animal Genetics, **Anais...**, Edimburgo, Reino Unido, 2010.

FORMIGONI, I.B.; FERRAZ, J.B.S.; ELER, J.P. Análise da sensibilidade e importância econômica da PP14 e “stayability” para diferentes circunstâncias de produção em rebanho de cria simulado. In: V Simpósio da Sociedade Brasileira de Melhoramento Animal. **Anais...** Pirassununga, SP, 2004.

HAO, K.; CHENG, L.; ROSENOW, C. *et al.* Estimation of genotype error rate using samples with pedigree information – an application on the GeneChip Mapping 10K array. **Genomics**, v.84, p.623-630, 2004.

JOHNSON, M.H.; NOTTER, D.R. Simulation of genetic control of reproduction in beef cows. I. Simulation Model. **Journal of Animal Science**, v.65, p.68-75, 1987.

LYNCH, M.; WALSH, B. **Genetics and analysis of quantitative traits**. Sunderland: Sinauer Associates, 1998. 978p.

KUEHN, L.A. **Implications of Connectedness in the Genetic Evaluation of Livestock**. 2005. 121p. Dissertação (Doctor of Philosophy In Animal and Poultry Sciences) – Virginia Polytechnic Institute and State University, 2005.

NEUENSCHWANDER, S.; HOSPITAL, F.; GUILLAUME, F. *et al.* quantiNemo: na individual –based program to simulate quantitative traits with explicit genetic architecture in a dynamic metapopulation. **Bioinformatics: Applications Note**, v.4, n.13, p.1552-1553, 2008.

NEVES, H.H.R.; CARVALHEIRO, R.; FRIES, L.A. *et al.* Uso combinado de sêmen sexado e acasalamento dirigido sobre uma população de bovinos de corte submetida a seleção: estudo de simulação. **Revista Brasileira de Zootecnia**, v.38, n.12, p.2368-2374, 2009.

PEREIRA, I.C. **Proposta de sistematização da simulação para fabricação em lotes**. Itajubá: Escola Federal de Engenharia de Itajubá, 2000. 134p. Dissertação (Mestrado em Engenharia de produção) - Escola Federal de Engenharia de Itajubá, 2000.

PEREIRA, I.G.; OLIVEIRA, H.N.; ROSA, G.J.M. Estudos de simulação de modelos lineares mistos com distribuição normal contaminada no melhoramento genético animal. **Revista Brasileira de Zootecnia**, v.36, n.5, p.1304-1315, 2007.

POLLAK, E.J.; VAN DER WERF, J.; QUAAS, R.L. Selection bias and multiple trait evaluation. **Journal of Dairy Science**, v.67, p.1590-1595, 1984.

QUAAS, R.L. Computing the diagonal elements and inverse of a large numerator relationship matrix. **Biometrics**, v.32, p.949-953, 1976.

REIS, R.L.; MUNIZ, J.A.; SILVA, F.F. et al. Estimativa de variância genética aditiva em populações selecionadas e não-selecionadas via simulação de Monte Carlo utilizando o software R. **Revista Ciência e Agrotecnologia**, v.33, n.1, p.285-291, jan./fev., 2009.

SORENSEN, D. FERNANDO, R.; GIANOLA, D. Inferring the trajectory of genetic variance in the course of artificial selection. **Genetics Research**, v.77, p.83-94, 2001.

SORENSEN D.; WAAGEPETERSEN, R. Model Selection for Prediction of Breeding Values. 7th WCCALP. **Anais...** 19-23 de Agosto, Montpellier, França, 2002.

SOUZA, G.H. **Avaliação de biotecnologias reprodutivas em programas de melhoramento utilizando simulação de dados**. Dissertação (Doutorado em Genética e Melhoramento) – Universidade Federal de Viçosa, Minas Gerais, 2008.

VERNIZ, V.; UTSUNOMIYA, A.T.H.; ONO, R.K. *et al.* LZ5 – Módulos para simulação de dados e avaliação genética animal. In: IV Simpósio de Ciências da UNESP – Dracena, 2008, Dracena. **Anais...**Dracena: São Paulo,2008.

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)