Universidade de São Paulo
Escola Superior de Agricultura "Luiz de Queiroz"

Estrutura populacional e acurácia da predição genômica em populações de aves: um estudo de simulação

Izally Carvalho Gervásio

Dissertação apresentada para obtenção do título de Mestra em Ciências. Área de concentração: Ciência Animal e Pastagens

Piracicaba
2020

Izally Carvalho Gervásio
Zootecnista

Estrutura populacional e acurácia da predição genômica em populações de aves: um estudo de simulação

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientador
Prof. Dr. **GERSON BARRETO MOURÃO**

Dissertação apresentada para obtenção do título de Mestra em Ciências. Área de concentração: Ciência Animal e Pastagens

Piracicaba
2020

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA – DIBD/ESALQ/USP

Gervásio, Izally Carvalho

Estrutura populacional e acurácia da predição genômica em populações de aves: um estudo de simulação / Izally Carvalho Gervásio. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2020.

80 p.

Dissertação (Mestrado) - - USP / Escola Superior de Agricultura “Luiz de Queiroz”.

1. Avicultura 2. População 3. Tamanho efetivo 4. Simulação I. Título

DEDICATÓRIA

"Aos meus pais Geraldo Osmil e Liberaci, pelo amor e apoio incondicional."

AGRADECIMENTOS

À Deus, que esteve comigo em todos os momentos e me iluminou sempre.

À Escola Superior de Agricultura "Luiz de Queiroz", principalmente ao Departamento de Zootecnia pela oportunidade.

À Capes, CNPq e a FAPESP pelo consentimento de bolsas e pelo apoio à pesquisa.

Ao Professor Dr. Gerson Barreto Mourão pela orientação, pelos conselhos e por todo aprendizado acadêmico, profissional e pessoal durante o período de mestrado, pois, um sonho fora do comum exige um orientador fora do comum, valeu Chefe!

Aos professores do Departamento de Zootecnia (LZT/ESALQ/USP), em especial aos professores, Drª Carla Maris Machado Bittar, Dr. José Eurico Possebon Cyrino e Dr. Luiz Lehmann Coutinho pelos ensinamentos, pelos aconselhamentos, vocês não só transmitem conhecimentos, mais inspiram novas atitudes.

A Drª Carla Cachoni Pizzolante e ao Dr. José Evandro de Moraes do Instituto de Zootecnia (IZ- Nova Odessa) pelo apoio, pela amizade e pelo aprendizado durante o mestrado.

Aos meus professores desde a minha infância até aqui, especialmente meus antigos orientadores da Universidade Federal de Lavras (UFLA), Profª Drª Ana Paula Peconick, Prof. Dr. Alessandro Torres Campos e Profª DrªSarah Laguna Conceição Meirelles, que foram verdadeiros mestres, transmitindo conhecimentos e experiências que não só me ajudaram no período de mestrado, mas que guardarei pelo resto da vida.

Aos colegas do Programa de Pós-graduação em Ciência Animal e Pastagens.

Aos alunos, ex-alunos e estagiários do Grupo de Estudos em Estatística e Melhoramento Genético Animal – GEMA/ESALQ/USP: Ariana Meira, Adhonis Lima, Brayan Dauria, Daysi Orzuza, Eduardo de Carvalho, Eula Carrara, Fabrício Pilonetto, Fátima Bogdanski, Flaviane Pagoto, Felipe Freitas, Felipe Silva, Giovanni Ladeira, Gregori Rovadoscki, Isabella Valeriano, José Eduardo Lopes, Juliana Petrini, Leonardo Mazzero, Letícia Sartori, Luiz Firmino Júnior, Mariana Pedroso, Marina Miquelini, Mayara Salvian, Paola Boscollo, Simone Pertille, Sofia Silva, Thaynara Teixeira e Vamilton Franzo, vocês foram um grande apoio e jamais esquecerei dos momentos de companheirismo, que juntos partilhamos, obrigada por enriquecerem minha vida profissional e pessoal, por serem minha segunda família, vocês estarão eternamente em meu coração, valeu família GEMA!

Aos alunos do Laboratório de Biotecnologia pelo apoio, ensinamentos, companheirismo e troca de experiências durante o mestrado, vocês contribuíram imensamente para essa conquista.

Aos amigos que fiz nesses dois anos intensos e também aos amigos das antigas, "os mio da tia Corina", os da UFLA, do Grupo de Melhoramento Animal e Biotecnologia (GMAB –

UFLA), a turma da Zootecnia 2013/1 e as maravilhosas da República Camponesas, obrigada pela amizade, pelos conselhos, por tudo, vocês estarão sempre no meu coração e na minha memória, e mesmo distantes vocês foram incentivo e apoio em muitas horas, e me deram força para que eu chegasse até aqui.

Á minha família, que apesar das nossas diferenças, nunca deixaram de me apoiar, agradeço por tudo que já fizeram e fazem por mim, é realmente estas diferenças que nos fazem completos, e não impedem de nos amarmos mutuamente, podem ter certeza, que agradeço e peço a Deus todos os dias por vocês.

Aos meus pais Geraldo Osmil e Liberaci, que me deram a vida e me ensinaram a vivê-la com dignidade, que iluminam meus caminhos obscuros com afeto e dedicação para que eu os trilhe sem medo e cheio de esperança, que se doaram inteiros e renunciaram aos seus sonhos, para que, muitas vezes, possa realizar os meus, meu eterno muito obrigada.

Enfim a todos que contribuíram de uma forma ou de outra para a realização deste trabalho.

MUITO OBRIGADA!

EPÍGRAFE

"… Ano passado eu morri, mas esse ano eu não morro..."

Emicida – AmarElo (Sample: Belchior – Sujeito de Sorte)

SUMÁRIO

RESUMO 9
ABSTRACT 10
1. INTRODUÇÃO 11
Referências 13
2. REVISÃO BIBLIOGRÁFICA 15
2.1 Fatores que afetam os programas de melhoramento genético animal 15
2.2 Tamanho efetivo da população (Ne) 18
2.3. Endogamia 21
2.4. Desequilíbrio de ligação (LD) 22
2.5. Seleção genômica (GS) 25
Referências 29
3. SIMULAÇÃO DE ESTRATÉGIAS DE SELEÇÃO E TAMANHO EFETIVO POPULACIONAL DE REBANHOS DE LINHAGENS CAIPIRAS NACIONAIS 33
Resumo 33
Abstract 34
3.1. Introdução 34
3.2. Material e métodos 35
3.2.1 Processo de simulação 36
3.2.2 Avaliações genéticas 40
3.2.2.1 Tamanho efetivo da população 40
3.2.2.2 Endogamia e homozigose 40
3.2.2.4 Tendências genéticos 41
3.3. Resultados 41
3.3.1. Endogamia e taxa de homozigose 41
3.3.2 – Tendência genética 44
3. 4 Discussão 46
3.4.1 Endogamia e Homozigose 46
3.4.2 Tendência genética 48
3.5 Conclusão 48
Referências 49

4. SIMULAÇÃO DE DIFERENTES TAMANHOS EFETIVOS POPULACIONAIS DE REBANHOS DE LINHAGENS CAIPIRAS NACIONAIS E OS IMPACTOS NO DESEQUILÍBRIO DE LIGAÇÃO E NA ACURÁCIA DE PREDIÇÃO GENÔMICA. 52
Resumo 52
Abstract 52
4.1 Introdução 53
4.2 Material e métodos 55
4.2.1 Processo de simulação 55
4.2.2 Desequilíbrio de ligação (LD) 59
4.2.3 População e fenótipos e genótipos 59
4.2.4 Predição 59
4.3 Resultados 61
4.3.1 Desequilíbrio de ligação 61
4.3.2 Acurácia e viés de predição “empíricos” 65
4.4 Discussão 66
4.4.1 Desequilíbrio de Ligação 66
4.4.2 Acurácia e viés de predição “empíricos” 68
4.5 Conclusão 70
Referências 71
ANEXOS 74

RESUMO

Estrutura populacional e acurácia da predição genômica em populações de aves: um estudo de simulação

Para o desenvolvimento sustentável, a produção de linhagens de aves nativas, seria uma alternativa de aliar a preservação da diversidade e variabilidade genética. Neste contexto, foram desenvolvidos dois estudos. O primeiro estudo objetivou-se entender melhor as estratégias de seleção e acasalamentos utilizados; e compreender se o Ne de rebanhos de aves caipiras nacionais em condições reais está adequado. Foi simulado uma geração histórica que por mil gerações teve um tamanho constante de 2000 indivíduos, ocorrendo um gargalo genético nas gerações subsequentes (1010 a 1020) para 400 animais e expandiu para 1000, 500, 200 e 100 animais de onde surgiram as fases (população recente) dos quatro cenários simulados e para três coeficientes de herdabilidade (0,15; 0,30 e 0,45). Na Fase 1 os indivíduos não passaram por nenhum processo de seleção por quarenta gerações (G_0 a G_40), posteriormente nas Fases 2 e 3, os indivíduos foram selecionados por características fenotípicas por 25 gerações (G_41 a G_65) e pelos valores genéticos preditos (BLUP-EBV) nas ultimas 15 gerações (G_66 a G_80), respectivamente. Os cenários 1, 2, 3 e 4 possuíram tamanhos efetivos da população de 1000, 640 e 640; 500, 320 e 320; 200, 128 e 128; 100, 64 e 64 nas Fases 1, 2 e 3, respectivamente. Foram calculados coeficientes de endogamia, a taxa de homozigose, as tendências genéticas e os ganhos genéticos. Para todos os cenários estudados, a geração zero (G_0) das três fases da população recente (Fase 1. Fase 2 e Fase 3) teve um coeficiente de endogamia igual a zero. Além disso, os valores obtidos para os coeficientes de endogamia aumentaram da primeira para a última geração de cada uma das fases 1, 2 e 3 estudadas, e os maiores coeficientes de endogamia foram obtidos nos cenários com menor Ne, assim como, os valores das taxas de homozigose. Os valores de ganhos genéticos foram maiores para o coeficiente de herdabilidade de 0,45 dos quatro cenários simulados com diferentes tamanhos efetivos da população. Pode-se concluir que o Ne dos rebanhos de linhagens caipiras deve ser aumentado; e as estratégias de seleção devem ser revisadas e aplicadas para a minimização dos níveis de endogamia e homozigose; para que os ganhos genéticos destas populações possam ser maximizados. No segundo objetivou-se compreender os impactos do Ne sobre o desequilíbrio de ligação (LD) e a acurácia de predição genômica. Foi realizado o mesmo processo de simulação do estudo anterior para três cenários e três coeficientes de herdabilidade (0,15; 0,30 e 0,45) simulados. Os cenários 1, 2, 3 possuíram tamanhos efetivos da população de 500, 320 e 320; 200, 128 e 128; 100, 64 e 64 nas Fases 1, 2 e 3, respectivamente. Foram calculados o desequilíbrio de ligação (LD), a acurácia e o viés de predição "empíricos". Os valores médios de r^2, a distâncias dos marcadores de 0-0,05 Kb, foram inferiores ou próximos a 0,30 nas Fases 1 e 2 dos cenários 1 e 2, enquanto que para o cenário 3 (menor Ne), os valores médios de r^2 foram superiores a 0,30 em todas as fases e coeficientes de herdabilidade simulados. Os valores de correlações entre os TBVs e EBVs variaram de 0,69 a 0,81 nos três cenários e herdabilidades simulados. As correlações entre TBVs e GEBVs variaram de 0,45 a 0,90; e foram menores no cenário 3. Os coeficientes de regressão estimados dos fenótipos simulados nos EBVs foram altos e maiores que 1; e as regressões dos fenótipos simulados nos GEBVs apresentaram valores menores que 1. Por isso, com base nos resultados deste estudo de simulação pode-se concluir que o LD das populações de linhagens caipiras da ESALQ pode ser considerado "útil" e eficiente para estudos genômicos e para que os valores de acurácia de predição sejam maiores será necessário o aumento do Ne nos rebanhos de linhagens caipira da ESALQ.

Palavras-chave: Avicultura, População, Tamanho efetivo, Simulação

ABSTRACT

Population structure and accuracy of genomic prediction in poultry populations: a simulation study

For sustainable development, the production of native bird lines would be an alternative to combine the preservation of genetic diversity and variability. In this context, two studies were developed. The first study aimed to better understand the selection and mating strategies used and understand whether the Ne of national free-range poultry flocks in real conditions is adequate. A historical generation was simulated, which for a thousand generations had a constant size of 2000 individuals, with a genetic bottleneck occurring in subsequent generations (1010 to 1020) for 400 animals and expanded to 1000, 500, 200 and 100 animals from which the phases emerged (population recent) of the four simulated scenarios and for three coefficients of heritability (0.15; 0.30 and 0.45). In Phase 1, individuals did not undergo any selection process for forty generations (G_0 to G_40), later in Phases 2 and 3, individuals were selected for phenotypic characteristics for 25 generations (G_41 to G_65) and the predicted genetic values (BLUP -EBV) in the last 15 generations (G_66 to G_80), respectively. Scenarios 1, 2, 3 and 4 had effective population sizes of 1000, 640 and 640; 500, 320 and 320; 200, 128 and 128; 100, 64 and 64 in Phases 1, 2 and 3, respectively. Inbreeding coefficients, homozygosity rate, genetic trends, and genetic gains were calculated. For all studied scenarios, the zero generation (G_0) of the three phases of the recent population (Phase 1. Phase 2 and Phase 3) had an inbreeding coefficient equal to zero. In addition, the values obtained for the inbreeding coefficients increased from the first to the last generation of each of the studied phases 1, 2, and 3, and the highest inbreeding coefficients were obtained in the scenarios with the lowest Ne, as well as the values of the homozygosity rates. The values of genetic gains were higher for the heritability coefficient of 0.45 of the four simulated scenarios with different effective population sizes. It can be concluded that the Ne of herds of free-range lineages must be increased; and selection strategies should be reviewed and applied to minimize levels of inbreeding and homozygosity; so that the genetic gains of these populations can be maximized. The second aimed to understand the impacts of Ne on the linkage disequilibrium (LD) and the accuracy of genomic prediction. The same simulation process of the previous study was performed for three scenarios and three coefficients of heritability (0.15; 0.30 and 0.45) simulated. Scenarios 1, 2, 3 had effective population sizes of 500, 320 and 320; 200, 128 and 128; 100, 64 and 64 in Phases 1, 2 and 3, respectively. Linkage disequilibrium (LD), accuracy, and "empirical" prediction bias were calculated. The average values of r^2, at distance from the markers of 0-0.05 Kb, were lower or close to 0.30 in Phases 1 and 2 of scenarios 1 and 2, while for scenario 3 (lowest Ne), the values mean r^2 were greater than 0.30 in all phases and simulated heritability coefficients. The correlation values between TBVs and EBVs varied from 0.69 to 0.81 in the three scenarios and simulated heritabilities. The correlations between TBVs and GEBVs ranged from 0.45 to 0.90 and were lower in scenario 3. The estimated regression coefficients of the simulated phenotypes in the EBVs were high and greater than 1; and the regressions of the simulated phenotypes in the GEBVs showed values less than 1. Therefore, based on the results of this simulation study, it can be concluded that the LD of the populations of ESALQ free-range strains can be considered "useful "and efficient for genomic and for the prediction accuracy values to be higher, it will be necessary to increase the Ne in herds of ESALQ free-range strains.

Keywords: Poultry farming, Population, Effective size, Simulation

1. INTRODUÇÃO

As aves domésticas (*Gallus gallus*) são nativas do continente asiático e foram introduzidas no continente americano durante o processo de colonização espanhola e portuguesa. No Brasil, inicialmente as aves eram criadas sem nenhum padrão de produção e foi apenas a partir da década de 40 que os produtores passaram a selecionar os melhores animais conforme as características fenotípicas desejáveis (Silva; Nakano, 1997).

A seleção desses animais fez com que na década de 50, surgissem novas linhagens de corte devido ao cruzamento das linhagens Cornish Branca/Preta e Plymouth Rock Branca/Barrada, que, ao longo dos anos foram sendo melhoradas, resultando no que conhecemos hoje como o frango de corte moderno. Além disso, para as criações de subsistência (pequenas propriedades), foram desenvolvidas linhagens adequadas para esse sistema de criação tais como, Plymouth Rock Barrada (carijós), Rhode Island Red (vermelhas) e New Hampshire Shamo Assel e linhagens caipiras nacionais (Figueiredo et al., 2003).

O primeiro programa de melhoramento genético de aves no Brasil teve início em 1957, em que o foco foi o melhoramento de frangos com aptidão para a produção de carne (Silva, 2009). Na década de 1960 a Escola Superior de Agricultura "Luiz de Queiroz" (ESALQ/ USP) e o Instituto de Zootecnia (IZ) deram início a projetos na mesma temática, e na década de 1980, foram desenvolvidas linhagens caipiras nacionais por meio da seleção massal originando o programa intitulado "Frango Feliz". Estas linhagens caipiras foram utilizadas para este estudo (Coelho et al., 2008), e atualmente, o programa possui controle completo do *pedigree* de no mínimo duas gerações de seleção (Rovadoscki et al. 2016).

Tradicionalmente, para a realização o melhoramento genético de aves, é necessário que os animais selecionados sejam precoces, tenham alta eficiência alimentar, possuam uma carcaça com alto rendimento de cortes nobres, sejam saudáveis, produtivos e prolíferos (Figueiredo et al., 2003). Para a obtenção de resultados satisfatórios, devem ser levados em consideração, as estimativas dos parâmetros genéticos e fenotípicos dos animais, o tamanho efetivo da população, a intensidade de seleção utilizada para machos e fêmeas do plantel, os métodos de seleção dentre outros, pois estes, têm uma grande contribuição no estabelecimento do melhoramento genético e auxiliam na seleção das melhores aves da população em estudo (Ledur et al., 1993).Com o desenvolvimento da biologia molecular, bem como das tecnologias genômicas, tornou-se possível acelerar e melhorar a acurácia do processo de seleção dos animais superiores. Dentre elas, a seleção genômica tem contribuído para a avaliação de características de difícil mensuração, que possuem elevado custo de análise, que tenham valores de herdabilidades baixos e que podem ser limitadas ao sexo, permitindo a seleção dos animais de forma efetiva (Ledur, 2010).

Mesmo assim, os estudos e programas de melhoramento genético sobre linhagens caipiras são escassos na literatura, especialmente os que englobam linhagens de postura, corte e de dupla aptidão, as quais tem uma pequena, porém crescente, participação no mercado avícola brasileiro (Silva, 2009; Rovadoscki et al., 2016). A criação de aves caipiras tem uma demanda crescente, com o objetivo de atender a sistemas alternativos de criação como, a agricultura silvo pastoril, propriedades de menor demanda comercial. Os consumidores que buscam uma alimentação mais saudável, exigem que os animais sejam criados em condições de bem-estar (Silva et al., 2003; Hellmeister Filho et al., 2003; Coelho et al., 2007; Vanhonacker; Verbeke, 2009) e que não tenham passado por processos industriais, o que é um dos requisitos na avicultura alternativa (Neves da Silva et al., 2003).

Neste contexto, a Escola Superior de Agricultura "Luiz de Queiroz" (ESALQ/USP) através dos trabalhos desenvolvidos pelo Grupo de Estudos em Estatística e Melhoramento genético animal (GEMA) tem contribuído para o desenvolvimento da avicultura alternativa, pelo desenvolvimento do programa de melhoramento genético de linhagens caipiras nacionais que em sua fase atual, possui controle completo de pedigree de no mínimo duas gerações de seleção; os primeiros resultados desse esforço podem ser encontrados em Rovadoscki et al. (2016). Por isso, para melhor entendimento das estratégias de seleção, acasalamentos e do tamanho efetivo da população (Ne) utilizados no plantel de linhagens caipiras em condições reais foi realizado o presente estudo de simulação.

Sendo assim, com este trabalho busca-se entender: a) os impactos das estratégias de seleção utilizadas e do tamanho efetivo da população sobre as medidas populacionais, como o coeficiente de endogamia, a taxa de homozigose e o ganho genético das populações de linhagens de aves caipiras simuladas; e b) a influência dos projetos de seleção e dos tamanhos efetivos da população (Ne) sobre o desequilíbrio de ligação (LD) e da acurácia de predição genômica dos indivíduos dos rebanhos de linhagens de aves caipiras simuladas.

Por isso, objetivou-se verificar, pelo processo de simulação os resultados das estratégias de seleção aplicadas e os diferentes tamanhos efetivos sobre medidas populacionais, desequilíbrio de ligação (LD) e acurácia de predição genômica de rebanhos simulados de aves, baseados nas populações de aves caipiras reais.

Referências

Coelho, A.A.D. et al.; Características da carcaça e da carne de genótipos de frangos caipiras. **Brazilian Journal of Food Technology**, Campinas, v. 10, p. 9- 15, 2007.

Coelho, A. A. D.; Savino, V. J. M.; Rosário, M. F. Frango feliz: caminhos para a avicultura alternativa. [S.l: s.n.], 2008.

Figueiredo, E. et al. Raças e linhagens de galinhas para criação comerciais e alternativas no Brasil. **Ministério da Agricultura, Pecuária e Abastecimento**, v. 347, n. 0100=8862, p. 8, 2003.

Hellmeister Filho, P. et al. Efeito de genótipo e do sistema de criação sobre o desempenho de frangos tipo caipira. **Revista Brasileira de Zootecnia**, v. 32, n. 6 suppl 2, p. 1883–1889, 2003.

Ledur, M. C. et. al, Parâmetros genéticos e fenotípicos para características produtivas em linhagens de poedeiras de ovos brancos. **Pesquisa Agropecuária Brasileira**, v. 28, n. 9, p. 1031-1037, 1993.

Ledur, C.M; Peixoto, O. J; Aplicações e Perspectivas da Genômica no Melhoramento Genético de Aves. **Embrapa Suínos e Aves**, mledur@cnpsa.embrapa.br. Acessado em: 28/10/2019.

Neves Da Silva, M. A. et al. Influence of the Rearing System on Performance, Physiological Condition and Behaviour of Broilers Lines | Influência do Sistema de Criação sobre o Desempenho, a Condição Fisiológica e o Comportamento de Linhagens de Frangos para Corte. **Revista Brasileira de Zootecnia**, v. 32, n. 1, p. 208–213, 2003.

Rovadoscki, G. A. et al. Genetic parameters for growth characteristics of free-range chickens under univariate random regression models. **Poultry Science**, v. 95, n. 9, p. 1989–1998, 2016.

Silva, R. D. M.; Nakano, M. Sistema caipira de criação de galinhas. [S.l: s.n.], 1998.

Silva, Evolução do melhoramento genético de aves no Brasil Genetic improvement of Poultry in Brazil. **Ceres**, v. 56, n. 3, p. 437–445, 2009.

Vanhonacker, F.; Verbeke, W. Buying higher welfare poultry products? Profiling flemish consumers who do and do not. **Poultry Science**, v. 88, n. 12, p. 2702–2711, 2009.

.

1. REVISÃO BIBLIOGRÁFICA

A Escola Superior de Agricultura "Luiz de Queiroz" tem contribuído desde meados da década de 1960 para a evolução do melhoramento genético de aves no Brasil. O projeto "Frango Feliz" foi idealizado por pesquisadores do Departamento de Genética ESALQ/USP, onde foram desenvolvidos diversos trabalhos com linhagens caipiras nacionais com aptidões para o corte e postura (Silva, 2009).

Um destes trabalhos foi realizado em 2013 em uma dissertação de mestrado, intitulada "Modelos de curvas de crescimento e regressão aleatória em linhagens nacionais de frango caipira", onde as linhagens experimentais de frangos caipiras (7P, Caipirão da ESALQ, Caipirinha da ESALQ e Carijó Barbado) foram avaliadas através de seus pesos. O estudo teve o objetivo de selecionar uma curva de crescimento que melhor descrevesse o padrão de crescimento para cada linhagem através de modelos específicos como, von Bertalanffy, Gompertz, Logística, Richards e Brody, pelo método dos Mínimos Quadrados Ordinários (QMO) e Quadrados Mínimos Ponderados (QMP). Este trabalho deu origem ao artigo publicado na revista *Poultry Science* intitulado "*Genetic parameters for growth characteristics of free-range chickens under univariate random regression models*" (Rovadoscki et al., 2016)

O setor avícola está sob responsabilidade do Departamento de Zootecnia (ESALQ/USP), onde foram realizados ensaios de eficiência alimentar com 144 machos de duas linhagens, Caipirão da ESALQ e Caipirinha da ESALQ e um trabalho para conclusão do mestrado intitulada "Parâmetros genéticos de características de produção e qualidade do ovo". Neste trabalho, foram avaliadas características de qualidade do ovo (a cor da casca, peso do ovo, altura do albúmen, cor da gema, unidades Haugh, resistência da casca, espessura da casca, peso da gema e peso da casca) e de produção de ovos, considerando idade na primeira postura e produção de ovos semanal de oito linhagens nacionais caipiras (Caipirão da ESALQ, Caipirinha da ESALQ, 7P, Carijó, Perna Longa, Carne Preta, Linhagem Azul e Baianinha da ESALQ) (Bogdanski, 2019).

Portanto, espera-se que o presente estudo de simulação de dados contribua para o entendimento dos eventos que podem ou não estar ocorrendo no rebanho de aves de linhagens caipiras de uma forma precisa e rápida; e possa ser base para a implementação e continuidade do programa de melhoramento destas linhagens em condições reais

2.1 Fatores que afetam os programas de melhoramento genético animal

O melhoramento genético animal é a união do processo de seleção aos acasalamentos direcionados, visando o aumento das frequências gênicas desejáveis para características de

importância econômica em uma determinada população. Para que este programa de melhoramento seja efetivo, deve-se escolher os melhores animais para reprodução, estabelecendo os métodos de cruzamento e seleção. Assim, a obtenção de resultados satisfatórios depende de diversos fatores, como: objetivos e estrutura do programa; coleta de dados, a importância econômica das características, parâmetros genéticos, tamanho da população, intensidade de seleção, métodos de seleção e progresso genético esperado (Lana, 2000).

Na avicultura, a pirâmide de produção apresenta três etapas que se complementam, no topo estão os Rebanhos Núcleo ou elite, responsáveis pelo melhoramento genético das linhas puras, pela seleção intensiva das características economicamente importantes; na parte central estão os Rebanhos Multiplicadores, responsáveis pela produção de matrizes macho e fêmea; e na base da pirâmide os Rebanhos Comerciais, que são os indivíduos destinados ao abate ou à produção de ovos. Dessa forma, o progresso genético é transferido para as gerações seguintes por meio de multiplicação, passando para as bisavós, avós, matrizes e finalmente, após cerca de quatro anos, chega-se ao produto final, os híbridos produzidos por meio de cruzamentos entre três ou quatro linhagens, denominados pintos de um dia, que são vendidos aos criadores comerciais de frangos de corte e de galinhas poedeiras (Ageitec, 2020).

Os parâmetros genéticos, tais como herdabilidade e correlações, tanto genética como fenotípica, são importantes para a estruturação de programas de melhoramento genético de várias espécies de animais. A herdabilidade (h^2) é a medida de consistência da relação entre o desempenho (valores fenotípicos) e os valores genéticos para uma característica em uma população. Portanto, a h^2 expressa o quanto da variabilidade de uma característica é devido à origem genética aditiva. (Bourdon, 2000).

A correlação genética é uma medida da força (consistência, confiabilidade) da relação entre os valores genéticos para uma característica e os valores genéticos para outra característica, ou seja, a seleção para uma determinada característica pode influenciar positivamente ou negativamente outras características. Além disso, a correlação genética pode ser uma medida de pleiotropia, em que um único par de alelos é responsável pelo determinismo de dois ou mais caracteres simultaneamente. (Bourdon, 2000). O uso desses parâmetros genéticos (herdabilidade e correlações genéticas) de forma eficiente em um programa de seleção em linhagens caipiras nacionais, por exemplo, tem como objetivo orientar, conduzir e até avaliar a seleção empregada nas gerações que compuseram essa linhagem (Gaya et al., 2006).

Deve-se definir o tamanho ideal e o tamanho efetivo da população, para melhor entendimento de seu efeito sobre a frequências gênicas. A importância destes conceitos torna-se mais clara à medida que se leva em consideração que é uma amostra de genes de uma determinada

população que será transmitida à próxima. Assim, a frequência gênica na progênie será influenciada pela variação amostral que será tanto maior quanto menor for o número de pais. (Filho, 1999).

Nesse contexto, a intensidade de seleção está relacionada com a proporção de indivíduos que serão utilizados na reprodução e pode ser obtida pela seguinte formula:

$i = \frac{\Delta S}{\sigma_p}$ e $\Delta S = \bar{X}s - \bar{X}\mathrm{p}$

Em que:

i intensidade de seleção

ΔS diferencial de seleção

σ_p desvio-padrão fenotípico da característica em estudo

$\bar{X}s$ média dos indivíduos selecionados

$\bar{X}p$ média da população.

Sendo assim, quanto menor o número de indivíduos selecionados para reprodução, maior será a média e consequentemente maior será o diferencial e a intensidade de seleção praticada (Júnior, 2010). Consequentemente todos estes fatores estão relacionados ao progresso genético esperado para a população em estudo. Por isso, o ganho genético representa a superioridade genética dos descendentes em relação à média da geração dos pais, dependendo da acurácia de predição, intensidade de seleção, variabilidade genética e intervalo de gerações, formando a equação chave para a predição do ganho genético:

$$\Delta G_{ano} = \frac{(r_{AA})(i)(\sigma_A)}{t}$$

ΔG_{ano} ganho genético por ano

r_{AA} acurácia de seleção

i intensidade de seleção

σ_A desvio padrão genético aditivo

t intervalo de gerações.

Essa equação mostra quão o ganho genético é diretamente proporcional à acurácia de seleção, à intensidade de seleção e à variância genética existente; e inversamente proporcional ao ntervalo de gerações, que é o tempo médio em anos entre o nascimento de um animal e de seus filhos. Ou seja, o progresso genético será maior quanto maior for a intensidade de seleção, a variabilidade genética e a acurácia de predição e menor quanto forem maiores o intervalo de gerações. Mas este processo não é tão simples, por exemplo, a diminuição do intervalo de gerações pode ocasionar o decréscimo da acurácia de predição, devido ao pequeno conjunto de informações

que estaria disponível para avaliação genética e o aumento da intensidade de seleção pode reduzir a variância genética aditiva, com consequentes prejuízos para a o progresso genético esperado (Júnior, 2010).

2.2 Tamanho efetivo da população (Ne)

O termo tamanho efetivo populacional foi originalmente formulado por Wright (1931) que o definiu como "o número de indivíduos capazes de se reproduzir numa população idealizada que apresentariam a mesma dispersão de frequências alélicas sob deriva genética aleatória ou a mesma quantidade de endo cruzamentos que a população em questão". O Ne é utilizado como um elemento teórico padronizador e, por isso, a população em estudo deve seguir alguns padrões para que possa ser comparada a uma população idealizada como: não ocorrerem sobreposições de gerações, o número de indivíduos intercruzantes deve ser o mesmo em todas as gerações, todos os indivíduos devem ser potencialmente intercruzantes, a união dos gametas deve ser aleatória, não deve estar havendo influência de seleção, mutação ou migração; a média de descendentes, por indivíduo adulto, de um indivíduo (Wright, 1938).

O valor do tamanho efetivo populacional na maioria das vezes costuma ser menor que o valor de N (tamanho populacional real), isto se deve a três fatores: a variância reprodutiva, quando os casais da população não possuem o mesmo número de descendentes; a proporção sexual, quando o número de machos e fêmeas é significativamente discrepante e; estabilidade populacional, quando o número de indivíduos da população pode variar entre as gerações. Para o cálculo do valor de Ne podem ser utilizados diversas fórmulas, considerando ou não os fatores mencionados anteriormente (Wright, 1938).

A estimativa do tamanho efetivo da população com base no número de pais e mães (Nes) segue o modelo de Wright (1931). Este método permite prever Nes sob várias premissas, incluindo acasalamento aleatório, ausência de seleção e variação aleatória do tamanho da ninhada entre os pais (Leroy et al. 2013). O Nes requer apenas o número estimado de machos reprodutores (Nm) e fêmeas (Nf) na população de referência usando a fórmula: $N_{es} = \frac{4\,N_m N_f}{N_m + N_f}$. O incremento da endogamia hipotética (ΔF) é inversamente proporcional ao número de Nes: $\Delta F = \frac{1}{2N_{es}}$.

Pham et al. 2016 em um estudo sobre conservação e diversidade genética envolvendo oito populações de aves nativas de Taiwan, puderam observar que os valores de Ne estimados com base no número de reprodutores e matrizes foram de 3,2 a 36,5 e os valores do coeficiente de endogamia foram de 1,37 a 15,63; mostrando que quanto maiores os valores do tamanho efetivo

populacional, menores eram os valores de coeficiente de endogamia (ΔF) como pode ser observado na Tabela 1.

Tabela 1. Informação de pedigree na primeira geração de oito populações de aves nativas no início do programa de conservação

População	Primeira geração				
	Ano	Reprodutor	Matriz	Ne_s	ΔF
B strain (BS)	1984	6	20	18,5	2,71
L2 strain (LS)	1984	5	28	17	2,95
Hsin-Yi (HY)	1984	7	15	19,1	2,62
Hua-Tung (HT)	1990	2	4	5,3	9,38
Ju-Chi (JC)	1986	12	38	36,5	1,37
Nagoya (NG)	1989	9	22	25,5	1,96
Quemoy (KM)	1995	1	4	3,2	15,63
Shek-Ki (KT)	1989	6	19	18,2	2,74

Ne_s tamanho efetivo populacional baseado no número de reprodutores e matrizes; e ΔF , taxa de endogamia hipotética (em porcentagem) para a população com tal tamanho efetivo populacional (Adaptado de Pham et al., 2016).

Em outras espécies, como bovinos, as estimativas do tamanho efetivo da população geralmente são baixas (Ne de aproximadamente 100), mesmo com o grande número de animais que constituem uma determinada raça. O Ne é inversamente relacionado ao coeficiente de endogamia, que foram elevados devido à intensa e precisa seleção (Rodriguez-Ramilo et al, 2015).

Nesse contexto, a diversidade genética nas populações pecuárias contribui significativamente para a sustentabilidade da produção animal. Além disso, permite que a produção animal se torne mais sensível às mudanças ambientais e as demandas do mercado consumidor. A perda da diversidade genética pode resultar em um platô na produção e pode resultar em perda de aptidão ou viabilidade na produção animal. Assim, Makanjuola et al. (2020) investigaram o coeficiente de endogamia (ΔF), coeficiente de parentesco (Δf) e o tamanho efetivo da população (Ne) como importantes indicadores quantitativos da diversidade genética e avaliaram os efeitos da recente implementação da seleção genômica sobre a perda de diversidade genética nas raças Holandesa e Jersey norte americanas. Estes autores concluíram que os coeficientes de endogamia e de parentesco aumentaram após a implementação da seleção genômica como mostrado na tabela 2:

Tabela 2. Estimação do coeficiente de endogamia por pedigree (ΔF PED, %), coeficiente de endogamia genômico (ΔF ROH_SNP1101, ΔF ROH_PLINK, ΔF GRM, %), coeficiente de parentesco por pedigree (Δf PED, %) e coeficiente de parentesco genômico (Δf SEG, %) por ano , por geração e seu respectivo tamanho efetivo da população (Ne) de 1990 a 2018 na raças Holandesa e Jersey norte americanas

	Holandesa			Jersey		
Item	Ano	Geração	Ne	Ano	Geração	Ne
ΔF PED[1]	0,15	0,75	66	0,13	0,67	75
ΔF ROH_SNP1101[1]	0,22	1,10	46	0,12	0,62	81
ΔF ROH_PLINK[1]	0,23	1,16	43	0,12	0,63	80
ΔF GRM[1]	0,20	1,02	49	0,11	0,59	85
Δf PED[1]	0,20	0,98	51	0,14	0,73	68
Δf SEG[1]	0,20	0,98	51	0,15	0,78	64

[1]Erros-padrões para a estimação dos coeficientes de endogamia estavam <0,0001. Adaptado de Makanjuola et al. 2020.

Na aquicultura, pode acontecer a diminuição do tamanho efetivos da população (Ne) durante a seleção em massa resultante das grandes variações do sucesso reprodutivo em espécies altamente fecundas e difundidas, como as ostras e os camarões, onde as ninhadas podem ser constituídas por relativamente poucos indivíduos. Por isso, manter um Ne grande é essencial para minimizar o efeito da endogamia e perda de variabilidade genética ao selecionar para um melhor desempenho industrial (Xu et al, 2019).

Tradicionalmente, o tamanho efetivo da população e o coeficiente de endogamia são estimados a partir de informações de pedigree, no entanto, é necessário a integralidade das informações para as estimativas de Ne e ΔF. Recentemente, dados genômicos obtidos a partir de chips de SNP de alta densidade oferecem a oportunidade das estimativas do tamanho efetivo populacional e do coeficiente de endogamia com informações de todo o genoma. Por isso, espera-se que as estimativas baseadas no genoma sejam mais precisas que as estimativas usando apenas informações de relacionamento entre os indivíduos (Keller et al. 2011).

As estimativas do tamanho efetivo da população foram realizadas em diferentes espécies, especialmente em bovinos leiteiros. Qanbari et al. (2010) estimaram o Ne em bovinos holandeses alemães e este foi de aproximadamente 103 por quatro gerações. Rodriguez-Ramilo et al (2015) relatou Ne de 74 em bovinos holandeses espanhóis. Zhang et al. (2015) estimou o tamanho efetivo da população de sete gerações em ovinos de raças sunitas, alemães, Merinos e Dorper e; relataram Ne de 207, 74 e 67 respectivamente. Contudo, poucos estudos estimaram o Ne em aves (Pasandideh et al. 2020).

2.3 Endogamia

O coeficiente de endogamia (F) ou o nível de endogamia se refere à probabilidade de que dois alelos escolhidos ao acaso em um loco sejam idênticos por descendência. Associado à endogamia está a taxa de endogamia e a depressão endogâmica. Esta por sua vez compromete os valores fenotípicos, como a fertilidade, por exemplo, causado pelo aumento da endogamia em uma população (Falconer; Mackay, 1996). O coeficiente de endogamia é afetado pelo tamanho da população, pois, populações com menores Ne possuem maiores valores de coeficientes de endogamia e; pela proporção desigual de machos e fêmeas na população, pois as práticas realizadas para o aumento da eficiência reprodutiva dos animais, podem aumentar o coeficiente de endogamia (Norberg, 2007).

Além do coeficiente de endogamia, podem ser calculados a depressão endogâmica e a taxa de endogamia. A depressão por endogamia depende tanto do nível de dominância de uma característica quanto do coeficiente de endogamia (F); quanto maior o nível de dominância de uma característica, maior a depressão por endogamia. A taxa de depressão por endogamia tem uma relação linear negativa com o coeficiente de endogamia para a maioria das características (Falconer, 1996).

A taxa de endogamia expressa o aumento no nível médio de endogamia em uma população de uma geração para a seguinte. Como o aumento da endogamia é não linear, a taxa de endogamia é expressa em relação ao quanto a população está afastada da endogamia total. Com isso, em populações menores, o número efetivo de animais em reprodução ou o tamanho efetivo da população (Ne) podem determinar o incremento esperado de endogamia por geração (taxa de endogamia) (Falconer; Mackay, 1996; Olusegun, A. et al. 2012)

O conjunto de propriedades BLUP (Best Linear Unbiased Prediction ou Melhor Predição Linear não Viesada), permite usar informações de parentesco para se estimar o valor genético do indivíduo (VG), resultando no aumento da correlação dos Estimated Breeding Values (EBV) ou Valores Genéticos Estimados (VGEs) entre os indivíduos relacionados. Consequentemente a probabilidade de selecionar indivíduos aparentados aumenta, levando à elevação da taxa de endogamia. Com isso, a incorporação de informações de marcadores genéticos ao BLUP provoca uma redução da taxa de endogamia, pois as estimativas dos valores genéticos genômicos (Genomic Breeding Value Estimation) são baseados em sua maioria nas informações do indivíduo do que nas informações de parentesco (Belonsky; Kennedy, 1988).

O nível do coeficiente de endogamia e o tamanho efetivo da população podem ajudar a examinar a diversidade genética nas diferentes espécies, como é o caso de um estudo realizado com

a raça bovina Hereford nos EUA. Os coeficientes de endogamia foram computados de 1900 a 2001, concluindo que a endogamia aumentou rapidamente de 1900 a 1945 e a partir de 1946 aumenta linearmente chegando ao máximo de 11,5% em 1966 (Cleveland et al., 2005).

Segundo konig et al. (2010), aumentos de endogamia e de ganho genético estão associados aos reprodutores selecionados, a partir da avaliação da endogamia em galinhas poedeiras por meio da aplicação de teoria da contribuição e do fluxo gênico, de modo que o número de reprodutores influenciou no valor genético para a produção de ovos. Para o nível mais baixo de endogamia de 10% o número ideal de reprodutores foi de 70 e a estimativa do valor genético para produção de ovos foi de 127,9. Quando os níveis de endogamia foram para 16%, o número de reprodutores diminui para 15 e o valor genético médio aumentou para 139,7.

Usando dados demográficos de diferentes ninhadas de coelhos, Ragab et al. (2015) estimaram o tamanho efetivo da população e o coeficiente de endogamia e; seu efeito em características como, tamanho de ninhada, acumulado ao longo do tempo. No estudo foram utilizados 47794 informações e um pedigree de 14622 animais. Algumas práticas no gerenciamento de acasalamento e seleção permitiram um aumento no coeficiente de endogamia menor que 0,01 por geração, com proporção de 25 machos para 125 fêmeas. O tamanho efetivo da população (Ne) foi de 57,3; mostrando o efeito da seleção no aumento do coeficiente de endogamia (Ragab et al, 2015).

Os dados de SNPs (single nucleotide polymorphisms) também permitem a estimativa de endogamia no nível do genoma. Assim, em um estudo de Martikainen et al. (2017) foram estimados coeficientes de endogamia associados a características reprodutivas de 19075 vacas finlandesas da raça Ayrshire genotipadas com um painel de SNP de baixa densidade (8K). Os coeficientes médios de endogamia dos animais foram de 0,02; 0,09 0,63 para F_{PED} (coeficiente de endogamia estimado por pedigree), F_{ROH} (coeficiente de endogamia baseado em "runs of homozigosity") e F_{PH} (coeficiente de endogamia baseado na porcentagem de SNPs homozigotos), respectivamente. Concluiu-se que as estimativas de endogamia baseadas no genoma diferem das estimativas baseadas no pedigree, e que estimativas de endogamia genômica podem estar associadas à fertilidade (Martikainen et al., 2017).

2.4 Desequilíbrio de ligação (LD)

O desequilíbrio de ligação ou Linkage Disequilibrium (LD) é a associação não aleatória de alelos em dois ou mais loci, ou seja, a quantificação do nível de informação entre diferentes marcadores (Aerts et al., 2007; Qanbari et al., 2010). Com isso, os métodos mais utilizados para

quantificar LD são o coeficiente de correlação (r^2) e o D', sendo que, seus valores podem variar de zero (sem LD) a um (LD completo). Os valores de D' podem indicar ausência ou presença de recombinação e o r^2 é definido como o coeficiente de correlação dos alelos ao quadrado em dois locus, podendo ser medido pelo quadrado da diferença dos haplótipos dividido pelo produto das frequências alélicas. O r^2 é o mais utilizado, pois os valores do coeficiente D' podem superestimar o desequilíbrio de ligação (Lu et al., 2012). O cálculo do coeficiente de correlação dos alelos ao quadrado em dois locus (r^2) é descrito pela seguinte equação:

$$r^2 = \frac{D^2}{f(A)f(a)f(B)f(b)}$$

Sendo que $D = f(AB) - f(A)f(B)$ e $f(AB)$ são frequências haplotípicas e; $f(a), f(A)$ $f(B), f(b)$ são frequências alélicas dos marcadores.

O frango (*Gallus gallus*), como a maioria das espécies aviárias, tem um cariótipo bem distinto, consistindo em muitos cromossomos pequenos (micro-cromossomos) e alguns cromossomos maiores (macro-cromossomos). O genoma possui cinco macrocromossomos (GGA1-5) que variam em tamanho de 50 a 200Mb, cinco cromossomos intermediários (GGA6-10) que variam de 20 a 40 Mb e 28 microcromossomos (GGA11-38) com média de aproximadamente 12Mb. O menor dos microcromossomos foi estimado em menos de 5 Mb. Os microcromossomso têm diferenças estruturais em comparação com os cromossomos maiores, como o maior conteúdo de GC, por exemplo (Megens et al., 2009).

As distâncias intergênicas nos microcromossomos são menores, assim como o tamanho médio dos introns, resultando em uma densidade genética muito maior em comparação aos macrocromossomos. Taxas de recombinação em microcromossomos (50-10kb/cM) são muito maiores comparadas com os macrocromossomos (aproximadamente 300 kb/cM). Possivelmente, isso é resultado da exigência de pelo menos um quiasma por cromossomo por meiose e facilitada por uma maior densidade de ligação de sítios ativos das enzimas (Groenen et al., 2009).

Nesse contexto, a extensão do LD em frangos de corte estimados pela genotipagem, mostrou valores de r^2 abaixo de 0,30, em média, para distancias entre 0,25 e 1 cM (Megens et al., 2009). Independente dos baixos níveis de LD em frangos de corte, Andreescu et al. (2007) reportaram que há sobreposição significativa no LD de pares de marcadores entre diferentes populações e; esses resultados sugerem que haplótipos sejam compartilhados entre linhagens de frangos.

A densidade ideal de marcadores para identificar esses QTLs na ligação e em estudos de associação é amplamente determinado pela extensão do desequilíbrio de ligação (LD) em torno

deles. Aerts et al. (2007) investigaram a extensão do LD em dois cromossomos em duas linhagens de frangos de corte. Os níveis de LD foram calculados para 33 e 36 marcadores nos cromossomos 10 e 28, respectivamente. Descobriram que o "LD útil" (ou seja, um valor de r^2 superior a 0,30) em diferentes distâncias entre os marcadores (0,8 a 4 cM) nas linhagens de frangos de corte. Concluíram que, pode haver diferenças grandes nos padrões de LD entre diferentes raças de galinhas e diferentes regiões genômicas.

Utilizando aves de linhagens comerciais e nativas para a investigação do desequilíbrio de ligação (LD), taxa de recombinação e do tamanho efetivo da população (Ne) nos macro e microcromossomos, pôde-se concluir que existem diferenças na taxa de recombinação e no Ne de micro e macrocromossomos, que podem ser observados na tabela 3.

Tabela 3. Diferentes estimativas na taxa de recombinação e Ne entre micro e macrocromossomos

População	REC_{micro}/REC_{macro}	Ne_{macro}	Ne_{micro}	Ne_{micro}/Ne_{macro}
Linhagem ovo branco	3,20	37,61	26,75	1,41
Linhagem ovo marrom	4,14	160,31	147,34	1,09
Linha de frango de corte	4,16	813,41	751,68	1,08
Matriz de corte E5	1,59	249,41	88,03	2,83
Matriz de corte A3	3,03	1097,47	738,9	1,49
Galinha da selva vermelha	2,63	2473,92	1447,01	1,71
Coruja barbuda	1,97	876,67	383,86	2,28
Galinha Frisian	3,96	634,39	558,06	1,14
Média	2,76			1,63

Diferenças na taxa de recombinação (Rec.) e tamanho efetivo da população (Ne) entre micro e macro cromossomos estimados da adaptação de LD a curva Sved. As diferenças nas taxas de recombinação foram estimadas usando o mesmo Ne para micro e macro cromossomos. Por sua vez, as diferenças em Ne foram baseadas em uma taxa de recombinação 4,5 vezes maior nos microcromossomos em comparação aos macrocromossomos. Adaptado de Mengens et al., 2009.

Em um estudo foram utilizados dois tipos de chips de genotipagem independentes de 36 e 57k para caracterizar a extensão e estrutura do LD no genoma de quatro populações de galinhas. Foram analisados 454 animais de dois estabelecimentos comerciais e de duas populações experimentais de linhas brancas e marrom. O número de SNPs utilizados no estudo foi de 19 a 37k nas linhas marrons e de 8 a 19k nas linhas brancas. Concluiu-se, que existe uma expressiva diferença de LD entre linhas brancas e marrons. Um valor médio de r^2 de 0,73± 0,36 foram observados em distâncias pareadas <25Kb para linhas brancas comerciais, e ocorreu um decréscimo para 0,60±0,38 com distâncias de 75 a 120 Kb. Por outro lado, o valor médio de r^2 foi

de 0,32±0,33 para SNPs com menos de 25 Kb de distância entre eles e caiu para 0,21± 0,26 a uma distância de 100Kb em linhas marrons comerciais, este padrão semelhante de LD aconteceu nas duas linhas brancas e nas duas marrons, enquanto entre as linhas brancas e marrons o padrão de LD foi claramente diferente (Qanbari et al., 2010).

Tanto os estudos de associação genômica ampla (GWAS) quanto a seleção genômica (GS) dependem do nível de associação de alelos em diferentes loci, isto é, desequilíbrio de ligação (LD), pelo genoma. Portanto, para as duas abordagens é importante a caracterização do LD. Por isso, no estudo de Fu et al. (2015) foi usado um nucleotídeo de polimorfismo único (SNP) de 60K, estimado-se a estrutura do LD e haplótipos em frangos de corte cruzados e seus constituintes puros (uma linha masculina e duas femininas). O nível de LD (medido por r^2) entre SNPs adjacentes nos cromossomos autossômicos de frango estudados variou de 0,34 a 0,40 nas linhas puras, e foi de 0,24 nas populações mestiças, com 28,4% dos pares de SNPs com r^2 maior que 0,30. Ou seja, quando comparadas as linhas puras e mestiças, estas por sua vez, mostraram um nível mais baixo de LD, tamanhos menores de blocos de haplótipos e menor homozigosidade de haplótipos nos micros cromossomos.

A seleção de frangos para crescimento rápido sempre foi acompanhada do aumento da deposição de gordura, especialmente gordura abdominal, que pode diminuir a eficiência alimentar e pode causar doenças. Para o processo reprodutivo a descoberta de genes candidatos associados à deposição de gordura abdominal é essencial. Por isso, para identificar esses genes candidatos foram aplicadas análises de desequilíbrio de ligação e assinatura de seleção usando chip de SNPs de 60K em duas linhas de frangos selecionados divergentemente para teor de gordura abdominal por 11 gerações. Usando estes SNPs após o controlo de qualidade (46.033 SNPs), os autores encontraram r^2 era de 0,06 a 0,14 na linha leve e de 0,07 a 0,13 na linha pesada, para todos os 28 cromossomos (exceto GGA16). Distâncias pareadas do SNP maiores que 25 Kb mostraram r^2 médio igual a 0,33 na linha leve e r^2 médio de 0,32 na linha pesada (Na et al., 2019).

2.5 Seleção genômica (GS)

O sequenciamento do genoma da galinha e sua subsequente liberação em domínio público em 2004, lançou a indústria de criação de aves para o campo da genômica. As principais empresas de criação de aves investigam ativamente o uso de informações genômicas para melhorar programas de melhoramento genético e tornar a seleção cada vez mais eficiente (Fulton, 2012). A predição genômica é uma ferramenta para prever o valor genético de animais com base em milhares de marcadores moleculares e é aplicada na criação de animais domésticos (Meuwissen et al., 2001).

A GS tem sido muito utilizada na indústria de criação de animais, especialmente na pecuária (Loberg; D¨urr, 2009), possuindo grande potencial em espécies, como suínos e aves (Yan et al., 2018).

A seleção genômica usando painéis de SNPs de alta densidade foi implementada pela primeira vez em programas de criação de gado leiteiro. Proporcionou inúmeros benefícios em termos de redução no intervalo de gerações, maior precisão na seleção de animais mais jovens e reduções nos custos associados ao teste de progênies de touros em programas de melhoramento tradicionais (Hayes et al., 2009a).

Várias características da indústria avícola diferem acentuadamente da pecuária leiteira e influenciam a maneira pela qual a seleção genômica pode ser usada para o aprimoramento genético na criação de aves. Os programas tradicionais de melhoramento genético em aves já possuem intervalos de geração curtos (várias gerações sobrepostas por ano, com seleção a cada seis semanas em frangos de corte, sem sobreposição de gerações anuais em postura). No caso de aves de postura existem uma margem para encurtar o intervalo e geração, mas não tanto quanto no caso de sistemas de criação de bovinos leiteiros, onde o intervalo poderia ser reduzido de seis anos a menos de três anos (Schaeffer, 2006; Schefres e Weigel, 2012).

Em programas de melhoramento de aves são gerados números expressivos de candidatos à seleção e a intensidade de seleção é alta. Dezenas de milhares de candidatos a seleção são gerados por linhagem por geração, com apenas 1 a 3% dos machos selecionados para reprodução. Nas aves, o progresso genético criado em linhas puras é disseminado através de uma abrangente pirâmide de multiplicação de cruzamentos afetando muitas aves de postura e frangos comerciais. Assim, uma única linhagem contribui com 25% dos genes em aves comerciais em comparação com 100% nos sistemas de produção de raça pura em gado leiteiro (Wolc et al., 2016).

Na avicultura, nenhum método eficiente de criopreservação de sêmen está disponível. Assim, machos são altamente selecionados para serem utilizados apenas localmente, e cada galo só pode inseminar um número limitado de galinhas por tempo limitado. Por isso, todos estes fatores devem ser considerados para a implementação de qualquer tecnologia, especialmente da seleção genômica (Wolc et al., 2016). Segundo Misztal et al. (2013) a metodologia utilizada para a seleção genômica em uma situação comercial depende de atenção a detalhes, usando uma metodologia reconhecida e ter o conhecimento de questões específicas para a população estudada.

Na seleção genômica, uma população de referência é genotipada e são feitos registros para a característica de interesse, para assim estimar os efeitos do SNP. Em seguida, os candidatos à seleção são genotipados e, combinando seus genótipos com os efeitos estimados, o EBV genômico

(GEBV) é estimado para os candidatos à seleção. Pode-se notar que a abordagem GS não requer registro de pedigree, essencial para o BLUP tradicional (Meuwissen et al., 2016).

Foram realizados vários experimentos para testar se a seleção genômica pode alcançar ganhos mais rápidos do que a seleção tradicional. Wolc et al. (2015) dividiram uma população em duas sublinhas; uma foi submetido à seleção fenotípica convencional e outra foi selecionada com base na predição genômica. O período experimental teve duração de três anos, período em que foram realizados quatro ciclos de seleção genômica e dois de seleção fenotípica. Posteriormente, as duas sub-linhas foram comparadas para várias características de desempenho relevantes economicamente para a produção de ovos. Conclui-se que a sub-linha selecionada pelo uso de informações genômicas superou a sub-linhas de seleção fenotípica para a maioria das 16 características inclusas no índice usado para seleção.

Mesmo que os dois programas de melhoramento citados no estudo anterior tenham sido projetados para atingir o mesmo coeficiente de endogamia por ano, Wolc et al. (2015) verificaram que a endogamia realizada por ano avaliada a partir da linhagem foi maior na sub-linha em que foi aplicada a seleção genômica do que na sub-linha selecionada convencionalmente. A seleção genômica pode ser utilizada para melhorar o desempenho de mestiços em um ambiente comercial e para características que não podem ser registradas nos rebanhos núcleos, como testes de desafio à doença. (Meuwissen et al. 2016).

Outros experimentos foram realizados em aves de postura e frangos de corte (Misztal et al., 2013; Heidaritabar et al., 2014; Wolc et al., 2015), para a verificação das premissas da seleção genômica em termos de maior precisão, maior resposta à seleção e oportunidades de redesenhar os programas de melhoramento para maximizar os benefícios dessa técnica de avaliação genética Essas experiencias iniciais forneceram conhecimentos sobre a aplicação prática dos programas de melhoramento contendo informações genômicas, como a genotipagem em larga escala com rastreamento adequado de amostras (Wolc et al. 2016).

O ssGBLUP (Single-step genomic best linear unbiased prediction) foi proposto por Misztal et al. (2009) e Christensen e Lund (2010), e se tornou um método para estimar valores genômicos de animais que possuem informações de pedigree ou não; e especialmente animais que não tenham dados fenotípicos conhecidos para um procedimento de avaliação genética. Este procedimento por sua vez, não requer várias etapas para calcular pseudo-fenótipos e permite extensões para modelos mais complexos (Vitezica et al., 2011). Por isso, comparada com uma avaliação de várias etapas, a execução de uma avaliação BLUP regular é modificada para o uso da matriz H, matriz de relacionamento genômico, e as outras etapas são eliminadas. Foram Legarra et al. (2009) e Christensen e Lund (2010) desenvolveram essa matriz e Aguilar et al. (2010)

demonstraram que uma metodologia de etapa única pode ser mais simples, rápida e precisa (Misztal et al., 2013).

A matriz de relacionamento genômico (matriz **H**) pode ser obtida começando com os animais genotipados e, em seguida, usando o pedigree para calcular as relações que envolvem os descendentes não genotipados dos animais que possuem informações genômicas (Meuwissen et al., 2011). Em um estudo com bovinos leiteiros, ssGBLUP produziu de 0 a 2% mais precisão do que os métodos de várias etapas (Legarra et al., 2014), mas para outras espécies, que são menos dominadas por grandes famílias de touros, a diferença de precisão entre ssGBLUP e os métodos de várias etapas pode ser maior. Mesmo assim, uma das falhas do método de etapa única (ssGBLUP) é que este não funciona para a estimativa não linear, embora algumas soluções para a única estimativa não linear tenham sido propostas na literatura (Liu et al., 2014; Legarra e Ducrocq, 2012).

Por causa das oportunidades limitadas para redução dos intervalos de geração, o principal benefício da seleção genômica sobre avaliações baseadas em pedigree em aves domésticas está relacionado com aumentos na precisão dos valores estimados de reprodução na puberdade e para características limitadas por sexo. Precisões dos GEBVs em aves de postura e frangos de corte para produção, qualidade do produto, características de reprodução e bem-estar foram avaliadas usando métodos de uma etapa (Chen et al., 2011a, 2011b), métodos bayesianos (Wolc et al., 2011b; Wang et al., 2013), Bayesiano LASSO (Liu et al., 2014), métodos não paramétricos (González-Recio et al., 2008), métodos que separam a variação genética em regiões codificantes e não codificantes (Abdollahi-Arpanahi et al., 2014) e abordagens que capitalizaram e incluíram informações de anotação (Morota et al., 2014).

Em todos esses estudos, os GEBVs foram mais precisos que os EBVs baseados em pedigree, mas não houve nenhuma superioridade clara entre os diferentes métodos baseados em genômica, ou seja, o método superou consistentemente outros métodos entre características e populações. O uso da predição genômica foi particularmente promissor para características limitadas por sexo, de difíceis mensuração, mensuração onerosa, ou medidas no final da vida dos animais. Para características para as quais os fenótipos eram desconhecidos nos candidatos à seleção (ou seja, produção e qualidade de ovos em galos), predições genômicas capturaram informações sobre termos de amostragem mendelianos e, assim, permitiu a seleção dentro da família, em contraste com as predições tradicionais baseadas no pedigree (Wolc et al., 2016).

Referências

Abdollahi-Arpanahi, R. et al. Dissection of additive genetic variability for quantitative traits in chickens using SNP markers. **Journal of Animal Breeding and Genetics**, v. 131, n. 3, p. 183–193, 2014.

Aerts, J. et al. Extent of linkage disequilibrium in chicken. Cytogenetic and Genome Research, v. 117, n. 1–4, p. 338–345, 2007.

Ageitec, Agência Embrapa de Informação tecnológica. Material genético. Acesso em: 21/07/2020.

Andreescu, C. et al. Linkage disequilibrium in related breeding lines of chickens. Genetics, v. 177, n. 4, p. 2161–2169, 2007.

Belonsky, G.M.; Kennedy, B.W. Selection on individual phenotype and best linear unbiased predictor of breeding value in a closed swine herd. Journal Animal Science, v.66, p.1124-1131, 1988.

Bourdon, R. M. Understanding animal breeding. New Jersey. Prentice – Hall, 2000.

Cleveland, M. A. et al. Changes in inbreeding of U.S. Herefords during the twentieth century. Journal of Animal Science, v. 83, n. 5, p. 992–1001, 2005.

Christensen, O. F.; LUND, M. S. Genomic prediction when some animals are not genotyped. Genetics Selection Evolution, v. 42, n. 3, p. 1–8, 2010.

Falconer, D.S.; Mackay, T.F.C. Introduction to Quantitative Genetics. 4th Edition, Prentice Hall, Essex, 1996.

Filho, K. E. **Melhoramento Genético Animal No Brasil: Fundamentos, História E Importância**. [s.l: s.n.] 2009.

Fu, W. et al. Linkage disequilibrium in crossbred and pure line chickens. **Genetics Selection Evolution**, v. 47, n. 1, p. 1–12, 2015.

Fulton, J. E. Genomic selection for poultry breeding. **Animal Frontiers**, v. 2, n. 1, p. 30–36, 2012.

Gaya, L. G. et al. Heritability and genetic correlation estimates for performance and carcass and body composition traits in a male broiler line. **Poultry Science**, v. 85, n. 5, p. 837–843, 2006.

González-Recio, O. et al. Nonparametric methods for incorporating genomic information into genetic evaluations: An application to mortality in broilers. **Genetics**, v. 178, n. 4, p. 2305–2313, 2008.

Hayes, B. J. et al. Accuracy of genomic breeding values in multi-breed dairy cattle populations. Genetics Selection Evolution, v. 41, n. 1, p. 1–9, 2009a.

Heidaritabar, M. et al. Systematic differences in the response of genetic variation to pedigree and genome-based selection methods. **Heredity**, v. 113, n. 6, p. 503–513, 2014.

Júnior, J. M. C. Melhoramento Genético Animal. **Livro EMBRAPA**, p. 197–208, 2010.

Keller, M. C.; Visscher, P. M.; Goddard, M. E. Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. **Genetics**, v. 189, n. 1, p. 237–249, 2011.

Konig, S. et al. Evaluation of inbreeding in laying hens by applying optimum genetic contribution and gene flow theory. Poultry Science, v. 89, n. 4, p. 658–667, 2010

Legarra, A.; Aguilar, I.; Misztal, I. A relationship matrix including full pedigree and genomic information. Journal of Dairy Science, v. 92, n. 9, p. 4656–4663, 2009.

Legarra, A.; Ducrocq, V. Computational strategies for national integration of phenotypic, genomic, and pedigree data in a single-step best linear unbiased prediction. **Journal of Dairy Science**, v. 95, n. 8, p. 4629–4645, 2012.

Legarra, A. Et Al. Single Step, a general approach for genomic selection. **Livestock Science**, v. 166, n. 1, p. 54–65, 2014.

Leroy, G. et al. Methods to estimate effective population size using pedigree data: Examples in dog, sheep, cattle, and horse. Genetics Selection Evolution, v. 45, n. 1, p. 1–10, 2013.

Liu, Z. et al. A single-step genomic model with direct estimation of marker effects. **Journal of Dairy Science**, v. 97, n. 9, p. 5833–5850, 2014.

Liu, T. et al. Accuracy of genomic prediction for growth and carcass traits in Chinese triple-yellow chickens. **BMC genetics**, v. 15, p. 110, 2014.

Loberg, A.; Dürr, J. Interbull survey on the use of genomic information. in: Proc. Interbull International Workshop on Genomic Information in Genetic Evaluations. Interbull Bulletin No. 39.: 3–14, 2009.

Lu, Y. et al. Polymorphisms in Wnt signaling pathway genes are significantly associated with chicken carcass traits. Poultry Science, v. 91, n. 6, p. 1299–1307, 2012.

Makanjuola, B. O. et al. Effect of genomic selection on rate of inbreeding and coancestry and effective population size of Holstein and Jersey cattle populations. **Journal of Dairy Science**, n. 1, 2020.

Martikainen, K. et al. Estimation of inbreeding depression on female fertility in the Finnish Ayrshire population. **Journal of Animal Breeding and Genetics**, v. 134, n. 5, p. 383–392, 2017.

Megens, H. J. et al. Comparison of linkage disequilibrium and haplotype diversity on macro- and microchromosomes in chicken. BMC Genetics, v. 10, p. 1–11, 2009.

Meuwissen, T. H et al. Predic- tion of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829, 2001.

Meuwissen, T. H. E.; Luan, T.; Woolliams, J. A. The unified approach to the use of genomic and pedigree information in genomic evaluations revisited. **Journal of Animal Breeding and Genetics**, v. 128, n. 6, p. 429–439, 2011.

Meuwissen, T.; Hayes, B.; Goddard, M. Genomic selection: A paradigm shift in animal breeding. **Animal Frontiers**, v. 6, n. 1, p. 6–14, 2016.

Misztal, I.; Legarra, A.; Aguilar, I. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. **Journal of Dairy Science**, v. 92, n. 9, p. 4648–4655, 2009.

Misztal (2013). Single-step Genome evaluation. p. 2530–2534, 2010.

Morota, G. et al. Genome-enabled prediction of quantitative traits in chickens using genomic annotation. **BMC Genomics**, v. 15, n. 1, p. 1–10, 2014.

Na, W. et al. Important candidate genes for abdominal fat content identified by linkage disequilibrium and fixation index information. Poultry Science, v. 98, n. 2, p. 581–589, 2019.

Norberg, E.; Sørensen, A. C. Inbreeding trend and inbreeding depression in the Danish populations of Texel, Shropshire, and Oxford Down. Journal of Animal Science, v. 85, n. 2, p. 299–304, 2007.

Olusegun, A. et al. We are Intech Open, the world' s leading publisher of Open Access books Built by scientists, for scientists TOP 1 %. **Intech**, v. i, n. tourism, p. 38, 2012.

Pasandideh, M.; Gholizadeh, M.; Mianji, G. R. Estimation of effective population size and genomic inbreeding coefficients in baluchi sheep using wide single nucleotide polymorphisms (SNPs). **Iranian Journal of Applied Animal Science**, v. 10, n. 2, p. 287–295, 2020.

Pham, M. H. et al. Monitoring of genetic diversity in Taiwan conserved chickens assessed by pedigree and molecular data. **Livestock Science**, v. 184, p. 85–91, 2016.

Qanbari, S. et al. The pattern of linkage disequilibrium in German Holstein cattle. **Animal Genetics**, v. 41, n. 4, p. 346–356, 2010.

Qanbari, S. et al. Linkage disequilibrium reveals different demographic history in egg laying chickens. **BMC Genetics**, v. 11, 2010.

Ragab, M.; Sánchez, J. P.; Baselga, M. Effective population size and inbreeding depression on litter size in rabbits. A case study. Journal of Animal Breeding and Genetics, v. 132, n. 1, p. 68–73, 2015.

Rodríguez-Ramilo, S. T. et al. Genome-Wide estimates of coancestry, inbreeding and effective population size in the spanish holstein population. **PLoS ONE**, v. 10, n. 4, p. 1–11, 2015.

Rovadoscki, G. A. et al. Genetic parameters for growth characteristics of free-range chickens under univariate random regression models. **Poultry Science**, v. 95, n. 9, p. 1989–1998, 2016.

Schaffner, S.F. et al. Calibrating a coalescent simulation of human genome sequence variation. Genome Research, 15, 1576–1583, 2005.

Schefers, J. M.; Weige, K. A. Genomic selection in dairy cattle: Integration of DNA testing into breeding programs. **Animal Frontiers**, v. 2, n. 1, p. 4–9, 2012.

Silva, Evolução do melhoramento genético de aves no Brasil Genetic improvement of Poultry in Brazil. **Ceres**, v. 56, n. 3, p. 437–445, 2009.

Vitezica, Z. G. et al. Bias in genomic predictions for populations under selection. Genetics Research. 93:357–366, 2011.

Xu, L. et al. Genetic diversity and effective population size in successive mass selected generations of black shell strain Pacific oyster (Crassostrea gigas) based on microsatellites and mtDNA data. **Aquaculture**, v. 500, p. 338–346, 2019.

Wang, C. et al. Accuracy of Genomic EBV Using an Evenly Spaced, Low-density SNP Panel in Broiler Chickens. **Animal Industry …**, p. 2620, 2011.

Wolc, A. et al. Breeding value prediction for production traits in layer chickens using pedigree or genomic relationships in a reduced animal model. **Genetics Selection Evolution**, v. 43, n. 1, p. 1–9, 2011.

Wolc, A. et al. Response and inbreeding from a genomic selection experiment in layer chickens. **Genetics Selection Evolution**, v. 47, n. 1, p. 1–12, 2015.

Wolc, A. et al. Implementation of genomic selection in the poultry industry. **Animal Frontiers**, v. 6, n. 1, p. 23–31, 2016.

Wright, S. Size of population and breeding structure in relation to evolution. Science 87:430-431, 1938.

Wright, S. Evolution in Mendelian Populations. **Genetics**, v. 16, p. 97-159, 1931.

Yan, Y. et al. Genomic prediction in a nuclear population of layers using single-step models. Poultry Science, v. 97, n. 2, p. 397–402, 2018.

Zhang, Q. et al. Estimation of inbreeding using pedigree, 50k SNP chip genotypes and full sequence data in three cattle breeds. **BMC Genetics**, v. 16, n. 1, p. 1–11, 2015.

3 SIMULAÇÃO DE ESTRATÉGIAS DE SELEÇÃO E TAMANHO EFETIVO POPULACIONAL DE REBANHOS DE LINHAGENS CAIPIRAS NACIONAIS

RESUMO

Para o desenvolvimento sustentável, a produção de linhagens de aves caipiras, seria uma alternativa de aliar a preservação da diversidade e variabilidade genética, especialmente a conservação de qualidades específicas desses animais, principalmente relacionadas ao crescimento lento e as qualidades organolépticas da sua carne. Neste contexto, para melhor entendimento das estratégias de seleção e acasalamentos utilizados; e compreender se o tamanho efetivo populacional de rebanhos de aves caipiras nacionais está adequado foi realizado este estudo de simulação. Foi simulado uma geração histórica que por mil gerações teve um tamanho constante de 2000 indivíduos, ocorrendo um gargalo genético nas gerações subsequentes (1010 a 1020) para 400 animais, a qual expandiu para 1000, 500, 200 e 100 animais de onde surgiram as fases (população recente) dos quatro cenários simulados e para três coeficientes de herdabilidade (0,15; 0,30 e 0,45). Na Fase 1 os indivíduos não passaram por nenhum processo de seleção por quarenta gerações (G_0 a G_40). Posteriormente nas Fases 2 e 3, os indivíduos foram selecionados por características fenotípicas por 25 gerações (G_41 a G_65) e pelos valores genéticos preditos (BLUP-EBV) nas últimas 15 gerações (G_66 a G_80), respectivamente. Os cenários 1, 2, 3 e 4 possuíram tamanhos efetivos da população de 1000, 640 e 640; 500, 320 e 320; 200, 128 e 128; 100, 64 e 64 nas Fases 1, 2 e 3, respectivamente. Foram calculados coeficientes de endogamia, taxa de homozigose, tendências genéticas e os ganhos genéticos para os quatro cenários e três coeficientes de herdabilidade simulados. Para todos os cenários estudados, a geração zero (G_0) das três fases da população recente (Fase 1, Fase 2 e Fase 3) teve um coeficiente de endogamia igual a zero. Além disso, os valores obtidos para os coeficientes de endogamia aumentaram da primeira para a última geração de cada uma das fases estudadas. Os maiores coeficientes de endogamia foram obtidos nos cenários com menor Ne, assim como, os valores das taxas de homozigose. Os valores de ganhos genéticos foram maiores para o coeficiente de herdabilidade de 0,45 dos quatro cenários simulados com diferentes tamanhos efetivos da população. Com base nos resultados desse estudo de simulação pode-se concluir que o tamanho efetivo da população (Ne) dos rebanhos de linhagens caipiras da ESALQ deve ser aumentado; e as estratégias de seleção devem ser revisadas e aplicadas para a minimização dos níveis de endogamia e homozigose. Dessa forma, os ganhos genéticos destas populações não seriam afetados e podem ser maximizados

Palavra-chave: endogamia; homozigose; populações; tamanho efetivo populacional

ABSTRACT

For sustainable development, the production of free-range lines would be an alternative to combine the preservation of genetic diversity and variability, especially the conservation of specific qualities of these animals, mainly related to the slow growth and the organoleptic qualities of their meat. In this context, for a better understanding of the selection and mating strategies used; and to understand if the effective population size of flocks of national free-range birds is adequate, this simulation study was carried out. A historical generation was simulated, which for a thousand generations had a constant size of 2000 individuals, with a genetic bottleneck occurring in subsequent generations (1010 to 1020) for 400 animals and expanded to 1000, 500, 200 and 100 animals from which the phases emerged (population recent) of the four simulated scenarios and for three coefficients of heritability (0.15; 0.30 and 0.45). In Phase 1, individuals did not undergo any selection process for forty generations (G_0 to G_40), later in Phases 2 and 3, individuals were selected for phenotypic characteristics for 25 generations (G_41 to G_65) and the predicted genetic values (BLUP -EBV) in the last 15 generations (G_66 to G_80), respectively. Scenarios 1, 2, 3 and 4 had effective population sizes of 1000, 640 and 640; 500, 320 and 320; 200, 128 and 128; 100, 64 and 64 in Phases 1, 2 and 3, respectively. Inbreeding coefficients, homozygosity rate, genetic trends, and genetic gains were calculated for the four scenarios and three simulated heritability coefficients. For all studied scenarios, the zero generation (G_0) of the three phases of the recent population (Phase 1. Phase 2 and Phase 3) had an inbreeding coefficient equal to zero. In addition, the values obtained for the inbreeding coefficients increased from the first to the last generation of each of the studied phases and the highest inbreeding coefficients were obtained in the scenarios with the lowest Ne, as well as the values of the homozygosity rates. The values of genetic gains were higher for the heritability coefficient of 0.45 of the four simulated scenarios with different effective population sizes. Based on the results of this simulation study, it can be concluded that the effective population size (Ne) of herds of free-range lineages must be increased; and selection strategies should be reviewed and applied to minimize levels of inbreeding and homozygosity; so that the genetic gains of these populations are not affected and can be maximized

Keyword: inbreeding; homozygosis; populations; effective population size

3.1 Introdução

A demanda global por alimentos tem aumentado devido ao crescimento da população mundial e pela melhoria nos padrões de vida (Teng et al 2019). Os consumidores, especialmente de países desenvolvidos, tem buscado uma alimentação mais saudável, exigindo que os animais tenham sido criados em condições de bem-estar favoráveis, em que as normas de proteção ambiental tenham sido respeitadas, que o produto final seja de qualidade e que não sejam industrializados (Silva et al., 2003; Hellmeister Filho et al., 2003; Coelho et al., 2007; Vanhonacker; Verbeke, 2009; Allais et al., 2018). Mesmo assim, estudos com aves nativas ainda são escassos na literatura, especialmente os que englobam linhagens de postura, corte e de dupla aptidão, por mais

que estas tenham uma pequena participação no mercado avícola brasileiro (Silva, 2009; Rovadoscki et al., 2016).

No contexto do desenvolvimento sustentável, a produção de linhagens de aves nativas seria uma alternativa de aliar a preservação da diversidade e variabilidade genética e uma adaptação às novas exigências do mercado rápido e fácil, preservando as qualidades específicas desses animais, principalmente relacionadas ao seu crescimento lento e as qualidades organolépticas da sua carne (Allais et al., 2018). A gestão da diversidade genética de uma raça doméstica é essencial para o uso sustentável de recursos genéticos e seu melhoramento (Falconer e Mackay, 1996). Por isso, um dos parâmetros a serem trabalhados seria o controle da endogamia, que é fundamental para a produção e conservação genética animal, pois induz à redução da diversidade genética (Joaquim et al., 2019).

Para isso, a estruturação populacional é crucial para a identificação e interpretação corretas das associações moleculares com a diversidade fenotípica, ou seja, e a conservação dos recursos genéticos é um tópico de interesse mundial durante as últimas décadas (Pritchard e Rosenberg, 1999**;** FAO, 2011). O progresso genético acelerado nas características produtivas de animais de criação é resultado da evolução da tecnologia reprodutiva, da aplicação de ferramentas genéticas modernas e da melhor logística. No entanto, um aumento no uso de raças altamente selecionadas e produtivas tem sido associado à perda de diversidade genética na maioria das espécies de animais domésticos (Dalvit et al., 2009; Ciampolini et al., 2013), especialmente na avicultura (Hillel et al., 2003; Blackburn 2006; Zanetti et al., 2011a, b; Viale et al., 2017).

Unindo o desenvolvimento sustentável e a preservação de recursos genéticos, a Escola Superior de Agricultura "Luiz de Queiroz" (ESALQ/USP) possui um pequeno plantel composto por oito linhagens caipiras nacionais especializadas para postura, corte e dupla aptidão. Foram desenvolvidas por um processo de seleção massal em um programa de melhoramento genético, que é conduzido desde a década de 80 e foi base para este estudo. Por isso, para melhor entendimento das estratégias de seleção e acasalamentos utilizados; e para compreender se o tamanho efetivo populacional desses rebanhos de aves caipiras nacionais da ESALQ está adequado foi realizado este estudo de simulação.

3.2 Material e métodos

O programa de melhoramento genético realizado na Escola Superior de Agricultura "Luiz de Queiroz" é composto por oito linhagens caipiras com aptidões para postura, corte e dupla aptidão. A linhagem Caipirão da ESALQ é selecionada para peso, com fenótipo variado para cor de plumagem; o 7P (pinto preto pesado de pasto de pescoço pelado de Piracicaba), selecionado

para peso, apresenta crescimento rápido, plumagem preta e pescoço pelado (gene Na); Caipirinha da ESALQ, de dupla aptidão (corte e postura), com crescimento lento, plumagem de cor variada e topete (gene Cr); a linhagem Azul, ave selecionada para postura, possui ovos de diversas tonalidades de azul em sua população e têm fenótipos variados para a cor da plumagem; o Carijó Barbada, ave de dupla aptidão (corte e postura), crescimento lento, plumagem barrada e apresenta barba e costeletas (gene Mb); a *Soo Ho*, linhagem exótica de dupla aptidão que possui a sua carne escurecida (carne preta) pelo processo de fibromelanose ou hiperpigmentação; a Baianinha da ESALQ e Pernalonga da ESALQ, aves selecionadas para características ornamentais. Todas pertencem ao projeto "Frango Feliz" da ESALQ e que hoje participam do projeto de melhoramento genético coordenado pelo Departamento de Zootecnia da ESALQ. Com isso, para que de forma mais rápida, precisa e economicamente viável, foi realizado um estudo de simulação para o melhor entendimento das ações que são executadas no programa de melhoramento em condições reais, de foma mais rápida, precisa e economicamente viável.

3.2.1 Processo de simulação

As populações das linhagens caipiras foram simuladas hipoteticamente em duas etapas através do software de simulação *QMSim*, que pode simular genomas gigantescos e complexas estruturas de pedigree, mimetizando as populações de animais domésticos. Ou seja, é basicamente um simulador que se baseia em famílias, que pode ou não levar em conta características evolutivas predefinidas, como o desequilíbrio de ligação (LD), mutações, gargalos genéticos e expansões (Sargolzaei; Schenkel, 2009).

Na etapa 1, a geração histórica foi constituída de 2000 indivíduos (1.000 machos e 1.000 fêmeas) e percorreu 1.000 gerações em tamanho constante. Na geração 1.010 ocorreu um gargalo genético, com 400 animais (200 machos e 200 fêmeas) para a formação do desequilíbrio de ligação; e constante até a geração 1.020 que se expandiu em diferentes tamanhos efetivos da população (Ne) até a geração 1.100 de onde advém a população recente que é dividida em três fases (Fases 1, 2 e 3) em quatro cenários simulados (etapa 2). Os cenários 1, 2, 3 e 4 possuíram tamanhos efetivos da população de 1.000, 640 e 640; 500, 320 e 320; 200, 128 e 128; 100, 64 e 64 nas fases um, dois e três (Fase 1, Fase 2 e Fase 3), respectivamente. Sendo assim, o cenário 4 é o mais próximo do número de animais efetivamente em reprodução nos rebanhos das linhagens caipiras pertencentes ao programa de melhoramento realizado em condições reais (Figura 1).

O processo de simulação percorreu 80 gerações para que se tivesse um cenário de um programa de melhoramento genético de aves que tenha sido desenvolvido ao longo desse período.

Assim, os indivíduos da primeira geração (G_0) à quadragésima geração (G_40) não foram selecionados (Fase 1). Nas 25 gerações subsequentes (entre G_41 e G_65 – Fase 2) os animais foram selecionados por características fenotípicas e a Fase 3 os indivíduos foram selecionados por valores genéticos estimados (EBV - BLUP), descrevendo hipoteticamente o início do uso do modelo animal nas avaliações genéticas dos animais domésticos. Os animais foram gerados a partir dos acasalamentos realizados na última geração da fase anterior (Figura 1).

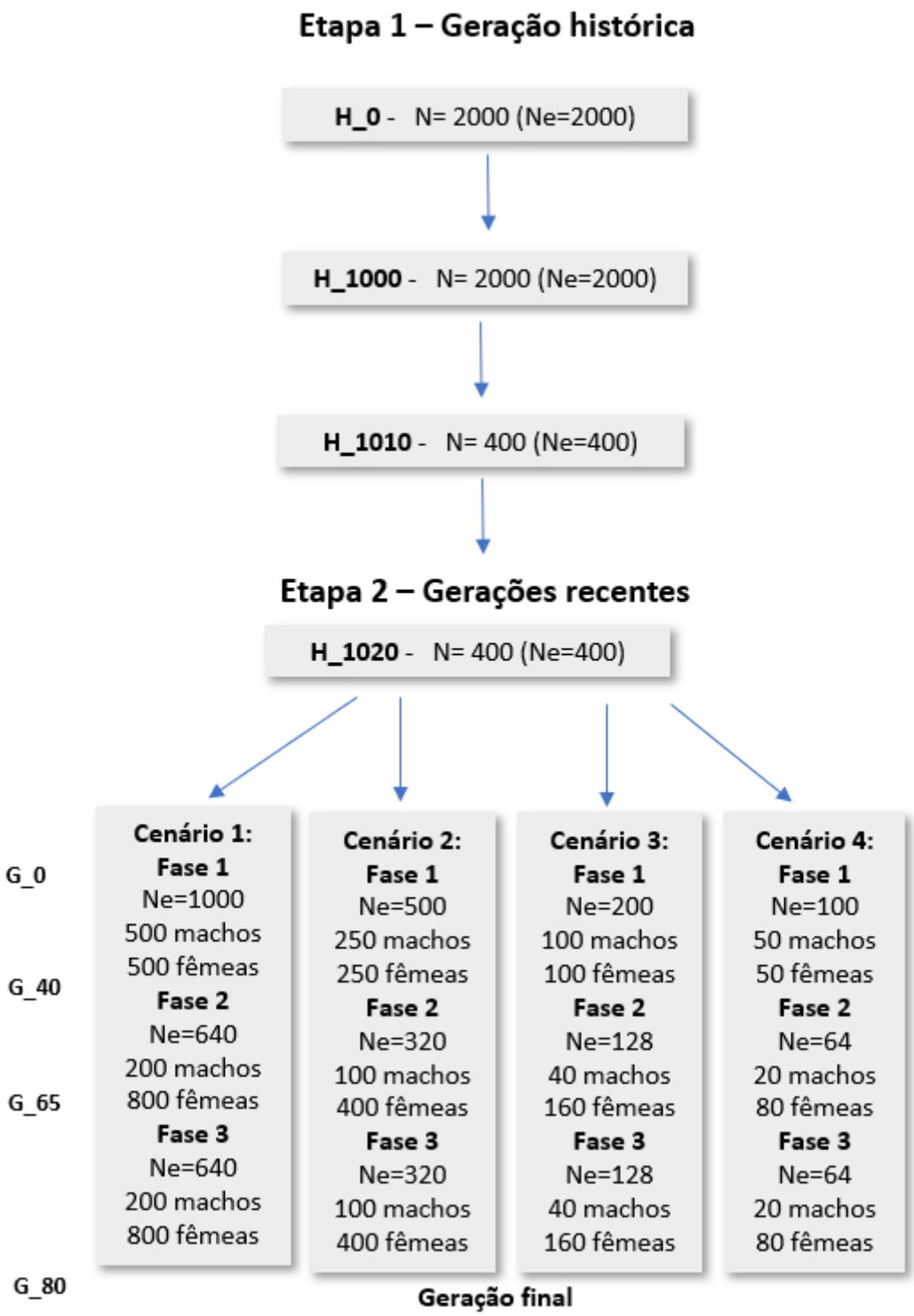
Na Fase 1 os acasalamentos foram construídos na proporção de um macho para uma fêmea (1:1) e nas Fases 2 e 3 os acasalamentos foram construídos com um macho para quatro fêmeas (1:4) a cada geração, estes por sua vez, são os acasalamentos utilizados no programa de melhoramento genético de linhagens caipiras base para este trabalho. Para melhor compreensão das características de importância econômica que são utilizadas em programas de melhoramento de aves, as populações também foram simuladas considerando três diferentes coeficientes de herdabilidade (0,15, 0,30 e 0,45), os quais são próximos aos coeficientes de herdabilidade estimados por Gaya et al. (2006), de 0,16 para a taxa de conversão alimentar, 0,33 para teor de gordura abdominal e 0,40 para a característica de peso aos 38 dias.

Foram realizadas 10 repetições do processo de simulação. A variância fenotípica utilizada foi de 1,0 e a geração histórica na etapa 1 foi constituída de 1100 gerações. Para que fossem geradas as informações da população recente (Etapa 2), o número de machos da última geração histórica foram de 500, 250, 100 e 50, respectivamente para os quatro cenários simulados. Diferente do que acontece com bovinos e equinos, que os machos deixam maior número de descendentes para as próximas gerações que as fêmeas, na avicultura, machos e fêmeas podem deixar descendentes de forma proporcional. Por isso, para cada casal acasalado no processo de simulação o tamanho de suas ninhadas foi de 8 filhotes, tendo uma proporção de machos nessas progênies de 0,5.

As relações de substituição de machos e fêmeas no plantel de aves são maiores que as utilizadas em bovinos, por exemplo. Por isso, para esse estudo, foram utilizadas relações de substituição de 1, que refere a troca de 100% do plantel a cada geração. Os projetos de seleção como descritos, foi aleatório na Fase 1, por Fenótipo na fase 2 e EBV na Fase 3 simuladas.

Para que os dados fossem ainda mais próximos das condições reais o genoma dos indivíduos foi simulado. O genoma consistiu em 28 pares de cromossomos autossomos de comprimentos semelhantes ao genoma real da galinha baseado na montagem de *Gallus gallus* (galGal6 - NCBI, 2020), totalizando 945 cM (centimorgans), para se criar um cenário mais verídico com relação ao número de loci de marcadores e QTLs fisicamente desvinculados. Os 600.012 marcadores de SNP foram distribuídos uniformemente e o número inicial de marcadores foi escolhido de tal forma que geraria uma densidade de loci bi alélicos segregantes com menor

frequência alélica (MAF)> 0,02: 600K. Os marcadores foram neutros em seu efeito sobre os QTLs. Um número de QTL foi simulado para gerar 336 loci segregantes com dois alelos e MAF> 0,02; cujas posições foram distribuídas aleatoriamente. Os efeitos alélicos aditivos foram amostrados aleatoriamente a partir da distribuição gama com parâmetro de forma igual a 0,4. A taxa de marcadores de genótipos em falta foi de 0,01 e a taxa de erro de genotipagem de marcadores foi de 0,005. Uma taxa de mutação recorrente de 10^{-5} para ambos os marcadores e QTLs foi considerada para estabelecer o equilíbrio mutação-deriva em gerações históricas. A mesma taxa de mutação também foi aplicada em todas as gerações subsequentes após as históricas. Os parâmetros e as etapas do processo de simulação estão descritos na Figura 1 e nos anexos desse estudo.



H_0: Geração histórica zero; H_1000: Geração histórica 1000; H_1010: Geração histórica 1010; H_1020: Geração histórica 1020; Fase 1, Fase 2 e Fase 3: Fases; Ne: Tamanho efetivo da população; G_0: Geração zero recente; G_40: Geração 40 recente; G_65: Geração 65 recente e G_80: Geração 80 recente.

Figura 1 - Representação esquemática das etapas de simulação

3.2.2 Avaliações genéticas

Os valores genéticos estimados (EBVs) foram baseados nos coeficientes de herdabilidades das características reais de peso corporal aos 38 dias, taxa de conversão alimentar e gordura abdominal e foram previstos usando a família de programas BLUPF90 (Misztal, 2002). Foi utilizado o seguinte modelo animal (Sargolzaei; Schenkel, 2009):

$$\boldsymbol{y} = \mathbf{1}\boldsymbol{\mu} + \boldsymbol{Za} + \boldsymbol{e}$$

sendo,

$\boldsymbol{y}$ é o vetor de fenótipos observados;

μ é a média geral e 1 é o vetor de "uns";

Z é uma matriz de incidência para os efeitos aleatórios de animal;

$\boldsymbol{a}$ é o vetor aleatório de efeitos genéticos aditivos, sendo a ~N (0, Aσ_a^2);

$\boldsymbol{e}$ é o vetor de efeitos aleatório dos resíduos.

A é a matriz que de relacionamento dos animais, via informação de pedigree; e $\boldsymbol{\sigma}_a^2$ é a variância genética aditiva.

3.2.2.1 Tamanho efetivo da população

O tamanho efetivo das populações foi calculado com base no plano de acasalamento, seguindo o modelo de Wright (1931) pela seguinte fórmula (Leroy et al, 2013; Pham et al, 2016):

$$N_{es} = \frac{4\, N_m N_f}{N_m + N_f}$$

Para a computação de Ne, foi requerido apenas o número estimado de machos reprodutores (Nm) e fêmeas (Nf) na população de referência.

3.2.2.2 Endogamia e homozigose

O coeficiente de endogamia populacional médio na geração t foi predito por:

$$Ft = 1 - \left(1 - \frac{1}{2Ne}\right)^t$$

onde Ne é o tamanho efetivo inicial da população.

Ao longo das gerações simuladas, a taxa de homozigose individual e por geração foi calculada pelo software *QMSim*, usando as informações descritas para a realização das análises de simulação.

3.2.2.4 Tendências genéticos

As tendências genéticas para as populações de aves na Fase 3 de cada cenário (cenários 1, 2, 3 e 4) e também os coeficientes de herdabilidade (0,15; 0,30 e 0,45) simulados foram obtidas a partir de uma regressão linear simples (PROC REG – SAS® 9.4) dos respectivos valores genéticos estimados (EBV-BLUP) em função de quinze gerações. Com isso, pode-se ter, o conhecimento sobre os ganhos genéticos que ocorriam hipoteticamente para as características que estavam passando pelo processo seletivo no programa de melhoramento genético de linhagens caipiras real que foi base para este estudo.

3.3 Resultados

3.3.1 Endogamia e taxa de homozigose

Para todos os cenários estudados, a geração zero (G_0) das três fases da população recente (Fase 1, Fase 2 e Fase 3) teve um coeficiente de endogamia igual a zero. Além disso, os valores obtidos para os coeficientes de endogamia aumentaram da primeira para a última geração de cada uma das fases 1, 2 e 3 estudadas, e os maiores coeficientes de endogamia foram obtidos nos cenários com menor Ne (Figura 2).

Na Fase 1, os coeficientes de endogamia foram semelhantes considerando as diferentes herdabilidades (0,15; 0,30 e 0,45), atingindo valores máximos de 0,02, 0,035, 0,08 e 0,17 para os cenários 1, 2, 3 e, respectivamente, na geração 40 (G_40). Na Fase 2, houve maior influência dos coeficientes de herdabilidade nos coeficientes de endogamia obtidos em comparação à Fase 1, variando de acordo com os diferentes tamanhos efetivos populacionais, atingindo os valores extremos de 0,032, 0,058, 0,12 e 0,21 para os cenários 1, 2 3 e 4, respectivamente, na geração 65 (G_65). Na Fase 3, para todos os cenários estudados, os maiores valores do coeficiente de endogamia foram considerando o fenótipo com coeficiente de herdabilidade de 0,15, atingindo valores de 0,05, 0,09, 0,2 e 0,3, na geração 80 (G_80) dos cenários 1, 2, 3 e 4, respectivamente. Os resultados dos coeficientes de endogamia acumulados nas Fases 1, 2 e 3 simuladas e a relação dos tamanhos efetivos da população, coeficientes de herdabilidade e coeficiente de endogamia médios

dos quatro cenários de rebanhos de aves caipiras simulados podem ser observados na Figura 2 e na Tabela 4.

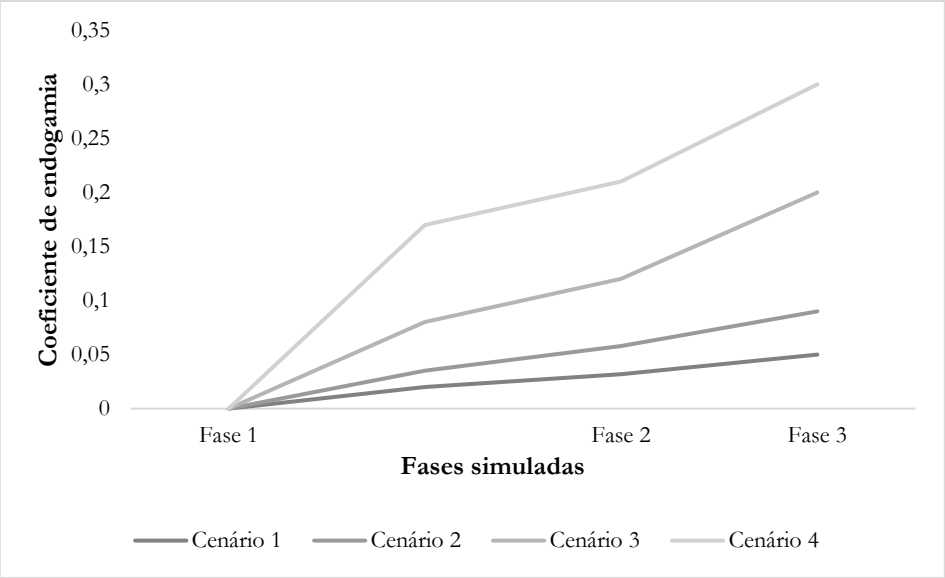


Figura 2 – Coeficientes de endogamia acumulados das Fases 1, 2 e 3 simuladas

Tabela 4. Relação dos tamanhos efetivos da população, coeficientes de herdabilidade e coeficiente de endogamia médios dos quatro cenários de rebanhos de aves caipiras simulados

Cenários	Cenário 1			Cenário 2			Cenário 3			Cenário 4		
Ne	1000, 640 e 640			500, 320 e 320			200, 128 e 128			100, 64 e 64		
	h^2			h^2			h^2			h^2		
Geração	0,15	0,30	0,45	0,15	0,30	0,45	0,15	0,30	0,45	0,15	0,30	0,45
0_10	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,01	0,01	0,02	0,02	0,02
11_20	0,01	0,01	0,01	0,01	0,01	0,01	0,03	0,03	0,03	0,06	0,06	0,06
21_30	0,01	0,01	0,01	0,02	0,02	0,02	0,05	0,05	0,05	0,10	0,10	0,10
31_40	0,02	0,02	0,01	0,03	0,03	0,03	0,07	0,07	0,07	0,14	0,14	0,14
41_50	0,00	0,01	0,01	0,01	0,01	0,01	0,02	0,02	0,02	0,04	0,05	0,05
51_60	0,01	0,02	0,02	0,03	0,03	0,04	0,07	0,08	0,06	0,12	0,14	0,14
61_70	0,01	0,01	0,02	0,03	0,03	0,03	0,07	0,07	0,06	0,12	0,12	0,12
71_80	0,03	0,02	0,02	0,06	0,05	0,04	0,15	0,10	0,09	0,22	0,17	0,13
DP	0,010	0,007	0,007	0,019	0,014	0,012	0,043	0,031	0,026	0,063	0,053	0,046

Ne: tamanho efetivo da população; h^2: herdabilidades: DP: desvio-padrão.

As taxas de homozigose variaram na geração histórica (H_0 a H_1010) nos quatro cenários simulados de 0,64 a 0,67. Na população recente nas Fases 1, 2 e 3 estes valores variaram de 0,71 a 0,77 no cenário 1; de 0,72 a 0,79 no cenário 2; de 0,73 a 0,84 no cenário 3; e de 0,75 a 0,89 no cenário 4, sendo os maiores valores da taxa de homozigose para o coeficiente de herdabilidade de 0,45 (Figura 3 e Tabela 5).

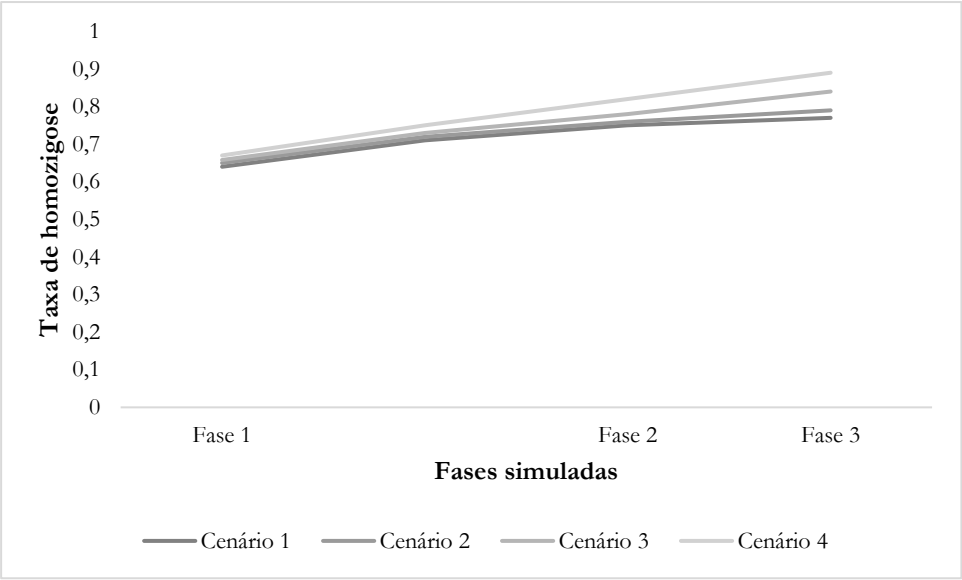


Figura 3 – Taxas de homozigose das Fases 1, 2 e 3 simuladas

Tabela 5. Relação dos tamanhos efetivos da população, coeficientes de herdabilidade e taxas de homozigose dos quatro cenários de rebanhos de aves caipiras simulados

Cenários	Cenário 1			Cenário 2			Cenário 3			Cenário 4		
Ne	1000, 640 e 640			500, 320 e 320			200, 128 e 128			100, 64 e 64		
	h^2			h^2			h^2			h^2		
Geração	0,15	0,30	0,45	0,15	0,30	0,45	0,15	0,30	0,45	0,15	0,30	0,45
0_10	0,71	0,71	0,71	0,72	0,72	0,72	0,73	0,73	0,73	0,75	0,74	0,75
11_20	0,71	0,71	0,71	0,72	0,72	0,72	0,74	0,74	0,74	0,76	0,76	0,76
21_30	0,71	0,71	0,71	0,72	0,72	0,72	0,74	0,74	0,74	0,77	0,77	0,77
31_40	0,71	0,71	0,71	0,72	0,72	0,72	0,75	0,75	0,75	0,78	0,78	0,78
41_50	0,71	0,71	0,71	0,73	0,73	0,73	0,76	0,76	0,76	0,79	0,80	0,80
51_60	0,72	0,72	0,73	0,74	0,75	0,75	0,78	0,79	0,78	0,82	0,83	0,83
61_70	0,73	0,73	0,75	0,75	0,76	0,77	0,80	0,81	0,80	0,85	0,86	0,86

71_80	0,74	0,75	0,77	0,78	0,78	0,79	0,84	0,84	0,84	0,89	0,89	0,89
Média	0,72	0,72	0,72	0,74	0,74	0,74	0,77	0,77	0,77	0,80	0,80	0,80
DP	0,012	0,014	0,022	0,020	0,022	0,027	0,034	0,037	0,034	0,045	0,048	0,049

Ne: tamanho efetivo da população; h^2: herdabilidades: DP: desvio-padrão.

3.3.2 Tendência genética

As tendências genéticas por geração dos quatro cenários simulados para a Fase 3; e para os três coeficientes de herdabilidade (0,15; 0,30 e 0,45) estão apresentadas na Figura 4. Esta fase foi escolhida dentre as oitenta gerações do processo de simulação devido ao projeto de seleção realizado nestas quinze gerações.

Os coeficientes de regressão foram significativamente diferentes de zero ($p<0,05$). Os valores genéticos foram maiores de acordo com o aumento dos coeficientes de herdabilidade e pelo número de gerações selecionadas. No cenário 1, os valores médios do EBV obtidos na geração 65 (G_65) à geração 80 (G_80) foram de 0,275 a 4,812; 0,516 a 8,093; e 1,032 a 12,976, para fenótipos dom herdabilidades de 0,15; 0,30 e 0,45; respectivamente. Esses valores foram semelhantes aos obtidos no cenário 2, em que os valores médios de EBV obtidos foram de 0,210 a 5,066 para herdabilidade 0,15; ,587 a 8,039 para herdabilidade 0,30; e 1,049 a 10,570 para herdabilidade 0,45, respectivamente da geração 65 a 80.

No cenário 3, considerando os fenótipos com coeficientes de herdabilidade de 0,15; 0,30 e 0,45; os EBVs médios obtidos de G_65 a G_80 foram de 0,160 a 4,312; 0,382 a 6,519 e de 0,6720 a 9,060, respectivamente. No cenário 4 os valores médios dos EBVs obtidos foram de 0,110 a 3,526 para herdabilidade de 0,15; 0,260 a 6,099 para herdabilidade 0,30; e 0,599 a 8,360 para herdabilidade 0,45. Assim, os maiores valores médios de EBV foram para o coeficiente de herdabilidade de 0,45 para todos os cenários, os valores de EBVs médios foram crescentes até a quinta geração sob seleção (G_69) e se estabilizaram em todos os cenários; e os valores médios dos EBVs obtidos no cenário 4 foram inferiores aos dos cenários 1, 2 e 3.

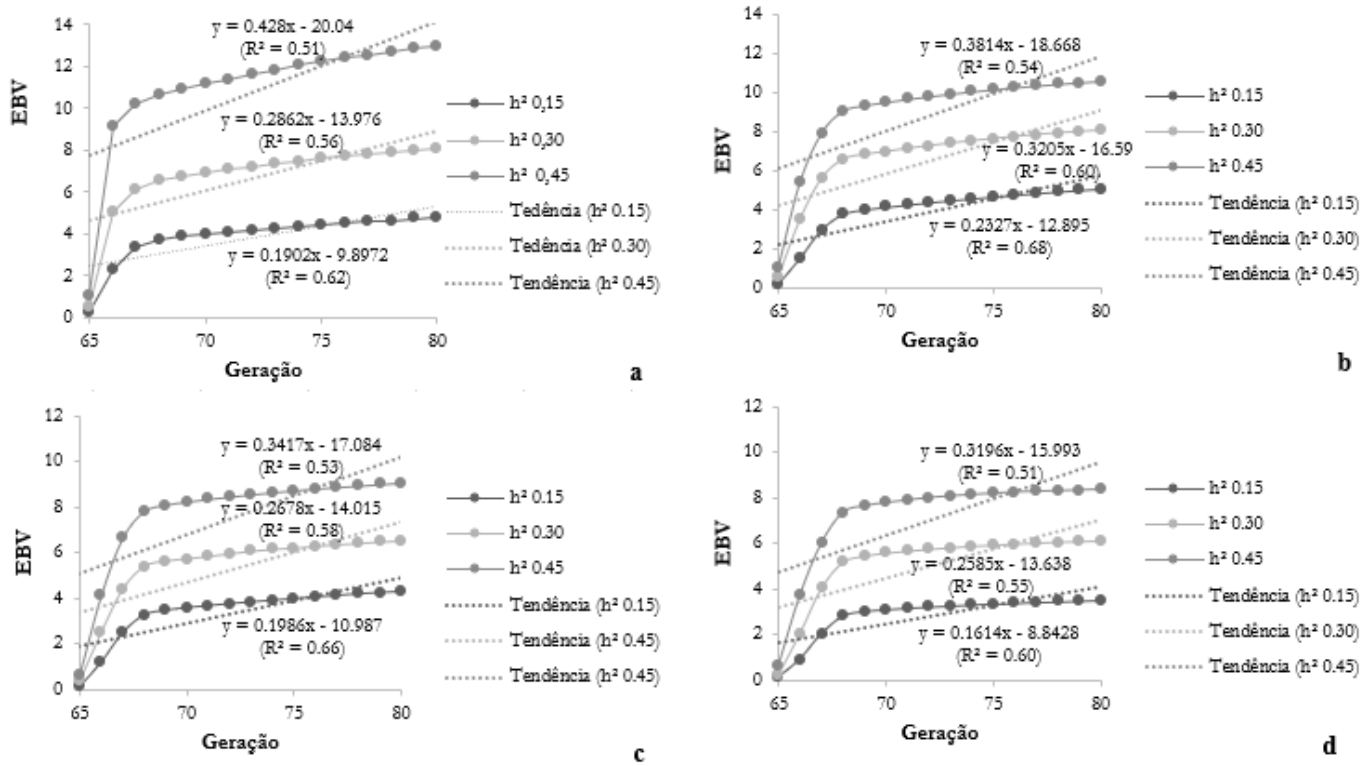


Figura 4 - Tendências genéticas para os coeficientes de herdabilidade: 0,15; 0,30 e 0,45 da Fase 3 dos cenários simulados; a: cenário 1; b: cenário 2; c: cenário 3; d: cenário **4**

Foi calculado o ganho genético (ΔG) para características com os coeficientes de herdabilidade simulados: 0,15; 0,30 e 0,45. A relação dos ganhos genéticos com os coeficientes de herdabilidade e os tamanhos efetivos dos rebanhos simulados na Fase 3 são apresentados na Tabela 6. O Cenário 1 com Ne de 640 apresentou os maiores valores de ganhos genéticos, sendo 0,19 para herdabilidade 0,15, 0,29 para herdabilidade 0,30; e 0,43 para herdabilidade 0,45. Para o cenário 2 de Ne igual a 320 os valores de ganhos genéticos foram de 0,23; 0,32 e 0,38 para os coeficientes de herdabilidade de 0,15; 0,30 e 0,45; respectivamente. Quando o tamanho efetivo da população simulado foi de 128 (cenário 3), os valores de ganhos genéticos obtidos foram de 0,20 para o coeficiente de herdabilidade baixo (0,15); 0,27 para a herdabilidade de 0,30 e 0,34 para o coeficiente de herdabilidade de alta magnitude (0,45). O cenário 4 obteve os menores valores de ganhos genéticos; e possuía um Ne de 64, por isso, apresentou valores de ganhos genéticos de 0,16; 0,26 e 0,32 para os três coeficientes de herdabilidade simulados (0,15; 0,30 e 0,45; respectivamente). Assim, os valores de ganhos genéticos foram maiores para o coeficiente de herdabilidade de 0,45 dos quatro cenários simulados com diferentes tamanhos efetivos da população.

Tabela 6. Relação dos ganhos genéticos, coeficientes de herdabilidade e os tamanhos efetivos populacionais dos rebanhos de aves caipiras simulados na Fase 3

Cenários	Cenário 1	Cenário 2	Cenário 3	Cenário 4
Ne	640	320	128	64
h^2		ΔG		
0,15	0,19	0,23	0,20	0,16
0,30	0,29	0,32	0,27	0,26
0,45	0,43	0,38	0,34	0,32
h^2		R^2		
0,15	0,643	0,630	0,622	0,572
0,30	0,622	0,568	0,546	0,506
0,45	0,651	0,545	0,630	0,501

Ne: tamanho efetivo da população; ΔG: ganho genético; R^2: coeficiente de determinação; h^2: herdabilidade.

3. 4 Discussão

3.4.1 Endogamia e Homozigose

As estimativas de endogamia geralmente são mensuradas pelo coeficiente de endogamia proposto por Wright (1922). Sendo assim, os resultados demonstram que populações de menores tamanhos efetivos da população (cenário 4) possuíam valores de coeficientes de endogamia elevados e que os coeficientes de herdabilidade das características que estiveram sob processo de seleção podem influenciar nestes valores como mostram os resultados para o coeficiente de herdabilidade de 0,15. Estes resultados são diferentes dos apresentados por Viale et al. (2017), que estudaram linhagens nativas italianas e encontraram resultados de coeficientes de endogamia (FIS) médios de 0,114, indicaram um efeito de endogamia de baixo nível. Além disso, os valores positivos encontrados implicaram que os indivíduos de uma população estavam mais relacionados do que o esperado em um modelo de acasalamento aleatório, pois um certo nível de endogamia ocorreu. Estes valores variaram de 0,008 para a linhagem Millefiori di Lonigo (ML), em que foram amostrados 33 indivíduos e 0,223 para linhagem Ermellinata di Rovigo (ER), que teve 103 indivíduos amostrados. Desse modo, para programas de melhoramento que buscam a conservação de linhagens nativas, como o de linhagens caipiras nacionais, é possível que estratégias de seleção adotadas e acasalamentos não aleatórios para a manutenção de padrões morfológicos das linhagens caipiras, podem causar um aumento do valores dos coeficientes de endogamia da população.

Marchesi et al. (2018) avaliaram indivíduos de uma linha paterna de frangos chamada TT, essa linhagem foi mantida sobre seleção de diversas características; e teve um tamanho populacional dos fundadores de 89 machos acasalados com 10 matrizes cada, que passaram por outros acasalamentos para a produção das próximas gerações (25 reprodutores acasalados com cinco matrizes cada um, na atualidade). Para esses indivíduos as estimativas dos coeficientes de endogamia foram realizadas com base em informações de pedigree, assim como neste trabalho, e variaram de 0,0423 a 0,0981. Assim, o Ne da população ancestral, baseado em 200 gerações passadas, foi maior, contando com 548 indivíduos, em comparação as gerações atuais. No entanto, os autores relataram que nas últimas cinco gerações, o Ne variou de 113 (população atual) para 157 galinhas. Estes valores de Ne são maiores que os rebanhos de linhagens caipiras hipoteticamente simulados nesse estudo no cenário 4. Para fins de conservação, FAO (2004), recomenda no mínimo 50 indivíduos para qualquer espécie de rebanho.

As estratégias de seleção utilizadas nas fases (1, 2 e 3) deste estudo influenciaram nos valores obtidos para os coeficientes de endogamia dos quatro cenários (Figura 2), especialmente no cenário 4, que representa hipoteticamente os rebanhos reais de linhagens caipiras. Com isso, um dos efeitos do aumento da endogamia é a redução das frequências de genótipos heterozigotos e a elevação de genótipos em homozigose, que pode ser observado pelo aumento das taxas de homozigose com a diminuição dos tamanhos efetivos da população ao longo dos cenários simulados (Figura 3 e Tabela 4). Valores de endogamia podem descrever mudanças na estrutura genética de uma população a favor da homozigose de um conjunto de genes, prejudicando a heterozigosidade do pool genético de indivíduos, implicando na perda de variabilidade genética. Isso pode influenciar negativamente as características produtivas, por isso, evitar acasalamentos de animais intimamente relacionados pode ser uma abordagem adequada para manter a endogamia sobre controle na população (Vostry et al. 2018).

Dessa forma, em populações domésticas e nativas, como linhagens caipiras da ESALQ, o Ne mínimo para evitar a depressão por endogamia a curto prazo e a própria extinção da espécie de ser de pelo menos 50, enquanto que o Ne de 500 (tamanho efetivo da população na Fase 1 do cenário 3), pode ser suficiente para manter o potencial evolutivo a longo prazo (Menwissen, 2009; Howard et al. 2017). Ou seja, valores de coeficiente de endogamia superiores a 10% (<10%) podem resultar em perdas no desempenho dos animais. Portanto, a diminuição constante do Ne pode provocar um consequente aumento da endogamia (Kiya et al. 2019), o que pode ter ocorrido nesse estudo.

3.4.2 Tendência genética

A estimação de tendências genéticas em uma população permite visualizar se os procedimentos de seleção foram efetivos e assegurar que a pressão de seleção foi direcionada para ás características de importância econômica, além de auxiliar na definição dos objetivos de seleção de programas de melhoramento genético (Weber et al.2009).

O tamanho efetivo da população (Ne), intervalo de gerações e a variabilidade genética influenciam potencialmente na seleção, consequentemente, no progresso genético. Além disso, a acurácia de seleção e a intensidade de seleção influenciam na resposta à seleção e devem ser otimizados separadamente, por serem componentes do ganho genético (ΔG). O aumento na intensidade de seleção, pode provocar o decréscimo do tamanho efetivo da população (Ne), pelo uso de um menor número de animais selecionados; e consequentemente, diminuir a resposta à seleção (Malhado et al., 2008; Muir, 2000).

Nas populações de animais domésticos, selecionadas efetivamente, o tamanho efetivo da população desempenha um papel importante, porque pode afetar a resposta a seleção e a sensibilidade aos efeitos da endogamia. O conhecimento do Ne pode facilitar o conhecimento de esquemas eficientes de seleção artificial. Portanto, o Ne deve ser maximizado para que não influencie nos ganhos genéticos dos indivíduos participantes dos programas de melhoramento genético (Meuwissen; Woolliams, 1994; Qanbari et al., 2010).

Assim, o cenário 4 deste estudo de simulação, foi escolhido para representar hipoteticamente os rebanhos de linhagens caipiras pertencentes a um programa de melhoramento genético em condições reais, e que pudesse ser comparado com os outros cenários que possuíam Ne maiores. Este cenário obteve os menores valores de EBVs médios em comparação aos cenários 1, 2 e 3, devido ao tamanho efetivo da população aplicado (Ne=64), valor próximo aos dos rebanhos de linhagens caipiras que foi base para o estudo. Entende-se que as diferenças nos valores de EBVs médios e dos ganhos genéticos ao longo das quinze gerações não foram devidas somente aos coeficientes de herdabilidades simulados (0,15; 0,30 e 0,45), mas foram influenciadas pelos tamanhos efetivos da população, na Fase 3.

3.5 Conclusão

Pode-se concluir que o tamanho efetivo da população (Ne) dos rebanhos de linhagens caipiras da ESALQ deve ser aumentado; e as estratégias de seleção devem ser revisadas e aplicadas para a minimização dos níveis de endogamia e homozigose, para que os ganhos genéticos destas populações não sejam afetados e possam ser maximizados.

Referências

Allais, S. et al. Mapping of QTL for chicken body weight, carcass composition, and meat quality traits in a slow-growing line. **Poultry Science**, v. 98, n. 5, p. 1960–1967, 2019.

Blackburn, H. D. The National Animal Germplasm Program: Challenges and opportunities for poultry genetic resources. Poultry Science, v. 85, n. 2, p. 210–215, 2006.

Ciampolini, R. et al. Investigation on the genetic variability of the american Pit Bull Terrier dogs belonging to an Italian breeder using microsatellite markers and genealogical data. Cytology and Genetics, v. 47, n. 4, p. 217–221, 2013.

Dalvit, C. et al. Genetic variation and population structure of Italian native sheep breeds undergoing in situ conservation. Journal of Animal Science, v. 87, n. 12, p. 3837–3844, 2009.

Falconer, D.S.; Mackay, T.F.C. Introduction to Quantitative Genetics. 4th Edition, Prentice Hall, Essex, 1996.

FAO. Molecular genetic characterization of animal genetic resources. FAO Animal Production and Health Guidelines. 2011.

Hillel, J. et al. Biodiversity of 52 chicken populations assessed by microsatellite typing of DNA pools. Genetic Selection Evolution. 35:533–557, 2003.

Howard, J. T. et al. Invited review: Inbreeding in the genomics era: Inbreeding, inbreeding depression, and management of genomic variability. Journal of Dairy Science, v. 100, n. 8, p. 6009–6024, 2017.

Joaquim, L. B. et al. Genomic structure of a crossbred Landrace pig population. PLoS ONE, v. 14, n. 2, p. 1–14, 2019.

Kiya, C. K. et al. Population structure of a nucleus herd of Dorper sheep and inbreeding effects on growth, carcass, and reproductive traits. Small Ruminant Research, v. 177, n. March, p. 141–145, 2019.

Leroy, G. et al. Methods to estimate effective population size using pedigree data: Examples in dog, sheep, cattle and horse. Genetics Selection Evolution, v. 45, n. 1, p. 1–10, 2013.

Malhado, C. H. M. et al. Progresso genético e estrutura populacional do rebanho Nelore no Estado da Bahia. Pesquisa Agropecuaria Brasileira, v. 43, n. 9, p. 1163–1169, 2008.

Marchesi, J. A. P. et al. Relationship of runs of homozygosity with adaptive and production traits in a paternal broiler line. Animal, v. 12, n. 6, p. 1126–1134, 2018.

Meuwissen, T. H. E.; Woolliams, J. A. Effective sizes of livestock populations to prevent a decline in fitness. Theoretical and Applied Genetics, v. 89, n. 7–8, p. 1019–1026, 1994.

Meuwissen, T. Genetic management of small populations: A review. Acta Agriculturae Scandinavica A: Animal Sciences, v. 59, n. 2, p. 71–79, 2009.

Misztal, I. et al. Blupf90 and Related Programs (Bgf90). p. 2001–2002, 2002.

Muir, W. M. The interaction of selection intensity, inbreeding depression, and random genetic drift on short- and long-term response to selection: Results using finite locus and finite population size models incorporating directional dominance. Journal of Animal Science, v. 79, n. E-Suppl_1, p. 1, 2000.

Pham, M. H. et al. Monitoring of genetic diversity in Taiwan conserved chickens assessed by pedigree and molecular data. Livestock Science, v. 184, p. 85–91, 2016.

Pritchard, J. K.; Rosenberg, N. A. Use of unlinked genetic markers to detect population stratification in association studies. American Journal of Human Genetics, v. 65, n. 1, p. 220–228, 1999.

Qanbari, S. et al. Linkage disequilibrium reveals different demographic history in egg laying chickens. **BMC Genetics**, v. 11, 2010.

Rovadoscki, G. A. et al. Genetic parameters for growth characteristics of free-range chickens under univariate random regression models. **Poultry Science**, v. 95, n. 9, p. 1989–1998, 2016.

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

SILVA, Evolução do melhoramento genético de aves no Brasil Genetic improvement of Poultry in Brazil. **Ceres**, v. 56, n. 3, p. 437–445, 2009.

Teng, J. et al. Performance of whole genome prediction for growth traits in a crossbred chicken population. Poultry Science, v. 98, n. 5, p. 1968–1975, 2019.

Viale, E. et al. Development and validation of a novel SNP panel for the genetic characterization of Italian chicken breeds by next-generation sequencing discovery and array genotyping. Poultry Science, v. 96, n. 11, p. 3858–3866, 2017.

Vostry, L. et al. Genetic diversity and effect of inbreeding on litter size of the Romanov sheep. Small Ruminant Research, v. 168, n. August, p. 25–31, 2018.

Weber, T. et al. Genetic parameters and genetic and phenotypic trends for pre-weaning productive and conformation traits for an Aberdeen Angus breed population. Revista Brasileira de Zootecnia, v. 38, n. 5, p. 832–842, 2009.

Wright, S. Evolution in Mendelian Populations. **Genetics**, v. 16, p. 97-159, 1931.

Zanetti, E. et al. Variation of genetic diversity over time in local Italian chicken breeds undergoing in situ conservation. Poultry Science, v. 90, n. 10, p. 2195–2201, 2011

Zanetti, E. et al. Using 2-DE for the differentiation of local chicken breeds. Proteomics, v. 11, n. 13, p. 2613–2619, 2011.

4 SIMULAÇÃO DE DIFERENTES TAMANHOS EFETIVOS POPULACIONAIS DE REBANHOS DE LINHAGENS CAIPIRAS NACIONAIS E OS IMPACTOS NO DESEQUILÍBRIO DE LIGAÇÃO E NA ACURÁCIA DE PREDIÇÃO GENÔMICA.

RESUMO

Para aliar o desenvolvimento sustentável à preservação de recursos genéticos, a produção de linhagens de aves nativas, seria uma alternativa, contudo, estudos genéticos para estas espécies ainda são escassos. Para compreender se o tamanho efetivo populacional (Ne) de rebanhos de aves caipiras da ESALQ está adequado e seus impactos sobre o desequilíbrio de ligação (LD) e a acurácia de predição genômica, foi realizado este estudo de simulação. Foi simulada uma geração histórica que por mil gerações teve um tamanho constante de 2000 indivíduos, ocorrendo um gargalo genético nas gerações subsequentes (1010 a 1020) para 400 animais e expandiu-se para 500, 200 e 100 animais de onde surgiram as fases (população recente) dos três cenários simulados e para três coeficientes de herdabilidade (0,15; 0,30 e 0,45). Na Fase 1 os indivíduos não passaram por nenhum processo de seleção por quarenta gerações (G_0 a G_40), posteriormente nas Fases 2 e 3, os indivíduos foram selecionados por características fenotípicas por 25 gerações (G_41 a G_65) e pelos valores genéticos preditos (BLUP-EBV) nas últimas 15 gerações (G_66 a G_80), respectivamente. Os cenários 1, 2, 3 possuíram tamanhos efetivos da população de 500, 320 e 320; 200, 128 e 128; 100, 64 e 64 nas Fases 1, 2 e 3, respectivamente. Foram calculados o desequilíbrio de ligação (LD), a acurácia e o viés de predição "empíricos" para os três cenários e para três coeficientes de herdabilidade simulados. Os valores médios de r^2, para distância dos marcadores de 0-0,05 Kb, foram inferiores ou próximos a 0,30 nas Fases 1 e 2 dos cenários 1 e 2, devido ao Ne (500 e 320 no cenário 1; e 200 e 128 no cenário 2) e as estratégias de seleção aplicados (Indivíduos sob nenhum processo de seleção na Fase 1 e seleção por características fenotípicas na Fase 2). Por outro lado, para o cenário 3, de menor tamanho efetivo populacional os valores médios de r^2 foram superiores a 0,30 em todas as fases e coeficientes de herdabilidade simulados. Os valores de correlações entre os TBVs e EBVs variaram de 0,69 a 0,81 para os cenários e herdabilidades simulados. As correlações entre TBVs e GEBVs variaram de 0,45 a 0,90, sendo apresentados os menores valores no cenário 3. Os coeficientes de regressão estimados para os fenótipos simulados e EBVs foram altos e maiores que 1, sendo os EBVs subestimados e as regressões dos fenótipos simulados nos GEBVs apresentando valores menores que 1, indicando que os GEBVs foram superestimados. Pode-se concluir que o LD das populações de linhagens caipiras da ESALQ pode ser considerado "útil" e eficiente para estudos de associação e predições genômicas que estão relacionados ao desequilíbrio de ligação e para que os valores de acurácia de predição sejam maiores será necessário o aumento do Ne destes rebanhos para que as avaliações genômicas não sejam influenciadas negativamente.

Palavras-Chave: avicultura, confiabilidade preditiva, desequilíbrio de ligação, simulação, seleção genômica.

Abstract

To combine sustainable development with the preservation of genetic resources, the production of native bird lines would be an alternative, but genetic studies for these species are still scarce. In this context, in order to understand whether the effective population size (Ne) of flocks of national free-range birds is adequate and the impacts on the linkage disequilibrium (LD) and the accuracy of genomic prediction, this simulation study was carried out. A historical generation was simulated, which for a thousand generations had a constant size of 2000 individuals, with a genetic bottleneck occurring in subsequent generations (1010 to 1020) for 400 animals and expanded to 500, 200 and 100 animals from which the phases emerged (recent population) of the three simulated scenarios and for three coefficients of heritability (0.15; 0.30 and 0.45). In Phase 1, individuals did not undergo any selection process for forty generations (G_0 to G_40), later in Phases 2 and 3, individuals were selected for phenotypic characteristics for 25 generations (G_41 to G_65) and the predicted genetic values (BLUP -EBV) in the last 15 generations (G_66 to G_80), respectively. Scenarios 1, 2, 3 had effective population sizes of 500, 320 and 320; 200, 128 and 128; 100, 64 and 64 in Phases 1, 2 and 3, respectively. Linkage disequilibrium (LD), accuracy and "empirical" prediction bias were calculated for the three scenarios and three simulated heritability coefficients. The average values of r^2, at distance from the markers of 0-0.05 Kb, were lower or close to 0.30 in Phases 1 and 2 of scenarios 1 and 2, due to the Ne (500 and 320 in scenario 1; and 200 and 128 in scenario 2) and the selection strategies applied (Individuals under no selection process in Phase 1 and selection by phenotypic characteristics in Phase 2), while for scenario 3, the lowest effective population size, the mean r^2 values were greater than 0.30 in all phases and simulated heritability coefficients. The correlation values between TBVs and EBVs ranged from 0.69 to 0.81 for the scenarios and simulated heritabilities. The correlations between TBVs and GEBVs varied from 0.45 to 0.90, with the lowest values being presented in scenario 3. The estimated regression coefficients for the simulated phenotypes and EBVs were high and greater than 1, with the EBVs being underestimated and the regressions of the phenotypes simulated in the GEBVs showed values less than 1, indicating that the GEBVs were overestimated. It can be concluded that the LD of populations of national free-range lineages of ESALQ can be considered "useful" and efficient for association studies and genomic predictions that are related to the linkage disequilibrium and for the accuracy values prediction values are higher, it will be necessary to increase the Ne of these herds so that the genomic evaluations are not negatively influenced.

Keywords: poultry farming, simulation, predictive reliability, linkage disequilibrium, genomic selection

4.1 Introdução

Nas últimas duas décadas a conservação de recursos genéticos tornou-se um tópico de interesse mundial (FAO, 2011). Com isso, o rápido progresso genético nas características economicamente viáveis na criação de animais pode ser observado como um resultado de

desenvolvimentos em biotecnologias reprodutivas, aplicação de ferramentas genéticas avançadas em programas de melhoramento; e melhor logística (Viale et al., 2017).

No entanto, o uso de raças altamente selecionadas e produtivas tem sido associada à perda de diversidade genética na maioria das espécies de animais domésticos, especialmente na avicultura (Dalvit et al., 2009; Ciampolini et al., 2013; Hillel et al., 2003; Blackburn 2006; Zanetti et al., 2011a, b). Nesse contexto, nos últimos anos, a valorização dos recursos genéticos para agricultura local, combinada com a preferência de consumidores por ovos e carnes destes animais, estimulou o estudo da diversidade genética de raças nativas, que podem ser avaliadas usando informações fenotípicas e genotípicas para priorizar a conservação (Fulton e Delany, 2003).

Na década de 1980, foi criado um programa de melhoramento genético com linhagens caipiras nacionais na escola Superior de Agricultura Luiz de Queiroz, intitulado "Frango Feliz", desenvolvidas utilizando-se seleção massal (Coelho et al., 2008), havendo, controle completo de pedigree de no mínimo duas gerações de seleção destas linhagens, contribuindo para a preservação de recursos genéticos.

O desequilíbrio de ligação (LD) pode ser utilizado de forma eficaz para determinar os reais genes responsáveis pela variação de características economicamente importantes através de associações em rebanhos de animais. Informações sobre a extensão do LD em regiões genômicas que abrigam (QTLs – Quantitative trait loci) é necessário para aplicações efetivas de predições genômicas, estudos de associação e na implementação de programas de melhoramento genéticos. Além disso, contribuindo para o entendimento das mudanças nos tamanhos efetivos históricos das populações (Qanbari et al. 2010).

O resultado satisfatório da seleção genômica depende da acurácia da predição do valor genético dos animais. Esta acurácia por sua vez, é influenciada por diversos fatores, como o desequilíbrio de ligação entre o QTL e o marcador, o tamanho efetivo da população, a herdabilidade da característica, a densidade de marcadores e a metodologia utilizada para a predição genômica (De Los Campos et al. 2013).

O presente estudo é baseado no programa de melhoramento genético composto por oito linhagens caipiras nacionais: Caipirão da ESALQ, Caipirinha da ESALQ, Azul, Baianinha da ESALQ, Pernalonga da ESALQ, 7P, Carijó barbada e SooHo. Em estudos anteriores foram estimados os parâmetros genéticos para características de ganho de peso e as curvas de crescimento destas linhagens caipiras nacionais (Rovadoscki et al. 2016). Por isso, para compreender se o tamanho efetivo populacional desses rebanhos de aves caipiras da ESALQ está adequado, além do seus impactos sobre o desequilíbrio de ligação (LD) e a acurácia de predição genômica, foi realizado este estudo de simulação.

4.2 Material e métodos

As oito linhagens caipiras têm aptidões para postura, corte e dupla aptidão. A linhagem Caipirão da ESALQ é selecionada para peso, com fenótipo variado para cor de plumagem; o 7P (pinto preto pesado de pasto de pescoço pelado de Piracicaba), selecionado para peso, apresenta crescimento rápido, plumagem preta e pescoço pelado (gene Na); Caipirinha da ESALQ, de dupla aptidão (corte e postura), com crescimento lento, plumagem de cor variada e topete (gene Cr); a linhagem Azul, ave selecionada para postura, possui ovos de diversas tonalidades de azul em sua população e têm fenótipos variados para a cor da plumagem; o Carijó Barbada, ave de dupla aptidão (corte e postura), crescimento lento, plumagem barrada e apresenta barba e costeletas (gene Mb); a *Soo Ho*, linhagem exótica de dupla aptidão que possui a sua carne escurecida (carne preta) pelo processo de fibromelanose ou hiperpigmentação; a Baianinha da ESALQ e Pernalonga da ESALQ, aves selecionadas para características ornamentais, todos pertencentes ao projeto "Frango Feliz da ESALQ e que hoje participam do projeto coordenado pelo Departamento de Zootecnia da ESALQ. Com isso, para que de forma mais rápida, precisa e economicamente viável, foi realizado este estudo de simulação para o melhor entendimento da influência do tamanho efetivo destas populações de linhagens caipiras no desequilíbrio de ligação (LD) e na acurácia de predição genômica em condições reais.

4.2.1 Processo de simulação

As populações das linhagens caipiras foram simuladas hipoteticamente em duas etapas através do software de simulação *QMSim*, que pode simular genomas gigantescos e complexas estruturas de pedigree, mimetizando as populações de animais domésticos, ou seja, é basicamente um simulador que se baseia em famílias, que pode ou não levar em conta características evolutivas predefinidas, como o desequilíbrio de ligação (LD), mutações, gargalos genéticos e expansões (Sargolzaei; Schenkel, 2009).

Na etapa 1, a geração histórica foi constituída de 2000 indivíduos (1.000 machos e 1.000 fêmeas) e percorreu 1.000 gerações em tamanho constante. Na geração 1.010 ocorreu um gargalo genético, com 400 animais (200 machos e 200 fêmeas) para a formação do desequilíbrio de ligação; e constante até a geração 1.020 que se expandiu em diferentes tamanhos efetivos da população (Ne) até a geração 1.100 de onde advém a população recente que é dividida em três fases (Fases 1,

2 e 3) em três cenários simulados (etapa 2). Os cenários 1, 2, 3 possuíram tamanhos efetivos da população de 500, 320 e 320; 200, 128 e 128; 100, 64 e 64 nas fases um, dois e três (Fase 1, Fase 2 e Fase 3), respectivamente. Sendo assim, o cenário 3 é o mais próximo do número de animais efetivamente em reprodução nos rebanhos das linhagens caipiras pertencentes ao programa de melhoramento realizado em condições reais (Figura 1).

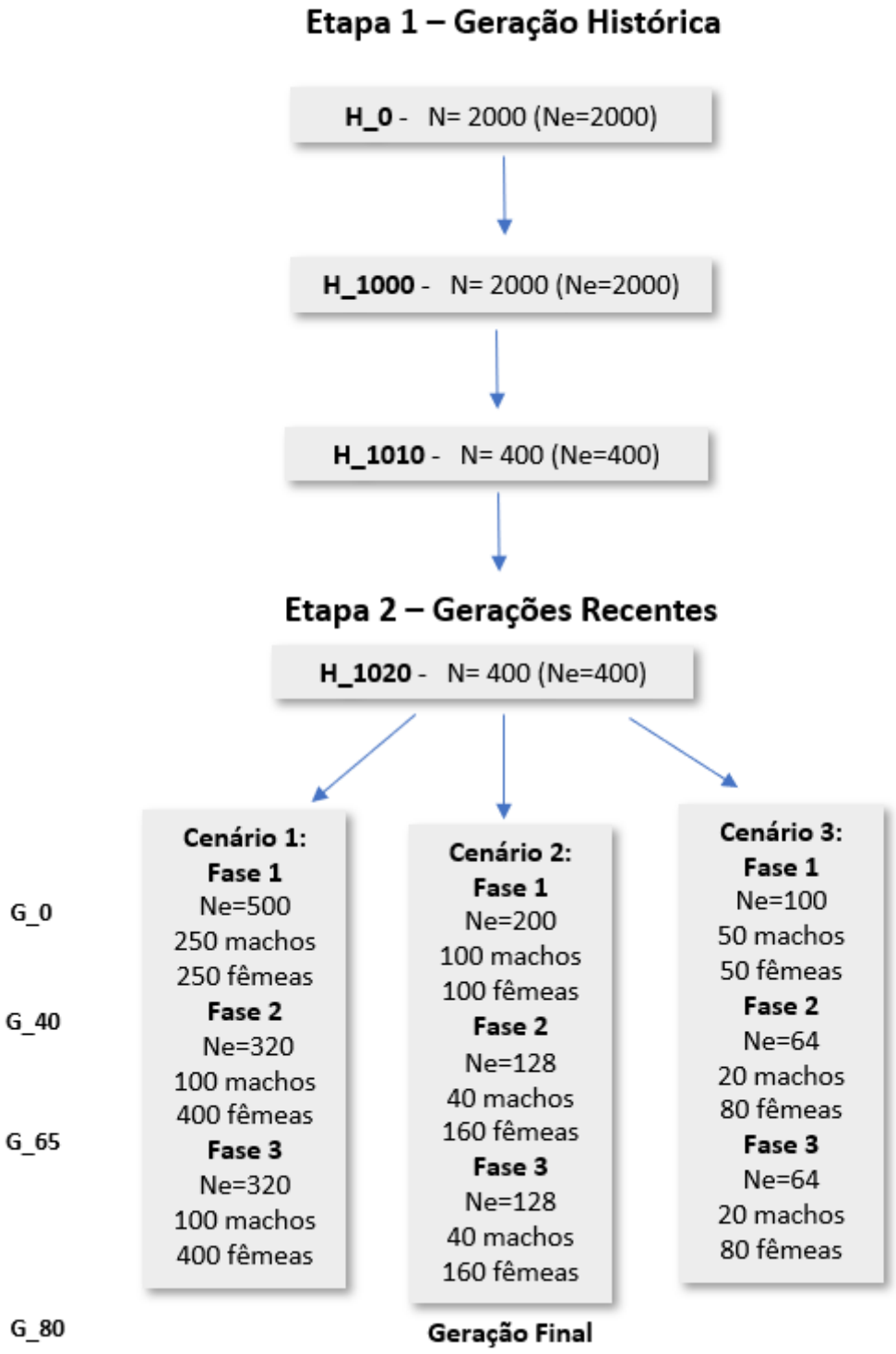
O processo de simulação percorreu 80 gerações para que se tivesse um cenário de um programa de melhoramento genético de aves que tenha sido desenvolvido ao longo desse período. Assim, os indivíduos da primeira geração (G_0) à quadragésima geração (G_40) não foram selecionados (Fase 1). Nas 25 gerações subsequentes (entre G_41 e G_65 – Fase 2) os animais foram selecionados por características fenotípicas e a Fase 3 os indivíduos foram selecionados por valores genéticos estimados (EBV - BLUP), descrevendo hipoteticamente o início do uso do modelo animal nas avaliações genéticas dos animais domésticos. Os animais foram gerados a partir dos acasalamentos realizados na última geração da fase anterior (Figura 5).

Na Fase 1 os acasalamentos foram construídos na proporção de um macho para uma fêmea (1:1) e nas Fases 2 e 3 os acasalamentos foram construídos com um macho para quatro fêmeas (1:4) a cada geração, estes por sua vez, são os acasalamentos utilizados no programa de melhoramento genético de linhagens caipiras base para este trabalho. Para melhor compreensão das características de importância econômica que são utilizadas em programas de melhoramento de aves, as populações também foram simuladas considerando três diferentes coeficientes de herdabilidade (0,15, 0,30 e 0,45), os quais são próximos aos coeficientes de herdabilidade estimados por Gaya et al. (2006), de 0,16 para a taxa de conversão alimentar, 0,33 para teor de gordura abdominal e 0,40 para a característica de peso aos 38 dias.

Foram realizadas 10 repetições do processo de simulação. A variância fenotípica utilizada foi de 1,0 e a geração histórica na etapa 1 foi constituída de 1100 gerações. Para que fossem geradas as informações da população recente (Etapa 2), o número de machos da última geração histórica foram de 250, 100 e 50, respectivamente para os três cenários simulados. Diferente do que acontece com bovinos e equinos, que os machos deixam maior número de descendentes para as próximas gerações que as fêmeas, na avicultura, machos e fêmeas podem deixar descendentes de forma proporcional, por isso, para cada casal acasalado no processo de simulação o tamanho de suas ninhadas foi de 8 filhotes, tendo uma proporção de machos nessas progênies de 0,5.

As relações de substituição de machos e fêmeas no plantel de aves são maiores que as utilizadas em bovinos, por exemplo, por isso, para este estudo, foram utilizadas relações de substituição de 1, que refere a troca de 100% do plantel a cada geração. Os projetos de seleção como descritos, foi aleatório na Fase 1, por Fenótipo na fase 2 e EBV - BLUP na Fase 3 simuladas.

Para que os dados fossem ainda mais próximos das condições reais, o genoma dos indivíduos foi simulado. O genoma consistiu em 28 pares de cromossomos autossomos de comprimentos semelhantes ao genoma real da galinha baseado na montagem de *Gallus gallus* (galGal6 - NCBI, 2020), totalizando 945 cM (centimorgans), para se criar um cenário mais verídico com relação ao número de loci de marcadores e QTLs fisicamente desvinculados. Os 600.012 marcadores de SNP foram distribuídos uniformemente e o número inicial de marcadores foi escolhido de tal forma que geraria uma densidade de loci bi alélicos segregantes com menor frequência alélica (MAF)> 0,02: 600K. Os marcadores foram neutros em seu efeito sobre os QTLs. Um número de QTL foi simulado para gerar 336 loci segregantes com dois alelos e MAF> 0,02; cujas posições foram distribuídas aleatoriamente. Os efeitos alélicos aditivos foram amostrados aleatoriamente a partir da distribuição gama com parâmetro de forma igual a 0,4. A taxa de marcadores de genótipos em falta foi de 0,01 e a taxa de erro de genotipagem de marcadores foi de 0,005. Uma taxa de mutação recorrente de 10^{-5} para ambos os marcadores e QTLs foi considerada para estabelecer o equilíbrio mutação-deriva em gerações históricas. A mesma taxa de mutação também foi aplicada em todas as gerações subsequentes após as históricas. Os parâmetros e as etapas do processo de simulação estão descritos na Figura 5 e nos anexos deste estudo.



H_0: Geração histórica zero; H_1000: Geração histórica 1000; H_1010: Geração histórica 1010; H_1020: Geração histórica 1020; Fase 1, Fase 2 e Fase 3: fases; Ne: Tamanho efetivo da população; G_0: Geração zero recente; G_40: Geração 40 recente; G_65: Geração 65 recente e G_80: Geração 80 recente.

Figura 5- Representação esquemática das etapas de simulação.

4.2.2 Desequilíbrio de ligação (LD)

O desequilíbrio de ligação (LD) foi realizado pelo cálculo do coeficiente de correlação dos alelos ao quadrado em dois locus (r^2) descrito pela seguinte equação:

$$r^2 = \frac{D^2}{f(A)f(a)f(B)f(b)}$$

Sendo que $D = f(AB) - f(A)f(B)$ e $f(AB)$ são frequências haplotípicas; e $f(a), f(A)$ $f(B), f(b)$ são frequências alélicas dos marcadores.

4.2.3 População e fenótipos e genótipos

Os fenótipos, genótipos e registros de pedigree foram obtidos através do processo de simulação (Fase 3) dos três cenários (1, 2 e 3) e dos coeficientes de herdabilidade avaliados (0,15; 0,30 e 0,45).

4.2.4 Predição

- **Valor genético verdadeiro**

Os valores genéticos verdadeiros (TBVs) foram calculados pelo software de simulação QMSim pelo modelo aleatório (Sargolzaei; Schenkel, 2009)

$$\boldsymbol{y} = \boldsymbol{1\mu} + \boldsymbol{Za} + \boldsymbol{e}$$

sendo,

y é o vetor de fenótipos observados;

μ é a média geral e 1 é o vetor de "uns";

Z é a matriz de incidência para os efeitos aleatórios do animal;

a é o vetor aleatório de feitos genéticos aditivos, sendo $a \sim N\ (0, \boldsymbol{A}\sigma_a^2)$;

e é o vetor de efeitos aleatórios dos resíduos.

$\boldsymbol{A}$ é a matriz de relacionamento dos animais, via informação de pedigree; e σ_a^2 é a variância genética aditiva.

- **Valores genéticos estimados**

Os valores genéticos estimados foram estimados para os fenótipos e os coeficientes de herdabilidade de 0,15; 0,30 e 0,45 simulados usando o programa da família BLUPF90 (Misztal,

2002) usando duas abordagens: 1) BLUP baseado no pedigree (PBLUP) e 2) "Single-step genomic BLUP" (ssGBLUP) usando genótipos simulados. O modelo misto univariado usado foi:

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{Za} + \boldsymbol{e}$$

sendo,

y os fenótipos simulados;

X é a matriz de incidência para os efeitos fixos;

b é o vetor de efeitos fixos (sexo);

Z é a matriz de incidência para os efeitos aleatórios de animal;

a é o vetor aleatório efeitos genéticos aditivos, sendo $a \sim N\ (0, \boldsymbol{A}\sigma_a^2)$ para PBLUP e $a \sim N\ (0, H\sigma^2)$ para ssGBLUP;

e é vetor de efeitos aleatórios dos resíduos.

A matriz **H** combina informações da matriz de relacionamento dos animais, via pedigree (A) e matriz genômica (**G**). A inversa de H foi calculada seguindo a abordagem de Aguilar et al. (2010) como:

$$\boldsymbol{H}^{-1} = \boldsymbol{A}^{-1} \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{G}^{-1} - \boldsymbol{A}_{22}^{-1} \end{bmatrix}$$

onde $\boldsymbol{A}^{-1}$ é o inverso de uma matriz de relacionamento com informações de pedigree; $\boldsymbol{G}^{-1}$ é o inverso de uma matriz genomica combinada; e $\boldsymbol{A}_{22}^{-1}$ é o inverso de uma matriz de relacionamento baseada em pedigree apenas de animais genotipados. A matriz **G** combinada foi obtida da seguinte forma:

$$\boldsymbol{G} = 0.95\boldsymbol{G}_{\boldsymbol{w}} + 0.05\boldsymbol{A}_{22}$$

como $\boldsymbol{A}_{22}$ é a matriz de relacionamento baseada no pedigree apenas de animais genotipados; $\boldsymbol{G}_{\boldsymbol{w}}$ é a matriz genômica obtida a seguir (VanRaden, 2008; Heidaritabar et al., 2016):

$$\boldsymbol{G}_{\boldsymbol{w}} = \frac{\boldsymbol{MM}'}{2\sum p_i(1 - p_i)}$$

onde $\boldsymbol{M}$ é a matriz SNP, codificada com 0, 1 ou 2; p_i é a frequência alélica do i-ésimo SNP.

- **Acurácia e viés de predição "empíricos"**

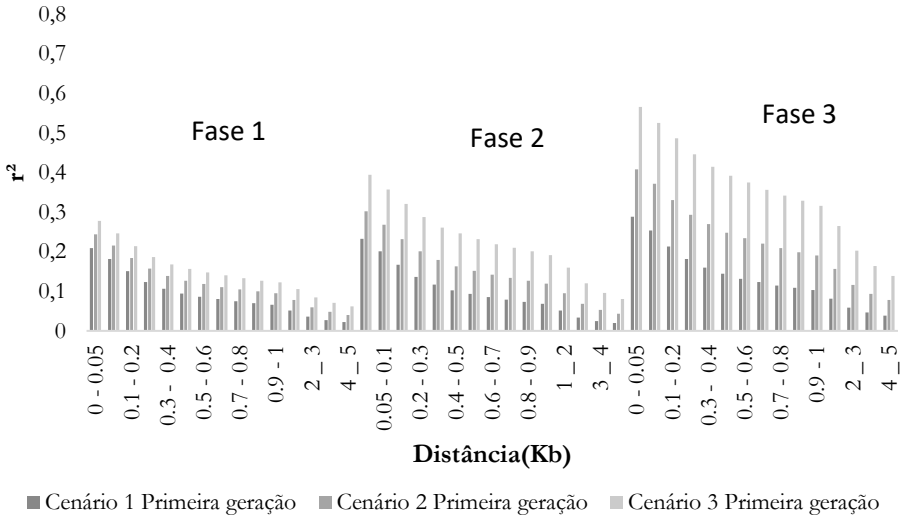
Considerou-se os animais genotipados da décima quinta geração da Fase 3 de simulação (G_80) como grupo de validação e os indivíduos das quatro greções anteriores, como grupo de treinamento (gerações 76 a 79). As acurácias de predição "empíricas" foram obtidas pela correlação (r) entre os TBVs e os EBVs e GEBVs ($cor(TBV, EBV), cor(TBV, GEBV)$). O viés de predição " empirico" dos EBVs e GEBVs foi adquirido pelo coeficiente de regressão dos fenótipos simulados, que foram centralizados pelo acréscimo de uma coluna de " uns" e da média nos EBVs e GEBVs ($b(fen, EBV, b(fen, GEBV)$), permitindo avaliar o grau de similaridade entre as previsões (Momen et al., 2018; Ramírez-Flores et al., 2017). Foram utilizados as médias e os erros-padrões das dez repetições do processo de simulação.

4.3 Resultados

4.3.1 Desequilíbrio de ligação

O desequilíbrio de ligação (LD) entre os marcadores SNP nas primeiras e últimas gerações das fases 1, 2 e 3; e para os coeficientes de herdabilidade de 0,15; 030 e 0,45 é apresentado na Figuras 6, 7 e 8. Percebe-se aumento do LD ao longo das gerações em todos os cenários, possivelmente devido a diminuição dos tamanhos efetivos da população. Além disso, o LD foi maior quando a seleção dos animais foi realizada pelos valores genéticos superiores.

Para o coeficiente de herdabilidade de 0,15; na Fase 1, o valor médio de r^2 para a geração zero (G_0) foi de 0,20 a uma distância de 0-0,05 kb para o cenário 1. Nos cenários 2 e 3 estes valores médios de r^2, às mesmas distâncias foi de 0,24 e 0,28, respectivamente, para a primeira geração da Fase 1. Na Fase 2, os valores médio de r^2 para a distância de 0-0,05 Kb foram de 0,21; 0,24 e 0,28 para os cenário 1, 2 e 3, respectivamente. Na Fase 3, os valore médios de r^2 foram ainda maiores chegando a 0,56 no cenário 3 (Figura 6). Após aplicação das estratégias de seleção os valores médios de r^2 atingiram valores de 0,23; 0,28 e 0,38 nas três fases simuladas (Fases 1, 2 e 3) do cenário 1, respectivamente. No cenário 2, os valores de r^2 médio para as Fase 1 foi de 0,30; nas Fases 2 e 3 foi de 0,41 e 0,58. O cenário 3 apresentou valores médios de r^2 de 0,39 para a fase 1; de 0,57 para a Fase 2 ; e de 0,76 para Fase 3, com coeficiente de herdabilidade de 0,15 (Figura 6).



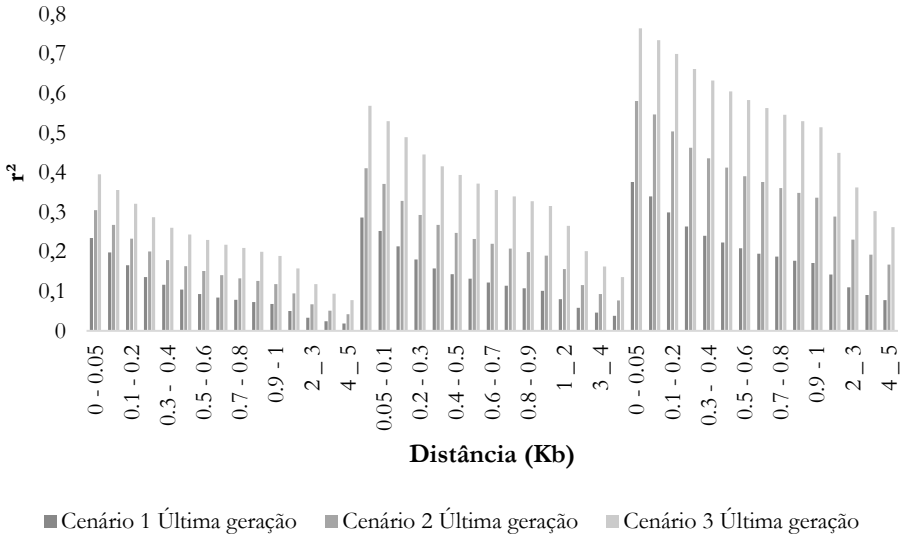
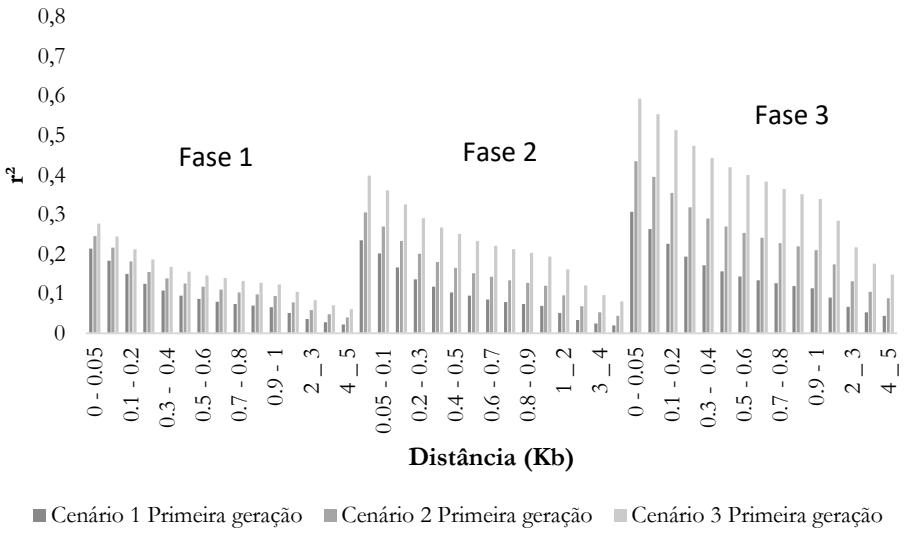


Figura 6 – Desequilíbrio de Ligação (LD) medido por r^2 plotado em função da distância entre os marcadores (Kb) em 80 gerações dos cenários 1, 2 e 3 para as três fases simuladas, para coeficiente de herdabilidade 0,15

Para o coeficiente de herdabilidade de 0,30; na Fase 1, o valor médio de r² para a geração zero (G_0) foi de 0,21 a uma distância de 0-0,05 kb para o cenário 1. Nos cenários 2 e 3 estes valores médios de r², às mesmas distâncias, foi de 0,25 e 0,28, respectivamente, para a primeira geração da Fase 1. Na Fase 2, os valores médio de r² para a distância de 0-0,05 Kb foram de 0,23; 0,30 e 0,40 para os cenário 1, 2 e 3, respectivamente. Na Fase 3, os valore médios de r² foram ainda maiores que os valores apresentados para o coeficiente de herdabilidade de 0,15, chegando a 0,59 no cenário 3 (Figura 7). Na última geração das Fases 1, 2 e 3 simuladas os valores médios de r² atingiram valores de 0,23; 0,30 e 0,37 no cenário 1, respectivamente. No cenário 2, os valores de r² médio para as Fase 1 foi de 0,30; nas Fases 2 e 3 foram de 0,44 e 0,56. O cenário 3 apresentou

valores médios de r² de 0,40 para a Fase 1; 0,59 para a Fase 2; e 0,75 para Fase 3 do coeficiente de herdabilidade 0,30 (Figura 7).



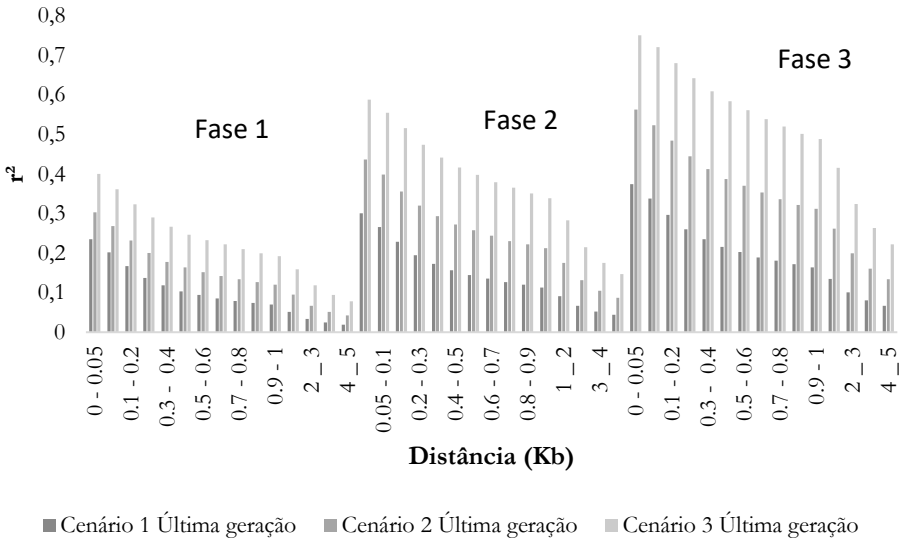


Figura 7 – Desequílibrio de Ligação (LD) medido por r² plotado em função da distância entre os marcadores (Kb) em 80 gerações dos cenários 1, 2 e 3 para as três fases simuladas, para coeficiente de herdabilidade 0,30.

Na Fase 1, para o coeficiente de herdabilidade 0,45, o valor médio de r² para a geração zero (G_0) foi de 0,21, a uma distância de 0-0,05 kb para o cenário 1. Nos cenários 2 e 3 estes valores médios de r² as mesmas distâncias foi de 0,25 e 0,28, respectivamente, para a primeira geração (G_0) da Fase 1. Na Fase 2, os valores médio de r² para a distância de 0-0,05 Kb foram de 0,23; 0,30 e 0,39 para os cenário 1, 2 e 3, respectivamente. Na Fase 3, os valore médios de r² foram

ainda maiores que os valores apresentados para os coeficientes de herdabilidade de 0,15 e 0,30; chegando a 0,61 no cenário 3 (Figura 8). Na última geração das Fases 1, 2 e 3 simuladas (G_40, G_65 e G_80) os valores médios de r² atingiram valores de 0,23; 0,28 e 0,38 no cenário 1, respectivamente. No cenário 2, os valores de r² médio para as Fase 1 foi de 0,30; nas Fases 2 e 3 foram de 0,41 e 0,54. O cenário 3 apresentou valores médios de r² de 0,40 para a Fase 1; de 0,60 para a Fase 2; e de 0,74 para Fase 3 para coeficiente de herdabilidade 0,45 (Figura 8).

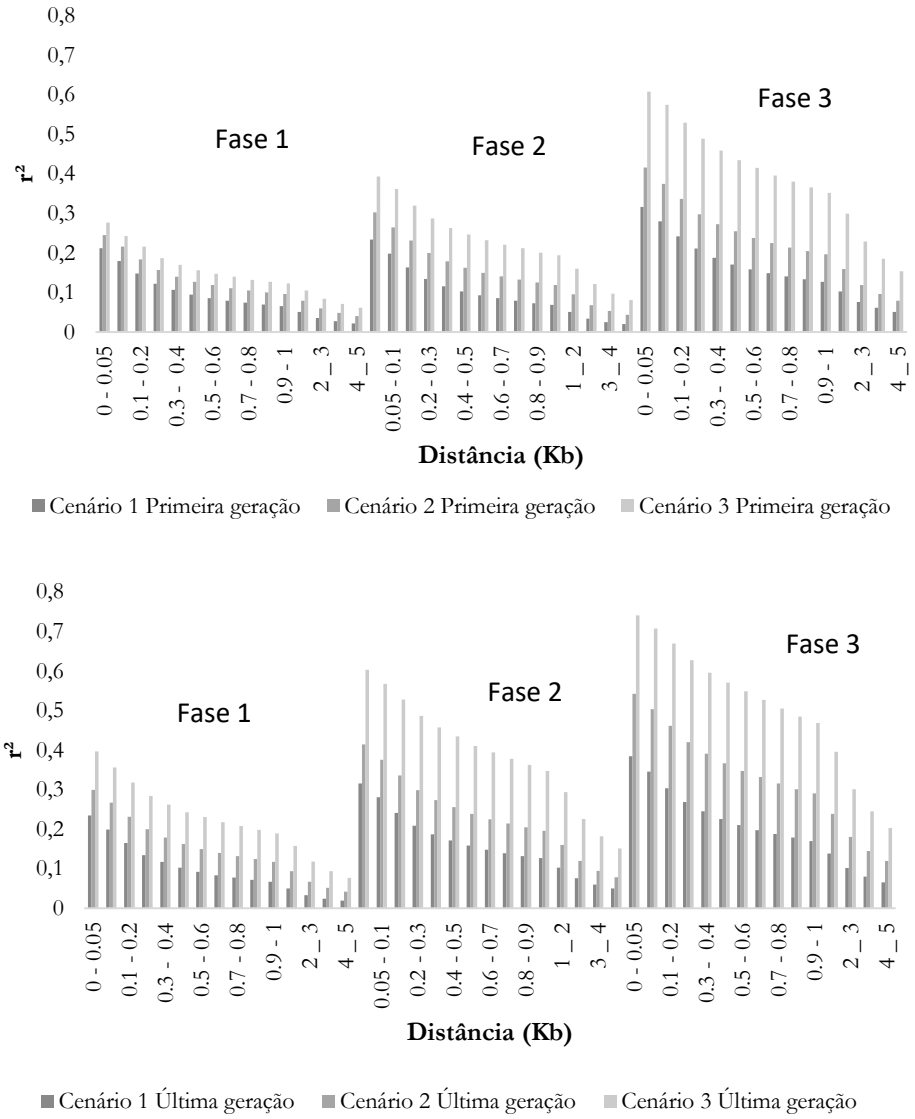


Figura 8 – Desequilíbrio de Ligação (LD) medido por r² plotado em função da distância entre os marcadores (Kb) em 80 gerações dos cenários 1, 2 e 3 para as três fases simuladas, para coeficiente de herdabilidade 0,45.

4.3.2 Acurácia e viés de predição "empíricos"

Os resultados das acurácias de predição e do viés de predição "empíricos" dos EBVs e GEBVs estão apresentados nas Tabelas 7 e 8. As correlações entre TBVs e EBVs apresentaram valores de 0,79 para a h² de 0,15; 0,80 e 0,81 para os coeficientes de herdabilidade de 0,30 e 0,45, respectivamente, no cenário de maior tamanho efetivo da população (Ne=320 – Cenário 1). Para o cenário 2, estes valores for a de 0,70; 0,71 e 0,72, respectivamente, para os três coeficientes de herdabilidade (0,15; 0,30 e 0,45). No cenário 3, com menor tamanho efetivo populacional (Ne=64), os valores de correlações variaram de 0,69 a 0,72 (Tabela 7). Com o acréscimo das informações genômicas, os valores das correlações entre os TBVs e GEBVs foram de 0,83; 0,87 e 0,90 para o cenário 1 e os três coeficientes de herdabilidades (0,15; 0,30 e 0,45), respectivamente. No cenário 2 (Ne=128), esses valores foram menores que os apresentados no cenário 1, mas crescentes em relação aos coeficientes de herdabilidades, variando de 0,77 a 0,81. Para o cenário 3, os valores foram menores que os cenários 1 e 2; e menores que as coerrelações entre TBVs e EBVs (Tabela 7).

Tabela 7. Acurácia de predição "empirica" do grupo de validação para os três cenários e coeficientes de herdabilidades simulados.

Características – $b(fen, EBV)$ / $b(fen, GEBV)$ ± EP				
PBLUP				
Cenários	Ne	h² 0,15	h² 0,30	h² 0,45
1	320	0,79±0,034	0,80±0,051	0,81±0,061
2	128	0,70±0,072	0,71±0,093	0,72±0,115
3	64	0,69±0,120	0,70±0,067	0,72±0,081
ssGBLUP				
Cenários	Ne	h² 0,15	h² 0,30	h² 0,45
1	320	0,83±0,091	0,87±0,110	0,90±0,067
2	128	0,77±0,082	0,80±0,079	0,81±0,078
3	64	0,45±0,174	0,47±0,068	0,54±0,096

Ne: tamanho efetivo da população; $b(fen, EBV)$ coeficiente de regressão dos fenótipos simulados nos EBV (valor genético estimado); $b(fen, GEBV)$ coeficiente de regressão do fenótipos simulados nos GEBV (valor genético genômico estimado); h²: herdabilidades; EP: erro-padrões.

Os coeficientes de regressão dos fenótipos simulados sobre os EBVs para o grupo de validação variaram de 4,79±0,079 a 2,71±0,021 para o cenário 1 (Ne=320); 4,86±0,125 a 2,75±0,033 para o cenário 2 (Ne=128); e 5,09±0,181 a 2,79±0,047 para o cenário 3 (Ne=64). Estes valores foram decrescentes de acordo com o aumento dos coeficientes de herdabilidade simulados (Tabela 8). Os coeficientes de regressão dos fenótipos simulados sobre os GEBVs apresentaram no cenário 1 valores de 0,77 para a herdabilidade de 0,15; e 0,80 e 1,05 para os coeficientes de herdabilidade de 0,30 e 0,45, respectivamente. No cenário 2, os valores dos coeficientes de regressão foi de 0,80 para a herdabilidade de 0,15, e chegou a 0,99 para o coeficiente de herdabilidade de 0,45. Para o cenário 3, esses valores foram de 0,81; 0,90 e 0,98 para os três coeficientes de herdabilidade simulados (0,15; 0,30 e 0,45), respectivamente (Tabela 8).

Tabela 8. Viés de predição "empírico" do grupo de validação para os três cenários e coeficientes de herdabilidades simulados

Características – $r = cov(TBV, EBV)$ / $r = cov(TBV, GEBV)$ ± EP				
PBLUP				
Cenários	Ne	h² 0,15	h² 0,30	h² 0,45
1	320	4,79±0,079	3,34±0,036	2,71±0,021
2	128	4,86±0,125	3,51±0,057	2,75±0,033
3	64	5,09±0,181	3,55±0,080	2,79±0,047
ssGBLUP				
Cenários	Ne	h² 0,15	h² 0,30	h² 0,45
1	320	0,77±0,005	0,80±0,027	1,05±0,034
2	128	0,80±0,008	0,83±0,010	0,99±0,010
3	64	0,81±0,012	0,90±0,007	0,98±0,027

Ne: tamanho efetivo da população; $r = cov(TBV, EBV)$: correlação entre TBV (valor genético verdadeiro) e EBV (valor genético estimado). $r = cov(TBV, GEBV)$ correlação entre TBV (valor genético verdadeiro) e GEBV (valor genético genômico estimado); h²: herdabilidades; EP: erro-padrões.

4.4 Discussão

4.4.1 Desequilíbrio de Ligação

Estudos de desequilibrio de ligação do genoma completo podem fornecer informações sobre a história da população. Além disso, o LD em regiões genômicas permitem aprender sobre o histórico de seleção e outras forças que possam ter causado a evolução da frequência gênica.

Nesse contexto, este estudo teve como objetivo entender os impactos das estratégias de seleção e dos tamanhos efetivos da população nos valores médios de r^2 para três cenários de populações simuladas com diferentes Ne e coeficientes de herdabilidade (0,15; 0,30 e 0,45). Assim, o cenário 3 representou o tamanho efetivo populacional dos rebanhos de linhagens caipiras da ESALQ que foi base para este estudo. Os resultados mostraram que a diminuição do Ne das populações podem impactar no aumento do desequilíbrio de ligação (LD) da população em estudo.

Uma vez que, vários estudos de LD foram realizados em aves, esses por sua vez, apresentaram valores de r^2 abaixo de 0,30, em média, para distâncias entre 0,25 e 1 cM (Aerts et al. 2007; Andreescu et al. 2007; Heifetz et al. 2005; Megens et al. 2009). Aerts et al. (2007) investigaram a extensão do LD em dois cromossomos em duas linhagens de frangos de corte. Os níveis de LD foram calculados para 33 e 36 marcadores nos cromossomos 10 e 28, respectivamente; descobrindo-se que o "LD útil" é um valor de r^2 superior a 0,30, para distâncias entre os marcadores de 0,8 a 4 cM em linhagens de frangos de corte. Os autores concluíram assim, que pode haver diferenças grandes nos padrões de LD entre diferentes raças de galinhas e diferentes regiões genômicas. No presente estudo, os valores médios de r^2, a distâncias dos marcadores de 0-0,05 Kb, foram inferiores ou próximos a 0,30 nas Fases 1 e 2 dos cenários 1 e 2, devido ao Ne (500 e 320 no cenário 1; e 200 e 128 no cenário 2) e as estratégias de seleção aplicados (indivíduos sob nenhum processo de seleção na Fase 1 e seleção por características fenotípicas na Fase 2). Para o cenário 3, de menor tamanho efetivo populacional os valores médios de r^2 foram superiores a 0,30 em todas as fases e coeficientes de herdabilidade simulados. Na Fase 3, com estratégia de seleção utilizada foram obtidos os maiores valores de r^2 em todos os cenários, mostrando mais uma vez, que a seleção é um fator que tende a aumentar o LD. Qanbari et al. (2010) encontraram valores médios de r^2 de 0,73± 0,36 em distâncias pareadas <25 Kb para linhas brancas comerciais, e ocorreu um decréscimo para 0,60±0,38 à distâncias de 75 a 120 Kb, valores aproximados aos que foram encontrados neste trabalho para a Fase 3 dos três cenários e os três coeficientes de herdabilidade simulados a menores distâncias (0 a 0,05 Kb).

Em outras espécies, Marchiori et al. (2019) realizaram um estudo com equinos da raça Quarto de Milha; e caracterizaram o desequilíbrio de ligação de linhagens de animais para trabalho e corrida pelo coeficiente de correlação (r^2). Além disso, o Ne das duas populações foi determinado utilizando informações de distâncias entre os SNPs. Foram utilizados 68 equinos de trabalho e 360 de corrida, totalizando 428 equinos de ambos os sexos. Assim, o r^2 genômico médio entre os pares de marcadores foi de 0,22 para a linha de trabalho e 0,27 para a linha de corrida. O Ne era de 60 e 50 animais efetivos na última geração das linhas de trabalho e corrida,

respectivamente. Estes autores concluíram que a maior extensão de LD (0,27) e Ne menor (50) foram encontrados na linha de corrida, pela estruturação populacional desta linha, o que ocorreu com o cenário 3 (Ne= 100 na Fase 1 e Ne=64 nas Fases 2 e 3) deste trabalho, que se baseia no Ne das linhagens caipiras nacionais em condições reais. Jasiellczuk et al. (2020) caracterizaram o desequilíbrio de ligação (LD) e tamanho efetivo da população (Ne) usando informações de nucleotídeos de polimorfismo único (SNPs) em 135 suínos Landrace polonoses (PL) e compararam esses dados com informações de suínos de raças nativas polonesas conservadas Puławska (PUL; n = 155), Złotnicka White (ZW; n = 141) e Złotnicka Spotted (ZS; n = 99). A distâncias de até 100 Kb as médias de r^2 foram de 0,33 para PL; 0,34 para PUL; 0,39 para ZW; e 0,45 para ZS. A distâncias de 4-5 Mb os valores médios de r^2 foram de 0,05; 0,07; 0,11 e 0,15 para PL, PUL, ZW w ZS, respectivamente. O Ne variou de 147 (ZS) a 263 (PL) para 100 gerações anteriores e de 23 (ZS) a 91 (PL) nas últimas cinco gerações, deixando evidente, que menores extensões de LD são encontradas em populações com maiores Ne, o que corrobora com o este estudo de simulação, em que nos cenários 1 e 2 com maiores Ne, as extensões de LD (r^2) foram menores, enquanto que, e para o cenário 3, que representa a população de linhagens caipiras as extensões de LD foram elevadas pela diminuição do tamanho efetivo populacional.

4.4.2 Acurácia e viés de predição "empíricos"

Os benefícios da seleção genômica sobre a seleção baseada nas informações de pedigree dos animais foi bem documentada a partir de trabalhos como este, baseados em simulações e estudos empiricos em sistemas de criação de bovino, suínos; e especialmente aves. Chu et al. (2019) compararam os procedimentos PBLUP e ssGPLUP para a avaliação da acurácia de predição de 3,5 gerações das aves com informações de pedigree incluindo 70.174 aves; e para análises genômicas, após o controle de qualidade, os dados continham 39.767 aves genotipadas com 50.562 SNPs para a construção da matriz de relacionamento genômico. Assim, estes autores, observaram um aumento relativo de precisão de 31,7% a 73,1% para ave genotipadas e de 6,3% a 14,9% para aves não genotipadas, ou seja, valores baseados no conjunto de dados completo e reduzido, mostrando que a incorporação de informações genômicas é eficiente.

Neste estudo, as correlações entre os TBV e o EBV genômico de indivíduos genotipados, para os cenários 1 e 2, apresentaram valores elevados em relação aos valores de correlações entre os TBVs e EBVs, exceto para o cenário 3, que possui um tamanho efetivo da população menor em relação aos outros cenários. O cenário 3, conformejá descrito anteriormente, representa de forma hipotética um rebanho de linhagens caipiras da ESALQ que é estudado em um programa

de melhoramento genético real, apesar de outros estudos de simulação terem mostrado que as correlações entre os TBV e os GEBVs foram significativamente maiores com o ssGBLUP do que com a metodologia PBLUP (Andonov et al, 2017, Christensen et al., 2012; Lourenço et al., 2013; Putz et al, 2018).

Assim, os cenários 1, 2 e 3 simulados possuem 3200, 1280 e 640 animais genotipados no grupo de validação, respectivamente. Por isso, no cenário 3 possui Ne base pequeno, contando com apenas 50 machos na última geração da população histórica, pode ter influenciado nos resultados de acurácia e viés de predição deste estudo. Os valores de correlações entre TBV e GEBV foram menores que os valores dos cenários 1 e 2, que possuem um maior número de animais genotipados na população. Conseguinte, o relacionamento dos indivíduos é importante para uma acurácia de predição aprimorada, pois quanto mais aparentados forem os indivíduos maiores serão os efeitos positivos na acurácia de predição (Habier et al, 2010). Por isso, o cenário 3 que possui um pequeno tamanho efetivo da população base pode por sua vez, apresentar um intenso desequilíbrio de ligação (LD), o que corrobora com os resultados de LD obtidos neste trabalho, demonstrando que os níveis elevados de LD podem significaar que os marcadores tem maiores chances de captura de QTLs (quantitative trait loci) (Yan et al, 2018).

Pelo viés de predição dos EBVs e dos GEBVs podem ser mensurados o grau de inflação ou não das predições genômicas. Se os valores dos coeficientes de regressão forem maiores que 1 o modelo pode estar subestimando os EBVs ou GEBVs, o que acontece nos tres cenários e para os três coeficientes de herdabilidade simulados neste trabalho pelos coeficientes de regressão dos fenótipos simulados nos EBVs (Tabela 8). Pode ocorrer a superestimação dos EBVs ou GEBVs se os valores dos coeficientes de regressão forem menores que 1, ou seja, as estimativas de regressões proximas a 1 podem ser um indicativo de que o preditor é não viesado em acordo com os valores dos coeficientes de regressão dos fenótipos simulados nos GEBVs deste trabalho para os três cenários. Por isso, o viés de predição seria um bom critério para a comparação de diferentes ferramentas de predição.

Dessa maneira, para que as predições sejam acuradas, as relações entre os animais dentro da população de treinamento devem ser fracas e a relação entre a população de treinamento e a população de validação necessitam ser altas. Em aves e suínos, por exemplo, o número de descendentes por machos é menor em relação ao número por as fêmeas. Com isso, a genotipagem das fêmeas pode influenciar substancialmente na acurácia, e consequentemente, no progresso genético da população estudada. No presente estudo, foram simulados dados genotípicos de machos e de fêmeas, o que pode ter influenciado nos elevados valores de acurácia de predição genômica dos cenários 1 e 2, uma vez que a metodologia ssGBLUP, pode modelar fenótipos e

genótipos de ambos os sexos quando os genótipos não estão disponívies para toda população (Lourenço et al, 2015).

Com isso, a predição genômica também é influenciada pelo tamanho efetivo da população (Ne). Hayes et al. (2009) relataram que, em uma população com Ne=100, como o cenário 3 simulado, seriam necessários 12.474 indivíduos com genótipos e fenótipos conhecidos para prever GEBVs de indivíduos não fenotipados na mesma população visando uma precisão de 0,7. Ou seja, para que a acurácia de predição do cenário 3 deste trabalho fosse maior, seria necessário um maior número de informações de animais com fenótipos e genótipos conhecidos. Estes autores também descreveram que a diminuição da herdabilidade pode diminuir a acurácia, porque a razão da variância fenotípica para a variância dos efeitos dos QTLs aumenta, assim como o número de segmentos cromossômicos efetivos, o que pode ser notado com o aumento dos coeficientes de herdabilidade simulados, ao longo dos três cenários (Tabela 7).

Nesse contexto, se o Ne é maximizado, o número de segmentos cromossômicos independentes é tambem grande, significando que a extensão do LD será limitada na população, o que mais uma vez pode ser observado nos resultados deste estudo de simulação do cenário 1 (Ne=320). Um número grande de marcadores e registros fenotípicos será necessário para capturar os efeitos dos QTLs e para prever os valores genéticos para indivíduos não fenotipados com precisão, particularmente se o coeficiente de herdabilidade foi baixo (Hayes et al., 2009b).

Portanto, a predição genômica é vantajosa em aumentar a precisão dos EBVs (GEBVs), permitindo a seleção de animais mais jovens e a consequente diminuição do intervalo de gerações (Hayes et al, 2009a). Em vista disso, diversos fatores podem influenciar na acurácia da predição genômica, como o tamanho da população de treinamento (Goddard, 2009; Habier et al., 2010), o coeficiente de herdabilidade da característica, a relação entre a população de treinamento e de validação (Wientjes et al., 2013), o tamanho efetivo da população (Goddard, 2009), o desequilíbrio de ligação (LD) entre os marcadores e QTLs (Habier et al., 2007; Wientjes et al., 2013), além da densidade de marcadores utilizada (Meuwissen e Goddard, 2010), número de QTL (Daetwyler et al., 2010), frequências alélicas menores de mutações causais e marcadores de predição (Druet et al., 2014; Wientjes et al., 2015a; Van Den Berg et al. 2019).

4.5 Conclusão

O LD das populações de linhagens caipiras da ESALQ pode ser considerado "útil" e eficiente para estudos de associação e predições genômicas que estão relacionados ao desequilíbrio de ligação, mesmo possuindo um tamanho efetivo da população pequeno, mas para que os valores de acurácia de predição sejam maiores será necessário o aumento do Ne destes rebanhos de

linhagens caipiras da ESALQ para que as avaliações genômicas não sejam influenciadas negativamente.

Referências

Aerts, J. et al. Extent of linkage disequilibrium in chicken. Cytogenetic and Genome Research, v. 117, n. 1–4, p. 338–345, 2007.

Andonov, S. et al. Accuracy of breeding values in small genotyped populations using different sources of external information—A simulation study. **Journal of Dairy Science**, v. 100, n. 1, p. 395–401, 2017.

Blackburn, H. D. The National Animal Germplasm Program: Challenges and opportunities for poultry genetic resources. Poultry Science, v. 85, n. 2, p. 210–215, 2006.

Christensen, O. F. et al. Single-step methods for genomic evaluation in pigs. **Animal**, v. 6, n. 10, p. 1565–1571, 2012.

Chu, T. T. et al. Use of genomic information to exploit genotype-by-environment interactions for body weight of broiler chicken in bio-secure and production environments. Genetics Selection Evolution, v. 51, n. 1, p. 1–13, 2019.

Ciampolini, R. et al. Investigation on the genetic variability of the american Pit Bull Terrier dogs belonging to an Italian breeder using microsatellite markers and genealogical data. Cytology and Genetics, v. 47, n. 4, p. 217–221, 2013.

Coelho, A. A. D.; Savino, V. J. M.; Rosário, M. F. Frango feliz: caminhos para a avicultura alternativa. [S.l: s.n.], 2008.

Dalvit, C. et al. Genetic variation and population structure of Italian native sheep breeds undergoing in situ conservation. Journal of Animal Science, v. 87, n. 12, p. 3837–3844, 2009.

Daetwyler, H. D. et al. Accuracy of estimated genomic breeding values for wool and meat traits in a multi-breed sheep population. Animal Production Science, v. 50, n. 11–12, p. 1004–1010, 2010.

De Los Campos, G. et al. Whole-genome regression and prediction methods applied to plant and animal breeding. Genetics, v. 193, n. 2, p. 327–345, 2013.

Druet, T.; Macleod, I. M.; Hayes, B. J. Toward genomic prediction from whole-genome sequence data: Impact of sequencing design on genotype imputation and accuracy of predictions. Heredity, v. 112, n. 1, p. 39–47, 2014.

FAO. Molecular genetic characterization of animal genetic resources. FAO Animal Production and Health Guidelines. 2011.

Fulton, J. E.; Delany, M. E. Olicy orum. v. 300, n. June, p. 1667–1668, 2003.

Goddard, M. E.; Hayes, B. J. Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nature Reviews Genetics, v. 10, n. 6, p. 381–391, 2009.

Habier, D.; Fernando, R. L.; Dekkers, J. C. M. The impact of genetic relationship information on genome-assisted breeding values. Genetics, v. 177, n. 4, p. 2389–2397, 2007.

Habier, D. et al. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. Genetics Selection Evolution, v. 42, n. 1, p. 1–12, 2010.

Hayes, B. J. et al. Accuracy of genomic breeding values in multi-breed dairy cattle populations. Genetics Selection Evolution, v. 41, n. 1, p. 1–9, 2009a.

Hayes, B. J.; Visscher, P. M.; Goddard, M. E. Increased accuracy of artificial selection by using the realized relationship matrix. Genetics Research, v. 91, n. 1, p. 47–60, 2009b.

Hillel, J. et al. Biodiversity of 52 chicken populations assessed by microsatellite typing of DNA pools. Genetic Selection Evolution. 35:533–557, 2003.

Jasielczuk, I. et al. Comparison of linkage disequilibrium, effective population size and haplotype blocks in Polish Landrace and Polish native pig populations. Livestock Science, v. 231, n. August 2019, p. 103887, 2020.

Lourenco, D. A. L. et al. Prediction accuracy for a simulated maternally affected trait of beef cattle using different genomic evaluation models. **Journal of Animal Science**, v. 91, n. 9, p. 4090–4098, 2013.

Lourenco, D. A. L. et al. Accuracy of estimated breeding values with genomic information on males, females, or both: An example on broiler chicken. Genetics Selection Evolution, v. 47, n. 1, p. 1–9, 2015.

Marchiori, C. M. et al. Linkage disequilibrium and population structure characterization in the cutting and racing lines of Quarter Horses bred in Brazil. Livestock Science, v. 219, n. August 2018, p. 45–51, 2019.

Megens, H. J. et al. Comparison of linkage disequilibrium and haplotype diversity on macro- and microchromosomes in chicken. BMC Genetics, v. 10, p. 1–11, 2009.

Meuwissen, T.; Goddard, M. Accurate prediction of genetic values for complex traits by whole-genome resequencing. Genetics, v. 185, n. 2, p. 623–631, 2010.

Misztal, I. et al. Blupf90 and Related Programs (Bgf90). p. 2001–2002, 2002.

Momen, M. et al. Predictive ability of genome-assisted statistical models under various forms of gene action. Scientific Reports, v. 8, n. 1, p. 1–11, 2018.

Putz, A. M. et al. A comparison of accuracy validation methods for genomic and pedigree-based predictions of swine litter size traits using Large White and simulated data. **Journal of Animal Breeding and Genetics**, v. 135, n. 1, p. 5–13, 2018.

Qanbari, S. et al. Linkage disequilibrium reveals different demographic history in egg laying chickens. **BMC Genetics**, v. 11, 2010.

Ramírez-Flores, F. et al. Accuracy of genomic values predicted using deregressed predicted breeding values as response variables. Revista Mexicana De Ciencias Pecuarias, v. 8, n. 4, p. 445–451, 2017

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

VanRaden, P. M. Efficient methods to compute genomic predictions. Journal of Dairy Science, v. 91, n. 11, p. 4414–4423, 2008.

Van Den Berg, I. et al. Predicting the effect of reference population on the accuracy of within, across, and multibreed genomic prediction. Journal of Dairy Science, v. 102, n. 4, p. 3155–3174, 2019.

Viale, E. et al. Development and validation of a novel SNP panel for the genetic characterization of Italian chicken breeds by next-generation sequencing discovery and array genotyping. Poultry Science, v. 96, n. 11, p. 3858–3866, 2017.

Wientjes, Y. C. J.; Veerkamp, R. F.; Calus, M. P. L. The effect of linkage disequilibrium and family relationships on the reliability of genomic prediction. Genetics, v. 193, n. 2, p. 621–631, 2013.

Wientjes, Y. C. J. et al. Impact of QTL properties on the accuracy of multi-breed genomic prediction. Genetics Selection Evolution, v. 47, n. 1, p. 1–16, 2015.

Yan, Y. et al. Genomic prediction in a nuclear population of layers using single-step models. Poultry Science, v. 97, n. 2, p. 397–402, 2018.

Zanetti, E. et al. Variation of genetic diversity over time in local Italian chicken breeds undergoing in situ conservation. Poultry Science, v. 90, n. 10, p. 2195–2201, 2011

Zanetti, E. et al. Using 2-DE for the differentiation of local chicken breeds. Proteomics, v. 11, n. 13, p. 2613–2619, 2011.

ANEXOS

Anexo I - Exemplo utilizado para simulação do capítulo um para o tamanho efetivo de 100 (Fase 1) e 64 (Fases 2 e 3); e para a herdabilidade de 0,15 (adaptado exemplo 7 do software QMSim - Sargolzaei; Schenkel, 2009).

```
  /******************************
 **     Global parameters     **
 *******************************/
title = "Example 7 - population expansion in a recent population";
nrep = 10;       //Number of replicates
h2    = 0.15;      //Heritability
qtlh2 = 0.15;     //QTL heritability
phvar = 1.0;      //Phenotypic variance
/******************************
 ** Historical population    **
 *******************************/
begin_hp;
   hg_size = 2000 [0]
             2000[1000]
             400 [1010]
             400 [1020]
             100 [1100];    //Size of the historical generations
   nmlhg   = 50;          //Number of males in the last generation
end_hp;
/******************************
 **       Populations        **
 *******************************/
  begin_pop = "p1";
     begin_founder;
          male [n = 50, pop = "hp"];
          female [n = 50, pop = "hp"];
     end_founder;
     ls = 8; //Litter size
     pmp = 0.5; //Proportion of male progeny
     ng = 40; //Number of generations
     md = rnd; //Mating design
     sr = 1; //Replacement ratio for sires
     dr = 1; //Replacement ratio for dams
     sd = rnd; //Selection design
     cd = rnd; //Culling design
     begin_popoutput;
        ld /maft 0.02 /gen 0 40;
        data;
        stat;
```

genotype/ snp_code/ gen 40;

```
      allele_freq /gen 40;
  end_popoutput;
end_pop;

begin_pop = "p2";
    begin_founder;
        male [n = 20, pop = "p1", gen= 40];
        female [n = 80, pop = "p1", gen=40];
    end_founder;
    ls = 8; //Litter size
    pmp = 0.5; //Proportion of male progeny
    ng = 25; //Number of generations
    md = rnd; //Mating design
    sr = 1; //Replacement ratio for sires
    dr = 1; //Replacement ratio for dams
    sd = phen /h; //Selection design
    cd = age; //Culling design
    begin_popoutput;
      ld /maft 0.02 /gen 0 25;
      data;
      stat;
      genotype/ snp_code/ gen 25;
      allele_freq /gen 25;
  end_popoutput;
end_pop;

begin_pop = "p3";
    begin_founder;
        male [n = 20, pop = "p2", gen= 25];
        female [n = 80, pop = "p2", gen=25];
    end_founder;
    ls = 8; //Litter size
    pmp = 0.5; //Proportion of male progeny
    ng = 15; //Number of generations
    md = rnd; //Mating design    sr = 1; //Replacement ratio for sires
    dr = 1; //Replacement ratio for dams
    sd = ebv /h; //Selection design
    cd = age; //Culling design
    ebv_est = blup;
    begin_popoutput;
      ld /maft 0.02 /gen 0 15;
      data;
      stat;
      genotype/ snp_code/ gen 15;
      allele_freq /gen 15;
  end_popoutput;
end_pop;

/******************************
 **        Genome          **
```

```
 ******************************/
begin_genome;
  begin_chr = 28;
    chrlen = 34;    //Chromosome length
    nmloci = 21429; //Number of markers
    mpos   = rnd;   //Marker positions
    nma    = all 2; //Number of marker alleles
    maf    = rnd;   //Marker allele frequencies
    nqloci = 12;    //Number of QTL
    qpos   = rnd;   //QTL positions
    nqa    = all 2; //Number of QTL alleles
    qaf    = rnd;   //QTL allele frequencies
    qae    = rndg 0.4; //QTL allele effects
  end_chr;
    mmutr     = 2.5e-5 /recurrent;  //Marker mutation rate
    qmutr     = 2.5e-5;      //QTL mutation rate
    r_mpos_g;        // Randomize marker positions across genome
    r_qpos_g;       // Randomize marker positions across genome
end_genome;

/******************************
 **      Output options      **
 ******************************begin_output;
  linkage_map;
  allele_effect;
  monitor_hp_homo /freq 100;
end_output;
```

Anexo II - Exemplo utilizado para simulação do capítulo dois para tamanho efetivo da população de 100 (Fase 1) e 64 (Fases 2 e 3); e herdabilidade de 0,15 (adaptado exemplo 7 do software *QMSim* - Sargolzaei; Schenkel, 2009).

```
  /******************************
 **     Global parameters     **
 ******************************/
title = "Example 7 - population expansion in a recent population";
nrep = 10;     //Number of replicates
h2    = 0.15;   //Heritability
qtlh2 = 0.15;  //QTL heritability
phvar = 1.0;   //Phenotypic variance

/******************************
 **  Historical population   **
 ******************************/
begin_hp;
  hg_size =  2000 [0]
             2000[1000]
             400 [1010]
             400 [1020]
             100 [1100];          //Size of the historical generations
  nmlhg   = 50;              //Number of males in the last generation
end_hp;

/******************************
 **        Populations        **
 ******************************/
 begin_pop = "p1";
    begin_founder;
        male [n = 50, pop = "hp"];
        female [n = 50, pop = "hp"];
    end_founder;
    ls = 8; //Litter size
    pmp = 0.5 /fix; //Proportion of male progeny
    ng = 40; //Number of generations
    md = rnd; //Mating design
    sr = 1; //Replacement ratio for sires
    dr = 1; //Replacement ratio for dams
    sd = rnd; //Selection design
    cd = rnd; //Culling design
    begin_popoutput;
      data;
      stat;
      genotype /snp_code /gen 40;
      allele_freq;
  end_popoutput;
end_pop;

begin_pop = "p2";
    begin_founder;
```

```
        male [n = 20, pop = "p1", gen= 40];
        female [n = 80, pop = "p1", gen=40];
    end_founder;
    ls = 8; //Litter size
    pmp = 0.5 /fix; //Proportion of male progeny
    ng = 25; //Number of generations
    md = rnd; //Mating design
    sr = 1; //Replacement ratio for sires
    dr = 1; //Replacement ratio for dams
    sd = phen /h; //Selection design
    cd = age; //Culling design
    begin_popoutput;
      data;
      stat;
      genotype /snp_code /gen 25;
      allele_freq;
  end_popoutput;
end_pop;

begin_pop = "p3";
    begin_founder;
        male [n = 20, pop = "p2", gen= 25];
        female [n = 80, pop = "p2", gen=25];
    end_founder;
    ls = 8; //Litter size
    pmp = 0.5/ fix; //Proportion of male progeny
    ng = 15; //Number of generations
    md = rnd; //Mating design
    sr = 1; //Replacement ratio for sires
    dr = 1; //Replacement ratio for dams
    sd = ebv /h; //Selection design
    cd = age; //Culling design
    ebv_est = blup /true_av;
    begin_popoutput;
      data;
      stat;
      genotype /snp_code /gen 11 12 13 14 15;
      allele_freq;
  end_popoutput;
end_pop;

/******************************
 **        Genome        **
 ******************************/
begin_genome;
  begin_chr = 28;
    chrlen = 34;     //Chromosome length
    nmloci = 2143;   //Number of markers
    mpos   = rnd;    //Marker positions
    nma    = all 2; //Number of marker allele
    maf    = rnd;   //Marker allele frequencies
```

```
    nqloci = 12;    //Number of QTL
    qpos   = rnd;   //QTL positions
    nqa    = all 2; //Number of QTL alleles
    qaf    = rnd;  //QTL allele frequencies
    qae    = rndg 0.4; //QTL allele effects
  end_chr;
  mmutr     = 2.5e-5 /recurrent; //Marker mutation rate
  qmutr     = 2.5e-5;  //QTL mutation rate
  r_mpos_g;       // Randomize marker positions across genome
  r_qpos_g;      // Randomize marker positions across genome
end_genome;

/******************************
 **      Output options      **
 ******************************/
begin_output;
  linkage_map;
  hp_stat;
  allele_effect;
  monitor_hp_homo;
end_output
```