

Brasília, DF / Maio, 2024

## Métodos ótimos de seleção genética no melhoramento do cafeeiro arábica



**Empresa Brasileira de Pesquisa Agropecuária  
Embrapa Café  
Ministério da Agricultura e Pecuária**

ISSN 1678-1694 / e-ISSN 2965-9159

# **Documentos 19**

Maio, 2024

## **Métodos ótimos de seleção genética no melhoramento do cafeeiro arábica**

*Marcos Deon Vilela de Resende  
Rodrigo Silva Alves  
Eveline Teixeira Caixeta  
Antonio Carlos Baião de Oliveira  
Antonio Alves Pereira  
Bruna Lopes Mariz*

**Embrapa Café**  
Brasília, DF  
2024

**Embrapa Café**  
Parque Estação Biológica (PqEB)  
Av. W3 Norte (final), Ed. Sede  
CEP: 70770-901, Brasília, DF  
www.embrapa.br/cafe  
www.embrapa.br/fale-conosco/sac

Comitê Local de Publicações

Presidente

*Lucas Tadeu Ferreira*

Vice-Presidente

*Jamilsen de Freitas Santos*

Secretária-Executiva

*Adriana Maria Silva Macedo*

Membros

*Anísio José Diniz*

*Carlos Henrique Siqueira de Carvalho*

*Helena Maria Ramos Alves*

*Lucilene Maria de Andrade*

*Mauricio Sergio Zacarias*

*Milene Alves de Figueiredo Carvalho*

*Omar Cruz Rocha*

*Rogério Novais Teixeira*

*Roseane Pereira Villela*

Revisão de texto

*Maria Cristina Ramos Jubé*

Normalização bibliográfica

*Márcia Maria de Souza Pereira*

Projeto gráfico

*Leandro Sousa Fazio*

Diagramação

*Thiago Farah Cavaton*

Foto da capa

*Carlos Alberto Alves Meira*

Publicação digital: PDF

#### **Todos os direitos reservados**

A reprodução não autorizada desta publicação, no todo ou em parte, constitui violação dos direitos autorais (Lei nº 9.610).

#### **Dados Internacionais de Catalogação na Publicação (CIP)**

Embrapa Café

---

Métodos ótimos de seleção genética no melhoramento do cafeeiro arábica / Marcos Deon Vilela de Resende ... [et al.]. – Brasília, DF : Embrapa Café, 2024.

PDF (58 p.) : il. color. (Documentos / Embrapa Café, e-ISSN 2965-9159 ; 19)

1. *Coffea arabica* 2. Melhoramento genético 3. Espécies autógamas. I. Alves, Rodrigo Silva. II. Caixeta, Eveline Teixeira. III. Oliveira, Antonio Carlos Baião de. IV. Pereira, Antonio Alves. V. Mariz, Bruna Lopes. VI. Série.

CDD (21. ed.) 633.73

---

*Rejane Maria de Oliveira Cechinel Darós* (CRB-1/2913)

© 2024 Embrapa

## Autores

---

**Marcos Deon Vilela de Resende**

Engenheiro-agrônomo, doutor em Genética, pesquisador da Embrapa Café, Brasília, DF

**Rodrigo Silva Alves**

Engenheiro florestal, doutor em Genética e Melhoramento, pós-doutorando no Instituto Nacional de Ciência e Tecnologia do Café, Lavras, MG

**Eveline Teixeira Caixeta**

Engenheira-agrônoma, doutora em Genética e Melhoramento, pesquisadora da Embrapa Café, Brasília, DF

**Antonio Carlos Baião de Oliveira**

Engenheiro-agrônomo, doutor em Genética e Melhoramento, pesquisador da Embrapa Café, Brasília, DF

**Antonio Alves Pereira**

Engenheiro-agrônomo, doutor em Fitopatologia, pesquisador da Empresa de Pesquisa Agropecuária de Minas Gerais (Epamig), Viçosa, MG

**Bruna Lopes Mariz**

Engenheira-agrônoma, doutoranda em Genética e Melhoramento na Universidade Federal de Viçosa, Viçosa, MG

Agradecemos à Fundação de Amparo  
à Pesquisa do Estado de Minas Gerais  
(Fapemig), ao Instituto Nacional de Ciência  
e Tecnologia do Café (INCT Café) e à  
Embrapa Café pelo apoio financeiro.

## Apresentação

---

Os principais fatores que ditam o sucesso de um programa de melhoramento são: (i) eficiência no processo seletivo (métodos ótimos de identificação e seleção de indivíduos superiores); (ii) estratégia adequada de melhoramento (método de seleção recorrente e delineamento de cruzamento para a criação de novos indivíduos superiores); (iii) curtas gerações de melhoramento, ou seja, seleção precoce maximizando o ganho genético por unidade de tempo. Este documento desenvolve a teoria e realiza recomendações práticas acerca desses fatores na cultura do cafeeiro arábica.

Os métodos ótimos de seleção para plantas autógamas perenes são: índice matriz-progênie-indivíduo para seleção na fase inicial; índice família-indivíduo-progênie para seleção precoce; Blup individual dentro de gerações de autofecundação; Blup de famílias e multipopulações dentro de geração; Blup de famílias e multigerações, entre e dentro de gerações; Blup total multigerações com genitores, populações e progênies; Blup individual com medidas repetidas e endogamia; Blup individual com matriz de covariância estruturada para medidas repetidas; Blup individual para a seleção para homeostase sob bienalidade de produção.

A seleção combinada envolvendo três gerações mostrou-se ligeiramente superior à seleção combinada envolvendo duas gerações. O índice que utiliza

toda a variabilidade genética disponível entre e dentro de populações e gerações, combinando informações dos genitores, populações, progênies e gerações (SIPPPG) permite também fazer a seleção já em  $F_3$  com base em 6 informações: 3 de populações em  $F_1$  a  $F_3$ , uma de progênies em  $F_3$  e os dois pais. Pode-se ter também informações de plantas dentro de  $F_3$ .

Para a recombinação em um programa de seleção recorrente intrapopulacional em população híbrida (oriunda da  $F_1$ ), não é necessário postergar a seleção para além da  $F_4$  (seleção precoce em  $F_2$ ,  $F_3$  ou  $F_4$ ). Pode-se adotar seleção precoce na geração  $F_3$ , explorando-se a grande variabilidade genética entre e dentro de linhagens  $F_3$ . A heterose para vários caracteres revela que realizar a seleção logo após a hibridação  $F_1$  ou  $F_2$ , associada à propagação vegetativa, deverá ser a estratégia mais eficiente, pois, além de capturar a heterose máxima, reduz o comprimento do ciclo seletivo. A estratégia de hibridação seguida de seleção e propagação vegetativa comercial deverá ser usada rotineiramente visando ganhos máximos em produtividade e a perpetuação de indivíduos resistentes a doenças e pragas. Essa metodologia de seleção genética de café arábica é inovadora e eficiente, e usa simultaneamente dados de propagação vegetativa e seminal.

*Antonio Fernando Guerra*  
Chefe-Geral da Embrapa Café

# Sumário

---

<b>Introdução</b>	7
<b>Importância do melhoramento genético do cafeeiro</b>	8
<b>Maximização da acurácia seletiva no melhoramento genético</b>	11
<b>Origem do termo acurácia e propostas pioneiras para uso no melhoramento genético</b>	11
<b>Melhoramento de espécies autógamas perenes</b>	12
Modelos genéticos e componentes de variação genotípica	13
Comparação entre processos seletivos	13
Estatísticas F hierárquicas de Wright no melhoramento de populações autógamas	16
Esquema hierárquico para a estimação de componentes de variância	18
<b>Métodos ótimos de seleção em café arábica</b>	21
Índice matriz-progênie-indivíduo para seleção inicial	21
Índice família-indivíduo-progênie para seleção precoce com acurácia máxima	26
Blup individual dentro de gerações em plantas autógamas perenes	30
Blup de famílias e multipopulações dentro de geração em plantas autógamas perenes	31
Blup de famílias e multigerações em plantas autógamas perenes	32
Blup multigerações com genitores, populações e progênies em plantas autógamas perenes	34
Blup individual com medidas repetidas e endogamia em plantas autógamas perenes	38
Blup individual com matriz de covariância estruturada em plantas autógamas perenes	40
Blup individual para seleção para homeostase sob bienalidade	41
Blup individual via índice família $F_3$ -indivíduo $F_3$ -progênie $F_{3,4}$ -indivíduo $F_4$ , multigerações, para seleção recorrente intrapopulacional em espécies autógamas perenes	41
<b>Estratégias de melhoramento do café arábica: planta autógama perene, com propagação seminal e clonal</b>	44
Comparação de métodos de seleção recorrente	45
Seleção recorrente recíproca individual (SRRI)	46
Eficiência seletiva por unidade de tempo e seleção genômica ampla (GWS)	46
<b>Seleção genômica ampla</b>	47
<b>Estratégias de melhoramento propostas para o café arábica</b>	48
Estratégia de Seleção Recorrente (SRI)	48
Estratégia de Seleção Recorrente Recíproca (SRR)	48
<b>Interação genótipos x ambientes</b>	49
<b>Métodos de seleção, de melhoramento e de condução das gerações de autofecundação em espécies autógamas</b>	50
<b>Considerações finais</b>	51
<b>Referências</b>	51

## Introdução

Os principais fatores que ditam o sucesso de um programa de melhoramento são: a) eficiência no processo seletivo (métodos ótimos de seleção); b) estratégia adequada de melhoramento (método de seleção recorrente e delineamento de cruzamento); c) curtas gerações de melhoramento, ou seja, seleção precoce maximizando o ganho genético por unidade de tempo.

A genética quantitativa no melhoramento do café arábica (*Coffea arabica*) é tipicamente única, pois trata-se de uma planta autógama perene, de propagação seminal e vegetativa. Dessa forma, não pode ser totalmente espelhada em nenhuma outra espécie de planta. As plantas autógamas perenes são raras, destacando-se algumas fruteiras como o pêssego, a nectarina, o damasco e o limão, as quais possuem propagação vegetativa. Em função dessa especificidade, a abordagem quantitativa do melhoramento do cafeeiro arábica não avançou muito, em termos mundiais, pois sempre se baseou no melhoramento de plantas autógamas anuais. Assim, visando preencher essa lacuna, foi realizado o presente trabalho, cujos objetivos são desenvolver a teoria e realizar recomendações práticas acerca dos métodos ótimos de seleção (visando à correta identificação dos indivíduos superiores) e das estratégias efetivas de melhoramento genético (visando à criação de novos indivíduos superiores) do cafeeiro arábica. Isso permite também otimizar o uso das informações experimentais.

A obtenção de dados experimentais fidedignos é abordada em um outro documento que acompanha este. No presente trabalho, foram derivados métodos ótimos de seleção para café arábica, de acordo com várias situações: índice matriz-progênie-indivíduo para seleção na fase inicial; índice família-indivíduo-progênie para seleção precoce; melhor previsão linear imparcial (Blup) individual dentro de gerações para maximizar a eficiência da seleção em cada geração de autofecundação; Blup de famílias e multipopulações dentro de geração para aumentar a eficiência da seleção em cada geração de autofecundação; Blup de famílias e multigerações para uso simultâneo dos dados de todas as gerações; Blup

total multigerações com genitores, populações e progênies para uso simultâneo de todas as informações disponíveis em um programa de melhoramento; Blup individual com medidas repetidas e endogamia em plantas autógamas perenes; Blup individual com matriz de covariância estruturada em plantas autógamas perenes, para o adequado uso de medidas repetidas; Blup individual para seleção para homeostase sob bienalidade de produção.

Na fase inicial do melhoramento, o índice matriz-progênie-indivíduo (seleção combinada tripla) mostrou-se o método mais acurado (acurácia de 91% em F<sub>5</sub>, para um caráter com  $h^2$  de 20%, avaliado em 20 indivíduos por família), dentre 7 métodos avaliados. Na estratégia de melhoramento via seleção recorrente, o índice genitores-indivíduo-progênie (seleção combinada envolvendo três gerações) mostrou-se acurado (acurácia de 92% em indivíduo em F<sub>4</sub>, usando dados também de F<sub>3</sub> e F<sub>5</sub>, para um caráter com  $h^2$  de 20%, avaliado em 20 indivíduos por família). Os resultados revelaram que não é necessário postergar a seleção, visando à recombinação em um programa de seleção recorrente intrapopulacional em população híbrida (oriunda da F<sub>1</sub>), para além da F<sub>4</sub> com testes em F<sub>5</sub>. Isso pode ser um bom indicador da eficiência da seleção precoce (em F<sub>2</sub>, F<sub>3</sub> ou F<sub>4</sub>). A seleção combinada envolvendo três gerações mostrou-se ligeiramente superior à seleção combinada envolvendo duas gerações. Os resultados do Blup multigerações, de famílias e indivíduos, para uso simultâneo dos dados de todas as gerações, indicaram que, com adequada experimentação, a seleção pode ser praticada já em F<sub>4</sub> (seleção de indivíduos) ou F<sub>5</sub> (seleção de famílias). Essa informação é relevante aos programas de seleção recorrente em espécies autógamas. O índice que utiliza toda a variabilidade genética disponível entre e dentro de populações e gerações, combinando informações dos genitores, populações, progênies e gerações (SIPPPG), permite computar o ganho nas linhagens completamente homocigotas, quando se efetua a seleção entre pais e entre populações F<sub>1</sub> ou F<sub>2</sub>. Permite também fazer a seleção já em F<sub>3</sub> com base em 6 informações: 3 de populações em F<sub>1</sub> a F<sub>3</sub>, 1 de progênies em F<sub>3</sub> e os 2 pais. Pode-se ter também informações de plantas dentro de F<sub>3</sub>.



Os melhoristas de plantas perenes necessitam encurtar os ciclos seletivos e ter maior agressividade no lançamento de cultivares. Acumular dados experimentais na mesma geração pouco impacta o melhoramento de plantas perenes. A velocidade de troca de cultivares melhoradas é um fator essencial na competitividade do agronegócio. Em culturas de propagação seminal via linhas puras é importante se ter uma alta rotatividade de cultivares em plantios comerciais homogêneos, visando reduzir a vulnerabilidade dos plantios aos ataques de novas pragas e doenças. Gerações curtas de melhoramento facilitam isso. Pode-se adotar alguma forma de seleção precoce já na geração F<sub>3</sub>, explorando-se a grande variabilidade genética entre e dentro de progênies F<sub>3</sub>. Tal variabilidade contempla 1,5 vez a variância genética aditiva ( $\sigma_a^2$ ), sendo que 0,5  $\sigma_a^2$  encontra-se dentro de linhagem e 1,0  $\sigma_a^2$  encontra-se entre linhagens. Assim, tal geração é adequada para seleção, pois 75% (1,5) da variação aditiva total (2  $\sigma_a^2$ ) que estará disponível em F<sub>∞</sub> já se encontra disponível em F<sub>3</sub>. Dessa forma, a seleção em F<sub>3</sub> por meio de um método preciso como o Blup é relevante.

A seleção mais precoce possível é disponibilizada pela seleção genômica (SG). A SG propicia uma forma de seleção precoce direta (SPD), pois atua precocemente sobre genes expressos na idade adulta. Assim, a SPD propiciada pela seleção genômica é especialmente importante para o melhoramento de espécies perenes como o cafeeiro. A seleção genômica ampla mostrou-se potencial para o melhoramento de *C. arabica* por possibilitar a obtenção dos valores genéticos genômicos com boa acurácia e pela maior eficiência seletiva, por unidade de tempo, em relação à seleção fenotípica, o que possibilita a redução no tempo necessário para completar o ciclo de seleção e obter novas cultivares. O conhecimento prévio da herança do caráter torna-se necessário para determinar com segurança a estratégia ideal de melhoramento e a geração ideal para seleção.

Em café arábica (Cilas et al., 1998), foi mostrada a existência de heterose para várias características de interesse, o que é um indício da existência de dominância alélica. Assim sendo, pode-se dizer que, para o melhoramento dessas características, dificilmente será vantajoso retardar a seleção para gerações avançadas. Pelo contrário, realizar a seleção logo após a hibridação (F<sub>1</sub>) ou em F<sub>2</sub>, associada à propagação vegetativa, deverá ser a estratégia mais eficiente, pois, além de capturar a heterose máxima, reduz por muitos anos o ciclo seletivo. Em cafeeiro, a estratégia de hibridação seguida de seleção e

propagação vegetativa deverá ser usada rotineiramente visando a ganhos máximos em produtividade e à perpetuação de indivíduos resistentes a doenças e pragas (insetos e nematoides)

## Importância do melhoramento genético do cafeeiro

As duas espécies de cafeeiro mais cultivadas no mundo são *Coffea arabica* L. (café arábica) e *Coffea canephora* Pierre (café robusta e café conilon), as quais respondem por cerca de 60% e 40% da produção mundial. No Brasil, a principal espécie plantada é *C. arabica* (Minas Gerais, Espírito Santo, São Paulo e Paraná), e *C. canephora* a que apresenta área de plantio mais restrita (Espírito Santo, Rondônia e Sul da Bahia). Em nível mundial, *C. arabica* é produzido principalmente no Brasil, Colômbia, Etiópia, Honduras, Peru<sup>1</sup>, e *C. canephora*, no Vietnã, Indonésia, Brasil, Uganda e Índia<sup>2</sup>.

O estado de Minas Gerais é o maior produtor de café do Brasil, responde por cerca de 50% da produção nacional e é uma das principais fontes de cafés especiais do País. Praticamente 100% das plantações são de café arábica, cultivado em quatro regiões produtoras: sul de Minas, Cerrado de Minas, chapadas de Minas e matas de Minas, que exportam seus cafés pelos portos de Santos, Rio de Janeiro e Vitória. As seguintes instituições atuam no melhoramento do café arábica em Minas Gerais: Empresa de Pesquisa Agropecuária de Minas Gerais (Epmig)/Universidade Federal de Viçosa (UFV)/Embrapa Café, em Viçosa; Epmig/Universidade Federal de Lavras (Ufla)/Embrapa Café, em Lavras; e Fundação Procafé/Embrapa Café, em Varginha.

*C. arabica* é uma espécie autógama e apresenta distribuição geográfica natural restrita aos países Etiópia (região sudeste), Sudão (região sudeste) e Quênia (região norte), em altitudes entre 1.000 m e 2.000 m. E *C. canephora* é uma espécie autoestéril, alógama e tem ampla distribuição geográfica, ocorrendo em regiões de pequena altitude no oeste e centro do continente africano (Carvalho et al., 1985). O *Coffea liberica* ocorre com maior frequência no sul da Guiné, Costa do Marfim, Libéria, Gana, Gabão, Congo e no norte de Angola. Suas formas costumam constituir elementos característicos da floresta alta e densa. Essa espécie também é autoestéril,

<sup>1</sup> Disponível em: <https://www.indexmundi.com/agriculture/?commodity=green-coffee&graph=arabica-Production>.

<sup>2</sup> Idem.

alógama e altamente polimórfica. É cultivada em pequena extensão na Libéria, Suriname, Filipinas, Indonésia e Malásia.

Em razão da importância da cultura do café para o Brasil, os programas de melhoramento do cafeeiro desenvolvidos no País apresentam importância fundamental como fator de aumento da produtividade e do desenvolvimento socioeconômico das regiões de cultivo. Tais programas têm sido desenvolvidos com sucesso para *C. arabica* em Minas Gerais (Epamig/UFV/Ufla), Paraná (Instituto de Desenvolvimento Rural do Paraná – IDR – Paraná) e São Paulo (Instituto Agrônomo – IAC) (Fazuoli, 1986; Carvalho; Fazuoli, 1993; Sera, 1998; Sakiyama et al., 1999; Mendes, 1999); e *C. canephora* (Fazuoli, 1986; Carvalho; Fazuoli, 1993; Ferrão et al., 1999), principalmente no Espírito Santo (Instituto Capixaba de Pesquisa, Assistência Técnica e Entensão Rural – Incaper) e em Rondônia (Embrapa Café).

Estimativas de parâmetros genéticos, como a herdabilidade individual, são raras em café, citando-se apenas as obtidas por Walyaro e Van Der Vossen (1979), no Quênia, Cilas et al. (1998), em Camarões, ambas para *C. arabica*; Leroy et al. (1994), na Costa do Marfim, e Cilas et al. (2000), na Costa do Marfim e Togo, ambas para *C. canephora*. Estimativas de parâmetros genéticos e fenotípicos, em nível de médias de linhagens, são comuns no Brasil, podendo-se citar Sera (1981, 1987) e Sera e Alves (1999). Neste documento, as estimativas de herdabilidade individual foram atualizadas.

Na Tabela 1, são apresentados os 55 países maiores produtores mundiais de café, discriminados pelo idioma principal de cada um deles e por espécie. Na Figura 1, tem-se o mapa destacando os países plantadores de café e quais espécies plantam ou possuem de forma nativa ou autóctone.

**Tabela 1.** Listagem dos 52 países maiores produtores mundiais de café, discriminados pelo idioma principal de cada um deles e por espécie.

País idioma francês	País idioma espanhol	País idioma inglês	País idioma português
1. Vietnã (robusta/arábica)	1. Colômbia (arábica)	1. Indonésia (robusta/arábica)	1. Brasil (robusta/arábica)
2. Costa do Marfim (robusta)	2. México (arábica)	2. Etiópia (arábica)	2. Timor Leste (arábica)
3. Camarões (robusta/arábica)	3. Honduras (arábica)	3. Índia (robusta/arábica)	3. Angola (robusta)
4. Madagascar (robusta)	4. Peru (arábica)	4. Uganda (robusta/arábica)	4. Moçambique (robusta/arábica)
5. Laos (robusta)	5. Guatemala (arábica/robusta)	5. Papua Nova Guiné (robusta/arábica)	<b>País idioma árabe</b>
6. Burundi (arábica)	6. Costa Rica (arábica)	6. Tanzânia (robusta/arábica)	1. Iêmen (arábica)
7. Rep. Dem. do Congo-Kinshasa (ex Zaire) (arábica)	7. Nicarágua (arábica)	7. Quênia (arábica)	
8. Guiné (robusta)	8. El Salvador (arábica)	8. Filipinas (robusta/arábica)	
9. Ruanda (arábica)	9. Equador (arábica/robusta)	9. Gana (robusta)	
10. Haiti (arábica)	10. Venezuela (arábica)	10. Serra Leoa (robusta)	
11. Togo (robusta)	11. Rep. Dominicana (arábica)	11. Malawi (arábica)	
12. Rep. Centro Africana (robusta)	12. Bolívia (arábica)	12. Libéria (robusta)	
13. Camboja (robusta/arábica)	13. Cuba (arábica)	13. Zâmbia (arábica)	
14. Congo (ex Congo Belga-Brazzaville) (robusta)	14. Panamá (arábica)	14. Zimbábue (arábica)	
15. Benin (robusta)	-	15. Nigéria (robusta)	
16. Gabão (robusta)	-	16. Sudão (arábica)	
-	-	17. África do Sul (robusta/arábica)	



**Figura 1.** Mapa destacando os países plantadores de café, quais espécies plantam e/ou possuem de forma nativa ou autóctone.

Verifica-se, pelo mapa, que o plantio do café está confinado aos países da faixa intertropical, entre os Trópicos de Capricórnio ao sul (acima do Sul do Brasil e acima do norte da África do Sul) e de Câncer ao norte (abaixo do norte do México e abaixo do norte da Índia). Esses países da faixa intertropical falam predominantemente quatro idiomas ocidentais: francês, inglês, espanhol e português.

Conforme a Tabela 1, constata-se: a grande maioria (94,0%) dos países de idioma francês plantadores de café plantam café canéfora e situam-se na África; a grande maioria (86,0%) dos países de idioma espanhol plantadores de café plantam o café arábica e situam-se nas Américas; os países de idioma inglês plantam de forma equilibrada os cafés canéfora e arábica e situam-se na África Oriental e Sudeste Asiático. Associado a essas informações, verifica-se que a teoria (e aplicação prática) do melhoramento genético do café canéfora se desenvolveu mais que a teoria (e aplicação prática) do melhoramento genético do café arábica. Isso ocorreu por causa das contribuições teóricas dos franceses as suas ex-colônias na África. Esse tipo de tratamento não foi realizado pelos espanhóis e ingleses. Assim, o melhoramento do café arábica não teve ajuda dos europeus e não está teoricamente (em termos de métodos de seleção) tão avançado quanto ao robusta, conforme melhorado por brasileiros e franceses.

O melhoramento do café canéfora (espécie alógama, perene, de propagação clonal) foi supostamente mais facilitado pelo uso de experiências obtidas com outras espécies com características similares, como, por exemplo, o eucalipto e espécies fruteiras

e frutíferas alógamas. Entretanto, a genética quantitativa no melhoramento do café arábica é tipicamente única, pois trata-se de uma planta autógama perene, de propagação seminal e vegetativa. Dessa forma, não pode ser espelhada em nenhuma outra espécie de planta. As plantas autógamas perenes mais cultivadas são: café arábica, pêsego, nectarina, damasco, limão, coco-anão, mamão papaia, leucena e estilosantes. As fruteiras pêsego, nectarina, damasco e limão possuem propagação vegetativa. As demais espécies aqui relatadas seguem o melhoramento tradicional de espécies autógamas anuais, tais quais a soja, o feijão, o arroz e o trigo. E isso não é totalmente adequado para espécies autógamas perenes, as quais são colhidas muitas vezes durante suas vidas. Uma abordagem inicial para o melhoramento dessas espécies autógamas perenes foi reportada por Resende e Medrado (1994). Em função do exposto, o presente documento aborda a teoria do melhoramento e a comparação entre métodos de seleção de plantas autógamas perenes, de propagação seminal e clonal. Isso permitirá a otimização do melhoramento genético do café arábica.

Os parâmetros e conceitos mais importantes em genética quantitativa e melhoramento de plantas são: ganho genético com a seleção ( $G_s$ ), acurácia seletiva ( $r_{\hat{g}g}$ ), herdabilidade ( $h^2$ ), variância genética ( $\sigma_g^2$ ) e valor genético ( $g$ ) do indivíduo, clone, linhagem, híbrido, progênie, família, variedade. O ganho genético com a seleção ( $G_s = k r_{\hat{g}g} \sigma_g$ ) mede os avanços genéticos e práticos obtidos com o melhoramento. A acurácia seletiva e a herdabilidade (per si um componente da própria acurácia) permitem

medir  $G_s$  de uma maneira preditiva. O valor genético predito ( $g$ ) e a acurácia estimada são essenciais em catálogos de avaliação genética sobre os quais são realizadas as tomadas de decisões de seleções. Nos dias atuais, o valor genético é um termo genérico que pode ser assim definido: valor genético aditivo (associado a indivíduo e genitor; a serem propagados por via sexuada); valor genotípico (associado a clone, linhagem, híbrido, progênie, família, variedade; a serem propagados identicamente). A acurácia também é útil para comparar métodos de seleção alternativos e será o elemento fundamental do presente documento.

## Maximização da acurácia seletiva no melhoramento genético

A qualidade da avaliação genética deve preferencialmente ser inferida com base na acurácia ( $r_{\hat{g}g}$ ) (Resende, 1995; Resende; Duarte, 2007). Como o objetivo final da avaliação genética é maximizar o ganho genético com a seleção ( $G_s$ ), um exame da fórmula para  $G_s$  (equação do melhorista, de Lush) pode nos dizer qual elemento é mais importante para atingir esse objetivo, olhando para  $G_s = k r_{\hat{g}g} \sigma_g$ , em que  $k$  é o diferencial de seleção padronizado. Desde que  $k$  e  $\sigma_g$  são constantes para uma determinada população, pode-se ver que a acurácia  $r_{\hat{g}g}$  é o único elemento capaz de impulsionar e aprimorar  $G_s$  nessa população. A  $r_{\hat{g}g}$  refere-se à correlação entre o valor genotípico estimado e o verdadeiro de uma cultivar. Essa propriedade foi derivada em conexão com as teorias dos índices de seleção e Blup (Hazel; Lush, 1942; Henderson, 1952, 1975, 1984; Resende, Higa, 1994; Resende, 2002, 2007, 2015; Mrode, 2014; Resende et al., 2014). O nome Blup

foi usado pela primeira vez por Goldberger (1962), na área de econometria e, posteriormente, foi adotado por Henderson (1973). Agora é o procedimento padrão para seleção no contexto dos modelos lineares mistos (LMM), juntamente com a máxima verossimilhança restrita desenvolvida por Patterson e Thompson (1971), e chamado de REML pela primeira vez por Corbeil e Searle (1976). Em LMM,  $r_{\hat{g}g} = (1 - PEV/Var(g))^{1/2}$ , em que  $PEV$  é a variância do erro de predição e  $Var(g) = \sigma_g^2$  é a variância genotípica.

## Origem do termo acurácia e propostas pioneiras para uso no melhoramento genético

A palavra acurácia está associada ao estado de ser ou não preciso ou exato, em conformidade com a verdade. Foi atestada pela primeira vez na década de 1660. Vem da palavra acurado, surgida na década de 1610, e significa feito com cuidado. Origina do latim *accuratus*, que significa preparado com cuidado, exatidão, elaboração. A noção de fazer algo com cuidado levou à noção de ser preciso (década de 1650). É uma palavra mais forte do que correta e mais fraca do que exata. Palavras relacionadas também são acuradamente e acuracidade. Como vem do latim, três das cinco línguas derivadas do latim possuem palavras semelhantes para acurácia, como mostrado abaixo junto com algumas (em cada coluna) palavras estatísticas sinônimas (Tabela 2). Em alemão, acurácia é expressa pela palavra "richtigkeit". Em galego (idioma muito próximo ao português, falado na Galícia, Espanha), é denominada acurácia.

**Tabela 2.** Palavras para designar acurácia em seis idiomas e termos estatísticos sinônimos.

Inglês	Português	Espanhol	Francês	Italiano	Romeno
Accuracy	Acurácia	-	Fidélité	Accuratezza	Acuratetea
Exactness	Exatidão	Exactitud	Justesse	-	-
Rightness	-	-	Légitimité	-	Legitimitate
Correctness	-	-	-	Correttezza	Corectitudine

A introdução do termo acurácia no melhoramento animal foi realizada por Lush et al. (1928) e Lush (1930, 1931, 1935) e incorporada nos livros pioneiros sobre genética quantitativa no melhoramento animal por Lush (1937, 1945, 1948), Bettini (1955) em italiano e Leroy (1960) em alemão. A inclusão do termo acurácia na fórmula de ganho com seleção (equação geral do melhoramento de Lush) foi introduzida por Hazel e Lush (1942). No melhoramento de plantas no Brasil, a introdução do termo acurácia (definida como correlação entre o valor genético verdadeiro e o índice usado para estimá-lo) foi realizada por Resende e Higa (1994), Resende et al. (1990, 1994), Resende (1995) e Resende et al. (1995a). Lush (1937, 1945, 1949) também criou o conceito de herdabilidade em termos da ação aditiva dos alelos.

Além disso, os termos valor genético, valor genotípico e capacidade de transmissão foram introduzidos por Lush (1935). Antes, o valor genético (breeding value) era denominado "breeding worth" (Lush, 1931), e depois foi denominado valor gênico ou valor genético por Lush (1948, 1949). Hoje em dia, os termos sinônimos "breeding value", valor genético e valor genético aditivo permanecem. O termo valor genético foi introduzido no melhoramento de plantas no Brasil por Resende et al. (1990, 1994) e Resende e Higa (1994). Naquela época, os termos valor genético e acurácia não eram aceitos pela comunidade de melhoristas de plantas do País, principalmente nas publicações, pelos revisores de artigos. O termo "breeding value" tem como sinônimos: valor genético aditivo, valor de melhoramento, valor em cruzamento, valor em propagação sexuada, valor em reprodução sexuada, capacidade de transmissão.

A palavra acurácia tem dois conceitos em estatística: a) acurácia das medições: significa reprodutibilidade ou repetibilidade de uma medição ou observação; b) acurácia de uma predição ou estimativa de uma característica em um indivíduo. Alguns dos artigos do Lush, de 1926 a 1930, podem ser descritos como o desenvolvimento de maneiras mais acuradas de medir características quantitativas como um meio de obter fenotipagem de alta precisão. Os artigos de 1931 em diante trataram da acurácia das predições na seleção e na genética quantitativa.

O desenvolvimento teórico da acurácia das medidas ocorreu na década de 1920, principalmente com os trabalhos de Shewhart, Fisher e Lush. O tratamento matemático foi dado por Shewhart (1926), o tratamento estatístico (baseado principalmente na análise de variância) foi dado por Fisher (1925) e o tratamento prático em melhoramento animal foi dado por Lush et al. (1928)

e Lush (1930). Lush fez uso dos conceitos de análise de variância e graus de liberdade de Fisher. O termo graus de liberdade significa o número efetivo de observações na população sobre as quais a estatística em questão é calculada. Ele também usou a fórmula de Shewhart, chamada Lei de Propagação de Erros, dada como:  $\sigma_o = (\sigma_t^2 + \sigma_e^2)^{1/2}$  em que  $\sigma_o$  é o desvio-padrão observado,  $\sigma_t^2$  é o verdadeiro desvio-padrão (erro provável ou esperado, abordado por Bravais (1846)) livre de erros de medição e  $\sigma_e^2$  é o desvio-padrão dos erros de medição (erro experimental). Uma vez que médias baseadas em duas ou mais repetições da medição seriam ainda mais precisas, a fórmula de Shewhart torna-se  $\sigma_o^2 = (\sigma_t^2 + \sigma_e^2/n)$  em que n é o número de repetições da medição feita em cada indivíduo. À medida que n é aumentado,  $\sigma_o$  aproxima-se mais de  $\sigma_t$ . Nenhuma quantidade de repetições da medição pode eliminar inteiramente o erro na medição. Por exemplo, no caso de  $\sigma_o=0.95$ , obtido como  $0.95 = (\sigma_t^2 + 0.79^2/11)^{1/2}$ , onde  $\sigma_t=0.92$ ,  $\sigma_e^2=0.792$  e n = 11 repetições, verifica-se que  $\sigma_o$  tende a  $\sigma_t=0.92$  quando n tende a infinito.

A palavra repetibilidade tem várias conotações em estatística: repetibilidade de uma medição; repetibilidade de uma característica; repetibilidade de uma prova ou avaliação genética, significando o mesmo que confiabilidade, que é dada pelo quadrado da acurácia da predição. Em estatística analítica, três grandezas podem ser especificadas: a observação (dados no vetor y); as estimativas (predições no vetor  $\hat{y}$ ); e o parâmetro ou valores verdadeiros (em um vetor t). A partir dessas quantidades, pode-se definir o seguinte, com base nas correlações: acurácia maximizada de  $\hat{y}$  para prever t (correlação entre  $\hat{y}$  e t); capacidade preditiva de  $\hat{y}$  para prever y (correlação entre  $\hat{y}$  e y); acurácia intrínseca de y para prever t (correlação entre y e t). Esta última correlação é a raiz quadrada da herdabilidade em nível individual em genética quantitativa. Tem-se também a igualdade (correlação entre  $\hat{y}$  e t) = (correlação entre  $\hat{y}$  e y) / (correlação entre y e t).

## Melhoramento de espécies autógamas perenes

Nas espécies em que ocorre a autofecundação natural, o modelo genético difere daquele modelo simplificado (em termos de componentes de médias e de variâncias) usado para as espécies alógamas

e, como consequência, as estratégias de melhoramento para essas espécies podem diferir daquelas estudadas e utilizadas para a maioria das espécies alógamas. Os modelos genéticos e parâmetros associados apresentados a seguir são válidos para espécies diploides e alopoliploides (como o café arábica). Em termos meióticos (citogenéticos e reprodutivos), o comportamento de indivíduos alopoliploides é similar àquele apresentado por indivíduos diploides. Isso ocorre porque os genomas em plantas alopoliploides são muito diferentes, de forma que são formados apenas bivalentes de cromossomos homólogos do mesmo genoma. Assim, em termos de modelos, em genética quantitativa, espécies alopoliploides podem ser tratadas como diploides (Wricke; Weber, 1986).

## Modelos genéticos e componentes de variação genotípica

Aqui são considerados dois modelos genéticos, baseados nos efeitos aditivo e de dominância, ignorando, portanto, a epistasia. O primeiro modelo, proposto por Cockerham (1954), é adequado a situações em que apenas duas linhagens (cultivares) são cruzadas, com objetivo de gerar variabilidade para a seleção. Gillois (1964) e Gallais (1967, 1970a, 1970b) decompueram a variância genética para populações em qualquer nível de endogamia  $F$ , empregando a covariância genética entre indivíduos endógamos (Malecot, 1948, 1966).

O segundo modelo (Cockerham, 1983) é mais geral e apropriado, também, para situações em que várias linhagens são inter cruzadas e a seleção é realizada na população segregante. Nesses modelos genéticos, a variabilidade genética total ( $\sigma_g^2$ ) em uma geração de endogamia equivale:

Modelo 1. (Cockerham, 1954, 1963)

$$\sigma_g^2 = (1 + F)\sigma_a^2 + (1 - F^2)\sigma_d^2 \quad (1)$$

Modelo 2. (Cockerham, 1983)

$$\sigma_g^2 = (1 + F)\sigma_a^2 + (1 - F)\sigma_d^2 + 4FD_1 + FD_2 + F(1 - F)H \quad (2)$$

em que  $F$  é o coeficiente de endogamia da geração de autofecundação considerada, associado ao número de gerações de autofecundação a partir dos cruzamentos iniciais;  $\sigma_a^2$  é a variância genética aditiva;  $\sigma_d^2$  é a variância genética de dominância;  $D_1$  é a covariância entre os efeitos aditivos dos alelos e os efeitos de dominância dos homozigotos;  $D_2$  é a variância dos efeitos de dominância dos homozigotos; e  $H$  é a soma de quadrados da depressão endogâmica de cada loco.

Pelo modelo 2, com  $F = 1$ ,  $\sigma_g^2 = 2\sigma_a^2 + 4D_1 + D_2$ .

É importante relatar que, na situação de cruzamentos entre apenas duas linhagens, em geral considera-se  $p = q = 0,5$  ( $p$  e  $q$  são as frequências dos alelos favoráveis e desfavoráveis na população resultante). Neste caso,  $D_1 = D_2 = 0$  e  $\sigma_a^2 = H$ , de forma que, se for empregada a expressão 2, obtém-se a expressão 1. No Brasil, a abordagem quantitativa aplicada ao melhoramento de espécies autógamas foi realizada em detalhes por Ramalho e Vencovsky (1978), Resende (2002, 2015) e Ramalho et al. (2012).

## Comparação entre processos seletivos

Para comparação entre processos seletivos em espécies autógamas, as variâncias e covariâncias empregadas são dadas pelas seguintes expressões, segundo o modelo 2 (Resende; Medrado, 1994):

a) Variância genética total:

$$C_{ggg} = \sigma_g^2 = (1 + F_g)\sigma_a^2 + (1 - F_g)\sigma_d^2 + 4F_g D_1 + F_g D_2 + F_g(1 - F_g)H.$$

b) Variância genética entre progênies:

$$C_{ggg'} = (1 + F_i)\sigma_a^2 + 2(F_g + F_i) D_1 + \frac{(1 - F_g)^2}{(1 - F_i)}\sigma_d^2 + [F_i + \frac{(F_g - F_i)^2}{2(1 - F_i)}] D_2 + \frac{F_i(1 - F_g)^2}{(1 - F_i)} H.$$

c) Covariância entre a geração de seleção e a população melhorada totalmente endógama: seleção entre progênies – plantio via sementes ou estacas clonais da geração  $\infty$ :

$$C_{tg\infty} = (1 + F_t)\sigma_a^2 + (F_g + 1 + 2F_t) D_1 + [(F_g + F_t)/2] D_2.$$

d) Covariância entre a geração de seleção e a imediatamente posterior: seleção entre progênies – plantio via sementes na geração  $g'$ :

$$C_{ggg'} = (1 + F_i)\sigma_a^2 + (F_g + F_{g'} + 2F_i) D_1 + \frac{(1 - F_g)(1 - F_{g'})}{(1 - F_i)}\sigma_d^2 + [F_i + \frac{(F_g - F_i)(F_{g'} - F_i)}{2(1 - F_i)}] D_2 + \frac{F_i + (1 - F_g)(1 - F_{g'})}{1 - F_i} H$$

e) Variância genética dentro de progênies ( $\sigma_{gd}^2$ ):

$$\sigma_{gd}^2 = C_{ggg} - C_{tgg}.$$

f) Covariância para o ganho com seleção massal na geração  $g$  e plantio da geração  $g'$ :

$$C_{ggg'} = (1 + F_g)\sigma_a^2 + (1 - F_g)\sigma_d^2 + (F_{g'} + 3F_g) D_1 + F_g D_2 + F_g(1 - F_g)H.$$

g) Covariância para o ganho com seleção massal na geração  $g$  e plantio da geração  $\infty$ :

$$C_{gg\infty} = (1 + F_g)\sigma_a^2 + (1 + 3F_g) D_1 + F_g D_2.$$

h) Covariância para o ganho com seleção dentro de progênies na geração  $g$  e plantio na geração  $g'$ :

$$CGD_{g'} = C_{ggg'} - C_{tgg'}.$$

i) Covariância para o ganho com seleção dentro de progênies na geração  $g$  e plantio na geração  $\infty$ :

$$CGD_{\infty} = C_{gg\infty} - C_{tg\infty}.$$

j) Covariância para o ganho com seleção massal na geração  $g$  e plantio clonal de indivíduos da própria geração  $g$ :

$$CMC = \sigma_g^2 = C_{ggg}$$

k) Covariância para o ganho com seleção entre progênies na geração  $g$  e plantio clonal de progênies da mesma geração:

$$CEC = C_{tgg}$$

l) Covariância para o ganho com seleção dentro de progênies na geração  $g$  e plantio clonal de indivíduos na mesma geração:

$$CDC = \sigma_{gd}^2$$

Os índices  $g$ ,  $t$  e  $g'$  referem-se à geração atual, à geração precedente a  $g$  e a qualquer geração posterior a  $g$ , respectivamente.

O melhoramento genético de espécies autógamas anuais e que não possuem ainda técnicas de propagação vegetativa dominadas fundamenta-se no cruzamento entre diferentes linhagens, para a geração de variabilidade genética adicional, e posterior seleção de novas linhagens, após várias gerações de autofecundação. Assim, as cultivares obtidas sempre representam genótipos totalmente homocigotos e, portanto, não apresentam qualquer variabilidade dentro delas.

As estratégias de melhoramento genético das espécies autógamas perenes deverão diferir das estratégias empregadas no melhoramento de espécies autógamas anuais. Isso ocorre em razão da possibilidade de fixação de genótipos superiores nas autógamas perenes em qualquer geração, por meio

da propagação vegetativa destes, ou propagação via sementes utilizando-se dos genitores permanentes. Em espécies autógamas anuais (que geralmente não permitem a propagação vegetativa), obrigatoriamente, deve-se caminhar à homocigose total, como única forma de fixação de genótipos superiores.

Em espécies autógamas perenes, os métodos ou estratégias de melhoramento podem ser comparados em função de dois fatores principais: a) diferentes estratégias de propagação (vegetativa em qualquer geração de autofecundação; via sementes em qualquer geração de autofecundação; via sementes na situação de linhagens totalmente endógamas) e b) geração ideal para seleção. Logicamente, essa comparação deve basear-se nos progressos genéticos esperados com seleção. Fixando a geração de seleção, os métodos de propagação podem ser comparados pela expressão:

$$G_s = k \frac{Cov(US, PM)}{(Var(US))^{1/2}}$$

em que  $k$  é o diferencial de seleção padronizado,  $Cov(US, PM)$  é a covariância entre a unidade de seleção e a população melhorada e  $Var(US)$  é a variância da unidade de seleção. Nessa situação (para uma geração de autofecundação específica),  $k$  e  $Var(US)$  são constantes para os diferentes métodos de propagação. Dessa forma, as comparações podem ser baseadas apenas em  $Cov(US, PM)$ . Essas covariâncias derivadas para os métodos de seleção massal, entre progênies e dentro de progênies (Resende; Medrado, 1994), são apresentadas nas Tabelas 3 e 4.

**Tabela 3.** Covariâncias\* unidade de seleção – população melhorada para os métodos: seleção massal (M), seleção entre progênies (E) e seleção dentro de progênies (D), associadas aos métodos de propagação via sementes na própria geração de plantio ( $g'$  = uma geração após a seleção) e na situação de linhagens totalmente endógamas ( $\infty$ ).

Geração de seleção	F**	M $g'$	M $\infty$	E $g'$	E $\infty$	D $g'$	D $\infty$
F <sub>2</sub>	0	$\sigma_a^2 + (1/2) \sigma_d^2$	$\sigma_a^2$	—	—	—	—
F <sub>3</sub>	1/2	$(3/2) \sigma_a^2 + (3/8) \sigma_d^2$	$(3/2) \sigma_a^2$	$\sigma_a^2 + (1/8) \sigma_d^2$	$\sigma_a^2$	$(1/2) \sigma_a^2 + (1/4) \sigma_d^2$	$(1/2) \sigma_a^2$
F <sub>4</sub>	3/4	$(7/4) \sigma_a^2 + (7/32) \sigma_d^2$	$(7/4) \sigma_a^2$	$(3/2) \sigma_a^2 + (3/32) \sigma_d^2$	$(2/3) \sigma_a^2$	$(1/4) \sigma_a^2 + (1/8) \sigma_d^2$	$(1/4) \sigma_a^2$
F <sub>5</sub>	7/8	$(15/8) \sigma_a^2 + (15/128) \sigma_d^2$	$(15/8) \sigma_a^2$	$(7/4) \sigma_a^2 + (7/128) \sigma_d^2$	$(7/4) \sigma_a^2$	$(1/8) \sigma_a^2 + (1/16) \sigma_d^2$	$(1/8) \sigma_a^2$
F $\infty$	1	—	—	—	$2 \sigma_a^2$	—	—

\* Consideraram-se cruzamentos entre apenas duas linhagens genitoras e grande número de indivíduos por progênie nos experimentos.

\*\* Coeficiente de endogamia associado à geração de seleção.

**Tabela 4.** Covariâncias\* unidade de seleção – população melhorada para os métodos de seleção: massal (M), entre progênies (E) e dentro de progênies (D), associadas à propagação vegetativa na geração de seleção.

Geração de seleção	F**	M	E	D
F <sub>2</sub>	0	$\sigma_a^2 + \sigma_d^2$	—	—
F <sub>3</sub>	1/2	$(3/2) \sigma_a^2 + (3/4) \sigma_d^2$	$\sigma_a^2 + (1/4) \sigma_d^2$	$(1/2) \sigma_a^2 + (1/2) \sigma_d^2$
F <sub>4</sub>	3/4	$(7/4) \sigma_a^2 + (7/16) \sigma_d^2$	$(3/2) \sigma_a^2 + (3/16) \sigma_d^2$	$(1/4) \sigma_a^2 + (1/4) \sigma_d^2$
F <sub>5</sub>	7/8	$(15/8) \sigma_a^2 + (15/64) \sigma_d^2$	$(7/4) \sigma_a^2 + (7/64) \sigma_d^2$	$(1/8) \sigma_a^2 + (1/8) \sigma_d^2$
F <sub>∞</sub>	1	—	$2 \sigma_a^2$	—

\* Foram considerados os cruzamentos entre apenas duas linhagens genitoras e grande número de indivíduos por progênie nos experimentos.

\*\* Coeficiente de endogamia associado à geração de seleção.

Fonte: Resende (2002).

Observa-se com base nas covariâncias que, de maneira geral, a melhor estratégia de propagação é a clonal, seguida daquela via sementes com plantio da geração g' e, por último, a propagação via sementes com plantio das linhagens totalmente endógamas (Tabelas 3 e 4). Logicamente, com a seleção realizada na geração infinita não existem diferenças entre os sistemas de propagação. Fica claro, então, que embora a espécie seja autógama, não compensa obter novas linhagens totalmente homocigotas, exceto se ficar definido F<sub>∞</sub> como a melhor geração para seleção.

É importante ressaltar que as inferências realizadas são válidas também para o método do índice multifeitos, cujo valor de  $Cov(US, PM)$  é função desses valores para os métodos da seleção entre progênies e dentro de progênies.

A definição da geração ideal de seleção é dependente da herança do caráter considerado, principalmente quanto ao grau médio de dominância. Considerando, inicialmente, os graus médios de dominância (g.m.d.) inerentes à dominância completa (g.m.d. = 1) e sobredominância (g.m.d. > 1), pode-se afirmar que a seleção deve ser praticada na geração F<sub>1</sub>, pois com seleção em gerações de autofecundação (F<sub>2</sub> por exemplo), o máximo que se consegue (selecionando o melhor indivíduo) é um indivíduo com valor genético/genotípico idêntico ao de F<sub>1</sub>. Assim, para casos de sobredominância e dominância completa, a estratégia de melhoramento deve restringir-se à realização de cruzamentos dialélicos ou fatoriais, entre genitores selecionados, com posterior seleção e propagação clonal do melhor híbrido. Entretanto, para geração de variabilidade genética adicional para futuras seleções, um esquema de seleção recorrente é recomendável, preferencialmente a seleção recorrente recíproca (Resende; Higa, 1990; Ramalho et al., 1999).

Com dominância parcial e ausência de dominância (herança totalmente aditiva), existe certa superioridade de genótipos homocigotos sobre heterocigotos e a condução de gerações de autofecundação pode ser uma boa opção. Nessa situação, a definição da geração ideal de seleção deve basear-se na estimação de ganhos genéticos com seleção. Com herança completamente aditiva, não existe depressão endogâmica, e as médias das diferentes gerações não se alteram. Considerando que a variância fenotípica pouco varia entre as gerações, depreende-se que as covariâncias unidades de seleção – população melhorada (Tabelas 3 e 4) são preponderantes na comparação dos ganhos. É imperativo dizer que, nesse caso, as covariâncias apresentadas são função apenas da variância genética aditiva (substituindo  $\sigma_d^2$  por 0 nas Tabelas 3 e 4) e, portanto, não existem diferenças quanto aos três sistemas de propagação. Verifica-se, assim, que nesse caso é vantajoso trabalhar com linhagens completamente endogâmicas e selecionar em F<sub>∞</sub>, pois  $Cov(US, PM) = 2 \sigma_a^2$  contra  $Cov(US, PM) = \sigma_a^2$  da geração F<sub>2</sub>.

Se existir algum nível de dominância alélica, existe depressão endogâmica e, portanto, as médias das diferentes gerações não serão as mesmas. Assim, para o caso de dominância parcial, a determinação da geração ideal de seleção deverá levar em conta também as diferentes médias das gerações. A média de cada geração de autofecundação é dada por  $M_t = M_o - 2F\Sigma p q d$  (Falconer, 1989) em que  $d_g$  é o valor genotípico do heterocigoto,  $M_o$  a média original e  $\Sigma$  indica somatório sobre todos os locos. Assim, a redução na média é diretamente proporcional ao coeficiente de endogamia, dependendo também das frequências alélicas, do nível de dominância e do número de locos controlando o caráter. Com dominância parcial,  $d_g$  varia no intervalo



aberto de zero a um, de forma que a depressão será mais baixa e alta para valores de  $d_g$  próximos de zero e um, respectivamente.

Analisando inicialmente as covariâncias e sabendo que para o grau médio de dominância 1,0 e  $p = q = 0,5$ ,  $\sigma_d^2 = 1/2 \sigma_a^2$  (Vencovsky; Barriga, 1992), verifica-se que elas aumentam com as gerações de autofecundação (Tabelas 3 e 4). Como as médias populacionais decrescem com as autofecundações, não fica claro, para casos de dominância parcial, qual a melhor geração para se efetuar a seleção, visando à propagação vegetativa ou via sementes com fixação dos genótipos na geração de seleção. Com dominância baixa, a geração ideal de seleção tenderá para  $F_\infty$ , conforme verificado para os casos de herança totalmente aditiva. Com dominância alta, a geração ideal de seleção tenderá para  $F_1$  ou  $F_2$  ( $S_0$ ), segundo o verificado nos casos de dominância completa ou sobredominância. Logicamente, com dominância intermediária, uma geração com  $F$  intermediário ( $F_3, F_4$ ) deverá ser a ideal.

Torna-se necessário, então, um conhecimento prévio da herança do caráter, para determinar com segurança a geração ideal para seleção. Em café arábica (Cilas et al., 1998), foi mostrada a existência de heterose para vários caracteres de interesse, o que é um indício da existência de dominância alélica. Assim sendo, pode-se dizer que, para o melhoramento desses caracteres, dificilmente será vantajoso retardar a seleção para gerações avançadas. Pelo contrário, realizar a seleção logo após a hibridação ( $F_1$ ) ou em  $F_2$ , associada à propagação vegetativa, deverá ser a estratégia mais eficiente. Em pessegueiro, a estratégia de hibridação seguida de seleção e propagação vegetativa já vem sendo usada rotineiramente (Bruckner, 1999).

## Estatísticas $F$ hierárquicas de Wright no melhoramento de populações autógamias

O presente tópico visa mostrar os fundamentos de genética de populações associados a um método de seleção que abarca toda a estrutura de um programa de melhoramento de plantas autógamias, incluindo todos os dados desde os pais, a geração  $F_1$  e  $F_2$  e, também, tanto os efeitos de progênie quanto de populações em todas as gerações endogâmicas a partir de  $F_3$ .

Suponha uma população global de linhagens de uma espécie. O cruzamento de algumas dessas linhagens aos pares gera populações  $F_1$  que serão conduzidas até  $F_\infty$ . Ocorre então a fragmentação

da população global em subpopulações, e essa diferenciação produz um coeficiente de endogamia em razão da subdivisão, denominado  $F_{ST}$  (Wright, 1943). A variação genética entre essas populações pode então ser expressa como função do  $F_{ST}$ . Esse coeficiente permite então desenvolver um índice de seleção com a inclusão do efeito de populações em todas as gerações desde a  $F_1$  até a  $F_\infty$ , por meio do uso de covariâncias entre gerações que fazem uso do  $F_{ST}$  e da herdabilidade total entre e dentro de populações. Esses três parâmetros podem ser estimados na geração  $F_3$  em um esquema hierárquico de progênie dentro de populações, conforme mostrado mais adiante.

Nessa situação conceitual, as diferentes populações podem ser vistas como fragmentos de uma mesma população global da espécie. Seleção ou algum outro fator de deriva conduzirá à diferenciação entre subpopulações. Essa estruturação pode ser vista em termos das estatísticas  $F$  de Wright. Populações que não se enquadram na definição teórica de panmíticas (grandes e sob cruzamentos aleatórios) são ditas populações estruturadas. A estruturação pode se dar através de subpopulações com limitada migração e fluxo gênico entre elas até subpopulações totalmente isoladas. A subdivisão das populações provoca um excesso de homozigose, ou seja, um efeito de endogamia em relação à espécie como um todo. Wright (1951) introduziu os parâmetros (denominados "estatísticas  $F$  hierárquicas")  $F_{ST}$ ,  $F_{IT}$  e  $F_{IS}$  para descrever estrutura genética, em que  $F_{ST}$  é a correlação de alelos entre indivíduos dentro de subpopulações. Refere-se, portanto, ao próprio  $F$  ou coeficiente de endocruzamento (de Wright) dentro das populações, provocado pela própria deriva genética associada à subdivisão. Percebe-se que existe uma associação entre  $F_{ST}$  e o coeficiente de coancestria entre indivíduos (Resende, 2002).

$F_{IT}$  é a correlação entre alelos dentro de indivíduos. Refere-se ao coeficiente de endocruzamento de indivíduos no contexto da população global, entre e dentro de subpopulações.

$F_{IS}$  é a correlação entre alelos dentro de indivíduos dentro de subpopulações. Refere-se a um coeficiente de endocruzamento de um indivíduo por causa de cruzamentos não aleatórios dentro de populações. Tem-se também  $F_{IS} = (F_{IT} - F_{ST}) / (1 - F_{ST})$ .

O coeficiente  $F_{IT}$  resulta da endogamia (cruzamentos não aleatórios) dentro da subpopulação ( $F_{IS}$ ) e da endogamia em razão da subdivisão ( $F_{ST}$ ) (Ollivier 1981, 2002). Tem-se que  $F_{IT} = F_{ST} + (1 - F_{ST}) F_{IS}$ . Essa expressão é fundamental no desenvolvimento teórico que se fará no desenvolvimento do novo índice de seleção. Se existir panmixia dentro

de subpopulações,  $F_{IS} = 0$  e, portanto,  $F_{IT} = F_{ST}$ . Se todas as subpopulações forem geneticamente idênticas,  $F_{ST} = 0$  e, portanto,  $F_{IS} = F_{IT}$ .

A subdivisão, diferenciação entre subpopulações ou estratificação da espécie como um todo, pode ser causada por deriva aleatória e/ou por seleção. A variação genética total ( $\sigma_{a0}^2$ ) exibida se todos os indivíduos fizessem parte de uma única população panmítica é reduzida dentro de subpopulações ( $\sigma_a^2$ ) com o processo de diferenciação entre subpopulações. Essa redução é ditada pela fração  $(1 - F_{ST}) \sigma_{a0}^2$ , produzindo  $\sigma_a^2 = (1 - F_{ST}) \sigma_{a0}^2$ .

Redução similar a  $\sigma_a^2 = (1 - F_{ST}) \sigma_{a0}^2$  ocorre por causa de cruzamentos preferenciais entre genitores selecionados por pressão seletiva natural, fato que conduz a maior correlação intraclasse ( $r_{pp}$ ) entre os descendentes e conseqüente redução na variabilidade da população. A seleção reduz a variação genética pela criação de desequilíbrio alélico (de Hardy-Weinberg e de ligação) em fase gamética. A seleção de genitores reduz a variação entre eles e, conseqüentemente, reduz a covariância entre os indivíduos da progênie, de forma que a herdabilidade calculada é subestimada (Resende, 2002). Então, pode ser escrito também que  $\sigma_a^2 = (1 - r_{pp}) \sigma_{a0}^2$ . A quantidade  $r_{pp}$  é causada por desequilíbrio alélico em fase gamética (D), dado por  $D = \sum_{i=1}^n (p_i - q_i)^2 / (4p_i q_i)$  (Burrows, 1971), em que  $p_i$  e  $q_i$  são as frequências dos alelos favoráveis no loco  $i$ , nas duas populações, respectivamente. Assim, assumindo que a diferenciação genética intraespecífica é por causa das diferenças em frequências alélicas, pode-se assumir também que  $\sigma_a^2 = (1 - D) \sigma_{a0}^2$ . Então, tendo-se estimado  $\sigma_a^2$  via componentes de variância e  $\sigma_{a0}^2$  via  $F_{ST}$ , pode-se estimar D via  $D = (\sigma_{a0}^2 - \sigma_a^2) / \sigma_{a0}^2 = 1 - \sigma_a^2 / \sigma_{a0}^2 = F_{ST}$ . Dessa forma, D fornece o grau de diferenciação ou divergência genética entre as populações, e maiores valores de D indicam maior divergência. A variância genética aditiva é relacionada ao desequilíbrio

alélico via  $\sigma_{a0}^2 = \sum_{i=1}^n 2p_i q_i (1 - D) \alpha_i^2$ , em que  $\alpha_i^2$  é o efeito de substituição alélica no loco  $i$ . Como  $D = F_{ST}$ , a diferença entre ambos refere-se apenas às suas assumidas causas, seleção ou deriva, respectivamente.

Considerando que há equilíbrio nas subpopulações,  $F_{ST} \cong 1 / (1 + 4N_e m)$ , em que  $m$  é a taxa de migração em cada subpopulação em cada geração, e  $N_e$  é o tamanho efetivo com valor igual para todas as subpopulações. Tendo-se estimado  $F_{ST}$ , pode-se estimar  $N_e m$  por  $N_e m = (1 - F_{ST}) / (4F_{ST})$ , o qual, assumindo valor  $< 1$ , significa que a deriva conduzirá a substancial diferenciação. Essa abordagem de  $F_{ST}$  fornece um outro estimador para a variância das frequências alélicas, ou seja,  $\sigma_p^2 = [\bar{p}(1 - \bar{p})] / (1 + 4N_e m)$ , mostrando que a variação entre subpopulações diminui à medida que  $N_e m$  aumenta.

A contribuição de um loco para a variância dentro de população é  $2pq\alpha^2$ . A contribuição média dentro de população é  $2[\bar{p}(1 - \bar{p}) - \sigma_p^2] \alpha^2$ . Mas a variância das frequências alélicas entre populações é dada por  $\sigma_p^2 = \bar{p}(1 - \bar{p}) F_{ST}$  e a variância aditiva média dentro de população é  $\sigma_a^2 = (1 - F_{ST}) \sum 2[\bar{p}(1 - \bar{p})] \alpha^2$ . A contribuição do efeito de substituição alélica para a média da população é  $2p\alpha$  e tem variância  $4\sigma_p^2 \alpha^2 = 4F_{ST} [\bar{p}(1 - \bar{p})] \alpha^2$ . Assim, a variância entre populações equivale a  $\sigma_{pop}^2 = 2F_{ST} \sum [\bar{p}(1 - \bar{p}) \alpha^2]$ . E a variância total, entre ( $\sigma_{pop}^2$ ) e dentro de populações, tem valor  $\sigma_{pop}^2 + \sigma_a^2 = (1 + F_{ST}) \sum [\bar{p}(1 - \bar{p}) \alpha^2]$ . Pode-se notar que o  $F_{ST}$  mede o aumento proporcional na variação da população total resultante da subdivisão em linhas endogâmicas (Wright, 1943). Em resumo, tem-se a seguinte distribuição da variabilidade entre e dentro de subpopulações, em que se percebe que  $F_{ST}$  é dado pela relação entre a variância entre subpopulações sobre a soma dela com duas vezes a variância dentro de populações, ou seja,  $\frac{2F_{ST} \sigma_{a0}^2}{2F_{ST} \sigma_{a0}^2 + 2(1 - F_{ST}) \sigma_{a0}^2} = \frac{2F_{ST} \sigma_{a0}^2}{2\sigma_{a0}^2} = F_{ST}$ .

Componente	Composição
Dentro de população	$(1 - F_{ST}) \sigma_{a0}^2$
Entre populações	$2F_{ST} \sigma_{a0}^2$
Total = entre + dentro	$(1 + F_{ST}) \sigma_{a0}^2$

Substituindo-se  $\sigma_p^2 = \bar{p}(1 - \bar{p}) F_{ST}$  em  $2[\bar{p}(1 - \bar{p}) - \sigma_p^2] \alpha^2$ , tem-se  $\sigma_a^2 = 2[\bar{p}(1 - \bar{p})] \alpha^2 - F_{ST} 2[\bar{p}(1 - \bar{p})] \alpha^2 = (1 - F_{ST}) 2[\bar{p}(1 - \bar{p})] \alpha^2$ .

Como tanto  $F_{ST}$  quanto  $D$  estimam o grau de diferenciação, substituindo  $F_{ST}$  por  $D$ , tem-se  $\sigma_a^2 = (1-D)2[\bar{p}(1-\bar{p})]\alpha^2$ . Também, de  $\sigma_p^2 = \bar{p}(1-\bar{p})F_{ST}$  tem-se que  $F_{ST} = \frac{\sigma_p^2}{\bar{p}(1-\bar{p})}$ . Esse é um estimador de  $F_{ST}$  via marcadores moleculares. O  $F_{ST}$  é também denominado  $Q_{ST}$ . Sob locos de efeitos aditivos e neutrais à seleção,  $Q_{ST}$  (calculado via dados fenotípicos) e  $F_{ST}$  (calculado via dados moleculares) se equivalem.

Existem duas formas de cruzamentos não aleatórios que produzem desequilíbrio e subdividem parcialmente uma população: seleção de genitores e cruzamentos preferenciais. Esses dois fatores produzem dois tipos de covariância ou correlação entre efeitos genéticos: correlação entre alelos de diferentes locos em um gameta (desequilíbrio em fase gamética ou de ligação); correlação entre alelos em pares de gametas que se unem (endogamia). A seleção de genitores gera desequilíbrio em fase gamética. Os cruzamentos

preferenciais geram ambos, endogamia e desequilíbrio em fase gamética. O desequilíbrio em fase gamética de locos não ligados é menos persistente do que o desequilíbrio em fase gamética de locos ligados (Falconer, 1989). Assim, parte da variação genética entre populações é variância em razão do desequilíbrio de ligação, a qual pode não persistir em longo prazo, especialmente em espécies alógamas.

## Esquema hierárquico para a estimação de componentes de variância

O  $F_{ST}$  pode ser estimado também via dados fenotípicos, conforme mostrado a seguir. Segundo o modelo  $Y_{ijk} = \mu + pop_i + b_j + (f/p)_{k(i)} + e$ , a análise de variância para um experimento envolvendo  $f$  progênes (f/p)  $F_{2:3}$  dentro de população (pop), avaliadas em  $b$  repetições, equivale a:

Fontes de variação	GL	QM	E(QM)
Blocos	b-1	$Q_b$	-
Populações	g-1	$Q_p$	$\sigma_e^2 + b\sigma_{f/p}^2 + bf\sigma_{pop}^2$
Progênes/Populações	(s-1)g	$Q_{fp}$	$\sigma_e^2 + b\sigma_{f/p}^2$
Resíduo	(b-1)(qs-1)	$Q_e$	$\sigma_e^2$

Associando a teoria da subdivisão com a estimação dos componentes de variância no modelo hierárquico, tem-se o seguinte quadro-resumo:

Entidade	Composição da variância	Estimador do componente de variância
Dentro de populações	$\sigma_a^2 = (1 - F_{ST})\sigma_{a0}^2$	$\sigma_a^2 = \sigma_{f/p}^2$
Entre populações	$\sigma_{pop}^2 = 2F_{ST}\sigma_{a0}^2$	$\sigma_{pop}^2$
Total = Entre + Dentro	$\sigma_{aT}^2 = (1 + F_{ST})\sigma_{a0}^2$	$\sigma_{aT}^2 = \sigma_{pop}^2 + \sigma_a^2$
Entre + 2 Dentro	$2\sigma_{a0}^2$	$\sigma_{pop}^2 + 2\sigma_a^2$

Proporção entre variâncias	Composição	Estimador
Entre / (Entre + 2 Dentro)	$\frac{2F_{ST}\sigma_{a0}^2}{2\sigma_{a0}^2} = F_{ST}$	$F_{ST} = \frac{\sigma_{pop}^2}{\sigma_{pop}^2 + 2\sigma_a^2} = Q_{ST}$
Entre/Total	$\rho_p = \frac{2F_{ST}}{1 + F_{ST}}$	$\rho_p = \frac{\sigma_p^2}{\sigma_{pop}^2 + \sigma_a^2}$
Herdabilidade dentro de população	-	$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2}$
Herdabilidade entre populações	-	$h_{pop}^2 = \frac{\sigma_{pop}^2}{\sigma_{pop}^2 + \sigma_a^2 + \sigma_e^2}$
Herdabilidade total	-	$h_{aT}^2 = \frac{\sigma_{pop}^2 + \sigma_a^2}{\sigma_{pop}^2 + \sigma_a^2 + \sigma_e^2}$

$2\sigma_{a0}^2$  é a variância genética aditiva original da espécie.

Considerando progênies  $F_{2:3}$ , os seguintes parâmetros podem ser estimados:

$\sigma_{pop}^2$ : variação genética entre populações.

$\sigma_a^2 = \sigma_{f/p}^2$ : variação genética aditiva média dentro de populações.

$F_{ST} = \frac{\sigma_{pop}^2}{\sigma_{pop}^2 + 2\sigma_a^2}$ : coeficiente de endogamia por causa da diferenciação entre populações.

$\sigma_{a0}^2 = \frac{\sigma_{pop}^2}{2F_{ST}} = \frac{\sigma_{pop}^2 + \sigma_a^2}{1 + F_{ST}} = \frac{\sigma_{aT}^2}{1 + F_{ST}}$ : variação genética aditiva total da população panmítica original.

$\sigma_{aT}^2 = \sigma_p^2 + \sigma_a^2 = (1 + F_{ST})\sigma_{a0}^2$ : variação genética aditiva total entre e dentro de populações.

$\sigma_y^2 = \sigma_{pop}^2 + \sigma_a^2 + \sigma_e^2 = (1 + F_{ST})\sigma_{a0}^2 + \sigma_e^2$ : variância fenotípica total entre e dentro de populações.

$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2}$ : herdabilidade individual dentro de população.

$h_{aT}^2 = \frac{\sigma_{pop}^2 + \sigma_a^2}{\sigma_{pop}^2 + \sigma_a^2 + \sigma_e^2} = \frac{(1 + F_{ST})\sigma_{a0}^2}{(1 + F_{ST})\sigma_{a0}^2 + \sigma_e^2}$ : herdabilidade total entre e dentro de população.

$h_{pop}^2 = \frac{\sigma_{pop}^2}{\sigma_y^2} = \frac{\rho_p \sigma_{aT}^2}{\sigma_y^2} = \rho_p h_{aT}^2 = \frac{2F_{ST}\sigma_{a0}^2}{(1 + F_{ST})\sigma_{a0}^2 + \sigma_e^2}$ :

herdabilidade entre populações.

$h_a^{2*} = h_{aT}^2 - \rho_p h_{aT}^2 = \frac{\sigma_a^2}{\sigma_{pop}^2 + \sigma_a^2 + \sigma_e^2}$ : herdabilidade modificada.

$\rho_p = \frac{\sigma_{pop}^2}{\sigma_{pop}^2 + \sigma_a^2} = \frac{h_{pop}^2}{h_{aT}^2} = \frac{2F_{ST}}{(1 + F_{ST})}$ : proporção das

variâncias entre populações e total.

O  $F_{ST}$  é também dado por:

$$F_{ST} = \frac{\sigma_{pop}^2}{\sigma_{pop}^2 + 2\sigma_a^2} = \frac{\rho_p}{\rho_p + 2(1 - \rho_p)} = \frac{\rho_p}{2 - \rho_p}$$

O  $F_{ST}$ , usando dados fenotípicos, foi calculado pela primeira vez por Wright (1943), e Lande (1992) o considerou sob uma base mais teórica. Para cálculo do  $F_{ST}$ , é necessária uma estimativa da variância aditiva dentro de populações. E isso só pode ser feito a partir da geração  $F_3$ . Mas a variância entre populações pode ser feita em  $F_2$ ,  $F_3$  ou qualquer outra geração.

Considerando simultaneamente as endogamias entre populações ( $F_{ST}$ ) e dentro de populações ( $F_{IS}$ ), tem-se a seguinte configuração para os casos de  $F_{IS} = 0$  e  $F_{IS} = 1$ .

Entidade	Componente	Composição	$F_{IS} = 1$	$F_{IS} = 0$
Dentro de população	Dentro de progênes	$(1 - F_{IS})(1 - F_{ST})\sigma_{a0}^2$	0	$(1 - F_{ST})\sigma_{a0}^2$
	Entre progênes	$2F_{IS}(1 - F_{ST})\sigma_{a0}^2$	$2(1 - F_{ST})\sigma_{a0}^2$	0
	Total = Entre + Dentro Progênes	$(1 + F_{IS})(1 - F_{ST})\sigma_{a0}^2$	$2\sigma_{a0}^2 - 2F_{ST}\sigma_{a0}^2$	$(1 - F_{ST})\sigma_{a0}^2$
Entre populações	-	$2F_{ST}\sigma_{a0}^2$	$2F_{ST}\sigma_{a0}^2$	$2F_{ST}\sigma_{a0}^2$
Total	Entre + Dentro populações	$(1 + F_{IT})\sigma_{a0}^2$	$2\sigma_{a0}^2$	$\sigma_{a0}^2 + F_{ST}\sigma_{a0}^2$

$$F_{IT} = F_{ST} + (1 - F_{ST})F_{IS}$$

Esse quadro mostra, de forma elegante, a estrutura hierárquica em três níveis, entre populações e entre progênes dentro de populações e dentro de progênes, da população global de melhoramento. Partindo-se da variabilidade total entre genitores ( $2\sigma_{a0}^2$ ), após a estruturação, recompõe-se ao final ( $F_{IS} = 1$ ), sem seleção, a mesma variabilidade original  $2\sigma_{a0}^2$ .

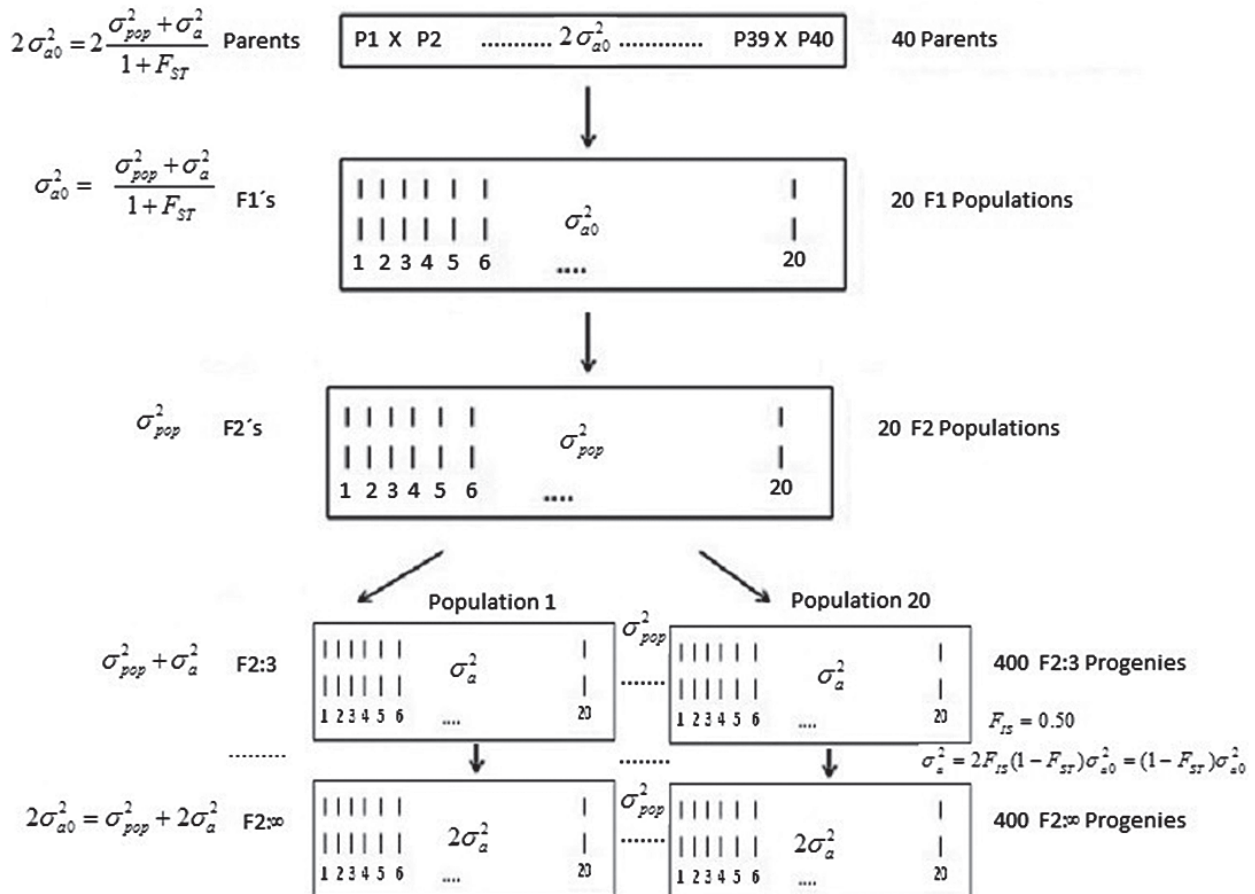
A Figura 2 ilustra a estrutura global do programa de melhoramento de uma espécie autógama, mostrando a distribuição da variabilidade genética total a ser aproveitada por um índice total entre e dentro

de populações e multigerações, derivado por Resende et al. (2015, 2016).

O índice total é dado por:

$$\hat{a} = \hat{b}_{pai1}P_1 + \hat{b}_{pai2}P_2 + \hat{b}_{F1}F_1 + \hat{b}_{F2}F_2 + \hat{b}_{F3pop}F_{3pop} + \hat{b}_{F4pop}F_{4pop} + \hat{b}_{F5pop}F_{5pop} + \hat{b}_{F6pop}F_{6pop} + \hat{b}_{F3prog}F_{3prog} + \hat{b}_{F4prog}F_{4prog} + \hat{b}_{F5prog}F_{5prog} + \hat{b}_{F6prog}F_{6prog}$$

o qual estima o valor genético aditivo de uma linhagem na geração  $F_{\infty}$ , considerando todas as informações fenotípicas corrigidas desde as gerações parentais (P1 e P2) até a geração  $F_6$ .



**Figura 2.** Estrutura global do programa de melhoramento de uma espécie autógama, mostrando a distribuição da variabilidade genética total a ser aproveitada por um índice total entre e dentro de populações e multigerações.

## Métodos ótimos de seleção em café arábica

### Índice matriz-progênie-indivíduo para seleção inicial

O melhoramento do café arábica baseia-se na avaliação de pelo menos três entidades genéticas: a planta matriz, a progênie dessa planta e os indivíduos dentro dessa progênie (Carvalho, 1952; Carvalho et al., 1959). O uso dessas três fontes de informação sempre foi objeto de muitos debates em termos de quais das três informações deveriam ser priorizadas nos programas de melhoramento do IAC, coordenados pelo Dr. Alcides Carvalho. Neste documento, essa questão é abordada em termos biométricos, com a ajuda da genética quantitativa e da estatística. São apresentados estimadores de valores genéticos e acurácias de vários métodos de seleção envolvendo essas três entidades. Dessa forma, procura-se indicar o melhor método, comparados via magnitudes da acurácia seletiva. Assim, otimiza-se o uso de informações de progênies em cada geração de autofecundação e das plantas matrizes da geração anterior.

Foi iniciado no IAC, em 1932, um extenso projeto de melhoramento do cafeeiro, compreendendo o estudo pormenorizado de progênies de cafeeiros selecionadas e de híbridos entre essas plantas matrizes. Carvalho et al. (1959) verificaram que a seleção de matrizes em plantios comerciais do café Bourbon Vermelho gerou progênies uniformes, e, portanto, a seleção subsequente teve que ser baseada na produção média das progênies, já que a seleção dentro de progênies não conduziria a nenhum progresso. Porém, no café Mundo Novo, a seleção entre progênies e dentro delas foi bastante efetiva. O estudo comparativo de todos os tratamentos mostrou a existência de uma razoável correlação entre o bom aspecto vegetativo (vigor vegetativo, diâmetro e altura) e a produtividade das progênies.

Essa variabilidade culminou, em longo prazo, com a obtenção da cultivar Catuaí (até hoje uma das melhores cultivares de café arábica do mundo), em geração F4 do cruzamento entre as cultivares Caturra (mutação caturra para pequeno porte: descoberta em Manhumirim, MG, Zona da Mata Mineira), e Mundo Novo. Essa criação da cultivar foi realizada da seguinte maneira, para 'Caturra Vermelho IAC 477' e 'caturra amarelo IAC

476'. Estas referem-se a duas cultivares de porte baixo, provavelmente originadas de uma ou duas mutações naturais de 'Bourbon Vermelho', de porte alto. A forma amarela pode também ter origem numa mutação natural do próprio 'Caturra Vermelho'. Ambas as variantes foram encontradas na Serra do Caparaó, na divisa dos estados de Minas Gerais e Espírito Santo. Sementes de plantas com frutos vermelhos ou amarelos foram introduzidas no IAC, provenientes do município de Siqueira Campos, ES, em 1937, onde foram selecionadas e liberadas a partir de 1949 e registradas no RNC em 1999, com as siglas Caturra Vermelho IAC 477 e Caturra Amarelo IAC 476 (Carvalho et al., 2007).

### Método 1: seleção de matrizes com base em várias colheitas

#### Escolha das matrizes

A escolha das árvores-mãe é o primeiro passo para iniciar um programa de melhoramento do cafeeiro. Depois de uma inspeção visual crítica, várias árvores-mãe elegíveis são selecionadas do material inicial e colocadas sob observação. Essas candidatas devem combinar tantas propriedades boas quanto possível. Além da capacidade de produção, fatores como regularidade ou persistência de produção (capacidade de permanecer entre as melhores em todas as colheitas), resistência, tolerância ou resiliência a doenças e pragas, forma e tamanho dos grãos e outros. Em um estágio mais avançado, as qualidades organolépticas também devem ser avaliadas. A produtividade relativa deve ser enfatizada mais do que a produtividade absoluta, ou seja, o rendimento de uma árvore-mãe é melhor comparado com o de um grupo de árvores vizinhas da mesma idade crescendo em condições semelhantes e expresso como uma porcentagem desta última (Ferreira, 1969), por um processo de seleção massal estratificada. A persistência pode ser avaliada como a diferença ou distância em relação ao melhor material em cada colheita. Pode ser avaliada também como proporção em relação ao melhor material, por meio da performance relativa (PR) (Resende, 2004, 2007). Nesse caso, a PR é uma medida de adaptabilidade temporal ao longo das colheitas.

Em vista da natureza autógama do café arábica, o sistema mais comum de melhoramento tem sido, na maioria dos países produtores de arábica, algum tipo de seleção de linhas puras. Além disso, o cruzamento artificial entre tipos de genitores selecionados está sendo aplicado a fim de produzir

uma combinação de certas características desejáveis. O ponto de partida de qualquer programa de melhoramento é a escolha das árvores-mãe dentro das melhores cultivares disponíveis. Testes com o propósito de descobrir as cultivares mais adaptadas foram feitos em larga escala em muitas regiões cafeeiras do mundo, por exemplo, no Brasil, Colômbia, El Salvador, Costa Rica, Tanzânia e outros países (Carvalho; Mônaco, 1969a, 1969b). Após a escolha das árvores-mãe, a etapa seguinte é o teste de suas progênies.

A seleção de árvores-mãe de destaque pode ser realizada em cafezais comerciais das cultivares mais adaptadas ou em parcelas experimentais estabelecidas para esse fim. A colheita dessas seleções pode ser feita por um certo número de anos para selecionar as de maior rendimento e depois estudar suas progênies, como já foi feito no Brasil e em outros países (Gilbert, 1939; Carvalho, 1952; Fernie, 1965). Na seleção dessas matrizes, deve-se levar em consideração o vigor vegetativo e a qualidade do café. Verificou-se que o vigor vegetativo do cafeeiro pode ser correlacionado com sua produção (Carvalho et al., 1961; Dhaliwal, 1965; Fernie, 1965). Uma vez que grande parte da variação na produtividade do café se deve às condições ambientais, uma correlação muito pobre ou mesmo negativa foi encontrada entre a produtividade da planta-mãe e a de sua progênie. Por essa razão, foi proposto que os registros de produção das árvores-mãe não devem cobrir um período muito longo antes que as seleções finais sejam feitas (Carvalho, 1952). Essa seleção deve ser baseada em apenas um registro de produção de um ano e também no vigor geral do crescimento em diâmetro e altura, que geralmente é uma indicação de que uma safra razoável será obtida no ano seguinte. Os valores genéticos aditivos das árvores-mãe são melhor avaliados pelo estudo de suas progênies em ensaios com repetição. A multiplicação vegetativa das árvores-mãe para testar seus valores genotípicos pode ser aplicada, mas isso talvez seja desnecessário, uma vez que a análise de suas progênies autofecundadas é muito mais fácil e eficiente.

Outras características das sementes, como tamanho, peso e densidade, além da qualidade da bebida também podem ser avaliadas. Depois que essas observações forem feitas e todas as árvores que não atendem totalmente aos requisitos forem eliminadas, deve-se prosseguir para o estágio muito mais importante que é o dos testes de progênies.

## Avaliação e seleção de matrizes

Nesse caso, as matrizes de populações segregantes ( $F_2$  por exemplo) são avaliadas repetidas vezes em várias colheitas e são estimadas as seguintes quantidades, para fins de seleção, conforme estendido de Resende e Silva (1991) e Resende et al. (1995), via incorporação da endogamia ( $F$ ) e da repetibilidade de medições repetidas ( $m$ ).

### Valores genotípicos das matrizes

$I_1 = \hat{g}_1 = \mu + b_{11}(y_m - \mu)$ , em que  $I_1$  é o índice de seleção associado ao método 1,  $\mu$  é a média geral do caráter,  $y_m$  é o valor fenotípico médio observado em uma matriz (medições de campo no caráter em questão) e  $b_{11}$  é o peso (da informação 1 do método 1) ou ponderador de  $y_m$  visando a predição do valor genotípico ( $g_1$ ) via  $I_1$ .

### Estimador do peso $b_{11}$

$$b_{11} = \frac{m(1 + F_m)h^2}{1 + F_m h^2 + (m - 1)(\rho + F_m h^2)}, \text{ em}$$

que  $m$  é o número de colheitas ou medições por planta,  $h^2$  é a herdabilidade individual,  $\rho$  é o coeficiente de repetibilidade individual e  $F_m$  é o coeficiente de endogamia da geração atual das matrizes ( $m$ ).

### Estimador da acurácia ( $r_{\hat{g}_1 g}$ ) da seleção, via o índice de seleção $I_1$

$$r_{\hat{g}_1 g} = \left[ \frac{m(1 + F_m)h^2}{1 + F_m h^2 + (m - 1)(\rho + F_m h^2)} \right]^{1/2} / (1 + F)^{1/2}.$$

Por exemplo, considerando a geração  $F_2$ , em que  $F_m = 0$ , e um caráter com  $h^2=0,30$  e  $\rho=0,50$  e praticando  $m=5$  colheitas tem-se:  $b_{11}=0,50$  e  $r_{\hat{g}_1 g} = 0,50$ , a qual é a acurácia projetada para a população totalmente endógama com  $F = 1$ . Considerando agora a geração  $F_3$ , em que  $F_m = 0,5$ , tem-se:  $b_{11}=0,60$  e  $r_{\hat{g}_1 g} = 0,548$ .

## Método 2: seleção de matrizes com base em suas progênies

O sucesso deste método de seleção no caso de *C. arabica* depende do grau de variabilidade genética encontrada na população original. Resultados positivos no melhoramento de *C. arabica* foram obtidos com esse método, principalmente quando as populações de café não foram previamente selecionadas. Progênies notáveis de várias cultivares foram isoladas por esse método em várias regiões cafeeiras do mundo (Carvalho; Monaco, 1969a, 1969b). Para poder avaliar melhor as progênies de árvores-mãe selecionadas, é altamente desejável testá-las simultaneamente em várias localidades.

Um resultado médio pode então ser obtido, o que permitirá a seleção de progênies bem adaptadas a uma série de diferentes condições ecológicas. Verificou-se que algumas progênies excelentes foram isoladas e amplamente cultivadas, apresentando a característica desejável de ampla adaptabilidade ecológica (Carvalho; Monaco, 1969a, 1969b).

Nesse caso, as progênies ( $F_3$ , por exemplo) das matrizes selecionadas na etapa anterior são avaliadas repetidas vezes em várias colheitas e são estimadas as seguintes quantidades, para fins de seleção de matrizes testadas (seleção das mães com base na avaliação das filhas ou progênies).

#### Valores genotípicos das matrizes

$I_2 = \hat{g}_2 = \mu + b_{21}(y_p - \mu_p)$ , em que  $I_2$  é o índice de seleção associado ao método 2,  $\mu$  é a média geral do caráter,  $y_p$  é o valor fenotípico médio observado na progênie e  $b_{21}$  é o peso ou ponderador de  $y_p$  visando à predição dos valores genéticos das matrizes ( $g_2$ ) via  $I_2$ .

#### Estimador do peso $b_{21}$

$$b_{21} = \frac{(1 + F_m) nmh^2}{(1 + F_i h^2) + (m - 1)(\rho + F_i h^2) + m(n - 1)r_g h^2}$$

em que  $n$  é o número de plantas por progênie,  $m$  é o número de colheitas ou medições por planta,  $h^2$  é a herdabilidade individual,  $\rho$  é o coeficiente de repetibilidade individual,  $r_g$  é o coeficiente de parentesco de Wright entre os indivíduos dentro de progênie, que equivale também ao coeficiente da variância genética aditiva, na variância genética entre progênies,  $r_g = 1 + F_m$ , em que  $F_m$  é o coeficiente de endogamia na geração das mães.

Estimador da acurácia ( $r_{\hat{g}_2 g}$ ) da seleção, via o índice de seleção  $I_2$

$$r_{\hat{g}_2 g} = \left[ \frac{(1 + F_m) nmh^2}{(1 + F_i h^2) + (m - 1)(\rho + F_i h^2) + m(n - 1)r_g h^2} \right]^{1/2} / (1 + F)^{1/2}$$

Por exemplo, considerando a geração  $F_3$ , em que  $r_g = 1$ ,  $F_m = 0$  e  $F_i = 0,5$  (coeficiente de endogamia de indivíduo da geração da progênie), e um caráter com  $h^2 = 0,30$ ,  $\rho = 0,50$ ,  $n = 20$  e praticando  $m = 5$  colheitas, tem-se:  $b_{21} = 0,93$  e  $r_{\hat{g}_2 g} = 0,682$  (Tabela 5).

### Método 3: seleção de matrizes via índice matriz-progênie

Neste caso, propõe-se a utilização de um índice de seleção usando simultaneamente a média da progênie e o valor fenotípico médio da matriz (seleção das mães com base na avaliação das próprias e de suas progênies).

#### Valores genotípicos das matrizes

$I_3 = \hat{g}_3 = (\mu_p + \mu_m)/2 + b_{31}(y_m - \mu_m) + b_{32}(y_p - \mu_p)$ , em que  $I_3$  é o índice de seleção associado ao método 3,  $\mu_p$  é a média geral do caráter avaliado nas progênies,  $\mu_m$  é a média geral do caráter avaliado nas matrizes,  $y_p$  é o valor fenotípico médio observado na progênie  $p$ ,  $y_m$  é o valor fenotípico médio observado na matriz  $m$ ,  $b_{31}$  é o peso ou ponderador de  $y_m$  visando à predição do valor genético ( $g_3$ ) via  $I_3$ , e  $b_{32}$  é o peso ou ponderador de  $y_p$  visando à predição do valor genético ( $g_3$ ) via  $I_3$ . Esse índice foi relatado por Lush (1935) e por Resende e Higa (1994), porém somente para espécie alógama e sem o uso de medidas repetidas. Assim, o presente índice derivado aqui é uma extensão para espécie autógama e com colheitas repetidas.

Os pesos  $b_{31}$  e  $b_{32}$  são estimados via  $b = P^{-1} C$

$$\begin{bmatrix} b_{31} \\ b_{32} \end{bmatrix} = \begin{bmatrix} P_{MM} & P_{MP} \\ P_{PM} & P_{PP} \end{bmatrix}^{-1} \begin{bmatrix} C_1 \\ C_2 \end{bmatrix}$$

em que  $P_{MM}$  é a variância fenotípica entre médias (de várias colheitas) de matrizes,  $P_{PP}$  é a variância fenotípica entre médias (de várias colheitas e plantas por progênie) de progênies,  $P_{MP} = P_{PM}$  é a covariância fenotípica entre dados médios das matrizes e médios de suas progênies representadas por  $n$  indivíduos,  $C_1$  é a covariância entre os valores genéticos paramétricos das matrizes e a média fenotípica das matrizes, e  $C_2$  é a covariância entre os valores genéticos paramétricos das matrizes e a média fenotípica das progênies.

Assim, obtendo-se a inversa  $P^{-1}$  tem-se  $\det(P) = P_{MM}P_{PP} - P_{MP}P_{PM} = P_{MM}P_{PP} - P_{MP}^2$ ;

$$A = \text{Adjunta de } P = \begin{bmatrix} P_{PP} & -P_{MP} \\ -P_{PM} & P_{MM} \end{bmatrix}$$

Numerador da inversa de  $P$ :

$$AC = \begin{bmatrix} P_{PP} & -P_{MP} \\ -P_{PM} & P_{MM} \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \end{bmatrix} = \begin{bmatrix} P_{PP}C_1 - P_{MP}C_2 \\ -P_{PM}C_1 + P_{MM}C_2 \end{bmatrix}$$

Dessa forma, o vetor de pesos é dado por  $AC/\det(P)$ .

$$\begin{bmatrix} b_{31} \\ b_{32} \end{bmatrix} = \frac{\begin{bmatrix} P_{PP}C_1 - P_{MP}C_2 \\ -P_{PM}C_1 + P_{MM}C_2 \end{bmatrix}}{\det(P)} = \frac{\begin{bmatrix} P_{PP}C_1 - P_{MP}C_2 \\ P_{MM}P_{PP} - P_{MP}^2 \\ -P_{PM}C_1 + P_{MM}C_2 \\ P_{MM}P_{PP} - P_{MP}^2 \end{bmatrix}}{\det(P)}$$

Fazendo-se  $b_{31} = 1$ , que é igual a  $b_{31} / b_{31}$ , tem-se que  $b_{32} = b_{32} / b_{31}$ , ou seja,

$$b_{32} = \frac{-P_{PM}C_1 + P_{MM}C_2}{P_{PP}C_1 - P_{MP}C_2}$$

$$\begin{bmatrix} b_{31} \\ b_{32} \end{bmatrix} = \frac{\begin{bmatrix} P_{PP}C_1 - P_{MP}C_2 \\ P_{MM}P_{PP} - P_{MP}^2 \\ -P_{PM}C_1 + P_{MM}C_2 \\ P_{MM}P_{PP} - P_{MP}^2 \end{bmatrix}}{\det(P)} \text{ ou } \begin{bmatrix} b_{31} \\ b_{32} \end{bmatrix} = \begin{bmatrix} 1 \\ \frac{-P_{PM}C_1 + P_{MM}C_2}{P_{PP}C_1 - P_{MP}C_2} \end{bmatrix}$$



Inserindo a composição genética dos elementos do índice, tem-se

$$\begin{bmatrix} C_1 \\ C_2 \end{bmatrix} = \begin{bmatrix} (1 + F_m)h^2 \\ (1 + F_m)h^2 \end{bmatrix}, \text{ assim}$$

$$\begin{bmatrix} b_{31} \\ b_{32} \end{bmatrix} = \begin{bmatrix} \frac{Var(p)(1 + F_m)h^2 - [(1 + F_m)h^2]^2}{Var(m)Var(p) - [(1 + F_m)h^2]^2} \\ \frac{Var(m)(1 + F_m)h^2 - [(1 + F_m)h^2]^2}{Var(m)Var(p) - [(1 + F_m)h^2]^2} \end{bmatrix}.$$

Estimador da acurácia ( $r_{\hat{g}_{3g}}$ ) da seleção, via o índice de seleção  $I_3$

$$r_{\hat{g}_{3g}} = \left( b_{31}(1 + F_m) + b_{32}(1 + F_m) \right)^{1/2} / (1 + F)^{1/2}.$$

em que

$$Var(p) = [(1 + F_i h^2) + (m - 1)(\rho + F_i h^2) + (n - 1)mr_g h^2] / nm, \text{ e}$$

$$Var(m) = 1 + F_m h^2 + (m - 1)(\rho + F_m h^2) / m.$$

#### Método 4: seleção individual no experimento

Avaliação e seleção de indivíduos do experimento de progênies, em várias colheitas, sem considerar os efeitos de famílias. Seleção massal dentro do conjunto de progênies.

Valores genotípicos de indivíduos

$$I_4 = \hat{g}_1 = \mu + b_{41}(y_m - \mu).$$

Esse método é análogo ao método 1, pois baseia-se na média de  $m$  safras por indivíduo. Porém, usa  $F_i$  em vez de  $F_m$ .

$$b_{41} = \frac{m(1 + F_i)h^2}{1 + F_i h^2 + (m - 1)(\rho + F_i h^2)}, \text{ em que}$$

$m$  é o número de colheita ou medição por planta,  $h^2$  é a herdabilidade individual,  $\rho$  é o coeficiente de repetibilidade individual, e  $F_i$  é o coeficiente de endogamia da geração atual.

Estimador da acurácia ( $r_{\hat{g}_{1g}}$ ) da seleção, via o índice de seleção  $I_1$

$$r_{\hat{g}_{1g}} = \left[ \frac{m(1 + F_i)h^2}{1 + F_i h^2 + (m - 1)(\rho + F_i h^2)} \right]^{1/2} / (1 + F)^{1/2}.$$

#### Método 5: seleção entre progênies, visando ao uso de sementes das próprias progênies

Valores genotípicos de progênie

$$I_5 = \hat{g}_5 = \mu + b_{51}(y_p - \mu_p).$$

Estimador do peso  $b_{51}$

$$b_{51} = \frac{[1 + (n - 1)r_g] / n}{[(1 + F_i h^2) + (m - 1)(\rho + F_i h^2) + (n - 1)mr_g h^2] / nm} h^2,$$

em que  $r_g = 1 + F_m$ .

Acurácia do método 5

$$r_{\hat{g}_{5g}} = \frac{\{[1 + (n - 1)r_g] / n\}^2}{[(1 + F_i h^2) + (m - 1)(\rho + F_i h^2) + (n - 1)mr_g h^2] / nm} h^2 / (1 + F)^{1/2}.$$

#### Método 6: seleção combinada dupla via índice progênie-indivíduo na mesma geração

Seleção combinada dupla baseada nas informações de progênies e de indivíduos dentro de progênies.

Valores genotípicos dos indivíduos

$$I_6 = \hat{g}_6 = \mu_p + b_{61}(y_p - \mu_p) + b_{62}(y_i - y_p),$$

Estimadores dos pesos  $b_{61}$  e  $b_{62}$

$$b_{61} = \frac{[1 + (n - 1)r_g] / n}{[(1 + F_i h^2) + (m - 1)(\rho + F_i h^2) + (n - 1)mr_g h^2] / nm} h^2, \text{ e}$$

$$b_{62} = \frac{(1 + F_i) - r_g}{[(1 + F_i h^2) + (m - 1)(\rho + F_i h^2) - mr_g h^2] / m} h^2.$$

Estimador da acurácia ( $r_{\hat{g}_{6g}}$ ) da seleção, via o índice de seleção  $I_6$

$$r_{\hat{g}_{6g}} = \left[ \frac{\{[1 + (n - 1)r_g] / n\}^2}{[(1 + F_i h^2) + (m - 1)(\rho + F_i h^2) + (n - 1)mr_g h^2] / nm} h^2 + \frac{\{(1 + F_i) - r_g\}^2 \left(\frac{n - 1}{n}\right)^2}{[(1 + F_i h^2) + (m - 1)(\rho + F_i h^2) - mr_g h^2] / m} h^2} \right]^{1/2} / (1 + F)^{1/2},$$

em que  $F$  é o coeficiente de endogamia na geração  $F_\infty$ , assim  $1 + F = 1 + 1 = 2$ .

#### Método 7: seleção combinada tripla de indivíduos via índice matriz-progênie-indivíduo

Seleção combinada tripla de indivíduos, via informações de progênies, de matrizes genitoras e de indivíduos dentro de progênie. Esse índice foi proposto por Resende e Higa (1994a), porém somente para espécies alógamas e sem o uso de medidas repetidas. Assim, o presente índice aqui derivado é uma extensão para espécies autógamas e com colheitas repetidas.

Valores genotípicos de indivíduos

$$I_7 = \hat{g}_7 = \mu_p + b_{71}(y_i - y_p) + b_{72}(y_m - \mu_m) + b_{73}(y_p - \mu_p).$$

Os pesos  $b_{71}$ ,  $b_{72}$  e  $b_{73}$  são estimados via

$$b = P^{-1} C$$

$$\begin{bmatrix} b_{71} \\ b_{72} \\ b_{73} \end{bmatrix} = \begin{bmatrix} P_{II} & P_{IM} & P_{IP} \\ P_{MI} & P_{MM} & P_{MP} \\ P_{PI} & P_{PM} & P_{PP} \end{bmatrix}^{-1} \begin{bmatrix} C_1 \\ C_2 \\ C_3 \end{bmatrix},$$

$$\begin{bmatrix} b_{71} \\ b_{72} \\ b_{73} \end{bmatrix} = \begin{bmatrix} Var(i) & (1 + F_m)h^2 & \{[1 + (n - 1)r_g] / n\} h^2 \\ (1 + F_m)h^2 & Var(m) & (1 + F_m)h^2 \\ \{[1 + (n - 1)r_g] / n\} h^2 & (1 + F_m)h^2 & Var(p) \end{bmatrix}^{-1} \begin{bmatrix} (1 + F_i)h^2 \\ (1 + F_m)h^2 \\ \{[1 + (n - 1)r_g] / n\} h^2 \end{bmatrix}$$

em que

$$Var(p) = [(1 + F_i h^2) + (m - 1)(\rho + F_i h^2) + (n - 1)m r_g h^2] / nm,$$

$$Var(m) = 1 + F_m h^2 + (m - 1)(\rho + F_m h^2) / m, \text{ e}$$

$$Var(i) = 1 + F_i h^2 + (m - 1)(\rho + F_i h^2) / m$$

*Estimador da acurácia ( $r_{\hat{g},g}$ ) da seleção, via o índice de seleção  $I_7$*

$$r_{\hat{g},g} = \{ [b_{71}(1+F_i)/h^2] + [b_{72}(1+F_m)/h^2] + [b_{73}\{[1+(n-1)r_g]/n\}/h^2] \}^{1/2} / (1+F)^{1/2},$$

em que F é o coeficiente de endogamia na  $F_\infty$ .

Os resultados para os sete métodos, em todas as gerações, são sumarizados nas Tabelas 5 a 8.

**Tabela 5.** Acurácias seletivas dos vários métodos para caracteres com herdabilidade = 0,10 e repetibilidade = 0,20.

	Seleção de matriz	Seleção de matriz -progênie	Índice matriz + progênie	Seleção individual experimento	Seleção entre progênies	Seleção combinada dupla	Seleção combinada tripla
Geração	Método 1	Método 2	Método 3	Método 4	Método 5	Método 6	Método 7
F3	0,373	0,658	0,660	0,428	0,658	0,688	0,705
F4	0,428	0,828	0,829	0,448	0,814	0,821	0,834
F5	0,448	0,901	0,903	0,458	0,882	0,883	0,894
F6	0,458	0,936	0,937	0,462	0,914	0,914	0,923
F7	0,462	0,952	0,954	0,464	0,929	0,930	0,937

**Tabela 6.** Acurácias seletivas dos vários métodos para caracteres com herdabilidade = 0,20 e repetibilidade = 0,40.

	Seleção de matriz	Seleção de matriz -progênie	Índice matriz + progênie	Seleção individual experimento	Seleção entre progênies	Seleção combinada dupla	Seleção combinada tripla
Geração	Método 1	Método 2	Método 3	Método 4	Método 5	Método 6	Método 7
F3	0,439	0,673	0,675	0,492	0,673	0,716	0,731
F4	0,492	0,840	0,842	0,511	0,826	0,837	0,850
F5	0,511	0,913	0,914	0,519	0,894	0,896	0,907
F6	0,519	0,947	0,948	0,523	0,925	0,926	0,935
F7	0,523	0,964	0,965	0,525	0,941	0,941	0,949

**Tabela 7.** Acurácias seletivas dos vários métodos para caracteres com herdabilidade = 0,30 e repetibilidade = 0,50.

	Seleção de matriz	Seleção de matriz -progênie	Índice matriz + progênie	Seleção individual experimento	Seleção entre progênies	Seleção combinada dupla	Seleção combinada tripla
Geração	Método 1	Método 2	Método 3	Método 4	Método 5	Método 6	Método 7
F3	0,500	0,682	0,684	0,548	0,682	0,741	0,755
F4	0,548	0,849	0,850	0,564	0,834	0,849	0,863
F5	0,564	0,921	0,921	0,571	0,901	0,905	0,916
F6	0,571	0,955	0,955	0,574	0,933	0,934	0,943
F7	0,574	0,971	0,972	0,576	0,948	0,948	0,956

**Tabela 8.** Acurácias seletivas dos vários métodos para caracteres com herdabilidade = 0,70 e repetibilidade = 0,90.

	Seleção de matriz	Seleção de matriz -progênie	Índice matriz + progênie	Seleção individual experimento	Seleção entre progênies	Seleção combinada dupla	Seleção combinada tripla
Geração	Método 1	Método 2	Método 3	Método 4	Método 5	Método 6	Método 7
F3	0,617	0,693	0,695	0,643	0,693	0,796	0,807
F4	0,643	0,858	0,859	0,651	0,844	0,876	0,892
F5	0,651	0,930	0,930	0,654	0,910	0,919	0,935
F6	0,654	0,963	0,964	0,656	0,941	0,944	0,957
F7	0,656	0,980	0,980	0,657	0,956	0,957	0,969

Na fase inicial do melhoramento, o índice matriz-progênie-indivíduo (seleção combinada tripla) mostrou-se o método mais acurado (acurácia de 91% em F<sub>5</sub>), para um caráter com h<sup>2</sup> de 20%, avaliado em 20 indivíduos, conforme a Tabela 6, dentre os sete métodos avaliados. Os resultados revelam também que não é necessário postergar a seleção além da F<sub>5</sub> ou S<sub>3</sub>.

## Índice família-indivíduo-progênie para seleção precoce com acurácia máxima

Em plantas perenes e animais, o índice de seleção usando a genealogia completa envolve múltiplas gerações, por exemplo, conforme o esquema a seguir.

Informações verticais	Informações colaterais
Avó (A)	—
Mãe (M)	Família da mãe (F <sub>M</sub> )
Indivíduo (I)	Família do indivíduo (F <sub>I</sub> )
Progênie (P)	—

O índice de seleção associado ao esquema acima, usando dados ajustados para os efeitos ambientais, tem então a seguinte forma (Viana; Resende, 2014; Resende, 2015):

$$\hat{a} = b_1 A + b_2 F_M + b_3 M + b_4 F_I + b_5 I + b_6 P .$$

$$\begin{bmatrix} (1/h^2) & 0,125 & 0,500 & 0,0625 & 0,250 & 0,125 \\ & (K_2/h^2) & 0,250 & 0,03125 & 0,125 & 0,0625 \\ & & (1/h^2) & 0,125 & 0,500 & 0,250 \\ Sim. & & & (K_4/h^2) & 0,250 & 0,125 \\ & & & & (1/h^2) & 0,500 \\ & & & & & (K_6/h^2) \end{bmatrix} \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hat{b}_3 \\ \hat{b}_4 \\ \hat{b}_5 \\ \hat{b}_6 \end{bmatrix} = \begin{bmatrix} 0,250 \\ 0,125 \\ 0,500 \\ 0,250 \\ 1,000 \\ 0,500 \end{bmatrix}$$

Na Tabela 9, são apresentados vários índices, partindo-se do índice completo envolvendo seis informações, conforme a equação acima, até índices contendo apenas uma informação, obtido pela retirada progressiva (partindo-se da informação associada a b<sub>1</sub>) das informações incluídas no índice total.

**Tabela 9.** Acurácias com o uso de índices de seleção com diferentes fontes de informação, para caracteres com  $h^2 = 0,20$  e  $h^2 = 0,40$  e vários (n) indivíduos por progênie.

Índice <sup>†</sup>	Elementos no índice	n = 5	n = 10	n = 20	n = 50	n = 100
6	Todos exceto avós	0,62	0,70	0,78	0,87	0,92
5		0,62	0,70	0,78	0,87	0,92
4	M, F <sub>I</sub> , I, P	0,62	0,69	0,77	0,87	0,92
3	F <sub>I</sub> , I, P	0,60	0,68	0,77	0,87	0,92
2	I e P	0,58	0,66	0,75	0,86	0,92
1P	P	0,46	0,59	0,71	0,85	0,91
1I	I	0,32	0,32	0,32		
F <sub>I</sub> I*	F <sub>I</sub> e I	0,48	0,51	0,53	0,57	0,58
IM	I e M	0,48	0,48	0,48		
IMFAM	M, F <sub>I</sub> e I	0,51	0,53	0,56	0,59	0,60
Índice <sup>‡</sup>	Elementos no índice	n = 5	n = 10	n = 20	n = 50	n = 100
6	Todos exceto avós	0,77	0,82	0,86	0,94	0,96
5		0,77	0,82	0,87	0,94	0,96
4	M, F <sub>I</sub> , I, P	0,77	0,82	0,87	0,94	0,96
3	F <sub>I</sub> , I, P	0,75	0,81	0,87	0,94	0,96
2	I e P	0,74	0,80	0,86	0,94	0,96
1P	P	0,60	0,73	0,83	0,93	0,96
1I	I	0,45				
F <sub>I</sub> I*	F <sub>I</sub> e I	0,66	0,67	0,68	0,70	0,70
IM	I e M	0,66	0,66			
IMFAM	M, F <sub>I</sub> e I	0,68	0,69	0,70	0,72	0,72

\*: Seleção combinada tradicional. †:  $h^2 = 0,20$ . ‡:  $h^2 = 0,40$ .

Fonte: Viana e Resende (2014) e Resende (2015).

Verifica-se que as acurácias não se reduzem muito pela eliminação no índice, das informações acima da geração dos pais do indivíduo-alvo. O índice 3 (Tabela 9), formado pelas informações do indivíduo, de sua família (que fornece informação sobre o valor genético de seus genitores) e de sua progênie, praticamente maximiza a acurácia seletiva. Assim, um índice contendo três informações (valores genéticos dos pais, valor fenotípico do indivíduo e informação da progênie do indivíduo) é suficiente para estimar, da melhor maneira possível, o valor genético aditivo dos indivíduos, concordando com VanRaden e Wiggans (1991).

Com tamanho da progênie maior que 20, apenas a informação da progênie (índice 1P) é suficiente. Nesse caso, essa informação contempla a contribuição advinda dos pais e a segregação mendeliana no próprio indivíduo.

Outro estudo interessante refere-se à correlação entre diferentes fontes de informação e o valor genético de um indivíduo. A contribuição de diferentes fontes de informação para a acurácia da seleção genética é apresentada na Tabela 10.

**Tabela 10.** Correlação entre diferentes fontes de informação e o valor genético de um indivíduo.

Herdabilidade	Genitor (2)	Genitor e avós (6)	Pedigree ascendente completo	Meio irmão (40)*	Indivíduo (1)	Irmão completo (40)*	Progênie (40)
0,1	0,22	0,27	0,29	0,36	0,32	0,58	0,71
0,2	0,32	0,37	0,39	0,41	0,45	0,64	0,82
0,3	0,39	0,43	0,45	0,44	0,55	0,66	0,87
0,4	0,45	0,49	0,50	0,45	0,63	0,67	0,90
0,5	0,50	0,53	0,54	0,46	0,71	0,68	0,92
0,6	0,55	0,57	0,57	0,47	0,77	0,69	0,94
0,7	0,59	0,61	0,61	0,47	0,84	0,69	0,95

\*: exceto o próprio indivíduo; (2): número de indivíduos, no caso 2.

Fonte: Viana e Resende (2014) e Resende (2015).

Com  $h^2$  igual a 0,20, a acurácia seletiva pelo uso do valor fenotípico médio dos dois genitores é 0,32. Acrescentando as informações dos quatro avós, a acurácia eleva-se a 0,37, e, usando todo o pedigree ascendente fórmula derivada por Searle (1963, 1964), a acurácia aumenta para 0,39. Assim, não é necessário usar informações além dos avós. Nessa mesma situação, o uso apenas da informação do indivíduo apresenta acurácia 0,45, a qual é maior que aquela (0,41) propiciada pelo uso da média de 40 meios irmãos. A informação que mais contribui é aquela da progênie (acurácia de 0,82) seguida pela de irmãos completos (acurácia de 0,64). Assim, o uso das informações do indivíduo, da família e da progênie praticamente conduz à maximização da acurácia seletiva. O uso de todo o pedigree ascendente apresenta mesma acurácia que o uso da média de quatro irmãos completos (Cunningham, 1969). Assim, o núcleo genealógico de um pedigree é calcado no tripé indivíduo, família e progênie.

Baseado no exposto na Tabela 9, o índice família-indivíduo-progênie (índice  $F_{MIP}$ ) ou genitores-indivíduo-progênie foi proposto (VanRaden; Wiggans, 1991; Resende; Higa, 1994) considerando o genitor representado pela média de sua progênie ou pela família ( $F_1$ ) do indivíduo, o próprio valor do indivíduo ( $I$ ), e a média de sua progênie ( $P$ ), respectivamente. Esse índice é dado por  $I = b_f F_1 + b_i I + b_p P$ , em que, para a geração  $F_3$  ( $F_1$  e  $I$  em  $F_3$ ) e  $P$  em  $F_4$ , tem-se

$$\begin{bmatrix} (K_f/h^2) & 1,00 & 1,00 \\ & (K_i) & 1,50 \\ & & (K_p/h^2) \end{bmatrix} \begin{bmatrix} \hat{b}_f \\ \hat{b}_i \\ \hat{b}_p \end{bmatrix} = \begin{bmatrix} 1,00 \\ 1,50 \\ 1,50 \end{bmatrix}.$$

A solução dessa equação fornecerá a estimativa do vetor de pesos  $b$ .

Para plantas autógamas perenes, como o café arábica, tem-se

$$K_i = \frac{(1 + F_i h^2) + (m-1)(\rho + F_i h^2) + (n-1)m r h^2}{nm}, \text{ com}$$

$n = 20$  indivíduos por progênie,  $m = 5$ ,  $h^2 = 0,20$  e  $\rho = 0,40$ , tem-se os resultados

$$\begin{bmatrix} \hat{b}_f \\ \hat{b}_i \\ \hat{b}_p \end{bmatrix} = \begin{bmatrix} 0,12 & (12\%) \\ 0,05 & (5\%) \\ 0,82 & (83\%) \end{bmatrix}.$$

Verifica-se, pelo vetor de pesos estimados, que o teste de progênie (83%) é a fonte de informação mais importante, seguido pela família (12%) do indivíduo e o próprio indivíduo (5%), respectivamente. O teste de progênie é capaz de informar ambos, contribuição parental e segregação mendeliana do indivíduo.

Com base no vetor de pesos, a acurácia é estimada por  $\hat{r}_{aa} = (\sum b_i g_i)^{1/2} = (b_1 g_1 + b_2 g_2 + b_3 g_3 + \dots + b_6 g_6)^{1/2}$ , para espécies alógamas, em que  $g_i$  é um elemento do vetor  $g$  (vetor à direita da igualdade no sistema matricial). Índices envolvendo um menor número de gerações são facilmente obtidos tomando-se submatrizes do sistema acima. Para espécies autógamas, a acurácia é estimada por  $\hat{r}_{aa} = (\sum b_i g_i)^{1/2} = (b_1 g_1 + b_2 g_2 + b_3 g_3 + \dots + b_6 g_6)^{1/2} / (1+F)^{1/2}$ .  $F$  refere-se ao coeficiente de endogamia dos indivíduos na geração infinita. Nesse exemplo, a estimativa da acurácia equivale a 0,844.

Para F<sub>4</sub> e F<sub>5</sub>, tem-se as seguintes composições matriciais:

$$\begin{bmatrix} (K_f/h^2) & 1,50 & 1,50 \\ & (K_i) & 1,75 \\ & & (K_p/h^2) \end{bmatrix} \begin{bmatrix} \hat{b}_f \\ \hat{b}_i \\ \hat{b}_p \end{bmatrix} = \begin{bmatrix} 1,50 \\ 1,75 \\ 1,75 \end{bmatrix} \text{ para F}_4;$$

$$\begin{bmatrix} (K_f/h^2) & 1,75 & 1,75 \\ & (K_i) & 1,875 \\ & & (K_p/h^2) \end{bmatrix} \begin{bmatrix} \hat{b}_f \\ \hat{b}_i \\ \hat{b}_p \end{bmatrix} = \begin{bmatrix} 1,75 \\ 1,875 \\ 1,875 \end{bmatrix} \text{ para F}_5.$$

Para F<sub>2</sub>, tem-se:

$$\begin{bmatrix} (K_f/h^2) & 0,00 & 0,00 \\ & (K_i) & 1,00 \\ & & (K_p/h^2) \end{bmatrix} \begin{bmatrix} \hat{b}_f \\ \hat{b}_i \\ \hat{b}_p \end{bmatrix} = \begin{bmatrix} 0,00 \\ 1,00 \\ 1,00 \end{bmatrix}.$$

Genericamente, por exemplo, tem-se para F<sub>3</sub>:

$$\begin{bmatrix} (K_f/h^2) & Cov(F_3F, F_3I) & Cov(F_3F, F_3I) \\ & (K_i) & Cov(F_4F, F_3I) \\ & & (K_p/h^2) \end{bmatrix} \begin{bmatrix} \hat{b}_f \\ \hat{b}_i \\ \hat{b}_p \end{bmatrix} = \begin{bmatrix} Cov(F_3F, F_3I) \\ Cov(F_3I, F_3I) \\ Cov(F_4F, F_3I) \end{bmatrix}$$

As expressões dependem também do coeficiente de endogamia F<sub>i</sub> dos indivíduos de cada geração e de r, a correlação genética aditiva intraclasse entre indivíduos de cada progênie, que equivale a 1 para progênies F<sub>3</sub>, 1,5 para F<sub>4</sub>, 1,75 para F<sub>5</sub> e 1,875 para F<sub>6</sub>. As seguintes quantidades são também definidas: n, m, h<sup>2</sup>; ρ : número de repetições ou indivíduos, de medidas repetidas, herdabilidade individual, repetibilidade.

O tripé possibilita então máxima acurácia na predição do valor genético real de um indivíduo a ser recombinado em um programa de seleção recorrente intrapopulacional em população híbrida (oriunda da F<sub>1</sub>) e pode ser um indicador da eficiência da seleção precoce (em F<sub>2</sub>, F<sub>3</sub>, F<sub>4</sub> ou F<sub>5</sub>).

Na Tabela 11, são apresentados os resultados para o índice Família-Indivíduo-Progênie (FIP) com indivíduos-alvos em F<sub>2</sub>, F<sub>3</sub>, F<sub>4</sub> e F<sub>5</sub>.

**Tabela 11.** Acurácias associadas ao índice FMIP nas várias gerações. Indivíduos-alvos em F<sub>2</sub>, F<sub>3</sub>, F<sub>4</sub> e F<sub>5</sub>.

Geração	Cenário 1	Cenário 2	Cenário 3	Cenário 4
F2	0,571	0,585	0,596	0,615
F3	0,833	0,844	0,85	0,859
F4	0,909	0,918	0,923	0,93
F5	0,946	0,953	0,958	0,964

Cenário 1: h<sup>2</sup>=0,10, ρ=0,20; Cenário 2: h<sup>2</sup>=0,20, ρ=0,40; Cenário 3: h<sup>2</sup>=0,30, ρ=0,50; Cenário 4: h<sup>2</sup>=0,70, ρ=0,90.  
Fonte: Viana e Resende (2014) e Resende (2015).

Na estratégia de melhoramento via seleção recorrente, o índice genitores-indivíduo-progênie (seleção combinada envolvendo três gerações) mostrou-se acurado (acurácia de 92% em indivíduo em F<sub>4</sub>, usando dados também de F<sub>3</sub> e F<sub>5</sub>, para um caráter com h<sup>2</sup> de 20%, avaliado em 20 indivíduos, conforme a Tabela 11, cenário 2. Os resultados revelam que não é necessário postergar a seleção visando

à recombinação em um programa de seleção recorrente intrapopulacional em população híbrida (oriunda da F<sub>1</sub>) para além da F<sub>4</sub> com testes em F<sub>5</sub>. Isso pode ser um bom indicador da eficiência da seleção precoce (em F<sub>2</sub>, F<sub>3</sub>, ou F<sub>4</sub>). A seleção combinada envolvendo três gerações (Tabela 12) mostrou-se ligeiramente superior à seleção combinada envolvendo duas gerações (Tabela 6).

**Tabela 12.** Acurácias em gerações acumuladas.

Geração	Cenário 1	Cenário 2	Cenário 3	Cenário 4
F <sub>3</sub>	0,658	0,673	0,682	0,693
F <sub>3</sub> ; F <sub>4</sub>	0,832	0,843	0,85	0,858
F <sub>3</sub> ; F <sub>4</sub> ; F <sub>5</sub>	0,908	0,917	0,923	0,93
F <sub>3</sub> ; F <sub>4</sub> ; F <sub>5</sub> ; F <sub>6</sub>	0,946	0,953	0,958	0,964

Cenário 1: h<sup>2</sup>=0,10, ρ=0,20; Cenário 2: h<sup>2</sup>=0,20, ρ=0,40; Cenário 3: h<sup>2</sup>=0,30, ρ=0,50; Cenário 4: h<sup>2</sup>=0,70, ρ=0,90.  
Fonte: Viana e Resende (2014) e Resende (2015).

A partir das equações de modelo misto, a equação para predição dos valores genéticos de cada indivíduo é dada por (Van-Raden; Wiggans, 1991; Resende; Rosa-Perez, 1999):

$$\hat{a}_{ij} = n_1 (MP) + n_2 (VIC) + n_3 (CP),$$

em que MF é a média do valor genético dos parentais, VIC é o valor individual corrigido para os efeitos fixos e CP é a contribuição da progênie.

Esse índice é similar ao índice anterior. Expandindo esse índice tem a forma:

$$\hat{a}_{ij} = 0,5 (\hat{a}_{pmi} + \hat{a}_{pjf}) + K(Y_{ij} - (X\hat{\beta}) - 0,5 (\hat{a}_{pmi} + \hat{a}_{pjf})) + \sum_i (wti (2 \hat{a}_{pi} - \hat{a}_{mi})) / d_{ij},$$

em que

$$n_1 (MF) = 0,5 (\hat{a}_{pmi} + \hat{a}_{pjf});$$

$$n_2 (VIC) = K(Y_{ij} - (X\hat{\beta}) - 0,5 (\hat{a}_{pmi} + \hat{a}_{pjf}));$$

$$n_3 (CP) = \sum_i (wti (2 \hat{a}_{pi} - \hat{a}_{mi})) / d_{ij},$$

em que a contribuição da progênie é definida como a média ponderada (por  $d_{ij}$ ) de duas vezes o valor genético de um indivíduo da progênie menos o valor genético do genitor "mate", com o qual o progenitor em questão foi cruzado. Ver detalhes em Resende e Perez (1999).

A lógica dessa equação é que os elementos da inversa da matriz de parentesco ( $A^{-1}$ ) são correlações parciais dados os valores genéticos dos pais. Assim, as correlações parciais entre avô-filho são iguais a zero em  $A^{-1}$ . Essa matriz tem elementos diferentes de zero fora da diagonal apenas entre o indivíduo e seus genitores, sua progênie e par ("mate") do genitor.

Outra forma de escrever a equação do índice é:

$$\hat{a}_{ij} = b_1(0,5\hat{a}_{pmi} + 0,5\hat{a}_{pjf}) + b_2(Y_{ij} - X\hat{\beta}) + b_3 \sum_i (w(\hat{a}_{pi} - 0,5\hat{a}_{mi})).$$

Usando a seleção genômica ampla (GWS) para avaliar o efeito da segregação mendeliana, tem-se:  $\hat{a}_{ij} = b_1(0,5\hat{a}_{pmi} + 0,5\hat{a}_{pjf}) + \hat{g}(GWS)$ . Assim, comparado com a abordagem tradicional, o preditor  $\hat{g}(GWS)$  substitui a avaliação individual ( $b_2(Y_{ij} - X\hat{\beta})$ ) e o teste de progênie ( $b_3 \sum_i (w(\hat{a}_{pi} - 0,5\hat{a}_{mi}))$ ) do indivíduo, informando sobre o efeito da segregação mendeliana.

## Blup individual dentro de gerações em plantas autógamas perenes

Com o REML/Blup, maior eficiência preditiva e de seleção pode ser alcançada, principalmente, no melhoramento de plantas perenes (Resende, 1999; Barbosa et al., 2005; Resende; Barbosa, 2005; Pedrozo et al., 2009). Os dados associados a autofecundações podem também ser analisados via REML/Blup. Em espécies autógamas é comum a

avaliação, em várias gerações de autofecundação, de linhagens obtidas a partir de cruzamentos entre dois genitores divergentes. Em vários programas de melhoramento nessas espécies, adota-se alguma forma de seleção precoce na geração  $F_3$ , explorando-se a grande variabilidade genética entre e dentro de linhagens  $F_3$ . Tal variabilidade contempla 1,5 vez a variância genética aditiva ( $\sigma_a^2$ ), sendo que 0,5  $\sigma_a^2$  encontra-se dentro de linhagem, e 1,0  $\sigma_a^2$  encontra-se entre linhagens. Assim, tal geração é adequada para seleção, pois 75% (1,5) da variação aditiva total ( $2\sigma_a^2$ ) que estará disponível em  $F_\infty$  já se encontra disponível em  $F_3$ . Dessa forma, a seleção em  $F_3$  por meio de um método preciso como o Blup é relevante.

O Blup para seleção de plantas individuais em linhagens  $F_3$  foi derivado por Resende (2004), considerando o delineamento de blocos ao acaso com várias plantas por parcela. Usando o índice multiefeitos derivado por Resende e Higa (1994), tem-se que o índice ótimo ou Blup para o caso balanceado e, considerando blocos como efeitos fixos, nesse caso, é dado por:

$$I = b_1\delta_{ijk} + b_2g_i + b_3c_j = b_1(Y_{ijk} - \bar{Y}_{ij.}) + b_2(\bar{Y}_{i.} - \bar{Y}...) + b_3(\bar{Y}_{.j} - \bar{Y}_{i.} - \bar{Y}...).$$

A versão matricial desse índice, usando equações de modelo misto, fornece o Blup generalizado para os casos balanceado e desbalanceado. Os coeficientes do índice, no caso de linhagens  $F_3$  advindas do cruzamento entre dois genitores para gerar a  $F_1$ , são dados por:

$$b_1 = \frac{(1/2) \sigma_a^2}{\sigma_\delta^2} \text{ é a herdabilidade do efeito de indivíduo dentro de parcela, } b_2 = \frac{(2nb+1)}{2nb} \sigma_a^2$$

é a herdabilidade do efeito de família, e

$$b_3 = \frac{[(1/2)/n] \sigma_a^2}{\sigma_c^2 + \sigma_\delta^2 / n} \text{ é o coeficiente de determinação do efeito de parcela.}$$

Os componentes de variância  $\sigma_f^2$ ,  $\sigma_p^2$ ,  $\sigma_\delta^2$  e  $\sigma_{pop}^2$  são: variância entre famílias, entre parcelas, dentro de parcelas e entre populações, respectivamente. As quantidades n, b e p referem-se aos números de indivíduos por parcela, número de blocos e número de famílias, respectivamente.

A estimação de  $\sigma_a^2$  usando dados apenas da geração  $F_3$  implica assumir 0,25  $\sigma_a^2$  tendendo a zero na variação entre progênies. Entretanto, mesmo sem essa suposição, a presença dessa pequena fração da variância de dominância ( $\sigma_d^2$ ) não deverá afetar o ranking pelo Blup, pois tal variância estará incluída (0,125  $\sigma_d^2$ ) também no numerador do peso ( $b_1$ ) dado ao componente dentro de linhagem, ao se obter 0,50 da variância genética entre linhagens no numerador de  $b_1$ . Esse índice ou Blup é adequado

também para a seleção envolvendo progêneses  $S_1$  de espécies alógamas.

Em plantas autógamas, geralmente são avaliadas simultaneamente  $p$  linhagens pertencentes a várias populações (pop) segregantes. Nesse caso, o Blup para seleção pelos efeitos genéticos aditivos está associado ao seguinte índice multiefeitos:

$$I_2 = b_1 \delta_{\mu i} + b_2 g_i + b_3 c_{\mu i} + b_4 p_{opi} = b_1 (Y_{ij} - \bar{Y}_{j..}) + b_2 (\bar{Y}_{i..} - \bar{Y}_{...}) + b_3 (\bar{Y}_{j..} - \bar{Y}_{...} - \bar{Y}_{i..} + \bar{Y}_{...}) + b_4 (\bar{Y}_{i..} - \bar{Y}_{...}),$$

em que

$$b_4 = \frac{\sigma_{pop}^2}{\sigma_{pop}^2 + \sigma_p^2 / p + \sigma_c^2 / bp + \sigma_s^2 / nbp}$$

é a herdabilidade de do efeito da média da população.

No caso em que as linhagens  $F_3$  são semeadas em linha e não há repetição, o índice ótimo dentro de população equivale a  $I_3 = b_5 (Y_{ij} - \bar{Y}_{i..}) + b_6 (\bar{Y}_{i..} - \bar{Y}_{...})$ , em que  $b_5 = b_1$  e  $b_6 = \frac{[2n+1]/(2n)}{\sigma_p^2 + \sigma_s^2 / n} \sigma_a^2$ . É importante re-

latar que, nesse caso, a seleção não é puramente genética, pois o experimento não teve repetição, fato que prejudica também a casualização. Numeradores dos pesos entre e dentro de famílias na seleção combinada em cada geração de endogamia são apresentados na Tabela 13.

**Tabela 13.** Numeradores dos pesos entre e dentro de famílias na seleção combinada em cada geração de endogamia.

Geração de seleção	Numerador do peso da família			Numerador do peso dentro de família	
	$F$	$1+F$	$r = \frac{2F}{1+F}$	$\frac{1+(2n-1)F}{n} \sigma_a^2$	$[(1-r)(1+F)] \sigma_a^2$
$F_2$	0	-	-	-	-
$F_3$	1/2	3/2	0,670	$\frac{1+(2n-1)(1/2)}{n} \sigma_a^2 = \frac{2n+1}{2n} \sigma_a^2$	$[(0,33)(3/2)] \sigma_a^2 = 0,50 \sigma_a^2$
$F_4$	3/4	7/4	0,857	$\frac{1+(2n-1)(3/4)}{n} \sigma_a^2 = \frac{6n+1}{4n} \sigma_a^2$	$[(0,143)(7/4)] \sigma_a^2 = 0,25 \sigma_a^2$
$F_5$	7/8	15/8	0,933	$\frac{1+(2n-1)(7/8)}{n} \sigma_a^2 = \frac{14n+1}{8n} \sigma_a^2$	$[(0,067)(15/8)] \sigma_a^2 = 0,13 \sigma_a^2$
$F_6$	15/16	31/16	0,968	$\frac{1+(2n-1)(15/16)}{n} \sigma_a^2 = \frac{30n+1}{16n} \sigma_a^2$	$[(0,032)(31/16)] \sigma_a^2 = 0,06 \sigma_a^2$
$F_\infty$	1	2	1	$\frac{1+(2n-1)}{n} \sigma_a^2 = 2 \sigma_a^2$	0

Fonte: Viana e Resende (2014) e Resende (2015).

O coeficiente de correlação genética ( $r$ ) entre irmãs endógamas (em conexão com indivíduos  $S_1$  ou  $F_3$  advindos de autofecundação) foi apresentado por Lush (1947, 1948, 1994), sendo dado por  $r = \frac{2F}{1+F}$ . O numerador do peso dado à informação do efeito da média de família foi dado por  $\frac{1+(n-1)r}{n} (1+F) \sigma_a^2$  (Resende, 1991). Substituindo o  $r$  na equação, tem-se  $\frac{(1+F)+(n-1)2F}{n} \sigma_a^2$ . Desdobrando, tem-se  $[\frac{(1+F)}{n} + 2F - \frac{2F}{n}] \sigma_a^2$ . Prosseguindo, tem-se  $[\frac{(1+F)+2Fn-2F}{n}] \sigma_a^2$ , seguido por  $\frac{1-F+2Fn}{n} \sigma_a^2$  e  $\frac{1+F(2n-1)}{n} \sigma_a^2$ , ou seja,  $\frac{1+(2n-1)F}{n} \sigma_a^2$ , conforme mostrado na Tabela. Para  $F_3$ ,  $F_4$ ,  $F_5$ ,  $F_6$  e  $F_\infty$  os numeradores são dados por  $\frac{2n+1}{2n} \sigma_a^2$ ,  $\frac{6n+1}{4n} \sigma_a^2$ ,  $\frac{14n+1}{8n} \sigma_a^2$ ,  $\frac{30n+1}{16n} \sigma_a^2$  e  $2 \sigma_a^2$ , respectivamente.

## Blup de famílias e multipopulações dentro de geração em plantas autógamas perenes

Os métodos de seleção em plantas autógamas têm dado os mesmos pesos para todas as progêneses, independentemente dos méritos das populações a que pertencem. E é sabido que existem populações melhores do que outras, e, portanto, suas progêneses devem receber maior peso na seleção.

O índice ótimo (Blup), combinando informações de populações e progêneses via o método bulk dentro de progêneses (BDP), para a seleção dentro de uma geração é dado por:

$$I = b_1 (\bar{Y}_{i..} - \bar{Y}_{...}) + b_2 (Y_{ij} - \bar{Y}_{i..}) = \frac{\rho_p \sigma_{at}^2}{\rho_p \sigma_{at}^2 + \rho_{\mu i} \sigma_{at}^2 / f + \sigma_c^2 / (fb)} (\bar{Y}_{i..} - \bar{Y}_{...}) + \frac{\rho_{\mu i} \sigma_{at}^2}{\rho_{\mu i} \sigma_{at}^2 + (\sigma_c^2 / b) - \rho_p \sigma_{at}^2} (Y_{ij} - \bar{Y}_{i..}),$$

$$= \frac{fb \rho_p h_{at}^2}{(1+F)h_{at}^2 + (b-1)\rho_{\mu i} h_{at}^2 + (fb-1)\rho_p h_{at}^2} (\bar{Y}_{i..} - \bar{Y}_{...}) + \frac{b \rho_{\mu i} h_{at}^2}{1+F h_{at}^2 + (b-1)\rho_{\mu i} h_{at}^2 - \rho_p h_{at}^2} (Y_{ij} - \bar{Y}_{i..})$$



em que  $\sigma_e^2$  é a variância residual (dentro de populações), e  $\rho_{gd}$  é a correlação genética intraclasse entre indivíduos das progênes dentro de população na geração g. As quantidades  $Y_{j..}$ ,  $\bar{Y}_{i..}$  e  $\bar{Y}_{...}$  são valores de médias de progênes, de médias de população e média geral, respectivamente; b é o número de repetições; e f é o número de famílias por população. Tem-se ainda que:

$h_{aT}^2$  é a herdabilidade total entre e dentro de populações;  $\rho_p = \frac{\sigma_{pop}^2}{\sigma_{aT}^2}$  é a proporção da variação genética

total ( $\sigma_{aT}^2 = \sigma_{pop}^2 + \sigma_a^2$ ) devida à variação entre populações ( $\sigma_{pop}^2$ ), em que  $\sigma_a^2$  é a variância genética aditiva média dentro de população;  $F^* = \frac{F_{IT}(1+F_{ST}) - 2F_{ST}}{(1+F_{ST})}$  é

o F de um indivíduo na geração g, corrigido para o F entre populações;  $\rho_{gd} = \frac{(1+F_{IT}^*)}{(1+F_{ST})}$ ;  $F_{IT}^* = (1-F_{ST})F_{IS}$ ;

$F_{ISg} = 0,00; 0,50; 0,75; 0,875; 0,9375$ , para F<sub>2</sub>, F<sub>3</sub>, F<sub>4</sub>, F<sub>5</sub> e F<sub>6</sub>, respectivamente. Define-se ainda o coeficiente de endogamia em razão da estruturação em

populações:  $F_{ST} = \frac{\sigma_p^2}{\sigma_p^2 + 2\sigma_a^2}$ .

Wright (1951) introduziu os parâmetros (denominados "estatísticas F hierárquicas")  $F_{ST}$ ,  $F_{IT}$  e  $F_{IS}$  para descrever estrutura genética, em que

- $F_{ST}$  é a correlação de alelos entre indivíduos dentro de subpopulações. Refere-se, portanto, ao próprio F ou coeficiente de endocruzamento (de Wright) dentro das populações, provocado pela própria deriva genética associada à subdivisão. Percebe-se que existe uma associação entre  $F_{ST}$  e o coeficiente de coancestria entre indivíduos (Resende, 2002).

- $F_{IT}$  é a correlação entre alelos dentro de indivíduos. Refere-se ao coeficiente de endocruzamento de indivíduos no contexto da população global, entre e dentro de subpopulações.

- $F_{IS}$  é a correlação entre alelos dentro de indivíduos dentro de subpopulações. Refere-se a um coeficiente de endocruzamento de um indivíduo devido a cruzamentos não aleatórios dentro de populações. Tem-se também  $F_{IS} = (F_{IT} - F_{ST}) / (1 - F_{ST})$ .

O coeficiente  $F_{IT}$  resulta da endogamia (cruzamentos não aleatórios) dentro da subpopulação ( $F_{IS}$ ) e da endogamia em razão da subdivisão ( $F_{ST}$ ) (Ollivier, 1981, 2002). Tem-se que  $F_{IT} = F_{ST} + (1 - F_{ST}) F_{IS}$ . Essa expressão é fundamental no desenvolvimento teórico que se fará no desenvolvimento do índice de seleção.

A quantidade  $(1 + F_{IT}^*)$  é o coeficiente da variância genética aditiva entre progênes e baseia-se em  $C_{igg}$ , conforme notação de Cockerham (1963), em

que os índices g, t e g' são referentes às gerações atual, precedente e qualquer geração posterior a g, respectivamente.

## Blup de famílias e multigerações em plantas autógamas perenes

Os métodos de melhoramento de plantas autógamas são: genealógico, bulk, descendente de semente única (SSD), BDP, índice de seleção com progênes e gerações (SIPG) e SIPPFG. Os métodos em autógamas ficaram estagnados por longa data. Mais recentemente, o SIPG método foi desenvolvido por Resende et al. (2015) para espécies autógamas anuais e perenes, usando a genealogia completa associada à condução da população via BDP. Usa as médias fenotípicas das progênes nas gerações F<sub>2:3</sub>, F<sub>2:4</sub>, F<sub>2:5</sub> e F<sub>2:6</sub> por meio do índice multigerações dado por  $\hat{a} = b_3 F_3 + \dots + b_6 F_6$ . Os pesos dados a cada geração são calculados por:

$$\begin{bmatrix} (K_3/h^2) & 1,00 & 1,00 & 1,00 \\ & (K_4/h^2) & 1,50 & 1,50 \\ Sim. & & (K_5/h^2) & 1,75 \\ & & & (K_6/h^2) \end{bmatrix} \begin{bmatrix} \hat{b}_3 \\ \hat{b}_4 \\ \hat{b}_5 \\ \hat{b}_6 \end{bmatrix} = \begin{bmatrix} 1,00 \\ 1,50 \\ 1,75 \\ 1,875 \end{bmatrix}.$$

Com medidas repetidas, o fator  $K_i$  para autógamas perenes (café arábica) é dado por:

$$K_i = \frac{(1 + F_i h^2) + (m-1)(\rho + F_i h^2) + (n-1)m r h^2}{nm}$$

São necessárias estimativas da repetibilidade  $\rho$  (melhor se estimada entre plantas F<sub>2</sub>) e da herdabilidade  $h^2$  (melhor se estimada via variância entre progênes F<sub>2:3</sub> / variância fenotípica).

Para plantas autógamas anuais, uma só observação ( $m = 1$ ) é feita para cada indivíduo, de forma que, na fórmula, a repetibilidade e m são tomados como 1 e, portanto:

$$K_i = \frac{(1 + F_i h^2) + (n-1)r h^2}{n}$$

também do coeficiente de endogamia F dos indivíduos de cada geração e de r, a correlação genética aditiva intraclasse entre indivíduos de cada progênie, que equivale a 1 para progênes F<sub>3</sub>, 1,5 para F<sub>4</sub>, 1,75 para F<sub>5</sub> e 1,875 para F<sub>6</sub>.

A Tabela 14 apresenta as acurácias com a seleção de famílias inteiras, em vários cenários (diferentes herdabilidades e repetibilidades), em cada geração de seleção. Os resultados da Tabela 14 mostram que a acurácia aumenta substancialmente da geração F<sub>3</sub> até a geração F<sub>5</sub>. Em F<sub>6</sub>, a acurácia praticamente se estabiliza.

**Tabela 14.** Acurácias baseadas em gerações individuais.

Geração	Cenário 1	Cenário 2	Cenário 3	Cenário 4
F <sub>3</sub>	0,658	0,673	0,682	0,693
F <sub>4</sub>	0,828	0,840	0,849	0,858
F <sub>5</sub>	0,901	0,913	0,921	0,930
F <sub>6</sub>	0,936	0,947	0,955	0,963

Cenário 1:  $h^2=0,10$ ,  $\rho=0,20$ ; Cenário 2:  $h^2=0,20$ ,  $\rho=0,40$ ; Cenário 3:  $h^2=0,30$ ,  $\rho=0,50$ ; Cenário 4:  $h^2=0,70$ ,  $\rho=0,90$ ;  $n = 20$ ;  $m = 5$ .  
 Fonte: Viana e Resende (2014) e Resende (2015).

Na Tabela 15, são apresentadas as acurácias com a seleção de famílias inteiras, por meio do índice multigerações acumuladas. Verifica-se a superioridade da seleção pelo índice, cuja acurácia aumenta com o acúmulo de gerações no índice para a seleção.

**Tabela 15.** Acurácias para o Blup de famílias em gerações acumuladas.

Geração	Cenário 1	Cenário 2	Cenário 3	Cenário 4
F <sub>3</sub>	0,658	0,673	0,682	0,693
F <sub>3</sub> ; F <sub>4</sub>	0,832	0,843	0,850	0,858
F <sub>3</sub> ; F <sub>4</sub> ; F <sub>5</sub>	0,908	0,917	0,923	0,930
F <sub>3</sub> ; F <sub>4</sub> ; F <sub>5</sub> ; F <sub>6</sub>	0,946	0,953	0,958	0,964

Cenário 1:  $h^2=0,10$ ,  $\rho=0,20$ ; Cenário 2:  $h^2=0,20$ ,  $\rho=0,40$ ; Cenário 3:  $h^2=0,30$ ,  $\rho=0,50$ ; Cenário 4:  $h^2=0,70$ ,  $\rho=0,90$ ;  $n = 20$ ;  $m = 5$ .  
 Fonte: Viana e Resende (2014) e Resende (2015).

Na Tabela 16, são apresentadas as acurácias com a seleção de famílias inteiras, por meio do índice multigerações acumuladas, fixando-se  $m$  e a repetibilidade em 1, e variando  $n$  de 1 a 80.

**Tabela 16.** Acurácias do Blup total de gerações acumuladas (no método bulk dentro de famílias), para um caráter com  $h^2 = 0,20$  e variados números  $n$  de indivíduos por família.

Geração	$n = 1$	$n = 5$	$n = 10$	$n = 20$	$n = 40$	$n = 80$
F <sub>3</sub>	0,51	0,58	0,62	0,66	0,68	0,69
F <sub>3</sub> ; F <sub>4</sub>	0,65	0,76	0,80	0,83	0,84	0,85
F <sub>3</sub> ; F <sub>4</sub> ; F <sub>5</sub>	0,75	0,85	0,88	0,90	0,92	0,93
F <sub>3</sub> ; F <sub>4</sub> ; F <sub>5</sub> ; F <sub>6</sub>	0,81	0,90	0,93	0,94	0,95	0,96

### Correlação genética entre cada geração e os valores genéticos das linhagens em $F_\infty$

As correlações genéticas (Tabela 17) entre entidades genéticas de cada geração e os valores genéticos das linhagens em  $F_\infty$  foram obtidas conforme a seguir.

*Correlação genética ao nível de famílias assumindo  $h^2$  igual a 1:*

$$\hat{r}_{aa_x} = \frac{Cov(F_j, F_\infty)}{\sqrt{Var(F_j)Var(F_\infty)}} = \frac{Var(F_j)}{\sqrt{Var(F_j)Var(F_\infty)}} = \frac{\sqrt{c(F_j)}}{\sqrt{1+F}}$$

em que  $c(F_j)$  é o coeficiente da variância genética entre famílias nas gerações de autofecundação,

equivalendo a 1, 1,5, 1,75 e 1,875 para F<sub>3</sub>, F<sub>4</sub>, F<sub>5</sub> e F<sub>6</sub>, respectivamente.

*Correlação genética ao nível de indivíduos assumindo  $h^2$  igual a 1:*

$$\hat{r}_{aa_x} = \frac{Cov(IF_j, F_\infty)}{\sqrt{Var(IF_j)Var(F_\infty)}} = \frac{Var(IF_j)}{\sqrt{Var(IF_j)Var(F_\infty)}} = \frac{\sqrt{c(IF_j)}}{\sqrt{1+F}}$$

em que  $c(IF_j)$  é o coeficiente da variância genética total entre indivíduos nas famílias nas gerações de autofecundação, equivalendo a 1, 1,5, 1,75, 1,875 e 1,9375 para F<sub>2</sub>, F<sub>3</sub>, F<sub>4</sub>, F<sub>5</sub> e F<sub>6</sub>, respectivamente.

**Tabela 17.** Correlações genéticas entre cada geração e os valores genéticos das linhagens em  $F_\infty$ .

Geração	$h^2 = 0,20$ ; $n = 5$ (Indivíduo)	$h^2 = 1$ (Entre famílias)	$h^2 = 1$ (Indivíduo)
F <sub>2</sub>	0,32	–	0,71
F <sub>3</sub>	0,56	0,71	0,87
F <sub>4</sub>	0,72	0,87	0,94
F <sub>5</sub>	0,78	0,94	0,97
F <sub>6</sub>	0,81	0,97	0,98

Fonte: Viana e Resende (2014) e Resende (2015).

Verifica-se que, com o aumento (via aumento do número de repetições) da herdabilidade da seleção na direção de 1 ( $h^2 = 1$  na Tabela 17), a acurácia (correlação genética) da seleção visando à geração  $F_\infty$  também cresce na direção de 1. De maneira geral, os resultados indicam que, com adequada experimentação, a seleção pode ser praticada já em F<sub>4</sub> (seleção de indivíduos) ou F<sub>5</sub> (seleção de famílias). Essa informação é relevante aos programas de seleção recorrente em espécies autógamas.

## Blup multigerações com genitores, populações e progênies em plantas autógamas perenes

Em plantas autógamas, geralmente são avaliadas várias populações oriundas do cruzamento de vários pares de linhagens, sendo gerados vários híbridos simples F<sub>1</sub> e suas correspondentes gerações F<sub>2</sub>. Plantas de cada população F<sub>2</sub> são tomadas para a colheita de sementes, e gerações endogâmicas são avançadas até a obtenção de linhagens, principalmente pelo método do BDP (Ramalho et al., 2012, 2013).

Os métodos de seleção empregados usam dados de gerações individuais (algumas vezes com o auxílio da genealogia, conforme Nunes et al. (2008), e somente recentemente foi proposto um método que usa simultaneamente dados de todas as gerações. Esse é o método BDP-Blup (Blup associado ao esquema Bulk Dentro de Progênies), proposto por Resende et al. (2015). Entretanto, esse método foi otimizado para a aplicação dentro de população, podendo ser aplicado também na mistura de progênies de várias populações. Um método de seleção (denominado SIPP PG) para abarcar toda a estrutura de um programa de melhoramento de plantas autógamas, incluindo todos os dados desde os pais, a geração F<sub>1</sub> e F<sub>2</sub> e, também, tanto os efeitos de

progênies quanto de populações, foi desenvolvido por Resende et al. (2016), pela inclusão dos efeitos de população e, também, dos pais e das gerações F<sub>1</sub> e F<sub>2</sub>. Tradicionalmente, os pais e essas duas gerações são analisados separadamente das gerações endogâmicas a partir de F<sub>3</sub>.

Esse índice utiliza toda a variabilidade genética disponível entre e dentro de populações e gerações, combinando informações dos genitores, populações, progênies e gerações.

O índice total é então dado por:

$$\hat{a} = \hat{b}_{pai1}P_1 + \hat{b}_{pai2}P_2 + \hat{b}_{F1}F_1 + \hat{b}_{F2}F_2 + \hat{b}_{F3pop}F_{3pop} + \hat{b}_{F4pop}F_{4pop} + \hat{b}_{F5pop}F_{5pop} + \hat{b}_{F6pop}F_{6pop} + \hat{b}_{F3prog}F_{3prog} + \hat{b}_{F4prog}F_{4prog} + \hat{b}_{F5prog}F_{5prog} + \hat{b}_{F6prog}F_{6prog}$$

o qual estima o valor genético aditivo de uma linhagem na geração  $F_\infty$ , considerando todas as informações fenotípicas corrigidas desde as gerações parentais (P<sub>1</sub> e P<sub>2</sub>) até a geração F<sub>6</sub>.

A Figura 2 ilustra a estrutura global do programa de melhoramento de uma espécie autógama e mostra a distribuição da variabilidade genética total a ser aproveitada pelo SIPP PG que será derivado. Detalhes podem ser vistos em Resende (2015).

As populações F<sub>1</sub> e F<sub>2</sub> somente possuem efeitos de população, mas não de progênie. Assim, o índice completo, de F<sub>1</sub> a F<sub>6</sub> e mais os dois pais, terá 12 fontes de informação e pesos. No caso do café arábica, poderá conter também plantas F<sub>2</sub> e, nesse caso, terá 13 fontes de informação. O método permite computar o ganho nas linhagens quando se efetua a seleção entre pais e entre populações F<sub>1</sub> ou F<sub>2</sub>. Permite também fazer a seleção em F<sub>3</sub> já com base em seis informações: três de populações de F<sub>1</sub> a F<sub>3</sub>, uma de progênies em F<sub>3</sub> e os dois pais. Pode-se ter também plantas dentro de F<sub>3</sub>.

O índice toma a seguinte forma:  $I = b_{Pop}(\bar{Y}_{i..} - \bar{Y}_{...})_g + b_{Prog}(Y_{.j.} - \bar{Y}_{i..})_g$ . O vetor de pesos  $b_{Pop}$  pondera os efeitos de população nas várias gerações de endogamia, e o vetor de pesos  $b_{Prog}$  pondera os efeitos de progênie nas várias gerações.

Uma vez que esses efeitos são independentes no modelo estatístico linear, os dois vetores podem ser estimados separadamente, gerando os 12 pesos necessários, conforme mostrado a seguir.

**Pesos dos efeitos de população**

$$\hat{b}_{pop} = \begin{bmatrix} \hat{b}_{F3pop} \\ \hat{b}_{F4pop} \\ \hat{b}_{F5pop} \\ \hat{b}_{F6pop} \\ \hat{b}_{pai1} \\ \hat{b}_{pai2} \\ \hat{b}_{F1} \\ \hat{b}_{F2} \end{bmatrix} = \text{Sim.} \begin{bmatrix} (K_{F3pop}/h_{ST}^2) & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} \\ & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} \\ & (K_{F4pop}/h_{ST}^2) & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} \\ & & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} \\ & (K_{F5pop}/h_{ST}^2) & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} \\ & & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} \\ & (K_{F6pop}/h_{ST}^2) & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} \\ & & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} \\ & (K_{pai1}/h_{ST}^2) & 0 & \frac{1}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} \\ & & \frac{1}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} \\ & (K_{pai2}/h_{ST}^2) & \frac{1}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} \\ & & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} \\ & (K_{F1}/h_{ST}^2) & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} \\ & & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} \\ & (K_{F2}/h_{ST}^2) & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} & \frac{2F_{ST}}{1+F_{ST}} \end{bmatrix}$$

em que  $K_{gpop} = (1 + \frac{F_{ST}(1+F_{ST})-2F_{ST}}{(1+F_{ST})} h_{ST}^2 + (b-1) \frac{1+(1-F_{ST})F_{BS}}{(1+F_{ST})} h_{ST}^2 + (f b - 1) \frac{2F_{ST}}{(1+F_{ST})} h_{ST}^2$ ;  $\rho_p = \frac{2F_{ST}}{1+F_{ST}}$ . A inclusão dos efeitos de população nas gerações F<sub>1</sub> e F<sub>2</sub> e também dos genitores (P<sub>1</sub> e P<sub>2</sub>) de cada população considerou que esses efeitos não são correlacionados com os efeitos de progênie dentro de populações (pois esses últimos efeitos são hierárquicos a populações), mas são correlacionados com os efeitos de populações nas várias gerações.

**Pesos dos efeitos de progênie em uma estrutura hierárquica**

$$\hat{b}_{prog} = \begin{bmatrix} \hat{b}_{F3prog} \\ \hat{b}_{F4prog} \\ \hat{b}_{F5prog} \\ \hat{b}_{F6prog} \end{bmatrix} = P^{-1}g = \text{Sim.} \begin{bmatrix} (K_{F3prog}/h_{ST}^2) & 1.00 & 1.00 & 1.00 & \frac{1.00}{1+(1-F_{ST})(0.50)} \\ & (K_{F4prog}/h_{ST}^2) & \frac{1+(1-F_{ST})(0.50)}{(1+F_{ST})} & \frac{1+(1-F_{ST})(0.50)}{(1+F_{ST})} & \frac{1+(1-F_{ST})(0.50)}{(1+F_{ST})} \\ & & (K_{F5prog}/h_{ST}^2) & \frac{1+(1-F_{ST})(0.75)}{(1+F_{ST})} & \frac{1+(1-F_{ST})(0.75)}{(1+F_{ST})} \\ & & & (K_{F6prog}/h_{ST}^2) & \frac{1+(1-F_{ST})(0.875)}{(1+F_{ST})} \end{bmatrix}$$

em que

$$K_{gpop} = \frac{1+F_{ST}h_{ST}^2 + (b-1)\rho_{gd}h_{ST}^2 - \rho_p h_{ST}^2}{b} = \frac{(1 + \frac{F_{ST}(1+F_{ST})-2F_{ST}}{(1+F_{ST})} h_{ST}^2 + (b-1) \frac{1+(1-F_{ST})F_{BS}}{(1+F_{ST})} h_{ST}^2 - \frac{2F_{ST}}{(1+F_{ST})} h_{ST}^2)}{b}$$

**Acurácia do índice**

O estimador da acurácia do índice completo é dado por  $\hat{r}_{\hat{a}a_{\infty}} = \sqrt{\hat{r}_{prog a_{\infty}}^2 + \hat{r}_{pop a_{\infty}}^2}$ , em que

$$\hat{r}_{pop a_{\infty}} = [\sum b_{Fxp} \rho_p / (1+F)]^{1/2} = [(b_{pai1} \rho_p + b_{pai2} \rho_p + b_{F1} \rho_p + b_{F2} \rho_p + b_{F3pop} \rho_p + b_{F4pop} \rho_p + b_{F5pop} \rho_p + b_{F6pop} \rho_p) / (1+F_{\infty})]^{1/2}$$

$$e \hat{r}_{prog a_{\infty}} = [\sum b_{Exp} \rho_{gd} / (1+F)]^{1/2} = [(b_{F3prog} \rho_{3d} + b_{F4prog} \rho_{4d} + b_{F5prog} \rho_{5d} + b_{F6prog} \rho_{6d}) / (1+F_{\infty})]^{1/2}$$

O componente  $\rho_{gd}$  (nesse caso, equivalente aos elementos do vetor g do sistema de equações para os coeficientes b) é o coeficiente de correlação entre a informação fenotípica da geração F<sub>g</sub> e o valor genético aditivo de uma linhagem na geração F<sub>∞</sub>. A acurácia do índice varia somente em função da herdabilidade individual dentro de populações (h<sup>2</sup>) e total (h<sup>2</sup><sub>aT</sub>) do caráter, do número de repetições ou plantas (b) por progênie, do número de progênies por população (f) e  $\rho_p$ . As eficiências relativas dos diferentes procedimentos de seleção podem então ser computadas via experimental ou via simulações. Via simulação é necessário variar h<sup>2</sup>, h<sup>2</sup><sub>aT</sub>, b, f e  $\rho_p$ . **Contribuição indireta das fontes de informação para o valor das linhagens**

Resultados de simulação (avaliação numérica) referente à contribuição indireta das fontes de informação para a seleção de linhagens são apresentados na Tabela 18.

**Tabela 18.** Contribuição indireta das fontes de informação para o valor das linhagens: herdabilidades indiretas, acurácias ( $r_{gg}$ ) e coeficientes de determinação ( $r^2_{gg}$ ).

Entre populações	$h^2$ indireta/ $r_{gg}$ / $r^2_{gg}$ †	Entre progênies	$h^2$ indireta/ $r_{gg}$ / $r^2_{gg}$ †	Entre genitores	$h^2$ indireta/ $r_{gg}$ / $r^2_{gg}$ †
F1	0,19 / 0,18 / 0,03	–	–	Pai1	0,14 / 0,15 / 0,02
F2	0,23 / 0,20 / 0,04	–	–	Pai2	0,14 / 0,15 / 0,02
F3	0,80 / 0,36 / 0,13	F3	0,40 / 0,40 / 0,16	–	–
F4	0,79 / 0,36 / 0,13	F4	0,51 / 0,56 / 0,31	–	–
F5	0,79 / 0,36 / 0,13	F5	0,56 / 0,63 / 0,40	–	–
F6	0,79 / 0,36 / 0,13	F6	0,58 / 0,66 / 0,44	–	–
Entre populações	$h^2$ indireta/ $r_{gg}$ / $r^2_{gg}$ ††	Entre progênies	$h^2$ indireta/ $r_{gg}$ / $r^2_{gg}$ ††	Entre genitores	$h^2$ indireta/ $r_{gg}$ / $r^2_{gg}$ ††
F1	0,33 / 0,29 / 0,08	–	–	Pai1	0,25 / 0,25 / 0,06
F2	0,38 / 0,31 / 0,10	–	–	Pai2	0,25 / 0,25 / 0,06
F3	0,89 / 0,47 / 0,22	F3	0,40 / 0,37 / 0,14	–	–
F4	0,85 / 0,46 / 0,21	F4	0,51 / 0,51 / 0,26	–	–
F5	0,88 / 0,47 / 0,22	F5	0,56 / 0,57 / 0,32	–	–
F6	0,88 / 0,47 / 0,22	F6	0,58 / 0,60 / 0,36	–	–
Entre populações	$h^2$ indireta/ $r_{gg}$ / $r^2_{gg}$ †††	Entre progênies	$h^2$ indireta/ $r_{gg}$ / $r^2_{gg}$ †††	Entre genitores	$h^2$ indireta/ $r_{gg}$ / $r^2_{gg}$ †††
F1	0,55 / 0,43 / 0,18	–	–	Pai1	0,41 / 0,37 / 0,14
F2	0,55 / 0,43 / 0,18	–	–	Pai2	0,41 / 0,37 / 0,14
F3	0,94 / 0,56 / 0,31	F3	0,40 / 0,32 / 0,10	–	–
F4	0,94 / 0,56 / 0,31	F4	0,51 / 0,44 / 0,19	–	–
F5	0,94 / 0,56 / 0,31	F5	0,56 / 0,49 / 0,24	–	–
F6	0,94 / 0,56 / 0,31	F6	0,58 / 0,52 / 0,27	–	–

†:  $h^2 = 0,20$ ;  $h^2_{aT} = 27\%$ ;  $\rho_p = 0,33$ . ††:  $h^2 = 0,20$ ;  $h^2_{aT} = 33\%$ ;  $\rho_p = 0,50$ . †††:  $h^2 = 0,20$ ;  $h^2_{aT} = 43\%$ ;  $\rho_p = 0,67$ .

Verifica-se que as maiores contribuições são dos efeitos de população F<sub>3</sub> a F<sub>6</sub>, seguido por progênies de F<sub>6</sub> a F<sub>3</sub>, F<sub>2</sub>, F<sub>1</sub> e genitores. Quando a variação entre populações é maior do que dentro de populações, aumenta-se a importância dos genitores, F<sub>1</sub> e F<sub>2</sub>, os quais atingem pesos similares aos associados às progênies. Esses pesos não são do índice, mas de cada fonte de informação caso ela fosse utilizada isoladamente na seleção visando ao ganho na linhagem a ser selecionada. Os maiores pesos dos efeitos de população das gerações F<sub>3</sub> a F<sub>6</sub> são pelo fato de elas terem sido avaliadas com grande número de repetições (60 = 3 x 20 famílias), ao passo que as progênies tiveram apenas três repetições.

Esses resultados são importantes nos estágios iniciais do ciclo de seleção, ocasiões essas em que ainda não se dispõe de informações de gerações mais avançadas. A Tabela 18 informa também as acurácias visando à inferência sobre os valores genéticos das futuras linhagens. No cenário em que a maioria da variação encontra-se entre populações,

as seguintes acurácias são proporcionadas por cada fonte de informação: 0,56 para os efeitos de população, 0,52 para o efeito de progênies F<sub>6</sub>; 0,49 para o efeito de progênies F<sub>5</sub>; 0,44 para o efeito de progênies F<sub>4</sub>; 0,43 para os efeitos de populações F<sub>1</sub> e F<sub>2</sub>; 0,37 para os efeitos de genitores; 0,32 para o efeito de progênies F<sub>3</sub>. Assim progênies F<sub>5</sub> e populações F<sub>1</sub> e F<sub>2</sub> são igualmente importantes para a seleção de futuras linhagens. Mesmas acurácias (em torno de 0,32 a 0,37) são contribuídas também pelos genitores P<sub>1</sub> e P<sub>2</sub> e progênies F<sub>3</sub>. Acurácias aproximadamente iguais (0,52 a 0,56) são contribuídas pelas populações em F<sub>3</sub> a F<sub>6</sub> e progênies F<sub>6</sub>. Os coeficientes de determinação ( $r^2_{gg}$ ) indicam as proporções da variação entre as futuras linhagens explicadas por cada fonte de informação. Os valores variaram de 2% (genitores) a 44% (progênies F<sub>6</sub>). Assim, os genitores são pouco informativos nesse contexto. Esses resultados são válidos para herdabilidade (de parcelas individuais) média dentro de população igual a 20%.

*Contribuição indireta das fontes de informação no índice multigerações e multipopulações (SIPPPG)*

Para comparação do novo índice com aquele proposto por Resende et al. (2015) envolvendo mistura das progênies de todas as populações como se fossem pertencentes a uma única

população contendo a variabilidade total, foram considerados seis cenários em termos da herdabilidade dentro de populações e herdabilidade total, respectivamente: 0,15 e 0,21; 0,15 e 0,26; 0,15 e 0,34; 0,30 e 0,39; 0,30 e 0,46; 0,30 e 0,56. Os resultados encontram-se na Tabela 19.

**Tabela 19.** Pesos das várias fontes de informação e acurácias ( $r_{gg}$ ) do índice multipopulações e multigerações (SIPPPG) para alguns cenários.

	Cenário					
$h^2$ Dentro	0,15	0,15	0,15	0,30	0,30	0,30
$h^2$ Total	0,21	0,26	0,34	0,39	0,46	0,56
$h^2$ Entre / $h^2$ Total	0,33	0,50	0,67	0,33	0,50	0,67
Pesos						
P1	0,009	0,013	0,029	0,008	0,012	0,045
P2	0,009	0,013	0,029	0,008	0,012	0,045
F1	0,008	0,009	0,016	0,006	0,008	0,021
F2	0,016	0,016	0,016	0,021	0,021	0,020
F3pop	0,228	0,235	0,230	0,237	0,241	0,224
F4pop	0,222	0,229	0,224	0,228	0,231	0,215
F5pop	0,219	0,226	0,221	0,223	0,227	0,211
F6pop	0,218	0,224	0,219	0,221	0,224	0,209
F3	0,109	0,120	0,139	0,095	0,109	0,130
F4	0,181	0,185	0,193	0,184	0,190	0,199
F5	0,239	0,239	0,238	0,277	0,272	0,264
F6	0,283	0,278	0,271	0,360	0,341	0,318
$r_{gg}$ SIPPPG	0,82	0,84	0,88	0,87	0,88	0,91
$r_{gg}$ BDP-Blup	0,81	0,81	0,81	0,86	0,85	0,84
Eficiência	1,01	1,04	1,09	1,01	1,04	1,08

Fonte: Viana e Resende (2014) e Resende (2015).

*SIPPPG: Índice Total Multigerações com Genitores, Populações ( $F_1$  a  $F_6$ ) e Progênies ( $F_3$  a  $F_6$ )*

Verifica-se que, quando a proporção da variação total em razão das populações é alta, o novo índice conduz a superioridade seletiva de até 9% (sobre o índice multigerações somente de progênies), e isso é um ganho considerável. A acurácia seletiva chega a atingir cerca de 90%, magnitude essa que é ótima, segundo Resende e Duarte (2007). Sobre a seleção tradicional com base em  $F_6$  (acurácia 0,71), a superioridade chega a 28% ( $0,91/0,71 = 1,28$ ). Na Tabela 20, são apresentadas as contribuições percentuais de cada fonte de informação no índice.

**Tabela 20.** Contribuições percentuais de cada fonte de informação para o índice multipopulações e multigerações (SI-PPPG), para valores da herdabilidade total variando de 0,21 a 0,56.

h2 Total	0,21	0,26	0,34	0,39	0,46	0,56	
$\rho_p$	0,33	0,50	0,67	0,33	0,50	0,67	Média
P1	0,5	0,7	1,6	0,4	0,6	2,4	1,0
P2	0,5	0,7	1,6	0,4	0,6	2,4	1,0
F1	0,5	0,5	0,9	0,3	0,4	1,1	0,6
F2	0,9	0,9	0,9	1,1	1,1	1,1	1,0
F3pop	13,1	13,2	12,6	12,7	12,8	11,8	12,7
F4pop	12,8	12,8	12,3	12,2	12,2	11,3	12,3
F5pop	12,6	12,6	12,1	11,9	12,0	11,1	12,1
F6pop	12,5	12,5	12,0	11,8	11,9	11,0	12,0
F3	6,3	6,7	7,6	5,1	5,8	6,8	6,4
F4	10,4	10,4	10,6	9,9	10,1	10,5	10,3
F5	13,7	13,4	13,0	14,8	14,4	13,9	13,9
F6	16,3	15,6	14,8	19,3	18,1	16,7	16,8
Total	100	100	100	100	100	100	100

Fonte: Resende (2015).

Verifica-se, pela última coluna da Tabela 20 que, dada a experimentação considerada (3 repetições e 20 progênies por população), as mais importantes fontes de informação (aquelas com os maiores pesos em porcentagem) foram, pela ordem: progênies em F<sub>6</sub>, progênies em F<sub>5</sub>, populações em F<sub>3</sub>, populações em F<sub>4</sub>, populações em F<sub>5</sub>, populações em F<sub>6</sub>, progênies em F<sub>4</sub>, progênies em F<sub>3</sub>, populações em F<sub>2</sub> e genitores, populações em F<sub>1</sub>. Nesse cenário, a avaliação de populações F<sub>2</sub> e de genitores é igualmente pouco importante, e a avaliação de F<sub>1</sub> não tem serventia. Logicamente essas conclusões são válidas dado todas as informações disponíveis ao final do avanço das gerações endogâmicas até F<sub>6</sub>.

## Blup individual com medidas repetidas e endogamia em plantas autógamas perenes

O procedimento ótimo generalizado é o REML/Blup, podendo ser também a estimação Bayesiana e HIML/HGBLup (Resende et al., 2018, Resende; Alves, 2020).

(a) *Avaliação de tratamentos genéticos g (clones, progênies, híbridos, linhagens) (com interação g x m)*

Modelo:  $y = Xb + Zg + Wgm + Tp + e$

em que b, gm, p são efeitos fixos (medições, blocos), da interação genótipos x medições e permanentes, respectivamente.

As equações de modelo misto para a predição dos efeitos genéticos sob o modelo acima equivalem a:

$$\begin{bmatrix} X'X & X'Z & X'W & X'T \\ Z'X & Z'Z + I\lambda_1 & Z'W & Z'T \\ W'X & W'Z & W'W + I\lambda_2 & W'T \\ T'X & T'Z & T'W & T'T + I\lambda_3 \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{g} \\ \hat{gm} \\ \hat{p} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ W'y \\ T'y \end{bmatrix},$$

$$\text{em que } \lambda_1 = \frac{1 - \rho - h_g^2 \frac{1 - r_{gm}}{r_{gm}}}{h_g^2}; \quad \lambda_2 = \frac{1 - \rho - h_g^2 \frac{1 - r_{gm}}{r_{gm}}}{h_g^2 \frac{1 - r_{gm}}{r_{gm}}};$$

$$\lambda_3 = \frac{1 - \rho - h_g^2 \frac{1 - r_{gm}}{r_{gm}}}{\rho - h_g^2}.$$

Assim, esse modelo usa simultaneamente a herdabilidade ( $h_g^2$ ), a repetibilidade ( $\rho$ ) e a correlação genética ( $r_{gm}$ ) por meio das medições, dadas por:

$$h_g^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_{gm}^2 + \sigma_p^2 + \sigma_e^2}; \quad \rho = \frac{\sigma_g^2 + \sigma_p^2}{\sigma_g^2 + \sigma_{gm}^2 + \sigma_p^2 + \sigma_e^2} \text{ e}$$

$$r_{gm} = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_{gm}^2}.$$

(b) *Avaliação de tratamentos genéticos g (clones, progênies, híbridos, linhagens) (sem interação g x m)*

Modelo:  $y = Xb + Zg + Tp + e$

As equações de modelo misto para a predição dos efeitos genéticos sob o modelo acima equivalem a:

$$\begin{bmatrix} X'X & X'Z & X'T \\ Z'X & Z'Z + I\lambda_4 & Z'T \\ T'X & T'Z & T'T + I\lambda_5 \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{g} \\ \hat{p} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ T'y \end{bmatrix},$$

$$\text{em que } \lambda_4 = \frac{1-\rho}{h_g^2} \text{ e } \lambda_5 = \frac{1-\rho}{\rho-h_g^2}.$$

Assim, esse modelo usa simultaneamente a herdabilidade e a repetibilidade, dadas por:

$$h_g^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_p^2 + \sigma_e^2} \text{ e } \rho = \frac{\sigma_g^2 + \sigma_p^2}{\sigma_g^2 + \sigma_p^2 + \sigma_e^2}.$$

(c) *Avaliação de indivíduos (pertencentes aos tratamentos) por seus valores genéticos aditivos a (com interação famílias x medições)*

Modelo:  $y = Xb + Za + Wgm + Tp + e$

As equações de modelo misto para a predição dos efeitos genéticos sob o modelo acima equivalem a:

$$\begin{bmatrix} X'X & X'Z & X'W & X'T \\ Z'X & Z'Z + A^{-1}\lambda_6 & Z'W & Z'T \\ W'X & W'Z & W'W + I\lambda_7 & W'T \\ T'X & T'Z & T'W & T'T + I\lambda_8 \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \\ \hat{gm} \\ \hat{p} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ W'y \\ T'y \end{bmatrix},$$

$$\text{em que } \lambda_6 = \frac{1-\rho-h_g^2 \frac{1-r_{gm}}{r_{gm}}}{h_a^2}; \lambda_7 = \frac{1-\rho-h_g^2 \frac{1-r_{gm}}{r_{gm}}}{h_g^2 \frac{1-r_{gm}}{r_{gm}}} \text{ e}$$

$$\lambda_8 = \frac{1-\rho-h_g^2 \frac{1-r_{gm}}{r_{gm}}}{\rho-h_a^2}.$$

Assim, esse modelo usa simultaneamente a herdabilidade no sentido restrito ( $h_a^2$ ), a herdabilidade entre progênies ( $h_g^2$ ), a repetibilidade ( $\rho$ ) e a correlação genética ( $r_{gm}$ ) por meio das medições, dadas por:

$$h_a^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_{gm}^2 + \sigma_p^2 + \sigma_e^2}; \rho = \frac{\sigma_a^2 + \sigma_p^2}{\sigma_a^2 + \sigma_{gm}^2 + \sigma_p^2 + \sigma_e^2} \text{ e}$$

$$r_{gm} = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_{gm}^2}.$$

A é a matriz com inclusão da endogamia F.

(d) *Avaliação de indivíduos (pertencentes aos tratamentos) por seus valores genéticos aditivos a (sem interação famílias x medições)*

Modelo:  $y = Xb + Za + Tp + e$

As equações de modelo misto para a predição dos efeitos genéticos sob o modelo acima equivalem a:

$$\begin{bmatrix} X'X & X'Z & X'T \\ Z'X & Z'Z + A^{-1}\lambda_9 & Z'T \\ T'X & T'Z & T'T + I\lambda_{10} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \\ \hat{p} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ T'y \end{bmatrix},$$

$$\text{em que } \lambda_9 = \frac{1-\rho}{h_a^2} \text{ e } \lambda_{10} = \frac{1-\rho}{\rho-h_a^2}.$$

Assim, esse modelo usa simultaneamente a herdabilidade e a repetibilidade, dadas por:

$$h_a^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_p^2 + \sigma_e^2} \text{ e } \rho = \frac{\sigma_a^2 + \sigma_p^2}{\sigma_a^2 + \sigma_p^2 + \sigma_e^2}.$$

(e) *Avaliação de indivíduos (pertencentes aos tratamentos) por seus valores genéticos aditivos a (com inclusão dos efeitos da interação indivíduo dentro de progênies com medição: efeito w)*

Modelo:  $y = Xb + Za + Wgm + Tp + Sw + e$  (com interação famílias x medições)

As equações de modelo misto para a predição dos efeitos genéticos sob o modelo acima equivalem a:

$$\begin{bmatrix} X'X & X'Z & X'W & X'T & X'S \\ Z'X & Z'Z + A^{-1}\lambda_{11} & Z'W & Z'T & Z'S \\ W'X & W'Z & W'W + I\lambda_{12} & W'T & W'S \\ T'X & T'Z & T'W & T'T + I\lambda_{13} & T'S \\ S'X & S'Z & S'W & S'T & S'S + I\lambda_{14} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \\ \hat{gm} \\ \hat{p} \\ \hat{w} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ W'y \\ T'y \\ S'y \end{bmatrix},$$

$$\text{em que } \lambda_{11} = \frac{1-\rho-h_g^2 \frac{1-r_{gm}}{r_{gm}} - h_w^2}{h_a^2}; \lambda_{12} = \frac{1-\rho-h_g^2 \frac{1-r_{gm}}{r_{gm}} - h_w^2}{h_g^2 \frac{1-r_{gm}}{r_{gm}}};$$

$$\lambda_{13} = \frac{1-\rho-h_g^2 \frac{1-r_{gm}}{r_{gm}} - h_w^2}{\rho-h_a^2} \text{ e } \lambda_{14} = \frac{1-\rho-h_g^2 \frac{1-r_{gm}}{r_{gm}} - h_w^2}{h_w^2}.$$

Assim, esse modelo usa simultaneamente a herdabilidade no sentido restrito ( $h_a^2$ ), a herdabilidade entre progênies ( $h_g^2$ ), a repetibilidade ( $\rho$ ), a correlação genética ( $r_{gm}$ ) por meio das medições e a herdabilidade intraindividual dentro de família ( $h_w^2$ ), dadas por:

$$h_a^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_{gm}^2 + \sigma_p^2 + \sigma_w^2 + \sigma_e^2}; \rho = \frac{\sigma_a^2 + \sigma_p^2}{\sigma_a^2 + \sigma_{gm}^2 + \sigma_p^2 + \sigma_w^2 + \sigma_e^2};$$

$$r_{gm} = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_{gm}^2} \text{ e } h_w^2 = \frac{\sigma_w^2}{\sigma_w^2 + \sigma_e^2}.$$

Sob esse modelo, o valor genético total em uma medição é predito por:  $\hat{a}_{total} = \hat{a} + \hat{gm} + \hat{w}$ . Mais detalhes são apresentados por Resende (2015).



(f) Avaliação de indivíduos (pertencentes aos tratamentos) por seus valores genéticos aditivos  $a$  (sem interação famílias  $\times$  medições e com inclusão dos efeitos da interação indivíduo com medição: efeito  $w^*$ )

Modelo:  $y = Xb + Za + Tp + e$

As equações de modelo misto para a predição dos efeitos genéticos sob o modelo acima equivalem a:

$$\begin{bmatrix} X'X & X'Z & X'T & X'S \\ Z'X & Z'Z + A^{-1}\lambda_{15} & Z'T & Z'S \\ T'X & T'Z & T'T + I\lambda_{16} & T'S \\ S'X & S'Z & S'T & S'S + I\lambda_{17} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \\ \hat{p} \\ \hat{w}^* \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ T'y \\ S'y \end{bmatrix},$$

em que  $\lambda_{15} = \frac{1 - \rho - h_{w^*}^2}{h_a^2}$ ;  $\lambda_{16} = \frac{1 - \rho - h_{w^*}^2}{\rho - h_a^2}$  e

$$\lambda_{17} = \frac{1 - \rho - h_{w^*}^2}{h_{w^*}^2}.$$

Assim, esse modelo usa simultaneamente a herdabilidade, a repetibilidade e a herdabilidade intraindividual total ( $h_{w^*}^2$ ), dadas por:

$$h_a^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_p^2 + \sigma_{w^*}^2 + \sigma_e^2}; \quad \rho = \frac{\sigma_a^2 + \sigma_p^2}{\sigma_a^2 + \sigma_p^2 + \sigma_{w^*}^2 + \sigma_e^2} \text{ e}$$

$$h_{w^*}^2 = \frac{\sigma_{w^*}^2}{\sigma_{w^*}^2 + \sigma_e^2}.$$

Sob esse modelo, o valor genético total em uma medição é predito por:  $\hat{a}_{total} = \hat{a} + \hat{w}^*$ . Parametricamente,  $w^* = gm + w$ .

## Blup individual com matriz de covariância estruturada em plantas autógamas perenes

Quando se possuem dados de produção de vários anos das plantas matrizes, deve-se examiná-los quanto à tendência de crescimento da produção anual, medida pelo coeficiente de regressão da produção nos vários anos (Stevens, 1949; Morales, 1951), regressões aleatórias via REML/Blup ou via inferência Bayesiana (Resende, 1997; Resende et al., 2001) ou matrizes de covariância estruturadas (Resende, 2002; Resende; Thompson, 2006) entre colheitas.

Com relação ao período de tempo em que se devem controlar as produções de um grupo de progênies, a fim de separar as melhores, os dados obtidos indicam que as melhores progênies se destacam pela sua produção média mais elevada, bem cedo, às vezes ao 4º ano de colheita. As melhores progênies continuam a ser colhidas por 12 a 15 anos, visando à seleção para longevidade de produção (Carvalho, 1952).

Constata-se uma pronunciada variabilidade anual no caso do café. Os rendimentos subsequentes de uma planta de café não constituem variáveis independentes, uma vez que um alto rendimento geralmente é seguido por uma colheita reduzida no próximo ano. Verificou-se que as primeiras quatro ou cinco colheitas não são tão variáveis como as subsequentes. Para as análises estatísticas, podem ser usados os rendimentos totais de quatro ou mais anos consecutivos ou os rendimentos acumulados de dois anos sucessivos. Sugere-se usar um número par de colheitas, pois um número ímpar delas pode ocasionalmente incluir mais números de colheitas em anos de baixa produção. Esse procedimento é suficiente para uma seleção adequada.

No caso de medidas repetidas em cada indivíduo (ou tratamento) ao longo do tempo, várias alternativas (Resende, 2007) existem para a predição de efeitos e modelagem da estrutura de correlação entre as referidas medidas: a) modelo univariado simplificado de repetibilidade, o qual assume que o caráter é o mesmo (correlação genética igual a um através do tempo) de uma medição para outra, que as correlações fenotípicas (repetibilidades) são de iguais magnitudes entre todos os pares de idade e que as variâncias (genética e residual) são homogêneas; b) modelo univariado de repetibilidade mais interação genótipos  $\times$  medições; c) modelo multivariado completo com matriz de covariância não estruturada, assumindo cada medida como um caráter diferente; d) modelo de regressão aleatória parcimonioso como aproximação do modelo multivariado; e) ajuste de curva *spline* cúbica ou, alternativamente, *spline* tipo B, no intervalo de idades considerado; f) modelos processo caráter, tal como o modelo autorregressivo com variâncias heterogêneas (ARH); g) modelos antedependência estruturados (SAD); h) modelo de correlação bandada (estrutura Toeplitz) com correlações específicas para cada intervalo entre medições; (i) estrutura de simetria composta (CS); j) estrutura de simetria composta com variâncias heterogêneas (CSH). Até o momento, a melhor estrutura encontrada foi a Toeplitz (Andrade et al., 2016), a qual parece adequada para lidar com a bialidade. Isto pode ser notado na estrutura Toeplitz entre 4 colheitas, com a bialidade, mostrada a seguir via correlações (C) nulas entre colheitas vizinhas.

$$C(\text{Toeplitz}) = \begin{bmatrix} C11 & 0 & C13 & 0 \\ 0 & C22 & 0 & C24 \\ C13 & 0 & C33 & 0 \\ 0 & C24 & 0 & C44 \end{bmatrix}.$$

## Blup individual para seleção para homeostase sob bienalidade

Um modelo adequado à avaliação de indivíduos de várias famílias na presença de medidas repetidas é dado por:  $y = X\mu + Za + Wgm + Sb + Tp + e$ , em que  $e = w + \eta$  e  $\sigma_e^2 = \sigma_w^2 + \sigma_\eta^2 \Rightarrow \sigma_\eta^2 = \sigma_e^2 - \sigma_w^2$ . Os componentes de médias do modelo são assim definidos:  $u$ ,  $a$ ,  $gm$ ,  $b$ ,  $p$ ,  $w$  e  $\eta$ : média geral, efeito genético aditivo individual, efeito da interação famílias x medições, efeito de blocos incompletos, efeito permanente de indivíduo, erro correlacionado e erro independente, respectivamente. Letras maiúsculas referem-se às matrizes de incidência dos efeitos do modelo.

Os seguintes coeficientes de herdabilidade no sentido restrito podem ser estimados:

herdabilidade longitudinal individual:

$$h_a^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_{gm}^2 + \sigma_b^2 + \sigma_p^2 + \sigma_w^2 + \sigma_\eta^2};$$

herdabilidade longitudinal de médias ajustada (para blocos) de família de irmãos completos (em que  $n$  e  $m$  são os números de indivíduos por família e de medições por indivíduo):

$$h_{amIG}^2 = \frac{(1/2)\sigma_a^2}{(1/2)\sigma_a^2 + \frac{\sigma_{gm}^2}{m} + \frac{\sigma_d^2}{n} + \frac{\sigma_w^2}{n} + \frac{\sigma_\eta^2}{nm}};$$

$$= \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{gm}^2}{m} + \frac{\sigma_d^2}{n} + \frac{\sigma_w^2}{n} + \frac{\sigma_\eta^2}{nm}}$$

herdabilidade intraindividual:  $h_w^2 = \frac{\sigma_w^2}{\sigma_w^2 + \sigma_\eta^2} = \frac{\sigma_w^2}{\sigma_e^2} = \bar{r}$ ,

em que  $\bar{r} = \frac{2\rho}{m(1-\rho)}$ : média das correlações residuais

entre pares de medições;  $\rho$ : correlação serial entre resíduos;  $\bar{r}\sigma_e^2 = \sigma_w^2$  e  $\sigma_\eta^2 = \sigma_e^2 - \bar{r}\sigma_e^2$ .

A correlação serial pode ser calculada a partir da estatística  $d$  de Durbin-Watson via  $\rho = \frac{2-d}{2}$  em que

$$d = \frac{\sum_{i=1}^n (r_i - r_{i-1})^2}{\sum_{i=1}^n r_i^2},$$

em que  $r_i$ : refere-se aos resíduos ordenados no tempo ou espaço (posição da observação).

O valor esperado de  $w$  é zero com variância  $\sigma_e^2/(1-\rho^2)$ , a qual é independente do tempo e permanece constante. Assim, trata-se de um processo estocástico estacionário. O valor observado da variável em qualquer tempo é distribuído em torno da

média zero e dentro de limites  $\pm \sigma_e / \sqrt{(1-\rho^2)}$ , os quais são conhecidos como limites homeostáticos.

O conceito de herdabilidade intraindividual foi cunhado por Sukhatme e Narain (1982) e tratado detalhadamente por Narain (1990, 2008).

A variação dentro de indivíduos não é puramente ambiental, mas há interação entre a genética do indivíduo e o microambiente. À medida em que se aumenta o número de medições sequenciais, a variabilidade é estabilizada. O comportamento do erro correlacionado de cada indivíduo permite inferir sobre a sua homeostase. Isso é importante para o melhoramento do cafeeiro e de outras fruteiras.

O valor genético aditivo predito é dado por:

$\hat{a}_{total} = \hat{a} + \hat{g}m + \hat{w}$ . E a interação indivíduo dentro

de família com microambiente pode ser estimada por:

$\hat{w} = h_w^2(1-h_d^2)(y-\hat{y}) = h_w^2(1-h_d^2)(dev)$ , em que  $h_d^2$  é

a herdabilidade dentro de família e  $dev = (y-\hat{y})$ . Outro modelo é:  $y = X\mu + Za + Sb + Tp + e^*$ , em que a interação  $gm$  não é ajustada e  $e^* = am + \eta$ , contempla o total da interação indivíduo com microambiente ( $am$ ).

## Blup individual via índice famíliaF<sub>3</sub>-indivíduoF<sub>3</sub>-progênieF<sub>3</sub>:4-indivíduoF<sub>4</sub>, multigerações, para seleção recorrente intrapopulacional em espécies autógamas perenes

A seleção combinada individual e os índices multiefeitos do modelo estatístico (Lush, 1945; Gallais 1980, 1989; Resende; Higa, 1994) são precursores do método Blup individual (Henderson, 1952; Henderson; Quaas, 1976). Na literatura francesa, o Blup (*Meilleure Prédiction Linéaire non Biaisée - MPLNB*) foi denominado inicialmente por *Régression Geno-Phénotypique Multivariable* (RGPM) (Vincourt; Gallais, 1983). O termo RGPM foi também usado por Baradat (1982), Gallais (1980, 1989), Resende et al. (1994) e Fall (1994). A equivalência entre RGPM e Blup foi também provada por Goffinet (1983), envolvendo o risco quadrático médio, equivalente à variância do erro de predição.

O *Estimateur Bayésien*, segundo Ronningen (1971), é equivalente aos métodos dos índices de seleção de Lush e das equações de modelos mistos de Henderson. Robertson (1955) apresentou índices de seleção combinada, derivados sem o uso dos coeficientes de caminhamento ou trilha e mostrou (também descrito por Lerner (1958)) que

o princípio da estimação Bayesiana de valores genéticos baseia-se na seleção combinada das informações a priori e experimental (verossimilhança). Jardine (1958) apresenta uma síntese dos vários índices de seleção combinada. Robertson (1955), Ronningen (1971), Lindley e Smith (1972), Dempfle (1977) e Gianola e Fernando (1986) estabeleceram as bases para o uso da inferência bayesiana na genética quantitativa. A estimação dos componentes de variância é realizada por *Residual Maximum Likelihood* (REML) ou *Maximum de Vraisemblance Résiduelle* (Ollivier, 2002).

O artigo de Gianola e Fernando (1986) foi uma publicação importante antes da era Monte Carlo Markov Chain (MCMC). Naquela época, a aplicação de métodos Bayesianos era tecnicamente árdua e exigia técnicas computacionais avançadas. A partir de 1990, os estatísticos introduziram os métodos MCMC (Gelfand; Smith, 1990), e isso marca o início de uma nova era para análise em genética quantitativa. MCMC é especialmente adequado para implementar modelos Bayesianos por abordagens baseadas em amostragem para calcular densidades marginais. Com dados normais, a Máxima Verossimilhança Marginal (MML) é idêntica à REML (Foulley et al., 1990). Tempelman e Gianola (1993) escreveram um artigo que representa um trabalho pioneiro sobre esse assunto na literatura de melhoramento animal. O MCMC é especialmente adequado para a implementação de modelos Bayesianos que agora se tornaram a principal corrente na genética quantitativa. O primeiro artigo de melhoramento animal a usar uma abordagem Bayesiana-MCMC para análise de dados foi Wang et al. (1993).

Estudos interessantes sobre vários índices de seleção combinada individual em espécies perenes no Brasil foram realizados por Resende et al. (1990), Resende (1991), Resende e Higa (1994) em eucalipto; Simeão et al. (2013, 2014) em forrageiras; e Barbosa et al. (2012) em cana-de-açúcar. Índices multigerações com informação de famílias e populações foram desenvolvidos (Resende et al., 2015, 2016) para espécies autógamas anuais e perenes. Índices multigerações em espécies autógamas, com informação de famílias e indivíduos simultaneamente, são desenvolvidos a seguir, para café arábica. Particularmente importante é o índice família<sub>F3</sub>-indivíduo<sub>F3</sub>-progênie<sub>F3:4</sub>-indivíduo<sub>F4</sub>, multigerações individual, para seleção recorrente intrapopulacional. Aparentemente, o índice multigerações individual para plantas autógamas perenes não foi ainda derivado. Esse tipo de índice parece ter sido considerado apenas por Fall (1994), no Senegal, para espécie autógama anual (amendoim),

com seleção em um sexo só, usando a abordagem RGPM. Segundo Resende et al. (1994) e Fall (1994), tem-se que um índice para a seleção de plantas individuais em linhagens  $F_4$  pode ser derivado, considerando o delineamento de blocos ao acaso com várias plantas por parcela. Usando o raciocínio do índice multifeitos derivado por Gallais (1980) e Resende e Higa (1994), tem-se que o índice ótimo ou Blup para o caso balanceado e, considerando blocos como efeitos fixos, nesse caso, é dado por:

$$I = b_1(\bar{Y}_{i.} - \bar{Y}_{..})_{F3} + b_2(\bar{Y}_{ij} - \bar{Y}_{i.})_{F3} + b_3(\bar{Y}_{i.} - \bar{Y}_{..})_{F4} + b_4(\bar{Y}_{ij} - \bar{Y}_{i.})_{F4}$$

A versão matricial desse índice, usando equações de modelo misto, fornece o Blup generalizado para os casos balanceado e desbalanceado. Os coeficientes do índice, no caso de linhagens  $F_3$  e  $F_4$  advindas do cruzamento entre dois genitores para gerar a  $F_1$ , são dados por:

$b = P^{-1}g$ , em que  $b$  é o vetor de pesos,  $P$  é a matriz de covariância fenotípica e  $g$  é o vetor de covariâncias entre o objetivo (valor genético aditivo na geração  $F_\infty$ ) da seleção e as fontes de informação entre e dentro de progênies  $F_3$  e  $F_4$ . Isso é equivalente ao Blup individual multigerações em espécies autógamas, obtido usando RGPM.

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \begin{bmatrix} \text{Var}(\bar{Y}_{i.} - \bar{Y}_{..})_{F3} & 0 & \text{Var}_g(\bar{Y}_{i.} - \bar{Y}_{..})_{F3} & 0 \\ 0 & \text{Var}(\bar{Y}_{ij} - \bar{Y}_{i.})_{F3} & 0 & \text{Var}_g(\bar{Y}_{ij} - \bar{Y}_{i.})_{F3} \\ \text{Var}_g(\bar{Y}_{i.} - \bar{Y}_{..})_{F3} & 0 & \text{Var}(\bar{Y}_{i.} - \bar{Y}_{..})_{F4} & 0 \\ 0 & \text{Var}_g(\bar{Y}_{ij} - \bar{Y}_{i.})_{F3} & 0 & \text{Var}(\bar{Y}_{ij} - \bar{Y}_{i.})_{F4} \end{bmatrix}^{-1} \begin{bmatrix} \text{Cov}(\bar{Y}_{i.} - \bar{Y}_{..})_{F3}, (\bar{Y}_{i.} - \bar{Y}_{..})_{F3} \\ \text{Cov}(\bar{Y}_{ij} - \bar{Y}_{i.})_{F3}, (\bar{Y}_{ij} - \bar{Y}_{i.})_{F3} \\ \text{Cov}(\bar{Y}_{i.} - \bar{Y}_{..})_{F4}, (\bar{Y}_{i.} - \bar{Y}_{..})_{F4} \\ \text{Cov}(\bar{Y}_{ij} - \bar{Y}_{i.})_{F4}, (\bar{Y}_{ij} - \bar{Y}_{i.})_{F4} \end{bmatrix}$$

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \begin{bmatrix} \text{Var}(\text{entre\_fam})_{F3} & 0 & \text{Var}_g(\text{entre\_fam})_{F3} & 0 \\ 0 & \text{Var}(\text{dentro\_fam})_{F3} & 0 & \text{Var}_g(\text{dentro\_fam})_{F3} \\ \text{Var}_g(\text{entre\_fam})_{F3} & 0 & \text{Var}(\text{entre\_fam})_{F4} & 0 \\ 0 & \text{Var}_g(\text{dentro\_fam})_{F3} & 0 & \text{Var}(\text{dentro\_fam})_{F4} \end{bmatrix}^{-1} \begin{bmatrix} \text{Var}_{g0}(\bar{Y}_{i.} - \bar{Y}_{..})_{F3} \\ \text{Var}_{g0}(\bar{Y}_{ij} - \bar{Y}_{i.})_{F3} \\ \text{Var}_{g0}(\bar{Y}_{i.} - \bar{Y}_{..})_{F4} \\ \text{Var}_{g0}(\bar{Y}_{ij} - \bar{Y}_{i.})_{F4} \end{bmatrix}$$

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \begin{bmatrix} \text{Var}(\text{entre\_fam})_{F3} & 0 & \text{Var}_g(\text{entre\_fam})_{F3} & 0 \\ 0 & \text{Var}(\text{dentro\_fam})_{F3} & 0 & \text{Var}_g(\text{dentro\_fam})_{F3} \\ \text{Var}_g(\text{entre\_fam})_{F3} & 0 & \text{Var}(\text{entre\_fam})_{F4} & 0 \\ 0 & \text{Var}_g(\text{dentro\_fam})_{F3} & 0 & \text{Var}(\text{dentro\_fam})_{F4} \end{bmatrix}^{-1} \begin{bmatrix} \text{Var}(\text{entre\_fam})_{F3} \\ \text{Var}(\text{dentro\_fam})_{F3} \\ \text{Var}(\text{entre\_fam})_{F4} \\ \text{Var}(\text{dentro\_fam})_{F4} \end{bmatrix}$$

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \begin{bmatrix} \text{Var}(\text{entre\_fam})_{F3} & 0 & 1\sigma_g^2 & 0 \\ 0 & \text{Var}(\text{dentro\_fam})_{F3} & 0 & 0,5\sigma_g^2 \\ 1\sigma_g^2 & 0 & \text{Var}(\text{entre\_fam})_{F4} & 0 \\ 0 & 0,5\sigma_g^2 & 0 & \text{Var}(\text{dentro\_fam})_{F4} \end{bmatrix}^{-1} \begin{bmatrix} 1\sigma_g^2 \\ 0,5\sigma_g^2 \\ 1,5\sigma_g^2 \\ 0,25\sigma_g^2 \end{bmatrix}$$

Dividindo por  $\sigma_y^2$  e depois dividindo por  $h^2$ , tem-se:

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \begin{bmatrix} k_i & 0 & 1 & 0 \\ 0 & k_i^* & 0 & 0,5 \\ 1 & 0 & k_i & 0 \\ 0 & 0,5 & 0 & k_i^* \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 0,5 \\ 1,5 \\ 0,25 \end{bmatrix},$$

em que  $K_i$  e  $K_i^*$ :

$$K_i = \frac{(1 + F_i h^2) + (m-1)(\rho + F_i h^2) + (n-1)m r h^2}{nm}$$

$$I = \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hat{b}_3 \\ \hat{b}_4 \end{bmatrix}^T \begin{bmatrix} (\bar{Y}_{i.} - \bar{Y}_{..})_{F3} \\ (\bar{Y}_{ij} - \bar{Y}_{i.})_{F3} \\ (\bar{Y}_{i.} - \bar{Y}_{..})_{F4} \\ (\bar{Y}_{ij} - \bar{Y}_{i.})_{F4} \end{bmatrix}$$

A acurácia é dada por:

$$r_{gg} = [(\hat{b}_1 + \hat{b}_2 0,5 + \hat{b}_3 1,5 + \hat{b}_4 0,25)^{1/2} / (1 + F_{\infty})]^{1/2}$$

Na Tabela 21, são apresentadas as várias categorias de índices de seleção combinada e multifeitos individuais, que equivalem aos Blup individuais para o caso desbalanceado.

**Tabela 21.** Índices de seleção combinada e multifeitos individuais: Blup individuais para o caso desbalanceado.

<b>Categoria</b>	<b>Literatura</b>	<b>Parentesco</b>	<b>Parâmetro genético e endogamia</b>
1. Animais perenes dioicos: índices combinados	Lush (1945, 1947); Bettini (1955); Robertson (1955); Le Roy (1960); Osbourne (1957) e Ronningen e Van Vleck (1985)	Meios irmãos, irmãos completos, dialélicos, fatoriais, hierárquicos	Herdabilidades; herdabilidades e repetibilidade (medidas repetidas)
2. Animais perenes dioicos: Blup individual	Henderson (1952); Ronningen (1971); Henderson; Quaas (1976) e Quaas e Pollack (1980)	Meios irmãos, irmãos completos, dialélicos, fatoriais, hierárquicos	Herdabilidades; herdabilidades e repetibilidade (medidas repetidas)
3. Plantas perenes alógamas	Wright (1962); Arbez et al. (1974); Nanson (1974); Resende (1991); Resende e Higa (1994); Resende; Fernandes (1999) e Resende (1999)	Meios irmãos, irmãos completos, dialélicos, fatoriais, hierárquicos	Herdabilidades
4. Plantas perenes alógamas	Resende et al. (1995); Resende (1997, 1999) e Resende et al. (2000)	Meios irmãos, irmãos completos, dialélicos, fatoriais, hierárquicos	Herdabilidades e repetibilidade (medidas repetidas)
5. Plantas anuais alógamas	Gallais (1980, 1989)	Meios irmãos (1 sexo), multifeitos, clonagem de irmãos, dialélicos, fatoriais, hierárquicos	Herdabilidades
6. Plantas anuais autóguas	Lush (1945); Weber (1982); Wricke; Weber (1986); Durel (1992) e Resende (2004)	Entre e dentro de progênies de autofecundações, gerações únicas	Herdabilidades, F
7. Plantas anuais autóguas	Resende et al. (2015) e Resende (2015)	Entre progênies de autofecundações, multigerações (BDP)	Herdabilidades, F
8. Plantas anuais autóguas	Resende et al. (2016) e Resende (2015)	Entre progênies de autofecundações, multigerações (BDP), multipopulações	Herdabilidades, F, F <sub>ST</sub> , F <sub>IS</sub> , F <sub>IT</sub>
9. Plantas anuais autóguas	Fall (1994)	Entre e dentro de progênies de autofecundações, multigerações (1 sexo)	Herdabilidades, F
10. Plantas perenes autóguas	<i>Este documento</i>	Entre e dentro de progênies de autofecundações, multigerações (BDP)	Herdabilidades, F
11. Plantas perenes autóguas	<i>Este documento</i>	Entre e dentro de progênies de autofecundações, multigerações (BDP)	Herdabilidades, repetibilidade (medidas repetidas), F

## Estratégias de melhoramento do café arábica: planta autógama perene, com propagação seminal e clonal

As etapas fundamentais em um programa de melhoramento genético são:

- (a) Definição dos objetivos do melhoramento e critérios de seleção;
- (b) Escolha do germoplasma: baseia-se nos objetivos do melhoramento e critérios de seleção em termos de características importantes nos ambientes disponíveis para plantio;
- (c) Planejamento da experimentação: delineamento experimental, números de plantas por parcela, de colheitas, de repetições e de locais de teste do germoplasma;
- (d) Modelagem e análise estatística de dados experimentais: seleção de modelos, testes de hipóteses e inferência sobre o controle genético dos caracteres via estimação de componentes de médias e de variâncias genética e ambiental (LS, REML, MCMC-BAYES, HIML), de coeficientes de determinação de efeitos genéticos e fenotípicos (herdabilidades, repetibilidade, determinação da interação genótipos x ambientes e da interação genótipos x colheitas) e de correlações genéticas e fenotípicas (entre caracteres, entre ambientes e entre colheitas);
- (e) Predição de componentes de médias e métodos de seleção: índices de seleção, Blup, Maximo A Posteriori – Bayes (MAP-BAYES), Hierarchical Generalized/Blup (HG-Blup); acurácia e precisão da predição e da seleção;
- (f) Delineamentos de cruzamento para recombinação, obtenção de progênies e composição dos efeitos aleatórios dos índices de seleção e Blup: policruzamentos (meios irmãos), pares simples (irmãos completos), dialélicos (meios irmãos e irmãos completos), fatoriais (meios irmãos e irmãos completos), delineamento V (meios irmãos e irmãos completos), autofecundações (irmãos endógamos, com endogamia  $F_i$ , própria de cada geração  $i$  de autofecundação). Esses delineamentos podem ser empregados nos níveis intrapopulacional, interpopulacional e multipopulacional;
- (g) Método de seleção recorrente: determinado a partir dos resultados da etapa (c), notadamente das estimativas e das significâncias das variâncias aditivas e de dominância;

(h) Estratégia de melhoramento: combinação entre métodos de seleção recorrente e delineamentos de cruzamentos intrapopulacionais e interpopulacionais. Também denominado métodos de melhoramento;

(i) Eficiência seletiva por unidade de tempo: seleção precoce e seleção genômica ampla também denominada seleção pangênômica precoce.

No melhoramento do café, dois tipos de cultivos podem ser almejados: sementes melhoradas; clones superiores. Em *C. arabica*, sementes melhoradas é o objetivo. No caso de *C. canephora*, sementes melhoradas e clones superiores são os objetivos perseguidos. E, nesse caso, quanto à sua importância relativa, clones superiores devem ser subordinados a sementes melhoradas. A reprodução vegetativa permite apenas a fixação de genótipos como clones. Novas combinações de genes só podem ser alcançadas por meio de cruzamentos. Na execução de um programa de melhoramento, ambos os objetivos são perseguidos simultaneamente porque são inseparáveis.

Heterose para o caráter rendimento foi encontrada em *C. arabica* na Costa Rica (Leon, 1965) e na Tanzânia (Fernie, 1965). No entanto, a produção comercial de sementes híbridas não está ainda disponível. Mas pode-se capturar a heterose via propagação de progênies em gerações precoces por meio de propagação vegetativa ou também por meio de sementes para a capitalização da heterose que ainda se manifesta nas gerações precoces nas seguintes proporções: 100% na  $F_1$ , 50% na  $F_2$ , 25% na  $F_3$  e 12,5% na  $F_4$  (Resende; Medrado, 1994). Como as plantas são permanentes em cada geração, podem ser colhidas sementes  $F_2$  nas plantas  $F_1$  (seleção e plantio de  $F_{1;2}$ ); colhidas sementes  $F_3$  nas plantas  $F_2$  (seleção e plantio de  $F_{2;3}$ ); colhidas sementes  $F_4$  nas plantas  $F_3$  (seleção e plantio de  $F_{3;4}$ ).

Para definição adequada das estratégias de melhoramento a serem empregadas, é necessário considerar a natureza do caráter sob melhoramento. Os caracteres de importância econômica podem ser divididos em caracteres qualitativos (monogênicos ou oligogênicos) e quantitativos (poligênicos). Os caracteres qualitativos podem ser melhorados de forma mais direta por meio da transferência de alelos/genes via as seguintes alternativas: identificação dos genótipos desejáveis e sua perpetuação via clonagem (essa modalidade tem sido praticada no melhoramento para resistência a certas doenças em plantas); retrocruzamentos (transferência de alelos); retrocruzamentos monitorados por marcadores moleculares (transferência de alelos); transgenia (transferência de genes).

O melhoramento genético de caracteres quantitativos, controlados por um grande número de genes e/ou sujeitos à elevada influência ambiental, apresentando herdabilidades moderadas a baixas, geralmente baseia-se em três estratégias globais: melhoramento e seleção dentro de população; melhoramento e seleção em populações sintéticas ou compostos; melhoramento do cruzamento ou híbrido entre indivíduos de diferentes populações. As estratégias de melhoramento correspondentes são respectivamente: seleção recorrente intrapopulacional (SRI); seleção recorrente intrapopulacional em população sintética (Srips); seleção recorrente recíproca (SRR) entre populações divergentes. A seleção recorrente, caracterizada por um processo seletivo em vários ciclos, é a principal força capaz de melhorar os caracteres quantitativos. Para tais caracteres, em razão do elevado número de genes, não é possível atingir o objetivo do melhoramento em apenas um ciclo seletivo.

Em plantas autógamas, os programas de melhoramento visam à obtenção de linhagens homogotas, por meio da hibridação entre linhagens, seguida por condução de gerações endogâmicas até a seleção de novas linhagens. O principal processo de melhoramento é a SRI em Srips. Essa é a principal estratégia utilizada na obtenção de segregantes transgressivos, visto que linhagens capturam apenas os efeitos genéticos aditivos. Nesse sentido, não há interesse em aumentar ou melhorar a heterose via SRR.

Seleção recorrente é qualquer processo cíclico de melhoramento que envolve a obtenção de progênies, a avaliação e a recombinação das melhores. Em consequência, é esperado um aumento da frequência dos alelos favoráveis do caráter sob seleção. Todo programa de melhoramento conduzido, em médio e longo prazos, pratica a seleção recorrente, pois certamente os melhores indivíduos obtidos são sempre utilizados nas hibridações da etapa seguinte do programa. A seleção recorrente foi inicialmente proposta para as plantas alógamas, mas o seu emprego tem se intensificado em várias espécies autógamas. Em consequência do sistema de reprodução das espécies autógamas, com o decorrer das sucessivas autofecundações, há o isolamento das progênies e dos indivíduos dentro das progênies, não sendo possível aproveitar os alelos favoráveis que estão em indivíduos diferentes, a não ser por meio do intercruzamento entre eles (Ramalho et al., 2012).

Em plantas autógamas, é aconselhável utilizar de 10 a 20 genitores para a formação da população base para o melhoramento. Para inter cruzar todos

os genitores, uma opção é o cruzamento múltiplo. No entanto, a melhor opção é proceder as hibridações direcionadas, simultaneamente com o processo de seleção, isto é, não obter a população base pura para depois iniciar a seleção e, sim, conduzir as duas atividades ao mesmo tempo, seleção de linhas puras e cruzamentos precoces para a geração das progênies para o próximo ciclo seletivo de seleção recorrente. Para obtenção da população base, pode ser efetuado um dialelo parcial entre 10 genitores, gerando 45 híbridos. Na geração  $F_1$ , os híbridos são avaliados e são selecionadas as 20 melhores populações. Uma mistura de sementes dessas 20 populações pode dar origem a população base  $S_0$ .

Havendo heterose e o interesse em capitalizá-la, outra estratégia é o melhoramento do cruzamento entre populações. Essa estratégia visa ao melhoramento da média do cruzamento entre duas ou mais populações (ou indivíduos) divergentes. E o melhoramento da média do cruzamento implica melhorar simultaneamente a média do valor genético aditivo interpopulacional das duas populações, bem como a heterose do cruzamento entre elas. A heterose é função direta da divergência genética entre as populações e do grau de dominância do caráter sob melhoramento. O processo de melhoramento da média do cruzamento e da heterose é a SRR ou interpopulacional (Comstock et al., 1949). A SRI aplicada às duas populações é também capaz de melhorar a média do cruzamento, mas não melhora a heterose, sendo assim um procedimento inferior quando o grau de dominância é considerável e existe divergência genética. Comparações entre a SRR e SRI para o melhoramento do híbrido interpopulacional são apresentadas por Wricke e Weber (1986), Gallais (1989), Resende (2002) e Resende e Barbosa (2005).

## Comparação de métodos de seleção recorrente

Em termos genéricos, as duas principais formas de seleção recorrente são a SRI e a SRR (Gallais, 1989). A SRI capitaliza os efeitos aditivos dos alelos e, portanto, é uma ferramenta muito poderosa para o melhoramento populacional (seleção para a capacidade geral de combinação). A SRR enfatiza o melhoramento do híbrido interpopulacional, por meio do melhoramento do valor genético aditivo e também da heterose.

A heterose ou vigor híbrido está associada à capacidade específica de combinação (CEC) entre

os genitores, sendo função da diversidade genética entre os progenitores e da dominância alélica nos locos que controlam o caráter de interesse. Dessa forma, um fator primordial na opção pela seleção para a capacidade geral de combinação (seleção recorrente intrapopulacional para os efeitos aditivos) ou pela seleção para capacidade específica de combinação (seleção recorrente recíproca enfatizando o híbrido interpopulacional) é o conhecimento da base genética do caráter de interesse, sobretudo quanto ao parâmetro grau médio de dominância. Determinados caracteres não exibem dominância, e, portanto, para o melhoramento deles, não se justifica a seleção para a CEC.

A heterose tem sido verificada para a maioria dos caracteres produtivos (herdabilidades baixas) em espécies vegetais perenes (Baudouin et al., 1997). O melhoramento da heterose ou do próprio híbrido ou cruzamento interpopulacional apresenta peculiaridades em relação ao melhoramento intrapopulacional, fato que demanda uma abordagem específica desse tema em melhoramento genético. Para o melhoramento do cruzamento entre duas populações, duas estratégias são competidoras: a SRI e a SRR. Em termos puramente genéticos, a SRR apresenta ganhos genéticos (por ciclo de seleção) superiores à SRI quando (Wricke; Weber, 1986; Gallais, 1989): existe sobredominância; a soma das frequências alélicas ( $p_1$  e  $p_2$ ) das duas populações é menor que 1 (neste caso, em qualquer nível de dominância, até mesmo na ausência). A SRI apresenta ganhos superiores à SRR em ausência de sobredominância e, simultaneamente, quando  $p_1 + p_2 > 1$ .

Entretanto, na comparação entre SRR e SRI para o melhoramento de plantas perenes, outras questões de ordem prática devem ser consideradas, tais como: o maior custo, o maior intervalo de gerações, a menor intensidade de seleção e o menor tamanho efetivo populacional associados à SRR. O método da SRR com recombinação dos próprios genitores foi proposto para o melhoramento do eucalipto (Resende; Higa, 1990), pois é a estratégia ideal para o melhoramento dos caracteres de crescimento (volume, diâmetro e altura) Tais caracteres apresentam dominância alélica e considerável heterose no cruzamento entre populações com divergência genética. De maneira geral, todos os programas de melhoramento que adotam a SRR têm obtido sucesso, tais quais aqueles praticados em milho, dendê, café robusta e aves (Resende; Barbosa, 2005).

## Seleção recorrente recíproca individual (SRRI)

Neste procedimento, a SRR é conduzida entre apenas dois genitores com alta CEC e alta média do cruzamento. Esses genitores são autofecundados gerando duas populações  $F_3$  ou  $S_1$ , que são melhoradas, uma em função da outra. Esse esquema tem sido empregado no melhoramento do dendê e do *Eucalyptus grandis* x *Eucalyptus urophylla*. Tem sido empregado também com sucesso no melhoramento do milho, com dois híbridos simples (Coors, 1999; Raposo et al., 2004). A escolha da SRRI é recomendada quando existe um cruzamento (família) muito superior a todos os demais.

De maneira complementar, genitores divergentes devem ser cruzados visando à obtenção de dois efeitos: heterose ou alta capacidade específica de combinação manifestada na progênie; alta variabilidade genética na progênie. A divergência mencionada deve ser para o próprio caráter de interesse. Assim, estimativas de CEC envolvendo cruzamentos são excelentes indicadores de útil divergência genética e podem ser usados na definição de grupos heteróticos (conforme realizado por Paula et al., 2020) e também na escolha de pares para a SRRI.

## Eficiência seletiva por unidade de tempo e seleção genômica ampla (GWS)

Os principais fatores que ditam o sucesso de um programa de melhoramento são: eficiência no processo seletivo (métodos ótimos de experimentação e seleção); estratégia adequada de melhoramento (método de seleção recorrente e delineamento de cruzamento); curtas gerações de melhoramento, ou seja, seleção precoce maximizando o ganho por unidade de tempo. Este documento aborda esses fatores.

Os melhoristas de plantas perenes necessitam encurtar os ciclos seletivos e ter maior agressividade no lançamento de cultivares. Acumular dados experimentais na mesma geração pouco impacta o melhoramento de plantas perenes. A velocidade de troca de cultivares melhoradas é um fator essencial na competitividade do agronegócio. Em culturas de propagação seminal via linhas puras, é importante se ter uma alta rotatividade de cultivares em plantios comerciais, visando reduzir a vulnerabilidade dos plantios aos ataques de novas pragas e doenças. Gerações curtas de melhoramento facilitam isso.

## Seleção genômica ampla

Dados genômicos são especialmente úteis para a seleção genômica ampla (GWS) ou seleção pangênômica (*sélection pangénomique*), conforme Montagnon et al., (2012), que permite a seleção na fase de plântula visando ao ganho genético na fase adulta.

A GWS propicia uma forma de SPD, pois, atua precocemente sobre genes expressos na idade adulta. Assim, a SPD propiciada pela seleção genômica é especialmente importante para o melhoramento de espécies perenes como o cafeeiro. A seleção genômica ampla mostrou-se potencial para o melhoramento de *C. arabica* por possibilitar a obtenção dos valores genéticos genômicos com boa acurácia e pela maior eficiência seletiva, por unidade de tempo, em relação à seleção fenotípica, e possibilitar também a redução no tempo necessário para completar o ciclo de seleção e obter novas cultivares (Resende et al., 2008; Grattapaglia; Resende, 2011; Azevedo et al., 2015).

Sousa et al. (2017) e Alkimim et al. (2020) demonstraram o potencial impacto da seleção genômica (GS) via marcadores Single Nucleotide Polimorphism (SNPs) nos programas de melhoramento do cafeeiro. A GS propicia eficiência e acurácia adequadas para aumentar a taxa de ganho genético por unidade de tempo. Após essa comprovação, a próxima etapa refere-se à incorporação da GS na rotina prática dos programas de melhoramento de cafeeiro. Na prática, a seleção não poderá basear-se exclusivamente em dados moleculares, visto que a maioria dos candidatos à seleção não possui informações genômicas. Assim, a via ideal de incorporação da genômica no melhoramento aplicado é por meio da metodologia Single Step-Genomic-Blup (SS-G-Blup), a qual coloca em avaliação simultânea todos os indivíduos candidatos à seleção, tanto aqueles genotipados, quanto aqueles não genotipados, produzindo um ranking único dos valores genéticos de todos os indivíduos. Nesse método, a acurácia seletiva é maior, pois, por meio das matrizes de parentesco genealógico (A) e genômico (G), uma classe de indivíduos ajuda a outra em termos de informação, a qual será ponderada adequadamente pelo método SS-G-Blup, propiciando maior precisão na seleção. Após a seleção genofenotípica, os indivíduos superiores serão cruzados (ou autofecundados) e recombinados, gerando plântulas que serão genotipadas e terão os seus valores genéticos preditos para fins de seleção precoce no próximo ciclo seletivo.

Essas plântulas genomicamente selecionadas serão levadas, via clonagem ou via progênies, para a avaliação de campo para a comprovação da eficácia seletiva da GWS, início de um novo ciclo de seleção e produção de sementes e/ou clones melhorados. Para dar continuidade a esse processo de inovação no melhoramento do cafeeiro, deve-se implementar e comparar alguns modelos single-step para a seleção genômica de características produtivas, agrônômicas e de resistência a doenças e pragas do cafeeiro.

*Esquemas de GWS em F<sub>2</sub> versus fenotípica em F<sub>3</sub>, F<sub>4</sub> e F<sub>5</sub>*

A fenotipagem visando à estimação dos efeitos de marcas na população genitora deve ser feita com informações em nível de médias de cultivares ou linhagens em várias repetições. Para isso poderão ser utilizados dados da avaliação de linhagens ou Valor, Cultivo e Uso (VCU's) do ciclo anterior de seleção.

Outra opção é a genotipagem de plantas F<sub>2</sub> e a avaliação fenotípica de progênies F<sub>2:3</sub> em várias repetições. Progênies F<sub>2:4</sub> avaliadas em várias repetições e vários locais poderão também ser utilizadas e submetidas à análise de associação com genótipos das plantas F<sub>2</sub>. Esse esquema tem a vantagem de mitigar os efeitos da interação genótipos x ambientes na estimação dos efeitos genéticos das marcas moleculares. Informações combinadas de F<sub>2:3</sub> e F<sub>2:4</sub> em um índice são ainda mais eficientes (Resende et al., 2015).

De posse de um catálogo de efeitos genéticos de marcas, todas as seleções posteriores demandarão apenas genotipagens e predições realizadas com base nelas e no catálogo de marcas. Para a realização de cruzamentos biparentais visando à predição das gerações F<sub>1</sub>, o procedimento relatado a seguir pode ser adotado (Resende et al., 2014; Resende, 2015).

Com o advento da genômica, o uso da capacidade específica de combinação no melhoramento pode ser realizado de maneira mais eficiente. A alocação de cruzamentos pode ser realizada incluindo capacidade geral ( $\alpha$ ) e específica ( $d$ ) de combinação, conforme a seguir, em que  $g$  é o valor genético do cruzamento:

$$\hat{g}_j = \sum_i^n pr_j^{22} \alpha_j + \sum_i^n pr_j^{12} d_j - \sum_i^n pr_j^{11} \alpha_j,$$

em que  $pr_j^{22}$ ,  $pr_j^{12}$  e  $pr_j^{11}$  são as probabilidades de ocorrência dos genótipos 22, 21 e 11 na descendência  $j$ ,  $\alpha_j$  são os efeitos aditivos de substituição alélica e  $d_j$  são os efeitos de dominância.



O objetivo da seleção de cruzamentos é selecionar populações com média alta e grande variabilidade genética. Uma opção para isso é selecionar pelo maior valor de  $u + a + d$ , em que  $u$  é a média geral do caráter sob seleção;  $a$  é relacionado aos efeitos aditivos dos alelos (função de  $\alpha$ );  $d$  é relacionado aos efeitos de dominância contribuídos pelos locos em heterozigose. Essa seleção que garante média alta e, por meio de maior  $d$ , contempla também maior variabilidade genética em  $F_\infty$ , como resultado do maior número de locos em heterozigose (Ramalho et al., 2012). Uma vez que não é possível a realização de todos os cruzamentos possíveis entre os genitores selecionados, esse método de predição genômica de híbridos não observados pode ser empregado visando à obtenção de  $u + a + d$ .

## Estratégias de melhoramento propostas para o café arábica

### Estratégia de Seleção Recorrente (SRI)

Ciclo 1

cruzar 10 genitores;

avaliar 45  $F_1$  (dialélico);

avaliar 20 pop  $F_2$  selecionadas (20 em 45);

avaliar 400 famílias  $F_{2:3}$ : selecionadas por seleção combinada tripla (entre  $F_1$  + entre 20 pop  $F_2$  + dentro (usar  $h^2 = 0.20$ ) 20 pop  $F_2$ );

10 indivíduos  $F_3$ : selecionados por seleção penta combinada (entre  $F_1$  + entre 20 pop  $F_2$  + dentro 20 pop  $F_2$  + entre 400 fam  $F_3$  + dentro 400 fam  $F_3$ ): recombinar e iniciar ciclo 2.

Opcionalmente ir até  $F_4$ ,  $F_5$  ou  $F_6$ .

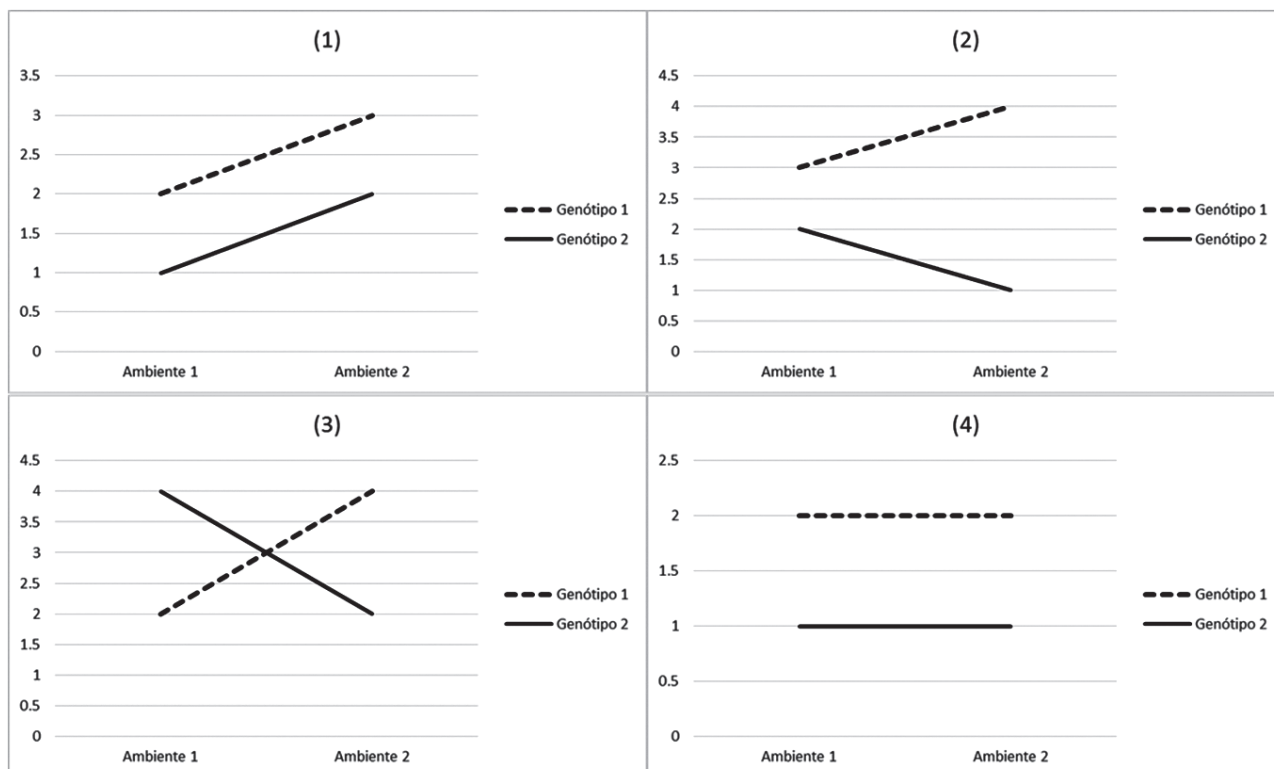
Ciclo 2

### Estratégia de Seleção Recorrente Recíproca (SRR)

Genitores 1–5		Genitores 6–10
	25 $F^1$	
$F^1$ _Rank1		$F^1$ _Rank2
$F^2$ _Rank1 (1 – 200)		$F^2$ _Rank2 (201–400)
200 $F^{2:3}$ _Rank1		200 $F^{2:3}$ _Rank2
	$F^{2:3}$ _Rank1 x $F^{2:3}$ _Rank2	
5 $F^{2:3}$ _Rank1		5 $F^{2:3}$ _Rank2
	25 $F^1$ – Ciclo 2	

## Interação genótipos x ambientes

Os vários genótipos devem ser avaliados em vários ambientes, visando estudar a interação genótipos x ambientes, definida como sendo o comportamento diferencial dos genótipos em ambientes diferentes. A Figura 3 ilustra o comportamento de dois genótipos em dois ambientes, conforme classificação original de Haldane (1946).



**Figura 3.** Possíveis tipos de interação genótipos x ambientes, considerando dois ambientes (Ambiente 1, Ambiente 2) e dois genótipos (Genótipo 1 e Genótipo 2). (1) Não há interação (retas paralelas inclinadas; diferença constante entre os genótipos nos ambientes; diferença na resposta fenotípica atribuída apenas a diferenças entre os ambientes; aditividade dos efeitos de ambientes). (2) Interação simples (retas não paralelas; efeitos de escala em razão de diferença de variabilidade genética nos ambientes, ou seja, heterogeneidade de variâncias nos ambientes). (3) Interação complexa (retas interceptas; inversão de ranks, ausência de correlação positiva entre as respostas dos genótipos nos ambientes). (4) Não há efeito ambiental (retas paralelas planas).

Fonte: Adaptado de Haldane (1946).

## Métodos de seleção, de melhoramento e de condução das gerações de autofecundação em espécies autógamas

Na Tabela 22, é apresentada o cronograma da evolução dos métodos de seleção e melhoramento e de condução das gerações de autofecundação em espécies autógamas.

**Tabela 22.** Cronograma dos métodos de seleção e melhoramento e de condução das gerações de autofecundação em espécies autógamas.

Método	Autor	Periódico	Observação
1. Conceito da prova ou teste de progênie e seleção genealógica	Vilmorin (1850)	Melhoramento da beterraba açucareira na França (avulso)	Contemporâneo a proposição da seleção natural por Darwin (1859), primo de Galton
2. Seleção de linhas puras	Johanssen (1903)	Avulso	Conceito de linhas puras, criação dos termos genes, genótipos, fenótipos e herdabilidade no sentido amplo
3. Método do pedigree ou genealógico	Vilmorin (1856)	Newman (1912)	Contemporâneo a proposição de Galton (1889), da correlação igual a $\frac{1}{2}$ entre pais e filhos
4. Bulk (Melange) e redescoberta do pedigree	Nilsson Ehle (1908) e Newman (1912)*	*Canadian Seed Grower Association	* De acordo com Gallais (1990)
5. Retrocruzamento	Harlan; Pope (1922)	Journal of Heredity	—
6. SSD	Goulden (1939)	Proceedings 7th International Genetic Congress	—
7. Bulk dentro de progênies (F <sub>2</sub> ou F <sub>3</sub> )	Frey (1954)	Agronomy Journal	—
8. Blup multigeracões no bulk dentro de progênies	Resende et al. (2015)	Crop Science	Publicado também em Resende et al. (2014) e Resende (2015)
9. Blup multigeracões e multipopulações e com genitores no bulk dentro de progênies	Resende et al. (2016)	Crop Science	Publicado também em Resende et al. (2014) e Resende (2015)
10. Métodos 8 e 9, para autógamas perenes (com medidas repetidas)	Resende et al. (2014) e Resende (2015)	Livros	—
11. Blup de plantas individuais entre e dentro de progênies de autofecundações, multigeracões (BDP)	Este documento	Este documento	—
12. Métodos 9 e 10, para autógamas perenes (com medidas repetidas), com efeitos aditivos e de dominância	Este documento	Este documento	—

Algumas datas importantes no melhoramento genético são:

700 a.C. – Polinização controlada em Tamareiras (*Phoenix dactylifera*) no Egito.

1769 – Evidência do vigor híbrido (heterose) a partir do cruzamento entre espécies.

1888 – Waldeyer propôs o termo cromossomo.

1900 – De Vries propõe o termo mutação.

1908–1909 – Hardy & Weinberg estabelecem as leis da panmixia.

1918 – Fisher introduz o modelo (infinitesimal) básico da genética quantitativa.

1921 – Wright desenvolve a teoria da genética de populações.

1948 – Malécot publica o livro *Les mathématiques de l'hérédité*.

1952 – Malécot modela a probabilidade de fixação de uma mutação (alelo).

1953 – Watson & Crick descobrem a estrutura em dupla hélice do DNA.

2000 – Meuwissen formula a seleção genômica.

## Considerações finais

O presente trabalho teve como objetivos desenvolver a teoria e realizar recomendações práticas acerca dos métodos ótimos de seleção, visando à correta identificação dos indivíduos superiores, e das estratégias efetivas de melhoramento, visando à criação de novos indivíduos superiores, genético do cafeeiro arábica. Foram derivados métodos ótimos de seleção para café arábica, de acordo com várias situações. Na fase inicial do melhoramento, o índice matriz-progênie-indivíduo (seleção combinada tripla) mostrou-se o método mais acurado (acurácia de 91% em F<sub>5</sub>, para um caráter com h<sup>2</sup> de 20%, avaliado em 20 indivíduos por família), dentre sete métodos avaliados. Na estratégia de melhoramento via seleção recorrente, o índice genitores-indivíduo-progênie (seleção combinada envolvendo três gerações) mostrou-se acurado (acurácia de 92% em indivíduo em F<sub>4</sub>, usando dados também de F<sub>3</sub> e F<sub>5</sub>, para um caráter com h<sup>2</sup> de 20%, avaliado em 20 indivíduos por família). Os resultados revelaram que não é necessário postergar a seleção, visando à recombinação em um programa de seleção recorrente intrapopulacional em população híbrida (oriunda da F<sub>1</sub>), para além da F<sub>4</sub> com testes em F<sub>5</sub>. Isso pode ser um bom indicador da eficiência da seleção precoce (em F<sub>2</sub>, F<sub>3</sub> ou F<sub>4</sub>). Pode-se adotar alguma forma de seleção precoce já na geração F<sub>3</sub>, explorando-se

a grande variabilidade genética entre e dentro de linhagens F<sub>3</sub>. A existência de heterose para vários caracteres de interesse revela que dificilmente será vantajoso retardar a seleção para gerações avançadas. Pelo contrário, realizar a seleção logo após a hibridação (F<sub>1</sub>) ou em F<sub>2</sub>, associada à propagação vegetativa, deverá ser a estratégia mais eficiente, pois, além de capturar a heterose máxima, reduz por muitos anos o comprimento do ciclo seletivo. Em cafeeiro, a estratégia de hibridação seguida de seleção e propagação vegetativa deverá ser usada rotineiramente visando a ganhos máximos em produtividade e à perpetuação de indivíduos resistentes a doenças e pragas (insetos e nematoides).

## Referências

- ALKIMIM, E. R.; CAIXETA, E. T.; SOUSA, T. V.; RESENDE, M. D. V.; SILVA, F. L.; SAKIYAMA, N. S.; ZAMBOLIM, L. Selective efficiency of genome-wide selection in *Coffea canephora* breeding. **Tree Genetics and Genomes**, v. 16, n. 41, Apr. 2020. DOI: <https://doi.org/10.1007/s11295-020-01433-3>.
- ANDRADE, V. T.; GONÇALVES, F. M. A.; NUNES, J. A. R.; BOTELHO, C. E. Statistical modeling implications for coffee progenies selection. **Euphytica**, v. 207, p. 177-189, Jan. 2016. DOI: <https://doi.org/10.1007/s10681-015-1561-6>.
- ANDREOTTI, C. A. **Chefs: Café**. São Paulo: Melhoramentos, 2014.
- ARBEZ, M.; BARADAT, P.; MAUGE, J. P.; MILLIER, C.; BADIA, J. Some problems related to use of selection indices in forest tree breeding. In: IUFRO JOINT MEETING, 1974, Stockholm. **Proceedings [...]** Stockholm: IUFRO, 1974. p. 97-116.
- AZEVEDO, C. F.; RESENDE, M. D. V.; SILVA, F. F. Ridge, Lasso and Bayesian additive-dominance genomic models. **BMC Genomics**, v. 6, p. 105, 2015.
- BARADAT, P. **Genétique quantitative, modèles statistiques et génétiques de base**: Fascicule I. [Bordeaux]: Inra, 1982. 210 p. Document polycopie, INRA Amélior. Des Arbres forestiers.
- BARBOSA, M. H. P.; RESENDE, M. D. V.; BRESSIANI, J. A.; SILVEIRA, L. C. L. Selection of sugarcane families and parents by Reml/Blup. **Crop Breeding and Applied Biotechnology**, v. 5, p. 443-450, 2005.
- BARBOSA, M. H. P.; RESENDE, M. D. V.; DIAS, L. A. S.; BARBOSA, G. V. S.; OLIVEIRA, R. A.; PETERNELLI, L. A.; DAROS, E. Genetic improvement of sugar cane

for bioenergy: the Brazilian experience in network research with RIDESA. **Crop Breeding and Applied Biotechnology**, Dec. 2012. DOI: <https://doi.org/10.1590/S1984-70332012000500010>.

BAUDOIN, L.; BARIL, C.; CLEMENTDEMANGE, A.; LEROY, T.; PAULIN, D. Recurrent selection of tropical tree crops. **Euphytica**, v. 96, n. 1, p. 101-114, 1997.

BETTINI, T. M. **Il miglioramento degli animali**: problemi e metodi. Firenze: [s.n.], 1955. 497 p.

BRAVAIS, A. Analyse mathématique sur les probabilités des erreurs de situation d'un point - 1846. **Memoires par divers Savans**, n. 9, p. 255-332, 1946.

BRUCKNER, C. H. Melhoria de fruteiras. In: BORÉM, A. (org.). **Melhoria de espécies cultivadas**. Viçosa, MG: Ed. da UFV, 1999. p. 679-714.

BURROWS, P. M. **Consequences of selection in finite populations**. North Carolina: State University at Raleigh, 1971.

CARVALHO, A. Melhoria do cafeeiro VI: estudo e interpretação para fins de seleção de produções individuais na variedade Bourbon. **Bragantia**, v. 12, n. 4-6, Jun. 1952. DOI: <https://doi.org/10.1590/S0006-87051952000200007>.

CARVALHO, A.; FAZUOLI, L. C. Café. In: FURLANI, A. M. C.; VIEGAS, G. P. O. **Melhoria de plantas no Instituto Agrônomo**. Campinas: Instituto Agrônomo, 1993. p. 29-76.

CARVALHO, A.; MEDINA-FILHO, H. P.; FAZUOLI, L. C. Evolução e melhoria do cafeeiro. In: COLÓQUIO SOBRE CITOGENÉTICA E EVOLUÇÃO DE PLANTAS, 10., 1984, Piracicaba. **Anais...** Ribeirão Preto: Sociedade Brasileira de Genética, 1985. p. 215-234.

CARVALHO, A.; MONACO, L. C. Breeding of *Coffea arabica*. In: FERWERDA, F. P.; WIT, F. (ed.). **Coffee: Coffea arabica L. and Coffea canephora Pierre ex Froehner**. Wageningen: Agricultural University, 1969a. p. 198-215.

CARVALHO, A.; MONACO, L. C. **O melhoria do café arábica**. 1969b. p. 198-216.

CARVALHO, A.; MONACO, L. C.; ANTUNES FILHO, H. Melhoria do cafeeiro XV: variabilidade observada em progênies de café. **Bragantia**, v. 18, n. 26, p. 373-386, 1959. DOI: <https://doi.org/10.1590/S0006-87051959000100026>.

CARVALHO, A.; SCARANARI, H. J.; ANTUNES FILHO, H.; MONACO, L. C. Melhoria do cafeeiro: XXII - resultados obtidos no ensaio de seleções regionais de Campinas. **Bragantia**, v. 20, n. 30, p. 711-740, 1961. DOI: <https://doi.org/10.1590/S0006-87051961000100030>.

CILAS, C.; BOUHARMONT, P.; BOCCARA, M.; ESKES, A. B.; BARADAT, P. Prediction of genetic value for coffee production in *Coffea arabica* from a half-diallel with lines and hybrids. **Euphytica**, v. 104, p. 49-59, 1998. DOI: <https://doi.org/10.1023/A:1018635216182>.

CILAS, C.; MONTAGNON, C.; BERTRAND, B.; GODIN, C. Wood elasticity of several *Coffea canephora* clones. A new trait to be included in selection schemes. **Agronomie**, v. 20, p. 439-444, 2000.

COCKERHAM, C. C. An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. **Genetics**, v. 39, p. 859-882, 1954.

COCKERHAM, C. C. Covariances of relatives from self fertilization. **Crop Science**, v. 23, p. 1177-1180, 1983.

COCKERHAM, C. C. Estimation of genetic variances. **Statistical Genetics and Plant Breeding**, v. 982, p. 53-94, June 1963.

COMSTOCK, R. E.; ROBINSON, H. F.; HARVEY, P. H. A breeding procedure designed to make maximum use of both general and specific combining ability. **Journal of the American Society of Agronomy**, 41, 360-367, 1949.

COORS, J. G. Selection methodology and heterosis. In: COORS, J. C.; PANDEY, P. (ed.). **Genetics and exploitation of heterosis in crops**, 1999. p. 225-245. DOI: [10.2134/1999.geneticsandexploitation](https://doi.org/10.2134/1999.geneticsandexploitation).

CORBEIL, R. R.; SEARLE, S. R. Restricted maximum likelihood (REML) estimation of variance components in the mixed model. **Technometrics**, v. 18, n. 1, p. 31-38, 1976.

CUNNINGHAM, E. P. **Animal breeding theory**: internordic licencist course in quantitative genetics. Norges Landbrukshøgskole: Universitetsforlaget, 1969. 544 p.

DARWIN, C. **On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life**. Londres: John Murray, 1859.

DEMPFLE, L. Relation entre Blup (Best Linear Unbiased Prediction) et estimateurs bayesiens. **Annales de Génétique et Sélection Animale**, v. 9, n. 1, p. 27-32, 1977. DOI: [10.1051/1297-9686-9-1-27](https://doi.org/10.1051/1297-9686-9-1-27).

DHALIWAL, T. S. Progress report of the project coffee breeding in Puerto Rico: 1954-55. **Miscellaneous publication**, v. 53, n. 4, 1965.

DUREL, C. E. Gains génétiques attendus après sélection sur index en seconde génération d'amélioration du pin maritime. **Revue Forestière Française**, v. 44, n. 4, p. 341-355, 1992. DOI: [10.4267/2042/26331](https://doi.org/10.4267/2042/26331).

- FALCONER, D. S. **Introduction to quantitative genetics**. 3rd ed. Harlow: Longman Scientific Technical, 1989.
- FALL, C. A. **Optimisation des schémas de selection pour l'adaptation physiologique a la sécheresse de l'arachide (A) (A hypogaea L.)**. Bambey, Sénégal: Centre National de Recherches, 1994. 100 p. Disponível em: <http://intranet.isra.sn/greenstone/collect/tmsira/index/assoc/HASH0128.dir/doc.pdf>. Acesso em: 23 nov. 2022.
- FAZUOLI, L.C. Genética e melhoramento do cafeeiro. In: RENA, A. B.; MALAVOLTA, E.; ROCHA, M.; YAMADA, T. **Cultura do cafeeiro: fatores que afetam a produtividade**. Piracicaba: Associação Brasileira para Pesquisa da Potassa e do Fosfato, 1986. p. 86-113.
- FERNIE, L. M. **Coffee breeding in Tanganyika, Lyamungu Coffee Research Station**. [Tanzânia], 1965. 3 p. Datilografado
- FERRÃO, R. G.; FONSECA, A. F. A.; FERRÃO, M. A. G. Programas de melhoramento genético de café robusta no Brasil. In: SIMPÓSIO DE ATUALIZAÇÃO EM GENÉTICA E MELHORAMENTO DE PLANTAS, 3., 1999, Lavras. **Anais...** Lavras: Ed. da Ufla, 1999. p. 50-64.
- FERWERDA, F. P. Breeding of *Coffea canephora*. In: FERWERDA, F. P.; WIT, F. (ed.). **Coffee: Coffea arabica L. and Coffea canephora Pierre ex Froehner**. Wageningen: Agricultural University, 1969. p. 216-241.
- FISHER, R. A. Theory of statistical estimation. **Proceedings of Cambridge Philosophical Society**, v. 22, p. 700-725, 1925.
- FOULLEY, J. L.; GIANOLA, D.; IM, S. Genetic evaluation for discrete polygenic traits in animal breeding. In: GIANOLA, D.; HAMMOND, K. (ed.). **Advances in statistical methods for genetic improvement of livestock**. Berlin: Springer-Verlag, 1990. p. 361-409.
- FREY, K. J. The use of F<sub>2</sub> lines in predicting the performance of F<sub>3</sub> selections in two barley crosses. **Agronomy Journal**, v. 46, p. 541-544, 1954.
- GALLAIS, A. Covariances entre apparentes quelconques avec linkage et epistasie. I. Expression generale. **Annales Génétique et de Selection Animale**, v. 2, p. 281-310, 1970a.
- GALLAIS, A. Covariances entre apparentes quelconques avec linkage et epistasie. II. Evolution en regime d'autofecondation. **Annales Génétique et de Selection Animale**, v. 2, n. 3, p. 417-427, 1970b.
- GALLAIS, A. Modele pour l' etude des effets genetiques. **Ann Amél Plantes**, 17, n. 3, p. 229-242, 1967.
- GALLAIS, A. Optimization of recurrent selection on the phenotypic value of doubled haploid lines. **Theoretical and applied genetics**, v. 77, n. 4, p. 501-504, 1989.
- GALLAIS, A. Sélection combinée et plans de croisements en amélioration des plantes. **Ann. Amél. Plantes**, v. 30, n. 4, p. 411-441, 1980.
- GALTON, F. **Natural inheritance**. London: Mac Millan, 1889.
- GELFAND, A. E.; SMITH, A. F. M. Sampling-based approaches to calculating marginal densities. **Journal of the American Statistical Association**, v. 85, p. 398-409, 1990.
- GIANOLA, D.; FERNANDO, R. L. Bayesian methods in animal breeding theory. **Journal of Animal Science**, v. 63, p. 217-244, 1986.
- GILBERT, S. M. Selection within *Coffea arabica* in Tanganyika Territory. **The East African Agricultural Journal**, v. 4, n. 4, p. 249-253, 1939. DOI: <https://doi.org/10.1080/03670074.1939.11663870>.
- GILLOIS, M. **La relation d'identite en genetique**. 1964. 295 f. These (Docteur) – Universidade de Paris, Paris.
- GOFFINET, B. **Risque quadratique et sélection: quelques résultats appliqués à la sélection animale et végétale**. 1983. Thèse (Docteur Ingénieur) – Université Paul Sabatier, Toulouse.
- GOLDBERGER, A. S. Best linear unbiased prediction in the generalized linear regression model. **Journal of the American Statistical Association**, v. 57, p. 369-375. DOI: 10.1080/01621459.1962.10480665.
- GOULDEN, C. H. Problems in plant selection. In: INTERNATIONAL GENETICS CONGRESS, 7th, 1939, Edinburg. **Proceedings...** Edinburg: Edinburg University Press, 1939. p. 132-133.
- GRATTAPAGLIA, D.; RESENDE, M. D. V. Genomic selection in forest tree breeding. **Tree Genetics & Genomes**, v. 7, p. 241-255, 2011.
- HALDANE, J. B. S. The interaction of nature and nurture. **Annals of Eugenics**, n. 13, p. 197-205, 1946.
- HARLAN, H. V.; POPE, M. N. The use and value of back-crosses in small-grain breeding. **Journal of Heredity**, v. 13, n. 7, p. 319-322, July 1922.
- HAZEL, L. N.; LUSH, J. L. The efficiency of three methods of selection. **Journal of Heredity**, v. 33, p. 393-399, 1942.
- HENDERSON, C. R. Sire evaluation and genetic trends. **Journal Animal Science**, Jan, 1973. DOI: <https://doi.org/10.1093/ansci/1973.Symposium.10>.
- HENDERSON, C. R. Best linear unbiased estimation and prediction under a selection model. **Biometrics**, v. 31, n. 2, p. 423-447, June 1975.

- HENDERSON, C. R. Specific and general combining ability. In: GOWEN, J. W. **Heterosis**. New York: Hafner Publishing, 1952. p. 352.
- HENDERSON, C. R. Estimation of variances and covariances under multiple trait models. **Journal of Dairy Science**, v. 67, n. 7, p. 1581-1589, 1984.
- HENDERSON, C. R. Specific and general combining ability. In: GOWEN, J. M. (ed.). **Heterosis**, Ames: Iowa State College Press, 1952. p. 352-370.
- HENDERSON, C. R.; QUAAS, R. L. Multiple trait evaluation using relatives records. **Journal of Animal Science**, v. 43, n. 6, p. 1188-1197, 1976.
- JARDINE, R. Animal breeding and the estimation of genetic value. **Heredity**, v. 12, n. 4, p. 499-511, 1958. DOI: <https://doi.org/10.1038/hdy.1958.50>.
- JOHANSEN, W. **Über Erblichkeit in populationen und in reinen Linien**. Jena: Gustav Fisher, 1903. 289 p.
- LANDE, R. Neutral theory of quantitative genetic variance in an island model with local extinction and colonization. **Evolution**, v. 46, n. 2, p. 381-389, Apr. 1992. DOI: <https://doi.org/10.1111/j.1558-5646.1992.tb02046.x>.
- LE ROY, H. L. **Statistische methoden der populationsgenetik**. Basel: Birkhauser, 1960. 397 p.
- LEON, J. Hibridación intervarietal en *Coffea arabica* Trabajos realizados en el Instituto Interamericano de Ciências Agrícolas, Turrialba. In: INFORME de la Primera Reunión del Grupo Técnico de Trabajo de la FAO sobre producción y protección del café celebrada en Rio de Janeiro, Brasil, 23-30 de octubre de 1965. Roma, 1966. p. 24-25.
- LERNER, I. M. **The genetic basis of selection**. New York: Willey, 1958. 298 p.
- LEROY, T.; MONTAGNON, C.; CILAS, C.; CHARRIER, A.; ESKEs, A. B. Reciprocal recurrent selection applied to *Coffea canephora*. II. Estimation of genetic parameters. **Euphytica**, v. 74, p. 121-128, 1994. DOI: <https://doi.org/10.1007/BF00033776>.
- LINDLEY, D. V.; SMITH, A. F. Bayes estimates for the linear model. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 34, n. 1, p. 1-18, 1972. DOI: <https://doi.org/10.1111/j.2517-6161.1972.tb00885.x>.
- LUSH, J. L. A study of the accuracy of measurements of dairy cattle. **Journal of Agricultural Research**, v. 41, p. 37, Dec. 1930.
- LUSH, J. L. **Animal breeding plans**. 3rd ed. Ames, IA: Iowa State College Press, 1945.
- LUSH, J. L. **Animal breeding plans**. Ames: Collegiate Press, 1937.
- LUSH, J. L. Family merit and individual merit as bases for selection: Part I. **The American Naturalist**, v. 81, n. 799, 241-261, July-Aug. 1947.
- LUSH, J. L. Heritability of quantitative characters in farm animals. **Hereditas**, Dec. 1949. DOI: <https://doi.org/10.1111/j.1601-5223.1949.tb03347.x>.
- LUSH, J. L. Progeny test and individual performance as indicators of an animal's breeding value. **Journal of Dairy Science**, v. 18, n. 1, p. 1-19, 1935. DOI: [https://doi.org/10.3168/jds.S0022-0302\(35\)93109-5](https://doi.org/10.3168/jds.S0022-0302(35)93109-5).
- LUSH, J. L. **The genetics of population**. Ames: [Iowa State College], 1948.
- LUSH, J. L. **The genetics of populations: Special Report 94**. Ames: [Iowa State College], 1994.
- LUSH, J. L. The number of daughters necessary to prove a sire. **Journal of Dairy Science**, v. 14, p. 209-220, May 1931. DOI: [https://doi.org/10.3168/jds.S0022-0302\(31\)93466-8](https://doi.org/10.3168/jds.S0022-0302(31)93466-8).
- LUSH, J. L.; CHRISTENSEN, F. W.; WILSON, C. V.; BLACK, W. H. The accuracy of cattle weights. **Journal of Agricultural Research**, v. 36, p. 551-580, 1928.
- MALECOT, G. **Les mathematiques de l'heredite**. Paris: Masson, 1948. 80 p.
- MALECOT, G. **Probabilites et heredite**. Paris: Presses Universitaires de France, 1966.
- MENDES, A. N. G. Métodos de melhoramento aplicados na cultura do cafeeiro. In: SIMPÓSIO DE ATUALIZAÇÃO EM GENÉTICA E MELHORAMENTO DE PLANTAS, 3., 1999, Lavras. **Anais...** Lavras: Ed. da Ufla, 1999. p. 18-35.
- MONTAGNON, C.; CUBRY, P.; LEROY, T. Amélioration génétique du caféier *Coffea canephora* Pierre: connaissances acquises, stratégies et perspectives. **Cahiers Agricultures**, v. 21, p. 143-53, 2012. DOI: [10.1684/agr.2012.055](https://doi.org/10.1684/agr.2012.055).
- MORALES, A. **Análise estatística do ensaio de variedades de café**: Parte II. v. 11, n. 1-3, p. 4549, jan-mar. 1951. DOI: <https://doi.org/10.1590/S0006-87051951000100006>.
- MRODE, R. A. **Linear models for the prediction of animal breeding values**. Wallingford: CAB International, 2014.
- NANSON, A. Tables comparatives de l'efficiencia de la selección individuelle inter-familles et intra-familles par rapport à la sélection combinée. **Biométrie-Praximétrie**, v. 14, 1-2, p. 1-11, 1974.
- NARAIN, P. **Statistical genetics**. New York: John Wiley, 1990.

- NARAIN, P. Statistics-genetic considerations in longitudinal data analysis. **Indian Journal of Agricultural Science**, v. 62, n. 2, p. 138-148, Jan. 2008.
- NEWMAN, L. H. **Plant breeding in scandinavia**. Ottawa: Canadian Seed Grower Association, 1912.
- NILSSO EHLE, H. **Einige ergebnisse von kreuzungen bei hafer und weizen**. Berlingska Boktryckeriet, 1908.
- NUNES, J. A. R.; MORETO, A. L.; RAMALHO, M. A. P. Using genealogy to improve selection efficiency of pedigree method. **Scientia Agricola**, v. 65, n. 1, p. 25-30, Feb. 2008. DOI: <https://doi.org/10.1590/S0103-90162008000100004>.
- OLLIVIER, L. Les objectifs de la sélection porcine en France. **Journées Recherch Porcine en France**, Fév. 1981.
- OLLIVIER, L. **Elements de genétique quantitative**. Paris: Institut National de La Recherche Agronomique, 2002. 184 p.
- OSBOURNE, R. The use of sire and dam family averages in increasing the efficiency of selective breeding under a hierarchical mating system. **Heredity**, v. 11, p. 93-116, Apr. 1957. DOI: <https://doi.org/10.1038/hdy.1957.7>.
- PATTERSON, H. D.; THOMPSON, R. Recovery of inter-block information when block sizes are unequal. **Biometrika**, v. 58, n. 3, p. 545-554, Dec. 1971. DOI: <https://doi.org/10.1093/biomet/58.3.545>.
- PAULA, T.; BRASILEIRO, B. P.; CURSI, D. E.; FREITAS, E. G.; SANTOS, J. M.; RESENDE, M. D. V.; KIMBENG, C.; BARBOSA, M. H. Establishment of gene pools for systematic heterosis exploitation in Sugarcane. **Agronomy Journal**, May 2020. DOI: <https://doi.org/10.1002/agj2.20299>.
- PEDROZO, C. A.; BENITES, F. R. G.; BARBOSA, M. H. P.; RESENDE, M. D. V.; SILVA, F. L. Eficiência de índices de seleção utilizando a metodologia REML/Blup no melhoramento da cana-de-açúcar. **Scientia Agraria**, v. 10, n. 1, p. 31-36, 2009.
- QUAAS, R. L.; POLLAK, E. J. Mixed model methodology for farm and ranch beef cattle testing programs. **Journal of Animal Science**, v. 51, n. 6, p. 1277-1287, Dec. 1980. DOI: <https://doi.org/10.2527/jas1981.5161277x>.
- RAMALHO, M. A. P.; ABREU, Â. de F. B.; SANTOS, J. B. de; NUNES, J. A. R. **Aplicações da genética quantitativa no melhoramento de plantas autógamas**. Lavras: Ed. da Ufla, 2012.
- RAMALHO, M. A. P.; VENCOVSKY, R. Estimação dos componentes da variância genética em plantas autógamas. **Ciência e Prática**, v. 2, p. 117-140, 1978.
- RAMALHO, M. A. P.; CARVALHO, B. L.; NUNES, J. A. R. Perspectives for the use of quantitative genetics in breeding of autogamous plants. **ISRN Genetics**, Art. 718127, p. 1-6, 2013. DOI: <http://dx.doi.org/10.5402/2013/718127>.
- RAMALHO, M. A. P.; GONÇALVES, F. M. A.; SOUZA SOBRINHO, F. Emprego da seleção recorrente no melhoramento do cafeeiro. In: SIMPÓSIO DE ATUALIZAÇÃO EM GENÉTICA E MELHORAMENTO DE PLANTAS, 3., 1999, Lavras. **Anais...** Lavras: Ed. da Ufla, 1999. p. 66-81.
- RAPOSO, F. V.; RAMALHO, M. A. P.; RIBEIRO, P. H. E. Alterations in heterosis of maize populations derived from single-cross hybrids after a reciprocal recurrent selection. **Crop Breeding and Applied Biotechnology**, v. 4, n. 1, p. 74-80, Mar. 2004.
- RESENDE, M. D. V. de. Avanços da genética biométrica florestal. In: ENCONTRO SOBRE TEMAS DE GENÉTICA E MELHORAMENTO: GENÉTICA BIOMÉTRICA VEGETAL. **Anais...** Piracicaba: Esalq, 1967. p. 20-46. Editado por: Gerhard Bandel, Natal Antonio Vello, José Branco de Miranda Filho.
- RESENDE, M. D. V. de. Correções nas expressões do progresso genético com seleção em função da amostragem finita dentro de famílias e populações e implicações no melhoramento florestal. **Boletim de Pesquisa Florestal**, n. 22/23, p. 61-77, jan./dez. 1991.
- RESENDE, M. D. V. de. **Genética biométrica e estatística no melhoramento de plantas perenes**. Brasília, DF: Embrapa Informação Tecnológica, 2002.
- RESENDE, M. D. V. de. **Genética quantitativa e de populações**. Visconde do Rio Branco: Suprema, 2015.
- RESENDE, M. D. V. de. **Inferência bayesiana e simulação estocástica (amostragem de Gibbs) na estimação de componentes de variância e valores genéticos em plantas perenes**. Colombo: Embrapa Florestas, 2000.
- RESENDE, M. D. V. de. **Matemática e estatística na análise de experimentos e no melhoramento genético**. Colombo: Embrapa Florestas. 561 p.
- RESENDE, M. D. V. de. Melhoramento de essências florestais. In: BOREM, A. **Melhoramento de espécies cultivadas**. Viçosa, MG: Ed. da UFV, 1999. p. 589-647.
- RESENDE, M. D. V. de. **Métodos estatísticos ótimos na análise de experimentos de campo**. Colombo: Embrapa Florestas, 2004.
- RESENDE, M. D. V. de. **Predição de valores genéticos, componentes de variância, delineamentos de cruzamento e estrutura de populações no melhoramento florestal**. Curitiba: Ed. da Universidade Federal do Paraná, 1999.
- RESENDE, M. D. V. de; ALVES, R. S. Linear, generalized, hierarchical, bayesian and random regression mixed models in genetics/genomics in plant



breeding. **Functional Plant Breeding Journal**, v. 2, n. 2, n. 11, p. 121-152, July/Dec. 2020. DOI: <https://doi.org/10.35418/2526-4117/v2n2a1>.

RESENDE, M. D. V. de; ALVES, R. S. Statistical significance, selection accuracy and experimental precision in plant breeding. **Crop Breeding and Applied Biotechnology**, v. 22, n. 3, Art. e42712238, 2022.

RESENDE, M. D. V. de; ARAÚJO, A. J. de; SAMPAIO, P. de T. B.; WIECHETECH, M. S. S. Acurácia seletiva, intervalos de confiança e variâncias de ganhos genéticos associados a 22 métodos de seleção em *Pinus caribaea* var. *hondurensis*. **Floresta**, v. 25, n. 1/2, p. 3-16, 1995. DOI: <http://dx.doi.org/10.5380/rf.v25i12.6444>.

RESENDE, M. D. V. de; AZEVEDO, C. F.; SILVA, F. F.; NASCIMENTO, M.; GOIS, I. B.; ALVES, R. S. **Modelos hierárquicos generalizados lineares mistos (HGLMM), Máxima verossimilhança hierárquica (HIML) e HG-Blup**: Unificação das três classes de inferência (frequentista, fisheriana e bayesiana) na análise estatística em biometria e genética. Visconde do Rio Branco: Suprema, 2018. 150 p.

RESENDE, M. D. V. de; BARBOSA, M. H. P. **Melhoramento genético de plantas de propagação assexuada**. Colombo: Embrapa Florestas, 2006. 130 p.

RESENDE, M. D. V. de; Delineamento de experimentos de seleção para a maximização da acurácia seletiva e progresso genético. **Revista Árvore**, v. 19, p. 479-500, 1995.

RESENDE, M. D. V. de; DUARTE, J. B. Precisão e controle de qualidade em experimentos de avaliação de cultivares. **Pesquisa Agropecuária Tropical**, v. 37, n. 3, p. 182-194, 2007.

RESENDE, M. D. V. de; FURLANI JÚNIOR, E.; MORAES, M. L. T. de; FAZUOLI, L. C. Estimativas de parâmetros genéticos e predição de valores genotípicos no melhoramento do cafeeiro pelo procedimento REML/Blup. **Bragantia**, v. 60, n. 3, p. 185-193, 2001. DOI: <https://doi.org/10.1590/S0006-87052001000300005>.

RESENDE, M. D. V. de; HIGA, A. R. Estratégias de melhoramento para *Eucalyptus* visando a seleção de híbridos. **Boletim de Pesquisa Florestal**, n. 21, p. 49-60, dez. 1990.

RESENDE, M. D. V. de; HIGA, A. R. Maximização da eficiência da seleção em testes de progênies de *Eucalyptus* através da utilização de todos os efeitos do modelo matemático. **Pesquisa Florestal Brasileira**, n. 28/29, p. 37-55, 1994.

RESENDE, M. D. V. de; HIGA, A. R.; LAVORANTI, O. J. Regressão geno-fenotípica multivariada e maximização do progresso genético em programas de melhoramento de *Eucalyptus*. **Boletim de Pesquisa Florestal**, n. 28/29, p. 57-71, jan./dez. 1994.

RESENDE, M. D. V. de; LOPES, P. S.; SILVA, R. L.; PIRES, I. E. Seleção genômica ampla (GWS) e maximização da eficiência do melhoramento genético. **Pesquisa Florestal Brasileira**, v. 56, p. 63-78, 2008. Disponível em: <https://pfb.cnpf.embrapa.br/pfb/index.php/pfb/article/view/63>. Acesso em: 26 jul. 2022.

RESENDE, M. D. V. de; MEDRADO, M. J. S. Aspectos metodológicos no melhoramento genético de *Leucena leucocephala*, uma espécie florestal autógama. In: CONGRESSO BRASILEIRO SOBRE SISTEMAS AGROFLORESTAIS, 1., 1994, Porto Velho. **Anais...** Colombo: EMBRAPA-CNPf, 1994. p. 233-348. (EMBRAPA-CNPf. Documentos, 27).

RESENDE, M. D. V. de; OLIVEIRA, E. B.; HIGA, A. R. Utilização de índices de seleção no melhoramento do *eucalyptus*. **Boletim de Pesquisa Florestal**, n. 21, p. 1-13, 1990.

RESENDE, M. D. V. de; PEREZ, J. R. H. R. Melhoramento animal: predição de valores genéticos pelo modelo animal Blup em bovinos de leite, bovinos de corte, ovinos e suínos. **Archives of Veterinary Science**, v. 4, n. 1, p. 17-30, 1999. DOI: <http://dx.doi.org/10.5380/avs.v4i1.3772>.

RESENDE, M. D. V. de; RAMALHO, M. A. P.; CARNEIRO, P. C. S.; CARNEIRO, J. E. S.; BATISTA, L. G.; GOIS, I. B. Selection index with parents, populations, progenies, and generations effects in autogamous plant breeding. **Crop Science**, v. 56, p. 530-546, Mar. 2016. DOI: <https://doi.org/10.2135/cropsci2015.05.0303>.

RESENDE, M. D. V. de; RAMALHO, M. A. P.; GUILHERME, S. R.; ABREU, A. F. B. Multigeneration index in the within progenies bulk method for breeding of self-pollinated plants. **Crop Science**, v. 55, p. 1202-1211, May 2015. DOI: <https://doi.org/10.2135/cropsci2014.08.0580>.

RESENDE, M. D. V. de; SILVA, F. F.; AZEVEDO, C. F. **Estatística matemática, biométrica e computacional**: modelos mistos, multivariados, categóricos e generalizados (REML/Blup), Inferência bayesiana, regressão aleatória, seleção genômica, qtl-gwas, estatística espacial e temporal, competição, sobrevivência. Viçosa, MG: Ed. da UFV, 2014. 882 p.

RESENDE, M. D. V. de; SILVA, H. D. Estratégia de melhoramento genético para erva-mate baseada no coeficiente de repetibilidade. In: CONGRESSO FLORESTAL E DO MEIO AMBIENTE DO PARANÁ, 1991, Curitiba. **Anais...** Curitiba: Sociedade Paranaense de Engenheiros Florestais, 1991. p. 241-251.

RESENDE, M. D. V. de; THOMPSON, R. **Multivariate spatial statistical analysis of multiple experiments and longitudinal data**. Colombo: Embrapa Florestas, 2003.

- RESENDE, M. D. V. de; THOMPSON, R.; WELHAM, S. J. Multivariate spatial statistical analysis of longitudinal data in perennial crops. **Revista de Matemática e Estatística**, v. 24, n. 1, p. 147-169, 2006.
- ROBERTSON, A. Prediction equations in quantitative genetics. **Biometrics**, v. 11, n. 1, p. 95-98, Mar. 1955.
- RONNINGEN, K. Some properties of the selection index derived by "Henderson's mixed model method". **Zeitschrift für Tierzüchtung und Züchtungsbiologie**, v. 88, p. 186, 1971. DOI: <https://doi.org/10.1111/j.1439-0388.1971.tb01365.x>.
- RONNINGEN, K.; VAN VLECK, L. D. Selection index theory with practical applications. In: CHAPMAN, A. B. **General and Quantitative Genetics**: World Animal Science: [Part] A4. 1985. p. 187-222.
- SAKIYAMA, N. S.; PEREIRA, A. A.; ZAMBOLIM, L. Melhoramento do café arábica. In: BORÉM, A. (ed.). **Melhoramento de espécies cultivadas**. Viçosa, MG: Imprensa Universitária, 1999. p. 189-204.
- SEARLE, S. R. Progeny-tests of sire and son. **Journal of Dairy Science**, v. 47, n. 4, p. 414, Apr. 1964.
- SEARLE, S. R. The efficiency of ancestor records in animal selection. **Heredity**, v. 18, p. 351-360, 1963. DOI: <https://doi.org/10.1038/hdy.1963.35>.
- SERA, T. Desafios no melhoramento genético do cafeeiro. In: SIMPÓSIO DE ATUALIZAÇÃO EM GENÉTICA E MELHORAMENTO DE PLANTAS, 2., 1998, Lavras. **Anais...** Lavras: Ed. da Ufla, 1998. p. 105-122.
- SERA, T. **Estimação dos componentes da variância e do coeficiente de determinação genotípica da produção de grãos de café (Coffea arabica L.)**. 1981. 62 f. Dissertação (Mestrado) – Escola Superior de Agricultura "Luiz de Queiroz", Piracicaba. DOI: <https://doi.org/10.11606/D.11.1981.tde-20220208-045248>.
- SERA, T. **Possibilidades de emprego de seleção nas colheitas iniciais de café (Coffea arabica, L. cv. Acaiá)**. 1987. 147 f. Tese (Doutorado) – Escola Superior de Agricultura "Luiz de Queiroz", Piracicaba. DOI: <https://doi.org/10.11606/T.11.2020.tde-20200111-144842>.
- SERA, T.; ALVES, S. J. Melhoramento genético de plantas perenes. In: DESTRO, D.; MONTALVÁN, R. (ed.). **Melhoramento genético de plantas**. Londrina: Ed. da UEL, 1999. p. 369-422.
- SHEWHART, W. A. Correction of data for errors of measurement. **The Bell System Technical Journal**, v. 5, n. 1, p. 11-26, Jan. 1926. DOI: [10.1002/j.1538-7305.1926.tb00106.x](https://doi.org/10.1002/j.1538-7305.1926.tb00106.x).
- SIMEÃO-RESENDE, R. M.; CASLER, M. D.; RESENDE, M. D. V. de. Genomic selection in forage breeding: accuracy and methods. **Crop Science**, v. 54, p. 143-156, Jan. 2014. DOI: [10.2135/cropsci2013.05.0353](https://doi.org/10.2135/cropsci2013.05.0353).
- SIMEÃO-RESENDE, R. M.; CASLER, M. D.; RESENDE, M. D. V. de. Selection methods in forage breeding: a quantitative appraisal. **Crop Science**, v. 53, p. 1925-1936, 2013. DOI: [10.2135/cropsci2013.03.0143](https://doi.org/10.2135/cropsci2013.03.0143).
- SOUSA, T. V.; CAIXETA, E. T.; ALKIMIM, E. R.; BERTRAND, B. Early selection enabled by the implementation of genomic selection in *Coffea arabica* breeding. **Front Plant Science**, v. 9, p. 1-12, 2017. DOI: <https://doi.org/10.3389/fpls.2018.01934>.
- STEVENS, W. L. Análise estatística do ensaio de variedades de café. **Bragantia**, v. 9, p. 103-23, 1949. DOI: <https://doi.org/10.1590/S0006-87051949000200002>.
- SUKHATME, P. V.; NARAIN, P. A possible genetic interpretation of auto-regulatory mechanism in models for protein deficiency. **Proceedings of the Indian National Science Academy-Part B: Biological Sciences**, v. 48, n. 6, p. 748-754, 1982.
- TEMPELMAN, R. J.; GIANOLA, D. Marginal maximum likelihood estimation of variance components in Poisson mixed models using Laplacian integration. **Genetics Selection Evolution**, v. 25, 305-319, 1993.
- VANRADEN, P. M.; WIGGANS, G. R. Derivation, calculation, and use of national animal model information. **Journal of Dairy Science**, v. 74, n. 8, p. 2737-2746, Aug. 1991. DOI: [https://doi.org/10.3168/jds.S0022-0302\(91\)78453-1](https://doi.org/10.3168/jds.S0022-0302(91)78453-1).
- VENCOVSKY, R.; BARRIGA, P. **Genética biométrica no fitomelhoramento**. Ribeirão Preto: Sociedade Brasileira de Genética, 1992. 496 p.
- VIANA, A. P.; RESENDE, M. D. V. **Genética quantitativa no melhoramento de fruteiras**. Rio de Janeiro: Interciência, 2014.
- VILMORIN, L. L. Notes sur la creation d'une nouvelle race de betterave a sucre. Considerations sur l'heredite des vegetaux. **Comptes Rendus des séances hebdomadaires de l'Académie des Sciences**, v. 43, p. 871-874, 1856.
- VINCOURT, P.; GALLAIS, A. Sur la recherche de critères de sélection: La régression géno-phenotypique. **Agronomie**, v. 3, n. 9, p. 827-830, 1983.
- WALYARO, D. J.; VAN DER VOSSEN, H. A. M. Early determination of yield potential in arabica coffee by applying index selection. **Euphytica**, v. 28, p. 465-472, 1979. DOI: <https://doi.org/10.1007/BF00056606>.
- WANG, C. S.; RUTLEDGE, J. J.; GIANOLA, D. Marginal inferences about variance components in a mixed linear

model using Gibbs sampling. **Genetics Selection Evolution**, v. 25, p. 41- 62, 1993.

WEBER, W. E. Selection in segregating generations of autogamous species. I. Selection response for combined selection. **Euphytica**, v. 31, p. 493-50, Sept. 1982. DOI: <https://doi.org/10.1007/BF00021668>.

WRICKE, G.; WEBER, W. E. **Quantitative genetics and selection in plant breeding**. Berlin: Walter de Gruyter, 1986. 406 p.

WRIGHT, J. W. **Genetics of forest tree improvement**. Rome: FAO, 1962. 399 p.

WRIGHT, S. An analysis of local variability of flower color in *Linanthus parryae*. **Genetics**, v. 28, n. 2, p. 139-156, Mar. 1943. DOI: 10.1093/genetics/28.2.139.

WRIGHT, S. Fisher and Ford on "the Sewall Wright effect". **American Scientist**, v. 39, n. 3, p. 452-479, July, 1951.

WRIGHT, S. The genetical structure of populations. **Annals of Eugenics**, v. 15, p. 323-354, 1951. DOI: <http://dx.doi.org/10.1111/j.1469-1809.1949.tb02451.x>.



MINISTÉRIO DA  
AGRICULTURA E  
PECUÁRIA

