

**TIAGO DE SOUZA MARÇAL**

**ANÁLISES DIALÉLICAS ENVOLVENDO AS CAPACIDADES GERAL E ESPECÍFICA  
RECÍPROCAS E ENSAIOS MULTIAMBIENTES E MULTIGERAÇÕES VIA  
MODELOS MISTOS**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Doctor Scientiae*.

Orientador: José Eustáquio de Souza Carneiro

Coorientador: Pedro Crescêncio S. Carneiro

**VIÇOSA - MINAS GERAIS  
2019**

**Ficha catalográfica preparada pela Biblioteca Central da Universidade  
Federal de Viçosa - Câmpus Viçosa**

T

M298a  
2019  
Marçal, Tiago de Souza, 1992-  
Análises dialélicas envolvendo as capacidades geral e  
específica recíprocas e ensaios multiambientes e multigerações  
via modelos mistos / Tiago de Souza Marçal. – Viçosa, MG,  
2019.

138f. : il. ; 29 cm.

Orientador: José Eustáquio de Souza Carneiro.  
Tese (doutorado) - Universidade Federal de Viçosa.  
Inclui bibliografia.

1. Melhoramento genético. 2. Estatística matemática.  
3. Análise multivariada . I. Universidade Federal de Viçosa.  
Departamento de Fitotecnia. Programa de Pós-Graduação em  
Genética e Melhoramento. II. Título.

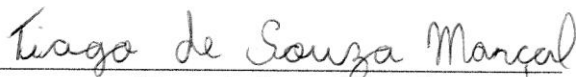
CDD 22 ed. 631.52

**TIAGO DE SOUZA MARÇAL**

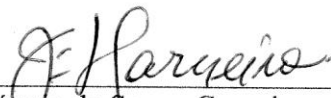
**ANÁLISES DIALÉLICAS ENVOLVENDO AS CAPACIDADES GERAL E ESPECÍFICA  
RECÍPROCAS E ENSAIOS MULTIAMBIENTES E MULTIGERAÇÕES VIA  
MODELOS MISTOS**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Doctor Scientiae*.

APROVADA: 27 de setembro de 2019.



\_\_\_\_\_  
Tiago de Souza Marçal  
Autor



\_\_\_\_\_  
José Eustáquio de Souza Carneiro  
Orientador

A Deus, aos meus pais, Sebastião e Marcelina, meu irmão Fábio e minha esposa Dinorah pelo apoio incondicional para que eu chegasse até aqui.

**DEDICO**

## **AGRADECIMENTOS**

A Deus, por me conceder a vida e persistência para superar os desafios que surgem ao longo da mesma.

Aos meus pais Sebastião Marçal e Marcelina Maria Marçal, ao meu irmão Fábio de Souza Marçal, à minha cunhada Rogéria Lourdes da Silva Marçal, aos meus sobrinhos Pedro Henrique da Silva Marçal e Gabriel da Silva Marçal, ao grande amigo José Henrique Soler Guilhen, pelo apoio, dedicação, compreensão nos momentos difíceis, e pelo amor.

A minha amada esposa Dinorah Moraes de Souza Marçal e toda sua família, pela compreensão nos momentos difíceis, atenção e motivação.

A Universidade Federal de Viçosa e ao Programa de Pós-Graduação em Genética e Melhoramento, pela oportunidade de realização do doutorado.

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa de estudos.

Ao meu orientador José Eustáquio de Souza Carneiro, pelo conhecimento transmitido, pela disponibilidade e dedicação nas orientações, pelas críticas, sugestões, pelos conselhos na vida profissional e pessoal e pela escola de melhoramento vegetal.

Ao meu coorientador professor Pedro Crescêncio Souza Carneiro, pelo conhecimento transmitido, pela disponibilidade e dedicação nas orientações, pelas críticas, sugestões e pelos conselhos na vida profissional e pessoal e pelos ensinamentos sobre docência.

Ao professor Rodrigo Lima de Oliveira, pelo conhecimento transmitido, pela disponibilidade, apoio e pela oportunidade de aplicação de ferramentas biométricas no melhoramento do milho.

Ao pesquisador Lauro José Moreira Guimarães e toda a equipe da EMBRAPA Milho e Sorgo pela visão empresarial do melhoramento do milho e pelos desafios propostos que alavancaram meu entendimento sobre a aplicação da metodologia de modelos mistos no melhoramento vegetal.

Aos amigos do Laboratório de Biometria, do Programa Feijão e da UFV pela troca de conhecimentos e pelos momentos de descontração.

A todos os professores que tive em minha vida, especialmente aos professores Ruimário Inácio Coelho, Adésio Ferreira e Marcia da Silva Ferreira por terem me iniciado na carreira científica.

A todos, que de forma direta ou indireta, contribuíram para a realização deste trabalho e para a minha formação profissional.

**MUITO OBRIGADO!**

## RESUMO

MARÇAL, Tiago de Souza, D.Sc., Universidade Federal de Viçosa, setembro de 2019. **Análises dialéticas envolvendo as capacidades geral e específica recíprocas e ensaios multiambientes e multigerações via modelos mistos.** Orientador: José Eustáquio de Souza Carneiro. Coorientador: Pedro Crescêncio Souza Carneiro.

Na análise dialética, utilizando modelos mistos, é possível estimar as variâncias da capacidade geral de combinação (CGC), capacidade específica de combinação (CEC), CGC recíproca (RCGC), CEC recíproca (RCEC), mesmo para conjuntos de dados desbalanceados. Entretanto, devido à interação genótipos por ambientes ( $G \times A$ ), estas informações podem mudar de um ambiente para outro. Para algumas espécies autógamas, devido à interação  $G \times A$  e a limitação de sementes  $F_1$ , os dialelos são conduzidos em experimentos sequenciais denominados aqui de *multi-environment and multi-generation trials* (MEGT). A abordagem mais recente sugere que a forma mais lógica, do ponto de vista genético, para modelar dados sob interação  $G \times A$  é tratar um mesmo caractere avaliado em uma série de ambientes como caracteres diferentes. Neste sentido, assume-se uma estrutura de (co)variância genotípica capaz de acomodar a heterogeneidade de variâncias e covariâncias. A estimação dos parâmetros de variância na análise dialética via modelos mistos pode ser realizada maximizando-se a função de verossimilhança residual via algoritmos numéricos como: *expectation maximization* (EM), *average information* (AI) e EM-AI. Assim, o objetivo deste trabalho foi descrever os detalhes para se realizar a análise dialética envolvendo as capacidades geral e específica recíprocas, bem como análise dialética envolvendo ensaios multiambientes e multigerações via modelos mistos usando os algoritmos EM, AI e EM-AI. O primeiro capítulo, publicado no periódico *Crop Science* 59(4):1494-1503, 2019, (Estimação das variâncias dos efeitos recíprocos de capacidade geral e específica de combinação por meio do algoritmo EM-AI) teve como foco a avaliação da eficiência computacional dos algoritmos EM, AI e EM-AI na estimação de componentes de variância para três modelos dialéticos (Modelo 1 de Griffing, Modelo 1 de Griffing + RCGC e Modelo 3, um modelo dialético geral com os efeitos de CGC, CEC, RCGC, RCEC e parentais). Os resultados revelaram que o algoritmo AI falhou em convergir para todos os modelos, enquanto os demais algoritmos convergiram normalmente (EM e EM-AI). Dentre os dois últimos, o algoritmo EM-AI foi o mais eficiente computacionalmente e o modelo dialético geral (Modelo 3) forneceu estimativas mais precisas dos componentes de variância. Assim, o algoritmo EM-AI e o modelo geral têm

potencial para uso em análises dialélicas no melhoramento de plantas. O segundo capítulo (Análise dialélica multiambiental e multigerações no melhoramento de plantas autógamas) focou na avaliação da eficiência computacional dos algoritmos EM, AI e EM-AI e escolha de modelo para a análise dialélica em MEGT. A combinação entre diferentes estruturas de (co)variância possibilitou a definição de quatro modelos. Para avaliar o potencial das populações segregantes foi utilizado um índice baseado na distância genótipos-ideótipo (IGI). Os resultados mostraram que os algoritmos AI e EM-AI foram os mais eficientes computacionalmente quando a convergência foi verificada. Além disso, o modelo que assume heterogeneidade de variâncias para os efeitos genéticos e não genético, bem como correlação uniforme para os efeitos genéticos foi o mais adequado. Assim, a modelagem de estruturas de (co)variância na análise dialélica em MEGT via algoritmo AI associada à utilização do IGI têm potencial para uso nos programas de melhoramento de plantas autógamas.

Palavras-chave: Capacidade combinatória. Máxima verossimilhança. Estruturas de (co)variância.

## ABSTRACT

MARÇAL, Tiago de Souza, D.Sc., Universidade Federal de Viçosa, September, 2019. **Diallel analyzes involving the reciprocal general and specific combining ability and multi-environment and multi-generation trials via mixed models.** Adviser: José Eustáquio de Souza Carneiro. Co-Adviser: Pedro Crescêncio Souza Carneiro.

In diallel analysis, using mixed models, it is possible to estimate the variance components of general combining ability (GCA), specific combining ability (SCA), reciprocal GCA (RGCA), and reciprocal SCA (RSCA), even in unbalanced datasets. However, due to the genotype by environment (G×E) interaction, this information may change through environments. For some autogamous species, due to the G×E interaction and limitation of F<sub>1</sub> seeds, the diallel trials are sequential *i.e.* in multi-environment and multi-generations trials (MEGT). The most recent approach suggests that the most logical way to model data under G×E interaction, from the genetic point of view, is to treat the same trait evaluated through several environments as different traits. In this context, a genotypic (co)variance structure is assumed to modeling the heterogeneity of variances and covariances. The estimation of variance parameters in diallel analysis via mixed models can be performed by maximizing the restricted likelihood by numerical algorithms such as: expectation maximization (EM), average information (AI), and EM-AI. Thus, the objective of this work was to describe the details for performing the diallel analysis involving the reciprocal general and specific combining ability, as well as diallel analysis involving multi-environment and multi-generation trials via mixed models using the EM, AI and EM-AI algorithms. The first chapter, published in the periodical Crop Science 59(4):1494-1503, 2019, (Estimation of variance for reciprocal general and specific combining ability effects by EM-AI algorithm) focused on the evaluation of computational efficiency of the EM, AI and EM-AI algorithms in the estimation of variance components for three diallel models (Griffing's Model 1, Griffing's Model 1 + RGCA, and Model 3, a general diallel model with the effects of GCA, SCA, RGCA, RSCA, and parents). The results revealed that the AI algorithm failed to converge for the three diallel models, while the other algorithms converged normally (EM and EM-AI). Of the last two, the EM-AI algorithm was the most computationally efficient and the general diallel model (Model 3) provided more accurate estimates of variance components. Thus, the EM-AI algorithm and the general model have potential for use in diallel analysis in plant breeding. The second chapter (Diallel analysis for multi-environment and multi-

generation trials in the autogamous plant breeding) focused on the computational efficiency evaluation of the EM, AI, and EM-AI algorithms and choice of model for the MEGT diallel analysis. The combination of different (co)variance structures allowed the fit of four models. To assess the potential of segregating populations, an index based on genotypes-ideotype distance was used (GID). The results showed that the AI and EM-AI algorithms were the most computationally efficient when convergence was verified. In addition, the model that assumes variance heterogeneity for genetic and non-genetic effects, as well as a uniform correlation for genetic effects, was the most appropriate. Thus, the modeling of the (co)variance structures of the diallel analysis of MEGT data via AI algorithm associate to the GID, have potential for use in diallel analysis in autogamous plant breeding.

Keywords: Combining ability. Maximum likelihood. (Co)variance structures.

## SUMÁRIO

<b>INTRODUÇÃO GERAL .....</b>	<b>13</b>
<b>REFERÊNCIAS.....</b>	<b>16</b>
<b>CAPÍTULO 1 .....</b>	<b>21</b>
<b>ESTIMATION OF VARIANCE FOR RECIPROCAL GENERAL AND SPECIFIC     COMBINING ABILITY EFFECTS BY EM-AI ALGORITHM.....</b>	<b>21</b>
<b>ABSTRACT .....</b>	<b>22</b>
<b>1. INTRODUCTION .....</b>	<b>23</b>
<b>2. MATERIALS AND METHODS.....</b>	<b>26</b>
<b>2.1. MODEL .....</b>	<b>27</b>
<b>2.2. RESIDUAL LOG-LIKELIHOOD .....</b>	<b>29</b>
<b>2.3. ALGORITHMS AND COMPUTATIONAL STRATEGIES .....</b>	<b>31</b>
<b>2.4. EXPECTATION MAXIMIZATION.....</b>	<b>31</b>
<b>2.5. AVERAGE INFORMATION.....</b>	<b>32</b>
<b>2.6. EM-AI ALGORITHM.....</b>	<b>33</b>
<b>2.7. DATASETS AND ANALYSES OF EXAMPLES.....</b>	<b>35</b>
<b>3. RESULTS .....</b>	<b>36</b>
<b>3.1. CONVERGENCE HISTORY AND COMPUTATIONAL EFFICIENCY OF THE     ALGORITHMS.....</b>	<b>36</b>
<b>3.2. VARIANCE COMPONENTS .....</b>	<b>41</b>
<b>4. DISCUSSION.....</b>	<b>42</b>
<b>4.1. CONVERGENCE HISTORY AND COMPUTATIONAL EFFICIENCY OF THE     ALGORITHMS.....</b>	<b>42</b>
<b>4.1. VARIANCE COMPONENTS .....</b>	<b>43</b>
<b>5. CONCLUSIONS.....</b>	<b>45</b>
<b>6. REFERENCES .....</b>	<b>45</b>

<b>7. SUPPLEMENTAL MATERIALS .....</b>	<b>50</b>
<b>CAPÍTULO 2 .....</b>	<b>77</b>
<b>ANÁLISE DIALÉLICA MULTIAMBIENTES E MULTIGERAÇÕES NO     MELHORAMENTO DE PLANTAS AUTÓGAMAS.....</b>	<b>77</b>
<b>RESUMO.....</b>	<b>78</b>
<b>1. INTRODUÇÃO .....</b>	<b>80</b>
<b>2. MATERIAL E MÉTODOS .....</b>	<b>84</b>
<b>2.1. FORMA GERAL DO MODELO PARA ANÁLISE DIALÉLICA EM MEGT .....</b>	<b>84</b>
<b>2.2. ESTIMAÇÃO DAS MATRIZES DE (CO)VARIÂNCIA .....</b>	<b>86</b>
<b>2.2.1. ALGORITMOS E ESTRATÉGIAS COMPUTACIONAIS .....</b>	<b>89</b>
<b>2.2.2. ALGORITMO EM.....</b>	<b>90</b>
<b>2.2.3. ALGORITMO AI.....</b>	<b>91</b>
<b>2.2.4. ALGORITMO EM-AI.....</b>	<b>93</b>
<b>2.3. APLICAÇÃO A DADOS EXPERIMENTAIS.....</b>	<b>94</b>
<b>2.3.1. MODELO PARA A ANÁLISE DO DIALELO PARCIAL EM MEGT.....</b>	<b>95</b>
<b>2.3.2. AVALIAÇÃO DOS ALGORITMOS PARA ANÁLISE DO DIALELO         PARCIAL EM MEGT .....</b>	<b>98</b>
<b>2.3.3. ESCOLHA DO MELHOR MODELO PARA A ANÁLISE DO DIALELO         PARCIAL EM MEGT .....</b>	<b>98</b>
<b>2.3.4. INFORMAÇÕES GENÉTICAS DA ANÁLISE DO DIALELO PARCIAL EM         MEGT.....</b>	<b>99</b>
<b>2.3.5. POTENCIAL DAS POPULAÇÕES SEGREGANTES VIA DIALELO EM         MEGT.....</b>	<b>100</b>
<b>3. RESULTADOS .....</b>	<b>101</b>
<b>3.1. EFICIÊNCIA COMPUTACIONAL PARA A ANÁLISE DO DIALELO PARCIAL     EM MEGT.....</b>	<b>101</b>

<b>3.2. ESCOLHA DO MODELO PARA A ANÁLISE DO DIALELO PARCIAL EM MEGT .....</b>	<b>103</b>
<b>3.3. ANÁLISE DO DIALELO PARCIAL EM MEGT .....</b>	<b>104</b>
<b>3.3.1. POTENCIAL DAS POPULAÇÕES SEGREGANTES PARA A EXTRAÇÃO DE LINHAGENS .....</b>	<b>107</b>
<b>4. DISCUSSÃO .....</b>	<b>110</b>
<b>4.1. EFICIÊNCIA COMPUTACIONAL PARA A ANÁLISE DO DIALELO PARCIAL EM MEGT .....</b>	<b>110</b>
<b>4.2. ESCOLHA DO MODELO PARA A ANÁLISE DO DIALELO PARCIAL EM MEGT .....</b>	<b>112</b>
<b>4.3. ANÁLISE DO DIALELO PARCIAL EM MEGT .....</b>	<b>113</b>
<b>4.3.1. POTENCIAL DAS POPULAÇÕES SEGREGANTES PARA EXTRAÇÃO DE LINHAGENS .....</b>	<b>115</b>
<b>5. CONCLUSÃO .....</b>	<b>117</b>
<b>6. REFERÊNCIAS .....</b>	<b>117</b>
<b>7. MATERIAL SUPLEMENTAR .....</b>	<b>125</b>
<b>CONCLUSÕES GERAIS .....</b>	<b>138</b>

## INTRODUÇÃO GERAL

Embora tenhamos escapado das previsões sombrias feitas por Thomas Malthus, que acreditava que o crescimento populacional suplantaria a capacidade produtiva do planeta, a humanidade poderá se deparar novamente com grandes desafios face às mudanças climáticas e ao crescimento populacional (LONG, ORT, 2010; FOLEY *et al.*, 2011; RAY *et al.*, 2013). De acordo com o relatório da FAO, na segunda metade deste século a população humana atingirá 9,1 bilhões de habitantes e a produção deve ser incrementada em 70% para suportar este aumento populacional (FAO, 2009). Além disso, o nosso planeta ainda pode estar sujeito a severas mudanças climáticas em escala global como: aumento da temperatura e da concentração atmosférica de CO<sub>2</sub> bem como secas mais frequentes (LONG, ORT, 2010).

Essas mudanças poderão alterar de maneira drástica o sistema produtivo atual no futuro, em um cenário onde é esperado aumento na produção das principais culturas responsáveis pela segurança alimentar (RAY *et al.*, 2013). Neste sentido, o melhoramento de plantas integrado a outras áreas da ciência (GODWIN *et al.*, 2019) pode oferecer valiosas contribuições para superar os desafios futuros. Neste contexto, destaca-se a genética quantitativa e a estatística, pois a maioria dos caracteres de interesse econômico são de natureza quantitativa, o que implica em uso de análises estatísticas para a tomada de decisão. Isto porque a união da estatística com genética quantitativa permite ao melhorista estimar os parâmetros genéticos e não genéticos úteis no momento da seleção/recombinação (CRUZ, 2012).

A estimação dos parâmetros genéticos requer um delineamento genético dentre os quais destaca-se o dialelo. Dialelo é um esquema ordenado de acasalamento para avaliar o desempenho dos pais em cruzamento, comumente usados no melhoramento de plantas e no estudo herança de características quantitativas (GRIFFING, 1956a; GRIFFING, 1956b; MÖHRING *et al.*, 2011). A partir do inter cruzamento de  $p$  linhagens parentais são obtidas  $p^2$  combinações, que são divididas em  $p(p - 1)/2$  combinações híbridas diretas,  $p(p - 1)/2$  combinações híbridas recíprocas e  $p$  linhagens parentais (GRIFFING, 1956a). A análise do dialelo completo com  $p^2$  combinações permite estimar os efeitos da capacidade geral de combinação (CGC), capacidade de combinação específica (CEC), CGC recíproca (RCGC) e CEC recíproca (RCEC) (GRIFFING, 1956b; COCKERHAM, WEIR, 1977). Esses efeitos informam sobre a frequência de alelos favoráveis, diversidade genética, contribuição de genes extranucleares e interação de genes nucleares e extranucleares bem como interação entre genes extranucleares, respectivamente (COCKERHAM,

WEIR, 1977; VIANA, MATTA, 2003; MOURA *et al.*, 2018). Destes, os dois últimos efeitos (RCGC e RCEC) são especialmente importantes para melhoramento de plantas com ênfase na obtenção de clones ou híbridos superiores (YAO *et al.*, 2013; FAN *et al.*, 2014; RUKUNDO *et al.*, 2017).

Embora a análise dialélica forneça valiosas informações aos melhoristas a respeito da capacidade combinatória dos parentais avaliados em cruzamentos, sabe-se que estas podem mudar em função do ambiente (OGUT *et al.*, 2014; PÁDUA *et al.*, 2016; MOURA *et al.*, 2018). Estas mudanças são atribuídas à interação genótipos por ambientes (G×A). Assim, fica evidente a importância da avaliação dos dialelos em vários locais, anos ou safras.

Os ensaios realizados em um conjunto de ambientes, denominados de *multi-environment trials* (MET) (SMITH *et al.*, 2001; SMITH *et al.*, 2005; SMITH *et al.*, 2015), são rotineiros no estudo da interação G×A (SMITH *et al.*, 2001; SMITH *et al.*, 2005; OGUT *et al.*, 2014; SMITH *et al.*, 2015). Entretanto, apesar da relevância da interação G×A nos programas de melhoramento, a avaliação de dialelos em MET tem sido menos frequente.

Nos programas melhoramento de plantas autógamas, devido à maior praticidade, os ensaios envolvem a avaliação de sucessivas gerações de endogamia e são realizados de forma sequencial (RESENDE *et al.*, 2015; BATISTA *et al.*, 2017). Neste contexto, diferentes gerações são avaliadas em diferentes ambientes (anos, safras ou locais), o que permite caracterizar tal conjunto de ensaios como *multi-environment and multi-generation trials* (MEGT).

Smith *et al.* (2001), trabalhando com dados MET, propuseram modelar a estrutura de (co)variância genotípica ao invés de particionar o efeito genotípico nas frações de comportamento geral dos genótipos e interação G×A. Além disso, Smith *et al.* (2001) e Smith *et al.* (2005) sugeriram a utilização da abordagem de modelos mistos pela maior flexibilidade para modelar as estruturas de (co)variância genotípica e residual, conectar experimentos que possuem desbalanceamento estatístico e/ou genético.

Embora existam vantagens significativas associadas ao procedimento REML, seu uso requer algoritmos numéricos para estimar os componentes de variância (THOMPSON *et al.*, 2005; DIFFEY *et al.*, 2017). Por meio dessa abordagem, estes parâmetros devem ser estimados usando algoritmos baseados nas derivadas de primeira e/ou segunda ordem da função de verossimilhança residual (THOMPSON *et al.*, 2005; THOMPSON, 2008).

O algoritmo *expectation maximization* (EM) pode ser obtido rearranjando-se as derivadas de primeira ordem da função de verossimilhança residual (SEARLE *et al.*, 1992; THOMPSON *et al.*, 2005). Esse algoritmo foi proposto por Dempster *et al.* (1977) e geralmente apresenta convergência monótona, isto é, garante aumento progressivo do  $\log(L)$  e mantém os parâmetros no espaço paramétrico automaticamente (THOMPSON *et al.*, 2005; MEYER, 2006; MEYER, 2008). Entretanto, esse algoritmo geralmente requer grande número de iterações para convergir (THOMPSON *et al.*, 2005; MEYER, 2006).

Os algoritmos baseados nas derivadas de segunda ordem da função de verossimilhança residual apresentam convergência quadrática, isto é, convergem em poucas iterações (GILMOUR *et al.*, 1995; JOHNSON, THOMPSON, 1995; MEYER, SMITH, 1996; ZHOU, STEPHENS, 2012; ZHOU, STEPHENS, 2014). O esquema iterativo associado a estes algoritmos requer uma matriz de informação (geralmente a matriz de informação hessiana ou esperada) (PATTERSON, THOMPSON, 1971; GILMOUR *et al.*, 1995; JOHNSON, THOMPSON, 1995). No entanto, as expressões para o cálculo dessas matrizes envolve traços matriciais complexos (GILMOUR *et al.*, 1995; JOHNSON, THOMPSON, 1995). Neste contexto, Johnson, Thompson (1995) e Gilmour *et al.* (1995) propuseram o algoritmo *average information* (AI), que é baseado na matriz de informação média.

Devido à sua maior eficiência computacional (GILMOUR *et al.*, 1995; JOHNSON, THOMPSON, 1995) este algoritmo se popularizou e vem sendo utilizado na implementação de vários *softwares* como ASReml, Wombat, AIREMLF90 e MTG2 (MEYER, 2006; MISZTAL, 2008; LEE, VAN DER WERF, 2016; GILMOUR, 2019). Além da maior eficiência computacional, o algoritmo AI ainda possibilita a estimação de grande número de parâmetros, permite trabalhar com grandes conjuntos de dados e é facilmente generalizável para estruturas de (co)variância complexas (GILMOUR *et al.*, 1995; THOMPSON, 2008; GILMOUR, 2019). Contudo, em algumas situações o algoritmo AI pode não convergir (MEYER, 2006; KNIGHT, 2008; DIFFEY *et al.*, 2017). Neste contexto, MEYER (2006) sugere a combinação entre os algoritmos *parameter expanded expectation maximization* (PXEM) e AI, visando explorar as vantagens complementares destes algoritmos (maior estabilidade e rápida convergência, respectivamente). Entretanto, apesar do algoritmo PXEM ser um pouco mais eficiente que o algoritmo *expectation maximization* (EM) (KNIGHT, 2008; MEYER, 2008), sua implementação requer o cálculo de um traço matricial adicional que pode ser caro computacionalmente

(KNIGHT, 2008; DIFFEY *et al.*, 2017). Neste sentido, Marçal *et al.* (2019) sugeriram a utilização da combinação entre os algoritmos EM e AI.

Assim, o objetivo deste trabalho foi descrever os detalhes para se realizar a análise dialéctica envolvendo as capacidades geral e específica recíprocas e ensaios multiambientes e multigerações via modelos mistos usando os algoritmos EM, AI e a combinação entre os algoritmos EM e AI (EM-AI).

## REFERÊNCIAS

BATISTA, L. G. *et al.* Multigeneration index in the selection of common bean inbred families. **Crop Science**, v. 57, n. 5, p. 2354-2360, June. 2017. DOI: 10.2135/cropsci2016.08.0665. Disponível em: <https://dl.sciencesocieties.org/publications/cs/abstracts/57/5/2354?access=0&view=pdf>. Acesso em: 25 agosto 2019.

COCKERHAM, C. C.; WEIR, B. S. Quadratic analyses of reciprocal crosses. **Biometrics**, v. 33, n. 1, p. 187-203, Mar. 1977. DOI: 10.2307/2529312. Disponível em: <https://www.jstor.org/stable/2529312>. Acesso em: 25 agosto 2019.

CRUZ, C. D. **Princípios de genética quantitativa**. 1. ed. Viçosa:Editora UFV, 2012. 394 p.

DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via the EM algorithm. **Journal of the Royal Statistical Society: Series B**, v. 39, n. 1, p. 1-38, 1977. Disponível em: <https://www.jstor.org/stable/2984875>. Acesso em: 25 agosto 2019.

DIFFEY, S. M. *et al.* A new REML (parameter expanded) EM algorithm for linear mixed models. **Australian & New Zealand Journal of Statistics**, v. 59, n. 4, p. 433-448, Nov. 2017. DOI: 10.1111/anzs.12208. Disponível em: <https://onlinelibrary.wiley.com/doi/full/10.1111/anzs.12208>. Acesso em: 25 agosto 2019.

FAN, X. M. *et al.* Reciprocal diallel crosses impact combining ability, variance estimation, and heterotic group classification. **Crop Science**, v. 54, n. 1, p. 89-97, Dec. 2014. DOI: 10.2135/cropsci2013.06.0393. Disponível em: <https://dl.sciencesocieties.org/publications/cs/abstracts/54/1/89>. Acesso em: 25 agosto 2019.

FAO. **Global agriculture towards 2050**. Roma, 2009. 4 p. Disponível em: [http://www.fao.org/fileadmin/templates/wsfs/docs/Issues\\_papers/HLEF2050\\_Global\\_Agriculture.pdf](http://www.fao.org/fileadmin/templates/wsfs/docs/Issues_papers/HLEF2050_Global_Agriculture.pdf). Acesso em: 25 agosto 2019.

FOLEY, J. A. *et al.* Solutions for a cultivated planet. **Nature**, v. 478, n. 7369, p. 337-342, Oct. 2011. DOI: 10.1038/nature10452. Disponível em: <https://www.nature.com/articles/nature10452>. Acesso em: 25 agosto 2019.

GILMOUR, A. R.; THOMPSON, R.; CULLIS, B. R. Average information REML: an efficient algorithm for variance parameter estimation in linear mixed models. **Biometrics**, v. 51, n. 4, p. 1440-1450, Dec. 1995. DOI: 10.2307/2533274. Disponível em: <https://www.jstor.org/stable/2533274>. Acesso em: 25 agosto 2019.

GILMOUR, A. R. Average information residual maximum likelihood in practice. **Journal of Animal Breeding and Genetics**, v. 136, n. 5, p. 262-272, Apr. 2019. DOI: 10.1111/jbg.12398. Disponível em: <https://onlinelibrary.wiley.com/doi/full/10.1111/jbg.12398>. Acesso em: 25 agosto 2019.

GODWIN, I. D. *et al.* Technological perspectives for plant breeding. **Theoretical and Applied Genetics**, v. 132, n. 3, p. 555-557, Mar. 2019. DOI: 10.1007/s00122-019-03321-4. Disponível em: <https://link.springer.com/article/10.1007/s00122-019-03321-4>. Acesso em: 25 agosto 2019.

GRIFFING, B. A generalised treatment of the use of diallel crosses in quantitative inheritance. **Heredity**, v. 10, n. 1, p. 31-50, Apr. 1956a. DOI: 10.1038/hdy.1956.2. Disponível em: <https://www.nature.com/articles/hdy19562>. Acesso em: 25 agosto 2019.

GRIFFING, B. Concept of general and specific combining ability in relation to diallel crossing systems. **Australian Journal of Biological Sciences**, v. 9, n. 4, p. 463-493, June. 1956b. DOI: 10.1071/BI9560463. Disponível em: <http://www.publish.csiro.au/bi/pdf/bi9560463>. Acesso em: 25 agosto 2019b.

JOHNSON, D. L.; THOMPSON, R. Restricted maximum likelihood estimation of variance components for univariate animal models using sparse matrix techniques and average information. **Journal of dairy science**, v. 78, n. 2, p. 449-456, Aug. 1995. DOI: 10.3168/jds.S0022-0302(95)76654-1. Disponível em: <https://www.journalofdairyscience.org/article/S0022-0302%2895%2976654-1/abstract>. Acesso em: 25 agosto 2019.

KNIGHT, E. **Improved iterative schemes for REML estimation of variance parameters in linear mixed models**. 2008. 290 f. Tese (Doutorado) – School of Agriculture, Food and Wine, University of Adelaide, Australia, 2008. Disponível em: <https://digital.library.adelaide.edu.au/dspace/bitstream/2440/49425/10/02whole.pdf>. Acesso em: 25 agosto 2019.

LEE, S. H.; VAN DER WERF, J. H. J. MTG2: An efficient algorithm for multivariate linear mixed model analysis based on genomic information. **Bioinformatics**, v. 32, n. 9, p. 1420-1422, May. 2016. DOI: 10.1093/bioinformatics/btw012. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/26755623>. Acesso em: 25 agosto 2019.

LONG, S. P.; ORT, D. R. More than taking the heat: crops and global change. **Current Opinion in Plant Biology**, v. 13, n. 3, p. 240-247, Jun. 2010. DOI: 10.1016/j.pbi.2010.04.008. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1369526610000518?via%3Dihub>. Acesso em: 25 agosto 2019.

MARÇAL, T. S. *et al.* Estimation of variance for reciprocal general and specific combining ability effects by EM-AI algorithm. **Crop Science**, v. 59, n. 4, p. 1494-1503, Apr. 2019. DOI: 10.2135/cropsci2018.09.0555. Disponível em: <https://dl.sciencesocieties.org/publications/cs/abstracts/59/4/1494>. Acesso em: 25 agosto 2019.

MEYER, K. Parameter expansion for estimation of reduced rank covariance matrices. **Genetics Selection Evolution**, v. 40, n. 1, p. 3-24, Jan. 2008. DOI: 10.1051/gse:2007032. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2674917/>. Acesso em: 25 agosto 2019.

MEYER, K.  $PX \times AI$ : Algorithmic for better convergence in restricted maximum likelihood estimation. *In*: WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION, 8., 2006, Belo Horizonte. **Resumos [...]**. Belo Horizonte, 2006. p. 4.

MEYER, K.; SMITH, S. P. Restricted maximum likelihood estimation for animal models using derivatives of the likelihood. **Genetics Selection Evolution**, v. 28, n. 1, p. 23-49, Mar. 1996. DOI: 10.1186/1297-9686-28-1-23. Disponível em: <https://gsejournal.biomedcentral.com/articles/10.1186/1297-9686-28-1-23>. Acesso em: 25 agosto 2019.

MISZTAL, I. Reliable computing in estimation of variance components. **Journal of animal breeding and genetics**, v. 125, n. 6, p. 363-370, Nov. 2008. DOI: 10.1111/j.1439-0388.2008.00774.x. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1439-0388.2008.00774.x>. Acesso em: 25 agosto 2019.

MÖHRING, J.; MELCHINGER, A. E.; PIEPHO, H. P. REML-based diallel analysis. **Crop science**, v. 51, n. 2, p. 470-478, Nov. 2011. DOI: 10.2135/cropsci2010.05.0272. Disponível em: <https://dl.sciencesocieties.org/publications/cs/abstracts/51/2/470?access=0&view=pdf>. Acesso em: 25 agosto 2019.

MOURA, L. M. *et al.* Combining ability of common bean parents in different seasons, locations and generations. **Euphytica**, v. 214, n. 10, p. 181, Oct. 2018. DOI: 10.1007/s10681-018-2259-3. Disponível em: <https://link.springer.com/article/10.1007/s10681-018-2259-3>. Acesso em: 25 agosto 2019.

OGUT, F. *et al.* Genetic analysis of diallel progeny test data using factor analytic linear mixed models. **Forest Science**, v. 60, n. 1, p. 119-127, Feb. 2014. DOI: 10.1007/s10681-016-1722-2v. Disponível em: <https://academic.oup.com/forestscience/article/60/1/119/4583734>. Acesso em: 25 agosto 2019.

PÁDUA, J. M. V. *et al.* A multi-environment trials diallel analysis provides insights on the inheritance of fumonisin contamination resistance in tropical maize. **Euphytica**, v. 211, n. 3, p. 277-285, Oct. 2016. DOI: 10.1007/s10681-016-1722-2. Disponível em: <https://link.springer.com/article/10.1007/s10681-016-1722-2>. Acesso em: 25 agosto 2019.

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: 10.1093/biomet/58.3.545. Disponível em: <https://www.jstor.org/stable/2334389>. Acesso em: 25 agosto 2019.

RAY, D. K. *et al.* Yield trends are insufficient to double global crop production by 2050. **PloS one**, v. 8, n. 6, p. e66428, June. 2013. DOI: 10.1371/journal.pone.0066428. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0066428>. Acesso em: 25 agosto 2019.

RESENDE, M. D. V. *et al.* Multigeneration index in the within-progenies bulk method for breeding of self-pollinated plants. **Crop Science**, v. 55, n. 3, p. 1202-1211, Apr. 2015. DOI: 10.2135/cropsci2014.08.0580. Disponível em: <https://dl.sciencesocieties.org/publications/cs/abstracts/55/3/1202>. Acesso em: 25 agosto 2019.

SEARLE, S. R.; CASELLA, G.; MCCULLOCH, C. E. **Variance Components**. John Wiley & Sons, 1992. 537 p.

SMITH, A. B.; CULLIS, B. R.; THOMPSON, R. Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. **Biometrics**, v. 57, n. 4, p. 1138-1147, Dec. 2001. DOI: 10.1111/j.0006-341X.2001.01138.x. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.0006-341X.2001.01138.x>. Acesso em: 25 agosto 2019.

SMITH, A. B.; CULLIS, B. R.; THOMPSON, R. The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. **The Journal of Agricultural Science**, v. 143, n. 6, p. 449-462, Dec. 2005. DOI: 10.1017/S0021859605005587. Disponível em: <https://www.cambridge.org/core/journals/journal-of-agricultural-science/article/analysis-of-crop-cultivar-breeding-and-evaluation-trials-an-overview-of-current-mixed-model-approaches/149BDADEE1D04E52616E74A1191A8F69>. Acesso em: 25 agosto 2019.

SMITH, A. B. *et al.* Factor analytic mixed models for the provision of grower information from national crop variety testing programs. **Theoretical and applied genetics**, v. 128, n. 1, p. 55-72, Jan. 2015. DOI: 10.1007/s00122-014-2412-x. Disponível em: <https://link.springer.com/article/10.1007/s00122-014-2412-x>. Acesso em: 25 agosto 2019.

THOMPSON, R.; BROTHERSTONE, S.; WHITE, I. M. S. Estimation of quantitative genetic parameters. **Philosophical Transactions of the Royal Society B: Biological Sciences**, v. 360, n. 1459, p. 1469-1477, July. 2005. DOI: 10.1098/rstb.2005.1676. Disponível em: <https://royalsocietypublishing.org/doi/10.1098/rstb.2005.1676>. Acesso em: 25 agosto 2019.

THOMPSON, R. Estimation of quantitative genetic parameters. **Proceedings of the Royal Society B: Biological Sciences**, v. 275, n. 1635, p. 679-686, Jan. 2008. DOI: 10.1098/rspb.2007.1417. Disponível em: <https://royalsocietypublishing.org/doi/10.1098/rspb.2007.1417>. Acesso em: 25 agosto 2019.

VIANA, J. M. S.; MATTA, Frederico de Pina. Analysis of general and specific combining abilities of popcorn populations, including selfed parents. **Genetics and Molecular Biology**, v. 26, n. 4, p. 465-471, Dec. 2003. DOI: 10.1590/S1415-47572003000400010. Disponível em: [http://www.scielo.br/scielo.php?pid=S1415-47572003000400010&script=sci\\_arttext&tlng=es](http://www.scielo.br/scielo.php?pid=S1415-47572003000400010&script=sci_arttext&tlng=es). Acesso em: 25 agosto 2019.

YAO, W. H. *et al.* Diallel analysis models: A comparison of certain genetic statistics. **Crop science**, v. 53, n. 4, p. 1481-1490, June. 2013. DOI: 10.2135/cropsci2013.01.0027. Disponível em: <https://dl.sciencesocieties.org/publications/cs/abstracts/53/4/1481?access=0&view=pdf>. Acesso em: 25 agosto 2019.

ZHOU, X.; STEPHENS, M. Efficient algorithms for multivariate linear mixed models in genome-wide association studies. **Nature methods**, v. 11, n. 4, p. 407-409, Apr. 2014. DOI: 10.1038/nmeth.2848. Disponível em: <https://www.nature.com/articles/nmeth.2848>. Acesso em: 25 agosto 2019.

ZHOU, X.; STEPHENS, M. Genome-wide efficient mixed-model analysis for association studies. **Nature genetics**, v. 44, n. 7, p. 821-824, June. 2012. DOI: 10.1038/ng.2310. Disponível em: <https://www.nature.com/articles/ng.2310>. Acesso em: 25 agosto 2019.

**CAPÍTULO 1**  
**ESTIMATION OF VARIANCE FOR RECIPROCAL GENERAL AND SPECIFIC**  
**COMBINING ABILITY EFFECTS BY EM-AI ALGORITHM**

**VIÇOSA - MINAS GERAIS**  
**2019**

## ABSTRACT

In diallel analysis, using mixed models, it is possible to estimate the variance components of general combining ability (GCA), specific combining ability (SCA), reciprocal GCA (RGCA), reciprocal SCA (RSCA), and parents, even in unbalanced datasets. These variances can be estimated by likelihood maximization via numerical algorithms (e.g., expectation maximization [EM] and average information [AI]), which have different advantages. Thus, the objective of this study was to describe and implement the EM-AI algorithm (combination of EM and AI) in R software to estimate the variance components of RGCA, RSCA, and parents effects in diallel analysis. Two real datasets and three diallel models (Griffing's Model 1, Griffing's Model 1 + RGCA, and Model 3, a general diallel model with the effects of GCA, SCA, RGCA, RSCA, and parents) were used to evaluate the efficiency of the algorithms AI, EM, and EM-AI. The AI algorithm failed to converge for the three diallel models in both datasets. The other algorithms (EM and EM-AI) converged normally, and the estimated variance components with these algorithms were similar for the three diallel models in both datasets. However, the EM-AI algorithm was more efficient than the EM algorithm, and the general diallel model (Model 3) provided more accurate estimates of variance parameters. Thus, the EM-AI algorithm with routine implemented in R has potential for use in diallel analyses in plant breeding.

**Key words:** Diallel analysis, linear mixed models, extranuclear genes.

## 1. INTRODUCTION

Diallel is an orderly mating scheme for evaluating the performance of parents in crosses, commonly used in plant breeding and inheritance studies of quantitative traits (Griffing, 1956a, 1956b; Möhring et al., 2011). From the crosses among  $p$  parents are obtained  $p^2$  combinations, which are divided into  $p(p - 1)/2$  crosses,  $p(p - 1)/2$  reciprocal crosses, and  $p$  parents (Griffing, 1956a; Möhring et al., 2011). Griffing (1956b) proposed models for analyzing four different diallel designs, known as Methods 1 (parents +  $F_1$  crosses +  $F_1$  reciprocal crosses), 2 (parents +  $F_1$  crosses), 3 ( $F_1$  crosses +  $F_1$  reciprocal crosses), and 4 ( $F_1$  crosses). These models enable estimation of the effects of the general combining ability (GCA) and specific combining ability (SCA). Beyond these effects, the models described in Methods 1 and 3 include the reciprocal effect (Griffing, 1956b).

Cockerham and Weir (1977) defined reciprocal effect as the reciprocal SCA (RSCA) and extended the Griffing model to include the effect of reciprocal GCA (RGCA). According to the same authors, the RGCA effect is caused by the contribution of extranuclear genes, whereas the RSCA effect is caused by the interaction among nuclear and extranuclear genes, as well as between extranuclear genes. These effects are especially important for plant breeding with an emphasis on obtaining superior clones or hybrids (Yao et al., 2013; Fan et al., 2014; Rukundo et al., 2017). Rukundo et al. (2017) emphasized the importance of extranuclear genes for breeding of vegetatively propagated crops, since a superior cross can be propagated identically (clonal propagation). In addition, Yao et al. (2013) reported that when large RGCA (named maternal effect [MAT] by the authors) and/or RSCA (named reciprocal effect [REC] by the authors) effects exist, there is a high risk of missing superior maize hybrids when diallel designs are used without reciprocal crosses.

The genetic interpretation of the parameters in the model provided by Cockerham and Weir (1977) is valid only for diallel analysis of Griffing's Method 3 (i.e., without the inclusion of parents in the analysis). Yao et al. (2013) affirmed that including parents in diallel analysis (Griffing's Method 1) biased the estimates of GCA and SCA effects for grain yield in maize (*Zea mays* L.). Additionally, Möhring et al. (2011) reported overestimated variance components of GCA and SCA using the model described in Griffing's Method 1. In this context, Möhring et al. (2011) adapted the model proposed by Cockerham and Weir (1977) to Griffing's Method 1 with a model that separates the effect of parents from the effect of crosses.

The genetic effects of the diallel model can be taken as fixed or random (Griffing, 1956a, 1956b; Möhring et al., 2011). When these effects are considered as random, the main interest resides in estimates of the variance components. Originally, Griffing (1956b) and Cockerham and Weir (1977) suggested estimating the variance components via a method of moments (ANOVA method). However, this approach is limited when you want to fit complex block structures, spatial error structures, and/or heterogeneous variances (Patterson and Thompson, 1971; Smith et al., 2005; Möhring et al., 2011). In this context, the residual maximum likelihood (REML) is an alternative to ANOVA for estimating variance components linked to complex models.

The REML method was developed by Patterson and Thompson (1971) for estimating variance components in incomplete block designs. Posteriorly, the authors found relations between REML and mixed model equations, which allowed the REML method to be generalized to other models (Thompson, 2008), such as diallel. Furthermore, the REML method keeps estimates of the variance components within a parametric space and facilitates analysis of correlated and/or unbalanced data (Xiang and Li, 2001; Knight, 2008; Thompson, 2008; Tenesa and Haley, 2013). Although there are significant advantages associated with the REML

procedure, its use requires numerical algorithms for estimating the variance components (Thompson et al., 2005; Diffey et al., 2017). By means of this approach, the variance components can be estimated using algorithms based on first or second differentials of the residual log-likelihood (Thompson et al., 2005; Thompson, 2008).

The expectation maximization (EM) algorithm was developed by manipulating the first differentials of the residual log-likelihood (Searle et al., 1992; Mrode and Thompson, 2005; Thompson et al., 2005). This algorithm was developed by Dempster et al. (1977) and generally provides monotone convergence (i.e., the log-likelihood is guaranteed to increase at each iterative step and automatically keeps the variance components within the parametric space; Thompson et al., 2005; Meyer, 2006, 2008). However, this method may provide slow convergence and generally requires a large number of iterations to converge (Thompson et al., 2005; Meyer, 2006). Foulley and Van Dyk (2000) improved the EM algorithm by making it faster, which was named parameter expanded EM (PXEM).

The algorithms based on second differentials of the residual log-likelihood provide quadratic convergence (i.e., within a few iterations; Gilmour et al., 1995; Johnson and Thompson, 1995; Meyer and Smith, 1996; Zhou and Stephens, 2012, 2014). However, the iterative scheme in these algorithms requires an information matrix (generally Hessian or Fisher information; Patterson and Thompson, 1971; Gilmour et al., 1995; Johnson and Thompson, 1995). Nevertheless, the equations for producing these matrices involve complex traces (Gilmour et al., 1995; Johnson and Thompson, 1995). In this context, Johnson and Thompson (1995) and Gilmour et al. (1995) proposed the AI algorithm, which is based on the average information matrix. However, Meyer (2006) reported situations where AI fails to converge, especially in

analysis involving several random effects, numerous traits, or bad starting points for variance parameters of the residual log-likelihood.

Meyer (2006) proposed the PXEM-AI algorithm (combination of PXEM and AI) with the goal of circumventing limitations of the algorithms mentioned above. This technique brings together the complementary advantages of the PXEM and AI algorithms (i.e., high stability and fast convergence, respectively). However, the PXEM algorithm requires additional trace computation of the off-diagonal terms of the inverse coefficient matrix of the mixed model equations ( $C^{XZ}$ ) that can be computationally more expensive than the EM algorithm (Knight, 2008; Diffey et al., 2017).

Thus, the objective of this study was to describe and implement an algorithm in R software that involves a combination of the EM and AI algorithms to estimate the variance components of the RGCA, RSCA, and parental effects based on the diallel analysis model described by Möhring et al. (2011) and adapted from Cockerham and Weir (1977).

## **2. MATERIALS AND METHODS**

In the sections below, we will present the diallel model described by Möhring et al. (2011) and adapted from Cockerham and Weir (1977), without covariance between the GCA and parental effects. Furthermore, we also implemented the EM, AI, and EM-AI algorithms in a routine available in R to fit this model using REML (see Routine S1 in the supplemental material <<https://dl.sciencesocieties.org/publications/cs/abstracts/59/4/1494>>).

## 2.1. MODEL

The diallel model described by Cockerham and Weir (1977), considering the randomized block design, is shown below:

$$y_{ijk} = \mu + g_i + g_j + s_{ij} + d_i - d_j + r_{ij} + \delta_k + e_{ijk} \quad [1]$$

where  $y_{ijk}$  is the phenotypic value of the  $ij$ th cross in the  $k$ th block,  $\mu$  is the overall mean,  $g_i$  and  $g_j$  are random effects of GCA from the  $i$ th and  $j$ th parents, respectively,  $s_{ij}$  is the random effect of SCA between the  $i$ th and the  $j$ th parents with  $s_{ij} = s_{ji}$ ,  $d_i$  and  $d_j$  are the random effects of RGCA from the  $i$ th and  $j$ th parents, respectively,  $r_{ij}$  is the random effect of RSCA of the  $ij$ th cross with  $r_{ij} = -r_{ji}$ ,  $\delta_k$  is the fixed effect of block, and  $e_{ijk}$  corresponds to random error.

A genetic interpretation of variance components of the model shown above (Eq. [1]) given by Cockerham and Weir (1977) is valid only for diallel analysis of Griffing's Method 3 (i.e., without parents). However, if parents were included in the trial, one may use the model proposed by Möhring et al. (2011), called Model 3 (general diallel model) by the authors. This model, excluding covariance between GCA and the parents, can be written in matrix form (Eq. [2]):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{Z}_g\mathbf{g} + \mathbf{Z}_s\mathbf{s} + \mathbf{Z}_d\mathbf{d} + \mathbf{Z}_r\mathbf{r} + \mathbf{Z}_p\mathbf{p} + \mathbf{e} \quad [2]$$

where  $\mathbf{y}$  is the vector of phenotypic values;  $\boldsymbol{\tau}$  is the vector of fixed effects divided into blocks, mean of crosses, and mean of parents;  $\mathbf{g}$  is the vector of random GCA effects;  $\mathbf{s}$  is the vector of random SCA effects;  $\mathbf{d}$  is the vector of random RGCA effects;  $\mathbf{r}$  is the vector of random RSCA effects;  $\mathbf{p}$  is the vector of random parents effects;  $\mathbf{e}$  is a random error vector; and  $\mathbf{X}$ ,  $\mathbf{Z}_g$ ,  $\mathbf{Z}_s$ ,  $\mathbf{Z}_d$ ,  $\mathbf{Z}_r$ , and  $\mathbf{Z}_p$  are the design matrices of the effects  $\boldsymbol{\tau}$ ,  $\mathbf{g}$ ,  $\mathbf{s}$ ,  $\mathbf{d}$ ,  $\mathbf{r}$ , and  $\mathbf{p}$ , respectively.

The vector of solutions  $\beta = (\hat{\tau}, \tilde{\mathbf{g}}, \tilde{\mathbf{s}}, \tilde{\mathbf{d}}, \tilde{\mathbf{r}}, \tilde{\mathbf{p}})$  for fixed and random effects of Model 3 (Eq. [2]) can be obtained from Eq. [3]. Equation [3] (shown below) is derived by maximizing the joint density function (Henderson et al., 1959):

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}_g & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}_s & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}_d & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}_r & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}_p \\ \mathbf{Z}'_g\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'_g\mathbf{R}^{-1}\mathbf{Z}_g + \mathbf{G}_g^{-1} & \mathbf{Z}'_g\mathbf{R}^{-1}\mathbf{Z}_s & \mathbf{Z}'_g\mathbf{R}^{-1}\mathbf{Z}_d & \mathbf{Z}'_g\mathbf{R}^{-1}\mathbf{Z}_r & \mathbf{Z}'_g\mathbf{R}^{-1}\mathbf{Z}_p \\ \mathbf{Z}'_s\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'_s\mathbf{R}^{-1}\mathbf{Z}_g & \mathbf{Z}'_s\mathbf{R}^{-1}\mathbf{Z}_s + \mathbf{G}_s^{-1} & \mathbf{Z}'_s\mathbf{R}^{-1}\mathbf{Z}_d & \mathbf{Z}'_s\mathbf{R}^{-1}\mathbf{Z}_r & \mathbf{Z}'_s\mathbf{R}^{-1}\mathbf{Z}_p \\ \mathbf{Z}'_d\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'_d\mathbf{R}^{-1}\mathbf{Z}_g & \mathbf{Z}'_d\mathbf{R}^{-1}\mathbf{Z}_s & \mathbf{Z}'_d\mathbf{R}^{-1}\mathbf{Z}_d + \mathbf{G}_d^{-1} & \mathbf{Z}'_d\mathbf{R}^{-1}\mathbf{Z}_r & \mathbf{Z}'_d\mathbf{R}^{-1}\mathbf{Z}_p \\ \mathbf{Z}'_r\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'_r\mathbf{R}^{-1}\mathbf{Z}_g & \mathbf{Z}'_r\mathbf{R}^{-1}\mathbf{Z}_s & \mathbf{Z}'_r\mathbf{R}^{-1}\mathbf{Z}_d & \mathbf{Z}'_r\mathbf{R}^{-1}\mathbf{Z}_r + \mathbf{G}_r^{-1} & \mathbf{Z}'_r\mathbf{R}^{-1}\mathbf{Z}_p \\ \mathbf{Z}'_p\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'_p\mathbf{R}^{-1}\mathbf{Z}_g & \mathbf{Z}'_p\mathbf{R}^{-1}\mathbf{Z}_s & \mathbf{Z}'_p\mathbf{R}^{-1}\mathbf{Z}_d & \mathbf{Z}'_p\mathbf{R}^{-1}\mathbf{Z}_r & \mathbf{Z}'_p\mathbf{R}^{-1}\mathbf{Z}_p + \mathbf{G}_p^{-1} \end{bmatrix} \begin{bmatrix} \hat{\tau} \\ \tilde{\mathbf{g}} \\ \tilde{\mathbf{s}} \\ \tilde{\mathbf{d}} \\ \tilde{\mathbf{r}} \\ \tilde{\mathbf{p}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'_g\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'_s\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'_d\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'_r\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'_p\mathbf{R}^{-1}\mathbf{y} \end{bmatrix} \quad [3]$$

We assume the following structures of (co)variances for the model defined in Eq. [2]:

$$\begin{bmatrix} \mathbf{g} \\ \mathbf{s} \\ \mathbf{d} \\ \mathbf{r} \\ \mathbf{p} \\ \mathbf{e} \end{bmatrix} \sim N \left( \mathbf{0}, \begin{bmatrix} \mathbf{G}_g & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_s & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{G}_d & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{G}_r & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{G}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{R} \end{bmatrix} \right) \quad [4]$$

$$\text{Var}(\mathbf{y}) = \mathbf{V} = \mathbf{Z}_g\mathbf{G}_g\mathbf{Z}'_g + \mathbf{Z}_s\mathbf{G}_s\mathbf{Z}'_s + \mathbf{Z}_d\mathbf{G}_d\mathbf{Z}'_d + \mathbf{Z}_r\mathbf{G}_r\mathbf{Z}'_r + \mathbf{Z}_p\mathbf{G}_p\mathbf{Z}'_p + \mathbf{R} \quad [5]$$

where  $\mathbf{G}_g$  is a (co)variance matrix of the random effects of GCA with  $\mathbf{G}_g = \mathbf{I}_g\sigma_g^2$ ,  $\mathbf{G}_s$  is a (co)variance matrix of the random effects of SCA with  $\mathbf{G}_s = \mathbf{I}_s\sigma_s^2$ ,  $\mathbf{G}_d$  is a (co)variance matrix of the random RGCA effects with  $\mathbf{G}_d = \mathbf{I}_d\sigma_d^2$ ,  $\mathbf{G}_r$  is a (co)variance matrix of the random RSCA effects with  $\mathbf{G}_r = \mathbf{I}_r\sigma_r^2$ ,  $\mathbf{G}_p$  is a (co)variance matrix of the random effects of parents with  $\mathbf{G}_p = \mathbf{I}_p\sigma_p^2$ ,  $\mathbf{R}$  is a (co)variance matrix of the residuals with  $\mathbf{R} = \mathbf{I}_n\sigma^2$ ,  $\mathbf{V}$  is a (co)variance matrix

of  $\mathbf{y}$ ,  $\mathbf{I}_g$ ,  $\mathbf{I}_s$ ,  $\mathbf{I}_d$ ,  $\mathbf{I}_r$ ,  $\mathbf{I}_p$ , and  $\mathbf{I}_n$  are identity matrices associated with the (co)variance matrices of the effects  $\mathbf{g}$ ,  $\mathbf{s}$ ,  $\mathbf{d}$ ,  $\mathbf{r}$ ,  $\mathbf{p}$  and  $\mathbf{e}$ , respectively.

## 2.2. RESIDUAL LOG-LIKELIHOOD

The residual log-likelihood assumes that  $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\tau}, \mathbf{V})$  and can be written according to the equation given by Johnson and Thompson (1995):

$$-2\log(L) = \text{const.} + \log|\mathbf{V}| + \log|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| + \mathbf{y}'\mathbf{P}\mathbf{y} \quad [6]$$

where  $\log(L)$  is the logarithm of the residual likelihood (Patterson and Thompson, 1971),  $\text{const.}$  is a constant of residual likelihood, and  $\mathbf{P}$  is an orthogonal projector of  $\mathbf{y}$  in  $\mathbf{X}$  (Gilmour et al., 1995).

The matrix  $\mathbf{P}$  can be defined as

$$\begin{aligned} \mathbf{P} &= \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1} \\ &= \mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{W}\mathbf{C}^{-1}\mathbf{W}'\mathbf{R}^{-1} \end{aligned} \quad [7]$$

$$\mathbf{W} = [\mathbf{X} : \mathbf{Z}_g : \mathbf{Z}_s : \mathbf{Z}_d : \mathbf{Z}_r : \mathbf{Z}_p] \quad [8]$$

The following likelihood equation can be used without requiring inversion of  $\mathbf{V}$  in Eq. [6]:

$$\begin{aligned} -2\log(L) &= \text{const.} + N\log(\sigma^2) + N_g\log(\sigma_g^2) \\ &\quad + N_s\log(\sigma_s^2) + N_d\log(\sigma_d^2) + N_r\log(\sigma_r^2) \\ &\quad + N_p\log(\sigma_p^2) + \log|\mathbf{C}| + \mathbf{y}'\mathbf{P}\mathbf{y} \end{aligned} \quad [9]$$

where  $\sigma^2$ ,  $\sigma_g^2$ ,  $\sigma_s^2$ ,  $\sigma_d^2$ ,  $\sigma_r^2$ , and  $\sigma_p^2$  correspond to the variance components of the residue, GCA, SCA, RGCA, RSCA, and the parents, respectively;  $N$ ,  $N_g$ ,  $N_s$ ,  $N_d$ ,  $N_r$ , and  $N_p$  correspond to the

total number of observations and number of GCA, SCA, RGCA, RSCA, and parental effects, respectively;  $\log|\mathbf{C}|$  is the logarithm of the determinant of the left-hand side ( $\mathbf{C}$ ) of the mixed model equations given in Eq. [3].

The generalized first- and second-order differentials of the residual log-likelihood can be found in Gilmour et al. (1995) and Johnson and Thompson (1995). Furthermore, the same authors also provide expressions for obtaining the observed (Hessian), expected (Fisher information), and average information matrices.

The first differentials of Eq. [6] for Model 3 (Eq. [2]) shown below were developed from the differentials proposed by Johnson and Thompson (1995):

$$\frac{\partial \log(L)}{\partial \sigma_q^2} = -\frac{1}{2} \begin{bmatrix} tr(\mathbf{PZ}_g \mathbf{Z}'_g) - \mathbf{y}' \mathbf{PZ}_g \mathbf{Z}'_g \mathbf{Py} \\ tr(\mathbf{PZ}_s \mathbf{Z}'_s) - \mathbf{y}' \mathbf{PZ}_s \mathbf{Z}'_s \mathbf{Py} \\ tr(\mathbf{PZ}_d \mathbf{Z}'_d) - \mathbf{y}' \mathbf{PZ}_d \mathbf{Z}'_d \mathbf{Py} \\ tr(\mathbf{PZ}_r \mathbf{Z}'_r) - \mathbf{y}' \mathbf{PZ}_r \mathbf{Z}'_r \mathbf{Py} \\ tr(\mathbf{PZ}_p \mathbf{Z}'_p) - \mathbf{y}' \mathbf{PZ}_p \mathbf{Z}'_p \mathbf{Py} \\ tr(\mathbf{P}) - \mathbf{y}' \mathbf{P} \mathbf{Py} \end{bmatrix} = -\frac{1}{2} \begin{bmatrix} \frac{N_g}{\sigma_g^2} - \frac{tr(\mathbf{C}^{gg})}{\sigma_g^4} - \frac{\mathbf{g}'\mathbf{g}}{\sigma_g^4} \\ \frac{N_s}{\sigma_s^2} - \frac{tr(\mathbf{C}^{ss})}{\sigma_s^4} - \frac{\mathbf{s}'\mathbf{s}}{\sigma_s^4} \\ \frac{N_d}{\sigma_d^2} - \frac{tr(\mathbf{C}^{dd})}{\sigma_d^4} - \frac{\mathbf{d}'\mathbf{d}}{\sigma_d^4} \\ \frac{N_r}{\sigma_r^2} - \frac{tr(\mathbf{C}^{rr})}{\sigma_r^4} - \frac{\mathbf{r}'\mathbf{r}}{\sigma_r^4} \\ \frac{N_p}{\sigma_p^2} - \frac{tr(\mathbf{C}^{pp})}{\sigma_p^4} - \frac{\mathbf{p}'\mathbf{p}}{\sigma_p^4} \\ \frac{N - r(\mathbf{X})}{\sigma^2} - \frac{1}{\sigma^2} \left[ \sum_q N_q - \sum_q \frac{\mathbf{C}^{qq}}{\sigma_q^2} \right] - \frac{\mathbf{e}'\mathbf{e}}{\sigma^4} \end{bmatrix} \quad [10]$$

where the index  $q$  corresponds to a generalization of the random effects;  $tr(.)$  is an operator of the matrix trace; and  $\mathbf{C}^{gg}$ ,  $\mathbf{C}^{ss}$ ,  $\mathbf{C}^{dd}$ ,  $\mathbf{C}^{rr}$ , and  $\mathbf{C}^{pp}$  correspond to the partition of  $\mathbf{C}^{-1}$  (inverse of left-hand side of the mixed model equations given in Eq. [3]), corresponding to  $\mathbf{g}$ ,  $\mathbf{s}$ ,  $\mathbf{d}$ ,  $\mathbf{r}$ , and  $\mathbf{p}$ , respectively.

### 2.3. ALGORITHMS AND COMPUTATIONAL STRATEGIES

In this study, we used the EM, AI, and EM-AI algorithms, the sparse matrices (Bates and Maechler, 2017), and the “working” variables (Gilmour et al., 1995; Johnson and Thompson, 1995).

The convergence criterion was defined according to Mrode and Thompson (2005). These authors monitored the stabilization of all variance components and  $\log(L)$  of the model to four decimal places. The initial values of the variance components ( $\sigma_g^2$ ,  $\sigma_s^2$ ,  $\sigma_d^2$ ,  $\sigma_r^2$ , and  $\sigma_p^2$ ) were  $1.0\sigma^2$ , with  $\sigma^2 = 0.5\sigma_y^2$  and  $\sigma_y^2$  is the simple variance of the response (y) (Butler et al., 2009).

### 2.4. EXPECTATION MAXIMIZATION

The EM algorithm is defined in terms of the first differentials of residual log-likelihood (Eq. [10]), as suggested by Mrode and Thompson (2005) and by Searle et al. (1992).

The variance components can be updated using the following iterative equations:

$$\text{EM} \left\{ \begin{array}{l} \hat{\sigma}_g^{2(k+1)} = \frac{\tilde{\mathbf{g}}^{(k)'} \tilde{\mathbf{g}}^{(k)} + \text{tr}[\mathbf{C}^{gg(k)}]}{N_g} \\ \hat{\sigma}_s^{2(k+1)} = \frac{\tilde{\mathbf{s}}^{(k)'} \tilde{\mathbf{s}}^{(k)} + \text{tr}[\mathbf{C}^{ss(k)}]}{N_s} \\ \hat{\sigma}_d^{2(k+1)} = \frac{\tilde{\mathbf{d}}^{(k)'} \tilde{\mathbf{d}}^{(k)} + \text{tr}[\mathbf{C}^{dd(k)}]}{N_d} \\ \hat{\sigma}_r^{2(k+1)} = \frac{\tilde{\mathbf{r}}^{(k)'} \tilde{\mathbf{r}}^{(k)} + \text{tr}[\mathbf{C}^{rr(k)}]}{N_r} \\ \hat{\sigma}_p^{2(k+1)} = \frac{\tilde{\mathbf{p}}^{(k)'} \tilde{\mathbf{p}}^{(k)} + \text{tr}[\mathbf{C}^{pp(k)}]}{N_p} \\ \hat{\sigma}^{2(k+1)} = \frac{\tilde{\mathbf{e}}^{(k)'} \mathbf{y}}{N - r(\mathbf{X})} \end{array} \right. \quad [11]$$

where the index  $k$  corresponds to the iterations counter.

In this iterative scheme, the estimator of the residual variance was obtained from Searle et al. (1992, see p. 282).

## 2.5. AVERAGE INFORMATION

The AI algorithm is obtained from the second differentials of residual log-likelihood (Eq. [6]) according to Gilmour et al. (1995) and Johnson and Thompson (1995). The AI matrix can be defined using Eq. [12], shown below (Smith et al., 2005; Knight, 2008):

$$\text{AI} = \frac{1}{2} \mathbf{Q}' \mathbf{P} \mathbf{Q} \quad [12]$$

where  $\mathbf{Q}$  is the matrix of working variables (Gilmour et al., 1995; Johnson and Thompson, 1995).

The matrix  $\mathbf{Q}$  for the diallel model in Eq. [2] is defined in Eq. [13]:

$$\mathbf{Q} \left\{ \begin{array}{l} \mathbf{Z}_g \mathbf{Z}'_g \mathbf{P} \mathbf{y} = \frac{1}{\hat{\sigma}_g^2} \mathbf{Z}_g \tilde{\mathbf{g}} \\ \mathbf{Z}_s \mathbf{Z}'_s \mathbf{P} \mathbf{y} = \frac{1}{\hat{\sigma}_s^2} \mathbf{Z}_s \tilde{\mathbf{s}} \\ \mathbf{Z}_d \mathbf{Z}'_d \mathbf{P} \mathbf{y} = \frac{1}{\hat{\sigma}_d^2} \mathbf{Z}_d \tilde{\mathbf{d}} \\ \mathbf{Z}_r \mathbf{Z}'_r \mathbf{P} \mathbf{y} = \frac{1}{\hat{\sigma}_r^2} \mathbf{Z}_r \tilde{\mathbf{r}} \\ \mathbf{Z}_p \mathbf{Z}'_p \mathbf{P} \mathbf{y} = \frac{1}{\hat{\sigma}_p^2} \mathbf{Z}_p \tilde{\mathbf{p}} \\ \mathbf{P} \mathbf{y} = \frac{1}{\hat{\sigma}^2} \tilde{\mathbf{e}} \end{array} \right. \quad [13]$$

The information matrix in the AI algorithm cited in Eq. [12] can also be obtained using sparse Cholesky factorization of the extended mixed model equations, as suggested by Meyer (1997).

The variance component updates can be obtained with the following iterative equation according to Lee and van der Werf (2016):

$$\Theta^{(k+1)} = \Theta^{(k)} + [\text{AI}^{(k)}]^{-1} \frac{\partial \log(L)}{\partial \Theta_i} |_{\Theta^{(k)}} \quad [14]$$

where  $\Theta$  is a column vector of estimated variance components,  $k$  is the iterations counter,  $\partial \log(L) / \partial \Theta_i$  is a column vector of the first derivatives of the log-likelihood with respect to each variance component, and AI is the average information matrix.

## 2.6. EM-AI ALGORITHM

The EM-AI algorithm was obtained by combination of EM and AI algorithms. The stages of EM-AI algorithm are shown on Fig. 1. Initially, a previously defined iterations number is performed using the EM steps to refine the initial estimates of variance components. In this study, 25 EM steps were used, as suggested by Bates and Sarkar (2005). Subsequently, these estimates are provided to AI algorithm, in which the estimation process is concluded. To evaluate the computational efficiency, the convergence of EM-AI algorithm was verified in each EM step (1–25).

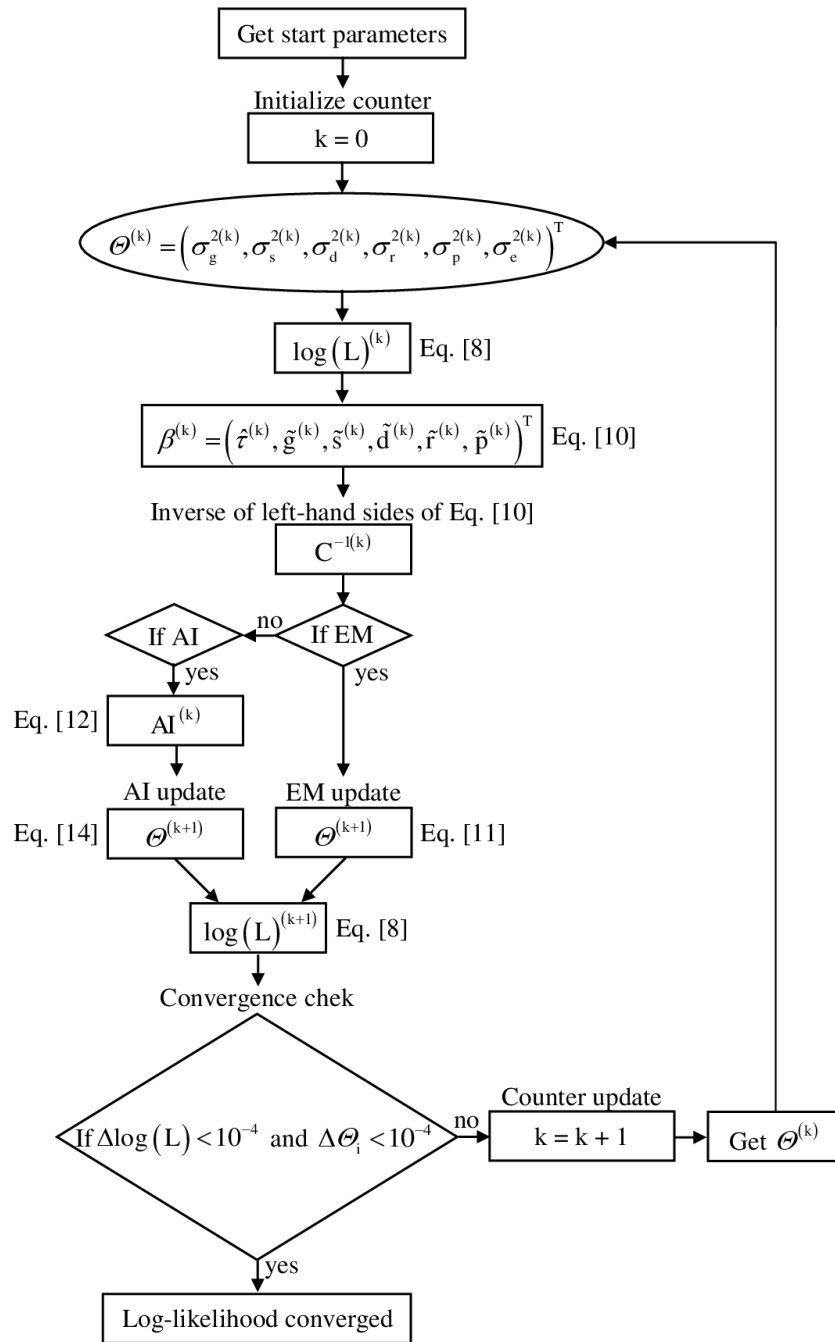


Fig. 1. Flowchart for each algorithm (expectation maximization [EM], average information [AI], and EM-AI) implemented in R, exemplifying a fit of Model 3.

## 2.7. DATASETS AND ANALYSES OF EXAMPLES

The variance components for two datasets were estimated using the REML algorithms (EM, AI, and EM-AI). The two experiments measured the flowering time of five plants per plot from a diallel design of Method 1 of eight inbred varieties (*Nicotiana rustica* L.). These field trials were conducted in randomized blocks design with two replicates in 1951 and 1952, respectively (Hayman, 1954a, 1954b). The first trial was conducted by Möhring et al. (2011) as a demonstration of the model in Eq. [2]. In this case, Hayman (1954b) multiplied the trait 10 times. The second trial was used by Cockerham and Weir (1977) as a demonstration of the model in Eq. [1], whereby the authors removed data from inbred varieties.

For comparison purposes, beyond Model 3 (Eq. [2]) given by Möhring et al. (2011), we further fit the diallel model described in Griffing's Method 1 (Griffing, 1956b; Möhring et al., 2011) and the diallel model described by Montesinos-López et al. (2009), which we call Model 1 and Model 1 + RGCA, respectively. We also implemented the last two models (Model 1 and Model 1 + RGCA) in the available routine. The three models (Model 1, Model 1 + RGCA, and Model 3) were fitted to the datasets using the proposed Routine S1 available in R (see Routine S1 in the supplemental material).

To evaluate the effectiveness of the EM algorithm for refining the initial values of the variance component vector ( $\Theta$ ) provided to the AI algorithm in the EM-AI algorithm, we estimated Pearson's correlation coefficients between  $\Theta^{(k)}$  ( $k$  ranging from 0 to 25) and converged  $\Theta$  for three diallel models (Model 1, Model 1 + RGCA, and Model 3) in both datasets (Dataset 1 and Dataset 2).

The analyses were performed using a laptop with a 64-bit Windows operating system, 2.5 GHz Intel Core i5 CPU, and 4-GB RAM. The average computation time (based on five replications) required to fit the diallel models to the data was recorded with the rbenchmark package (Kusnierczyk, 2015). Beyond that, we further computed the empirical rate of convergence ( $r_e$ ) of the algorithms using Eq. [15], as suggested by Knight (2008):

$$r_e = \lim_{\rightarrow +\infty} \frac{\sqrt{\sum_{i=1}^q [\Theta_i^{(k+1)} - \Theta_i^{(k)}]^2}}{\sqrt{\sum_{i=1}^q [\Theta_i^{(k)} - \Theta_i^{(k-1)}]^2}} \quad [15]$$

where  $\Theta_i$  is the vector of variance parameters of length  $q$ .

The correlation coefficients were estimated and graphs were created using R (R Core Team, 2017).

### 3. RESULTS

#### 3.1. CONVERGENCE HISTORY AND COMPUTATIONAL EFFICIENCY OF THE ALGORITHMS

The AI algorithm failed to converge for the three diallel models (Model 1, Model 1 + RGCA, and Model 3) in both datasets due to a singularity in the AI matrix (see Note S1 in the supplemental material). In addition, this algorithm updated some variance components outside the parametric space (negative values) after the first iteration for the three diallel models in Dataset 1 (see Supplemental Tables S2, S5, and S8) and Dataset 2 (see Supplemental Tables S11, S14, and S17). The other algorithms (EM and EM-AI) converged normally for the three diallel models in both datasets (see parameter convergence history in Supplemental Tables S1, S3, S4, S6, S7, S9, S10, S12, S13, S15, S16, and S18).

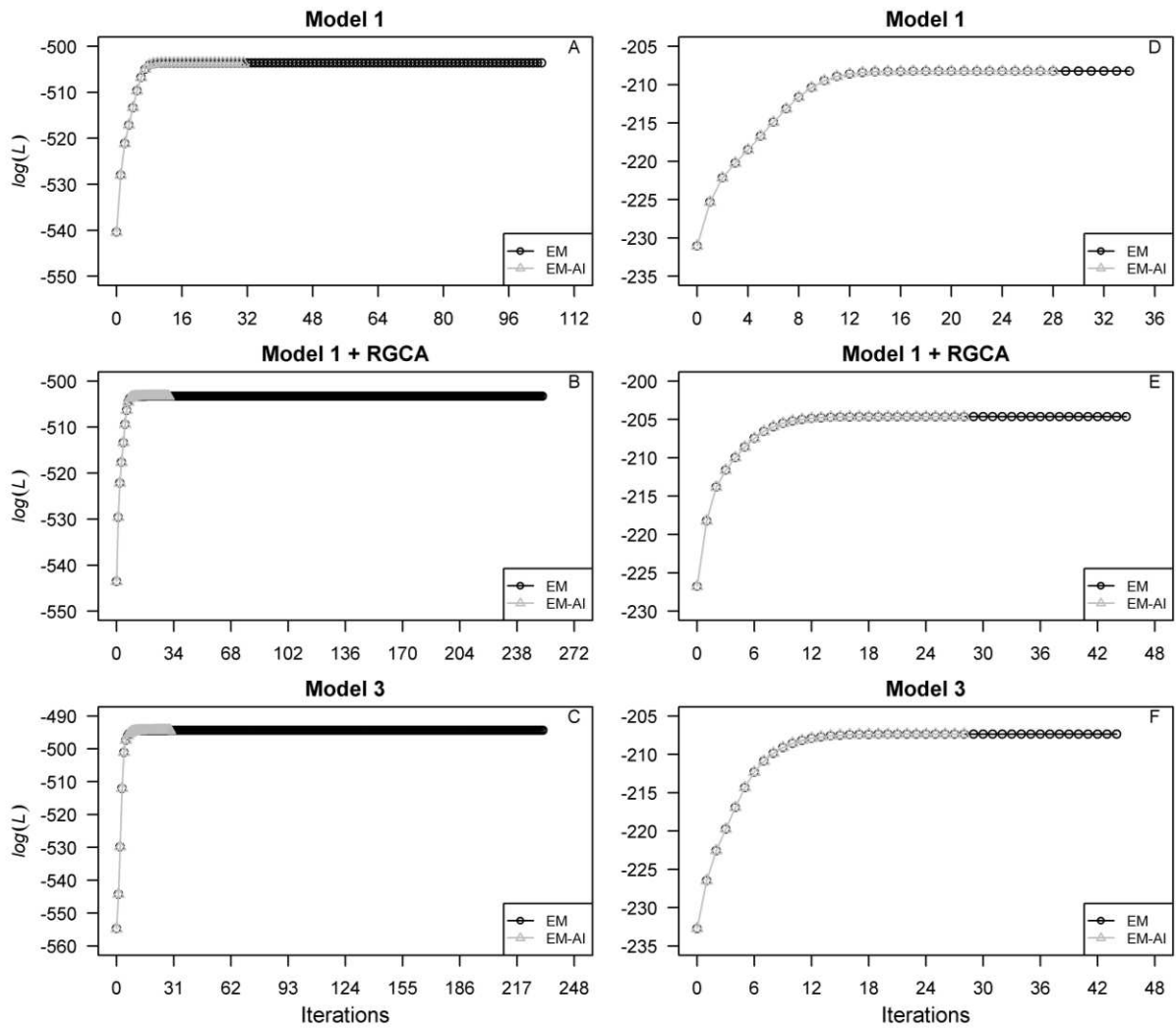


Fig. 2. Convergence history of logarithm of the residual likelihood [ $\log(L)$ ] for three diallel models (Model 1, Model 1 + reciprocal general combining ability [RGCA], and Model 3) with two datasets ([A–C] Dataset 1 and [D–F] Dataset 2), comparing expectation maximization (EM) and EM- average information (AI) algorithms.

The convergence history of  $\log(L)$  for EM and EM-AI algorithms is shown in Fig. 2 for Model 1, Model 1 + RGCA, and Model 3 in both datasets. The EM algorithm increased  $\log(L)$  rapidly after first iterations and significantly improved the initial values of variance components provided to the AI algorithm in the EM-AI algorithm (Fig. 2; Supplemental Tables S3, S6, S9, S12, S15, and S18). After the initial iterations, the EM algorithm slowly increased  $\log(L)$ ,

requiring a higher number of iterations for convergence with Dataset 1 and a moderate number with Dataset 2 (Fig. 2).

Table 1. Variance component estimates, log-likelihood, number of iterations, empirical rate of convergence, and average computation time for different diallel models and algorithms in two datasets.

Dataset <sup>†</sup>	Parameter <sup>‡</sup>	Model 1 <sup>§</sup>		Model 1 + RGCA <sup>§</sup>		Model 3 <sup>§</sup>	
		EM	EM-AI	EM	EM-AI	EM	EM-AI
1	$\hat{\sigma}_g^2$	1276.5245	1276.5246	1276.5246	1276.5246	927.8089	927.8089
	$\hat{\sigma}_s^2$	1108.9554	1108.9556	1108.9562	1108.9556	361.9043	361.9039
	$\hat{\sigma}_d^2$			11.6732	11.6734	11.6731	11.6734
	$\hat{\sigma}_r^2$	64.8975	64.8980	41.5528	41.5513	43.0911	43.0893
	$\hat{\sigma}_p^2$					10,002.0469	10,002.0462
	$\hat{\sigma}^2$	422.9799	422.9794	422.9782	422.9794	416.8259	416.8273
	$\log(L)$	-503.6195	-503.6195	-503.2784	-503.2784	-494.3903	-494.3903
	Iterations <sup>¶</sup>	104	31	253	31	231	29
	$r_e$	0.9165	0.0436	0.9645	0.0437	0.9617	<0.0001
	Time (s)	5.8100	2.0780	16.2740	2.4300	15.3960	2.2560
2	$\hat{\sigma}_g^2$	5.4831	5.4831	5.4831	5.4831	6.4100	6.4100
	$\hat{\sigma}_s^2$	4.4007	4.4007	4.4007	4.4007	4.4575	4.4576
	$\hat{\sigma}_d^2$			1.3053	1.3053	1.3054	1.3053
	$\hat{\sigma}_r^2$	4.6836	4.6836	2.0729	2.0730	2.0718	2.0720
	$\hat{\sigma}_p^2$					17.3302	17.3302
	$\hat{\sigma}^2$	3.3698	3.3697	3.3698	3.3697	3.3739	3.3738
	$\log(L)$	-208.2439	-208.2439	-204.6187	-204.6187	-207.3536	-207.3536
	Iterations <sup>¶</sup>	34	28	45	28	44	28
	$r_e$	0.6432	0.0024	0.7577	0.0042	0.7590	<0.0001
	Time (s)	2.2120	1.9720	3.0200	1.8960	3.0180	2.3340

<sup>†</sup> Dataset 1, data available in Hayman (1954b); Dataset 2, data available in Hayman (1954a).

<sup>‡</sup>  $\hat{\sigma}_g^2$ , variance of general combining ability;  $\hat{\sigma}_s^2$ , variance of specific combining ability;  $\hat{\sigma}_d^2$ , variance of reciprocal general combining ability;  $\hat{\sigma}_r^2$ , variance of reciprocal specific combining ability;  $\hat{\sigma}_p^2$ , variance of parents;  $\hat{\sigma}^2$ , variance of residuals;  $\log(L)$ , logarithm of the residual likelihood;  $r_e$ , empirical rate of convergence; time, average computing time based in five replications (s).

<sup>§</sup> Model 1, model described by Griffing's Method 1 (1956a); Model 1 + reciprocal general combining ability (RGCA), model described by Montesinos-López et al. (2009); Model 3, model described by Möhring et al. (2011); EM, expectation maximization algorithm; EM-AI, combination of EM and average information (AI) algorithms.

<sup>¶</sup> If EM-AI algorithm, iterations = 25 EM steps + required AI steps.

The EM-AI algorithm, considering 25 EM steps, converged in fewer iterations than EM algorithm for all diallel models in both datasets (Table 1, Fig. 2). In addition, the EM-AI algorithm further provided a lower empirical rate of convergence ( $r_c$ ) and lower average computation time than the EM algorithm (Table 1).

When varying the number of EM steps from 1 to 25 in the EM-AI algorithm, the EM-AI algorithm did not converge for any of the three diallel models in the two datasets used (Table 2) when the number of EM steps was  $\leq 2$ . By contrast, using six EM steps, the EM-AI algorithm converged for all models with both datasets (Table 2). Furthermore, with 12 EM steps, the number of iterations of the AI algorithm required for convergence in the EM-AI algorithm in each diallel model was constant for Dataset 1 and was reduced by one for Dataset 2 (Table 2).

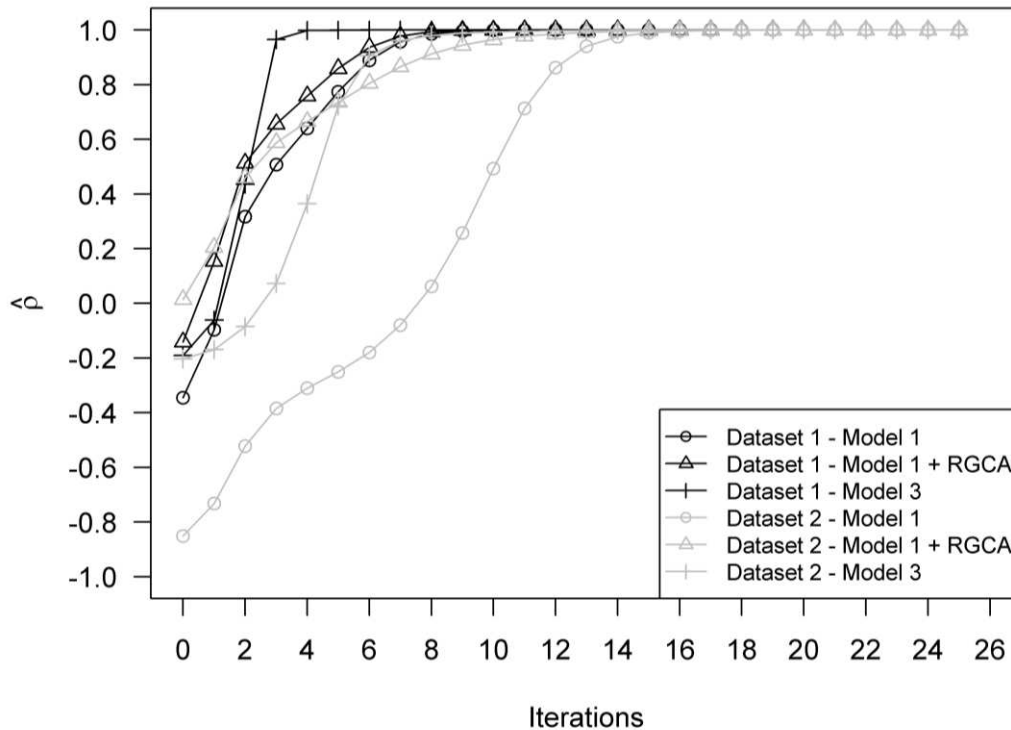


Fig. 3. Pearson's correlations between the current and converged variance components for the initial 25 expectation maximization (EM) steps for each diallel model in both datasets. RGCA, reciprocal general combining ability.

Table 2. Number of average information (AI) iterations required for convergence for different numbers of expectation maximization (EM) steps (1–25) on the two datasets.

EM steps <sup>†</sup>	Dataset 1 <sup>‡</sup>			Dataset 2 <sup>‡</sup>		
	Model 1	Model 1+ RGCA	Model 3	Model 1	Model 1+ RGCA	Model 3
1	fail	fail	fail	fail	fail	fail
2	fail	fail	fail	fail	fail	fail
3	13	fail	fail	fail	8	fail
4	10	fail	fail	fail	7	fail
5	9	9	7	fail	6	9
6	8	8	6	8	6	7
7	7	7	5	7	5	6
8	6	7	5	6	5	6
9	7	7	5	5	5	5
10	7	7	4	5	4	5
11	7	6	4	5	4	5
12	6	6	4	4	4	4
13	6	6	4	4	4	4
14	6	6	4	4	4	4
15	6	6	4	4	4	4
16	6	6	4	3	3	4
17	6	6	4	3	3	4
18	6	6	4	3	3	4
19	6	6	4	3	3	3
20	6	6	4	3	3	3
21	6	6	4	3	3	3
22	6	6	4	3	3	3
23	6	6	4	3	3	3
24	6	6	4	3	3	3
25	6	6	4	3	3	3

<sup>†</sup> EM steps, number of iterations of expectation maximization algorithm;

<sup>‡</sup> Dataset 1, data available in Hayman (1954b); Dataset 2, data available in Hayman (1954a); Model 1, model described by Griffing’s Method 1 (1956a); Model 1 + reciprocal general combining ability (RGCA), model described by Montesinos-López et al. (2009); Model 3, model described by Möhring et al. (2011).

A comparison between the initial values for variance components and the converged variance components indicates a low correlation for Model 1 + RGCA in Dataset 2 or negative correlations for the other models in both datasets (Fig. 3). However, the EM algorithm produced

a higher correlation between the current variance components and converged variance components to approximately one in the first 25 iterations (Fig. 3). Although we used 25 EM steps in the combined algorithm (EM-AI), using 12 EM iterations produced a high correlation ( $>0.80$ ) with converged variance components (Fig. 3).

### 3.2. VARIANCE COMPONENTS

The estimates of the variance components provided by the EM and EM-AI algorithms for the three models (Model 1, Model 1 + RGCA, and Model 3) with the two datasets are shown in Table 1. Contrastingly, the variance components provided by the AI algorithm are not shown in Table 1 because this algorithm did not converge in any situation.

In each of the three diallel models, the variance components estimated with EM or EM-AI algorithms were similar for both datasets (Table 1). However, the number of iterations, empirical rate of convergence, and average computation time were lower for the EM-AI algorithm than the EM algorithm (Table 1). These results indicate that the EM-AI algorithm was more efficient than the EM algorithm. Thus, a comparison of the variance components for these models will be based only on the results from the EM-AI algorithm (Table 1). Given the estimates of variance components supplied by the EM-AI algorithm, Model 1 and Model 1 + RGCA provided similar variance components estimates for the GCA ( $\hat{\sigma}_g^2$ ) and SCA ( $\hat{\sigma}_s^2$ ) with both datasets. However, we verified significant changes for the same parameters by comparing Model 3 with the other models (Model 1 and Model 1 + RGCA) with Dataset 1. We also verified that the RSCA variance component ( $\hat{\sigma}_r^2$ ) was largest in Model 1 for both datasets (i.e., the model without RGCA, Table 1).

## 4. DISCUSSION

### 4.1. CONVERGENCE HISTORY AND COMPUTATIONAL EFFICIENCY OF THE ALGORITHMS

The failure of the AI algorithm to converge was reported in the literature (Meyer, 2006; Knight, 2008; Diffey et al., 2017). According to Diffey et al. (2017), algorithms based on the observed (Hessian), expected (Fisher information), or average information matrices cannot guarantee the residual log-likelihood will increase. Furthermore, the same authors further comment that these algorithms may not update the variance components within the parametric space. According to Diffey et al. (2017), either problem may result in the algorithm failing to converge. Meyer (2006) reported different causes for failure of the AI algorithm to converge. However, the bad starting points for the variance components and the complexity of the models may be considered as the main causes of failure (Knight, 2008; Misztal, 2008). In addition, Zhou and Stephens (2014) reported that the Newton–Raphson algorithm (similar to the AI algorithm) can fail to converge when bad starting points for variance components are given.

The fast increase of  $\log(L)$  during the initial iterations of the EM algorithm and the subsequent slow increase are explained by a change in the rate of convergence. Knight (2008) showed the behavior of the empirical rate of convergence, whereby this parameter increased with more iterations. This result justifies the largest number of iterations required for convergence with the EM than with EM-AI algorithms. A greater number of iterations with the EM algorithm justifies the greater average computation time required to reach convergence. Johnson and Thompson (1995) and Gilmour et al. (1995) reported a similar result when they compared the EM and AI algorithms. In addition, a large number of iterations can increase the computational

cost because at each iteration, it is necessary to invert the left-hand side of the mixed model equations, given in Eq. [3].

The shorter computation time required for convergence with the EM-AI algorithm is indicative of its greater computational efficiency. However, this difference should be even greater for analyses involving larger datasets, such as diallel trials evaluated in several environments.

The EM-AI algorithm converged in a few iterations, even when we used few EM steps. This combination of algorithms provides fast convergence because it combines the stability of EM with the fast convergence of the AI algorithm. Zhou and Stephens (2014) and Meyer (2006) used a similar combination of algorithms and reported similar results.

The EM algorithm provided significantly higher correlation between the current and converged variance components in the first 12 iterations. Using the EM algorithm, Meyer (2006) observed that the first iterations provided variance components close to the maximum of  $\log(L)$ . This behavior can be attributed to the properties of monotonic convergence and variance component updates within the parametric space of the EM algorithm (Meyer, 2006; Knight, 2008; Diffey et al., 2017).

#### **4.1. VARIANCE COMPONENTS**

The lower magnitude of the variance component estimates of the effects of GCA ( $\hat{\sigma}_g^2$ ) and SCA ( $\hat{\sigma}_s^2$ ) observed for Model 3 compared with the other models (Model 1 and Model 1 + RGCA) for the first dataset can be attributed to the inclusion of parents when estimating the variance components of GCA ( $\hat{\sigma}_g^2$ ) and SCA ( $\hat{\sigma}_s^2$ ) in Model 1 and Model 1 + RGCA, as these estimates are free of the parents effect only in Model 3. Similar results were reported by Möhring et al. (2011) when Models 1 and 3 were compared. These authors argued that the high variability

between parents determined the discrepancy in the estimated GCA ( $\hat{\sigma}_g^2$ ) and SCA ( $\hat{\sigma}_s^2$ ) variance components for these models.

Yao et al. (2013) also reported a bias in estimates of the effects of GCA and SCA in corn crops using the Griffing's models that include parents during estimation. It is worth noting that the model described by Möhring et al. (2011) provides estimates of variance components of GCA and SCA that are free from the parent effect, even when these were included in the experiments. Nevertheless, parent information is used when estimating the residual variance. Thus, the model described by Möhring et al. (2011) is more suitable than the model described in Griffing's Method 1, in particular when the variability between parents is high. According to Griffing (1956a, 1956b), the effect of epistasis also biases estimates of the variance components for the GCA and SCA effects.

The larger magnitude of the RSCA variance component ( $\hat{\sigma}_r^2$ ) in Model 1 compared with the other models (Model 1 and Model 3 + RGCA, Table 1) can be attributed to the lack of RGCA in Model 1. Therefore, the estimated RSCA variance component ( $\hat{\sigma}_s^2$ ) in Model 1 is biased. This result, together with biased estimates of the GCA ( $\hat{\sigma}_g^2$ ) and SCA ( $\hat{\sigma}_s^2$ ) variance components, indicates advantages of using Model 3 in diallel analysis in detriment to the other two models (Model 1 and Model 1 + RGCA), which include parents in the experiments.

Extranuclear genes contribute to the RGCA variance component ( $\hat{\sigma}_d^2$ ), whereas the RSCA variance component ( $\hat{\sigma}_r^2$ ) is caused by the interaction between nuclear and extranuclear genes, as well as between different extranuclear genes. These effects are especially important for plant breeding with an emphasis on obtaining superior clones or hybrids, allowing the definition of the cross directions (i.e., defining which parent to use as female or male). Moreover, Model 3

facilitates estimation of parameters that provide a better understanding of the genetic effects involved in genetic control of the characters under analysis (i.e., the variance of the GCA, SCA, RGCA, and RSCA effects). This contributes to the definition of more adequate strategies for improving these traits.

It should be emphasized that, according to Cockerham and Weir (1977), the GCA ( $\hat{\sigma}_g^2$ ) and SCA ( $\hat{\sigma}_s^2$ ) variance components are confounded with the RGCA and RSCA variance components, respectively. Thus, to estimate the variance components due to nuclear genes ( $\hat{\sigma}_n^2$ ) and the interaction between parental nuclear genes ( $\hat{\sigma}_t^2$ ), the following estimators can be used:  $\hat{\sigma}_n^2 = \hat{\sigma}_g^2 - \hat{\sigma}_d^2$  and  $\hat{\sigma}_t^2 = \hat{\sigma}_s^2 - \hat{\sigma}_r^2$ , respectively. In addition, Fan et al. (2014) reported that the GCA ( $\hat{\sigma}_g^2$ ) and SCA ( $\hat{\sigma}_s^2$ ) variance components are confounded with the variances of reciprocal effects, called variance of the maternal effect and variance of the nonmaternal effect by the authors.

## 5. CONCLUSIONS

The EM-AI algorithm was more efficient than EM and AI algorithms for diallel models. The RGCA and RSCA variance components and the parent effects impact the GCA and SCA variance components, especially when they exhibited large magnitude. The EM-AI algorithm implemented in R has potential use in plant breeding and diallel analysis with the goal of obtaining superior hybrids or clones.

## 6. REFERENCES

Bates, D., and M. Maechler. 2017. Matrix: Sparse and dense matrix classes and methods. R

Package Version 1.2-12. Comprehensive R Arch. Network.

- <https://mran.microsoft.com/snapshot/2018-03-30/web/packages/Matrix/index.html>, accessed: October 15<sup>th</sup>, 2017.
- Bates, D., and D. Sarkar. 2005. lme4: Linear mixed-effects models using S4 classes. R Package Version 0.95-1. Comprehensive R Arch. Network. <https://www.rdocumentation.org/packages/lme4/versions/0.95-1/topics/lmerControl>, accessed: November 22<sup>th</sup>, 2017.
- Butler, D., B.R. Cullis, A.R. Gilmour, and B.J. Gogel. 2009. ASReml-R reference manual. Release 3.0. State Queensland, Dep. Prim. Ind. Fish., Brisbane, QLD, Australia.
- Cockerham, C.C., and B.S. Weir. 1977. Quadratic analyses of reciprocal crosses. *Biometrics* 33:187–203. doi:10.2307/2529312
- Dempster, A.P., N.M. Laird, and D.B. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B Methodol.* 39:1–38.
- Diffey, S.M., A.B. Smith, A.H. Welsh, and B.R. Cullis. 2017. A new REML (parameter expanded) EM algorithm for linear mixed models. *Aust. N. Z. J. Stat.* 59:433–448. doi:10.1111/anzs.12208
- Fan, X.M., Y.D. Zhang, W.H. Yao, Y.Q. Bi, L. Liu, H.M. Chen, and M.S. Kang. 2014. Reciprocal diallel crosses impact combining ability, variance estimation, and heterotic group classification. *Crop Sci.* 54:89–97. doi:10.2135/cropsci2013.06.0393
- Foulley, J.L., and D.A. Van Dyk. 2000. The PX-EM algorithm for fast stable fitting of Henderson's mixed model. *Genet. Sel. Evol.* 32:143–163. doi:10.1186/1297-9686-32-2-143
- Gilmour, A.R., R. Thompson, and B.R. Cullis. 1995. Average information REML: An efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics* 51:1440–1450. doi:10.2307/2533274

- Griffing, B. 1956a. A generalised treatment of the use of diallel crosses in quantitative inheritance. *Heredity* 10:31–50. doi:10.1038/hdy.1956.2
- Griffing, B. 1956b. Concept of general and specific combining ability in relation to diallel crossing systems. *Aust. J. Biol. Sci.* 9:463–493. doi:10.1071/BI9560463
- Hayman, B.I. 1954a. The analysis of variance of diallel tables. *Biometrics* 10:235–244. doi:10.2307/3001877
- Hayman, B.I. 1954b. The theory and analysis of diallel crosses. *Genetics* 39:789–809.
- Henderson, C.R., O. Kempthorne, S.R. Searle, and C.M. Von Krosigk. 1959. The estimation of environmental and genetic trends from records subject to culling. *Biometrics* 15:192–218. doi:10.2307/2527669
- Johnson, D.L., and R. Thompson. 1995. Restricted maximum likelihood estimation of variance components for univariate animal models using sparse matrix techniques and average information. *J. Dairy Sci.* 78:449–456. doi:10.3168/jds.S0022-0302(95)76654-1
- Knight, E. 2008. Improved iterative schemes for REML estimation of variance parameters in linear mixed models. Ph.D. diss., Univ. Adelaide, Adelaide, SA, Australia.
- Kusnierczyk, W. 2015. rbenchmark: Benchmarking routine for R. R Package version 1.0.0. Comprehensive R Arch. Network. <https://cran.r-project.org/web/packages/rbenchmark/index.html>, accessed: December 22<sup>th</sup>, 2018.
- Lee, S.H., and J.H.J. van der Werf. 2016. MTG2: An efficient algorithm for multivariate linear mixed model analysis based on genomic information. *Bioinformatics* 32:1420–1422. doi:10.1093/bioinformatics/btw012
- Meyer, K. 1997. An ‘average information’ restricted maximum likelihood algorithm for estimating reduced rank genetic covariance matrices or covariance functions for animal

- models with equal design matrices. *Genet. Sel. Evol.* 29:97–116. doi:10.1186/1297-9686-29-2-97
- Meyer, K. 2006.  $PX \times AI$ : Algorithmic for better convergence in restricted maximum likelihood estimation. Paper presented at the 8th World Congress on Genetics Applied to Livestock Production, Belo Horizonte, Brazil. 13–18 Aug. 2006. *Commun.* 24-00.
- Meyer, K. 2008. Parameter expansion for estimation of reduced rank covariance matrices. *Genet. Sel. Evol.* 40:3–24. doi:10.1051/gse:2007032
- Meyer, K., and S.P. Smith. 1996. Restricted maximum likelihood estimation for animal models using derivatives of the likelihood. *Genet. Sel. Evol.* 28:23–49. doi:10.1186/1297-9686-28-1-23
- Misztal, I. 2008. Reliable computing in estimation of variance components. *J. Anim. Breed. Genet.* 125:363–370. doi:10.1111/j.1439-0388.2008.00774.x
- Möhring, J., A.E. Melchinger, and H.P. Piepho. 2011. REML-based diallel analysis. *Crop Sci.* 51:470–478. doi:10.2135/cropsci2010.05.0272
- Montesinos-López, O.A., Á.A. Mastache-Agunas, I. Luna-Spinoza, C.M. Hernández-Suárez, and G. Hernández-Lira. 2009. Mejor predictor lineal e insesgado familiar de aptitud combinatoria general en experimentos parciales de cruzas dialélicas con efectos maternos. *Agric. Téc. Méx.* 35:131–146.
- Mrode, R.A., and R. Thompson, editors. 2005. Linear models for the prediction of animal breeding values. Cromwell Press, Trowbridge, UK. doi:10.1079/9780851990002.0000
- Patterson, H.D., and R. Thompson. 1971. Recovery of inter-block information when block sizes are unequal. *Biometrika* 58:545–554. doi:10.1093/biomet/58.3.545

- R Core Team. 2017. R: A language and environment for statistical computing. R Found. Stat. Comput., Vienna, Austria.
- Rukundo, P., H. Shimelis, M. Laing, and D. Gahakwa. 2017. Combining ability, maternal effects, and heritability of drought tolerance, yield and yield components in sweetpotato. *Front. Plant Sci.* 7:1–14. doi:10.3389/fpls.2016.01981
- Searle, S.R., G. Casella, and C.E. McCulloch. 1992. Variance components. John Wiley & Sons, New York.
- Smith, A.B., B.R. Cullis, and R. Thompson. 2005. The analysis of crop cultivar breeding and evaluation trials: An overview of current mixed model approaches. *J. Agric. Sci.* 143:449–462. doi:10.1017/S0021859605005587
- Tenesa, A., and C.S. Haley. 2013. The heritability of human disease: Estimation, uses and abuses. *Nat. Rev. Genet.* 14:139–149. doi:10.1038/nrg3377
- Thompson, R. 2008. Estimation of quantitative genetic parameters. *Proc. Biol. Sci.* 275:679–686. doi:10.1098/rspb.2007.1417
- Thompson, R., S. Brotherstone, and I.M.S. White. 2005. Estimation of quantitative genetic parameters. *Philos. Trans. R. Soc., B* 360:1469–1477. doi:10.1098/rstb.2005.1676
- Xiang, B., and B. Li. 2001. A new mixed analytical method for genetic analysis of diallel data. *Can. J. For. Res.* 31:2252–2259. doi:10.1139/x01-154
- Yao, W.H., Y.D. Zhang, M.S. Kang, H.M. Chen, L. Liu, L.J. Yu, and X.M. Fan. 2013. Diallel analysis models: A comparison of certain genetic statistics. *Crop Sci.* 53:1481–1490. doi:10.2135/cropsci2013.01.0027
- Zhou, X., and M. Stephens. 2012. Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* 44:821–824. doi:10.1038/ng.2310

Zhou, X., and M. Stephens. 2014. Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nat. Methods* 11:407–409. doi:10.1038/nmeth.2848

## 7. SUPPLEMENTAL MATERIALS

### 7.1. ROUTINE

Routine S1. Proposed routine to run diallel analysis with reciprocal crosses by REML-BLUP approach via EM-AI algorithm applied to R software.

The routine is available online in file ‘Routine’ <<https://dl.sciencesocieties.org/publications/cs/abstracts/59/4/1494>>. The data file used should be in the following sequence: crosses, blocks, male, female, grouping.crosses, cross.classes, and traits. Although, the names of columns do not need necessarily to be the same described before.

- ✓ crosses: Numerical code for distinguishing the levels of crosses.
- ✓ blocks: Numerical code for distinguishing the levels of blocks.
- ✓ male: Numerical code for distinguishing the levels of male.
- ✓ female: Numerical code for distinguishing the levels of female.
- ✓ grouping.crosses: Numerical code for distinguishing the levels of crosses. The reciprocal crosses assume the same code of the direct crosses. This column is used to design the incidence matrix of SCA.
- ✓ crosses.class: Numerical code for distinguishing the classes of cross, i.e. code 0 for parents, code 1 for crosses, and code -1 for reciprocal crosses. This column is used to design the incidence matrix of RSCA.
- ✓ traits: represents the traits evaluated, g.e.  $y_1, y_2$  to  $y_i$ .

If any doubt persists to see the examples in ‘Routine’.

## 7.2. NOTES

### Note S1. Singularity in Average Information Matrix

The AI matrix presented singularity in six instances ([S\_1] to [S\_6]) for the three diallel models in both datasets. Therefore, we observed abnormal termination of procedure due singularity in the rounds 4, 2, 2, 3, 3, 3 for AI matrices shown [S\_1] to [S\_6], respectively.

$$AI_{\substack{\text{Model 1} \\ \text{Dataset 1}}} = \begin{array}{cccc|c} & GCA & RGCA & SCA & Residuals & \\ \hline & 4.422298E-13 & 3.800412E-13 & 0.000000E+00 & 1.796558E-13 & GCA \\ & 3.800412E-13 & -8.603293E-11 & -5.289725E-38 & -3.822762E-11 & RGCA \\ & 1.876230E-39 & -1.108801E-38 & -3.682326E-41 & -6.117327E-39 & SCA \\ & 1.796558E-13 & -3.822762E-11 & -1.763242E-38 & -1.698569E-11 & Residuals \end{array} \quad [S_1]$$

$$AI_{\substack{\text{Model 1+RGCA} \\ \text{Dataset 1}}} = \begin{array}{ccccc|c} & GCA & RGCA & SCA & RSCA & Residuals & \\ \hline & 2.828333E-08 & 1.063576E-29 & 2.377981E-08 & -1.615587E-27 & 1.125689E-08 & GCA \\ & -8.406475E-30 & -1.993396E-23 & -2.591186E-29 & -2.491745E-24 & -6.229488E-25 & RGCA \\ & 2.377981E-08 & 4.415156E-29 & 1.601805E-06 & 8.077936E-27 & 7.124888E-07 & SCA \\ & 3.502799E-29 & -2.491745E-24 & -1.374663E-27 & -2.008682E-15 & -5.021704E-16 & RSCA \\ & 1.125689E-08 & -6.229083E-25 & 7.124888E-07 & -5.021704E-16 & 3.169342E-07 & Residuals \end{array} \quad [S_2]$$

$$AI_{\substack{\text{Model 3} \\ \text{Dataset 1}}} = \begin{array}{cccccc|c} & GCA & RGCA & SCA & RSCA & Parents & Residuals & \\ \hline & 4.538620E-06 & 2.398137E-28 & 7.564367E-07 & 1.518652E-25 & -4.936158E-40 & 1.891092E-07 & GCA \\ & -1.454062E-28 & -1.993396E-23 & 2.634593E-29 & -2.491745E-24 & -1.676358E-47 & -6.229297E-25 & RGCA \\ & 7.564367E-07 & -1.045043E-27 & 4.615674E-05 & -1.809458E-25 & -8.035606E-41 & 1.153918E-05 & SCA \\ & 2.410759E-25 & -2.491745E-24 & -7.634754E-26 & -2.008682E-15 & -1.116660E-45 & -5.021704E-16 & RSCA \\ & 5.877968E-28 & 2.782969E-35 & 1.088598E-28 & 3.478711E-36 & 1.876268E-07 & 9.381342E-08 & Parents \\ & 1.891092E-07 & -6.226234E-25 & 1.153918E-05 & -5.021704E-16 & 9.381342E-08 & 2.931703E-06 & Residuals \end{array} \quad [S_3]$$

$$AI_{\substack{\text{Model 1} \\ \text{Dataset 2}}} = \begin{array}{cccc|c} & GCA & RGCA & SCA & Residuals & \\ \hline & -3.802719E-10 & -2.738895E-09 & -7.031035E-24 & -1.172200E-09 & GCA \\ & -2.738895E-09 & -2.107946E-08 & 2.852119E-21 & -9.002786E-09 & RGCA \\ & 1.693500E-21 & 8.470740E-21 & 4.584532E-01 & 1.146133E-01 & SCA \\ & -1.172200E-09 & -9.002786E-09 & 1.146133E-01 & 2.865332E-02 & Residuals \end{array} \quad [S_4]$$

$$AI_{\substack{\text{Model 1+RGCA} \\ \text{Dataset 2}}} = \begin{array}{ccccc|c} & GCA & RGCA & SCA & RSCA & Residuals & \\ \hline & -3.802719E-10 & -4.391585E-20 & -2.738895E-09 & 1.214963E-20 & -1.172200E-09 & GCA \\ & -3.091666E-20 & 1.255340E+00 & 1.969353E-20 & 1.569175E-01 & 3.922937E-02 & RGCA \\ & -2.738895E-09 & 2.688999E-20 & -2.107946E-08 & 4.494899E-21 & -9.002786E-09 & SCA \\ & -6.013939E-20 & 1.569175E-01 & 1.355251E-20 & 1.254965E+00 & 3.137413E-01 & RSCA \\ & -1.172200E-09 & 3.922937E-02 & -9.002786E-09 & 3.137413E-01 & 7.843533E-02 & Residuals \end{array} \quad [S_5]$$

$$AI_{\substack{\text{Model 3} \\ \text{Dataset 2}}} = \begin{bmatrix} 6.834482\text{E-02} & 4.770490\text{E-18} & 1.139080\text{E-02} & -1.067939\text{E-17} & 3.526483\text{E-38} & 2.847701\text{E-03} \\ -1.127570\text{E-17} & 1.255340\text{E+00} & 6.532318\text{E-18} & 1.569175\text{E-01} & 7.346840\text{E-40} & 3.922938\text{E-02} \\ 1.139080\text{E-02} & -4.987330\text{E-18} & 3.577550\text{E-01} & 3.155028\text{E-17} & -5.510130\text{E-40} & 8.943874\text{E-02} \\ -3.469447\text{E-18} & 1.569175\text{E-01} & -3.447763\text{E-17} & 1.254965\text{E+00} & -1.262738\text{E-40} & 3.137414\text{E-01} \\ -2.899706\text{E-38} & 8.035606\text{E-41} & 4.177798\text{E-39} & 4.017803\text{E-41} & 9.709104\text{E-03} & 4.854552\text{E-03} \\ 2.847701\text{E-03} & 3.922937\text{E-02} & 8.943874\text{E-02} & 3.137413\text{E-01} & 4.854552\text{E-03} & 1.032223\text{E-01} \end{bmatrix} \begin{matrix} GCA \\ RGCA \\ SCA \\ RSCA \\ Parents \\ Residuals \end{matrix} \quad [S_6]$$

### 7.3. TABLES

Table S1. Convergence history for the variance components and  $\log(L)$  of the Model 1 via EM algorithm for dataset 1.

Iterations	$\hat{\sigma}_g^2$	$\hat{\sigma}_s^2$	$\hat{\sigma}_r^2$	$\hat{\sigma}^2$	$\log(L)$
0 <sup>†</sup>	167.5077	167.5077	167.5077	1675.0775	-540.3835
1 <sup>‡</sup>	617.4592	189.3851	141.1183	1443.2457	-527.9881
2	1046.9901	215.6041	120.7506	1055.0034	-521.1097
3	1202.2586	277.7570	104.3461	891.1994	-517.1226
4	1249.2223	380.4936	91.9624	791.2867	-513.3164
5	1269.0307	520.7672	82.6301	698.4776	-509.6750
6	1280.1460	681.0526	75.6803	610.8340	-506.7744
7	1285.1963	829.9993	70.7486	539.3658	-504.9698
8	1285.5430	943.1124	67.5298	489.9266	-504.1039
9	1283.5500	1016.4080	65.6270	459.8496	-503.7723
10	1281.1770	1059.1700	64.6054	442.9895	-503.6643
11	1279.3042	1082.6210	64.1081	433.9516	-503.6324
12	1278.0721	1095.0534	63.8985	429.2030	-503.6234
13	1277.3401	1101.5325	63.8382	426.7172	-503.6209
14	1276.9315	1104.8858	63.8519	425.4034	-503.6201
15	1276.7128	1106.6221	63.9001	424.6926	-503.6199
16	1276.5993	1107.5279	63.9626	424.2923	-503.6198
17	1276.5421	1108.0082	64.0294	424.0528	-503.6197
18	1276.5143	1108.2704	64.0956	423.8978	-503.6197
19	1276.5017	1108.4202	64.1589	423.7885	-503.6197
20	1276.4967	1108.5117	64.2183	423.7048	-503.6196
21	1276.4955	1108.5723	64.2735	423.6364	-503.6196
22	1276.4961	1108.6160	64.3246	423.5778	-503.6196
23	1276.4975	1108.6502	64.3717	423.5262	-503.6196
24	1276.4993	1108.6785	64.4150	423.4800	-503.6196
25	1276.5012	1108.7029	64.4548	423.4380	-503.6196
26	1276.5030	1108.7246	64.4914	423.3998	-503.6196
27	1276.5047	1108.7441	64.5250	423.3649	-503.6196
28	1276.5063	1108.7618	64.5559	423.3329	-503.6196

29	1276.5078	1108.7779	64.5842	423.3036	-503.6196
30	1276.5092	1108.7927	64.6102	423.2767	-503.6196
31	1276.5105	1108.8062	64.6341	423.2520	-503.6196
32	1276.5116	1108.8186	64.6560	423.2294	-503.6196
33	1276.5127	1108.8300	64.6761	423.2086	-503.6196
34	1276.5137	1108.8404	64.6945	423.1896	-503.6196
35	1276.5146	1108.8500	64.7114	423.1721	-503.6195
36	1276.5154	1108.8588	64.7269	423.1561	-503.6195
37	1276.5162	1108.8668	64.7411	423.1414	-503.6195
38	1276.5169	1108.8742	64.7542	423.1279	-503.6195
39	1276.5175	1108.8810	64.7661	423.1156	-503.6195
40	1276.5181	1108.8872	64.7771	423.1042	-503.6195
41	1276.5186	1108.8929	64.7872	423.0939	-503.6195
42	1276.5191	1108.8981	64.7964	423.0843	-503.6195
43	1276.5196	1108.9029	64.8049	423.0756	-503.6195
44	1276.5200	1108.9073	64.8126	423.0676	-503.6195
45	1276.5204	1108.9113	64.8197	423.0602	-503.6195
46	1276.5207	1108.9150	64.8263	423.0535	-503.6195
47	1276.5210	1108.9184	64.8322	423.0473	-503.6195
48	1276.5213	1108.9215	64.8377	423.0417	-503.6195
49	1276.5216	1108.9243	64.8427	423.0365	-503.6195
50	1276.5218	1108.9270	64.8473	423.0317	-503.6195
51	1276.5221	1108.9293	64.8516	423.0274	-503.6195
52	1276.5223	1108.9315	64.8554	423.0234	-503.6195
53	1276.5225	1108.9335	64.8590	423.0197	-503.6195
54	1276.5226	1108.9354	64.8622	423.0163	-503.6195
55	1276.5228	1108.9371	64.8652	423.0133	-503.6195
56	1276.5229	1108.9386	64.8680	423.0104	-503.6195
57	1276.5231	1108.9400	64.8705	423.0079	-503.6195
58	1276.5232	1108.9413	64.8728	423.0055	-503.6195
59	1276.5233	1108.9425	64.8749	423.0033	-503.6195
60	1276.5234	1108.9436	64.8768	423.0013	-503.6195
61	1276.5235	1108.9446	64.8786	422.9995	-503.6195
62	1276.5236	1108.9455	64.8802	422.9978	-503.6195
63	1276.5237	1108.9464	64.8817	422.9963	-503.6195
64	1276.5238	1108.9472	64.8830	422.9949	-503.6195
65	1276.5238	1108.9479	64.8843	422.9936	-503.6195
66	1276.5239	1108.9485	64.8854	422.9924	-503.6195
67	1276.5239	1108.9491	64.8865	422.9913	-503.6195
68	1276.5240	1108.9497	64.8874	422.9903	-503.6195
69	1276.5240	1108.9502	64.8883	422.9894	-503.6195
70	1276.5241	1108.9506	64.8891	422.9886	-503.6195
71	1276.5241	1108.9510	64.8899	422.9878	-503.6195

72	1276.5242	1108.9514	64.8905	422.9871	-503.6195
73	1276.5242	1108.9518	64.8912	422.9865	-503.6195
74	1276.5242	1108.9521	64.8917	422.9859	-503.6195
75	1276.5242	1108.9524	64.8923	422.9853	-503.6195
76	1276.5243	1108.9527	64.8927	422.9849	-503.6195
77	1276.5243	1108.9529	64.8932	422.9844	-503.6195
78	1276.5243	1108.9531	64.8936	422.9840	-503.6195
79	1276.5243	1108.9533	64.8940	422.9836	-503.6195
80	1276.5244	1108.9535	64.8943	422.9833	-503.6195
81	1276.5244	1108.9537	64.8946	422.9829	-503.6195
82	1276.5244	1108.9539	64.8949	422.9826	-503.6195
83	1276.5244	1108.9540	64.8951	422.9824	-503.6195
84	1276.5244	1108.9541	64.8954	422.9821	-503.6195
85	1276.5244	1108.9543	64.8956	422.9819	-503.6195
86	1276.5244	1108.9544	64.8958	422.9817	-503.6195
87	1276.5244	1108.9545	64.8960	422.9815	-503.6195
88	1276.5245	1108.9546	64.8962	422.9813	-503.6195
89	1276.5245	1108.9547	64.8963	422.9812	-503.6195
90	1276.5245	1108.9548	64.8964	422.9810	-503.6195
91	1276.5245	1108.9548	64.8966	422.9809	-503.6195
92	1276.5245	1108.9549	64.8967	422.9808	-503.6195
93	1276.5245	1108.9550	64.8968	422.9807	-503.6195
94	1276.5245	1108.9550	64.8969	422.9806	-503.6195
95	1276.5245	1108.9551	64.8970	422.9805	-503.6195
96	1276.5245	1108.9551	64.8971	422.9804	-503.6195
97	1276.5245	1108.9552	64.8972	422.9803	-503.6195
98	1276.5245	1108.9552	64.8972	422.9802	-503.6195
99	1276.5245	1108.9552	64.8973	422.9802	-503.6195
100	1276.5245	1108.9553	64.8974	422.9801	-503.6195
101	1276.5245	1108.9553	64.8974	422.9800	-503.6195
102	1276.5245	1108.9553	64.8975	422.9800	-503.6195
103	1276.5245	1108.9554	64.8975	422.9799	-503.6195
104	1276.5245	1108.9554	64.8975	422.9799	-503.6195

<sup>†</sup> and <sup>‡</sup> represents initial values for variance components and first iteration of the EM algorithm, respectively.

Table S2. Convergence history for the variance components and  $\log(L)$  of the Model 1 via AI algorithm for dataset 1.

Iterations	$\hat{\sigma}_g^2$	$\hat{\sigma}_s^2$	$\hat{\sigma}_r^2$	$\hat{\sigma}^2$	$\log(L)$
0 <sup>†</sup>	167.5077	167.5077	167.5077	1675.0770	-540.3835
1 <sup>‡</sup>	353.0179	1569.4560	-363.6132	-1912.4120	NA
2	793.9722	7346.8700	-2206.8510	-14513.7900	NA

3	-47237.2100	-990525.6000	-769301.2000	2232348.0000	NA
4 <sup>§</sup>	-51642310.0000	-943351700.0000	-792493600.0000	2122558000.0000	NA

<sup>†</sup>, <sup>‡</sup> and <sup>§</sup> represents initial values for variance components, first iteration of the AI algorithm, and fail to converge due singularity in AI matrix, respectively.

Table S3. Convergence history for the variance components and  $\log(L)$  of the Model 1 via EM-AI algorithm for dataset 1.

Iterations	$\hat{\sigma}_g^2$	$\hat{\sigma}_s^2$	$\hat{\sigma}_r^2$	$\hat{\sigma}^2$	$\log(L)$
0 <sup>†</sup>	167.5077	167.5077	167.5077	1675.0775	-540.3835
1 <sup>‡</sup>	617.4592	189.3851	141.1183	1443.2457	-527.9881
2	1046.9901	215.6041	120.7506	1055.0034	-521.1097
3	1202.2586	277.7570	104.3461	891.1994	-517.1226
4	1249.2223	380.4936	91.9624	791.2867	-513.3164
5	1269.0307	520.7672	82.6301	698.4776	-509.6750
6	1280.1460	681.0526	75.6803	610.8340	-506.7744
7	1285.1963	829.9993	70.7486	539.3658	-504.9698
8	1285.5430	943.1124	67.5298	489.9266	-504.1039
9	1283.5500	1016.4080	65.6270	459.8496	-503.7723
10	1281.1770	1059.1700	64.6054	442.9895	-503.6643
11	1279.3042	1082.6210	64.1081	433.9516	-503.6324
12	1278.0721	1095.0534	63.8985	429.2030	-503.6234
13	1277.3401	1101.5325	63.8382	426.7172	-503.6209
14	1276.9315	1104.8858	63.8519	425.4034	-503.6201
15	1276.7128	1106.6221	63.9001	424.6926	-503.6199
16	1276.5993	1107.5279	63.9626	424.2923	-503.6198
17	1276.5421	1108.0082	64.0294	424.0528	-503.6197
18	1276.5143	1108.2704	64.0956	423.8978	-503.6197
19	1276.5017	1108.4202	64.1589	423.7885	-503.6197
20	1276.4967	1108.5117	64.2183	423.7048	-503.6196
21	1276.4955	1108.5723	64.2735	423.6364	-503.6196
22	1276.4961	1108.6160	64.3246	423.5778	-503.6196
23	1276.4975	1108.6502	64.3717	423.5262	-503.6196
24	1276.4993	1108.6785	64.4150	423.4800	-503.6196
25	1276.5012	1108.7029	64.4548	423.4380	-503.6196
26 <sup>§</sup>	1276.5130	1108.9094	64.8957	422.9862	-503.6195
27	1276.5240	1108.9537	64.8979	422.9797	-503.6195
28	1276.5245	1108.9555	64.8980	422.9794	-503.6195
29	1276.5245	1108.9556	64.8980	422.9794	-503.6195
30	1276.5246	1108.9556	64.8980	422.9794	-503.6195
31	1276.5246	1108.9556	64.8980	422.9794	-503.6195

$\dagger$ ,  $\ddagger$  and  $\S$  represents initial values for variance components, first iteration of the EM algorithm and first iteration of the AI algorithm, respectively.

Table S4. Convergence history for the variance components and  $\log(L)$  of the Model 1 + RGCA via EM algorithm for dataset 1.

Iterations	$\hat{\sigma}_g^2$	$\hat{\sigma}_d^2$	$\hat{\sigma}_s^2$	$\hat{\sigma}_r^2$	$\hat{\sigma}^2$	$\log(L)$
0 <sup>†</sup>	167.5077	167.5077	167.5077	167.5077	1675.0775	-543.5475
1 <sup>‡</sup>	617.4592	78.2859	189.3851	144.1255	1416.6692	-529.5908
2	1049.0936	48.3021	216.9843	124.1736	1027.0948	-522.1382
3	1204.6085	33.6947	282.9308	107.0105	863.1787	-517.6222
4	1251.8830	25.7295	392.4954	93.5279	761.4507	-513.3392
5	1271.9793	20.9175	541.5116	83.1134	666.4277	-509.3666
6	1282.9049	17.8118	708.5549	75.1330	578.1695	-506.3316
7	1287.2761	15.7549	858.7158	69.1775	508.5184	-504.5440
8	1286.9034	14.3893	968.3515	64.8958	462.3407	-503.7425
9	1284.4701	13.4784	1036.7167	61.8784	435.6507	-503.4574
10	1281.9215	12.8570	1075.1867	59.7155	421.6886	-503.3695
11	1280.0059	12.4170	1095.4994	58.0909	414.9885	-503.3418
12	1278.7681	12.0928	1105.7624	56.8009	412.1266	-503.3302
13	1278.0273	11.8457	1110.7331	55.7259	411.1965	-503.3229
14	1277.5983	11.6526	1112.9967	54.7978	411.1977	-503.3171
15	1277.3498	11.4995	1113.9046	53.9773	411.6222	-503.3122
16	1277.2016	11.3770	1114.1485	53.2403	412.2177	-503.3080
17	1277.1081	11.2788	1114.0780	52.5715	412.8621	-503.3044
18	1277.0448	11.2004	1113.8678	51.9601	413.4983	-503.3014
19	1276.9983	11.1381	1113.6040	51.3982	414.1015	-503.2988
20	1276.9617	11.0893	1113.3279	50.8796	414.6625	-503.2966
21	1276.9312	11.0518	1113.0588	50.3993	415.1797	-503.2947
22	1276.9048	11.0238	1112.8049	49.9533	415.6546	-503.2930
23	1276.8815	11.0039	1112.5690	49.5379	416.0904	-503.2915
24	1276.8605	10.9907	1112.3513	49.1502	416.4905	-503.2903
25	1276.8413	10.9832	1112.1510	48.7875	416.8583	-503.2891
26	1276.8238	10.9805	1111.9668	48.4476	417.1970	-503.2881
27	1276.8077	10.9819	1111.7972	48.1285	417.5095	-503.2873
28	1276.7928	10.9866	1111.6409	47.8283	417.7984	-503.2865
29	1276.7790	10.9941	1111.4966	47.5456	418.0660	-503.2858
30	1276.7662	11.0039	1111.3632	47.2789	418.3144	-503.2851
31	1276.7543	11.0156	1111.2395	47.0270	418.5454	-503.2845
32	1276.7432	11.0288	1111.1246	46.7887	418.7607	-503.2840
33	1276.7328	11.0432	1111.0177	46.5632	418.9617	-503.2836
34	1276.7231	11.0586	1110.9180	46.3494	419.1497	-503.2831
35	1276.7140	11.0747	1110.8249	46.1465	419.3259	-503.2827

36	1276.7055	11.0912	1110.7378	45.9538	419.4912	-503.2824
37	1276.6975	11.1081	1110.6561	45.7707	419.6466	-503.2821
38	1276.6899	11.1252	1110.5795	45.5965	419.7930	-503.2818
39	1276.6828	11.1424	1110.5074	45.4307	419.9309	-503.2815
40	1276.6760	11.1595	1110.4395	45.2727	420.0612	-503.2813
41	1276.6696	11.1765	1110.3754	45.1221	420.1844	-503.2810
42	1276.6636	11.1933	1110.3149	44.9784	420.3011	-503.2808
43	1276.6579	11.2099	1110.2577	44.8413	420.4117	-503.2806
44	1276.6524	11.2261	1110.2035	44.7104	420.5167	-503.2805
45	1276.6473	11.2420	1110.1521	44.5853	420.6164	-503.2803
46	1276.6423	11.2575	1110.1033	44.4657	420.7113	-503.2802
47	1276.6377	11.2726	1110.0569	44.3514	420.8016	-503.2800
48	1276.6332	11.2873	1110.0128	44.2420	420.8876	-503.2799
49	1276.6289	11.3016	1109.9707	44.1373	420.9697	-503.2798
50	1276.6249	11.3155	1109.9307	44.0370	421.0480	-503.2797
51	1276.6210	11.3289	1109.8925	43.9410	421.1229	-503.2796
52	1276.6173	11.3418	1109.8560	43.8490	421.1944	-503.2795
53	1276.6137	11.3544	1109.8212	43.7609	421.2628	-503.2794
54	1276.6103	11.3665	1109.7879	43.6764	421.3283	-503.2793
55	1276.6071	11.3781	1109.7560	43.5954	421.3909	-503.2793
56	1276.6040	11.3894	1109.7256	43.5176	421.4510	-503.2792
57	1276.6010	11.4002	1109.6964	43.4431	421.5086	-503.2791
58	1276.5981	11.4107	1109.6684	43.3715	421.5638	-503.2791
59	1276.5954	11.4207	1109.6416	43.3029	421.6167	-503.2790
60	1276.5927	11.4304	1109.6158	43.2369	421.6675	-503.2790
61	1276.5902	11.4397	1109.5912	43.1736	421.7163	-503.2789
62	1276.5878	11.4486	1109.5675	43.1128	421.7631	-503.2789
63	1276.5854	11.4572	1109.5448	43.0544	421.8081	-503.2789
64	1276.5832	11.4655	1109.5229	42.9983	421.8514	-503.2788
65	1276.5810	11.4734	1109.5019	42.9444	421.8929	-503.2788
66	1276.5789	11.4811	1109.4818	42.8926	421.9328	-503.2788
67	1276.5770	11.4884	1109.4624	42.8428	421.9712	-503.2787
68	1276.5750	11.4954	1109.4438	42.7950	422.0082	-503.2787
69	1276.5732	11.5022	1109.4259	42.7490	422.0437	-503.2787
70	1276.5714	11.5087	1109.4087	42.7047	422.0779	-503.2787
71	1276.5697	11.5150	1109.3921	42.6622	422.1107	-503.2786
72	1276.5681	11.5210	1109.3762	42.6212	422.1424	-503.2786
73	1276.5665	11.5267	1109.3609	42.5819	422.1728	-503.2786
74	1276.5650	11.5323	1109.3461	42.5440	422.2021	-503.2786
75	1276.5635	11.5376	1109.3319	42.5076	422.2303	-503.2786
76	1276.5621	11.5427	1109.3182	42.4726	422.2575	-503.2786
77	1276.5607	11.5476	1109.3051	42.4388	422.2836	-503.2785
78	1276.5594	11.5523	1109.2924	42.4064	422.3088	-503.2785

79	1276.5581	11.5569	1109.2802	42.3751	422.3331	-503.2785
80	1276.5569	11.5612	1109.2685	42.3451	422.3564	-503.2785
81	1276.5558	11.5654	1109.2572	42.3162	422.3789	-503.2785
82	1276.5546	11.5695	1109.2463	42.2883	422.4006	-503.2785
83	1276.5536	11.5733	1109.2358	42.2615	422.4215	-503.2785
84	1276.5525	11.5771	1109.2257	42.2357	422.4416	-503.2785
85	1276.5515	11.5806	1109.2159	42.2108	422.4609	-503.2785
86	1276.5505	11.5841	1109.2066	42.1869	422.4796	-503.2785
87	1276.5496	11.5874	1109.1975	42.1638	422.4976	-503.2785
88	1276.5487	11.5906	1109.1888	42.1416	422.5149	-503.2785
89	1276.5478	11.5936	1109.1804	42.1203	422.5316	-503.2784
90	1276.5470	11.5966	1109.1723	42.0997	422.5477	-503.2784
91	1276.5462	11.5994	1109.1646	42.0798	422.5632	-503.2784
92	1276.5454	11.6021	1109.1571	42.0607	422.5781	-503.2784
93	1276.5447	11.6047	1109.1498	42.0423	422.5925	-503.2784
94	1276.5439	11.6073	1109.1429	42.0246	422.6064	-503.2784
95	1276.5432	11.6097	1109.1361	42.0076	422.6198	-503.2784
96	1276.5426	11.6120	1109.1297	41.9911	422.6327	-503.2784
97	1276.5419	11.6143	1109.1234	41.9753	422.6451	-503.2784
98	1276.5413	11.6164	1109.1174	41.9600	422.6571	-503.2784
99	1276.5407	11.6185	1109.1116	41.9453	422.6686	-503.2784
100	1276.5401	11.6205	1109.1060	41.9311	422.6797	-503.2784
101	1276.5396	11.6224	1109.1007	41.9175	422.6904	-503.2784
102	1276.5390	11.6243	1109.0955	41.9043	422.7008	-503.2784
103	1276.5385	11.6261	1109.0905	41.8916	422.7107	-503.2784
104	1276.5380	11.6278	1109.0856	41.8794	422.7203	-503.2784
105	1276.5375	11.6294	1109.0810	41.8676	422.7296	-503.2784
106	1276.5371	11.6310	1109.0765	41.8563	422.7385	-503.2784
107	1276.5366	11.6326	1109.0722	41.8453	422.7471	-503.2784
108	1276.5362	11.6341	1109.0680	41.8348	422.7554	-503.2784
109	1276.5358	11.6355	1109.0640	41.8246	422.7634	-503.2784
110	1276.5354	11.6368	1109.0601	41.8148	422.7711	-503.2784
111	1276.5350	11.6382	1109.0564	41.8054	422.7786	-503.2784
112	1276.5346	11.6394	1109.0528	41.7963	422.7857	-503.2784
113	1276.5343	11.6407	1109.0493	41.7875	422.7927	-503.2784
114	1276.5339	11.6418	1109.0460	41.7791	422.7993	-503.2784
115	1276.5336	11.6430	1109.0428	41.7709	422.8057	-503.2784
116	1276.5333	11.6441	1109.0397	41.7631	422.8119	-503.2784
117	1276.5329	11.6451	1109.0367	41.7555	422.8179	-503.2784
118	1276.5326	11.6461	1109.0338	41.7482	422.8237	-503.2784
119	1276.5324	11.6471	1109.0310	41.7411	422.8292	-503.2784
120	1276.5321	11.6481	1109.0283	41.7344	422.8346	-503.2784
121	1276.5318	11.6490	1109.0257	41.7278	422.8397	-503.2784

122	1276.5316	11.6499	1109.0232	41.7215	422.8447	-503.2784
123	1276.5313	11.6507	1109.0208	41.7154	422.8495	-503.2784
124	1276.5311	11.6515	1109.0185	41.7096	422.8541	-503.2784
125	1276.5308	11.6523	1109.0162	41.7039	422.8586	-503.2784
126	1276.5306	11.6531	1109.0141	41.6985	422.8629	-503.2784
127	1276.5304	11.6538	1109.0120	41.6932	422.8671	-503.2784
128	1276.5302	11.6545	1109.0100	41.6882	422.8711	-503.2784
129	1276.5300	11.6552	1109.0081	41.6833	422.8749	-503.2784
130	1276.5298	11.6558	1109.0062	41.6786	422.8786	-503.2784
131	1276.5296	11.6564	1109.0044	41.6740	422.8822	-503.2784
132	1276.5294	11.6571	1109.0027	41.6697	422.8857	-503.2784
133	1276.5293	11.6576	1109.0010	41.6654	422.8890	-503.2784
134	1276.5291	11.6582	1108.9994	41.6614	422.8922	-503.2784
135	1276.5289	11.6587	1108.9978	41.6574	422.8953	-503.2784
136	1276.5288	11.6593	1108.9963	41.6537	422.8983	-503.2784
137	1276.5286	11.6598	1108.9949	41.6500	422.9012	-503.2784
138	1276.5285	11.6603	1108.9935	41.6465	422.9040	-503.2784
139	1276.5283	11.6607	1108.9921	41.6431	422.9067	-503.2784
140	1276.5282	11.6612	1108.9908	41.6398	422.9093	-503.2784
141	1276.5281	11.6616	1108.9896	41.6367	422.9118	-503.2784
142	1276.5279	11.6620	1108.9884	41.6336	422.9142	-503.2784
143	1276.5278	11.6624	1108.9872	41.6307	422.9165	-503.2784
144	1276.5277	11.6628	1108.9861	41.6279	422.9187	-503.2784
145	1276.5276	11.6632	1108.9850	41.6252	422.9209	-503.2784
146	1276.5275	11.6636	1108.9840	41.6225	422.9230	-503.2784
147	1276.5274	11.6639	1108.9829	41.6200	422.9250	-503.2784
148	1276.5273	11.6642	1108.9820	41.6175	422.9269	-503.2784
149	1276.5272	11.6646	1108.9810	41.6152	422.9288	-503.2784
150	1276.5271	11.6649	1108.9801	41.6129	422.9306	-503.2784
151	1276.5270	11.6652	1108.9793	41.6107	422.9323	-503.2784
152	1276.5269	11.6655	1108.9784	41.6086	422.9340	-503.2784
153	1276.5268	11.6658	1108.9776	41.6066	422.9356	-503.2784
154	1276.5267	11.6660	1108.9768	41.6046	422.9372	-503.2784
155	1276.5267	11.6663	1108.9761	41.6027	422.9387	-503.2784
156	1276.5266	11.6665	1108.9754	41.6009	422.9401	-503.2784
157	1276.5265	11.6668	1108.9747	41.5991	422.9415	-503.2784
158	1276.5265	11.6670	1108.9740	41.5974	422.9429	-503.2784
159	1276.5264	11.6672	1108.9733	41.5958	422.9442	-503.2784
160	1276.5263	11.6675	1108.9727	41.5942	422.9454	-503.2784
161	1276.5263	11.6677	1108.9721	41.5927	422.9466	-503.2784
162	1276.5262	11.6679	1108.9715	41.5912	422.9478	-503.2784
163	1276.5261	11.6681	1108.9709	41.5898	422.9489	-503.2784
164	1276.5261	11.6683	1108.9704	41.5884	422.9500	-503.2784

165	1276.5260	11.6684	1108.9699	41.5871	422.9510	-503.2784
166	1276.5260	11.6686	1108.9694	41.5858	422.9520	-503.2784
167	1276.5259	11.6688	1108.9689	41.5846	422.9530	-503.2784
168	1276.5259	11.6690	1108.9684	41.5834	422.9540	-503.2784
169	1276.5258	11.6691	1108.9680	41.5823	422.9549	-503.2784
170	1276.5258	11.6693	1108.9675	41.5812	422.9557	-503.2784
171	1276.5257	11.6694	1108.9671	41.5801	422.9566	-503.2784
172	1276.5257	11.6695	1108.9667	41.5791	422.9574	-503.2784
173	1276.5257	11.6697	1108.9663	41.5781	422.9582	-503.2784
174	1276.5256	11.6698	1108.9659	41.5771	422.9589	-503.2784
175	1276.5256	11.6699	1108.9655	41.5762	422.9596	-503.2784
176	1276.5255	11.6701	1108.9652	41.5753	422.9604	-503.2784
177	1276.5255	11.6702	1108.9649	41.5745	422.9610	-503.2784
178	1276.5255	11.6703	1108.9645	41.5736	422.9617	-503.2784
179	1276.5254	11.6704	1108.9642	41.5729	422.9623	-503.2784
180	1276.5254	11.6705	1108.9639	41.5721	422.9629	-503.2784
181	1276.5254	11.6706	1108.9636	41.5713	422.9635	-503.2784
182	1276.5253	11.6707	1108.9633	41.5706	422.9641	-503.2784
183	1276.5253	11.6708	1108.9631	41.5699	422.9646	-503.2784
184	1276.5253	11.6709	1108.9628	41.5693	422.9651	-503.2784
185	1276.5253	11.6710	1108.9625	41.5686	422.9656	-503.2784
186	1276.5252	11.6711	1108.9623	41.5680	422.9661	-503.2784
187	1276.5252	11.6711	1108.9621	41.5674	422.9666	-503.2784
188	1276.5252	11.6712	1108.9618	41.5669	422.9671	-503.2784
189	1276.5252	11.6713	1108.9616	41.5663	422.9675	-503.2784
190	1276.5251	11.6714	1108.9614	41.5658	422.9679	-503.2784
191	1276.5251	11.6714	1108.9612	41.5653	422.9683	-503.2784
192	1276.5251	11.6715	1108.9610	41.5648	422.9687	-503.2784
193	1276.5251	11.6716	1108.9608	41.5643	422.9691	-503.2784
194	1276.5251	11.6716	1108.9606	41.5638	422.9695	-503.2784
195	1276.5250	11.6717	1108.9604	41.5634	422.9698	-503.2784
196	1276.5250	11.6718	1108.9603	41.5629	422.9702	-503.2784
197	1276.5250	11.6718	1108.9601	41.5625	422.9705	-503.2784
198	1276.5250	11.6719	1108.9599	41.5621	422.9708	-503.2784
199	1276.5250	11.6719	1108.9598	41.5617	422.9711	-503.2784
200	1276.5250	11.6720	1108.9596	41.5614	422.9714	-503.2784
201	1276.5250	11.6720	1108.9595	41.5610	422.9717	-503.2784
202	1276.5249	11.6721	1108.9594	41.5607	422.9720	-503.2784
203	1276.5249	11.6721	1108.9592	41.5603	422.9722	-503.2784
204	1276.5249	11.6722	1108.9591	41.5600	422.9725	-503.2784
205	1276.5249	11.6722	1108.9590	41.5597	422.9727	-503.2784
206	1276.5249	11.6722	1108.9589	41.5594	422.9730	-503.2784
207	1276.5249	11.6723	1108.9588	41.5591	422.9732	-503.2784

208	1276.5249	11.6723	1108.9586	41.5588	422.9734	-503.2784
209	1276.5249	11.6724	1108.9585	41.5586	422.9736	-503.2784
210	1276.5248	11.6724	1108.9584	41.5583	422.9738	-503.2784
211	1276.5248	11.6724	1108.9583	41.5581	422.9740	-503.2784
212	1276.5248	11.6725	1108.9582	41.5578	422.9742	-503.2784
213	1276.5248	11.6725	1108.9581	41.5576	422.9744	-503.2784
214	1276.5248	11.6725	1108.9581	41.5574	422.9746	-503.2784
215	1276.5248	11.6726	1108.9580	41.5571	422.9748	-503.2784
216	1276.5248	11.6726	1108.9579	41.5569	422.9749	-503.2784
217	1276.5248	11.6726	1108.9578	41.5567	422.9751	-503.2784
218	1276.5248	11.6726	1108.9577	41.5565	422.9752	-503.2784
219	1276.5248	11.6727	1108.9577	41.5564	422.9754	-503.2784
220	1276.5248	11.6727	1108.9576	41.5562	422.9755	-503.2784
221	1276.5247	11.6727	1108.9575	41.5560	422.9757	-503.2784
222	1276.5247	11.6727	1108.9574	41.5558	422.9758	-503.2784
223	1276.5247	11.6728	1108.9574	41.5557	422.9759	-503.2784
224	1276.5247	11.6728	1108.9573	41.5555	422.9761	-503.2784
225	1276.5247	11.6728	1108.9573	41.5554	422.9762	-503.2784
226	1276.5247	11.6728	1108.9572	41.5552	422.9763	-503.2784
227	1276.5247	11.6728	1108.9571	41.5551	422.9764	-503.2784
228	1276.5247	11.6729	1108.9571	41.5549	422.9765	-503.2784
229	1276.5247	11.6729	1108.9570	41.5548	422.9766	-503.2784
230	1276.5247	11.6729	1108.9570	41.5547	422.9767	-503.2784
231	1276.5247	11.6729	1108.9569	41.5546	422.9768	-503.2784
232	1276.5247	11.6729	1108.9569	41.5545	422.9769	-503.2784
233	1276.5247	11.6729	1108.9569	41.5543	422.9770	-503.2784
234	1276.5247	11.6730	1108.9568	41.5542	422.9771	-503.2784
235	1276.5247	11.6730	1108.9568	41.5541	422.9772	-503.2784
236	1276.5247	11.6730	1108.9567	41.5540	422.9772	-503.2784
237	1276.5247	11.6730	1108.9567	41.5539	422.9773	-503.2784
238	1276.5247	11.6730	1108.9566	41.5538	422.9774	-503.2784
239	1276.5247	11.6730	1108.9566	41.5537	422.9775	-503.2784
240	1276.5246	11.6730	1108.9566	41.5537	422.9775	-503.2784
241	1276.5246	11.6730	1108.9565	41.5536	422.9776	-503.2784
242	1276.5246	11.6731	1108.9565	41.5535	422.9777	-503.2784
243	1276.5246	11.6731	1108.9565	41.5534	422.9777	-503.2784
244	1276.5246	11.6731	1108.9565	41.5533	422.9778	-503.2784
245	1276.5246	11.6731	1108.9564	41.5533	422.9778	-503.2784
246	1276.5246	11.6731	1108.9564	41.5532	422.9779	-503.2784
247	1276.5246	11.6731	1108.9564	41.5531	422.9780	-503.2784
248	1276.5246	11.6731	1108.9563	41.5531	422.9780	-503.2784
249	1276.5246	11.6731	1108.9563	41.5530	422.9781	-503.2784
250	1276.5246	11.6731	1108.9563	41.5529	422.9781	-503.2784

251	1276.5246	11.6731	1108.9563	41.5529	422.9782	-503.2784
252	1276.5246	11.6732	1108.9562	41.5528	422.9782	-503.2784
253	1276.5246	11.6732	1108.9562	41.5528	422.9782	-503.2784

<sup>†</sup> and <sup>‡</sup> represents initial values for variance components and first iteration of the EM algorithm, respectively.

Table S5. Convergence history for the variance components and  $\log(L)$  of the Model 1 + RGCA via AI algorithm for dataset 1.

Iterations	$\hat{\sigma}_g^2$	$\hat{\sigma}_d^2$	$\hat{\sigma}_s^2$	$\hat{\sigma}_r^2$	$\hat{\sigma}^2$	$\log(L)$
0 <sup>†</sup>	167.5077	167.5077	167.5077	167.5077	1675.0780	-543.5475
1 <sup>‡</sup>	353.0179	-1300.5068	1569.4556	-682.8790	-1912.4120	NA
2 <sup>§</sup>	793.9722	-70920.9614	7346.8699	-7844.3294	-14513.7860	NA

<sup>†</sup>, <sup>‡</sup> and <sup>§</sup> represents initial values for variance components, first iteration of the AI algorithm and fail in converge due singularity in AI matrix, respectively.

Table S6. Convergence history for the variance components and  $\log(L)$  of the Model 1 + RGCA via EM-AI algorithm for dataset 1.

Iterations	$\hat{\sigma}_g^2$	$\hat{\sigma}_d^2$	$\hat{\sigma}_s^2$	$\hat{\sigma}_r^2$	$\hat{\sigma}^2$	$\log(L)$
0 <sup>†</sup>	167.5077	167.5077	167.5077	167.5077	1675.0775	-543.5475
1 <sup>‡</sup>	617.4592	78.2859	189.3851	144.1255	1416.6692	-529.5908
2	1049.0936	48.3021	216.9843	124.1736	1027.0948	-522.1382
3	1204.6085	33.6947	282.9308	107.0105	863.1787	-517.6222
4	1251.8830	25.7295	392.4954	93.5279	761.4507	-513.3392
5	1271.9793	20.9175	541.5116	83.1134	666.4277	-509.3666
6	1282.9049	17.8118	708.5549	75.1330	578.1695	-506.3316
7	1287.2761	15.7549	858.7158	69.1775	508.5184	-504.5440
8	1286.9034	14.3893	968.3515	64.8958	462.3407	-503.7425
9	1284.4701	13.4784	1036.7167	61.8784	435.6507	-503.4574
10	1281.9215	12.8570	1075.1867	59.7155	421.6886	-503.3695
11	1280.0059	12.4170	1095.4994	58.0909	414.9885	-503.3418
12	1278.7681	12.0928	1105.7624	56.8009	412.1266	-503.3302
13	1278.0273	11.8457	1110.7331	55.7259	411.1965	-503.3229
14	1277.5983	11.6526	1112.9967	54.7978	411.1977	-503.3171
15	1277.3498	11.4995	1113.9046	53.9773	411.6222	-503.3122
16	1277.2016	11.3770	1114.1485	53.2403	412.2177	-503.3080
17	1277.1081	11.2788	1114.0780	52.5715	412.8621	-503.3044
18	1277.0448	11.2004	1113.8678	51.9601	413.4983	-503.3014
19	1276.9983	11.1381	1113.6040	51.3982	414.1015	-503.2988
20	1276.9617	11.0893	1113.3279	50.8796	414.6625	-503.2966
21	1276.9312	11.0518	1113.0588	50.3993	415.1797	-503.2947
22	1276.9048	11.0238	1112.8049	49.9533	415.6546	-503.2930

23	1276.8815	11.0039	1112.5690	49.5379	416.0904	-503.2915
24	1276.8605	10.9907	1112.3513	49.1502	416.4905	-503.2903
25	1276.8413	10.9832	1112.1510	48.7875	416.8583	-503.2891
26 <sup>§</sup>	1276.6798	11.7010	1109.6018	41.3758	422.7973	-503.2784
27	1276.5335	11.6734	1108.9904	41.5523	422.9738	-503.2784
28	1276.5249	11.6734	1108.9572	41.5513	422.9792	-503.2784
29	1276.5246	11.6734	1108.9557	41.5513	422.9794	-503.2784
30	1276.5246	11.6734	1108.9556	41.5513	422.9794	-503.2784
31	1276.5246	11.6734	1108.9556	41.5513	422.9794	-503.2784

<sup>†</sup>, <sup>‡</sup> and <sup>§</sup> represents initial values for variance components, first iteration of the EM algorithm and first iteration of the AI algorithm, respectively.

Table S7. Convergence history for the variance components and  $\log(L)$  of the Model 3 via EM algorithm for dataset 1.

Iterations	$\hat{\sigma}_g^2$	$\hat{\sigma}_d^2$	$\hat{\sigma}_s^2$	$\hat{\sigma}_r^2$	$\hat{\sigma}_p^2$	$\hat{\sigma}^2$	$\log(L)$
0 <sup>†</sup>	167.5077	167.5077	167.5077	167.5077	167.5077	1675.0775	-554.6942
1 <sup>‡</sup>	425.8857	78.2859	171.6276	144.1255	391.2504	1814.1861	-544.2267
2	694.4077	50.1276	166.5828	126.1971	1099.3908	1513.2609	-529.7613
3	845.1396	35.8660	162.8132	110.6486	3664.2506	1088.6475	-512.0291
4	910.3488	27.2528	165.3263	96.8936	7645.2864	703.7212	-501.0536
5	940.0120	21.8545	181.6424	85.7622	9415.2658	538.3067	-497.3186
6	950.0378	18.5470	209.1329	77.8922	9850.3221	478.1681	-495.9763
7	950.5674	16.4302	240.0467	72.3028	9951.2836	451.0424	-495.2862
8	947.5606	14.9999	269.1794	68.1707	9979.4410	435.4459	-494.8949
9	943.6577	13.9940	293.9740	65.0214	9990.5834	425.2886	-494.6770
10	939.9332	13.2662	313.6805	62.5649	9996.6505	418.5038	-494.5593
11	936.7809	12.7279	328.6016	60.6072	10000.4764	414.0603	-494.4966
12	934.2862	12.3229	339.5114	59.0136	10002.9568	411.2630	-494.4628
13	932.3969	12.0136	347.2866	57.6892	10004.5290	409.6026	-494.4438
14	931.0095	11.7748	352.7213	56.5671	10005.4781	408.7093	-494.4324
15	930.0130	11.5888	356.4610	55.5999	10006.0050	408.3225	-494.4250
16	929.3088	11.4433	358.9983	54.7537	10006.2517	408.2623	-494.4197
17	928.8171	11.3291	360.6955	54.0039	10006.3164	408.4069	-494.4156
18	928.4770	11.2397	361.8118	53.3325	10006.2659	408.6749	-494.4124
19	928.2436	11.1702	362.5301	52.7260	10006.1453	409.0121	-494.4098
20	928.0845	11.1167	362.9779	52.1741	10005.9844	409.3833	-494.4076
21	927.9769	11.0763	363.2433	51.6689	10005.8027	409.7659	-494.4057
22	927.9046	11.0466	363.3869	51.2042	10005.6130	410.1455	-494.4040
23	927.8566	11.0257	363.4500	50.7747	10005.4232	410.5136	-494.4026
24	927.8251	11.0122	363.4607	50.3765	10005.2382	410.8651	-494.4014
25	927.8048	11.0048	363.4382	50.0061	10005.0610	411.1976	-494.4003

26	927.7921	11.0025	363.3953	49.6605	10004.8930	411.5101	-494.3993
27	927.7845	11.0043	363.3406	49.3375	10004.7349	411.8026	-494.3984
28	927.7803	11.0096	363.2796	49.0348	10004.5867	412.0758	-494.3977
29	927.7783	11.0177	363.2163	48.7507	10004.4483	412.3305	-494.3970
30	927.7777	11.0282	363.1528	48.4835	10004.3192	412.5680	-494.3964
31	927.7781	11.0405	363.0908	48.2319	10004.1988	412.7894	-494.3958
32	927.7790	11.0543	363.0310	47.9946	10004.0865	412.9959	-494.3954
33	927.7803	11.0693	362.9740	47.7706	10003.9818	413.1888	-494.3949
34	927.7817	11.0851	362.9200	47.5588	10003.8841	413.3690	-494.3945
35	927.7832	11.1016	362.8690	47.3583	10003.7927	413.5377	-494.3942
36	927.7846	11.1186	362.8210	47.1685	10003.7073	413.6957	-494.3938
37	927.7861	11.1358	362.7759	46.9884	10003.6272	413.8440	-494.3935
38	927.7874	11.1531	362.7334	46.8176	10003.5521	413.9833	-494.3933
39	927.7887	11.1705	362.6936	46.6553	10003.4816	414.1143	-494.3930
40	927.7899	11.1878	362.6561	46.5011	10003.4153	414.2378	-494.3928
41	927.7911	11.2049	362.6209	46.3545	10003.3528	414.3543	-494.3926
42	927.7921	11.2217	362.5877	46.2149	10003.2939	414.4643	-494.3924
43	927.7931	11.2383	362.5564	46.0821	10003.2383	414.5683	-494.3922
44	927.7940	11.2545	362.5269	45.9555	10003.1857	414.6667	-494.3921
45	927.7948	11.2703	362.4991	45.8348	10003.1359	414.7600	-494.3919
46	927.7956	11.2857	362.4728	45.7197	10003.0888	414.8486	-494.3918
47	927.7963	11.3007	362.4479	45.6099	10003.0441	414.9326	-494.3917
48	927.7970	11.3152	362.4242	45.5051	10003.0016	415.0125	-494.3916
49	927.7976	11.3293	362.4019	45.4051	10002.9613	415.0885	-494.3915
50	927.7982	11.3429	362.3806	45.3095	10002.9229	415.1609	-494.3914
51	927.7987	11.3561	362.3604	45.2181	10002.8864	415.2298	-494.3913
52	927.7992	11.3688	362.3413	45.1309	10002.8516	415.2955	-494.3912
53	927.7997	11.3810	362.3230	45.0474	10002.8184	415.3582	-494.3912
54	927.8001	11.3928	362.3056	44.9676	10002.7867	415.4181	-494.3911
55	927.8006	11.4041	362.2890	44.8912	10002.7565	415.4753	-494.3910
56	927.8009	11.4150	362.2732	44.8181	10002.7277	415.5299	-494.3910
57	927.8013	11.4255	362.2581	44.7482	10002.7001	415.5821	-494.3909
58	927.8017	11.4356	362.2437	44.6813	10002.6738	415.6321	-494.3909
59	927.8020	11.4453	362.2300	44.6171	10002.6486	415.6799	-494.3908
60	927.8023	11.4545	362.2168	44.5558	10002.6245	415.7257	-494.3908
61	927.8026	11.4635	362.2042	44.4969	10002.6014	415.7695	-494.3908
62	927.8029	11.4720	362.1922	44.4406	10002.5793	415.8115	-494.3907
63	927.8031	11.4802	362.1807	44.3866	10002.5581	415.8517	-494.3907
64	927.8034	11.4881	362.1696	44.3349	10002.5378	415.8902	-494.3907
65	927.8036	11.4956	362.1591	44.2853	10002.5184	415.9272	-494.3906
66	927.8038	11.5028	362.1490	44.2377	10002.4998	415.9626	-494.3906
67	927.8040	11.5098	362.1393	44.1921	10002.4819	415.9966	-494.3906
68	927.8042	11.5164	362.1300	44.1484	10002.4648	416.0292	-494.3906

69	927.8044	11.5228	362.1211	44.1065	10002.4484	416.0605	-494.3906
70	927.8046	11.5289	362.1125	44.0663	10002.4326	416.0905	-494.3905
71	927.8048	11.5347	362.1043	44.0277	10002.4175	416.1193	-494.3905
72	927.8050	11.5403	362.0965	43.9907	10002.4030	416.1469	-494.3905
73	927.8051	11.5457	362.0889	43.9552	10002.3890	416.1735	-494.3905
74	927.8053	11.5509	362.0817	43.9211	10002.3756	416.1989	-494.3905
75	927.8054	11.5558	362.0748	43.8884	10002.3628	416.2234	-494.3905
76	927.8056	11.5605	362.0681	43.8570	10002.3505	416.2469	-494.3905
77	927.8057	11.5651	362.0617	43.8269	10002.3386	416.2694	-494.3905
78	927.8058	11.5694	362.0555	43.7980	10002.3273	416.2911	-494.3904
79	927.8060	11.5736	362.0496	43.7703	10002.3164	416.3119	-494.3904
80	927.8061	11.5776	362.0440	43.7436	10002.3059	416.3319	-494.3904
81	927.8062	11.5814	362.0385	43.7181	10002.2958	416.3511	-494.3904
82	927.8063	11.5851	362.0333	43.6935	10002.2861	416.3695	-494.3904
83	927.8064	11.5886	362.0283	43.6699	10002.2768	416.3873	-494.3904
84	927.8065	11.5920	362.0235	43.6473	10002.2679	416.4043	-494.3904
85	927.8066	11.5952	362.0188	43.6255	10002.2593	416.4207	-494.3904
86	927.8067	11.5983	362.0144	43.6046	10002.2511	416.4364	-494.3904
87	927.8068	11.6013	362.0101	43.5846	10002.2431	416.4515	-494.3904
88	927.8069	11.6041	362.0060	43.5653	10002.2355	416.4660	-494.3904
89	927.8069	11.6069	362.0020	43.5468	10002.2282	416.4800	-494.3904
90	927.8070	11.6095	361.9983	43.5290	10002.2212	416.4934	-494.3904
91	927.8071	11.6120	361.9946	43.5120	10002.2144	416.5063	-494.3904
92	927.8072	11.6144	361.9911	43.4955	10002.2079	416.5186	-494.3904
93	927.8072	11.6168	361.9877	43.4798	10002.2017	416.5305	-494.3904
94	927.8073	11.6190	361.9845	43.4646	10002.1957	416.5420	-494.3904
95	927.8074	11.6211	361.9814	43.4501	10002.1899	416.5530	-494.3904
96	927.8074	11.6232	361.9784	43.4361	10002.1844	416.5636	-494.3904
97	927.8075	11.6252	361.9755	43.4227	10002.1791	416.5737	-494.3904
98	927.8075	11.6270	361.9728	43.4098	10002.1739	416.5835	-494.3904
99	927.8076	11.6289	361.9701	43.3974	10002.1690	416.5929	-494.3904
100	927.8076	11.6306	361.9676	43.3855	10002.1643	416.6019	-494.3904
101	927.8077	11.6323	361.9651	43.3740	10002.1597	416.6106	-494.3904
102	927.8077	11.6339	361.9628	43.3630	10002.1554	416.6189	-494.3904
103	927.8078	11.6354	361.9605	43.3524	10002.1512	416.6269	-494.3904
104	927.8078	11.6369	361.9583	43.3423	10002.1471	416.6346	-494.3904
105	927.8079	11.6383	361.9562	43.3325	10002.1432	416.6420	-494.3904
106	927.8079	11.6397	361.9542	43.3231	10002.1395	416.6492	-494.3904
107	927.8080	11.6410	361.9523	43.3141	10002.1359	416.6560	-494.3904
108	927.8080	11.6423	361.9504	43.3054	10002.1325	416.6626	-494.3904
109	927.8080	11.6435	361.9486	43.2971	10002.1292	416.6689	-494.3903
110	927.8081	11.6446	361.9469	43.2891	10002.1260	416.6750	-494.3903
111	927.8081	11.6458	361.9452	43.2814	10002.1229	416.6809	-494.3903

112	927.8081	11.6468	361.9437	43.2740	10002.1199	416.6865	-494.3903
113	927.8082	11.6479	361.9421	43.2669	10002.1171	416.6919	-494.3903
114	927.8082	11.6488	361.9407	43.2600	10002.1144	416.6971	-494.3903
115	927.8082	11.6498	361.9392	43.2535	10002.1118	416.7021	-494.3903
116	927.8082	11.6507	361.9379	43.2471	10002.1092	416.7069	-494.3903
117	927.8083	11.6516	361.9366	43.2411	10002.1068	416.7115	-494.3903
118	927.8083	11.6524	361.9353	43.2352	10002.1045	416.7160	-494.3903
119	927.8083	11.6532	361.9341	43.2296	10002.1023	416.7203	-494.3903
120	927.8083	11.6540	361.9330	43.2242	10002.1001	416.7244	-494.3903
121	927.8084	11.6548	361.9318	43.2190	10002.0980	416.7283	-494.3903
122	927.8084	11.6555	361.9308	43.2140	10002.0960	416.7321	-494.3903
123	927.8084	11.6562	361.9297	43.2092	10002.0941	416.7358	-494.3903
124	927.8084	11.6568	361.9287	43.2046	10002.0923	416.7393	-494.3903
125	927.8084	11.6575	361.9278	43.2002	10002.0905	416.7427	-494.3903
126	927.8085	11.6581	361.9269	43.1959	10002.0888	416.7459	-494.3903
127	927.8085	11.6587	361.9260	43.1918	10002.0872	416.7490	-494.3903
128	927.8085	11.6593	361.9251	43.1879	10002.0856	416.7520	-494.3903
129	927.8085	11.6598	361.9243	43.1841	10002.0841	416.7549	-494.3903
130	927.8085	11.6603	361.9235	43.1805	10002.0826	416.7577	-494.3903
131	927.8085	11.6608	361.9228	43.1770	10002.0812	416.7604	-494.3903
132	927.8086	11.6613	361.9221	43.1736	10002.0799	416.7629	-494.3903
133	927.8086	11.6618	361.9214	43.1703	10002.0786	416.7654	-494.3903
134	927.8086	11.6622	361.9207	43.1672	10002.0774	416.7678	-494.3903
135	927.8086	11.6626	361.9201	43.1642	10002.0762	416.7701	-494.3903
136	927.8086	11.6631	361.9194	43.1614	10002.0750	416.7722	-494.3903
137	927.8086	11.6635	361.9188	43.1586	10002.0739	416.7744	-494.3903
138	927.8086	11.6638	361.9183	43.1559	10002.0728	416.7764	-494.3903
139	927.8087	11.6642	361.9177	43.1534	10002.0718	416.7783	-494.3903
140	927.8087	11.6646	361.9172	43.1509	10002.0708	416.7802	-494.3903
141	927.8087	11.6649	361.9167	43.1486	10002.0699	416.7820	-494.3903
142	927.8087	11.6652	361.9162	43.1463	10002.0690	416.7838	-494.3903
143	927.8087	11.6655	361.9157	43.1441	10002.0681	416.7854	-494.3903
144	927.8087	11.6658	361.9153	43.1420	10002.0673	416.7870	-494.3903
145	927.8087	11.6661	361.9148	43.1400	10002.0665	416.7886	-494.3903
146	927.8087	11.6664	361.9144	43.1380	10002.0657	416.7900	-494.3903
147	927.8087	11.6667	361.9140	43.1362	10002.0649	416.7915	-494.3903
148	927.8087	11.6669	361.9136	43.1344	10002.0642	416.7928	-494.3903
149	927.8087	11.6672	361.9132	43.1326	10002.0635	416.7942	-494.3903
150	927.8087	11.6674	361.9129	43.1310	10002.0629	416.7954	-494.3903
151	927.8088	11.6676	361.9125	43.1294	10002.0622	416.7967	-494.3903
152	927.8088	11.6679	361.9122	43.1278	10002.0616	416.7978	-494.3903
153	927.8088	11.6681	361.9119	43.1264	10002.0610	416.7990	-494.3903
154	927.8088	11.6683	361.9116	43.1249	10002.0605	416.8000	-494.3903

155	927.8088	11.6685	361.9113	43.1236	10002.0599	416.8011	-494.3903
156	927.8088	11.6687	361.9110	43.1223	10002.0594	416.8021	-494.3903
157	927.8088	11.6688	361.9107	43.1210	10002.0589	416.8031	-494.3903
158	927.8088	11.6690	361.9105	43.1198	10002.0584	416.8040	-494.3903
159	927.8088	11.6692	361.9102	43.1186	10002.0579	416.8049	-494.3903
160	927.8088	11.6693	361.9100	43.1175	10002.0575	416.8057	-494.3903
161	927.8088	11.6695	361.9097	43.1164	10002.0570	416.8066	-494.3903
162	927.8088	11.6696	361.9095	43.1154	10002.0566	416.8074	-494.3903
163	927.8088	11.6698	361.9093	43.1144	10002.0562	416.8081	-494.3903
164	927.8088	11.6699	361.9091	43.1134	10002.0558	416.8088	-494.3903
165	927.8088	11.6701	361.9089	43.1125	10002.0555	416.8096	-494.3903
166	927.8088	11.6702	361.9087	43.1116	10002.0551	416.8102	-494.3903
167	927.8088	11.6703	361.9085	43.1107	10002.0548	416.8109	-494.3903
168	927.8088	11.6704	361.9083	43.1099	10002.0544	416.8115	-494.3903
169	927.8088	11.6705	361.9082	43.1091	10002.0541	416.8121	-494.3903
170	927.8088	11.6706	361.9080	43.1084	10002.0538	416.8127	-494.3903
171	927.8088	11.6707	361.9079	43.1076	10002.0535	416.8133	-494.3903
172	927.8089	11.6708	361.9077	43.1069	10002.0532	416.8138	-494.3903
173	927.8089	11.6709	361.9076	43.1063	10002.0530	416.8143	-494.3903
174	927.8089	11.6710	361.9074	43.1056	10002.0527	416.8148	-494.3903
175	927.8089	11.6711	361.9073	43.1050	10002.0525	416.8153	-494.3903
176	927.8089	11.6712	361.9072	43.1044	10002.0522	416.8157	-494.3903
177	927.8089	11.6713	361.9070	43.1038	10002.0520	416.8162	-494.3903
178	927.8089	11.6714	361.9069	43.1033	10002.0518	416.8166	-494.3903
179	927.8089	11.6715	361.9068	43.1027	10002.0516	416.8170	-494.3903
180	927.8089	11.6715	361.9067	43.1022	10002.0514	416.8174	-494.3903
181	927.8089	11.6716	361.9066	43.1017	10002.0512	416.8178	-494.3903
182	927.8089	11.6717	361.9065	43.1012	10002.0510	416.8181	-494.3903
183	927.8089	11.6717	361.9064	43.1008	10002.0508	416.8185	-494.3903
184	927.8089	11.6718	361.9063	43.1003	10002.0506	416.8188	-494.3903
185	927.8089	11.6719	361.9062	43.0999	10002.0504	416.8192	-494.3903
186	927.8089	11.6719	361.9061	43.0995	10002.0503	416.8195	-494.3903
187	927.8089	11.6720	361.9060	43.0991	10002.0501	416.8198	-494.3903
188	927.8089	11.6720	361.9059	43.0987	10002.0500	416.8201	-494.3903
189	927.8089	11.6721	361.9059	43.0984	10002.0498	416.8203	-494.3903
190	927.8089	11.6721	361.9058	43.0980	10002.0497	416.8206	-494.3903
191	927.8089	11.6722	361.9057	43.0977	10002.0496	416.8208	-494.3903
192	927.8089	11.6722	361.9056	43.0974	10002.0494	416.8211	-494.3903
193	927.8089	11.6723	361.9056	43.0971	10002.0493	416.8213	-494.3903
194	927.8089	11.6723	361.9055	43.0968	10002.0492	416.8216	-494.3903
195	927.8089	11.6723	361.9054	43.0965	10002.0491	416.8218	-494.3903
196	927.8089	11.6724	361.9054	43.0962	10002.0490	416.8220	-494.3903
197	927.8089	11.6724	361.9053	43.0960	10002.0488	416.8222	-494.3903

198	927.8089	11.6725	361.9053	43.0957	10002.0487	416.8224	-494.3903
199	927.8089	11.6725	361.9052	43.0955	10002.0486	416.8226	-494.3903
200	927.8089	11.6725	361.9052	43.0952	10002.0486	416.8227	-494.3903
201	927.8089	11.6726	361.9051	43.0950	10002.0485	416.8229	-494.3903
202	927.8089	11.6726	361.9051	43.0948	10002.0484	416.8231	-494.3903
203	927.8089	11.6726	361.9050	43.0946	10002.0483	416.8232	-494.3903
204	927.8089	11.6726	361.9050	43.0944	10002.0482	416.8234	-494.3903
205	927.8089	11.6727	361.9049	43.0942	10002.0481	416.8235	-494.3903
206	927.8089	11.6727	361.9049	43.0940	10002.0481	416.8237	-494.3903
207	927.8089	11.6727	361.9049	43.0938	10002.0480	416.8238	-494.3903
208	927.8089	11.6727	361.9048	43.0936	10002.0479	416.8240	-494.3903
209	927.8089	11.6728	361.9048	43.0935	10002.0479	416.8241	-494.3903
210	927.8089	11.6728	361.9048	43.0933	10002.0478	416.8242	-494.3903
211	927.8089	11.6728	361.9047	43.0932	10002.0477	416.8243	-494.3903
212	927.8089	11.6728	361.9047	43.0930	10002.0477	416.8244	-494.3903
213	927.8089	11.6729	361.9047	43.0929	10002.0476	416.8245	-494.3903
214	927.8089	11.6729	361.9046	43.0927	10002.0476	416.8246	-494.3903
215	927.8089	11.6729	361.9046	43.0926	10002.0475	416.8247	-494.3903
216	927.8089	11.6729	361.9046	43.0925	10002.0475	416.8248	-494.3903
217	927.8089	11.6729	361.9046	43.0924	10002.0474	416.8249	-494.3903
218	927.8089	11.6729	361.9045	43.0922	10002.0474	416.8250	-494.3903
219	927.8089	11.6730	361.9045	43.0921	10002.0473	416.8251	-494.3903
220	927.8089	11.6730	361.9045	43.0920	10002.0473	416.8252	-494.3903
221	927.8089	11.6730	361.9045	43.0919	10002.0472	416.8253	-494.3903
222	927.8089	11.6730	361.9044	43.0918	10002.0472	416.8253	-494.3903
223	927.8089	11.6730	361.9044	43.0917	10002.0472	416.8254	-494.3903
224	927.8089	11.6730	361.9044	43.0916	10002.0471	416.8255	-494.3903
225	927.8089	11.6730	361.9044	43.0915	10002.0471	416.8256	-494.3903
226	927.8089	11.6731	361.9044	43.0915	10002.0470	416.8256	-494.3903
227	927.8089	11.6731	361.9043	43.0914	10002.0470	416.8257	-494.3903
228	927.8089	11.6731	361.9043	43.0913	10002.0470	416.8257	-494.3903
229	927.8089	11.6731	361.9043	43.0912	10002.0470	416.8258	-494.3903
230	927.8089	11.6731	361.9043	43.0911	10002.0469	416.8259	-494.3903
231	927.8089	11.6731	361.9043	43.0911	10002.0469	416.8259	-494.3903

<sup>†</sup> and <sup>‡</sup> represents initial values for variance components and first iteration of the EM algorithm, respectively.

Table S8. Convergence history for the variance components and  $\log(L)$  of the Model 3 via AI algorithm for dataset 1.

Iterations	$\hat{\sigma}_g^2$	$\hat{\sigma}_d^2$	$\hat{\sigma}_s^2$	$\hat{\sigma}_r^2$	$\hat{\sigma}_p^2$	$\hat{\sigma}^2$	$\log(L)$
0 <sup>†</sup>	167.5077	167.5077	167.5077	167.5077	167.5077	1675.0780	-554.6942
1 <sup>‡</sup>	387.9617	-1300.5068	1280.4744	-315.6386	3601.8500	-3381.3740	NA

2 <sup>§</sup>	632.7387	-70920.9614	9012.3077	-2924.5200	20561.1126	-34193.0240	NA
----------------	----------	-------------	-----------	------------	------------	-------------	----

<sup>†</sup>, <sup>‡</sup> and <sup>§</sup> represents initial values for variance components, first iteration of the AI algorithm and fail in converge due singularity in AI matrix, respectively.

Table S9. Convergence history for the variance components and  $\log(L)$  of the Model 3 via EM-AI algorithm for dataset 1.

Iterations	$\hat{\sigma}_g^2$	$\hat{\sigma}_d^2$	$\hat{\sigma}_s^2$	$\hat{\sigma}_r^2$	$\hat{\sigma}_p^2$	$\hat{\sigma}^2$	$\log(L)$
0 <sup>†</sup>	167.5077	167.5077	167.5077	167.5077	167.5077	1675.0775	-554.6942
1 <sup>‡</sup>	425.8857	78.2859	171.6276	144.1255	391.2504	1814.1861	-544.2267
2	694.4077	50.1276	166.5828	126.1971	1099.3908	1513.2609	-529.7613
3	845.1396	35.8660	162.8132	110.6486	3664.2506	1088.6475	-512.0291
4	910.3488	27.2528	165.3263	96.8936	7645.2864	703.7212	-501.0536
5	940.0120	21.8545	181.6424	85.7622	9415.2658	538.3067	-497.3186
6	950.0378	18.5470	209.1329	77.8922	9850.3221	478.1681	-495.9763
7	950.5674	16.4302	240.0467	72.3028	9951.2836	451.0424	-495.2862
8	947.5606	14.9999	269.1794	68.1707	9979.4410	435.4459	-494.8949
9	943.6577	13.9940	293.9740	65.0214	9990.5834	425.2886	-494.6770
10	939.9332	13.2662	313.6805	62.5649	9996.6505	418.5038	-494.5593
11	936.7809	12.7279	328.6016	60.6072	10000.4764	414.0603	-494.4966
12	934.2862	12.3229	339.5114	59.0136	10002.9568	411.2630	-494.4628
13	932.3969	12.0136	347.2866	57.6892	10004.5290	409.6026	-494.4438
14	931.0095	11.7748	352.7213	56.5671	10005.4781	408.7093	-494.4324
15	930.0130	11.5888	356.4610	55.5999	10006.0050	408.3225	-494.4250
16	929.3088	11.4433	358.9983	54.7537	10006.2517	408.2623	-494.4197
17	928.8171	11.3291	360.6955	54.0039	10006.3164	408.4069	-494.4156
18	928.4770	11.2397	361.8118	53.3325	10006.2659	408.6749	-494.4124
19	928.2436	11.1702	362.5301	52.7260	10006.1453	409.0121	-494.4098
20	928.0845	11.1167	362.9779	52.1741	10005.9844	409.3833	-494.4076
21	927.9769	11.0763	363.2433	51.6689	10005.8027	409.7659	-494.4057
22	927.9046	11.0466	363.3869	51.2042	10005.6130	410.1455	-494.4040
23	927.8566	11.0257	363.4500	50.7747	10005.4232	410.5136	-494.4026
24	927.8251	11.0122	363.4607	50.3765	10005.2382	410.8651	-494.4014
25	927.8048	11.0048	363.4382	50.0061	10005.0610	411.1976	-494.4003
26 <sup>§</sup>	927.8089	11.6991	361.9229	42.9023	10002.0842	416.7512	-494.3904
27	927.8089	11.6734	361.9039	43.0890	10002.0462	416.8272	-494.3903
28	927.8089	11.6734	361.9039	43.0893	10002.0462	416.8273	-494.3903
29	927.8089	11.6734	361.9039	43.0893	10002.0462	416.8273	-494.3903

<sup>†</sup>, <sup>‡</sup> and <sup>§</sup> represents initial values for variance components, first iteration of the EM algorithm and first iteration of the AI algorithm, respectively.

Table S10. Convergence history for the variance components and  $\log(L)$  of the Model 1 via EM algorithm for dataset 2.

Iterations	$\hat{\sigma}_g^2$	$\hat{\sigma}_s^2$	$\hat{\sigma}_r^2$	$\hat{\sigma}^2$	$\log(L)$
0 <sup>†</sup>	1.1019	1.1019	1.1019	11.0194	-231.0372
1 <sup>‡</sup>	3.1552	1.2239	1.2079	10.7472	-225.3157
2	4.9235	1.3082	1.3300	9.0904	-222.1422
3	5.6048	1.4463	1.5339	8.1443	-220.2229
4	5.7814	1.6377	1.8147	7.5215	-218.4944
5	5.8078	1.8759	2.1582	6.9634	-216.7139
6	5.7911	2.1579	2.5459	6.4014	-214.8997
7	5.7595	2.4772	2.9506	5.8390	-213.1567
8	5.7212	2.8188	3.3390	5.3036	-211.6142
9	5.6797	3.1586	3.6821	4.8255	-210.3768
10	5.6386	3.4695	3.9635	4.4259	-209.4864
11	5.6009	3.7311	4.1803	4.1120	-208.9129
12	5.5689	3.9355	4.3395	3.8781	-208.5801
13	5.5436	4.0857	4.4523	3.7113	-208.4037
14	5.5246	4.1913	4.5300	3.5960	-208.3166
15	5.5110	4.2632	4.5826	3.5181	-208.2760
16	5.5016	4.3111	4.6176	3.4664	-208.2577
17	5.4952	4.3426	4.6407	3.4325	-208.2498
18	5.4910	4.3632	4.6558	3.4103	-208.2464
19	5.4882	4.3765	4.6656	3.3959	-208.2449
20	5.4864	4.3851	4.6720	3.3866	-208.2443
21	5.4853	4.3907	4.6761	3.3806	-208.2440
22	5.4845	4.3943	4.6788	3.3767	-208.2439
23	5.4840	4.3966	4.6805	3.3743	-208.2439
24	5.4837	4.3981	4.6816	3.3726	-208.2439
25	5.4835	4.3990	4.6823	3.3716	-208.2439
26	5.4834	4.3996	4.6828	3.3709	-208.2439
27	5.4833	4.4000	4.6831	3.3705	-208.2439
28	5.4832	4.4003	4.6833	3.3702	-208.2439
29	5.4832	4.4004	4.6834	3.3701	-208.2439
30	5.4832	4.4005	4.6835	3.3700	-208.2439
31	5.4832	4.4006	4.6835	3.3699	-208.2439
32	5.4831	4.4006	4.6836	3.3698	-208.2439
33	5.4831	4.4007	4.6836	3.3698	-208.2439
34	5.4831	4.4007	4.6836	3.3698	-208.2439

<sup>†</sup> and <sup>‡</sup> represents initial values for variance components and first iteration of the EM algorithm, respectively.

Table S11. Convergence history for the variance components and  $\log(L)$  of the Model 1 via AI algorithm for dataset 2.

Iterations	$\hat{\sigma}_g^2$	$\hat{\sigma}_s^2$	$\hat{\sigma}_r^2$	$\hat{\sigma}^2$	$\log(L)$
0 <sup>†</sup>	1.1019	1.1019	1.1019	11.0194	-231.0372
1 <sup>‡</sup>	2.1801	8.4969	8.4016	-13.5191	NA
2	3.4529	45.4869	32.5232	-108.1726	NA
3 <sup>§</sup>	-265.1453	-949.3418	-584.7233	2360.9958	NA

<sup>†</sup>, <sup>‡</sup> and <sup>§</sup> represents initial values for variance components, first iteration of the AI algorithm and fail in converge due singularity in AI matrix, respectively.

Table S12. Convergence history for the variance components and  $\log(L)$  of the Model 1 via EM-AI algorithm for dataset 2.

Iterations	$\hat{\sigma}_g^2$	$\hat{\sigma}_s^2$	$\hat{\sigma}_r^2$	$\hat{\sigma}^2$	$\log(L)$
0 <sup>†</sup>	1.1019	1.1019	1.1019	11.0194	-231.0372
1 <sup>‡</sup>	3.1552	1.2239	1.2079	10.7472	-225.3157
2	4.9235	1.3082	1.3300	9.0904	-222.1422
3	5.6048	1.4463	1.5339	8.1443	-220.2229
4	5.7814	1.6377	1.8147	7.5215	-218.4944
5	5.8078	1.8759	2.1582	6.9634	-216.7139
6	5.7911	2.1579	2.5459	6.4014	-214.8997
7	5.7595	2.4772	2.9506	5.8390	-213.1567
8	5.7212	2.8188	3.3390	5.3036	-211.6142
9	5.6797	3.1586	3.6821	4.8255	-210.3768
10	5.6386	3.4695	3.9635	4.4259	-209.4864
11	5.6009	3.7311	4.1803	4.1120	-208.9129
12	5.5689	3.9355	4.3395	3.8781	-208.5801
13	5.5436	4.0857	4.4523	3.7113	-208.4037
14	5.5246	4.1913	4.5300	3.5960	-208.3166
15	5.5110	4.2632	4.5826	3.5181	-208.2760
16	5.5016	4.3111	4.6176	3.4664	-208.2577
17	5.4952	4.3426	4.6407	3.4325	-208.2498
18	5.4910	4.3632	4.6558	3.4103	-208.2464
19	5.4882	4.3765	4.6656	3.3959	-208.2449
20	5.4864	4.3851	4.6720	3.3866	-208.2443
21	5.4853	4.3907	4.6761	3.3806	-208.2440
22	5.4845	4.3943	4.6788	3.3767	-208.2439
23	5.4840	4.3966	4.6805	3.3743	-208.2439
24	5.4837	4.3981	4.6816	3.3726	-208.2439
25	5.4835	4.3990	4.6823	3.3716	-208.2439
26 <sup>§</sup>	5.4832	4.4007	4.6836	3.3697	-208.2439
27	5.4831	4.4007	4.6836	3.3697	-208.2439

28                      5.4831                      4.4007                      4.6836                      3.3697                      -208.2439

<sup>†</sup>, <sup>‡</sup> and <sup>§</sup> represents initial values for variance components, first iteration of the EM algorithm and first iteration of the AI algorithm, respectively.

Table S13. Convergence history for the variance components and  $\log(L)$  of the Model 1 + RGCA via EM algorithm for dataset 2.

Iterations	$\hat{\sigma}_g^2$	$\hat{\sigma}_d^2$	$\hat{\sigma}_s^2$	$\hat{\sigma}_r^2$	$\hat{\sigma}^2$	$\log(L)$
0 <sup>†</sup>	1.1019	1.1019	1.1019	1.1019	11.0194	-226.7678
1 <sup>‡</sup>	3.1552	1.1382	1.2239	1.0581	9.2722	-218.2248
2	4.9928	1.1892	1.3574	1.0329	7.3236	-213.8094
3	5.6559	1.2526	1.5984	1.0382	6.3622	-211.6129
4	5.7990	1.3012	1.9378	1.0652	5.8091	-209.9816
5	5.7939	1.3318	2.3412	1.1073	5.3777	-208.5903
6	5.7486	1.3500	2.7646	1.1628	4.9966	-207.4450
7	5.6946	1.3607	3.1608	1.2313	4.6613	-206.5660
8	5.6433	1.3663	3.4950	1.3113	4.3789	-205.9304
9	5.6001	1.3679	3.7540	1.3996	4.1505	-205.4876
10	5.5663	1.3662	3.9429	1.4916	3.9708	-205.1854
11	5.5416	1.3622	4.0760	1.5821	3.8312	-204.9824
12	5.5240	1.3565	4.1684	1.6669	3.7235	-204.8481
13	5.5119	1.3500	4.2325	1.7429	3.6405	-204.7610
14	5.5036	1.3433	4.2776	1.8086	3.5767	-204.7057
15	5.4979	1.3369	4.3097	1.8639	3.5277	-204.6712
16	5.4939	1.3312	4.3329	1.9091	3.4901	-204.6500
17	5.4911	1.3262	4.3499	1.9456	3.4613	-204.6372
18	5.4890	1.3219	4.3625	1.9745	3.4394	-204.6296
19	5.4875	1.3184	4.3719	1.9973	3.4226	-204.6251
20	5.4864	1.3155	4.3789	2.0149	3.4099	-204.6224
21	5.4856	1.3133	4.3842	2.0286	3.4002	-204.6209
22	5.4850	1.3115	4.3883	2.0391	3.3928	-204.6200
23	5.4845	1.3101	4.3913	2.0472	3.3873	-204.6194
24	5.4842	1.3090	4.3936	2.0534	3.3830	-204.6191
25	5.4839	1.3081	4.3953	2.0581	3.3798	-204.6189
26	5.4837	1.3074	4.3966	2.0617	3.3774	-204.6188
27	5.4836	1.3069	4.3976	2.0644	3.3755	-204.6188
28	5.4835	1.3066	4.3984	2.0665	3.3741	-204.6188
29	5.4834	1.3063	4.3989	2.0680	3.3731	-204.6187
30	5.4833	1.3060	4.3994	2.0692	3.3723	-204.6187
31	5.4833	1.3059	4.3997	2.0701	3.3716	-204.6187
32	5.4832	1.3057	4.3999	2.0708	3.3712	-204.6187
33	5.4832	1.3056	4.4001	2.0713	3.3708	-204.6187

34	5.4832	1.3056	4.4003	2.0717	3.3706	-204.6187
35	5.4832	1.3055	4.4004	2.0720	3.3704	-204.6187
36	5.4832	1.3055	4.4005	2.0723	3.3702	-204.6187
37	5.4832	1.3054	4.4005	2.0724	3.3701	-204.6187
38	5.4831	1.3054	4.4006	2.0726	3.3700	-204.6187
39	5.4831	1.3054	4.4006	2.0727	3.3700	-204.6187
40	5.4831	1.3054	4.4006	2.0727	3.3699	-204.6187
41	5.4831	1.3054	4.4007	2.0728	3.3699	-204.6187
42	5.4831	1.3054	4.4007	2.0728	3.3698	-204.6187
43	5.4831	1.3054	4.4007	2.0729	3.3698	-204.6187
44	5.4831	1.3053	4.4007	2.0729	3.3698	-204.6187
45	5.4831	1.3053	4.4007	2.0729	3.3698	-204.6187

<sup>†</sup> and <sup>‡</sup> represents initial values for variance components and first iteration of the EM algorithm, respectively.

Table S14. Convergence history for the variance components and  $\log(L)$  of the Model 1 + RGCA via AI algorithm for dataset 2.

Iterations	$\hat{\sigma}_g^2$	$\hat{\sigma}_d^2$	$\hat{\sigma}_s^2$	$\hat{\sigma}_r^2$	$\hat{\sigma}^2$	$\log(L)$
0 <sup>†</sup>	1.1019	1.1019	1.1019	1.1019	11.0194	-226.7678
1 <sup>‡</sup>	2.1801	1.3389	8.4969	5.9912	-13.5191	NA
2	3.4529	1.3093	45.4869	29.9269	-108.1726	NA
3 <sup>§</sup>	-265.1453	1.3054	-949.3418	-587.3339	2360.9958	NA

<sup>†</sup>, <sup>‡</sup> and <sup>§</sup> represents initial values for variance components, first iteration of the AI algorithm and fail in converge due singularity in AI matrix, respectively.

Table S15. Convergence history for the variance components and  $\log(L)$  of the Model 1 + RGCA via EM-AI algorithm for dataset 2.

Iterations	$\hat{\sigma}_g^2$	$\hat{\sigma}_d^2$	$\hat{\sigma}_s^2$	$\hat{\sigma}_r^2$	$\hat{\sigma}^2$	$\log(L)$
0 <sup>†</sup>	1.1019	1.1019	1.1019	1.1019	11.0194	-226.7678
1 <sup>†</sup>	3.1552	1.1382	1.2239	1.0581	9.2722	-218.2248
2	4.9928	1.1892	1.3574	1.0329	7.3236	-213.8094
3	5.6559	1.2526	1.5984	1.0382	6.3622	-211.6129
4	5.7990	1.3012	1.9378	1.0652	5.8091	-209.9816
5	5.7939	1.3318	2.3412	1.1073	5.3777	-208.5903
6	5.7486	1.3500	2.7646	1.1628	4.9966	-207.4450
7	5.6946	1.3607	3.1608	1.2313	4.6613	-206.5660
8	5.6433	1.3663	3.4950	1.3113	4.3789	-205.9304
9	5.6001	1.3679	3.7540	1.3996	4.1505	-205.4876
10	5.5663	1.3662	3.9429	1.4916	3.9708	-205.1854
11	5.5416	1.3622	4.0760	1.5821	3.8312	-204.9824
12	5.5240	1.3565	4.1684	1.6669	3.7235	-204.8481

13	5.5119	1.3500	4.2325	1.7429	3.6405	-204.7610
14	5.5036	1.3433	4.2776	1.8086	3.5767	-204.7057
15	5.4979	1.3369	4.3097	1.8639	3.5277	-204.6712
16	5.4939	1.3312	4.3329	1.9091	3.4901	-204.6500
17	5.4911	1.3262	4.3499	1.9456	3.4613	-204.6372
18	5.4890	1.3219	4.3625	1.9745	3.4394	-204.6296
19	5.4875	1.3184	4.3719	1.9973	3.4226	-204.6251
20	5.4864	1.3155	4.3789	2.0149	3.4099	-204.6224
21	5.4856	1.3133	4.3842	2.0286	3.4002	-204.6209
22	5.4850	1.3115	4.3883	2.0391	3.3928	-204.6200
23	5.4845	1.3101	4.3913	2.0472	3.3873	-204.6194
24	5.4842	1.3090	4.3936	2.0534	3.3830	-204.6191
25	5.4839	1.3081	4.3953	2.0581	3.3798	-204.6189
26 <sup>§</sup>	5.4833	1.3053	4.4007	2.0729	3.3697	-204.6187
27	5.4831	1.3053	4.4007	2.0730	3.3697	-204.6187
28	5.4831	1.3053	4.4007	2.0730	3.3697	-204.6187

<sup>†</sup>, <sup>‡</sup> and <sup>§</sup> represents initial values for variance components, first iteration of the EM algorithm and first iteration of the AI algorithm, respectively.

Table S16. Convergence history for the variance components and  $\log(L)$  of the Model 3 via EM algorithm for dataset 2.

Iterations	$\hat{\sigma}_g^2$	$\hat{\sigma}_d^2$	$\hat{\sigma}_s^2$	$\hat{\sigma}_r^2$	$\hat{\sigma}_p^2$	$\hat{\sigma}^2$	$\log(L)$
0 <sup>†</sup>	1.1019	1.1019	1.1019	1.1019	1.1019	11.0194	-232.7202
1 <sup>‡</sup>	3.0386	1.1382	1.2392	1.0581	1.4035	10.5281	-226.4721
2	5.1876	1.1694	1.3338	1.0217	1.8822	8.8688	-222.5475
3	6.2550	1.2155	1.4736	1.0039	2.8690	7.8087	-219.7336
4	6.5985	1.2591	1.6671	1.0020	4.7911	7.0832	-216.9259
5	6.6889	1.2936	1.9107	1.0135	7.8818	6.4202	-214.3240
6	6.7027	1.3208	2.2037	1.0399	11.3834	5.7974	-212.3236
7	6.6895	1.3425	2.5368	1.0836	14.0154	5.2755	-210.8879
8	6.6614	1.3582	2.8842	1.1448	15.4805	4.8679	-209.8500
9	6.6244	1.3679	3.2137	1.2209	16.2010	4.5520	-209.0985
10	6.5845	1.3720	3.5010	1.3079	16.5644	4.3041	-208.5578
11	6.5467	1.3719	3.7359	1.4012	16.7705	4.1074	-208.1718
12	6.5140	1.3686	3.9195	1.4958	16.9045	3.9506	-207.8993
13	6.4875	1.3632	4.0588	1.5870	17.0009	3.8253	-207.7102
14	6.4670	1.3567	4.1625	1.6713	17.0741	3.7255	-207.5819
15	6.4517	1.3497	4.2391	1.7462	17.1312	3.6464	-207.4969
16	6.4404	1.3428	4.2955	1.8108	17.1761	3.5841	-207.4420
17	6.4321	1.3365	4.3370	1.8649	17.2112	3.5354	-207.4073
18	6.4262	1.3308	4.3677	1.9093	17.2387	3.4976	-207.3859

19	6.4219	1.3258	4.3903	1.9452	17.2601	3.4684	-207.3728
20	6.4188	1.3217	4.4072	1.9737	17.2766	3.4459	-207.3649
21	6.4165	1.3182	4.4197	1.9962	17.2893	3.4287	-207.3602
22	6.4148	1.3154	4.4291	2.0137	17.2991	3.4155	-207.3575
23	6.4136	1.3132	4.4361	2.0273	17.3065	3.4055	-207.3558
24	6.4127	1.3114	4.4414	2.0378	17.3122	3.3979	-207.3549
25	6.4120	1.3100	4.4454	2.0459	17.3166	3.3921	-207.3543
26	6.4115	1.3090	4.4484	2.0521	17.3199	3.3877	-207.3540
27	6.4112	1.3081	4.4506	2.0568	17.3224	3.3843	-207.3538
28	6.4109	1.3075	4.4523	2.0604	17.3243	3.3818	-207.3537
29	6.4107	1.3070	4.4536	2.0632	17.3257	3.3799	-207.3537
30	6.4105	1.3066	4.4546	2.0653	17.3268	3.3784	-207.3536
31	6.4104	1.3063	4.4553	2.0669	17.3276	3.3773	-207.3536
32	6.4103	1.3061	4.4559	2.0681	17.3283	3.3764	-207.3536
33	6.4102	1.3059	4.4563	2.0690	17.3287	3.3758	-207.3536
34	6.4102	1.3057	4.4566	2.0697	17.3291	3.3753	-207.3536
35	6.4101	1.3056	4.4569	2.0703	17.3294	3.3749	-207.3536
36	6.4101	1.3056	4.4570	2.0707	17.3296	3.3747	-207.3536
37	6.4101	1.3055	4.4572	2.0710	17.3297	3.3745	-207.3536
38	6.4101	1.3055	4.4573	2.0712	17.3299	3.3743	-207.3536
39	6.4101	1.3054	4.4574	2.0714	17.3300	3.3742	-207.3536
40	6.4100	1.3054	4.4574	2.0715	17.3300	3.3741	-207.3536
41	6.4100	1.3054	4.4575	2.0716	17.3301	3.3740	-207.3536
42	6.4100	1.3054	4.4575	2.0717	17.3301	3.3740	-207.3536
43	6.4100	1.3054	4.4575	2.0718	17.3302	3.3739	-207.3536
44	6.4100	1.3054	4.4575	2.0718	17.3302	3.3739	-207.3536

<sup>†</sup> and <sup>‡</sup> represents initial values for variance components and first iteration of the EM algorithm, respectively.

Table S17. Convergence history for the variance components and  $\log(L)$  of the Model 3 via AI algorithm for dataset 2.

Iterations	$\hat{\sigma}_g^2$	$\hat{\sigma}_d^2$	$\hat{\sigma}_s^2$	$\hat{\sigma}_r^2$	$\hat{\sigma}_p^2$	$\hat{\sigma}^2$	$\log(L)$
0 <sup>†</sup>	1.1019	1.1019	1.1019	1.1019	1.1019	11.0194	-232.7202
1 <sup>‡</sup>	2.2542	1.3389	8.3957	6.0996	17.9010	-13.9527	NA
2	3.9716	1.3093	26.6739	24.2857	58.3775	-85.6080	NA
3 <sup>§</sup>	5.5915	1.3054	591.1688	588.7830	1190.1293	-2343.4716	NA

<sup>†</sup>, <sup>‡</sup> and <sup>§</sup> represents initial values for variance components, first iteration of the AI algorithm and fail in converge due singularity in AI matrix, respectively.

Table S18. Convergence history for the variance components and  $\log(L)$  of the Model 3 via EM-AI algorithm for dataset 2.

Iterations	$\hat{\sigma}_g^2$	$\hat{\sigma}_d^2$	$\hat{\sigma}_s^2$	$\hat{\sigma}_r^2$	$\hat{\sigma}_p^2$	$\hat{\sigma}^2$	$\log(L)$
0 <sup>†</sup>	1.1019	1.1019	1.1019	1.1019	1.1019	11.0194	-232.7202
1 <sup>‡</sup>	3.0386	1.1382	1.2392	1.0581	1.4035	10.5281	-226.4721
2	5.1876	1.1694	1.3338	1.0217	1.8822	8.8688	-222.5475
3	6.2550	1.2155	1.4736	1.0039	2.8690	7.8087	-219.7336
4	6.5985	1.2591	1.6671	1.0020	4.7911	7.0832	-216.9259
5	6.6889	1.2936	1.9107	1.0135	7.8818	6.4202	-214.3240
6	6.7027	1.3208	2.2037	1.0399	11.3834	5.7974	-212.3236
7	6.6895	1.3425	2.5368	1.0836	14.0154	5.2755	-210.8879
8	6.6614	1.3582	2.8842	1.1448	15.4805	4.8679	-209.8500
9	6.6244	1.3679	3.2137	1.2209	16.2010	4.5520	-209.0985
10	6.5845	1.3720	3.5010	1.3079	16.5644	4.3041	-208.5578
11	6.5467	1.3719	3.7359	1.4012	16.7705	4.1074	-208.1718
12	6.5140	1.3686	3.9195	1.4958	16.9045	3.9506	-207.8993
13	6.4875	1.3632	4.0588	1.5870	17.0009	3.8253	-207.7102
14	6.4670	1.3567	4.1625	1.6713	17.0741	3.7255	-207.5819
15	6.4517	1.3497	4.2391	1.7462	17.1312	3.6464	-207.4969
16	6.4404	1.3428	4.2955	1.8108	17.1761	3.5841	-207.4420
17	6.4321	1.3365	4.3370	1.8649	17.2112	3.5354	-207.4073
18	6.4262	1.3308	4.3677	1.9093	17.2387	3.4976	-207.3859
19	6.4219	1.3258	4.3903	1.9452	17.2601	3.4684	-207.3728
20	6.4188	1.3217	4.4072	1.9737	17.2766	3.4459	-207.3649
21	6.4165	1.3182	4.4197	1.9962	17.2893	3.4287	-207.3602
22	6.4148	1.3154	4.4291	2.0137	17.2991	3.4155	-207.3575
23	6.4136	1.3132	4.4361	2.0273	17.3065	3.4055	-207.3558
24	6.4127	1.3114	4.4414	2.0378	17.3122	3.3979	-207.3549
25	6.4120	1.3100	4.4454	2.0459	17.3166	3.3921	-207.3543
26 <sup>§</sup>	6.4100	1.3053	4.4576	2.0718	17.3303	3.3737	-207.3536
27	6.4100	1.3053	4.4576	2.0720	17.3302	3.3738	-207.3536
28	6.4100	1.3053	4.4576	2.0720	17.3302	3.3738	-207.3536

<sup>†</sup>, <sup>‡</sup> and <sup>§</sup> represents initial values for variance components, first iteration of the EM algorithm and first iteration of the AI algorithm, respectively.

**CAPÍTULO 2**  
**ANÁLISE DIALÉLICA MULTIAMBIENTES E MULTIGERAÇÕES NO**  
**MELHORAMENTO DE PLANTAS AUTÓGAMAS**

**VIÇOSA - MINAS GERAIS**  
**2019**

## RESUMO

O sucesso no melhoramento de plantas autógamas é atribuído ao desenvolvimento de linhagens elite com potencial para substituir as cultivares em uso. Neste contexto, é de suma importância descartar precocemente as populações de menor mérito genético. Dentre as ferramentas disponíveis para esta tomada de decisão, destaca-se o dialelo. Contudo, devido à interação genótipos por ambientes e a limitação de sementes  $F_1$ , os ensaios são conduzidos em *multi-environment and multi-generation trials* (MEGT). Para os *multi-environment trials* (MET) a metodologia de modelos mistos tem sido amplamente utilizada. Portanto, acredita-se que esta mesma metodologia, associada a um índice de seleção, possa ser aplicada também com sucesso em dados MEGT. Assim, objetivou-se com este trabalho descrever estratégias computacionalmente eficientes para modelar dados MEGT no contexto de análise dialélica via metodologia de modelos mistos, bem como aplicar o índice de seleção baseado na distância genótipos-ideótipo (IGI) aos efeitos de capacidade geral e específica de combinação visando à seleção de populações promissoras para a extração de linhagens. Foram utilizadas populações oriundas de cruzamentos biparentais entre dois grupos de linhagens de feijão-comum, em esquema de dialelo parcial. O grupo 1 foi composto por cinco linhagens de grão preto e o grupo 2 por sete linhagens de grão carioca. As populações juntamente com as linhagens parentais foram avaliadas em cinco ensaios em condição de campo sob o delineamento de blocos casualizados, nos quais se avaliou o caráter produtividade de grãos. Diferentes algoritmos [*expectation maximization* (EM), *average information* (AI) e EM-AI] e quatro modelos [considerando diferentes estruturas de (co)variância] foram utilizados para modelar os dados dialélicos em MEGT. O algoritmo EM convergiu para três dos quatro modelos utilizados. Em contrapartida, o algoritmo AI convergiu apenas para dois modelos. Além disso, o modelo que admite heterogeneidade de

variâncias e correlação uniforme para os efeitos genéticos bem como variâncias heterogêneas para o resíduo se adequou melhor aos dados MEGT, e o IGI se mostrou eficiente na seleção de populações promissoras para a extração de linhagens via análise dialélica em MEGT. Assim, conclui-se que o algoritmo AI e o IGI têm potencial para uso nos programas de melhoramento de plantas autógamas para a análise dialélica em MEGT.

Palavras chave: Capacidade combinatória, interação genótipos por ambientes, modelos mistos.

## 1. INTRODUÇÃO

Nos programas de melhoramento de plantas autógamas, três etapas se destacam: a escolha de genitores/obtenção de populações segregantes, o avanço das gerações de endogamia/seleção (obtenção de linhagens) e a avaliação das linhagens elite visando a recomendação de novas cultivares (Ramalho et al., 2012). Atenção especial deve ser dada à primeira etapa, pois o investimento em populações com baixo potencial genético pode onerar o programa de melhoramento sem prover retorno positivo, ou seja, não gerar linhagens superiores. Neste contexto, é de suma importância quantificar o potencial das populações segregantes em gerações precoces, buscando descartar aquelas de menor mérito genético.

Dentre as metodologias disponíveis para quantificar o potencial genético de populações segregantes, destaca-se a de dialelo. Na literatura são descritos alguns métodos para a análise dialélica, como o de Hayman (1954), Griffing (1956a), Gardner e Eberhart (1966), Cockerham e Weir (1977), Geraldi e Miranda Filho (1988) e Zhu e Weir (1994). Contudo, cabe ressaltar que em espécies autógamas a análise dialélica baseada nos métodos de Griffing (1956a, 1956b) tem sido utilizada com maior frequência (Bhullar et al., 1979; Patil e Chopde, 1981; Huang et al., 2015; Hei et al., 2016; Kaushik et al., 2018; Moura et al., 2018; Rocha et al., 2018a). A partir deste tipo de análise dialélica é possível estimar os efeitos da capacidade geral de combinação (CGC) e capacidade específica de combinação (CEC), os quais informam, respectivamente, sobre a frequência de alelos favoráveis nos genitores e a diversidade genética entre eles para o caráter em questão (Viana e Matta, 2003; Moura et al., 2018). Estas informações são úteis no momento da seleção das populações segregantes de maior potencial para a extração de linhagens.

Embora a análise dialélica forneça valiosas informações aos melhoristas a respeito da herança do caráter em estudo, bem como da capacidade combinatória para o mesmo, sabe-se que

essas podem mudar de um ambiente para outro (Patil e Chopde, 1981; Ogut et al., 2014; Pádua et al., 2016; Moura et al., 2018). Estas mudanças são atribuídas à interação genótipos por ambientes (G×A). Segundo Zhang et al. (2015), a interação G×A resulta da expressão diferencial dos genes envolvidos no controle do caráter em avaliação nas diferentes condições ambientais. Assim, fica evidente a importância da avaliação dos dialelos em vários locais, anos ou safras, visando uma escolha mais acertada das populações segregantes de maior mérito genético.

Os ensaios realizados em um conjunto de ambientes, denominados de *multi-environment trials* (MET) (Smith et al., 2001; Smith et al., 2005; Smith et al., 2015), são rotineiros no estudo da interação G×A (Smith et al., 2001; Smith et al., 2005; Kelly et al., 2007; Ogut et al., 2014; Smith et al., 2015) e para a seleção de genótipos que apresentem bom desempenho e estabilidade (Zhang et al., 2015). Apesar da relevância da interação G×A nos programas de melhoramento, a avaliação de dialelos em MET tem sido menos frequente. Também é oportuno ressaltar que, para algumas espécies autógamas, como trigo, soja e feijão, a baixa quantidade de sementes F<sub>1</sub>, obtidas a partir da hibridação dos genitores, é um fator limitante para a avaliação dos dialelos em MET (Bhullar et al., 1979; Cho e Scott, 2000; Moura et al., 2018). Neste contexto, alguns autores têm utilizado a geração F<sub>2</sub> (Bhullar et al., 1979; Zhu e Weir, 1994; Cho e Scott, 2000; Moura et al., 2018). Entretanto, Zhu e Weir (1994) mostraram que a contribuição dos desvios de dominância reduz à metade na geração F<sub>2</sub>, o que pode comprometer a detecção da divergência genética entre os genitores via CEC. Por outro lado, Bhullar et al. (1979) e Patil e Chopde (1981) mostraram que as gerações F<sub>2</sub> ou F<sub>3</sub> podem fornecer estimativas mais confiáveis de CGC que a geração F<sub>1</sub>. Tal fato pode estar associado à possibilidade de detecção de efeitos de genes recessivos nas gerações F<sub>2</sub> ou F<sub>3</sub>. Desta forma, é desejável que se combine as informações das gerações disponíveis para ampliar a quantidade de ensaios e melhorar a tomada de decisão com

relação ao descarte de populações com menor mérito genético para a derivação de linhagens em programas de melhoramento.

Nos programas melhoramento de plantas autógamas, devido a maior praticidade, os ensaios envolvem a avaliação de sucessivas gerações de endogamia ( $F_1$ ,  $F_2$ , etc), ou seja, são realizados de forma sequencial (Resende et al., 2015; Batista et al., 2017). Neste contexto, diferentes gerações são avaliadas em diferentes ambientes (anos, safras ou locais), o que permite caracterizar tal conjunto de ensaios como *multi-environment and multi-generation trials* (MEGT).

Smith et al. (2001), usando um grande conjunto de dados em MET de um programa de melhoramento de cevada do Sul da Austrália, consideraram a característica rendimento de grãos avaliada em cada ambiente como características diferentes. Estes autores assumiram variâncias e covariâncias genóticas heterogêneas entre os diferentes ambientes. Neste cenário, os valores genóticos são preditos para cada ensaio, o que pode dificultar a classificação dos genótipos na presença de interação  $G \times A$ . Portanto, a utilização de um índice de seleção capaz de combinar os valores genóticos dos múltiplos ensaios é desejável (Kelly et al., 2007). Além disso, Smith et al. (2001, 2005) sugeriram a utilização da abordagem de modelos mistos pela maior flexibilidade para modelar as estruturas de (co)variância genotípica e residual, visando conectar experimentos que possuem desbalanceamento estatístico e/ou genético.

Os modelos lineares mistos assumem que as matrizes de (co)variância associadas aos seus efeitos aleatórios são conhecidas (Henderson et al., 1959). Entretanto, elas não são conhecidas, mas podem ser substituídas por suas estimativas obtidas via o método de máxima verossimilhança residual (REML) por meio de algoritmos iterativos (Patterson e Thompson, 1971; Gilmour et al., 1995). Smith et al. (2001) sugeriram a utilização do algoritmo *average*

*information* (AI) para maximizar a função de verossimilhança residual ao se modelar dados de MET. Isto pode ser justificado pelas vantagens atribuídas a este algoritmo como: maior eficiência computacional, possibilidade de estimação de grande número de parâmetros, eficiente para grandes conjuntos de dados e fácil generalização para estruturas de (co)variância complexas (Gilmour et al., 1995; Thompson, 2008; Gilmour et al. 2019). Contudo, em algumas situações, o algoritmo AI pode não convergir (Meyer, 2006; Knighth, 2008; Diffey et al., 2017). Neste contexto, Meyer (2006) sugere a combinação entre os algoritmos *parameter expanded expectation maximization* (PXEM) e AI, visando explorar as vantagens complementares destes algoritmos (maior estabilidade e rápida convergência, respectivamente). Entretanto, apesar do algoritmo PXEM ser um pouco mais eficiente que o algoritmo *expectation maximization* (EM) (Knighth, 2008; Meyer, 2008), sua implementação requer o cálculo de um traço matricial adicional que pode ser caro computacionalmente (Knighth, 2008; Diffey et al., 2017). Neste sentido, Marçal et al. (2019) sugeriram a utilização da combinação entre os algoritmos EM e AI.

Tendo-se em vista a carência de informações disponíveis na literatura a respeito da análise dialética de dados MET (Ogut et al., 2014; Pádua et al., 2016) bem como a ausência de informação a respeito da análise dialética de dados MEGT, buscou-se com esse trabalho, trazer luz ao problema e subsidiar a literatura com informação a respeito da modelagem de dados MEGT no contexto de análise dialética.

Diante do exposto, objetivou-se descrever os detalhes da implementação dos algoritmos EM, AI e EM-AI para se realizar a análise dialética de dados MEGT, comumente realizados em espécies autógamas, utilizando a abordagem proposta por Smith et al. (2001, 2005) para dados MET. Além disso, adaptou-se o índice baseado na distância genótipos-ideótipo (Rocha et al., 2018c) para o contexto de análise dialética em MEGT.

## 2. MATERIAL E MÉTODOS

### 2.1. FORMA GERAL DO MODELO PARA ANÁLISE DIALÉLICA EM MEGT

Nos programas de melhoramento de plantas autógamas tem sido mais frequente o uso do dialelo de meia tabela (Griffing, 1956a) ou em esquema fatorial, comumente denominado de dialelo parcial (Geraldi e Miranda Filho, 1988). Nos dialelos de meia tabela, em geral são avaliadas  $p(p - 1)/2$  combinações híbridas juntamente com as  $p$  linhagens parentais. Enquanto, nos dialelos parciais, geralmente são avaliadas  $pq$  combinações híbridas, obtidas pelo cruzamento fatorial entre linhagens parentais de dois grupos ( $p$  e  $q$ ), juntamente com as  $p + q$  linhagens parentais. Independentemente do esquema dialélico utilizado, os tratamentos (combinações híbridas + linhagens parentais) são avaliados em experimentos adotando-se delineamento experimental apropriado. Entretanto, para se obter previsões de CGC e CEC, livres do efeito das linhagens parentais, deve-se separar a média dos cruzamentos da média das linhagens parentais e incluir as linhagens parentais em um efeito aleatório adicional (Möhrling et al., 2011; Marçal et al., 2019).

Considerando um dialelo de meia tabela, avaliado em MEGT, a forma geral do modelo linear misto para análise dialélica é apresentada na Eq. [1]. Neste modelo (Eq. [1]), a interação  $G \times A$  pode ser incluída modelando-se os efeitos genéticos aleatórios  $\mathbf{g}$ ,  $\mathbf{s}$  e  $\mathbf{p}$  com uma estrutura de (co)variância apropriada, ou seja, que admita correlação entre ensaios.

$$\mathbf{y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{Z}_g\mathbf{g} + \mathbf{Z}_s\mathbf{s} + \mathbf{Z}_p\mathbf{p} + \mathbf{Z}_b\mathbf{b} + \mathbf{e} \quad [1]$$

em que:  $\mathbf{y}$  o vetor de observações fenotípicas dos cruzamentos e linhagens parentais em todos os ensaios;  $\boldsymbol{\tau}$  é o vetor de efeitos fixos de ensaios, repetições dentro de ensaios, média dos cruzamentos dentro de cada ensaio e média das linhagens parentais dentro de cada ensaio;  $\mathbf{g}$  é o vetor de efeitos aleatórios da CGC dos genitores em cada ensaio;  $\mathbf{s}$  é o vetor de efeitos aleatórios

da CEC dos cruzamentos em cada ensaio;  $\mathbf{p}$  é o vetor de efeitos aleatórios das linhagens parentais em cada ensaio;  $\mathbf{b}$  é o vetor de efeitos aleatórios de blocos, associado a um delineamento experimental de blocos incompletos qualquer;  $\mathbf{e}$  é o vetor de erros aleatórios em cada ensaio e  $\mathbf{X}$ ,  $\mathbf{Z}_g$ ,  $\mathbf{Z}_s$ ,  $\mathbf{Z}_p$  e  $\mathbf{Z}_b$  são as matrizes de incidência dos efeitos  $\tau$ ,  $\mathbf{g}$ ,  $\mathbf{s}$ ,  $\mathbf{p}$  e  $\mathbf{b}$ , respectivamente. Cabe ressaltar que para a análise dialélica em esquema fatorial, basta particionar o efeito de  $\mathbf{g}$  do modelo acima em  $\mathbf{g} = (\mathbf{g}_1, \mathbf{g}_2)^T$ , em que  $\mathbf{g}_1$  e  $\mathbf{g}_2$  referem-se, respectivamente, aos efeitos de CGC dos grupos  $p$  e  $q$  de linhagens parentais.

O vetor de soluções  $\boldsymbol{\beta} = (\hat{\tau}, \tilde{\mathbf{g}}, \tilde{\mathbf{s}}, \tilde{\mathbf{p}}, \tilde{\mathbf{b}})^T$  para os efeitos fixos e aleatórios do modelo descrito anteriormente (Eq. [1]) pode ser obtido pela Eq. [2]. A Eq. [2], apresentada a seguir, pode ser obtida maximizando-se a função de distribuição conjunta de  $\mathbf{y}$  e  $\mathbf{u} = (\mathbf{g}, \mathbf{s}, \mathbf{p}, \mathbf{b})^T$  (Henderson et al., 1959):

$$\begin{bmatrix} \mathbf{X}^T \mathbf{R}^{-1} \mathbf{X} & \mathbf{X}^T \mathbf{R}^{-1} \mathbf{Z}_g & \mathbf{X}^T \mathbf{R}^{-1} \mathbf{Z}_s & \mathbf{X}^T \mathbf{R}^{-1} \mathbf{Z}_p & \mathbf{X}^T \mathbf{R}^{-1} \mathbf{Z}_b \\ \mathbf{Z}_g^T \mathbf{R}^{-1} \mathbf{X} & \mathbf{Z}_g^T \mathbf{R}^{-1} \mathbf{Z}_g + \mathbf{G}_g^{-1} & \mathbf{Z}_g^T \mathbf{R}^{-1} \mathbf{Z}_s & \mathbf{Z}_g^T \mathbf{R}^{-1} \mathbf{Z}_p & \mathbf{Z}_g^T \mathbf{R}^{-1} \mathbf{Z}_b \\ \mathbf{Z}_s^T \mathbf{R}^{-1} \mathbf{X} & \mathbf{Z}_s^T \mathbf{R}^{-1} \mathbf{Z}_g & \mathbf{Z}_s^T \mathbf{R}^{-1} \mathbf{Z}_s + \mathbf{G}_s^{-1} & \mathbf{Z}_s^T \mathbf{R}^{-1} \mathbf{Z}_p & \mathbf{Z}_s^T \mathbf{R}^{-1} \mathbf{Z}_b \\ \mathbf{Z}_p^T \mathbf{R}^{-1} \mathbf{X} & \mathbf{Z}_p^T \mathbf{R}^{-1} \mathbf{Z}_g & \mathbf{Z}_p^T \mathbf{R}^{-1} \mathbf{Z}_s & \mathbf{Z}_p^T \mathbf{R}^{-1} \mathbf{Z}_p + \mathbf{G}_p^{-1} & \mathbf{Z}_p^T \mathbf{R}^{-1} \mathbf{Z}_b \\ \mathbf{Z}_b^T \mathbf{R}^{-1} \mathbf{X} & \mathbf{Z}_b^T \mathbf{R}^{-1} \mathbf{Z}_g & \mathbf{Z}_b^T \mathbf{R}^{-1} \mathbf{Z}_s & \mathbf{Z}_b^T \mathbf{R}^{-1} \mathbf{Z}_p & \mathbf{Z}_b^T \mathbf{R}^{-1} \mathbf{Z}_b + \mathbf{G}_b^{-1} \end{bmatrix} \times \begin{bmatrix} \hat{\tau} \\ \tilde{\mathbf{g}} \\ \tilde{\mathbf{s}} \\ \tilde{\mathbf{p}} \\ \tilde{\mathbf{b}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T \mathbf{R}^{-1} \mathbf{y} \\ \mathbf{Z}_g^T \mathbf{R}^{-1} \mathbf{y} \\ \mathbf{Z}_s^T \mathbf{R}^{-1} \mathbf{y} \\ \mathbf{Z}_p^T \mathbf{R}^{-1} \mathbf{y} \\ \mathbf{Z}_b^T \mathbf{R}^{-1} \mathbf{y} \end{bmatrix} \quad [2]$$

Assumiu-se a seguinte estrutura de (co)variância para o modelo [1]:

$$\begin{bmatrix} \mathbf{g} \\ \mathbf{s} \\ \mathbf{p} \\ \mathbf{b} \\ \mathbf{e} \end{bmatrix} \sim N \left( \mathbf{0}, \begin{bmatrix} \mathbf{G}_g & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_s & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{G}_p & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{G}_b & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{R} \end{bmatrix} \right) \quad [3]$$

$$\mathbf{V} = \mathbf{Z}_g \mathbf{G}_g \mathbf{Z}_g^T + \mathbf{Z}_s \mathbf{G}_s \mathbf{Z}_s^T + \mathbf{Z}_p \mathbf{G}_p \mathbf{Z}_p^T + \mathbf{Z}_b \mathbf{G}_b \mathbf{Z}_b^T + \mathbf{R} \quad [4]$$

em que:  $\mathbf{G}_g$  é a matriz de (co)variância associada ao vetor de efeitos aleatórios da CGC, com  $\mathbf{G}_g = \mathbf{I}_g \otimes \mathbf{G}_{0g}$ ;  $\mathbf{G}_s$  é a matriz de (co)variância dos efeitos aleatórios de CEC, com  $\mathbf{G}_s = \mathbf{I}_s \otimes \mathbf{G}_{0s}$ ;  $\mathbf{G}_p$  é a matriz de (co)variância dos efeitos aleatórios das linhagens parentais, com  $\mathbf{G}_p = \mathbf{I}_p \otimes \mathbf{G}_{0p}$ ;  $\mathbf{G}_b$  é a matriz de (co)variância associada ao vetor de efeitos aleatórios de blocos, com  $\mathbf{G}_b = \bigoplus_{t=1}^{N_t} \mathbf{I}_{b_t} \sigma_{b_t}^2$ ;  $\mathbf{R}$  é a matriz de (co)variância dos erros aleatórios, com  $\mathbf{R} = \bigoplus_{t=1}^{N_t} \mathbf{I}_{n_t} \sigma_{e_t}^2$  e  $\mathbf{V}$  é a matriz de (co)variância associada ao vetor  $\mathbf{y}$ ;  $\mathbf{I}_g, \mathbf{I}_s, \mathbf{I}_p, \mathbf{I}_{b_t}$  e  $\mathbf{I}_{n_t}$  são matrizes identidade associadas às matrizes (co)variância dos efeitos aleatórios  $\mathbf{g}, \mathbf{s}, \mathbf{p}, \mathbf{b}$  e  $\mathbf{e}$ , respectivamente.

## 2.2. ESTIMAÇÃO DAS MATRIZES DE (CO)VARIÂNCIA

A estimação das matrizes de (co)variância associadas aos efeitos aleatórios da Eq. [1] ( $\mathbf{G}_g, \mathbf{G}_s, \mathbf{G}_p, \mathbf{G}_b$  e  $\mathbf{R}$ ) pode ser realizada maximizando-se a função de verossimilhança residual descrita por Johnson e Thompson (1995) (Eq. [5]). A equação apresentada pelos autores foi descrita de forma generalizada, para um modelo linear misto qualquer, como mostrado a seguir:

$$-2\log(L) = \text{const.} + \log|\mathbf{V}| + \log|\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X}| + \mathbf{y}^T \mathbf{P} \mathbf{y} \quad [5]$$

em que  $\log(L)$  é o logaritmo da função de verossimilhança residual (Patterson e Thompson, 1971), const. é uma constante da função de verossimilhança residual e  $\mathbf{P}$  é o projetor ortogonal.

A matriz  $\mathbf{P}$  presente na função de verossimilhança residual pode ser escrita como (Johnson e Thompson, 1995; Gilmour et al., 1995):

$$\begin{aligned}\mathbf{P} &= \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1} \\ &= \mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{W}\mathbf{C}^{-1}\mathbf{W}^T\mathbf{R}^{-1}\end{aligned}\quad [6]$$

$$\mathbf{W} = [\mathbf{X}, \mathbf{Z}_g, \mathbf{Z}_s, \mathbf{Z}_p, \mathbf{Z}_b] \quad [7]$$

em que  $\mathbf{C}^{-1}$  é a inversa do lado esquerdo do sistema de equações de modelos mistos apresentado na Eq. [2].

Considerando a Eq. [6] nota-se que a segunda expressão associada à matriz  $\mathbf{P}$  não utiliza a matriz  $\mathbf{V}^{-1}$ . Este fato é desejável, pois, a inversão da matriz  $\mathbf{V}$  consiste em um processo computacionalmente caro, devido à dimensão da mesma ( $N \times N$ ), em que  $N$  equivale ao número de observações do vetor  $\mathbf{y}$ . Portanto, buscou-se evitar a inversão da matriz  $\mathbf{V}$  reescrevendo-se a Eq. [5] de forma específica para a Eq. [1], como mostrado a seguir:

$$\begin{aligned}-2\log(L) &= \text{const.} + \log|\mathbf{R}| + \log|\mathbf{G}_g| + \log|\mathbf{G}_s| \\ &\quad + \log|\mathbf{G}_p| + \log|\mathbf{G}_b| + \log|\mathbf{C}| + \mathbf{y}^T\mathbf{P}\mathbf{y}\end{aligned}\quad [8]$$

em que  $\log|\mathbf{C}|$  é o logaritmo do determinante do lado esquerdo ( $\mathbf{C}$ ) do sistema de equações de modelos mistos apresentado na Eq. [2].

A partir da função de verossimilhança residual pode-se estimar as estruturas de (co)variância associadas aos efeitos aleatórios da Eq. [1], encontrando-se os zeros das derivadas de primeira ordem em relação aos parâmetros de cada estrutura ( $\mathbf{G}_g$ ,  $\mathbf{G}_s$ ,  $\mathbf{G}_p$ ,  $\mathbf{G}_b$  e  $\mathbf{R}$ ). As derivadas de primeira ordem da Eq.[5] ou Eq. [8] em relação aos parâmetros das estruturas de (co)variância associadas aos efeitos aleatórios da Eq. [1] são equivalentes e podem ser escritas como:

$$\begin{aligned}
U(\boldsymbol{\kappa}) &= \frac{\partial \log(L)}{\partial \kappa_i} = -\frac{1}{2} \begin{bmatrix} \text{tr}\left(\mathbf{PZ}_g \dot{\mathbf{G}}_{g(i)} \mathbf{Z}_g^T\right) - \mathbf{y}^T \mathbf{PZ}_g \dot{\mathbf{G}}_{g(i)} \mathbf{Z}_g^T \mathbf{Py} \\ \text{tr}\left(\mathbf{PZ}_s \dot{\mathbf{G}}_{s(i)} \mathbf{Z}_s^T\right) - \mathbf{y}^T \mathbf{PZ}_s \dot{\mathbf{G}}_{s(i)} \mathbf{Z}_s^T \mathbf{Py} \\ \text{tr}\left(\mathbf{PZ}_p \dot{\mathbf{G}}_{p(i)} \mathbf{Z}_p^T\right) - \mathbf{y}^T \mathbf{PZ}_p \dot{\mathbf{G}}_{p(i)} \mathbf{Z}_p^T \mathbf{Py} \\ \text{tr}\left(\mathbf{PZ}_b \dot{\mathbf{G}}_{b(i)} \mathbf{Z}_b^T\right) - \mathbf{y}^T \mathbf{PZ}_b \dot{\mathbf{G}}_{b(i)} \mathbf{Z}_b^T \mathbf{Py} \\ \text{tr}\left(\mathbf{P}\dot{\mathbf{R}}_{(i)}\right) - \mathbf{y}^T \mathbf{P}\dot{\mathbf{R}}_{(i)} \mathbf{Py} \end{bmatrix} \\
&= -\frac{1}{2} \begin{bmatrix} \text{tr}\left(\mathbf{G}_g^{-1} \dot{\mathbf{G}}_{g(i)}\right) - \text{tr}\left(\mathbf{C}^{gg} \mathbf{G}_g^{-1} \dot{\mathbf{G}}_{g(i)} \mathbf{G}_g^{-1}\right) - \mathbf{g}^T \mathbf{G}_g^{-1} \dot{\mathbf{G}}_{g(i)} \mathbf{G}_g^{-1} \mathbf{g} \\ \text{tr}\left(\mathbf{G}_s^{-1} \dot{\mathbf{G}}_{s(i)}\right) - \text{tr}\left(\mathbf{C}^{ss} \mathbf{G}_s^{-1} \dot{\mathbf{G}}_{s(i)} \mathbf{G}_s^{-1}\right) - \mathbf{s}^T \mathbf{G}_s^{-1} \dot{\mathbf{G}}_{s(i)} \mathbf{G}_s^{-1} \mathbf{s} \\ \text{tr}\left(\mathbf{G}_p^{-1} \dot{\mathbf{G}}_{p(i)}\right) - \text{tr}\left(\mathbf{C}^{pp} \mathbf{G}_p^{-1} \dot{\mathbf{G}}_{p(i)} \mathbf{G}_p^{-1}\right) - \mathbf{p}^T \mathbf{G}_p^{-1} \dot{\mathbf{G}}_{p(i)} \mathbf{G}_p^{-1} \mathbf{p} \\ \text{tr}\left(\mathbf{G}_b^{-1} \dot{\mathbf{G}}_{b(i)}\right) - \text{tr}\left(\mathbf{C}^{bb} \mathbf{G}_b^{-1} \dot{\mathbf{G}}_{b(i)} \mathbf{G}_b^{-1}\right) - \mathbf{b}^T \mathbf{G}_b^{-1} \dot{\mathbf{G}}_{b(i)} \mathbf{G}_b^{-1} \mathbf{b} \\ \text{tr}\left(\mathbf{R}^{-1} \dot{\mathbf{R}}_{(i)}\right) - \text{tr}\left(\mathbf{C}^{-1} \mathbf{W}^T \mathbf{R}^{-1} \dot{\mathbf{R}}_{(i)} \mathbf{R}^{-1} \mathbf{W}\right) - \mathbf{e}^T \mathbf{R}^{-1} \dot{\mathbf{R}}_{(i)} \mathbf{R}^{-1} \mathbf{e} \end{bmatrix} \quad [9]
\end{aligned}$$

em que:  $U(\boldsymbol{\kappa})$  é o vetor de escores associado às derivadas de primeira ordem da Eq. [8];  $\boldsymbol{\kappa}$  o vetor de parâmetros das estruturas de (co)variância associadas aos efeitos aleatórios do modelo apresentado na Eq. [1];  $\dot{\mathbf{G}}_{g(i)}$ ,  $\dot{\mathbf{G}}_{s(i)}$ ,  $\dot{\mathbf{G}}_{p(i)}$ ,  $\dot{\mathbf{G}}_{b(i)}$  e  $\dot{\mathbf{R}}_{(i)}$  são as derivadas de primeira ordem segundo a notação de Knight (2008) em relação aos parâmetros das estruturas de (co)variância associadas aos efeitos aleatórios  $\mathbf{g}$ ,  $\mathbf{s}$ ,  $\mathbf{p}$ ,  $\mathbf{b}$  e  $\mathbf{e}$ , respectivamente;  $\mathbf{C}^{gg}$ ,  $\mathbf{C}^{ss}$ ,  $\mathbf{C}^{pp}$  e  $\mathbf{C}^{bb}$  são partições de  $\mathbf{C}^{-1}$  associadas aos efeitos aleatórios  $\mathbf{g}$ ,  $\mathbf{s}$ ,  $\mathbf{p}$ , e  $\mathbf{b}$ , respectivamente.

Os zeros das derivadas apresentadas acima (Eq. [9]) podem ser aproximados via algoritmos iterativos baseados nas primeiras e segundas derivadas da função de log-verossimilhança residual (Eq. [8]) (Patterson e Thompson, 1971; Searle et al., 1992).

### 2.2.1. ALGORITMOS E ESTRATÉGIAS COMPUTACIONAIS

Nas próximas seções serão descritos os detalhes da implementação dos algoritmos EM (Dempster et al., 1977), AI (Gilmour et al., 1995; Johnson e Thompson, 1995) e EM-AI (Knight, 2008; Marçal et al., 2019) para estimar as matrizes de (co)variância associadas aos efeitos do modelo apresentado na Eq. [1] ( $\mathbf{G}_g$ ,  $\mathbf{G}_s$ ,  $\mathbf{G}_p$ ,  $\mathbf{G}_b$  e  $\mathbf{R}$ ).

Cabe ressaltar que a utilização da técnica de matrizes esparsas reduz drasticamente o esforço computacional do processo de estimação de matrizes de (co)variância via função de verossimilhança residual (Masuda et al., 2014). Portanto, utilizou-se esta técnica em um exemplo prático (Seção 2.3. Aplicação a Dados Experimentais), que será apresentado posteriormente, por meio do *software* R (R Core Team, 2019) utilizando-se as funcionalidades do pacote ‘Matrix’ (Bates and Maechler, 2019).

A estimação de componentes de variância via REML envolve algoritmos iterativos que iniciam o processo iterativo com os valores iniciais para os componentes de variância fornecidos pelo usuário. Por outro lado, o processo iterativo não encerra automaticamente, portanto, algum critério de convergência deve ser adotado para verificar se o algoritmo alcançou um ponto suficientemente próximo ao ponto de máximo da função de verossimilhança residual. Assim, o critério de convergência adotado foi uma combinação dos critérios utilizados por Meyer (2008) e Gilmour et al. (2009). Para o algoritmo EM, considerou-se que a convergência é atingida quando a variação em  $\log(L)$  for inferior a  $10^{-6}$  e a variação em cada um dos parâmetros for inferior a 1%. Já para o algoritmo AI, considerou-se que a convergência é atingida quando a variação em  $\log(L)$  for inferior a 0,002 e a variação em cada um dos parâmetros for inferior a 1%.

Os valores iniciais dos parâmetros das estruturas de (co)variância associadas aos efeitos aleatórios da Eq. [1], com exceção das correlações, podem ser estabelecidos com base na

variância amostral do vetor  $\mathbf{y}$  (Butler et al., 2009). Assim, podem-se assumir os respectivos valores para os componentes de variância, componentes de covariância e correlações como:  $0,10\sigma_y^2$ ,  $0,15\sigma_y^2$  e  $0,10$ , em que  $\sigma_y^2$  equivale à metade da variância amostral do vetor  $\mathbf{y}$ . Por outro lado, os valores iniciais dos componentes de variância residuais em cada ensaio podem assumir a metade da variância amostral de  $\mathbf{y}$  em cada ensaio (Butler et al., 2009).

### 2.2.2. ALGORITMO EM

Admitindo-se a forma mais geral para as estruturas de (co)variância associadas aos efeitos genéticos aleatórios da Eq. [1] ( $\mathbf{G}_g$ ,  $\mathbf{G}_s$  e  $\mathbf{G}_p$ ), tem-se que para cada uma delas são possíveis  $t(t + 1)/2$  parâmetros, em que  $t$  é o número de ensaios (Smith et al., 2001). Por outro lado, assumindo-se heterogeneidade de variâncias para as estruturas de (co)variância associadas aos efeitos aleatórios não genéticos ( $\mathbf{G}_b$  e  $\mathbf{R}$ ), tem-se que para cada uma delas são possíveis  $t$  parâmetros. Neste contexto, o algoritmo EM pode ser obtido pelo simples rearranjo das expressões apresentadas na Eq. [9] (Knight, 2008) ou de acordo com Meyer (2008).

A seguir são apresentadas as equações iterativas para a estimação dos parâmetros das matrizes de (co)variância associadas aos efeitos aleatórios da Eq. [1] descritas acima.

$$EM \Rightarrow \left[ \begin{array}{l} \mathbf{G}_{0g} \left\{ \begin{array}{l} \hat{\sigma}_{g(i)}^{2(k+1)} = \tilde{\mathbf{g}}^{T(k)} \dot{\mathbf{G}}_{g(i)} \tilde{\mathbf{g}}^{(k)} + tr \left( \mathbf{C}^{gg(k)} \dot{\mathbf{G}}_{g(i)} \right) / N_g \\ \hat{\sigma}_{g(i)}^{(k+1)} = \tilde{\mathbf{g}}^{T(k)} \dot{\mathbf{G}}_{g(i)} \tilde{\mathbf{g}}^{(k)} + tr \left( \mathbf{C}^{gg(k)} \dot{\mathbf{G}}_{g(i)} \right) / 2N_g \end{array} \right. \\ \\ \mathbf{G}_{0s} \left\{ \begin{array}{l} \hat{\sigma}_{s(i)}^{2(k+1)} = \tilde{\mathbf{s}}^{T(k)} \dot{\mathbf{G}}_{s(i)} \tilde{\mathbf{s}}^{(k)} + tr \left( \mathbf{C}^{ss(k)} \dot{\mathbf{G}}_{s(i)} \right) / N_s \\ \hat{\sigma}_{s(i)}^{(k+1)} = \tilde{\mathbf{s}}^{T(k)} \dot{\mathbf{G}}_{s(i)} \tilde{\mathbf{s}}^{(k)} + tr \left( \mathbf{C}^{ss(k)} \dot{\mathbf{G}}_{s(i)} \right) / 2N_s \end{array} \right. \\ \\ \mathbf{G}_{0p} \left\{ \begin{array}{l} \hat{\sigma}_{p(i)}^{2(k+1)} = \tilde{\mathbf{p}}^{T(k)} \dot{\mathbf{G}}_{p(i)} \tilde{\mathbf{p}}^{(k)} + tr \left( \mathbf{C}^{pp(k)} \dot{\mathbf{G}}_{p(i)} \right) / N_p \\ \hat{\sigma}_{p(i)}^{(k+1)} = \tilde{\mathbf{p}}^{T(k)} \dot{\mathbf{G}}_{p(i)} \tilde{\mathbf{p}}^{(k)} + tr \left( \mathbf{C}^{pp(k)} \dot{\mathbf{G}}_{p(i)} \right) / 2N_p \end{array} \right. \\ \\ \mathbf{G}_{0b} \left\{ \begin{array}{l} \hat{\sigma}_{b(i)}^{2(k+1)} = \tilde{\mathbf{b}}^{T(k)} \dot{\mathbf{G}}_{b(i)} \tilde{\mathbf{b}}^{(k)} + tr \left( \mathbf{C}^{bb(k)} \dot{\mathbf{G}}_{b(i)} \right) / N_b \\ \hat{\sigma}_{b(i)}^{(k+1)} = 0 \end{array} \right. \\ \\ \mathbf{R}_0 \left\{ \begin{array}{l} \hat{\sigma}_{e(i)}^{2(k+1)} = \tilde{\mathbf{e}}^{T(k)} \dot{\mathbf{R}}_{(i)} \tilde{\mathbf{e}}^{(k)} + tr \left( \mathbf{W} \mathbf{C}^{-1(k)} \mathbf{W}^T \dot{\mathbf{R}}_{(i)} \right) / N \\ \hat{\sigma}_{e(i)}^{(k+1)} = 0 \end{array} \right. \end{array} \right. \quad [10]$$

em que  $N$ ,  $N_g$ ,  $N_s$ ,  $N_p$  e  $N_b$  correspondem ao número total de observações e número de efeitos de CGC, CEC, linhagens parentais e de blocos, respectivamente.

### 2.2.3. ALGORITMO AI

O algoritmo AI baseia-se em um esquema iterativo (Eq. [11]) que envolve as derivadas de primeira e segunda ordem da Eq. [8] (Johnson e Thompson, 1995). Este esquema é similar ao utilizado originalmente por Patterson e Thompson (1971), porém, a matriz de informação de Fisher é substituída pela matriz de informação média (Gilmour et al., 1995; Johnson e Thompson, 1995).

Por meio da Eq. [11], adaptada de Knight (2008), os parâmetros das estruturas de (co)variância associadas aos efeitos aleatórios da Eq. [1], apresentadas na seção anterior, podem ser estimados de forma iterativa.

$$\boldsymbol{\kappa}^{(k+1)} = \boldsymbol{\kappa}^{(k)} + \lambda^{(k)} \nabla_{AI}^{-1}(\boldsymbol{\kappa}^{(k)}) U(\boldsymbol{\kappa}^{(k)}) \quad [11]$$

em que  $\boldsymbol{\kappa}^{(k+1)}$  é o vetor de parâmetros na iteração  $k + 1$ ,  $\boldsymbol{\kappa}^{(k)}$  é o vetor de parâmetros na iteração  $k$ ,  $\lambda^{(k)}$  é o peso para restringir a solução dos parâmetros ao espaço paramétrico na iteração  $k$ , podendo variar de 0,10 a 1,00 (Knight, 2008),  $\nabla_{AI}^{-1}(\boldsymbol{\kappa}^{(k)})$  é a matriz AI na iteração  $k$  e  $U(\boldsymbol{\kappa}^{(k)})$  é o vetor de escores associados às derivadas de primeira ordem da Eq. [8] na iteração  $k$ .

Os diferenciais generalizados da Eq. [5] para a obtenção da matriz AI foram fornecidos por Gilmour et al. (1995). A matriz de informação média pode ser obtida por meio da Eq. [12] como mostrado a seguir (Smith et al., 2005; Knight, 2008):

$$\nabla_{AI}(\boldsymbol{\kappa}^{(k)}) = \frac{1}{2} \mathbf{Q}^{T(k)} \mathbf{P}^{(k)} \mathbf{Q}^{(k)} \quad [12]$$

em que  $\mathbf{Q}$  é a matriz de ‘*working variables*’ associada aos parâmetros do vetor de parâmetros  $\boldsymbol{\kappa}$ .

A matriz  $\mathbf{Q}$ , referente ao modelo dialélico apresentado na Eq. [1], pode ser obtida por meio da Eq. [13] apresentada a seguir:

$$\mathbf{Q}^{(k)} = [\mathbf{q}_g^{(k)}, \mathbf{q}_s^{(k)}, \mathbf{q}_p^{(k)}, \mathbf{q}_b^{(k)}, \mathbf{q}_e^{(k)}] \Rightarrow \begin{cases} \mathbf{q}_{g(i)}^{(k)} = \mathbf{Z}_g \dot{\mathbf{G}}_{g(i)} \mathbf{G}_g^{-1(k)} \tilde{\mathbf{g}}^{(k)} \\ \mathbf{q}_{s(i)}^{(k)} = \mathbf{Z}_s \dot{\mathbf{G}}_{s(i)} \mathbf{G}_s^{-1(k)} \tilde{\mathbf{s}}^{(k)} \\ \mathbf{q}_{p(i)}^{(k)} = \mathbf{Z}_p \dot{\mathbf{G}}_{p(i)} \mathbf{G}_p^{-1(k)} \tilde{\mathbf{p}}^{(k)} \\ \mathbf{q}_{b(i)}^{(k)} = \mathbf{Z}_b \dot{\mathbf{G}}_{b(i)} \mathbf{G}_b^{-1(k)} \tilde{\mathbf{b}}^{(k)} \\ \mathbf{q}_{e(i)}^{(k)} = \dot{\mathbf{R}}_{(i)} \mathbf{R}^{-1(k)} \tilde{\mathbf{e}}^{(k)} \end{cases} \quad [13]$$

em que  $\mathbf{q}_g^{(k)}$ ,  $\mathbf{q}_s^{(k)}$ ,  $\mathbf{q}_p^{(k)}$ ,  $\mathbf{q}_b^{(k)}$  e  $\mathbf{q}_e^{(k)}$  são as matrizes contendo as ‘*working variables*’ para os parâmetros das estruturas de (co)variância associadas aos efeitos aleatórios da Eq. [1] na iteração  $k$ . Note que, o número de colunas de cada matriz  $\mathbf{q}$  é equivalente ao número de parâmetros estimados na respectiva matriz de (co)variância.

#### 2.2.4. ALGORITMO EM-AI

O algoritmo EM-AI foi obtido pela combinação dos algoritmos EM e AI. Geralmente esta combinação de algoritmos [EM-*type* (EM ou PXEM) e AI] é realizada na seguinte sequência: inicia-se o processo iterativo com o primeiro algoritmo (EM-*type*) e finaliza com o algoritmo AI (Meyer, 2006; Meyer, 2008; Marçal et al., 2019). Nesta abordagem, o número de passos EM deve ser pré-estabelecido (Bates e Sarkar, 2005; Marçal et al., 2019). Entretanto, neste trabalho, utilizou-se um esquema iterativo similar ao descrito por Knigth (2008) que alterna os dois algoritmos, tomando-se como referência o p-valor associado à estatística  $S^{(k)}$  (Cullis et al., 2004). O p-valor é obtido a partir da distribuição de qui-quadrado ( $\chi^2$ ) com  $N_{\kappa}$  graus de liberdade (sendo  $N_{\kappa}$  o número de parâmetros no vetor  $\kappa$ ) (Cullis et al., 2004; Knigth, 2008).

$$S^{(k)} = U(\kappa^{(k)})^T \nabla_{\text{AI}}^{-1}(\kappa^{(k)}) U(\kappa^{(k)}) \quad [14]$$

em que  $S^{(k)}$  é a estatística do teste na iteração  $k$ .

O esquema iterativo inicia-se com o algoritmo AI a cada iteração. Caso o p-valor associado à estatística  $S^{(k)}$  seja inferior a um valor crítico pré-definido, os parâmetros são atualizados pelo algoritmo EM. Knigth (2008) denominou este esquema iterativo de *score criterion* e sugeriu tomar o valor de 0,50 como valor crítico para o p-valor da estatística  $S^{(k)}$ .

### 2.3. APLICAÇÃO A DADOS EXPERIMENTAIS

Para a realização deste trabalho foram utilizadas populações oriundas de cruzamentos biparentais entre dois grupos de linhagens de feijão-comum, em esquema de dialelo parcial. O grupo 1 foi composto por cinco linhagens de grão preto (L1-L20, L2-Xamego, L3-BRS Expedito, L4-BRS Valente e L5-Diamante Negro) e o grupo 2 por sete linhagens de grão carioca (L6-BRSMG Uai, L7-BRS Estilo, L8-VC12, L9-VC20, L10-CNFC10720, L11-MAI1813 e L12-VC16). Além do tipo dos grãos, os dois grupos são divergentes quanto à arquitetura de plantas, resistência a doenças e produtividade de grãos (Moura et al., 2018).

Foram conduzidos cinco ensaios distribuídos em dois locais (Coimbra-MG e Viçosa-MG), três anos (2012, 2013 e 2014), duas safras (Seca e Inverno) e três gerações (F<sub>1</sub>, F<sub>2</sub> e F<sub>3</sub>) em blocos casualizados com três repetições (Tabela 1).

Tabela 1. Descrição dos ensaios para análise dialélica em MEGT<sup>1</sup>

Ensaio	Ambientes			Gerações	Parcelas	Médias (kg ha <sup>-1</sup> )
	Local	Ano	Safra			
1	Coimbra	2012	Inverno	F <sub>1</sub>	3 linhas x 1 metro	2965,63
2	Coimbra	2013	Seca	F <sub>1</sub>	3 linhas x 1 metro	4541,73
3	Coimbra	2013	Seca	F <sub>2</sub>	4 linhas x 4 metros	3879,74
4	Viçosa	2013	Seca	F <sub>2</sub>	4 linhas x 4 metros	3061,54
5	Coimbra	2014	Inverno	F <sub>3</sub>	4 linhas x 4 metros	3699,32

<sup>1</sup>MEGT, *multi-environment and multi-generation trials*.

Em cada ensaio foram avaliadas 35 populações e 12 linhagens parentais. As parcelas nos ensaios um e dois foram constituídas por três linhas de um metro (m), devido à menor disponibilidade de sementes na geração F<sub>1</sub>. Em contrapartida, as parcelas nos ensaios três, quatro e cinco foram constituídas por quatro linhas de quatro metros, devido a maior disponibilidade de

sementes nas gerações F<sub>2</sub> e F<sub>3</sub> e à necessidade de maior número de plantas para representar as populações em estudo. A densidade de plantio (15 sementes m<sup>-1</sup>) e o espaçamento entre linhas (0,5 m) foram os mesmos nos cinco ensaios.

A variável avaliada nos ensaios descritos anteriormente foi a produtividade de grãos (PG) em kg ha<sup>-1</sup>.

### 2.3.1. MODELO PARA A ANÁLISE DO DIALELO PARCIAL EM MEGT

A análise para o esquema dialelo parcial em MEGT foi realizada com base no modelo linear misto apresentado na Eq. [15]. Neste modelo, o efeito de CGC apresentado na Eq. [1] é particionado nos efeitos de CGC<sub>1</sub> (**g**<sub>1</sub>) e CGC<sub>2</sub> (**g**<sub>2</sub>) para os grupos de linhagens parentais *p* e *q*, respectivamente (Geraldi e Miranda Filho, 1988; Bernardo, 2010).

$$\mathbf{y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{Z}_{g_1}\mathbf{g}_1 + \mathbf{Z}_{g_2}\mathbf{g}_2 + \mathbf{Z}_s\mathbf{s} + \mathbf{Z}_p\mathbf{p} + \mathbf{e} \quad [15]$$

em que: **y** o vetor de observações fenotípicas das populações e linhagens parentais em todos os ensaios; **τ** é o vetor de efeitos fixos de ensaios, repetições dentro de ensaios, média das populações dentro de cada ensaio e média das linhagens parentais dentro de cada ensaio; **g**<sub>1</sub> é o vetor de efeitos aleatórios da capacidade geral de combinação das linhagens do grupo 1 (CGC<sub>1</sub>) em cada ensaio; **g**<sub>2</sub> é o vetor de efeitos aleatórios da capacidade geral de combinação das linhagens do grupo 2 (CGC<sub>2</sub>) em cada ensaio; **s** é o vetor de efeitos aleatórios da capacidade específica de combinação (CEC) das linhagens do grupo 1 com as linhagens do grupo 2 em cada ensaio; **p** é o vetor de efeitos aleatórios das linhagens parentais em cada ensaio; **e** é o vetor de erros aleatórios em cada ensaio e **X**, **Z**<sub>g<sub>1</sub></sub>, **Z**<sub>g<sub>2</sub></sub>, **Z**<sub>s</sub> e **Z**<sub>p</sub> são as matrizes de incidência dos efeitos **τ**, **g**, **s**, **d**, **r** e **p**, respectivamente.

O vetor de soluções  $\boldsymbol{\beta} = (\hat{\boldsymbol{\tau}}, \tilde{\mathbf{g}}_1, \tilde{\mathbf{g}}_2, \tilde{\mathbf{s}}, \tilde{\mathbf{p}})^T$  para os efeitos fixos e aleatórios do modelo descrito na Eq. [15] pode ser obtido pela Eq. [16]. A Eq. [16], apresentada a seguir, pode ser obtida maximizando-se a função de distribuição conjunta de  $\mathbf{y}$  e  $\mathbf{u} = (\mathbf{g}_1, \mathbf{g}_2, \mathbf{s}, \mathbf{p})^T$  (Henderson et al., 1959):

$$\begin{bmatrix} \mathbf{X}^T \mathbf{R}^{-1} \mathbf{X} & \mathbf{X}^T \mathbf{R}^{-1} \mathbf{Z}_{g_1} & \mathbf{X}^T \mathbf{R}^{-1} \mathbf{Z}_{g_2} & \mathbf{X}^T \mathbf{R}^{-1} \mathbf{Z}_s & \mathbf{X}^T \mathbf{R}^{-1} \mathbf{Z}_p \\ \mathbf{Z}_{g_1}^T \mathbf{R}^{-1} \mathbf{X} & \mathbf{Z}_{g_1}^T \mathbf{R}^{-1} \mathbf{Z}_{g_1} + \mathbf{G}_{g_1}^{-1} & \mathbf{Z}_{g_1}^T \mathbf{R}^{-1} \mathbf{Z}_{g_2} & \mathbf{Z}_{g_1}^T \mathbf{R}^{-1} \mathbf{Z}_s & \mathbf{Z}_{g_1}^T \mathbf{R}^{-1} \mathbf{Z}_p \\ \mathbf{Z}_{g_2}^T \mathbf{R}^{-1} \mathbf{X} & \mathbf{Z}_{g_2}^T \mathbf{R}^{-1} \mathbf{Z}_{g_1} & \mathbf{Z}_{g_2}^T \mathbf{R}^{-1} \mathbf{Z}_{g_2} + \mathbf{G}_{g_2}^{-1} & \mathbf{Z}_{g_2}^T \mathbf{R}^{-1} \mathbf{Z}_s & \mathbf{Z}_{g_2}^T \mathbf{R}^{-1} \mathbf{Z}_p \\ \mathbf{Z}_s^T \mathbf{R}^{-1} \mathbf{X} & \mathbf{Z}_s^T \mathbf{R}^{-1} \mathbf{Z}_{g_1} & \mathbf{Z}_s^T \mathbf{R}^{-1} \mathbf{Z}_{g_2} & \mathbf{Z}_s^T \mathbf{R}^{-1} \mathbf{Z}_s + \mathbf{G}_s^{-1} & \mathbf{Z}_s^T \mathbf{R}^{-1} \mathbf{Z}_p \\ \mathbf{Z}_p^T \mathbf{R}^{-1} \mathbf{X} & \mathbf{Z}_p^T \mathbf{R}^{-1} \mathbf{Z}_{g_1} & \mathbf{Z}_p^T \mathbf{R}^{-1} \mathbf{Z}_{g_2} & \mathbf{Z}_p^T \mathbf{R}^{-1} \mathbf{Z}_s & \mathbf{Z}_p^T \mathbf{R}^{-1} \mathbf{Z}_p + \mathbf{G}_p^{-1} \end{bmatrix} \times \begin{bmatrix} \hat{\boldsymbol{\tau}} \\ \tilde{\mathbf{g}}_1 \\ \tilde{\mathbf{g}}_2 \\ \tilde{\mathbf{s}} \\ \tilde{\mathbf{p}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T \mathbf{R}^{-1} \mathbf{y} \\ \mathbf{Z}_{g_1}^T \mathbf{R}^{-1} \mathbf{y} \\ \mathbf{Z}_{g_2}^T \mathbf{R}^{-1} \mathbf{y} \\ \mathbf{Z}_s^T \mathbf{R}^{-1} \mathbf{y} \\ \mathbf{Z}_p^T \mathbf{R}^{-1} \mathbf{y} \end{bmatrix} \quad [16]$$

Assumiu-se a seguinte estrutura de (co)variância para a Eq. [15] (Tabela 2):

$$\begin{bmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \\ \mathbf{s} \\ \mathbf{p} \\ \mathbf{e} \end{bmatrix} \sim N \left( \mathbf{0}, \begin{bmatrix} \mathbf{G}_{g_1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_{g_2} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{G}_s & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{G}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{R} \end{bmatrix} \right) \quad [17]$$

$$\mathbf{V} = \mathbf{Z}_{g_1} \mathbf{G}_{g_1} \mathbf{Z}_{g_1}^T + \mathbf{Z}_{g_2} \mathbf{G}_{g_2} \mathbf{Z}_{g_2}^T + \mathbf{Z}_s \mathbf{G}_s \mathbf{Z}_s^T + \mathbf{Z}_p \mathbf{G}_p \mathbf{Z}_p^T + \mathbf{R} \quad [18]$$

em que:  $\mathbf{G}_{g_1}$  é a matriz de (co)variância dos efeitos aleatórios de CGC<sub>1</sub>, com  $\mathbf{G}_{g_1} = \mathbf{I}_{g_1} \otimes \mathbf{G}_{0g_1}$ ;  $\mathbf{G}_{g_2}$  é a matriz de (co)variância dos efeitos aleatórios de CGC<sub>2</sub>, com  $\mathbf{G}_{g_2} = \mathbf{I}_{g_2} \otimes \mathbf{G}_{0g_2}$ ;  $\mathbf{G}_s$  é a matriz de (co)variância dos efeitos aleatórios de CEC, com  $\mathbf{G}_s = \mathbf{I}_s \otimes \mathbf{G}_{0s}$ ;  $\mathbf{G}_p$  é a matriz de (co)variância dos efeitos aleatórios das linhagens parentais, com  $\mathbf{G}_p = \mathbf{I}_p \otimes \mathbf{G}_{0p}$ ;  $\mathbf{R}$  é a matriz de

(co)variância dos erros aleatórios, com  $\mathbf{R} = \sigma_e^2 \mathbf{I}_n$  ou  $\mathbf{R} = \bigoplus_{t=1}^{N_t} \mathbf{I}_{n_t} \sigma_{e_t}^2$  (Tabela 2),  $\mathbf{V}$  é a matriz de (co)variância de  $\mathbf{y}$ ,  $\mathbf{I}_{g_1}$ ,  $\mathbf{I}_{g_2}$ ,  $\mathbf{I}_s$ ,  $\mathbf{I}_p$  e  $\mathbf{I}_n$  são matrizes identidade associadas às matrizes (co)variância dos efeitos aleatórios  $\mathbf{g}_1$ ,  $\mathbf{g}_2$ ,  $\mathbf{s}$ ,  $\mathbf{p}$  e  $\mathbf{e}$ , respectivamente. Cabe ressaltar que os detalhes necessários para a estimação das matrizes de (co)variância associadas aos efeitos aleatórios da Eq. [15] via algoritmos EM, AI e EM-AI, são apresentados na seção 7.1. do Material Suplementar.

Na Tabela 2 são mostradas as estruturas de (co)variância testadas para o modelo apresentado na Eq. [15]. As diferentes combinações das estruturas de (co)variância possibilitaram quatro variações do modelo apresentado na Eq. [15]. Para maiores detalhes a respeito das estruturas de (co)variância apresentadas na Tabela 2 consulte o Material Suplementar (seção 7.2.).

Tabela 2. Resumo das estruturas de (co)variância ajustadas aos efeitos aleatórios descritos na equação Eq. [1]

Estrutura de (co)variância	Modelos <sup>†</sup>			
	1	2	3	4
$\mathbf{G}_{g_1}$	DIAG	DIAG	CORH	US
$\mathbf{G}_{g_2}$	DIAG	DIAG	CORH	US
$\mathbf{G}_s$	DIAG	DIAG	CORH	US
$\mathbf{G}_p$	DIAG	DIAG	CORH	US
$\mathbf{R}$	VC	DIAGH	DIAGH	DIAGH

<sup>†</sup>DIAG, *diagonal*; VC, *variance component*; CORH, *uniform correlation and variances heterogeneous*; US, *unstructured*; DIAGH, *diagonal heterogeneous*.

### 2.3.2. AVALIAÇÃO DOS ALGORITMOS PARA ANÁLISE DO DIALELO PARCIAL EM MEGT

Os modelos descritos na Tabela 2 foram ajustados ao conjunto de dados MEGT, considerando-se a característica produtividade de grãos, utilizando-se os algoritmos EM, AI e EM-AI. As análises foram realizadas por meio de um *laptop* com sistema operacional Windows de 64 bit, processador Intel Core i5 CPU de 2.5-GHz e 4-GB memória RAM. O tempo médio de execução de cada análise (baseado em cinco repetições) foi registrado pelo pacote *rbenchmark* (Kusnierczyk, 2012). Além disso, também se registrou o  $\log(L)$  do ponto crítico da função de verossimilhança residual e a taxa empírica de convergência (Eq. [19]) como sugerido por Knigh (2008).

$$r_e = \lim_{\rightarrow +\infty} \frac{\sqrt{\sum_{i=1}^{N_k} [\kappa_i^{(k+1)} - \kappa_i^{(k)}]^2}}{\sqrt{\sum_{i=1}^{N_k} [\kappa_i^{(k)} - \kappa_i^{(k-1)}]^2}} \quad [19]$$

Todas as análises foram realizadas por meio do software R (R Core Team, 2019).

### 2.3.3. ESCOLHA DO MELHOR MODELO PARA A ANÁLISE DO DIALELO PARCIAL EM MEGT

A comparação dos modelos testados foi realizada por meio do Critério de Informação de Akaike (AIC), no qual, o modelo que melhor se ajusta aos dados é aquele que apresenta menor valor de AIC (Akaike, 1974).

$$AIC = -2\log(L) + 2p \quad [20]$$

em que:  $\log(L)$  é o logaritmo do ponto de máximo da função de verossimilhança residual e  $p$  é o número de parâmetros estimados pelo modelo.

Além do critério AIC, também foi utilizado o teste de razão de verossimilhanças (LRT) para testar as hipóteses de homogeneidade das variâncias residuais (modelo 1 vs modelo 2), nulidade de (co)variâncias entre ensaios (modelo 2 vs modelo 3) e homogeneidade de (co)variâncias entre ensaios (modelo 3 vs modelo 4) via distribuição de qui-quadrado ( $\chi^2$ ) com  $\nu$  graus de liberdade, onde  $\nu$  é a diferença entre o número de parâmetros dos modelos comparados.

#### **2.3.4. INFORMAÇÕES GENÉTICAS DA ANÁLISE DO DIALELO PARCIAL EM MEGT**

A partir do melhor modelo (Tabela 2), foram estimados os componentes de variância de  $CGC_1$  ( $\sigma_{g_1}^2$ ),  $CGC_2$  ( $\sigma_{g_2}^2$ ) e  $CEC$  ( $\sigma_s^2$ ) para cada ensaio. Fazendo uma analogia ao dialelo de meia tabela, os componentes de variância associados à capacidade geral de combinação ( $\sigma_{g_1}^2$  e  $\sigma_{g_2}^2$ ) podem ser considerados de natureza aditiva, enquanto o componente de variância associado à capacidade específica de combinação ( $\sigma_s^2$ ) pode ser atribuído aos desvios de dominância, assumindo-se ausência de epistasia (Griffing, 1956a; Griffing, 1956b; Torres e Geraldi, 2007).

Em posse dos componentes de variância apresentados anteriormente pode-se obter a importância relativa dos efeitos aditivos ( $\theta$ ), envolvidos no controle genético do caráter produtividade de grãos avaliado neste estudo. O estimador de  $\theta$  utilizado no presente trabalho foi reportado por Oliveira et al. (2019) adaptado de Torres e Geraldi (2007), modificado de Baker (1978).

$$\theta = \frac{\sigma_{g_1}^2 + \sigma_{g_2}^2}{\sigma_{g_1}^2 + \sigma_{g_2}^2 + \sigma_s^2} \quad [21]$$

### 2.3.5. POTENCIAL DAS POPULAÇÕES SEGREGANTES VIA DIALELO EM MEGT

Considerando a característica produtividade de grãos, as populações segregantes de maior potencial para o uso nos programas de melhoramento de plantas autógamas são aquelas que combinam alta CEC com pelo menos um dos genitores, envolvido no cruzamento, de alta CGC em que o dialelo foi avaliado em um único ensaio. Neste cenário, diz-se que a população apresenta alta frequência de alelos favoráveis e diversidade genética. Entretanto, considerando os efeitos de CGC e CEC, preditos a partir de análise dialélica em MEGT, a seleção de populações segregantes pode ser dificultada em razão da interação G×A.

Visando capitalizar a interação G×A nos efeitos de CGC ( $\mathbf{g}_1$  e  $\mathbf{g}_2$ ) e CEC ( $\mathbf{s}$ ) foram utilizados índices baseados na distância genótipos-ideótipo adaptados de Rocha et al. (2018c) apresentados a seguir (Eq. [22], Eq. [23] e Eq. [24]):

$$I_{g_{1(i)}} = \frac{1/\sum_j^t (\tilde{\mathbf{g}}_{1(ij)} - v_j)^2}{\sum_{i=1}^{N_{g_1}} \left[ 1/\sum_j^t (\tilde{\mathbf{g}}_{1(ij)} - v_j)^2 \right]} \quad [22]$$

$$I_{g_{2(i)}} = \frac{1/\sum_j^t (\tilde{\mathbf{g}}_{2(ij)} - v_j)^2}{\sum_{i=1}^{N_{g_2}} \left[ 1/\sum_j^t (\tilde{\mathbf{g}}_{2(ij)} - v_j)^2 \right]} \quad [23]$$

$$I_{s(i)} = \frac{1/\sum_j^t (\tilde{\mathbf{s}}_{(ij)} - v_j)^2}{\sum_{i=1}^{N_s} \left[ 1/\sum_j^t (\tilde{\mathbf{s}}_{(ij)} - v_j)^2 \right]} \quad [24]$$

em que  $I_{g_{1(i)}}$ ,  $I_{g_{2(i)}}$  e  $I_{s(i)}$  correspondem à probabilidade espacial (Rocha et al., 2018b), associada ao índice genótipos-ideótipo, para cada genitor do grupo 1, grupo 2 e população;  $\tilde{\mathbf{g}}_{1(ij)}$ ,  $\tilde{\mathbf{g}}_{2(ij)}$  e  $\tilde{\mathbf{s}}_{(ij)}$

são os BLUPs da CGC<sub>1</sub>, CGC<sub>2</sub> e CEC em cada ensaio, respectivamente;  $\nu_j$  é o ideótipo para cada ensaio;  $t$  é o número de ensaios e  $\Sigma$  é o operador de somatório.

Neste trabalho, o ideótipo para a característica produtividade de grãos foi definido como o máximo em cada ensaio para os índices  $I_{g_{1(i)}}$ ,  $I_{g_{2(i)}}$  e  $I_{s(i)}$ .

### 3. RESULTADOS

#### 3.1. EFICIÊNCIA COMPUTACIONAL PARA A ANÁLISE DO DIALELO PARCIAL EM MEGT

O algoritmo EM convergiu para os modelos 1, 2 e 4 (Tabela 3). Em contrapartida, o algoritmo AI falhou em convergir para os modelos 1 e 4. Além disso, foi detectado singularidade na matriz de informação média ( $\nabla_{AI}$ ) logo na primeira iteração para a estimação das estruturas de (co)variância associadas aos efeitos aleatórios do modelo 4, o que impossibilitou a continuidade do processo iterativo. Cabe ressaltar, que a matriz  $\nabla_{AI}$  associada ao modelo 4 não atendeu a propriedade de ser positiva definida e apresentou singularidade mesmo com as soluções REML fornecidas pelo algoritmo EM, portanto, o uso do algoritmo EM-AI também foi inviabilizado. Por outro lado, o algoritmo EM-AI não foi utilizado no modelo 1, pois o p-valor (0,07) da estatística  $S^{(k)}$  (Eq. [14]), utilizando as soluções do algoritmo EM, foi inferior ao nível crítico estabelecido (0,50) para a utilização do algoritmo AI no processo iterativo do esquema *score criterion* (Tabela 3). A análise com o modelo 3 foi realizada apenas pelo algoritmo AI, pois, não foram obtidas expressões explícitas, rearranjando-se as derivadas de primeira ordem da estrutura de (co)variância CORH, para a estimação dos parâmetros da estrutura CORH.

Tabela 3. Logaritmo do ponto de máximo da função de verossimilhança residual [ $\log(L)$ ], taxa empírica de convergência ( $r_e$ ), tempo médio de execução (Tempo), em segundos, número de iterações (Iterações) e relação entre o tempo médio de execução e número de iterações (Tempo/Iteração) das análises considerando os modelos 1, 2, 3 e 4 via os algoritmos EM, AI e EM-AI

Algoritmos	Variáveis	Modelos <sup>†</sup>			
		1	2	3	4
EM	$\log(L)$	-4695,49	-4646,63		-4626,89
	$r_e$	1,00	1,00		1,00
	Tempo (s)	2507,82	4143,13		5887,63
	Iterações	23168	23117		27273
	Tempo/Iteração (s)	0,11	0,18		0,22
AI	$\log(L)$		-4646,62	-4637,79	
	$r_e$		0,52	0,56	
	Tempo (s)	$\nabla_{AI(Singular)}$	5,62	7,36	$\nabla_{AI(Singular)}$
	Iterações		12	13	
	Tempo/Iterações (s)		0,47	0,57	
EM-AI	$\log(L)$		-4646,62		
	$r_e$		0,50		
	Tempo (s)		5,89		$\nabla_{AI(Singular)}$
	Iterações <sup>‡</sup>		3/11		
	Tempo/Iterações (s)		0,42		
	p-valor	0,07	0,66		

<sup>†</sup>Modelos descritos na Tabela 2, <sup>‡</sup>Número de iterações EM/Número de iterações AI.

O algoritmo EM apresentou  $r_e$  próxima de 1,00 e número de iterações superior a 20000 nos três modelos ajustados (Tabela 3). Em contrapartida, observou-se  $r_e$  próxima de 0,50 e número de iterações igual ou inferior a 14 (3/11) para os modelos 2 e 3 ajustados via algoritmos AI ou EM-AI.

O tempo médio de execução (Tempo) variou de 2507,82 a 5887,63 segundos enquanto o tempo médio de execução por iteração (Tempo/Iteração) variou de 0,11 a 0,22 segundos para os três modelos ajustados via algoritmo EM (Tabela 3). Além disso, verificou-se incremento do tempo médio de execução (Tempo) bem como da relação Tempo/Iteração com o aumento da complexidade dos modelos (1, 2 e 4) ajustados via algoritmo EM, isto é, com o aumento do

número de parâmetros (Tabela 3, Material Suplementar – seção 7.2.). De maneira similar, o algoritmo AI apresentou maiores tempo médio de execução e relação Tempo/Iteração para o modelo 3 quando comparado com o modelo 2.

Considerando o modelos 2, os algoritmos AI e EM-AI apresentaram menor tempo médio de execução que o algoritmo EM, embora, a relação Tempo/Iteração deste último algoritmo tenha sido inferior à dos demais algoritmos (Tabela 3). O algoritmo AI apresentou Tempo de execução superior em 1,74 segundos superior para o modelo 3 quando comparado ao modelo 2.

### 3.2. ESCOLHA DO MODELO PARA A ANÁLISE DO DIALELO PARCIAL EM MEGT

O número de parâmetros variou entre 21 (modelo 1) e 65 (modelo 4) com incremento do  $\log(L)$  para modelos com maior número de parâmetros (Tabela 4). Entretanto, analisando o AIC, verificou-se que o modelo de melhor ajuste foi o modelo 3, devido ao menor valor de AIC (Tabela 4). O modelo selecionado (modelo 3) apresenta correlação uniforme ao longo dos ensaios para os efeitos genéticos e variâncias heterogêneas tanto para os efeitos genéticos quanto não genético (Tabela 2, Material Suplementar – seção 7.2.).

Tabela 4. Resumo dos modelos ajustados para os dados MEGT: número de parâmetros (NP), logaritmo do ponto de máximo da função de verossimilhança residual [ $\log(L)$ ], teste de razão de verossimilhança (LRT) e critério de informação de Akaike (AIC)

Modelos	NP	$\log(L)$	LRT <sup>†</sup>	AIC
1	21	-4695,49 <sup>‡</sup>		9432,99
2	25	-4646,62	97,74 (p < 0,01)	9343,25
3	29	-4637,79	17,66 (p < 0,01)	9333,58
4	65	-4626,89 <sup>‡</sup>	21,80 (p = 0,97)	9383,78

<sup>†</sup>Comparação de modelos via LRT na seguinte sequencia: modelo 2 vs modelo 1, modelo 3 vs modelo 2 e modelo 4 vs modelo 3. <sup>‡</sup>Informações obtidas a partir do algoritmo EM (Tabela 3).

Considerando os valores de LRT (Tabela 4), verificou-se que esta escolha (modelo 3 no caso) pode ainda ser reforçada pela rejeição das hipóteses de homogeneidade de variâncias residuais (modelo 2 vs modelo 1: 97,74 ( $p < 0,01$ )), nulidade de covariância entre ensaios (modelo 3 vs modelo 2: 17,66 ( $p < 0,01$ )) e não rejeição da hipótese de homogeneidade de covariâncias entre ensaios (modelo 4 vs modelo 3: 21,80 ( $p = 0,97$ )).

### 3.3. ANÁLISE DO DIALELO PARCIAL EM MEGT

A partir do modelo 3 foram obtidas as estimativas de variância e de correlação para os ensaios (Tabela 5). Dentre os componentes de variância apresentados, pode-se verificar que dois deles se aproximaram de zero ( $\hat{\sigma}_{g_1}^2$  no ensaio 1 e  $\hat{\sigma}_s^2$  no ensaio 5) e podem ser considerados *boundary* (próximos de zero). Além disso, é notória a heterogeneidade de variâncias para os efeitos genéticos ( $\mathbf{g}_1$ ,  $\mathbf{g}_2$ ,  $\mathbf{s}$ ,  $\mathbf{p}$ ) e não genético ( $\mathbf{e}$ ). Cabe ressaltar que a relação entre a maior e a menor variância residual superou o valor 7 ( $457217,50/60964,89 = 7,50$ ).

Tabela 5. Estimativas de parâmetros obtidas a partir do algoritmo AI para a produtividade de grãos em feijoeiro, considerando o modelo 3 para a análise dialélica em MEGT

Estimativas de parâmetros <sup>†</sup>	Ensaio					$\hat{\rho}^\ddagger$
	1	2	3	4	5	
$\hat{\sigma}_{g_1}^2$	0,00	56921,95	6550,51	40778,76	22385,89	0,23
$\hat{\sigma}_{g_2}^2$	448667,90	19408,96	25708,22	115950,90	5126,24	0,64
$\hat{\sigma}_s^2$	48904,95	1267,98	3674,24	95491,69	0,00	0,76
$\hat{\sigma}_p^2$	343553,10	50753,27	133573,80	135610,40	24029,84	0,42
$\hat{\sigma}_e^2$	221749,60	282658,90	60964,89	457217,50	204374,50	
$\theta$	0,90	0,98	0,90	0,62	1,00	

<sup>†</sup>  $\hat{\sigma}_{g_1}^2$ , componente de variância da CGC<sub>1</sub>;  $\hat{\sigma}_{g_2}^2$ , componente de variância da CGC<sub>2</sub>;  $\hat{\sigma}_s^2$ , componente de variância da CEC;  $\hat{\sigma}_p^2$ , componente de variância de linhagens parentais;  $\hat{\sigma}_e^2$ , componente de variância do resíduo;  $\theta$ , importância relativa dos efeitos aditivos no controle do caráter produtividade de grãos.

‡As correlações obtidas são indexadas aos respectivos efeitos aleatórios  $\mathbf{g}_1$ ,  $\mathbf{g}_2$ ,  $\mathbf{s}$ ,  $\mathbf{p}$  como mostrado a seguir:  $\hat{\rho}_{g_1}$ , correlação uniforme associada ao efeito de CGC<sub>1</sub>;  $\hat{\rho}_{g_2}$ , correlação uniforme associada ao efeito de CGC<sub>2</sub>;  $\hat{\rho}_s$ , correlação uniforme associada ao efeito de CEC;  $\hat{\rho}_p$ , correlação uniforme associada ao efeito de linhagens parentais.

Os valores de correlação estimados, com base no modelo 3 que admite correlação uniforme, foram todos positivos e variaram de 0,23 a 0,76. Dentre as correlações, somente aquelas associadas aos efeitos genéticos de CGC<sub>1</sub> e de linhagens parentais apresentaram valor inferior a 0,50. Cabe ressaltar que a estatística  $\theta$  foi superior a 0,50 para todos os ensaios.

A partir das estimativas dos parâmetros do modelo 3 foram preditos os efeitos de CGC das linhagens parentais (CGC<sub>1</sub> e CGC<sub>2</sub>) bem como os efeitos de CEC das combinações híbridas por meio da expressão Eq. [16] (Tabelas 6 e 7). Cabe ressaltar que os efeitos preditos de CGC<sub>1</sub> no ensaio 1 e CEC no ensaio 5 foram próximos de zero.

Tabela 6. BLUPs dos efeitos de CGC para a produtividade de grãos em feijoeiro, considerando o modelo 3 para a análise dialélica em MEGT

Linhagens parentais <sup>†</sup>	Ensaio				
	1	2	3	4	5
L1	0,00	-52,04	36,82	-225,81	-178,30
L2	0,00	-327,29	-68,73	-81,44	-19,61
L3	0,00	91,40	41,87	64,95	137,78
L4	0,00	7,34	-69,05	176,13	94,36
L5	0,00	280,59	59,09	66,17	-34,22
L6	-903,94	-146,18	-216,90	-503,87	-63,46
L7	1069,75	93,75	81,01	436,91	94,70
L8	-686,94	-57,58	46,40	-68,18	-32,85
L9	46,35	-122,73	-164,69	-199,03	-44,77
L10	242,89	38,86	154,76	-44,40	3,66
L11	28,00	30,64	-28,92	131,25	0,50
L12	203,89	163,24	128,35	247,32	42,21

†As linhagens parentais de 1 a 5 pertencem ao grupo 1 (CGC<sub>1</sub>) enquanto as linhagens parentais de 6 a 12 pertencem ao grupo 2 (CGC<sub>2</sub>).

Tabela 7. BLUPs dos efeitos de CEC para a produtividade de grãos em feijoeiro, considerando o modelo 3 para a análise dialélica em MEGT

Combinações híbridas	Ensaio				
	1	2	3	4	5
L1xL6	69,10	8,33	17,67	68,58	0,00
L1xL7	69,29	9,10	13,68	88,35	0,00
L1xL8	-104,51	-8,94	-7,91	-145,56	0,00
L1xL9	-59,11	-12,62	-24,63	-141,23	0,00
L1xL10	-290,59	-36,50	-63,52	-304,57	0,00
L1xL11	-107,89	-19,10	-30,78	-202,37	0,00
L1xL12	54,51	8,59	18,73	69,08	0,00
L2xL6	-57,57	-9,31	-16,40	-110,98	0,00
L2xL7	-120,98	-17,09	-29,87	-168,63	0,00
L2xL8	-26,10	-7,37	-4,77	-79,64	0,00
L2xL9	300,52	37,24	58,55	336,47	0,00
L2xL10	35,98	2,13	4,65	-31,65	0,00
L2xL11	-19,96	0,38	-3,77	73,75	0,00
L2xL12	219,44	28,40	46,27	289,75	0,00
L3xL6	-194,81	-28,32	-52,26	-247,37	0,00
L3xL7	-238,80	-35,06	-67,03	-318,83	0,00
L3xL8	102,25	19,61	34,23	243,21	0,00
L3xL9	-36,00	-6,31	-13,05	-62,39	0,00
L3xL10	14,81	1,78	11,76	-6,70	0,00
L3xL11	134,71	18,97	35,19	161,21	0,00
L3xL12	64,76	12,59	25,59	88,42	0,00
L4xL6	-82,66	-9,41	-14,56	-111,32	0,00
L4xL7	217,17	24,38	43,92	218,11	0,00
L4xL8	-102,69	-9,39	-17,69	-18,09	0,00
L4xL9	-84,33	-11,58	-17,46	-62,93	0,00
L4xL10	238,59	40,48	71,76	453,24	0,00
L4xL11	-87,85	-10,97	-27,88	-94,59	0,00
L4xL12	-141,48	-24,91	-53,72	-280,74	0,00
L5xL6	15,27	0,60	-3,70	-3,14	0,00
L5xL7	218,11	35,27	63,54	399,46	0,00
L5xL8	140,15	12,50	17,35	83,73	0,00
L5xL9	-135,59	-15,92	-22,93	-138,92	0,00
L5xL10	-30,35	-13,63	-23,95	-236,06	0,00
L5xL11	168,91	26,03	47,87	243,88	0,00
L5xL12	-142,30	-9,98	-14,91	-51,57	0,00

### 3.3.1. POTENCIAL DAS POPULAÇÕES SEGREGANTES PARA A EXTRAÇÃO DE LINHAGENS

Considerando os índices de seleção para CGC ( $I_{g_1}$  e  $I_{g_2}$ ), verificou-se que as linhagens L5, L3 e L4 apresentaram melhor desempenho para a CGC<sub>1</sub>, enquanto as linhagens L7, L12 e L10 foram as que se destacaram para CGC<sub>2</sub> (Figura 1). Por outro lado, das 35 populações oriundas dos cruzamentos realizados entre as linhagens do grupo 1 com as linhagens do grupo 2, apenas 13 se destacaram pelo índice  $I_s$ , apresentando estimativas positivas de CEC em todos os ensaios (L4xL10, L5xL7, L2xL9, L2xL12, L5xL11, L4xL7, L3xL8, L3xL11, L5xL8, L1xL7, L3xL12, L1xL6 e L1xL12) (Figura 2). Além disso, cabe ressaltar que os ensaios 1 e 4 se destacaram com desvios positivos de maior magnitude para os efeitos de CEC nas 13 populações selecionadas.

Das 13 populações selecionadas utilizando-se o índice ( $I_s$ ) para CEC, 11 se destacam (L4xL10, L5xL7, L2xL12, L5xL11, L4xL7, L3xL8, L3xL11, L5xL8, L1xL7, L3xL12 e L1xL12), pois além de possuírem complementaridade (alta CEC), também possuem, em sua genealogia, pelo menos um dos genitores com alta CGC, ou seja, genitores com alta frequência de alelos favoráveis.

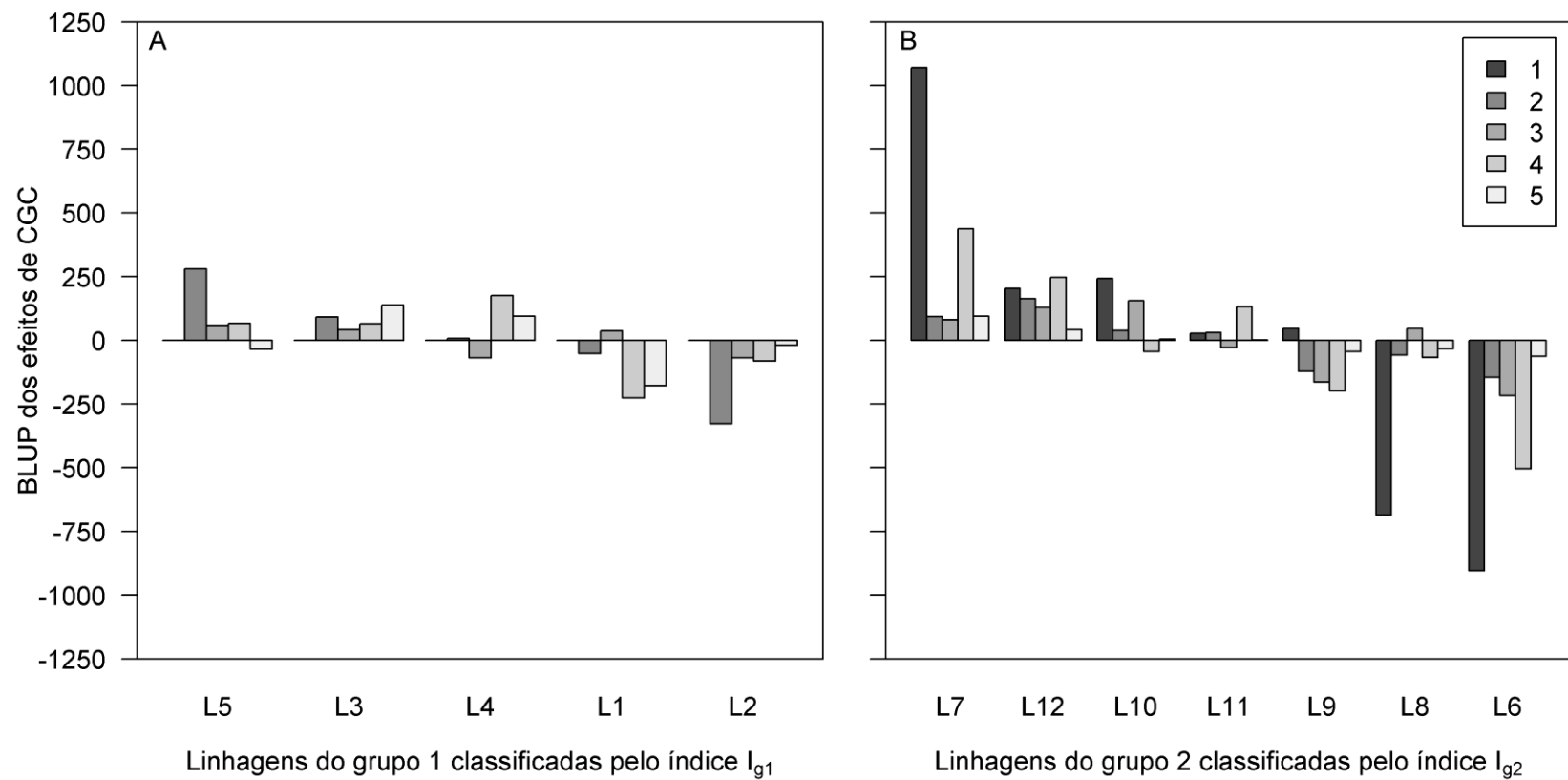


Figura 1. BLUPs de  $CGC_1$  (A) e  $CGC_2$  (B) para a característica produtividade de grãos nos ensaios 1, 2, 3, 4 e 5 das linhagens do grupo 1 e 2 classificadas pelos índices  $I_{g1}$  e  $I_{g2}$ , respectivamente.

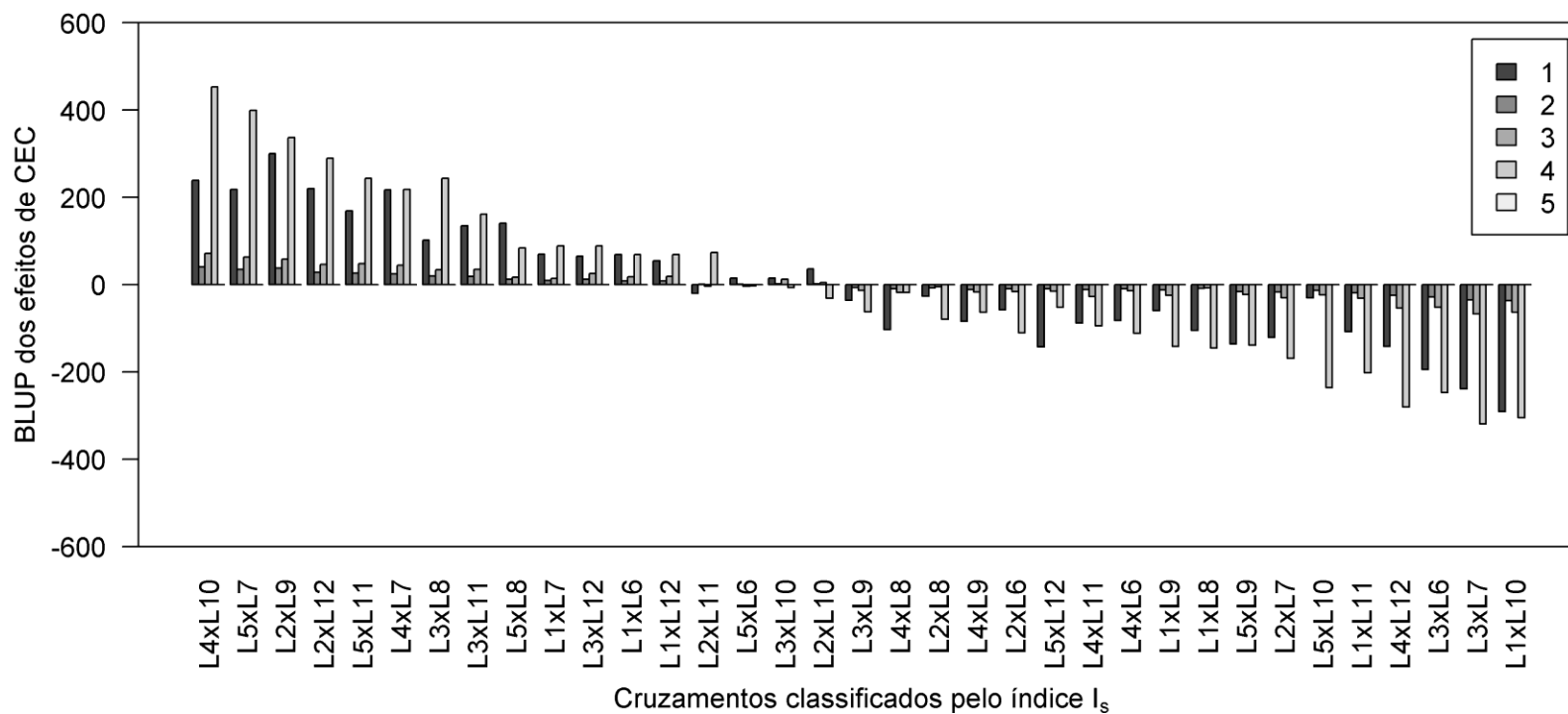


Figura 2. BLUP de CEC para a característica produtividade de grãos nos ensaios 1, 2, 3, 4 e 5 das populações oriundas dos cruzamentos entre linhagens do grupo 1 com linhagens do grupo 2 classificadas pelo índice  $I_s$ .

## 4. DISCUSSÃO

### 4.1. EFICIÊNCIA COMPUTACIONAL PARA A ANÁLISE DO DIALELO PARCIAL EM MEGT

Embora o algoritmo EM não garanta a convergência, esta era esperada para os modelos analisados, devido à propriedade de aumento monótono do  $\log(L)$  em direção ao ponto de máxima verossimilhança (Meyer, 2006; Knight, 2008; Meyer, 2008). Por outro lado, a falha do algoritmo AI em convergir para alguns dos modelos analisados também era um resultado esperado. Na literatura, este problema foi reportado em algumas oportunidades (Meyer, 2006; Knight, 2008; Diffey et al., 2017; Marçal et al., 2019). Diferente do algoritmo EM, o algoritmo AI não garante aumento do  $\log(L)$  a cada iteração (Meyer, 2006; Diffey et al., 2017). Além disso, o algoritmo AI pode atualizar parâmetros fora do espaço paramétrico bem como matrizes de (co)variância e/ou matriz  $\nabla_{AI}$  não positivas definidas (Jensen et al., 1997; Knight, 2008; Meyer, 2008; Diffey et al., 2017; Marçal et al., 2019; Gilmour et al., 2019). Estes problemas podem levar à falha de convergência do algoritmo AI. Meyer (2006) reportou diversas situações nas quais o algoritmo AI pode falhar em convergir. Entretanto, *bad start points* para os parâmetros a serem estimados e a complexidade dos modelos podem ser consideradas as principais causas de falha de convergência (Knight, 2008; Misztal, 2008).

A impossibilidade do uso do algoritmo EM-AI nas análises dos modelos 1 e 4 pode ser atribuída ao p-valor inferior ao nível crítico pré-estabelecido (p-valor < 0,50), o que impede a atualização de parâmetros pelo algoritmo AI, e pela violação da pressuposição associada à matriz  $\nabla_{AI}$  (positiva definida). Knight (2008) sugeriu o valor crítico de 0,50 para um esquema robusto de estimação de parâmetros via REML pelos algoritmos EM e AI usando o *score criterion*. Além disso, a mesma autora relata que quando a matriz  $\nabla_{AI}$  não é positiva definida a atualização de

parâmetros pode ser ilegal (fora do espaço paramétrico) ou em uma direção diferente do máximo global da função de verossimilhança residual. Neste sentido, Misztal (2008) fez algumas recomendações, visando à obtenção de estimativas realistas dos componentes de variância quando se utiliza o algoritmo AI, como refinar o modelo adicionando ou removendo efeitos não genéticos.

A lentidão para a convergência do algoritmo EM, devido ao elevado número de iterações, pode ser justificada pela elevada taxa empírica de convergência observada. Resultado similar foi reportado por Knigth (2008) na estimação de parâmetros da estrutura de (co)variância US para dados MET. O elevado número de iterações do algoritmo EM para atingir a convergência em análises envolvendo estruturas de (co)variância complexas não é raro e foi relatado na literatura em algumas oportunidades (Jensen et al., 1997; Knigth, 2008; Meyer, 2008). Neste cenário, o custo computacional da análise inviabiliza o uso deste algoritmo para grandes conjuntos de dados e modelos com grande número de parâmetros a serem estimados (Knigth, 2008; Gilmour et al., 2019). Em contrapartida, o algoritmo AI apresentou menor taxa empírica de convergência e convergiu rapidamente para os modelos 2 e 3. Segundo Thompson (2008) e Gilmour et al. (2019), essa é uma característica dos algoritmos baseados nas derivadas de segunda ordem da função de verossimilhança residual, como o algoritmo AI. Além disso, esse algoritmo tem sido recomendado para análises envolvendo estruturas de (co)variância complexas e para grandes conjuntos de dados devido a sua maior eficiência computacional (Gilmour et al., 1995; Jensen et al., 1997; Thompson et al., 2005; Thompson, 2008; Gilmour et al., 2019) e facilidade de derivação para modelos mais gerais.

O menor custo computacional por iteração do algoritmo EM em relação ao AI também foi reportado por Gilmour et al. (1995). Entretanto, este benefício é suplantado pelo elevado número

de iterações do algoritmo EM. Neste sentido, Meyer (1987) reparametrizou o algoritmo EM para um modelo linear misto específico e reduziu, em cerca de 214%, o número de iterações. Contudo, a reparametrização do algoritmo EM para modelos mais gerais pode ser complexa, portanto, deve-se, primeiramente, lançar mão do algoritmo AI ou da combinação entre os algoritmos EM e AI.

#### **4.2. ESCOLHA DO MODELO PARA A ANÁLISE DO DIALELO PARCIAL EM MEGT**

Do ponto vista estatístico, diferentes critérios como LRT, critério de informação de Akaike (AIC) e critério de informação Bayesiano (BIC) podem ser utilizados na escolha de modelos. Estes critérios são baseados na qualidade de ajuste e na parcimônia dos modelos testados. Thompson et al. (2003) utilizam o LRT para justificar a escolha do modelo fator analítico (FA) em detrimento ao modelo padrão ( $G + G \times A$ ) na modelagem de dados MET. Enquanto Kelly et al. (2007), trabalhando com dados MET de diferentes culturas, optou pelo critério AIC na escolha do melhor modelo. Por outro lado, Smith et al. (2015), trabalhando com dados MET do ensaio nacional de variedades de trigo da Austrália, relataram que tanto o LRT quanto o AIC levam a escolha de modelos mais parametrizados que o BIC. Contudo, Ogut et al. (2014) avaliaram diferentes estruturas de (co)variância para os efeitos de CGC e CEC na análise dialélica de dados MET de *Pinus taeda* L. e utilizaram o critério AIC na escolha do melhor modelo. Assim, no presente trabalho, optou-se pelo mesmo critério adotado por Ogut et al. (2014) para a análise dialélica dos dados MEGT.

O modelo selecionado (modelo 3) demonstra a homogeneidade de correlações entre ensaios para os efeitos genéticos do modelo dialélico, devido ao único parâmetro de correlação da estrutura CORH. Pádua et al. (2016), estudando a herança de fumonisina na cultura do milho,

também reportaram correlação uniforme para efeitos genéticos de CGC e CEC para um dialelo avaliado em MET.

A escolha do modelo 3 pode ser atribuída a dois fatores. O primeiro deles é a parcimônia, pois este modelo possui, aproximadamente, metade do número de parâmetros do modelo 4 (US). O segundo fator, está associado ao aumento modesto do  $\log(L)$  quando se compara o modelo 4 com o modelo 3. Assim, a combinação destes dois fatores favoreceram o modelo 3 em detrimento ao modelo 4. Smith et al. (2001) propuseram o uso da estrutura de (co)variância US para modelar o efeito genético em MET. Contudo, os mesmos autores relataram que esta estrutura pode ser ineficiente e instável quando se trabalha com múltiplos ambientes, devido ao elevado número de parâmetros associado a esta estrutura. Neste sentido, alternativas mais parcimoniosas são desejáveis.

A natureza heterogênea da estrutura de (co)variância dos resíduos foi verificada pelo LRT (Tabela 4). Pádua et al. (2014) e Ogut et al. (2014) também reportaram variâncias residuais heterogêneas trabalhando com dialelo em MET. Por outro lado, Rocha et al. (2019) reportaram a melhoria na qualidade de ajuste de um modelo linear misto estruturado com resíduos heterogêneos para ensaios de avaliação de progênies de feijão carioca em experimentos sequencias (com grau crescente de endogamia) em MEGT. Estes resultados corroboram a superioridade do procedimento REML em relação à ANOVA devido à flexibilidade para se modelar as estruturas de (co)variância.

#### **4.3. ANÁLISE DO DIALELO PARCIAL EM MEGT**

O algoritmo AI gerou estimativas próximas de zero (*boundary*) para dois componentes de variância. Isso ocorre porque o algoritmo AI não garante atualização de parâmetros dentro do

espaço paramétrico (Meyer, 2006; Knigh, 2008; Misztal, 2008; Diffey et al., 2017; Gilmour, 2009; Marçal et al., 2019). Assim, a implementação do algoritmo AI no presente trabalho utilizou as recomendações sugeridas por Gilmour (2019) para restringir as soluções dos parâmetros no espaço paramétrico.

A heterogeneidade de variâncias para os efeitos genéticos ( $\mathbf{g}_1$ ,  $\mathbf{g}_2$ ,  $\mathbf{s}$ ,  $\mathbf{p}$ ) pode ser atribuída à interação G×A caracterizada como expressão diferencial de genes frente às diferentes condições ambientais (Kelly et al., 2007; Zhang et al., 2015). Além da heterogeneidade de variâncias, a falta de correlação entre o desempenho dos genótipos ao longo dos ensaios também é atribuída à interação G×A. Neste sentido, Smith et al. (2001, 2005) e Smith et al. (2005) sugeriram considerar um determinado caráter avaliado em diferentes ambientes como caracteres diferentes. Neste cenário, segundo os mesmos autores, a interação G×A pode ser modelada de forma mais realista admitindo-se heterogeneidade de variâncias e correlação entre pares de ambientes. O valor superior a 0,50 das correlações (Tabela 5) associadas aos efeitos de CGA<sub>2</sub> e de CEC nos diferentes ensaios sugere que a interação G×A deve atuar predominantemente de maneira simples nestes efeitos. Segundo Van Eeuwijk (2016) ensaios com correlação genética alta tem menor proporção de interação complexa. Por outro lado, a correlação de baixa magnitude associada ao efeito de linhagens parentais indica interação G×A predominantemente complexa.

Heterogeneidade de variâncias residuais em MEGT, conforme observado neste trabalho, sugere que as condições ambientais distintas (efeito de anos, locais ou safras) também afetam a magnitude da variância ambiental (Tabela 5). Resultado similar foram reportados por Rocha et al. (2019) para ensaios de avaliação de progênies de feijão carioca em experimentos sequenciais (MEGT) e por Pádua et al. (2014) e Ogut et al. (2014) em análise dialélica de dados MET. Neste sentido, a relação entre as variâncias residuais de magnitude extrema pode violar a premissa de

homogeneidade de variâncias requerida quando se utiliza a abordagem da ANOVA para análise de dados em MEGT ou MET. Pimentel-Gomes (2009) sugere, como critério prático, que a relação entre as variâncias residuais de magnitude extrema não deve superar o valor 7 para que a análise conjunta possa ser realizada sem prejuízos em testes de hipóteses e estimativas de parâmetros. Estes resultados reforçam a relevância do método REML, devido a sua maior flexibilidade para se modelar as estruturas de (co)variância. Além disso, este método pode ser considerado uma generalização da ANOVA para situações mais complexas.

A estatística ( $\theta$ ) foi superior a 0,5 em todos os ensaios, indicando predomínio de efeitos aditivos no controle do caractere produtividade de grãos nas populações de feijão preto estudadas.

Os BLUPs, próximos de zero, observados para os efeitos de CGC<sub>1</sub> e CEC resultam dos componentes de variância *boundary* para os respectivos efeitos. Isso ocorre porque os efeitos aleatórios preditos pelo sistema de equações de modelos mistos são ponderados pela herdabilidade da característica associada aos referidos efeitos (Henderson et al., 1959; Searle et al., 1992).

#### **4.3.1. POTENCIAL DAS POPULAÇÕES SEGREGANTES PARA EXTRAÇÃO DE LINHAGENS**

A interação G×A tem sido uma das principais causas de dificuldade em todas as etapas de um programa de melhoramento, seja nas etapas de formação da população base, seleção e recomendação de novas cultivares. A interação G×A resulta da expressão diferencial de genes em função de variação ambiental (Zhang et al., 2015). Assim, esperam-se diferentes estimativas de parâmetros genéticos, bem como diferentes predições para os efeitos de CGC e CEC na análise

dialélica de dados MEGT. Neste contexto, a utilização dos índices baseados na distância genótipos-ideótipo permitiu capitalizar a interação G×A na escolha das populações segregantes de maior potencial, bem como simplificou a escolha das mesmas. Além disso, a seleção por meio destes índices capitaliza tanto a adaptabilidade quanto a estabilidade produtiva ao longo dos ambientes, pois a seleção leva em consideração tanto a magnitude dos valores genotípicos quanto as suas variações ao longo dos ambientes. Segundo Rocha et al. (2018c), esta abordagem leva em consideração a estabilidade e os altos valores genotípicos. Neste contexto, os ensaios multiambientes assumem importância essencial para o melhoramento de plantas, pois além de permitirem a quantificação da interação G×A também possuem papel fundamental na seleção de genótipos com boa adaptação e estabilidade (Smith et al., 2001; Smith et al., 2005; Zhang et al., 2015).

A utilização dos índices baseados na distância genótipos-ideótipo empregando-se os valores preditos para CGC<sub>1</sub>, CGC<sub>2</sub> e CEC, permitiram tratar a interação G×A de forma mais adequada em MEGT, pois além de capitalizar a interação G×A nos efeitos genéticos do dialelo, também permitiram identificar as populações que apresentaram alta frequência de alelos favoráveis e divergência genética. Estes índices ( $I_{g_1}$ ,  $I_{g_2}$  e  $I_s$ ) foram eficientes na identificação de 11 populações de feijoeiro com potencial para a extração de linhagens superiores.

A abordagem de modelos mistos para análise dialélica em MEGT associada aos índices genótipos-ideótipo ( $I_{g_1}$ ,  $I_{g_2}$  e  $I_s$ ) é promissora para mitigar as dificuldades relativas à interação nos programas de melhoramento de plantas autógamas.

## 5. CONCLUSÃO

Quando a convergência é atingida, ao se modelar dados MEGT, os algoritmos AI e EM-AI são computacionalmente mais eficientes em maximizar a função de verossimilhança residual que o algoritmo EM.

O algoritmo AI é mais apropriado para modelar dados em MEGT devido a sua maior generalidade na estimação de estruturas de (co)variância.

O modelo linear misto 3, que assume heterogeneidade de variâncias para os efeitos genéticos e não genético, bem como correlação uniforme para os efeitos genéticos do dialelo em MEGT, se mostrou eficiente para modelar dados MEGT em feijão comum.

Os índices baseados na distância genótipos-ideótipo se mostraram ferramentas poderosas na seleção de populações promissoras para o uso nos programas de melhoramento de plantas autógamas quando se trabalha com dialelo em MEGT.

## 6. REFERÊNCIAS

- Akaike, H. 1974. New look at the statistical model identification. *IEEE Trans. Automat. Contr.* AC-19:716–723. doi:10.1109/TAC.1974.1100705
- Baker, R.J. 1978. Issues in diallel analysis. *Crop Sci.* 18:533–536. doi:10.2135/cropsci1978.0011183X001800040001x
- Bates, D., and D. Sarkar. 2005. lme4: Linear mixed-effects models using Eigen and Eigenpack. R Package Version 0.95-1. Comprehensive R Arch. Network. <https://www.rdocumentation.org/packages/lme4/versions/0.95-1/topics/lmerControl>, accessed: August 2<sup>nd</sup>, 2019.

- Bates, D., and M. Maechler. 2019. Matrix: Sparse and dense matrix classes and methods. R Package Version 1.2-17. Comprehensive R Arch. Network. <https://cran.r-project.org/web/packages/Matrix/index.html>, accessed: August 2<sup>nd</sup>, 2019.
- Batista, L.G., R.S.R. Anjos, N.L. Poersch, R.S. Nalin, P.C.S. Carneiro, J.E.S. Carneiro, and M.D.V. Resende. 2017. Multigeneration index in the selection of common bean inbred families. *Crop Sci.* 57:2354–2360. doi:10.2135/cropsci2016.08.0665
- Bernardo, R. 2010. Breeding for quantitative traits in plants 2nd ed. Stemma Press, Woodbury, MN.
- Bhullar, G.S., K.S. Gill, and A.S. Khehra. 1979. Combining ability analysis over F<sub>1</sub>-F<sub>5</sub> generations in diallel crosses of bread wheat. *Theor. Appl. Genet.* 55:77–80. doi:10.1007/BF00285194
- Butler, D., B.R. Cullis, A.R. Gilmour, and B.J. Gogel. 2009. ASReml-R reference manual. Release 3.0. State Queensland, Dep. Prim. Ind. Fish., Brisbane, QLD, Australia.
- Cho, Y., and R.A. Scott. 2000. Combining ability of seed vigor and seed yield in soybean. *Euphytica* 112:145–150. doi:10.1023/A:10038279
- Cockerham, C.C., and B.S. Weir. 1977. Quadratic analyses of reciprocal crosses. *Biometrics* 33:187–203. doi:10.2307/2529312
- Cullis, B.R., A.B. Smith, and R. Thompson. 2004. Perspectives of ANOVA, REML and a general linear mixed model. In *Methods and Models in Statistics*. Imperial College Press, 53–94. doi:10.1142/9781860945410\_0004
- Dempster, A.P., N.M. Laird, and D.B. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B Methodol.* 39:1–38.

- Diffey, S.M., A.B. Smith, A.H. Welsh, and B.R. Cullis. 2017. A new REML (parameter expanded) EM algorithm for linear mixed models. *Aust. N. Z. J. Stat.* 59:433–448. doi:10.1111/anzs.12208
- Gardner, C.O., and S.A. Eberhart. 1966. Analysis and interpretation of the variety cross diallel and related populations. *Biometrics* 22:439–452. doi: 10.2307/2528181
- Geraldi, I.O., and J.B. Miranda Filho. 1988. Adapted models for the analysis of combining ability of varieties in partial diallel crosses. *Braz. J. Genet.* 11:419–430.
- Gilmour, A.R. 2019. Average information residual maximum likelihood in practice. *J. Anim. Breed. Genet.* 136:262–272. doi:10.1111/jbg.12398
- Gilmour, A.R., B.J. Gogel, B.R. Cullis, and R. Thompson. 2009. *ASReml user guide*. VSN Int., Hemel Hempstead, UK.
- Gilmour, A.R., R. Thompson, and B.R. Cullis. 1995. Average information REML: An efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics* 51:1440–1450. doi:10.2307/2533274
- Griffing, B. 1956a. A generalised treatment of the use of diallel crosses in quantitative inheritance. *Heredity* 10:31–50. doi:10.1038/hdy.1956.2
- Griffing, B. 1956b. Concept of general and specific combining ability in relation to diallel crossing systems. *Aust. J. Biol. Sci.* 9:463–493. doi:10.1071/BI9560463
- Hayman, B.I. 1954. The analysis of variance of diallel tables. *Biometrics* 10:235–244. doi:10.2307/3001877
- Hei, N., S. Hussein, and M. Laing. 2016. Heterosis and combining ability analysis of slow rusting stem rust resistance and yield and related traits in bread wheat. *Euphytica* 207:501–514. doi:10.1007/s10681-015-1526-9

- Henderson, C.R., O. Kempthorne, S.R. Searle, and C.M. Von Krosigk. 1959. The estimation of environmental and genetic trends from records subject to culling. *Biometrics* 15:192–218. doi:10.2307/2527669
- Huang, M., L.-Y. Chen, and Z.-Q. Chen. 2015. Diallel analysis of combining ability and heterosis for yield and yield components in rice by using positive loci. *Euphytica* 205:37–50. doi:10.1007/s10681-015-1381-8
- Jensen, J., E.A. Mantysaari, P. Madsen, and R. Thompson. 1997. Restricted maximum likelihood estimation of (co)variance components in multivariate mixed linear models using average information. *Ind. J. Agric. Stat.* 49:215–236.
- Johnson, D.L., and R. Thompson. 1995. Restricted maximum likelihood estimation of variance components for univariate animal models using sparse matrix techniques and average information. *J. Dairy Sci.* 78:449–456. doi:10.3168/jds.S0022-0302(95)76654-1
- Kaushik, P., M. Plazas, J. Prohens, S. Vilanova, and P. Gramazio. 2018. Diallel genetic analysis for multiple traits in eggplant and assessment of genetic distances for predicting hybrids performance. *Plos One* 13:e0199943. doi:10.1371/journal.pone.0199943
- Kelly, A.M., A.B. Smith, J.A. Eccleston, and B.R. Cullis. 2007. The accuracy of varietal selection using factor analytic models for multi-environment plant breeding trials. *Crop Sci.* 47:1063–1070. doi:10.2135/cropsci2006.08.0540
- Knight, E. 2008. Improved iterative schemes for REML estimation of variance parameters in linear mixed models. Ph.D. diss., Univ. Adelaide, Adelaide, SA, Australia.
- Kusnierczyk, W. 2012. rbenchmark: Benchmarking routine for R. R Package version 1.0.0. Comprehensive R Arch. Network. <https://cran.r-project.org/web/packages/rbenchmark/index.html>, accessed: August 2<sup>nd</sup>, 2019.

- Marçal, T.S., J.R.A.S.C. Rocha, F.V. Salvador, R.S.R. Anjos, A.C. Silva, P.C.S. Carneiro, and J.E.S. Carneiro. 2019. Estimation of variance for reciprocal general and specific combining ability effects by EM-AI algorithm. *Crop Sci.* 59:1494–1503. doi:10.2135/cropsci2018.09.0555
- Masuda, Y., T. Baba, and M. Suzuki. 2014. Application of supernodal sparse factorization and inversion to the estimation of (co)variance components by residual maximum likelihood. *J. Anim. Breed. Genet.* 131:227–236. doi:10.1111/jbg.1205825135
- Meyer, K. 1987. Restricted maximum likelihood to estimate variance components for mixed models with two random factors. *Genet. Sel. Evol.* 19:49–68. doi:10.1186/1297-9686-19-1-49
- Meyer, K. 2006.  $PX \times AI$ : Algorithmic for better convergence in restricted maximum likelihood estimation. Paper presented at the 8th World Congress on Genetics Applied to Livestock Production, Belo Horizonte, Brazil. 13–18 Aug. 2006. *Commun.* 24-00.
- Meyer, K. 2008. Parameter expansion for estimation of reduced rank covariance matrices. *Genet. Sel. Evol.* 40:3–24. doi:10.1051/gse:2007032
- Misztal, I. 2008. Reliable computing in estimation of variance components. *J. Anim. Breed. Genet.* 125:363–370. doi:10.1111/j.1439-0388.2008.00774.x
- Möhring, J., A.E. Melchinger, and H.P. Piepho. 2011. REML-based diallel analysis. *Crop Sci.* 51:470–478. doi:10.2135/cropsci2010.05.0272
- Moura, L.M., R.S.R. Anjos, R.O. Batista, N.M. Vale, C.D. Cruz, J.E.S. Carneiro, J.C. Machado, and P.C.S. Carneiro. 2018. Combining ability of common bean parents in different seasons, locations and generations. *Euphytica* 214:181–194. doi:10.1007/s10681-018-2259-3

- Ogut F., C. Maltecca, M.M. R. Whetten, S. McKeand, and F. Isik. 2014. Genetic analysis of diallel progeny test data using factor analytic linear mixed models. *For. Sci.* 60:119–127. doi:10.1007/s10681-016-1722-2v
- Oliveira, I.C.M., T.S. Marçal, K.C. Bernardino, P.C.O. Ribeiro, R.A.C. Parrella, P.C.S. Carneiro, R.E. Schaffert, and J.E.S. Carneiro. 2019. Combining ability of biomass sorghum lines for agroindustrial characters and multitrait selection of photosensitive hybrids for energy cogeneration. *Crop Sci.* 59:1554–1566. doi:10.2135/cropsci2018.11.0693
- Pádua, J.M.V., K.O.G. Dias, M.M. Pastina, J.C. Souza, V.A.V. Queiroz, R.V. Costa, M.B.P. Silva, C.A.G. Ribeiro, C.T. Guimarães, S.A. Gezan, and L.J.M. Guimarães. 2019. A multi-environment trials diallel analysis provides insights on the inheritance of fumonisin contamination resistance in tropical maize. *Euphytica* 211:277–285. doi:10.1007/s10681-016-1722-2
- Patil, V.D., and P.R. Chopde. 1981. Combining ability analysis over environments in diallel crosses of linseed (*Linum usitatissimum* L.). *Theor. Appl. Genet.* 60:339–343. doi:10.1007/BF00264325
- Patterson, H.D., and R. Thompson. 1971. Recovery of inter-block information when block sizes are unequal. *Biometrika* 58:545–554. doi:10.1093/biomet/58.3.545
- Pimentel-Gomes, F. 2009. Curso de estatística experimental 15ª ed. Fealq, Piracicaba.
- R Core Team. 2019. R: A language and environment for statistical computing. R Found. Stat. Comput., Vienna, Austria.
- Ramalho, M.P., Â.F.B. Abreu, J.B. Santos, and J.A.R. Nunes. 2012. Aplicações da genética quantitativa no melhoramento de plantas autógamas. Editora UFLA, Lavras.

- Resende, M.D.V., M.A.P. Ramalho, S.R. Guilherme, and A.F.B. Abreu. 2015. Multigeneration index in the within progenies bulk method for breeding of self-pollinated plants. *Crop Sci.* 55:1202–1211. doi:10.2135/cropsci2014.08.0580
- Rocha, G.A.F., F.A.C. Pereira, and N.A. Vello. 2018a. Potential of soybean crosses in early inbreeding generations for grain yield. *Crop Breed. Appl. Biotech.* 18:267–275. doi:10.1590/1984-70332018v18n3a40
- Rocha, J.R.A.S.C., J.C. Machado, and P.C.S. Carneiro. 2018b. Multitrait index based on factor analysis and ideotype-design: Proposal and application on elephant grass breeding for bioenergy. *Glob. Change Biol. Bioenergy* 10:52–60. doi:10.1111/gcbb.12443
- Rocha, J.R.A.S.C., K.V. Nunes, A.L.N. Carneiro, T.S. Marçal, F.V. Salvador, P.C.S. Carneiro, and J.E.S. Carneiro. 2019. Selection of superior inbred progenies toward the common bean ideotype. *Agron. J.* 111:1181–1189. doi:10.2134/agronj2018.12.0761
- Rocha, J.R.A.S.C., T.S. Marçal, F.V. Salvador, A.C. Silva, J.C. Machado, and P.C.S. Carneiro. 2018c. Genetic insights into elephantgrass persistence for bioenergy purpose. *Plos One* 13: e0203818. doi:10.1371/journal.pone.0203818
- Searle, S.R., G. Casella, and C.E. McCulloch. 1992. *Variance components*. John Wiley & Sons, New York.
- Smith, A.B., A. Ganesalingam, H. Kuchel, and B.R. Cullis. 2015. Factor analytic mixed models for the provision of grower information from national crop variety testing programs. *Theor. Appl. Genet.* 128:55–72. doi:10.1007/s00122-014-2412-x
- Smith, A.B., B.R. Cullis, and R. Thompson. 2001. Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. *Biometrics* 57:1138–1147. doi:10.1111/j.0006-341X.2001.01138.x

- Smith, A.B., B.R. Cullis, and R. Thompson. 2005. The analysis of crop cultivar breeding and evaluation trials: An overview of current mixed model approaches. *J. Agric. Sci.* 143:449–462. doi:10.1017/S0021859605005587
- Thompson, R. 2008. Estimation of quantitative genetic parameters. *Proc. Biol. Sci.* 275:679–686. doi:10.1098/rspb.2007.1417
- Thompson, R., B.R. Cullis, A.B. Smith, and A.R. Gilmour. 2003. A sparse implementation of the average information algorithm for factor analytic and reduced rank variance models. *Aust. N. Z. J. Stat.* 45:445–459. doi:10.1111/1467-842X.00297
- Thompson, R., S. Brotherstone, and I.M.S. White. 2005. Estimation of quantitative genetic parameters. *Philos. Trans. R. Soc., B* 360:1469–1477. doi:10.1098/rstb.2005.1676
- Torres, E.A., and I.O. Geraldi. 2007. Partial diallel analysis of agronomic characters in rice (*Oryza sativa* L.). *Genet. Mol. Biol.* 30:605–613. doi:10.1590/S1415-47572007000400018
- Van Eeuwijk, F.A., D.V. Bustos-Korts, and M. Malosetti. 2016. What should students in plant breeding know about the statistical aspects of genotype  $\times$  environment interactions? *Crop Sci.* 56:2119–2140. doi:10.2135/cropsci2015.06.0375
- Viana, J.M.S., and F.P. Matta. 2003. Analysis of general and specific combining abilities of popcorn populations, including selfed parents. *Genet. Mol. Biol.* 26:465–471. doi:10.1590/S1415-47572003000400010
- Zhang, X., P. Pérez-Rodríguez, K. Semagn, Y. Beyene, R. Babu, M.A. López-Cruz, F. San Vicente, M. Olsen, E. Buckler, J.-L. Jannink, B.M. Prasanna, and J. Crossa. 2015. Genomic prediction in biparental tropical maize populations in water-stressed and well-watered environments using low-density and GBS SNPs. *Heredity* 114:291–299. doi:10.1038/hdy.2014.99

Zhu, J., and B.S. Weir. 1994. Analysis of cytoplasmic and maternal effects I. A genetic model for diploids plant seeds and animals. *Theor. Appl. Genet.* 89:153–159. doi:10.1007/BF00225135

## 7. MATERIAL SUPLEMENTAR

### 7.1. ESTIMAÇÃO DAS MATRIZES DE (CO)VARIÂNCIA DO MODELO PARA ANÁLISE DO DIALELO PARCIAL EM MEGT

A estimação das matrizes de (co)variância, associadas aos efeitos aleatórios da Eq. [15] para os quatro modelos especificados na Tabela 2, foi realizada maximizando-se a função de verossimilhança residual. A seguir é apresentado o logaritmo da função de verossimilhança residual para a Eq. [15].

$$\begin{aligned}
 -2\log(L) = & \text{const.} + \log|\mathbf{R}| + N_{g_1} \log|\mathbf{G}_{0g_1}| \\
 & + N_{g_2} \log|\mathbf{G}_{0g_2}| + N_s \log|\mathbf{G}_{0s}| \\
 & + N_p \log|\mathbf{G}_{0p}| + \log|\mathbf{C}| + \mathbf{y}^T \mathbf{P} \mathbf{y}
 \end{aligned} \tag{S_1}$$

em que  $N$ ,  $N_{g_1}$ ,  $N_{g_2}$ ,  $N_s$  e  $N_p$  correspondem ao número total de observações e número de efeitos de CGC<sub>1</sub>, CGC<sub>2</sub>, CEC e de linhagens parentais, respectivamente;  $\log|\mathbf{C}|$  é o logaritmo do determinante do lado esquerdo ( $\mathbf{C}$ ) do sistema de equações de modelos mistos apresentado na Eq. [16];  $\mathbf{P}$  é o projetor ortogonal (Eq. [6]).

A partir da função de verossimilhança residual apresentada acima (Eq. [S\_1]) pode-se estimar as matrizes de (co)variância apresentadas acima ( $\mathbf{G}_{0g_1}$ ,  $\mathbf{G}_{0g_2}$ ,  $\mathbf{G}_{0s}$ ,  $\mathbf{G}_{0p}$  e  $\mathbf{R}$ ), encontrando-se os zeros das derivadas de primeira ordem associadas aos parâmetros de cada matriz. Cabe ressaltar que as derivadas para as matrizes de (co)variância apresentadas acima irão variar conforme a estrutura de (co)variância adotada para as mesmas. Assim, para maiores

detalhes a respeito da derivação das estruturas de (co)variância apresentadas na Tabela 2 consulte a seção 7.3.

As derivadas de primeira ordem para Eq.[S\_1] ou Eq. [5] são apresentadas a seguir:

$$U(\boldsymbol{\kappa}) = -\frac{1}{2} \begin{bmatrix} tr\left(\mathbf{G}_{g_1}^{-1} \dot{\mathbf{G}}_{g_1(i)}\right) - tr\left(\mathbf{C}^{g_1 g_1} \mathbf{G}_{g_1}^{-1} \dot{\mathbf{G}}_{g_1(i)} \mathbf{G}_{g_1}^{-1}\right) - \mathbf{g}_1^T \mathbf{G}_{g_1}^{-1} \dot{\mathbf{G}}_{g_1(i)} \mathbf{G}_{g_1}^{-1} \mathbf{g}_1 \\ tr\left(\mathbf{G}_{g_2}^{-1} \dot{\mathbf{G}}_{g_2(i)}\right) - tr\left(\mathbf{C}^{g_2 g_2} \mathbf{G}_{g_2}^{-1} \dot{\mathbf{G}}_{g_2(i)} \mathbf{G}_{g_2}^{-1}\right) - \mathbf{g}_2^T \mathbf{G}_{g_2}^{-1} \dot{\mathbf{G}}_{g_2(i)} \mathbf{G}_{g_2}^{-1} \mathbf{g}_2 \\ tr\left(\mathbf{G}_s^{-1} \dot{\mathbf{G}}_{s(i)}\right) - tr\left(\mathbf{C}^{ss} \mathbf{G}_s^{-1} \dot{\mathbf{G}}_{s(i)} \mathbf{G}_s^{-1}\right) - \mathbf{s}^T \mathbf{G}_s^{-1} \dot{\mathbf{G}}_{s(i)} \mathbf{G}_s^{-1} \mathbf{s} \\ tr\left(\mathbf{G}_p^{-1} \dot{\mathbf{G}}_{p(i)}\right) - tr\left(\mathbf{C}^{pp} \mathbf{G}_p^{-1} \dot{\mathbf{G}}_{p(i)} \mathbf{G}_p^{-1}\right) - \mathbf{p}^T \mathbf{G}_p^{-1} \dot{\mathbf{G}}_{p(i)} \mathbf{G}_p^{-1} \mathbf{p} \\ tr\left(\mathbf{R}^{-1} \dot{\mathbf{R}}_{(i)}\right) - tr\left(\mathbf{C}^{-1} \mathbf{W}^T \mathbf{R}^{-1} \dot{\mathbf{R}}_{(i)} \mathbf{R}^{-1} \mathbf{W}\right) - \mathbf{e}^T \mathbf{R}^{-1} \dot{\mathbf{R}}_{(i)} \mathbf{R}^{-1} \mathbf{e} \end{bmatrix} \quad [\text{S}_2]$$

em que:  $U(\boldsymbol{\kappa})$  é o vetor de escores associado as derivadas de primeira ordem da Eq. [S\_1];  $\boldsymbol{\kappa}$  o vetor de parâmetros das estruturas de (co)variância associadas aos efeitos aleatórios do modelo apresentado na Eq. [15];  $\dot{\mathbf{G}}_{g_1(i)}$ ,  $\dot{\mathbf{G}}_{g_2(i)}$ ,  $\dot{\mathbf{G}}_{s(i)}$ ,  $\dot{\mathbf{G}}_{p(i)}$  e  $\dot{\mathbf{R}}_{(i)}$  são as derivadas de primeira ordem (seção 7.3.) em relação aos parâmetros das matrizes de (co)variância associadas aos efeitos aleatórios  $\mathbf{g}_1$ ,  $\mathbf{g}_2$ ,  $\mathbf{s}$ ,  $\mathbf{p}$ , e  $\mathbf{e}$ , respectivamente (Tabela 2);  $\mathbf{C}^{g_1 g_1}$ ,  $\mathbf{C}^{g_2 g_2}$ ,  $\mathbf{C}^{ss}$ , e  $\mathbf{C}^{pp}$  são partições de  $\mathbf{C}^{-1}$  associadas aos efeitos aleatórios  $\mathbf{g}_1$ ,  $\mathbf{g}_2$ ,  $\mathbf{s}$ , e  $\mathbf{p}$ , respectivamente.

Os zeros das diferenciais apresentadas acima (Eq. [S\_2]) podem ser aproximados via algoritmo EM, AI ou EM-AI apresentados na seção Material e Métodos. A seguir serão apresentadas as equações iterativas do algoritmo EM bem como a matriz  $\mathbf{Q}$  do algoritmo AI para a Eq. [15], considerando os modelos especificados na Tabela 2.

### 7.1.1. ALGORITMO EM

O algoritmo EM pode ser obtido pelo simples rearranjo das expressões apresentadas em Eq. [S\_2]. A seguir são apresentados as equações iterativas para a estimação dos parâmetros das matrizes de (co)variância dos modelos 1, 2 e 4 apresentados na Tabela 2. Contudo, não foram encontradas equações explícitas para os parâmetros associados à estrutura de (co)variância CORH do modelo 3 rearranjando-se as expressões apresentadas na Eq. [S\_2], portanto, este modelo foi ajustado apenas pelo algoritmo AI. Na Eq. [S\_3] são apresentadas as equações iterativas para estimação dos parâmetros das matrizes de (co)variância do modelo 1. Estas equações divergem das equações iterativas para o modelo 2 apenas quanto ao estimador dos parâmetros da matriz  $\mathbf{R}_0$  apresentados na Eq. [S\_4].

$$EM_{M_1} \left\{ \begin{array}{l} \mathbf{G}_{0g_1} \left\{ \hat{\sigma}_{g_{1(i)}}^{2(k+1)} = \left[ \tilde{\mathbf{g}}_1^{T(k)} \dot{\mathbf{G}}_{g_{1(i)}} \tilde{\mathbf{g}}_1^{(k)} + tr \left( \mathbf{C}^{g_1 g_1(k)} \dot{\mathbf{G}}_{g_{1(i)}} \right) \right] / N_{g_1} \\ \mathbf{G}_{0g_2} \left\{ \hat{\sigma}_{g_{2(i)}}^{2(k+1)} = \left[ \tilde{\mathbf{g}}_2^{T(k)} \dot{\mathbf{G}}_{g_{2(i)}} \tilde{\mathbf{g}}_2^{(k)} + tr \left( \mathbf{C}^{g_2 g_2(k)} \dot{\mathbf{G}}_{g_{2(i)}} \right) \right] / N_{g_2} \\ \mathbf{G}_{0s} \left\{ \hat{\sigma}_{s_{(i)}}^{2(k+1)} = \left[ \tilde{\mathbf{s}}^{T(k)} \dot{\mathbf{G}}_{s_{(i)}} \tilde{\mathbf{s}}^{(k)} + tr \left( \mathbf{C}^{ss(k)} \dot{\mathbf{G}}_{s_{(i)}} \right) \right] / N_s \\ \mathbf{G}_{0p} \left\{ \hat{\sigma}_{p_{(i)}}^{2(k+1)} = \left[ \tilde{\mathbf{p}}^{T(k)} \dot{\mathbf{G}}_{p_{(i)}} \tilde{\mathbf{p}}^{(k)} + tr \left( \mathbf{C}^{pp(k)} \dot{\mathbf{G}}_{p_{(i)}} \right) \right] / N_p \\ \mathbf{R}_0 \left\{ \hat{\sigma}_e^{2(k+1)} = \left[ \tilde{\mathbf{e}}^{T(k)} \tilde{\mathbf{e}}^{(k)} + tr \left( \mathbf{W} \mathbf{C}^{-1(k)} \mathbf{W}^T \right) \right] / N \end{array} \right. \quad [S_3]$$

$$\hat{\sigma}_{e_{(i)}}^{2(k+1)} = \tilde{\mathbf{e}}^{T(k)} \dot{\mathbf{R}}_{(i)} \tilde{\mathbf{e}}^{(k)} + tr \left( \mathbf{W} \mathbf{C}^{-1(k)} \mathbf{W}^T \dot{\mathbf{R}}_{(i)} \right) / N \quad [S_4]$$

Na Eq. [S\_5] são apresentadas as equações iterativas para estimação dos parâmetros das matrizes de (co)variância do modelo 4.

$$EM_{M_4} \left[ \begin{array}{l}
\mathbf{G}_{0g_1} \left\{ \begin{array}{l}
\hat{\sigma}_{g_1(i)}^{2(k+1)} = \tilde{\mathbf{g}}_1^T \dot{\mathbf{G}}_{g_1(i)} \tilde{\mathbf{g}}_1^{(k)} + tr \left( \mathbf{C}^{g_1 g_1(k)} \dot{\mathbf{G}}_{g_1(i)} \right) / N_{g_1} \\
\hat{\sigma}_{g_1(i)}^{(k+1)} = \tilde{\mathbf{g}}_1^T \dot{\mathbf{G}}_{g_1(i)} \tilde{\mathbf{g}}_1^{(k)} + tr \left( \mathbf{C}^{g_1 g_1(k)} \dot{\mathbf{G}}_{g_1(i)} \right) / 2N_{g_1}
\end{array} \right. \\
\mathbf{G}_{0g_2} \left\{ \begin{array}{l}
\hat{\sigma}_{g_2(i)}^{2(k+1)} = \tilde{\mathbf{g}}_2^T \dot{\mathbf{G}}_{g_2(i)} \tilde{\mathbf{g}}_2^{(k)} + tr \left( \mathbf{C}^{g_2 g_2(k)} \dot{\mathbf{G}}_{g_2(i)} \right) / N_{g_2} \\
\hat{\sigma}_{g_2(i)}^{(k+1)} = \tilde{\mathbf{g}}_2^T \dot{\mathbf{G}}_{g_2(i)} \tilde{\mathbf{g}}_2^{(k)} + tr \left( \mathbf{C}^{g_2 g_2(k)} \dot{\mathbf{G}}_{g_2(i)} \right) / 2N_{g_2}
\end{array} \right. \\
\mathbf{G}_{0s} \left\{ \begin{array}{l}
\hat{\sigma}_{s(i)}^{2(k+1)} = \tilde{\mathbf{s}}^T \dot{\mathbf{G}}_{s(i)} \tilde{\mathbf{s}}^{(k)} + tr \left( \mathbf{C}^{ss(k)} \dot{\mathbf{G}}_{s(i)} \right) / N_s \\
\hat{\sigma}_{s(i)}^{(k+1)} = \tilde{\mathbf{s}}^T \dot{\mathbf{G}}_{s(i)} \tilde{\mathbf{s}}^{(k)} + tr \left( \mathbf{C}^{ss(k)} \dot{\mathbf{G}}_{s(i)} \right) / 2N_s
\end{array} \right. \\
\mathbf{G}_{0p} \left\{ \begin{array}{l}
\hat{\sigma}_{p(i)}^{2(k+1)} = \tilde{\mathbf{p}}^T \dot{\mathbf{G}}_{p(i)} \tilde{\mathbf{p}}^{(k)} + tr \left( \mathbf{C}^{pp(k)} \dot{\mathbf{G}}_{p(i)} \right) / N_p \\
\hat{\sigma}_{p(i)}^{(k+1)} = \tilde{\mathbf{p}}^T \dot{\mathbf{G}}_{p(i)} \tilde{\mathbf{p}}^{(k)} + tr \left( \mathbf{C}^{pp(k)} \dot{\mathbf{G}}_{p(i)} \right) / 2N_p
\end{array} \right. \\
\mathbf{R}_0 \left\{ \begin{array}{l}
\hat{\sigma}_{e(i)}^{2(k+1)} = \tilde{\mathbf{e}}^T \dot{\mathbf{R}}_{(i)} \tilde{\mathbf{e}}^{(k)} + tr \left( \mathbf{W} \mathbf{C}^{-1(k)} \mathbf{W}^T \dot{\mathbf{R}}_{(i)} \right) / N \\
\hat{\sigma}_{e(i)}^{(k+1)} = 0
\end{array} \right.
\end{array} \right. \quad [S_5]$$

### 7.1.2. MATRIZ Q DO ALGORITMO AI

A matriz  $\mathbf{Q}$ , referente ao modelo dialéxico apresentado na Eq. [15], pode ser obtida por meio da Eq. [S\_6] apresentadas a seguir:

$$\mathbf{Q}^{(k)} = \left[ \mathbf{q}_{g_1}^{(k)}, \mathbf{q}_{g_2}^{(k)}, \mathbf{q}_s^{(k)}, \mathbf{q}_p^{(k)}, \mathbf{q}_e^{(k)} \right] \Rightarrow \left\{ \begin{array}{l}
\mathbf{q}_{g_1(i)}^{(k)} = \mathbf{Z}_{g_1} \dot{\mathbf{G}}_{g_1(i)} \mathbf{G}_{g_1}^{-1(k)} \tilde{\mathbf{g}}_1^{(k)} \\
\mathbf{q}_{g_2(i)}^{(k)} = \mathbf{Z}_{g_2} \dot{\mathbf{G}}_{g_2(i)} \mathbf{G}_{g_2}^{-1(k)} \tilde{\mathbf{g}}_2^{(k)} \\
\mathbf{q}_{s(i)}^{(k)} = \mathbf{Z}_s \dot{\mathbf{G}}_{s(i)} \mathbf{G}_s^{-1(k)} \tilde{\mathbf{s}}^{(k)} \\
\mathbf{q}_{p(i)}^{(k)} = \mathbf{Z}_p \dot{\mathbf{G}}_{p(i)} \mathbf{G}_p^{-1(k)} \tilde{\mathbf{p}}^{(k)} \\
\mathbf{q}_{e(i)}^{(k)} = \dot{\mathbf{R}}_{(i)} \mathbf{R}^{-1(k)} \tilde{\mathbf{e}}^{(k)}
\end{array} \right. \quad [S_6]$$

em que  $\mathbf{q}_{g_1}^{(k)}$ ,  $\mathbf{q}_{g_2}^{(k)}$ ,  $\mathbf{q}_s^{(k)}$ ,  $\mathbf{q}_p^{(k)}$  e  $\mathbf{q}_e^{(k)}$  são as matrizes contendo as ‘*working variables*’ para os parâmetros das estruturas de (co)variância (Tabela 2) associadas aos efeitos aleatórios da Eq. [15] na iteração  $k$ . Note que, o número de colunas de cada matriz  $\mathbf{q}$  é equivalente ao número de parâmetros estimados na respectiva matriz de (co)variância.

### 7.1.3. ESPAÇO PARAMÉTRICO

No presente trabalho, assumiu-se valor de 0,50 para o peso  $\lambda$  (Eq. [11]) na iteração onde um ou mais parâmetros foram atualizados fora do espaço paramétrico pelo algoritmo AI e 1,0 caso contrário. Além disso, seguiu-se as recomendações sugeridas por Gilmour et al. (2019) para atualização dos parâmetros via algoritmo AI. Cabe ressaltar, que restrições específicas foram aplicadas aos parâmetros *boundary*, visando evitar estimativas fora do espaço paramétrico e garantir à convergência.

Considerando os componentes de variância, estabeleceu-se que os valores fora do espaço paramétrico, após a utilização do fator  $\lambda$ , na iteração  $k + 1$  devem ser substituídos pelo valor correspondente na iteração  $k$  dividido por 15. Essa penalização faz com que o parâmetro *boundary* se aproxime da fronteira do espaço paramétrico sem ultrapassá-la. Contudo, se o problema recorrer, e o componente de variância assumir valor inferior a  $10^{-7}$  na iteração  $k + 1$ , o parâmetro então é fixado próximo à fronteira do espaço paramétrico com seu valor correspondente da iteração  $k$ . Por outro lado, estabeleceu-se atualização restrita pelo fator  $\pm 0,30$  para correlações com valor absoluto inferior a 0,70 a cada iteração. Contudo, correlações com valor absoluto  $\geq 0,70$  na iteração  $k$  que apresentaram atualizações fora do intervalo de -1 a 1 na iteração  $k + 1$  foram atualizadas pela expressão apresentada na Eq. [S\_7]. De maneira similar à apresentada para os componentes de variância *boundary*, essa penalização faz com que a

correlação *boundary* se aproxime da fronteira do espaço paramétrico sem ultrapassá-la. Entretanto, se o problema recorrer, e a correlação assumir valor absoluto superior a 0,9999 na iteração  $k + 1$ , então fixa-se a correlação próximo à fronteira do espaço paramétrico com seu valor correspondente da iteração  $k$ .

$$\rho^{(k+1)} = \rho^{(k)} + \frac{1 \pm \rho^{(k)}}{2\rho^{(k)}} \quad [S_7]$$

## 7.2. ESTRUTURAS DE (CO)VARIÂNCIA UTILIZADAS NO MODELO PARA ANÁLISE DO DIALELO PARCIAL EM MEGT

No presente trabalho, foram utilizadas diferentes estruturas de (co)variância (VC, DIAG, DIAGH, CORH e US), para modelar os efeitos genéticos e não genéticos do modelo dialélico apresentado na Eq. [15]. Na Tabela 2 estas estruturas foram especificadas para os efeitos aleatórios da Eq. [15] e as diferentes combinações entre elas possibilitou a definição de quatro variações do modelo dialélico apresentado na Eq. [15]. Aqui estas estruturas serão apresentadas de maneira genérica considerando cinco ensaios, porém, a generalização destas estruturas para os efeitos aleatórios da Eq. [15] é imediata, bastando-se indexar os parâmetros das estruturas de (co)variância aos efeitos aleatórios da Eq. [15].

### 7.2.1. ESTRUTURA DE (CO)VARIÂNCIA VC

Esta estrutura de (co)variância é a mais simples dentre as ajustadas. Ela assume homogeneidade de variâncias e possui um único parâmetro ( $\sigma^2$ ). Esta estrutura foi utilizada para modelar os resíduos da Eq. [15] (Tabela 2).

$$\mathbf{R} = \mathbf{I}_n \sigma^2 \quad [S_8]$$

em que  $\mathbf{I}_n$  é uma matriz identidade de dimensão equivalente ao número de parcelas avaliadas nos cinco ensaios.

### 7.2.2. ESTRUTURA DE (CO)VARIÂNCIA DIAGH

Esta estrutura de (co)variância assume heterogeneidade de variâncias e covariâncias nulas e foi utilizada para modelar os resíduos da Eq. [15] (Tabela 2). Ela guarda similaridade com a estrutura de (co)variância DIAG, entretanto, além de assumir heterogeneidade de variâncias ela permite modelar os resíduos de experimentos com número de repetições diferentes, ou seja, conectar experimentos que apresentam desbalanceamento estatístico. O número de parâmetros associados a esta estrutura de (co)variância é equivalente ao número de ensaios utilizados. No presente trabalho, foram estimados cinco parâmetros quando se utilizou esta estrutura ( $\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2$  e  $\sigma_5^2$ ).

$$\mathbf{R} = \bigoplus_{t=1}^{N_t} \mathbf{I}_{n_t} \sigma_t^2 = \begin{bmatrix} \mathbf{I}_{n_1} \sigma_1^2 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n_2} \sigma_2^2 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{n_3} \sigma_3^2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_{n_4} \sigma_4^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_{n_5} \sigma_5^2 \end{bmatrix} \quad [\text{S}_9]$$

em que  $\mathbf{I}_{n_1}, \mathbf{I}_{n_2}, \mathbf{I}_{n_3}, \mathbf{I}_{n_4}$  e  $\mathbf{I}_{n_5}$  são matrizes identidade de dimensão equivalente ao número de parcelas avaliadas em cada um dos cinco ensaios, respectivamente.

### 7.2.3. ESTRUTURA DE (CO)VARIÂNCIA DIAG

Esta estrutura de (co)variância assume heterogeneidade de variâncias e covariâncias nulas e foi utilizada para modelar os efeitos genéticos da Eq. [15] (Tabela 2). O número de parâmetros associados a esta estrutura de (co)variância é equivalente ao número de ensaios utilizados. No

presente trabalho, foram estimados cinco parâmetros quando se utilizou esta estrutura ( $\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2$  e  $\sigma_5^2$ ).

$$\mathbf{G} = \mathbf{I} \otimes \mathbf{G}_0 = \mathbf{I} \otimes \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 & 0 \\ 0 & 0 & 0 & \sigma_4^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma_5^2 \end{bmatrix} \quad [\text{S}_{10}]$$

em que  $\mathbf{I}$  é uma matriz identidade de dimensão equivalente ao número de níveis do efeito aleatório modelado.

#### 7.2.4. ESTRUTURAS DE (CO)VARIÂNCIA CORH

Esta estrutura de (co)variância assume heterogeneidade de variâncias e correlação uniforme entre ensaios e foi utilizada para modelar os efeitos genéticos da Eq. [15] (Tabela 2). O número de parâmetros associados a esta estrutura de (co)variância é equivalente ao número de ensaios utilizados mais a correlação uniforme para os ensaios. No presente trabalho, foram modelados seis parâmetros quando se utilizou esta estrutura ( $\rho, \sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2$  e  $\sigma_5^2$ ).

$$\mathbf{G} = \mathbf{I} \otimes \mathbf{G}_0 = \mathbf{I} \otimes \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \rho\sigma_1\sigma_3 & \rho\sigma_1\sigma_4 & \rho\sigma_1\sigma_5 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 & \rho\sigma_2\sigma_3 & \rho\sigma_2\sigma_4 & \rho\sigma_2\sigma_5 \\ \rho\sigma_1\sigma_3 & \rho\sigma_2\sigma_3 & \sigma_3^2 & \rho\sigma_3\sigma_4 & \rho\sigma_3\sigma_5 \\ \rho\sigma_1\sigma_4 & \rho\sigma_2\sigma_4 & \rho\sigma_3\sigma_4 & \sigma_4^2 & \rho\sigma_4\sigma_5 \\ \rho\sigma_1\sigma_5 & \rho\sigma_2\sigma_5 & \rho\sigma_3\sigma_5 & \rho\sigma_4\sigma_5 & \sigma_5^2 \end{bmatrix} \quad [\text{S}_{11}]$$

em que  $\mathbf{I}$  é uma matriz identidade de dimensão equivalente ao número de níveis do efeito aleatório modelado.

### 7.2.5. ESTRUTURAS DE (CO)VARIÂNCIA US

Esta estrutura de (co)variância assume heterogeneidade de variâncias e covariâncias e foi utilizada para modelar os efeitos genéticos da Eq. [15] (Tabela 2). O número de parâmetros associados a esta estrutura de (co)variância é equivalente a  $t(t + 1)/2$ , sendo  $t$  o número de ensaios. No presente trabalho, foram estimados quinze parâmetros quando se utilizou esta estrutura ( $\sigma_1^2, \sigma_{12}, \sigma_2^2, \sigma_{13}, \sigma_{23}, \sigma_3^2, \sigma_{14}, \sigma_{24}, \sigma_{34}, \sigma_4^2, \sigma_{15}, \sigma_{25}, \sigma_{35}, \sigma_{45}$  e  $\sigma_5^2$ ).

$$\mathbf{G} = \mathbf{I} \otimes \mathbf{G}_0 = \mathbf{I} \otimes \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{15} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} & \sigma_{24} & \sigma_{25} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 & \sigma_{34} & \sigma_{35} \\ \sigma_{14} & \sigma_{24} & \sigma_{34} & \sigma_4^2 & \sigma_{45} \\ \sigma_{15} & \sigma_{25} & \sigma_{35} & \sigma_{45} & \sigma_5^2 \end{bmatrix} \quad [\text{S}_{12}]$$

em que  $\mathbf{I}$  é uma matriz identidade de dimensão equivalente ao número de níveis do efeito aleatório modelado.

Note que, diferente da estrutura de CORH, que assume correlação homogênea para todos os pares de ensaios, a estrutura US assume covariâncias/correlações heterogenias para cada par de ensaios.

### 7.3. DERIVADAS DAS ESTRUTURAS DE (CO)VARIÂNCIA UTILIZADAS NO MODELO PARA ANÁLISE DO DIALELO PARCIAL EM MEGT

A estimação de parâmetros via REML envolve as derivadas de primeira e segunda ordem da função de verossimilhança residual, dependendo do algoritmo utilizado, em relação aos parâmetros a serem estimados. As derivadas para as estruturas de efeitos aleatórios da Eq. [15] foram apresentadas na Eq. [S\_2], entretanto, não foram mostradas as derivadas de primeira ordem das estruturas de (co)variância, associadas aos efeitos aleatórios da Eq. [15], em relação

aos parâmetros a serem estimados ( $\dot{\mathbf{G}}_{g_{1(i)}}, \dot{\mathbf{G}}_{g_{2(i)}}, \dot{\mathbf{G}}_{s_{(i)}}, \dot{\mathbf{G}}_{p_{(i)}} e \dot{\mathbf{R}}_{(i)}$ ). Assim, serão apresentados aqui as derivadas de forma genérica para as estruturas de (co)variância especificadas na Tabela 2 considerando cinco ensaios, porém, a generalização destas estruturas para os efeitos aleatórios da Eq. [15] é imediata, bastando-se indexar as derivadas aos parâmetros das estruturas de (co)variância associadas aos efeitos aleatórios da Eq. [15].

### 7.3.1. DERIVADA PARA A ESTRUTURA DE (CO)VARIÂNCIA VC

A estrutura de (co)variância VC foi apresentada na Eq. [S\_8]. Como já foi mencionado, ela possui apenas um parâmetro ( $\sigma^2$ ). A derivada em relação a este parâmetro é apresentada a seguir:

$$\dot{\mathbf{R}} = \mathbf{I}_n \quad [S_{13}]$$

em que  $\dot{\mathbf{R}}$  representa a derivada de primeira ordem da matriz  $\mathbf{R}$  (Eq. [S\_8]) em relação ao parâmetro ( $\sigma^2$ ).

### 7.3.2. DERIVADAS PARA A ESTRUTURA DE (CO)VARIÂNCIA DIAGH

A estrutura de (co)variância DIAGH foi apresentada na Eq. [S\_9]. Como já foi mencionado, ela possui cinco parâmetros ( $\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2 e \sigma_5^2$ ). A seguir, será apresentada apenas a derivada com relação ao primeiro parâmetro ( $\sigma_1^2$ ), pois as demais derivadas são triviais.

$$\dot{\mathbf{R}}_{(i)} = \begin{bmatrix} \mathbf{I}_{n_i} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \quad [S_{14}]$$

em que  $\dot{\mathbf{R}}_{(1)}$  representa a derivada de primeira ordem da matriz  $\mathbf{R}$  (Eq. [S\_9]) em relação ao parâmetro ( $\sigma_1^2$ ).

### 7.3.3. DERIVADAS PARA A ESTRUTURA DE (CO)VARIÂNCIA DIAG

A estrutura de (co)variância DIAG foi apresentada na Eq. [S\_10]. Como já foi mencionado, ela possui cinco parâmetros ( $\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2$  e  $\sigma_5^2$ ). A seguir, será apresentada apenas a derivada com relação ao primeiro parâmetro ( $\sigma_1^2$ ), pois as demais derivadas são triviais.

$$\dot{\mathbf{G}}_{(1)} = \mathbf{I} \otimes \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad [\text{S}_{15}]$$

em que  $\dot{\mathbf{G}}_{(1)}$  representa a derivada de primeira ordem da matriz  $\mathbf{G}$  (Eq. [S\_10]) em relação ao parâmetro ( $\sigma_1^2$ ).

### 7.3.4. DERIVADAS PARA A ESTRUTURA DE (CO)VARIÂNCIA CORH

A estrutura de (co)variância CORH foi apresentada na Eq. [S\_11]. Como já foi mencionado, ela possui seis parâmetros ( $\rho, \sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2$  e  $\sigma_5^2$ ). A seguir, serão apresentadas apenas as derivadas de primeira ordem com relação aos parâmetros ( $\rho$  e  $\sigma_1^2$ ), pois as demais derivadas são triviais.

$$\dot{\mathbf{G}}_{(1)} = \mathbf{I} \otimes \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2 & \sigma_1\sigma_3 & \sigma_1\sigma_4 & \sigma_1\sigma_5 \\ \sigma_1\sigma_2 & \sigma_2^2 & \sigma_2\sigma_3 & \sigma_2\sigma_4 & \sigma_2\sigma_5 \\ \sigma_1\sigma_3 & \sigma_2\sigma_3 & \sigma_3^2 & \sigma_3\sigma_4 & \sigma_3\sigma_5 \\ \sigma_1\sigma_4 & \sigma_2\sigma_4 & \sigma_3\sigma_4 & \sigma_4^2 & \sigma_4\sigma_5 \\ \sigma_1\sigma_5 & \sigma_2\sigma_5 & \sigma_3\sigma_5 & \sigma_4\sigma_5 & \sigma_5^2 \end{bmatrix} \quad [\text{S}_{-16}]$$

$$\dot{\mathbf{G}}_{(2)} = \mathbf{I} \otimes \begin{bmatrix} 1 & \frac{\rho\sigma_2}{2\sqrt{\sigma_1}} & \frac{\rho\sigma_3}{2\sqrt{\sigma_1}} & \frac{\rho\sigma_4}{2\sqrt{\sigma_1}} & \frac{\rho\sigma_5}{2\sqrt{\sigma_1}} \\ \frac{\rho\sigma_2}{2\sqrt{\sigma_1}} & 0 & 0 & 0 & 0 \\ \frac{\rho\sigma_3}{2\sqrt{\sigma_1}} & 0 & 0 & 0 & 0 \\ \frac{\rho\sigma_4}{2\sqrt{\sigma_1}} & 0 & 0 & 0 & 0 \\ \frac{\rho\sigma_5}{2\sqrt{\sigma_1}} & 0 & 0 & 0 & 0 \end{bmatrix} \quad [\text{S}_{-17}]$$

em que  $\dot{\mathbf{G}}_{(1)}$  e  $\dot{\mathbf{G}}_{(2)}$  representam as derivadas de primeira ordem da matriz  $\mathbf{G}$  (Eq. [S\_11]) com relação aos parâmetros ( $\rho$  e  $\sigma_1^2$ ), respectivamente.

### 7.3.5. DERIVADAS PARA A ESTRUTURA DE (CO)VARIÂNCIA US

A estrutura de (co)variância US foi apresentada na Eq. [S\_12]. Como já foi mencionado, ela possui quinze parâmetros ( $\sigma_1^2, \sigma_{12}, \sigma_2^2, \sigma_{13}, \sigma_{23}, \sigma_3^2, \sigma_{14}, \sigma_{24}, \sigma_{34}, \sigma_4^2, \sigma_{15}, \sigma_{25}, \sigma_{35}, \sigma_{45}$  e  $\sigma_5^2$ ). A seguir, serão apresentadas apenas as derivadas de primeira ordem com relação aos parâmetros ( $\sigma_1^2$  e  $\sigma_{12}$ ), pois as demais derivadas são triviais.

$$\dot{\mathbf{G}}_{(1)} = \mathbf{I} \otimes \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad [\text{S}_{-18}]$$

$$\dot{\mathbf{G}}_{(2)} = \mathbf{I} \otimes \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad [\text{S}_{19}]$$

em que  $\dot{\mathbf{G}}_{(1)}$  e  $\dot{\mathbf{G}}_{(2)}$  representam as derivadas de primeira ordem da matriz  $\mathbf{G}$  (Eq. [S\_12]) com relação aos parâmetros ( $\sigma_1^2$  e  $\sigma_{12}$ ), respectivamente.

## CONCLUSÕES GERAIS

Os algoritmos AI e EM-AI se mostraram alternativas computacionalmente eficientes para estimar os componentes de variância em análises dialélicas.

Os componentes de variância dos efeitos de RCGC, RCEC e linhagens parentais impactam os componentes de variância de CGC e CEC, especialmente quando exibem grande magnitude. Além disso, o modelo que inclui os efeitos de RCGC e RCEC apresenta potencial para o uso em programas de melhoramento focados na obtenção de híbridos ou clones superiores.

O modelo linear misto que assume heterogeneidade de variâncias para os efeitos genéticos e não genético, bem como correlação uniforme para os efeitos genéticos do dialelo se mostrou adequado para modelar dados MEGT em feijão comum. Assim, modelar as estruturas de (co)variância dos efeitos genéticos e não genéticos do dialelo em MEGT permite um estudo mais detalhado da interação  $G \times A$ , sendo esta uma estratégia recomendada para os programas de melhoramento de plantas autógamas.

A análise dialélica via modelos mistos é uma ferramenta poderosa para estimação de parâmetros genéticos e apresenta potencial para uso nos programas de melhoramento. Isto porque, o procedimento REML pode ser facilmente derivado para modelos mais gerais, além de possibilitar a estimação de parâmetros genéticos sob desbalanceamento estatístico e/ou genético e modelar estruturas de (co)variância complexas.