



**GABRIELLE CARVALHO PEREIRA**

**SELEÇÃO GENÔMICA AMPLA (GWS) SOB ASSIMETRIA  
PARA RESISTÊNCIA À PODRIDÃO DA ESPIGA EM  
MILHO**

**LAVRAS - MG  
2021**

**GABRIELLE CARVALHO PEREIRA**

**SELEÇÃO GENÔMICA AMPLA (GWS) SOB ASSIMETRIA PARA RESISTÊNCIA À  
PODRIDÃO DA ESPIGA EM MILHO**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do programa de Pós-graduação em Genética e Melhoramento de Plantas, área de concentração em Genética e Melhoramento de Plantas, para obtenção do título de Mestre.

Prof. Dr. Renzo Garcia Von Pinho  
Orientador

Prof. Dr. Márcio Balestre  
Coorientador  
(*in memoriam*)

**LAVRAS - MG  
2021**

**Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca  
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).**

Pereira, Gabrielle Carvalho.

Seleção Genômica Ampla (GWS) sob assimetria para  
resistência à podridão da espiga em milho / Gabrielle Carvalho  
Pereira. - 2021.

42 p.

Orientador(a): Renzo Garcia Von Pinho.

Coorientador(a): Marcio Balestre.

Dissertação (mestrado acadêmico) - Universidade Federal de  
Lavras, 2021.

Bibliografia.

1. Predição Genômica. 2. Modelo bayesiano. 3. *Fusarium  
verticillioides*. I. Von Pinho, Renzo Garcia. II. Balestre, Marcio. III.  
Título.

O conteúdo desta obra é de responsabilidade do(a) autor(a) e de seu orientador(a).

**GABRIELLE CARVALHO PEREIRA**

**SELEÇÃO GENÔMICA AMPLA (GWS) SOB ASSIMETRIA PARA RESISTÊNCIA À  
PODRIDÃO DA ESPIGA EM MILHO**

**GENOMIC WIDE SELECTION (GWS) UNDER ASYMMETRY ASSOCIATE TO  
EAR ROT RESISTANCE IN MAIZE**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento de Plantas, área de concentração em Genética e Melhoramento de Plantas, para obtenção do título de Mestre.

APROVADA em 29 de janeiro de 2021.

Prof. Dr. Júlio Silvío de Sousa Bueno Filho      UFLA  
Dra. Marcela Pedroso Mendes Resende      UFG



Prof. Dr. Renzo Garcia Von Pinho  
Orientador

Prof. Dr. Márcio Balestre  
Coorientador  
(*in memoriam*)

**LAVRAS - MG  
2021**

## AGRADECIMENTOS

Agradeço primeiramente a Deus, por sempre guiar meu caminho e permitir que eu chegasse até aqui.

Ao coorientador Márcio Balestre (*in memoriam*), por todo apoio, ensinamentos e incentivo durante todo o projeto. Você foi fundamental para que esse trabalho acontecesse. Admiro muito o profissional que você foi e tenho certeza que deixou o seu legado aqui na Terra.

Ao professor e orientador Renzo, pela oportunidade, apoio, ensinamentos e paciência.

Aos meus pais, Eunice e Dehon, pela educação, amor e paciência, e por sempre me apoiarem nas minhas decisões. À minha irmã, Cristiane, pela compreensão, carinho e por todos os conselhos.

Ao professor Júlio e professor Diógenes, pelo apoio, dedicação e contribuições no trabalho.

Aos amigos e colegas do programa de genética e melhoramento de plantas, pela amizade e parceria durante o mestrado. À Luciana pela amizade e por sempre estar disposta em ajudar. Aos amigos, Eric e Wilson, pela paciência, contribuição e ensinamentos durante a condução desse trabalho.

À Universidade Federal de Lavras (UFLA) e ao Programa de Pós-Graduação em Genética e Melhoramento de Plantas (PPGMP), pela oportunidade e recursos durante o mestrado.

À Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), pela concessão da bolsa de estudos.

Enfim, agradeço a todas as pessoas que passaram pelo meu caminho durante essa etapa da minha vida e contribuíram para a concretização desse ciclo.

## RESUMO

Considerada uma cultura de grande impacto econômico, o milho pode ter a sua produtividade afetada pelo patógeno *Fusarium verticillioides*, que pode causar danos efetivos, como os grãos ardidos e micotoxinas, que são substâncias carcinogênicas, e podem afetar a saúde humana e animal. Muitos estudos têm sido feitos visando à obtenção de genótipos resistentes a esse patógeno. Trabalhos relatam que a resistência a essas doenças é controlada por genes de herança quantitativa, sendo a seleção fenotípica dificultada nesses caracteres, devido à baixa herdabilidade e alta influência do ambiente. Entre as ferramentas mais utilizadas nos programas de melhoramento de plantas, a Seleção Genômica Ampla (GWS) apresenta alta eficácia na seleção de genótipos superiores. Alguns caracteres de caráter quantitativo podem apresentar distribuição assimétrica, principalmente dados de resistência às doenças de plantas. Quando isso ocorre, nem sempre a transformação dos dados é uma alternativa eficaz, sendo indicada a utilização de modelos que lidem bem com a assimetria. Portanto nesse trabalho objetivou-se verificar a eficiência na utilização de modelo misto normal assimétrico bayesiano na predição de dados com distribuição assimétrica, em comparação ao GBLUP. As análises fenotípicas foram feitas em Lavras-MG e Uberlândia-MG, sendo avaliados três caracteres: porcentagem de grãos ardidos, proporção de espigas com sintomas e escore de incidência de podridão de espiga. Após verificar a assimetria dos dados, foi feita a transformação destes como formas de corrigir a não normalidade, mas ainda assim os dados continuaram a apresentar assimetria. Na análise dos parâmetros estimados, os caracteres grãos ardidos e escore apresentaram maior herdabilidade em comparação à proporção de espigas doentes, portanto esses caracteres podem ser utilizados na obtenção de genótipos resistentes à podridão da espiga causada por *Fusarium verticillioides*. Nas análises com os modelos GBLUP e o Modelo Assimétrico Bayesiano, foi observada uma alta herdabilidade e também correlação para os caracteres analisados sob o Modelo Assimétrico Bayesiano, diferente do GBLUP, que obteve menor herdabilidade e menor correlação. A alta correlação e boa predição genômica apresentada pelo Modelo Assimétrico Bayesiano leva a inferir que esse modelo é eficaz na análise de dados com distribuição assimétrica.

**Palavras-chave:** *Fusarium verticillioides*. Predição genômica. Distribuição assimétrica.

## ABSTRACT

Maize is a crop of great economic impact, but has its productivity affected by the *fusarium verticillioides* pathogen, which can cause rotten kernels and mycotoxins. In addition to all the management that must be done to control this disease, the use of resistant genotypes is the most effective. Several studies report that resistance to these diseases is controlled by genes of quantitative inheritance, and phenotypic selection is difficult in these characters, due to low heritability and high influence of the environment. Among the most used tools in plant breeding programs, the Wide Genomic Selection (GWS) is highly effective in selecting superior genotypes. Some characters of quantitative character may present skew normal distribution, mainly on resistance to plant diseases. When this occurs, data transformation is not always an effective alternative, and the use of models that deal with this skew normal is recommended. Therefore, this work aimed to verify the efficiency in the use of Mixed Normal Asymmetric Bayesian Model in the prediction of data with skew normal distribution and by GBLUP. Phenotypic analyzes were performed in the Lavras and Uberlândia environments and three characters were evaluated: percentage of rotten kernels, proportion of diseased ears and ear rot score. After verifying the data, the transformation was made as a way to correct non-normality, but even so the data presented skew normal distribution. In the analysis of the estimated parameters, the characters rotten kernels and score showed greater heritability compared to the proportion of diseased ears, so these characters can be used to obtain genotypes resistant to ear rot caused by *fusarium verticillioides*. In the analyzes with the GBLUB and the Bayesian Asymmetric Model, a high heritability and correlation were observed for the characters analyzed under the Bayesian Asymmetric Model, different from GBLUP, which obtained a lower heritability and less correlation. The high correlation and good genomic prediction presented by the Bayesian Asymmetric Model leads to infer that this model is effective in analyzing data with asymmetric distribution.

**Keywords:** *Fusarium verticillioides*. Genomic Prediction. Skew Normal.

## LISTA DE FIGURAS

- Figura 1 – Gráficos com distribuição assimétrica para os dados de *Fusarium verticillioides*...30
- Figura 2 – Gráficos com os dados transformados para o caráter grãos ardidos na análise conjunta e na análise individual de Lavras e Uberlândia.....31
- Figura 3 – Árvore filogenética construída pelo método Neighbor-Joining mostrando os diferentes *backgrounds* das linhagens utilizadas no estudo de seleção genômica.....32

## LISTA DE TABELAS

- Tabela 1 - Variância aditiva ( $V_a$ ), Variância Residual ( $V_e$ ), Beta, Herdabilidade ( $h^2$ ), Correlação e Delta obtidos pelo GBLUP e Modelo Misto Normal Assimétrico Bayesiano (MAB) para grãos ardidos, nas análises de Lavras, Uberlândia e Conjunta.....33
- Tabela 2 - Variância aditiva ( $V_a$ ), Variância Residual ( $V_e$ ), Beta, Herdabilidade ( $h^2$ ), Correlação e Delta obtidos pelo GBLUP e Modelo Misto Normal Assimétrico Bayesiano (MAB) para proporção de espigas doentes, nas análises de Lavras, Uberlândia e Conjunta.....34
- Tabela 3 - Variância aditiva ( $V_a$ ), Variância Residual ( $V_e$ ), Beta, Herdabilidade ( $h^2$ ), Correlação e Delta obtidos pelo GBLUP e Modelo Misto Normal Assimétrico Bayesiano (MAB) para score, nas análises de Lavras, Uberlândia e Conjunta.....35

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO .....</b>	<b>10</b>
<b>2</b>	<b>REVISÃO BIBLIOGRÁFICA .....</b>	<b>12</b>
<b>2.1</b>	<b>A cultura do milho.....</b>	<b>12</b>
<b>2.2</b>	<b>Podridão da espiga .....</b>	<b>12</b>
<b>2.3</b>	<b><i>Fusarium verticillioides</i> .....</b>	<b>13</b>
<b>2.4</b>	<b>Controle genético de <i>Fusarium verticillioides</i> .....</b>	<b>14</b>
<b>2.5</b>	<b>Seleção Genômica Ampla (GWS).....</b>	<b>15</b>
<b>2.6</b>	<b>Distribuição assimétrica.....</b>	<b>17</b>
<b>2.7</b>	<b>G-BLUP .....</b>	<b>19</b>
<b>2.8</b>	<b>Modelo Misto Normal Assimétrico .....</b>	<b>19</b>
<b>2.9</b>	<b>Modelo Misto via Inferência Bayesiana.....</b>	<b>21</b>
<b>3</b>	<b>MATERIAL E MÉTODOS .....</b>	<b>23</b>
<b>3.1</b>	<b>Fenotipagem.....</b>	<b>23</b>
<b>3.2</b>	<b>Genotipagem .....</b>	<b>24</b>
<b>3.3</b>	<b>Análise estatística.....</b>	<b>25</b>
<b>3.4</b>	<b>Teste de normalidade .....</b>	<b>25</b>
<b>3.5</b>	<b>Distribuição assimétrica.....</b>	<b>25</b>
<b>3.6</b>	<b>Seleção genômica ampla via G-BLUP .....</b>	<b>26</b>
<b>3.7</b>	<b>Modelo misto normal assimétrico Bayesiano .....</b>	<b>26</b>
<b>3.8</b>	<b>Distribuições a priori para os parâmetros do modelo .....</b>	<b>27</b>
<b>4</b>	<b>RESULTADOS E DISCUSSÃO .....</b>	<b>30</b>
<b>4.1</b>	<b>Dados com distribuição assimétrica.....</b>	<b>30</b>
<b>4.2</b>	<b>Caracterização das linhagens .....</b>	<b>31</b>
<b>4.3</b>	<b>Predição por G-BLUP e modelo misto normal assimétrico.....</b>	<b>32</b>
<b>5</b>	<b>CONCLUSÕES.....</b>	<b>37</b>
	<b>REFERÊNCIAS .....</b>	<b>38</b>

## 1 INTRODUÇÃO

O milho é uma das principais culturas produzidas no Brasil, com previsão para atingir uma produção aproximada de 102,3 milhões de toneladas na safra 2020/21 (CONAB, 2020). Até o final dos anos 1990, as doenças no milho não eram vistas como problema primário na cultura. Com as mudanças nos sistemas de produção, como expansão da fronteira agrícola, aumento da época de cultivo (safrinha), utilização de plantio direto sem o uso de rotação de culturas, a inadequação no uso dos insumos agrícolas, mudanças climáticas globais, entre outras práticas, as doenças se tornaram fatores limitantes à produtividade. Entre as doenças que reduzem a produtividade está a podridão da espiga, causada pelo *Fusarium verticillioides*. Esse fungo causa perdas na produtividade e deterioração na qualidade dos grãos, causando as micotoxinas, que trazem grande prejuízo para a saúde animal e humana, os quais consomem produtos com milho (ODRIOZOLA et al., 2005). Essas micotoxinas são carcinogênicas, teratogênicas e mutagênicas.

Várias práticas agronômicas têm sido utilizadas como forma de reduzir a incidência da podridão em plantas de milho, mas uma das práticas mais eficientes é o uso de cultivares resistentes. Trabalhos utilizando técnicas de melhoramento genético têm sido feitos visando a utilização das cultivares resistentes como seleção genômica e associação genômica ampla.

Nos últimos anos, a genotipagem utilizando SNPs espalhados pelo genoma, possibilitou o uso de ferramentas de alta eficácia, como a GWAS (Análise de Associação Genômica Ampla) e a GS (Seleção Genômica). A GWS é uma ferramenta de capacidade de predição de dados não observados, através do uso de marcadores moleculares. Os marcadores mais utilizados na seleção genômica geralmente são os SNPs e os DArTs. Os marcadores DArTs apresentam a vantagem de detecção da variação em milhares de locos gênicos a um custo relativamente baixo.

Alguns caracteres quantitativos podem apresentar distribuições assimétricas quando os dados são obtidos de múltiplas fontes ou contém observações isoladas (LI et al., 2015). Nesses casos, seria indicada a transformação dos dados, porém nem sempre essa é uma forma eficaz. Segundo Azzalini e Capitanio (1999), a transformação não fornece informação útil para entender o mecanismo de geração dos dados podendo dificultar a interpretação. A transformação para um conjunto de dados pode frequentemente não ser aplicável para outros conjuntos de dados e quando a suposição de homocedasticidade é necessária, algumas vezes a transformação requerida difere da transformação para alcançar a normalidade.

A GWS através da Inferência Bayesiana tem sido utilizada nos programas de

melhoramento genético. Modelos de regressão não lineares com erros normais são muito aplicados, porém quando ocorre assimetria dos dados, a suposição de normalidade não se aplica (ROSSI; SANTOS, 2014). O modelo bayesiano é um modelo estatístico robusto que pode ser trabalhado com caracteres fenotípicos que apresentam distribuição assimétrica. Doenças como a podridão de espiga em milho podem apresentar assimetria. Portanto, neste trabalho foi utilizado o modelo misto normal assimétrico em dados de doenças de podridão de espiga no âmbito da seleção genômica ampla. O objetivo foi verificar se o modelo misto normal assimétrico bayesiano é eficiente na capacidade preditiva dos dados, comparado ao GBLUP.

## 2 REVISÃO BIBLIOGRÁFICA

### 2.1 A cultura do milho

O milho (*Zea mays* L.) é uma espécie pertencente à família Gramineae/Poaceae, sendo uma cultura produzida em várias partes do mundo e de grande importância econômica. É utilizada na alimentação humana e animal, devido a sua qualidade nutricional. Está entre as principais culturas produzidas no país, ficando atrás apenas da soja. Na safra 2020/21 a produção deve atingir 102,3 milhões de toneladas. O Brasil está como segundo maior exportador, sendo os Estados Unidos o maior exportador de milho, e terceiro maior produtor, ficando atrás da China e dos Estados Unidos, que lideram a produção (CONAB, 2021).

No Brasil, a produção de milho ocorre em duas safras. A segunda safra, também conhecida como safrinha, ocorre principalmente na região Centro-Oeste, sendo plantada após a colheita da soja. A produção de milho na safrinha tem sido superior à produção da safra, em algumas regiões. O plantio de milho após a colheita da soja possibilitou o aumento da produtividade da cultura no Brasil, porém contribuiu também para o aumento da proliferação de pragas e doenças na lavoura; isso representa um grande desafio para os pesquisadores agrícolas. Portanto além de fazer o correto manejo de pragas e doenças, várias pesquisas têm sido feitas com o intuito de obter novas tecnologias que visem a maior produtividade do milho.

### 2.2 Podridão da espiga

Várias são as doenças que atacam a cultura do milho, causando danos na produção e consequente perda na produtividade. As mudanças globais, tanto no clima quanto no sistema de produção, têm agravado a incidência de patógenos na cultura. A podridão da espiga, que é responsável por causar grãos ardidos e pela produção de micotoxinas no milho, é uma das doenças mais graves na cultura. Entre os fungos causadores de podridões de espiga, estão *Stenocarpella maydis* e *S. macrospora*, causadores da podridão de diplodia ou podridão branca, *Fusarium verticillioides* (*Fusarium moniliforme*), *F. subglutinans*, *F. proliferatum* e *F. graminearum* causadores da podridão de fusarium ou podridão rosada da espiga, *Giberella zeae* (anamorfo *F. graminearum*), agente causal da podridão vermelha ou da ponta da espiga, *Penicillium oxalicum*, causador da podridão de espiga por *Penicillium* e *Aspergillus flavus*, causador da podridão de *Aspergillus* e de grãos armazenados (FANTIN; DUARTE, 2009; CASELA et al., 2006; REIS et al., 2004). As podridões destacam-se, no mundo, entre as mais

importantes doenças que atacam a cultura do milho por causarem redução de produção e de qualidade de grãos (COSTA et al., 2005; VIANA, 2009).

Entre os danos causados à produtividade por esses fungos, destaca-se a redução do peso dos grãos, redução da germinação e do vigor das sementes e redução da qualidade do produto pelo aumento da deterioração deles. Esses fungos são produtores de micotoxinas, as quais são metabólitos secundários. Mesmo após o controle do fungo, a micotoxina continua no alimento. Os principais efeitos desses metabólitos tóxicos são, em longo prazo, a acumulação no organismo, tanto de humanos como de animais. Os fungos do gênero *Stenocarpella* spp. não têm sido, tradicionalmente, considerados entre os principais produtores de micotoxinas em milho (COSTA et al., 2013).

A infecção que ocorre na espiga resulta em grãos ardidos, que são assim classificados quando infectados por fungos e ocorre a descoloração de, pelo menos, um quarto da superfície dos grãos. A coloração desses grãos pode variar de marrom claro a roxo ou de vermelho claro a vermelho intenso, dependendo do patógeno presente (PINTO, 2005).

Esses fungos podem contaminar as espigas na pré-colheita, causando os grãos ardidos e na pós-colheita, causando grãos mofados durante o armazenamento e beneficiamento (PINTO, 2005).

No controle são indicadas algumas práticas culturais, como tratamento de sementes, colheita antecipada, rotação de culturas, além da utilização de cultivares resistentes.

### **2.3 *Fusarium verticillioides***

A podridão de espiga causada pelo fungo *Fusarium verticillioides* é caracterizada como podridão rosada da espiga. O agente etiológico é endofítico e penetra nos grãos através do ponto de inserção da semente com a espiga (CHAMBERLAIN et al., 1993), alojando abaixo do pericarpo e prosseguindo disseminação pela semente (BULLERMAN; TSAI, 1994). Esses patógenos apresentam elevado número de plantas hospedeiras alternativas, sendo então considerados parasitas não especializados.

A disseminação dos esporos se dá através do vento e de insetos e o período de maior suscetibilidade ocorre de 7 a 10 dias após a polinização dos estigmas. Essa doença se desenvolve em ambiente seco, com temperatura elevada no início da cultura e condições úmidas no florescimento. O cultivo do milho em monocultura e plantio direto favorece a sobrevivência, a manutenção e a multiplicação do inóculo destes fungos (ZAMBOLIM et al., 2000).

Os sintomas podem ser apresentados como uma pigmentação rosa (*F. verticillioides*) ou roxa (*F. graminearum*) entre os grãos, sendo que as espigas que não dobram após a maturidade fisiológica dos grãos e aquelas com mau empalhamento são as mais suscetíveis. Pode desenvolver uma massa cotonosa avermelhada, recobrando os grãos infectados ou a área da palha atingida. Quando a infecção ocorre através do pedúnculo da espiga, todos os grãos podem ser infectados, mas a infecção só se desenvolverá naqueles que apresentarem alguma injúria no pericarpo. O desenvolvimento desses fungos nas espigas é paralisado quando o teor de umidade dos grãos atinge 18 a 19%. Embora esses fungos sejam frequentemente isolados das sementes, essas não são a principal fonte de inóculo. Como esses fungos possuem a fase saprofítica ativa, sobrevivem e se multiplicam no resto de cultura de milho, sendo essa a fonte principal de inóculo. Outro sintoma são os grãos ardidos, que podem causar reduções nos conteúdos de carboidratos, de proteínas e de açúcares totais (PINTO, 2005). No momento da comercialização esse sintoma causa prejuízo ao produtor, pois a incidência de grãos ardidos é descontada na venda. Como padrão de qualidade, adota-se, em algumas agroindústrias, a tolerância máxima de 6% para grãos ardidos, em lotes comerciais de milho.

As micotoxinas produzidas pelo fungo *Fusarium verticillioides* são chamadas de fumonisinas, das quais foram identificados os tipos B1, B2, B3 e B4 (MUSSE; PLATTNER, 1997; PROCTOR et al., 2006). De 28 fumonisinas estudadas, a B1 é a mais prejudicial, podendo causar leucoencefalomácia em equinos (LEME), edema pulmonar em suínos, câncer hepático em ratos, redução do desenvolvimento, imunossupressão, problemas cardíacos, degeneração e necrose hepática em aves, além de estar associada ao câncer de esôfago em humanos (NORRED; VOS, 1994; MARASAS, 1996).

A rotação de culturas é uma prática essencial utilizada na agricultura que ajuda a reduzir a podridão da espiga na cultura do milho, assim como o tratamento de sementes, colheita precoce e o armazenamento dos grãos sob condições de umidade relativa adequada. Mas o uso de cultivares resistentes é uma das práticas mais eficientes no controle da doença.

#### **2.4 Controle genético de *Fusarium verticillioides***

A resistência à podridão da espiga, causada pelo fungo *Fusarium verticillioides*, é encontrada em vários estudos que visam identificar o mecanismo de ação dos genes de resistência à essa doença. Em seu trabalho, Pérez-Brito et al. (2001) testando duas populações F2 de milho, de 238 e 206 indivíduos respectivamente, com um cruzamento simples entre linhagens resistentes e suscetíveis para resistência à podridão da espiga, encontrou baixa

herdabilidade para o caráter ( $h^2 = 0.26-0.42$ ). Já Robertson-Hoyt et al. (2006) ao testar duas populações para a contaminação por fumonisina e resistência à podridão da espiga, encontrou herdabilidade alta, de 80% para a resistência à podridão e de 86% para a contaminação por fumonisina. Uma herdabilidade baixa pode indicar que uma proporção significativa da variação do caráter é devida ao ambiente ou ao erro experimental, que um grande número de genes controlam o caráter e que as diferenças relativas entre os valores genotípicos dependem do ambiente (interação genótipo x ambiente) (ATLIN et al., 2000; HOLLAND et al., 2003).

A seleção assistida por marcadores tem sido utilizada na avaliação fenotípica de doenças, utilizando cruzamentos biparentais, nos quais os pais devem ser divergentes para a característica em avaliação (ROBERTSON-HOYT et al., 2006; SINGH, 2015). Posteriormente, é feita a fenotipagem da população em diferentes locais e anos, e os dois pais são testados com um amplo número de marcadores, que cobrem todo o genoma. Depois, é feita a genotipagem dos indivíduos da população, utilizando os marcadores polimórficos previamente identificados em ambos os pais. Os dados genotípicos são utilizados para construir um mapa de ligação para a população, e, por fim, os dados genotípicos e os dados fenotípicos são analisados para detectar associação entre os marcadores e os QTL (SINGH, 2015). Em estudos de mapas de ligação em populações biparentais, locus de caráter quantitativo (QTL) para podridão da espiga por *Fusarium* tem pequenos efeitos e não são consistentes entre populações (PÉREZ-BRITO et al. 2001). Grandes bases genéticas e uma forte influência do ambiente têm dificultado a localização precisa do QTL, reduzindo assim a eficiência da seleção assistida por marcadores (MAS) (ROBERTSON-HOYT et al., 2006). Aumentar o tamanho da população e o número de marcadores, melhorar a fenotipagem e usar dados de múltiplos ambientes, ajudam a superar tais limitações (ROBERTSON et al., 2005).

O desenvolvimento de métodos de genotipagem e estudos de seleção genômica ampla têm se mostrado eficientes na identificação de caracteres de baixa herdabilidade.

## **2.5 Seleção Genômica Ampla (GWS)**

O surgimento dos marcadores moleculares na década de 80 possibilitou a criação de importantes ferramentas no melhoramento genético, auxiliando a obtenção de genótipos superiores. Entre essas ferramentas tem-se a Seleção Genômica Ampla (GWS), proposta por Meuwissen et al., (2001), que consiste na predição dos valores genéticos genômicos dos indivíduos utilizando uma ampla gama de marcadores genéticos. A predição é feita utilizando dados fenotípicos e genotípicos; assim os dados genotípicos são utilizados para inferir sobre

os dados que não foram fenotipados. É utilizada uma população de treinamento, em que se têm os dados genotípicos e fenotípicos reais, e a população de validação, em que terá os dados preditos.

Na GWS é necessário que haja desequilíbrio de ligação (DL) entre o QTL (Locus de caráter quantitativo) e o marcador, e somente marcadores em LD podem explicar a variação genética. Como se utiliza uma ampla gama de marcadores, a probabilidade de se encontrar um QTL em LD com o marcador é alta. Desequilíbrio de ligação (LD) é uma associação não aleatória de alelos em uma população, ou seja, é a ocorrência de gametas com frequências diferentes daquelas esperadas sob segregação independente (GUPTA et al., 2005).

A GWS ganha destaque no melhoramento de plantas, comparada com outras ferramentas, como a SAM (Seleção Assistida por Marcadores). A SAM utiliza marcadores moleculares ligados ao locus de interesse para seleção de indivíduos superiores, mas captura apenas QTLs de grande efeito, não explicando toda a variação genética do indivíduo. Já a GWS detecta tanto QTLs de grande efeito como de pequeno efeito, além de utilizar muitos marcadores que cobrem todo o genoma, o que aumenta a chance de encontrar um marcador em LD com o QTL.

Na seleção genômica, o efeito dos marcadores é estimado utilizando os dados genotípicos e fenotípicos da população de treinamento, e um modelo é ajustado para estimar os efeitos dos parâmetros. Assim, é calculado o valor genético genômico (GEBV) da população de validação. Com a estimativa do valor genético genômico e o valor fenotípico dos indivíduos se têm a acurácia da predição genômica. Um esquema de validação cruzada é utilizado para validar o modelo, através da população de treinamento e da população teste. Na validação cruzada, divide-se a população em  $k$  grupos de mesmo tamanho e em cada um dos grupos retira-se  $k$  indivíduos para formar a população de validação. Em cada subgrupo é feita a correlação entre o valor genético predito e valor fenotípico dos indivíduos avaliados (BIAGIOTTI, 2018).

A Seleção Genômica começou sendo implantada no melhoramento animal e passou a ser utilizada também em plantas, principalmente para selecionar genótipos com caracteres de herança quantitativa, visto que estes sofrem grande influência do ambiente e são controlados por muitos genes. Na literatura são encontrados muitos trabalhos em que se obteve sucesso com a GWS no melhoramento vegetal. Fritsche-Neto et al. (2012) aplicaram a GWS em milho e obtiveram uma alta acurácia e ganhos genéticos significativos. A GWS também foi utilizada para explorar a resistência ao nematoide do cisto em soja (BAO et al., 2014) e para selecionar uma alta resistência a *Stenocarpella maydis* em linhagens de milho (DOS

SANTOS et al., 2016).

O marcador DArT - Diversity Arrays Technology (JACCOUD et al., 2001; WENZL et al., 2004), apresenta a vantagem de detecção da variação em milhares de locos gênicos sem a necessidade de sequenciamento prévio (WENZL et al., 2006) e tem sido utilizado na GWS, já que apresenta baixo custo de genotipagem por loco, e é um procedimento rápido e robusto, não sendo necessário o uso de géis de agarose ou poliacrilamida. Essa metodologia permite gerar uma gama de marcadores, com ampla cobertura do genoma (JACCOUD et al., 2001). Os marcadores DArTs são bialélicos e apresentam dominância. Considerado um marcador confiável e útil em pesquisas em melhoramento genético utilizando desequilíbrio de ligação e estudos de associação, o DArT é encontrado em estudos de várias culturas, entre elas, o *Eucalyptus* (SANSALONI et al., 2010), trigo (ORABI et al., 2014) e cevada (LEX et al., 2014).

Devido ao grande número de efeitos a serem estimados, simultaneamente, a partir de um pequeno número de observações, o tamanho amostral e a colinearidade do desequilíbrio de ligação entre os marcadores e a interpretação dos resultados, são necessários modelos e metodologias estatísticas adequadas na predição dos efeitos dos marcadores (CANTOR; LANGE; SINSHEIMER, 2010; PEARSON; MANOLIO, 2008; RESENDE et al., 2008). Assim, entre os modelos que mais têm sido utilizados na GWS, se encontram os modelos mistos, como GBLUP (*Genomic Best Linear Unbiased Prediction*), e os modelos com inferência bayesiana. No GBLUP, os marcadores assumem uma única distribuição normal, diferente do BayesB, em que os marcadores de maior e menor efeito são alocados em duas distribuições diferentes (LI; HENGDE et al., 2015).

## 2.6 Distribuição assimétrica

Distribuições assimétricas são caracterizadas por apresentarem o terceiro momento não nulo, ou seja, apresentam assimetria de cauda. Nesses casos há modelos baseados em classes de distribuição assimétrica, que são discutidas por Azzalini (1985). A distribuição normal assimétrica apresenta o parâmetro  $\Delta$ , que controla a assimetria, quando  $\Delta = 0$ , a distribuição é simétrica, se  $\Delta < 0$ , ocorre assimetria à esquerda e se  $\Delta > 0$ , ocorre assimetria à direita. Frequentemente, a distribuição assimétrica é consequência de uma mistura de dados de duas ou mais subpopulações, gerando convolução de densidades (PEWSEY, 2000).

Em modelos tradicionais a análise de variância deve atender aos quatro pressupostos, que são aditividade, independência dos erros, normalidade e homocedasticidade dos erros.

Quando algumas dessas pressuposições são violadas, a confiabilidade de todos os testes paramétricos, tais como a análise de variância, a comparação de médias e a análise de regressão ficam comprometidas, pois ocorrem alterações na probabilidade de ocorrência do erro tipo I (rejeição da hipótese de nulidade quando ela é verdadeira) e tipo II (não rejeição da hipótese de nulidade quando ela é falsa), podendo levar a falsas conclusões a respeito dos efeitos de tratamento (STEEL et al., 1997; MARTIN; STORCK, 2008; STORCK et al., 2011). Quando ocorre violação da pressuposição da normalidade dos erros é desejável que os testes apresentem controle das taxas de erros tipo I e II. Esse tipo de característica é conhecido por robustez do procedimento (BORGES; FERREIRA, 2003).

Em estudos genéticos de plantas ou animais é possível encontrar caracteres quantitativos com distribuição não normal, pois os dados são obtidos de múltiplas fontes ou contém observações isoladas (LI et al., 2015). Vários caracteres fenotípicos têm distribuição assimétrica, por exemplo, florescimento feminino, florescimento masculino, o intervalo entre florescimento feminino e masculino, medida categórica das doenças em escala e dados contáveis (PÉREZ RODRIGUES et al., 2018),

Segundo Varona et al. (2008), as distribuições assimétricas observadas para alguns fenótipos são resultado de fatores ambientais e os dados são utilizados para resíduos de distribuições não simétricas. Nesses casos, tanto a transformação dos dados da variável resposta ou um modelo que lida com resposta assimétrica podem ser usados, assim como modelos lineares mistos com distribuição assimétrica dos resíduos. Segundo Azzalini e Capitanio, (1999), a transformação dos dados nem sempre é eficaz, pois não fornece informação útil para entender o mecanismo de geração dos dados, pode dificultar a interpretação, principalmente quando cada variável é transformada usando diferentes funções. Além disso, em um conjunto de dados pode frequentemente não ser aplicável para outros conjuntos de dados e quando a suposição de homocedasticidade é necessária, algumas vezes a transformação requerida difere da transformação para alcançar a normalidade. A violação desse pressuposto pode reduzir o poder da análise, uma vez que dados assimétricos podem diminuir o poder do teste.

Supondo que  $\varphi_n(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma})$  e  $\Phi_n(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma})$ , sendo a função de densidade de probabilidade (pdf) e a função de densidade cumulativa (cdf), respectivamente, da distribuição normal  $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  sendo essa avaliada em  $\mathbf{x}$ . Quando  $\boldsymbol{\mu} = \mathbf{0}$ , e  $\boldsymbol{\Sigma} = \mathbf{I}_n$  ( $n \times n$  matriz de identidade), essas funções são consideradas como  $\varphi_n(\mathbf{x})$  e  $\Phi_n(\mathbf{x})$ .

Um vetor normal segue uma distribuição n-variada normal com vetor de locação  $\boldsymbol{\mu} \in \mathbb{R}^n$ , matriz de covariância  $\boldsymbol{\Sigma}$  ( $n \times n$  matriz identidade positiva) e com matriz de assimetria  $\boldsymbol{\Delta}$  ( $n$

x k), se sua função densidade de probabilidade é dada por:

$$f(y|y;\mu;\Delta) = 2^n \varphi_n(y|\mu, \Sigma + \Delta\Delta^T) \Phi_n(\Delta^T(\Sigma + \Delta\Delta^T)^{-1}) \quad (1.1)$$

$$(y - \mu)|0, (\mathbf{I}_k + \Delta^T \Sigma^{-1} \Delta)^{-1} \quad (1.2)$$

## 2.7 G-BLUP

Como falado anteriormente, o G-BLUP (VANRADEN et al., 2009) é um dos modelos utilizados na seleção genômica, utilizando a relação entre os indivíduos estimados pelos marcadores moleculares para prever o valor genético genômico. Nesse modelo BLUP genômico, todos os efeitos dos marcadores são normalmente distribuídos e alocados em uma única distribuição. É um método muito utilizado no melhoramento animal e vegetal, podendo ser aplicado para caracteres de herança poligênica e que apresentam normalidade dos dados. Diferente do BLUP tradicional, o GBLUP avalia cada segregação em nível individual e não em nível médio, assim permite obter uma melhor predição do indivíduo (RESENDE et al., 2008).

O modelo é basicamente descrito pela seguinte equação (HENDERSON, 1984):

$$Y = X\beta + Z\alpha + \varepsilon \quad (1.3)$$

Em que Y é o vetor das observações, Xβ é o vetor de efeitos fixos, Zα é o vetor de efeito aleatório e ε é o vetor de resíduos.

## 2.8 Modelo Misto Normal Assimétrico

Um modelo misto é aquele que apresenta tanto fatores de efeitos fixos como aleatórios, além do erro experimental e da constante μ. Quando um modelo é considerado misto, a parte aleatória do modelo não compõe a esperança dos dados, mas sua variância. Essa característica implica em estimar os componentes da variância fenotípica. Quando se adota um modelo linear misto existe a possibilidade de se fazer a predição de efeitos aleatórios, na presença de efeitos fixos, através dos BLUP's (best linear unbiased prediction) que são de grande valia em genética e melhoramentos (CAMARINHA, 2003).

Os modelos lineares mistos são comumente usados quando os dados assumem

distribuição normal, porém podem não apresentar robustez quando há dados com assimetria. Apesar dos modelos mistos serem mais usados quando assumem distribuição normal simétrica, há na literatura alguns casos em que os modelos mistos foram usados em distribuição assimétrica. O modelo misto tradicional é dado por (ARELLANO-VALE et al., 2005):

$$Y_j = X_j\beta + Z_jb_j + e_j, j = 1, \dots, m \quad (1.4)$$

$$Y \sim N(X_j\beta + Z_jb_j, r) \quad (1.5)$$

Em que,  $Y$  é um  $(n_j \times 1)$  vetor de resposta, com unidade de amostra  $j$ ,  $j = 1, \dots, m$ .  $X_j$  de dimensão  $(n_j \times p)$  é a matriz de efeitos fixos,  $\beta$  de dimensão  $(p \times 1)$  é um vetor de coeficiente de regressão da média populacional, que são os efeitos fixos,  $Z_j$  de dimensão  $(n_j \times q)$  é a matriz correspondente a  $(q \times 1)$  vetor de efeitos aleatórios  $b_j$ ,  $e_j$  de dimensão  $(n_j \times 1)$  é o vetor de resíduos. Normalmente assumindo que  $b_j$  e  $e_j$  são independentes, considerando:

$$b_j \sim N_q(\mathbf{0}, \mathbf{D}) \text{ e } e_j \sim N_{n_j}(\mathbf{0}, \psi_j) \quad (1.6)$$

Sendo  $Y \sim N(X\beta + Zb, r)$ . Em que,  $\mathbf{D} = \mathbf{D}(\alpha)$  e  $\psi_j = \psi_j(\gamma)$ ,  $j = 1, \dots, m$ , são matrizes de dispersão, associadas com a variância dentro e entre indivíduos, os quais são dependentes em parâmetros reduzidos e desconhecidos,  $\alpha$  e  $\gamma$ , respectivamente (ARELLANO et al., 2005). O modelo misto tradicional é muito utilizado, porém quando os dados apresentam assimetria, esse modelo pode não ser eficiente. Muitos estudos têm sido feitos utilizando modelos que lidam bem com a assimetria.

O modelo misto normal assimétrico segue o mesmo modelo misto tradicional, com  $b_j$  e  $e_j$  independentes, considerando:

$$b_j \sim SN_q(\mathbf{0}, \mathbf{D}, \Delta_b) \text{ e } e_j \sim N_{n_j}(\mathbf{0}, \psi_j, \Delta_e) \quad (1.7)$$

Sendo  $Y \sim SN(X\beta + Zb, \Delta, r)$ , sendo  $\Delta$  a matriz de assimetria  $n \times k$ .

Os modelos mistos são bastante usados na avaliação de dados fenotípicos nos programas de melhoramento genético, pois dados fenotípicos geralmente provém de avaliações de diferentes anos agrícolas e diferentes localidades, podendo apresentar

desbalanceamento (MOHRING; PIEPHO, 2009).

## 2.9 Modelo Misto via Inferência Bayesiana

Antes de se realizar uma análise dos dados é necessário escolher adequadamente um modelo que se ajusta ao comportamento dos dados em estudo. A metodologia Bayesiana tem sido usada na estimação dos parâmetros genéticos, pois ela mede o grau de incerteza que se tem sobre a ocorrência de determinado evento do espaço amostral, por meio de probabilidades.

A inferência Bayesiana é realizada com dados amostrais e informações a priori sobre o parâmetro e o cálculo da distribuição a posteriori do parâmetro. A informação a priori, que pode ser informativa ou não informativa, é dada pela densidade de probabilidade  $P(\theta)$ , que expressa o conhecimento do pesquisador sobre o parâmetro a ser estimado, antes de serem obtidos os dados. A informação a priori não informativa é considerada quando o pesquisador não tem muita informação a ser incorporada no estudo (JEFFREYS, 1961). É necessária a especificação de uma distribuição conjunta para os dados amostrais,  $Y = \{y_1, y_2, \dots, y_n\}$ , denominada de função de verossimilhança, ou  $L(\theta|y)$ , a qual representa a informação sobre o parâmetro avaliado,  $\theta$ , obtida dos dados. A operação bayesiana,  $L(\theta|y) p(\theta)$ , resulta na distribuição condicional a posteriori  $p(\theta|y)$ , que contém toda a informação probabilística a respeito do parâmetro. Assim, toda a inferência relativa ao parâmetro é realizada por meio dessa distribuição, e seus valores centrais, como média e mediana, fornecem as estimativas de interesse, que serão idênticas às obtidas com procedimentos frequentistas quando a distribuição a priori é não informativa (PAULINO; TURKMAN; MURTEIRA, 2003). Na estatística bayesiana é utilizado o Teorema de Bayes para quantificar a informação sobre o parâmetro em estudo, a densidade a posteriori de  $\theta$ ,  $P(\theta|y)$ , que é expresso por:

$$P(\theta|Y) = \frac{L(Y|\theta)P(\theta)}{\int L(Y|\theta)P(\theta)d\theta} \quad (1.8)$$

Sendo  $Y = \{y_1, y_2, \dots, y_n\}$ , o vetor de observações,  $L(Y|\theta)$  a função de verossimilhança e  $P(\theta)$  a densidade a priori de  $\theta$ . O denominador é uma constante, não depende de  $\theta$ . Então se tem o teorema como:  $P(\theta|Y) \propto L(Y|\theta) P(\theta)$ , sendo que  $\propto$  representa proporcionalidade. Na expressão acima,  $L(Y|\theta) = \prod_i f(Y_i|\theta)$  é a função de verossimilhança.

Em estudos genômicos, como GWAS e GS, a inferência bayesiana é utilizada, pois apresenta vantagens em relação a outros métodos, já que prediz o efeito de todos os locos

simultaneamente (BayesA) e assume variâncias específicas para todos os locos (BayesA e BayesB) (Júnior, Oliveira et al. 2017). Essa metodologia também permite inferir os graus de incerteza em relação a algum parâmetro genético, por meio de um modelo a priori, que considera tanto as observações, quanto os parâmetros, como aleatórios (SOLLERO et al., 2015).

Como dito anteriormente, vários caracteres estudados em melhoramento genético apresentam distribuição assimétrica, e a inferência bayesiana via modelos mistos tem sido usada no estudo desses caracteres. Em estudos de seleção genômica, em que se utilizam uma alta densidade de marcadores em relação ao número de fenótipos da população ( $p \gg n$ ), os métodos bayesianos apresentam maior eficácia em relação aos métodos de regressão clássicos (PÉREZ et al. 2018). Campos (2011) utilizou a abordagem bayesiana para estimar parâmetros de modelos de regressão não linear com distribuição assimétrica, utilizando técnicas MCMC (Método Monte Carlo via Cadeias de Markov). De La Cruz; Branco (2009) utilizando dados de um grupo de mulheres grávidas, com inferência Bayesiana para modelos de regressão não linear, levando em consideração a assimetria dos dados, encontraram que o melhor método para os erros é a distribuição assimétrica.

### 3 MATERIAL E MÉTODOS

#### 3.1 Fenotipagem

A avaliação dos experimentos para incidência da podridão da espiga foi feita em dois locais, Lavras e Uberlândia, na safra 2012/2013. Foram avaliadas 238 linhagens, que foram conduzidas em blocos incompletos, 10 tratamentos por bloco (oito tratamentos regulares e dois tratamentos comuns), com três repetições. A parcela era composta por uma linha de 3 metros e um espaçamento de 0,7 metros. O experimento não foi balanceado devido à perda de algumas parcelas.

Os fungos de *F. verticillioides* foram obtidos de espigas que continham o patógeno em campos experimentais, de Lavras e Uberlândia, e posteriormente foram replicados no laboratório de patologia de sementes da Universidade Federal de Lavras (UFLA). O inóculo do patógeno foi cultivado em meio de cultura por sete dias antes da inoculação. No dia da inoculação, uma suspensão do conídios foi então ajustada a uma concentração de  $1 \times 10^6$  conídios  $\text{mL}^{-1}$  contando esporos na câmara de Neubauer. Foi utilizada a metodologia de Clements et al. (2003) com algumas modificações na inoculação do patógeno (foi utilizado 1 mL da suspensão conidial ao invés de 5 mL). A inoculação foi feita 15 dias depois que as plantas tiveram emitido o estilo estigma. Foi inoculado 1 mL da suspensão em cada espiga.

A incidência da podridão da espiga foi avaliada pela proporção de espigas que apresentaram os sintomas. A presença do patógeno na espiga foi determinada por cada espiga que apresentou mais de 25% dos grãos com os devidos sintomas.

Foram avaliados três caracteres em relação à podridão da espiga, consistindo em porcentagem de grãos ardidos (ARD), proporção de espigas com sintomas de podridão (PROP) e escore de incidência de podridão de espiga (SCO).

A avaliação da porcentagem de grãos ardidos foi realizada de acordo com o procedimento proposto no decreto n. 11 de 04/12/96, que estabeleceu uma amostra de 230 g de grãos por parcela para separação visual e determinação da porcentagem de grãos apresentando descoloração em mais de um quarto da superfície total. A avaliação da proporção de espigas com sintomas de podridão foi realizada por meio do cálculo da proporção de espigas com sintomas característicos das doenças em relação ao número total de espigas por parcela. E o escore de incidência de podridão de espiga foi realizado por meio de uma escala diagramática proposta por Reid et al. (2002). Os valores dessa escala variam de 1 a 7 e incluem as seguintes categorias: 1 (0%); 2 (1-3%); 3 (4-10%); 4 (11-25%); 5 (26-50%);

6 (51-75%); e 7 (maior que 75%), sendo a porcentagem referente à incidência de podridão na espiga.

### 3.2 Genotipagem

A avaliação dos experimentos para incidência da podridão da espiga foi feita em dois locais, Lavras e Uberlândia, na safra 2012/2013. Foram avaliadas 238 linhagens, que foram conduzidas em blocos incompletos, 10 tratamentos por bloco (oito tratamentos regulares e dois tratamentos comuns), com três repetições. A parcela era composta por uma linha de 3 metros e um espaçamento de 0,7 metros. O experimento não foi balanceado devido à perda de algumas parcelas.

Foram utilizadas 470 linhagens provenientes de uma empresa privada de melhoramento e produção de sementes de milho. Devido a não amplificação de algumas amostras, foram obtidos dados de 447 linhagens. Foi feita a extração do DNA das amostras coletadas no experimento, instalado em 2013, na área experimental da Universidade Federal de Lavras, no município de Lavras – MG. O experimento, realizado em delineamento em blocos incompletos, continha 47 blocos com 10 linhagens cada, duas repetições e parcela constituída de duas fileiras de 3 m de comprimento com espaçamento de 0,6 m entre fileiras. Foram semeadas 25 sementes por linha e, posteriormente, foi feito o desbaste de três plantas em estádio V2 na parcela.

A extração de DNA foi realizada seguindo protocolo específico da empresa Diversity Arrays Technology. As amostras de DNA devidamente diluídas foram enviadas para a empresa Diversity Arrays Technology para genotipagem. As linhagens foram genotipadas utilizando 23153 marcadores DArTs. As marcas foram avaliadas como —1 para a presença e —0 para ausência do alelo e — quando não foi possível a leitura com confiabilidade pelo *software*. Marcadores perdidos foram imputados pela função *A.mat*, método mean, do pacote *rrBLUP* do software R. Para capturar as relações genéticas entre os indivíduos, foi elaborada a matriz de parentesco aditiva (**A**) seguindo a proposta de Vitezica; Varona; Legarra (2013):

$$A = \frac{W_A W_A'}{\sum pq} \quad (1.9)$$

Em que **p** é a frequência do alelo favorável e **q** a frequência do alelo desfavorável; **WA** é a matriz de desvio dos marcadores.

A matriz descrita acima foi utilizada para construir uma árvore filogenética pelo

método *neighbor-joining* (SAITOU; NEI, 1987). A função *nj* do pacote *ape* (PARADIS et al., 2004) do *software R* (R CORE TEAM, 2016) foi utilizada na construção da árvore.

### 3.3 Análise estatística

Na análise estatística foram usados modelos mistos para ajuste de médias, via máxima verossimilhança restrita (REML), utilizando o algoritmo EM (expectation – maximization) (PATTERSON; THOMPSON, 1971). As médias foram corrigidas considerando os efeitos de repetições, ambientes, interação genótipos x ambientes, e o resíduo, adotando-se o modelo:

$$Y = X\beta + Tg + \Omega b + W\delta + e \quad (2.0)$$

Em que  $Y$  é o vetor de observações  $n \times 1$ ;  $X$  é uma matriz  $n \times p$  de incidência de feitos fixos (repetição dentro de local e local);  $T$  é uma matriz  $n \times q$  de incidência de efeitos genéticos;  $\Omega$  é uma matriz de incidência dos efeitos aleatórios de blocos dentro de locais e repetições;  $W$  é uma matriz de incidência dos efeitos de interação linhagens x ambientes; e  $g$ ,  $b$  e  $\delta$ , são vetores dos efeitos relacionados a  $X$ ,  $T$ ,  $\Omega$ ,  $W$ , respectivamente e  $e$  são os efeitos residuais. A distribuição dos efeitos  $g$ ,  $b$ ,  $\delta$  e  $e$  são assumidas como  $N(0, \sigma_g^2)$ ,  $N(0, \sigma_b^2)$ ,  $N(0, \sigma_\delta^2)$  e  $N(0, \sigma_e^2)$ , respectivamente.

### 3.4 Teste de normalidade

Para confirmar a não normalidade dos dados, foi feito teste de normalidade Shapiro – Wilk (SS SHAPIRO, MB WILK. 1965), com os dados fenotípicos referentes ao patógeno *fusarium*. Esse teste utiliza uma estatística  $W$  que testa se uma amostra  $n$  provém de uma distribuição normal. Quando *p-value* é menor que 0,05, a hipótese nula de normalidade é rejeitada e os dados assumem a não normalidade.

### 3.5 Distribuição assimétrica

Como forma de corrigir a não normalidade dos dados, foi feita a transformação dos dados de *fusarium*, utilizando a transformação Box-Cox (BOX; COX, 1964). Foi feita a estimação do parâmetro  $\lambda$  para cada subconjunto de marcadores e depois foi calculada a média desses parâmetros. Posteriormente, foi feito um histograma para verificar se a distribuição dos dados era simétrica. Para a transformação dos dados, o parâmetro  $\lambda$  foi utilizado na seguinte equação:

$$yt_i = \frac{y_i^\lambda - 1}{\lambda} \quad (2.1)$$

Após a transformação dos dados, foi feito novamente o teste de Shapiro-Wilk para confirmar a normalidade dos dados transformados.

### 3.6 Seleção genômica ampla via G-BLUP

O modelo GBLUP, que é um modelo que apresenta eficácia na seleção genômica ampla, foi utilizado nesse trabalho para fazer a predição genômica e assim estimar os componentes de variância e a herdabilidade dos dados fenotípicos.

$$Y = Xb + Zu + e \quad (2.2)$$

Em que:

$Y$ : vetor dos dados fenotípicos;

$b$ : vetor de efeito fixo;

$u$ : vetor de efeito aditivo, que contém o valor genético genômico de todos os indivíduos genotipados;

$X$ : matriz de efeito fixo;

$Z$ : matriz de efeito aleatório, que relaciona os indivíduos aos fenótipos.

### 3.7 Modelo misto normal assimétrico Bayesiano

A predição genômica também foi realizada nesse trabalho por meio do modelo misto normal assimétrico, via inferência bayesiana, através do seguinte modelo:

$$Y = X\beta + Za + \varepsilon \quad (2.3)$$

Onde:

$Y$  tem dimensão  $n \times 1$  é um vetor de respostas;

$X$  tem dimensão  $n \times p$  é uma matriz de incidência dos efeitos fixos;

$\beta$  tem dimensão  $p \times 1$  é o vetor dos efeitos fixos;

$Z$  tem dimensão  $n \times q_a$  é uma matriz de incidência dos efeitos aditivos;

$\mathbf{a}$  tem dimensão  $q_a \times 1$  é o vetor dos efeitos aditivos;

$\boldsymbol{\varepsilon}$  é o vetor de erros aleatórios ou resíduos de dimensão  $n \times 1$ .

Assim, para o modelo misto normal assimétrico supõe-se que:

$$\mathbf{a} \sim SN_q(0, S_a, \Delta_a) \text{ e } \boldsymbol{\varepsilon} \sim SN_n(0, \Omega_a, \Delta_\varepsilon) \quad (2.4)$$

Em que  $S_a$  e  $\Omega_a$  são matrizes de covariâncias entre os efeitos aleatórios  $\mathbf{a}_1 \dots \mathbf{a}_q$ , e o resíduo  $\boldsymbol{\varepsilon}_1 \dots \boldsymbol{\varepsilon}_n$ ; Os parâmetros  $\Delta_a \in \mathfrak{R}^{q \times q}$  e  $\Delta_\varepsilon \in \mathfrak{R}^{n \times n}$  são matrizes diagonais com os elementos  $\delta_a 1, \dots, \delta_a q_a$  e  $\delta_{\varepsilon 1}, \dots, \delta_{\varepsilon n}$ , respectivamente, correspondentes aos parâmetros de assimetria. Portanto:  $\Delta_a = \delta_a I_{q_a}$  e  $\Delta_\varepsilon = \delta_\varepsilon I_n$ , com  $\delta_a \in \mathfrak{R}$  e  $\delta_\varepsilon \in \mathbb{R}$ . Sendo assim obtemos a forma do modelo hierárquico do modelo misto da seguinte maneira:

$$Y | \beta, \mathbf{a}, \sigma^2, \delta_\varepsilon \sim SN_n(X\beta + Z\mathbf{a}, \sigma_\varepsilon^2 I_n, \delta_\varepsilon I_n) \quad (2.5)$$

$$a | \sigma_a^2, \delta_a \sim SN_{q_a}(0, S_a, \delta_a I_{q_a}) \quad (2.6)$$

Assim, a densidade condicional do vetor  $Y$ , a verossimilhança, é dada por:

$$f(y | \beta, \mathbf{a}, \sigma_\varepsilon^2, \delta_\varepsilon) = 2^n \phi_n(y | X\beta + Z\mathbf{a}, (\sigma_\varepsilon^2 + \delta_\varepsilon^2) I_n) \quad (2.7)$$

$$\Phi_n\left(\frac{\delta_\varepsilon}{\sigma_\varepsilon^2 + \delta_\varepsilon^2} (y - X\beta - Z\mathbf{a}) | 0, \frac{\delta_\varepsilon}{\sigma_\varepsilon^2 + \delta_\varepsilon^2} I_n\right) \quad (2.8)$$

Seja  $\mathbf{y} = (y_1, \dots, y_n)^T$  e  $\mathbf{t} = (t_1, \dots, t_n)^T$ . Segue que a função de verossimilhança para  $\theta = (\mu, \sigma^2, \Delta, t)^T$  é dada por:

$$L(\theta) = \frac{1}{\Sigma^{n/2}} \exp\left(-\frac{1}{2\Sigma} \sum_{i=1}^n (y_i - \mu - \Delta t_i)^2\right) \exp\left(-\frac{1}{2} \sum_{i=1}^n t_i^2\right) \prod_{i=1}^n I_{t_i > 0} \quad (2.9)$$

### 3.8 Distribuições a priori para os parâmetros do modelo

Quando se faz inferência bayesiana, deve se especificar as distribuições a priori para todos os parâmetros desconhecidos do modelo para garantir as distribuições a posteriori próprias. Assim, considera-se uma distribuição normal multivariada para o vetor de

parâmetros  $\beta$ , com densidade:

$$\pi(\beta|\beta_0, S_\beta) = \frac{1}{2\pi^{p/2}\sqrt{|S_\beta|}} \exp\left[-\frac{1}{2}(\beta - \beta_0)^T S_\beta^{-1}(\beta - \beta_0)\right] \quad (3.0)$$

$$\beta \sim N_p(\beta_0, S_\beta) \quad (3.1)$$

Para o parâmetro  $t$ , será considerada uma distribuição normal multivariada com o vetor de parâmetros  $t$ , com densidade:

$$\pi(t|\mu_t, I\sigma_t^2) = \frac{1}{2\pi^{p/2}\sqrt{|I\sigma_t^2|}} \exp\left[-\frac{1}{2}(t - \mu_t)^T (\sigma_t^2)^{-1}(t - \mu_t)\right] \quad (3.2)$$

$$t \sim N(\mu_t, I\sigma_t^2) \quad (3.3)$$

Para o parâmetro  $\Delta$ , foi considerada uma distribuição normal multivariada com o vetor de parâmetros  $\Delta$  com densidade:

$$\pi(\Delta|\mu_\Delta, I\sigma_\Delta^2) = \frac{1}{2\pi^{p/2}\sqrt{|I\sigma_\Delta^2|}} \exp\left[-\frac{1}{2}(\Delta - \mu_\Delta)^T I(\sigma_\Delta^2)^{-1}(\Delta - \mu_\Delta)\right] \quad (3.4)$$

$$\Delta \sim N(\mu_\Delta, I\sigma_\Delta^2) \quad (3.5)$$

Para o parâmetro  $\mu$ , foi considerada uma distribuição normal multivariada com o vetor de parâmetros  $\mu$ , com densidade:

$$\pi(\mu|\mu_\mu, I\sigma_\mu^2) = \frac{1}{2\pi^{p/2}\sqrt{|I\sigma_\mu^2|}} \exp\left[-\frac{1}{2}(\mu - \mu_\mu)^T I(\sigma_\mu^2)^{-1}(\mu - \mu_\mu)\right] \quad (3.6)$$

$$\mu \sim N(\mu_\mu, I\sigma_\mu^2) \quad (3.7)$$

Para o parâmetro  $\sigma^2$ , foi considerada uma distribuição normal multivariada com o vetor de parâmetros  $\sigma^2$ , com densidade:

$$\sigma_e^2 \sim X^{-2} \text{escalada} (v_e, s_e^2) \quad (3.8)$$

Assim como, a distribuição a priori para a variância genotípica, tem-se

$$\sigma_g^2 \sim X^{-2} \text{escalada}(v_g, S_g^2) \quad (3.9)$$

Para os componentes de variância  $\sigma_e^2$  e  $\sigma_g^2$ , são atribuídas distribuições a priori qui quadrado escaladas invertidas:

$$P(\sigma^2|v, S^2) = \left(\frac{vS^2}{2}\right)^{v/2} \frac{\exp\left(-\frac{vS^2}{2\sigma^2}\right)}{\Gamma\left(\frac{v}{2}\right) (\sigma^2)^{1+\frac{v}{2}}} \quad (4.0)$$

$$\sigma^2 \sim \chi^{-2} \text{escalada}(v_e, S_e^2) \quad (4.1)$$

Onde  $v$  representa os graus de liberdade,  $S^2$  é o parâmetro de escala e  $\Gamma(\cdot)$  é a função gama. Os graus de liberdade  $v_e$  e  $v_g$  bem como os parâmetros de escala ( $S_e^2$  e  $S_g^2$ ) para  $\sigma_e^2$  e  $\sigma_g^2$ , respectivamente, são considerados iguais a zero, resultando em  $\frac{1}{\sigma_e^2}$  e  $\frac{1}{\sigma_g^2}$ , consideradas também como distribuições a priori não informativas. Como a priori atribuída a  $\sigma_g^2$  não é informativa, a incerteza em relação ao parâmetro genético é determinada pela função de verossimilhança, ou seja, a partir dos dados experimentais.

A densidade conjunta a posteriori é:

$$p(\mu, t, \Delta, \sigma^2|y) = \frac{1}{\Sigma^{n/2}} (\sigma^2)^{-(v_e+1)} \exp\left(-\frac{1}{2\Sigma} \sum_{i=1}^n (y_i - \mu - \Delta t_i)^2\right) \exp\left(-\frac{1}{2} \sum_{i=2}^n t_i^2\right) \prod_{i=1}^n I_{t_i > 0} \\ \exp\left(-\frac{1}{2} \sum_{i=1}^n (y_i - \mu - \Delta t_i)^2 - \frac{(\mu - \mu_\mu)^2}{2\sigma_\mu^2} - \frac{(\Delta - \mu_\Delta)^2}{2\sigma_\Delta^2} - \frac{(t - \mu_t)^2}{2\sigma_t^2} - \frac{(S_e^2)^2}{\sigma^2}\right) \quad (4.2)$$

Foi calculada a convergência das cadeias para os parâmetros estimados, pela teoria das Cadeias de Markov, utilizando o pacote *BOA* (*Bayesian Output Analysis*), pelo software *R*.

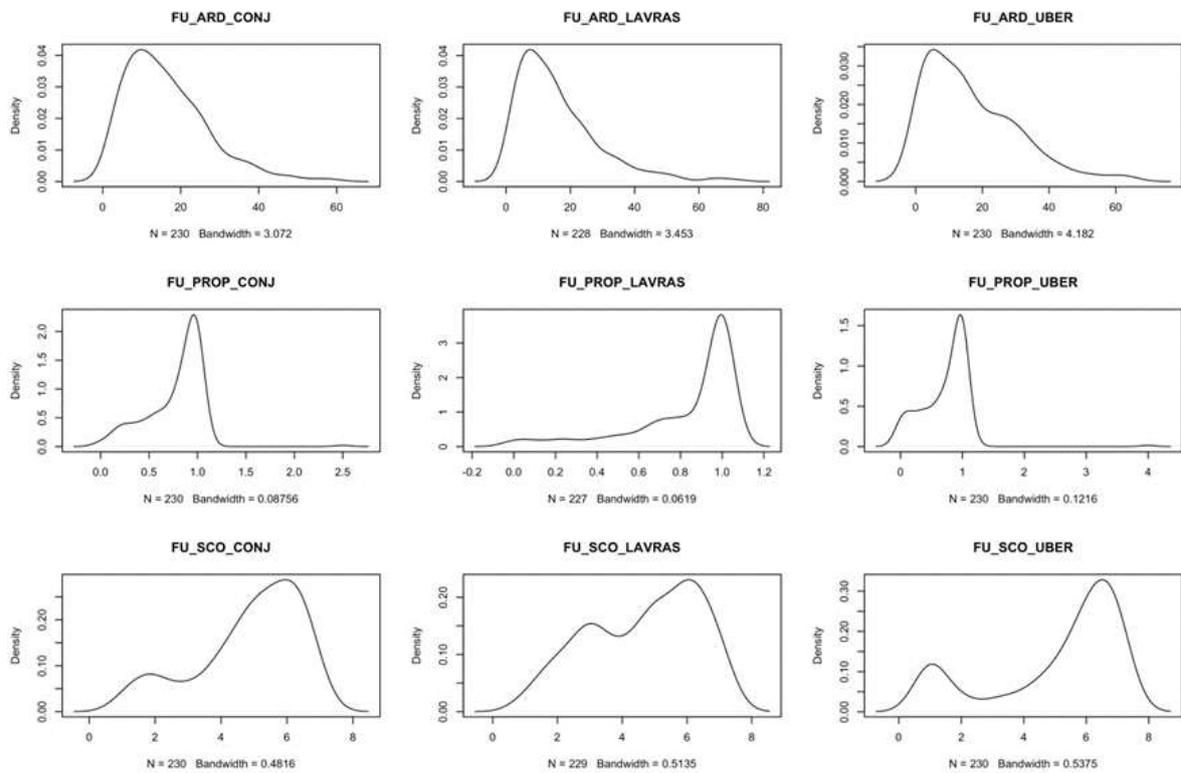
## 4 RESULTADOS E DISCUSSÃO

### 4.1 Dados com distribuição assimétrica

Na análise dos dados de doenças a distribuição apresentou assimetria. Como pode ser observado na Figura 1, a doença *fusarium* apresentou distribuição assimétrica à direita para o caractere grãos ardidos, enquanto que para o caractere escore, observa-se assimetria à esquerda. Em relação à proporção de espigas com sintomas, pode-se observar assimetria à esquerda para o ambiente de Lavras, e assimetria à direita para o ambiente de Uberlândia. Os gráficos de densidade foram obtidos utilizando a função *density*, no software R. De acordo com o teste de normalidade Shapiro Wilk, os dados originais referentes a grãos ardidos não seguem distribuição normal ( $p < 0,05$ ).

$$W = 0.92235, p\text{-value} = 1.27e-09$$

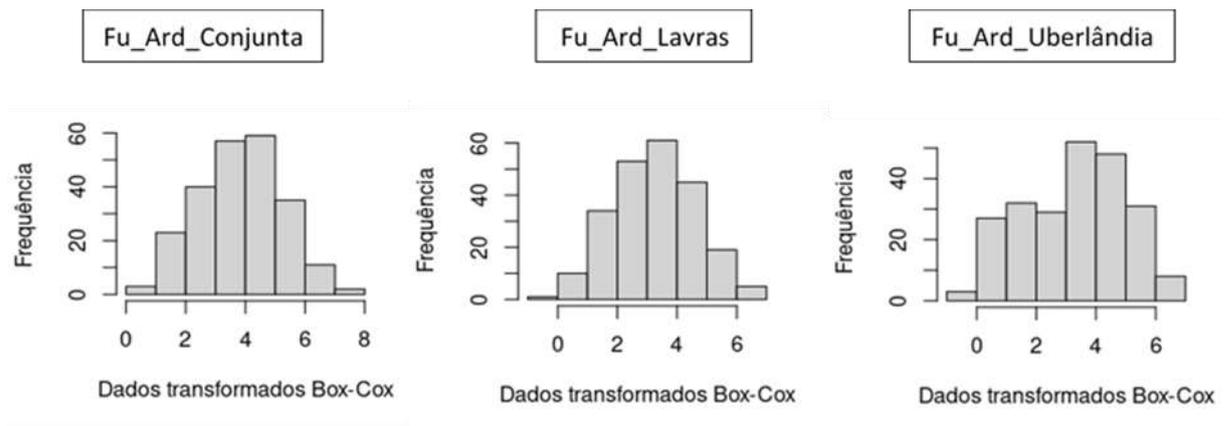
Figura 1 - Gráficos com distribuição assimétrica para os dados de *Fusarium verticillioides*.



Fonte: Do autor (2021).

Com a transformação Box-Cox, verificou-se que os dados de grãos ardidos referentes à análise conjunta e de Lavras passaram a apresentar distribuição simétrica, mas na análise de Uberlândia, a transformação reduziu a assimetria, porém os dados continuaram não sendo normalmente distribuídos, como pode ser observado na Figura 2.

Figura 2 - Gráficos com os dados transformados para o caráter grãos ardidos, na análise conjunta, e na análise individual de Lavras e Uberlândia.

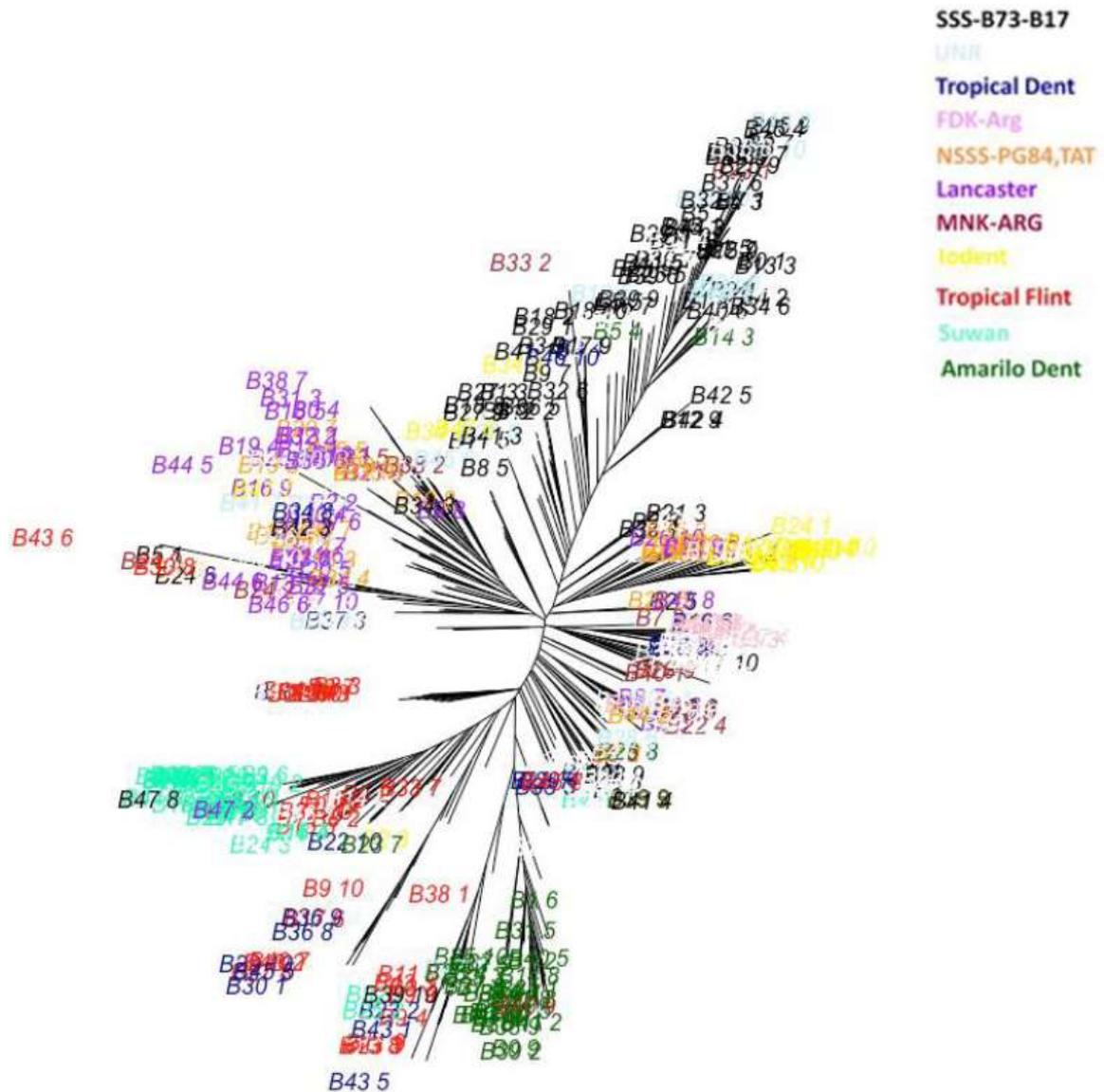


Fonte: Do autor (2021).

## 4.2 Caracterização das linhagens

As linhagens utilizadas nesse trabalho vieram de diferentes ambientes, sendo algumas linhagens de regiões tropicais e outras de regiões temperadas. Essa diferença de origem das linhagens pode ser uma das causas da distribuição assimétrica para as doenças de podridão de espiga. Por meio da árvore filogenética abaixo, construída pelo método *neighbor-joining*, pode ser observado que as linhagens utilizadas no estudo de seleção foram agrupadas em diferentes subgrupos (DE JONG; GUILHERME, et al., 2018). As linhagens derivadas do grupo *Stiff Stalk Synthetic* (SSS-B73-B17) e as linhagens pertencentes ao grupo UNR foram agrupadas na parte superior da árvore filogenética. Já as linhagens do subgrupo temperado, o qual inclui as linhagens pertencentes aos subgrupos FDK-Arg, *Non Stiff Stalk Synthetic* (NSSS-PG84,TAT), *Lancaster*, MNK-ARG e *Iodent* foram agrupadas na parte mediana da árvore. E, na parte inferior da figura, podem ser observadas as linhagens dos grupos *Tropical Dent*, *Tropical Flint*, *Suwan* e *Amarilo Dent*, as quais pertencem ao subgrupo tropical de linhagens.

Figura 3 - Árvore filogenética construída pelo método Neighbor-Joining mostrando os diferentes *backgrounds* das linhagens utilizadas no estudo de seleção genômica.



Fonte: (DE JONG; GUILHERME, et al., 2018)

#### 4.3 Predição por G-BLUP e modelo misto normal assimétrico

A predição dos fenótipos foi feita utilizando o total de marcadores DArTs. Foram obtidos os parâmetros de variância aditiva, variância residual, beta e a herdabilidade. A variância aditiva refere-se à parte herdável da variância genotípica do indivíduo, sendo de grande importância na seleção de genótipos superiores. A partir da relação entre a variância aditiva e a variância fenotípica obtém-se a herdabilidade. De acordo com a Tabela 1 em relação a grãos ardidos, pelo método G-BLUP, em Uberlândia foi o local que se obteve o

maior valor da variância aditiva, de 96.62, com herdabilidade de 0.41. Pode-se observar uma grande diferença entre a variância aditiva pelo método GBLUP e pelo Modelo Assimétrico Bayesiano, onde a variância aditiva em Uberlândia foi de 3,16 e a herdabilidade foi 0,83. O modelo misto assimétrico apresentou maior herdabilidade em comparação ao GBLUP, para o caractere grãos ardidos, devido à variância residual que foi menor em relação à variância aditiva, nos dois locais. A correlação entre o valor genético observado e o valor predito refere-se à capacidade preditiva do modelo, mas além do modelo, a correlação também sofre influência da herdabilidade e da arquitetura genética do caráter e da densidade dos marcadores. De acordo com as tabelas abaixo, a correlação foi proporcional ao valor da herdabilidade para todos os caracteres em estudo, nos dois modelos, sendo maior no modelo misto normal assimétrico, o que mostra a eficiência do modelo em prever os valores genéticos. Para as três análises, Beta ( $\beta$ ) e os parâmetros de variância foram maiores no modelo sem assimetria. O parâmetro de assimetria  $\Delta$  apresentou assimetria positiva também nas três análises, o que indica uma assimetria à direita para o caractere grãos ardidos.

Tabela 1 - Variância aditiva (Va), Variância Residual (Ve), Beta, Herdabilidade ( $h^2$ ), Correlação e Delta obtidos pelo GBLUP e Modelo Misto Normal Assimétrico Bayesiano (MAB) para grãos ardidos, nas análises de Lavras, Uberlândia e Conjunta.

<b>Grãos ardidos</b>						
Lavras						
	Va	Ve	$\beta$	$h^2$	Correlação	Delta ( $\Delta$ )
<b>GBLUP</b>	56.03	129.54	15.89	0.3	0.35	
<b>MAB</b>	1.88	1.09	1.57	0.63	0.99	20.14
Uberlândia						
	Va	Ve	$\beta$	$h^2$	Correlação	Delta ( $\Delta$ )
<b>GBLUP</b>	96.62	130.87	16.65	0.41	0.39	
<b>MAB</b>	3.16	0.63	1	0.83	0.99	20.54
Conjunta						
	Va	Ve	$\beta$	$h^2$	Correlação	Delta ( $\Delta$ )
<b>GBLUP</b>	67.83	72.01	16.34	0.48	0.42	
<b>MAB</b>	9.5	1.48	3.73	0.86	0.99	14.36

Fonte: Do autor (2021).

Em relação ao caractere proporção de espigas doentes (Tabela 2) não houve uma diferença significativa entre os dois modelos para os parâmetros obtidos, podendo-se observar uma herdabilidade próxima a 0,6 nos dois ambientes, e uma alta correlação nos dois modelos, para as duas análises individuais e a conjunta. O parâmetro de assimetria foi menor, comparado ao caractere grãos ardidos, e na análise conjunta foi observada uma assimetria

negativa, à esquerda.

Tabela 2 - Variância aditiva (Va), Variância Residual (Ve), Beta, Herdabilidade ( $h^2$ ), Correlação e Delta obtidos pelo GBLUP e Modelo Misto Normal Assimétrico Bayesiano (MAB) para proporção de espigas doentes, nas análises de Lavras, Uberlândia e Conjunta.

<b>Proporção de espigas doentes</b>						
Lavras						
	Va	Ve	$\beta$	$h^2$	Correlação	Delta ( $\Delta$ )
<b>GBLUP</b>	0.03	0.02	0.82	0.56	0.62	
<b>MAB</b>	0.08	0.04	0.75	0.64	0.72	0.07
Uberlândia						
	Va	Ve	$\beta$	$h^2$	Correlação	Delta ( $\Delta$ )
<b>GBLUP</b>	0.07	0.04	0.71	0.65	0.64	
<b>MAB</b>	0.13	0.06	0.82	0.68	0.74	0.1
Conjunta						
	Va	Ve	$\beta$	$h^2$	Correlação	Delta ( $\Delta$ )
<b>GBLUP</b>	0.04	0.02	0.76	0.63	0.69	
<b>MAB</b>	0.07	0.05	0.81	0.61	0.74	-0.04

Fonte: Do autor (2021).

Em relação ao caracter escore, pode-se observar uma pequena diferença entre as variâncias obtidas nas três análises pelos dois modelos, mas ainda assim a herdabilidade foi superior pelo modelo misto normal assimétrico, assim como para os outros caracteres em estudo, obtendo uma maior correlação entre o valor observado e o valor genético predito. Na análise conjunta e no ambiente de Uberlândia foi observada uma assimetria negativa para esse caracter (Tabela 3).

Tabela 3 - Variância aditiva ( $V_a$ ), Variância Residual ( $V_e$ ), Beta, Herdabilidade ( $h^2$ ), Correlação e Delta obtidos pelo GBLUP e Modelo Misto Normal Assimétrico Bayesiano (MAB) para escore, nas análises de Lavras, Uberlândia e Conjunta.

<b>Escore</b>						
Lavras						
	$V_a$	$V_e$	$\beta$	$h^2$	Correlação	Delta ( $\Delta$ )
<b>GBLUP</b>	1.09	2.13	4.64	0.33	0.38	
<b>MAB</b>	2.63	1.22	3.78	0.68	0.74	1.14
Uberlândia						
	$V_a$	$V_e$	$\beta$	$h^2$	Correlação	Delta ( $\Delta$ )
<b>GBLUP</b>	2.66	1.89	5.07	0.58	0.61	
<b>MAB</b>	2.09	0.54	6.66	0.79	0.94	-2.17
Conjunta						
	$V_a$	$V_e$	$\beta$	$h^2$	Correlação	Delta ( $\Delta$ )
<b>GBLUP</b>	1.45	1.16	4.85	0.55	0.6	
<b>MAB</b>	0.85	0.29	6.13	0.74	0.94	-1.64

Fonte: Do autor (2021).

Os caracteres grãos ardidos e escore apresentaram maior herdabilidade em relação à proporção de espigas doentes. Em estudos de seleção genômica para *Stenocarpella maydis*, patógeno que também causa podridão de espiga em milho, também foram encontrados resultados semelhantes, com maior herdabilidade para grãos ardidos em comparação aos outros caracteres (DOS SANTOS et al., 2016). Mas isso não descarta as análises feitas com proporção de espigas e escore, visto que são sintomas relevantes para a identificação da doença.

Na resistência a doenças de plantas, há plantas que apresentam resistência controlada por muitos genes, que é poligênica, e resistência controlada por um ou poucos genes, que é oligogênica. No estudo de genética quantitativa, sabe-se que caracteres controlados por muitos genes apresentam uma distribuição simétrica. Na literatura podem ser encontrados trabalhos que relatam que a resistência à podridão da espiga é de herança quantitativa (BRAUNER et al., 2017). Estudos de mapeamento com populações biparentais mostram que a resistência à podridão da espiga causada por *fusarium* é controlada por genes menores com pequenos efeitos, que variam entre ambientes e não são consistentes entre as populações (MESTERHÁZY et al., 2012). Mas mesmo sendo de caráter quantitativo, pode ser encontrada assimetria na distribuição dos dados, como foi encontrado nesse trabalho.

O modelo GBLUP assume que os efeitos dos genes possuem distribuição normal, com variância constante, diferente da distribuição que os dados de podridão de espiga apresentaram nesse trabalho. Em casos de doenças de plantas é comum a assimetria dos

dados, e o modelo assimétrico com inferência bayesiana lida bem com essa assimetria. Na literatura há trabalhos que relatam que a estatística GBLUP apresenta uma maior correlação quando se tem uma alta densidade de marcadores e quando se tem muitos genes de pequenos efeitos, comparada à estatística bayesiana, em que se pode apresentar uma alta correlação mesmo para uma baixa densidade de marcadores (ZHANG et al., 2019). Pérez Rodrigues et al. (2018) concluiu, em seu trabalho, que o modelo de regressão genômica bayesiano lidou bem com a assimetria dos dados fenotípicos comparado ao modelo de regressão clássico. Nesse trabalho, em que os dados são obtidos de análises de doenças, o modelo assimétrico bayesiano apresentou alta correlação para os caracteres, sendo mais interessante do que utilizar o GBLUP ou fazer a transformação dos dados.

## 5 CONCLUSÕES

Os caracteres grãos ardidos e escore foram os que apresentaram maior herdabilidade, comparados aos outros caracteres, podendo ser utilizados para selecionar indivíduos resistentes ao patógeno *Fusarium verticillioides*.

As análises para grãos ardidos apresentaram maior assimetria, comparada aos outros caracteres. Em estudos de resistência a doenças é comum encontrar distribuição assimétrica dos dados, mesmo em casos de herança quantitativa. Nesses casos é importante considerar um modelo que lida bem com a assimetria.

O modelo misto normal assimétrico apresentou boa correlação dos dados, ou seja, apresentou uma boa predição. Assim, pode-se concluir que este é um modelo que lida bem com a assimetria.

## REFERÊNCIAS

- ARELLANO-VALLE, R. B.; BOLFARINE, H.; LACHOS, V. H. Skew-normal linear mixed models. **Journal of data science**, v. 3, n. 4, p. 415-438, 2005.
- ATLIN, G. N. et al. Selection response in subdivided target regions. **Crop science**, v. 40, n. 1, p. 7-13, 2000.
- AZZALINI, A. A class of distributions which includes the normal ones. **Scandinavian journal of statistics**, p. 171-178, 1985.
- AZZALINI, A.; CAPITANIO, A. Statistical applications of the multivariate skew normal distribution. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, v. 61, n. 3, p. 579-602, 1999.
- BAO, Y. et al. Potential of association mapping and genomic selection to explore PI 88788 derived soybean cyst nematode resistance. **The Plant Genome**, v. 7, n. 3, p. plantgenome 2013.11.0039, 2014.
- BIAGIOTTI, D. "Associação e seleção genômica ampla em ovinos santa inês para características relacionadas a resistência à endoparasitas." 2018.
- BORGES, L. C.; FERREIRA, D. F. Poder e taxas de erro tipo I dos testes Scott-Knott, Tukey e Student-Newman-Keuls sob distribuições normal e não normais dos resíduos. **Revista de matemática e estatística**, v. 21, n. 1, p. 67-83, 2003.
- BOX, G. E.; COX, D. R. An analysis of transformations. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 26, n. 2, p. 211-243, 1964.
- BRAUNER, P. C. et al. Low validation rate of quantitative trait loci for Gibberella ear rot resistance in European maize. **Theoretical and applied genetics**, v. 130, n. 1, p. 175-186, 2017.
- BULLERMAN, L. B.; TSAI, W. Y. J. Incidence and Levels of Fusarium moniliforme, Fusarium proliferatum and Fumonisin in Corn and Corn-Based Foods and Feeds1. **Journal of food Protection**, v. 57, n. 6, p. 541-546, 1994.
- BULLERMAN, L. B.; DRAUGHON, F. Ann. Fusarium moniliforme and Fumonisin symposium. **Journal of Food Protection**, v. 57, n. 6, p. 28, 1994.
- CAMARINHA FILHO, J. A. Nota Metodológica sobre Modelos Lineares Mistos. **Notas de aula, Universidade Federal do Paraná, Departamento de Estatística**, 2002.
- CAMPOS, A. M. **Uma abordagem bayesiana para modelos não lineares na presença de assimetria e heteroscedasticidade**. 2011. Tese de Doutorado. Universidade de São Paulo.
- CANTOR, R. M.; LANGE, K.; SINSHEIMER, J. S. Priorizing GWAS results: A review of statistical methods and recommendations for their application. **The American Journal of Human Genetics**, v. 86, n. 1, p. 6-22, jan. 2010.

CASELA, C. R.; FERREIRA, A. da S.; PINTO, N. F. J. de A. Doenças na cultura do milho. **Circular técnica 83, Embrapa Milho e Sorgo (CNPMS)**, Sete Lagoas, 14p. dez. 2006.

CHAMBERLAIN, W. J. et al. Levels of fumonisin B1 in corn naturally contaminated with aflatoxins. **Food and Chemical Toxicology**, v. 31, n. 12, p. 995-998, 1993.

CHEN, Z.; CHEN, J. Tournament screening cum EBIC for feature selection with high-dimensional feature spaces. **Science in China Series A: Mathematics**, v. 52, n. 6, p. 1327-1341, 2009.

CLEMENTS, M. J. et al. Evaluation of inoculation techniques for Fusarium ear rot and fumonisin contamination of corn. **Plant Disease**, v. 87, n. 2, p. 147-153, 2003.

CONAB - Companhia Nacional de Abastecimento. Acompanhamento da safra brasileira de grãos, v. 8 - Safra 2020/21, n. 4 - Quarto levantamento, janeiro 2021.

COSTA, R. V.; CASELA, C. R.; COTA, V. L. Podridões do colmo e das raízes. **Agência de informação-EMBRAPA**, 2005.

COSTA, R. V.; COTA, L. V.; Silva, D. D. Doenças Causadas por Fungos do Gênero *Stenocarpella* spp. (*Diplodia* spp.) em Milho. Brasília: Circular Técnica, Embrapa, 2013.

DE JONG, G. et al. Genome-wide association analysis of ear rot resistance caused by *Fusarium verticillioides* in maize. **Genomics**, v. 110, n. 5, p. 291-303, 2018.

DE LA CRUZ, R.; BRANCO, M. D. Bayesian analysis for nonlinear regression model under skewed errors, with application in growth curves. **Biometrical Journal: Journal of Mathematical Methods in Biosciences**, Weinheim, v. 51, n. 4, p. 588-609. 2009.

DE RESENDE, M. D. V. et al. Seleção genômica ampla (GWS) e maximização da eficiência do melhoramento genético. **Pesquisa florestal brasileira**, n. 56, p. 63-63, 2008.

DOS SANTOS, J. P. R. et al. Genomic selection to resistance to *Stenocarpella maydis* in maize lines using DArTseq markers. **BMC genetics**, v. 17, n. 1, p. 1-10, 2016.

FANTIN, G. M.; DUARTE, A. P. Manejo de doenças na cultura do milho safrinha. Campinas: **Instituto Agrônomo**, 99p. 2009.

FILHO, D. F. et al. Tournaments between markers as a strategy to enhance genomic predictions. **PloS one**, v. 14, n. 6, p. e0217283, 2019.

FRITSCHÉ-NETO, R. et al. Seleção genômica ampla e novos métodos de melhoramento do milho. **Revista Ceres**, v. 59, n. 6, p. 794-802, 2012.

DE JONG, Guilherme et al. Genome-wide association analysis of ear rot resistance caused by *Fusarium verticillioides* in maize. **Genomics**, v. 110, n. 5, p. 291-303, 2018.

GUPTA, P. K.; RUSTGI, S.; KULWAL, P. L. Linkage disequilibrium and association studies in higher plants: Present status and future prospects. **Plant molecular Biology**, Amsterdam, n. 57, p. 461- 485, jan. 2005.

HOLLAND, J. B.; NYQUIST, W. E.; CERVANTES-MARTÍNEZ, C. T. Estimating and interpreting heritability for plant breeding: an update. **Plant breeding reviews**, v. 22, 2003.

JACCOUD, D. et al. Diversity arrays: a solid state technology for sequence information independent genotyping. *Nucleic Acids Research*, **Oxford**, v. 29, n. 4, p. 1-7, Feb. 2001.

JEFFREYS, H. **Theory of probability**. Oxford: Clarendon, 1961. 325p.

JÚNIOR, O. et al. **Associação genômica ampla (GWAS) aplicada a características reprodutivas de novilhas da raça Nelore**. 2017. Tese de Doutorado. Universidade de São Paulo.

LANUBILE, A. et al. Molecular basis of resistance to Fusarium ear rot in maize. **Frontiers in plant science**, v. 8, p. 1774, 2017.

LEX, J. et al. Genome-wide association studies of agronomic and quality traits in a set of German winter barley (*Hordeum vulgare* L.) cultivars using Diversity Arrays Technology (DArT). **Journal of applied genetics**, v. 55, n. 3, p. 295-305, 2014.

LI, H.; WANG, J.; BAO, Z. A novel genomic selection method combining GBLUP and LASSO. **Genetica**, v. 143, n. 3, p. 299-304, 2015.

LI, Z.; MÖTTÖNEN, J.; SILLANPÄÄ, M. J. A robust multiple-locus method for quantitative trait locus analysis of non-normally distributed multiple traits. **Heredity**, v. 115, n. 6, p. 556-564, 2015.

MARASAS, W. F. Fumonins: history, world-wide occurrence and impact. In: JACKSON, L. S.; De VRIES, J. W.; BULLERMAN, L. B. (Eds.). *Fumonins in food*. New York: Plenum Press, 1996. P. 1-17.

MARTIN, T. N.; STORCK, L. Análise das pressuposições do modelo matemático em experimentos agrícolas no delineamento blocos ao acaso. In: MARTIN, T. N.; ZIECH, M. F. (org). **Sistemas de Produção Agropecuária**. Curitiba: UTFPR. p. 177-196. 2008.

MESTERHAZY, A.; LEMMENS, M.; REID, L. M. Breeding for resistance to ear rots caused by Fusarium spp. in maize—a review. **Plant Breeding**, v. 131, n. 1, p. 1-19, 2012.

MEUWISSEN, T. H.; HAYES, B. J.; GODDARD, M. E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, v. 157, n. 4, p. 1819-1829, 2001.

MÖHRING, J.; PIEPHO, H.-P. Comparison of weighting in two-stage analysis of plant breeding trials. **Crop Science**, v. 49, n. 6, p. 1977-1988, 2009.

MUSSER, S. M.; PLATTNER, R. D. Fumonisin composition in cultures of *Fusarium moniliforme*, *Fusarium proliferatum* and *Fusarium nygamai*. **Journal of Agricultural and Food Chemistry**, Easton, v. 45, n. 4, p. 1169-1173, 1997.

NORRED, W. P.; VOSS, K. A. Toxicity and role of fumonisins in animal diseases and human esophageal cancer. **Journal of Food Protection**, v. 57, n. 6, p. 522-527, 1994.

ODRIOZOLA, E. et al. Diplodia maydis: a cause of death of cattle in Argentina. **New Zealand Veterinary Journal**, v. 53, n. 2, p. 160-161, 2005.

ORABI, J.; JAHOR, A.; BACKES, G. Changes in allelic frequency over time in European bread wheat (*Triticum aestivum* L.) varieties revealed using DArT and SSR markers. **Euphytica**, v. 197, n. 3, p. 447-462, 2014.

PARADIS, E.; CLAUDE, J.; STRIMMER, K. APE: analyses of phylogenetics and evolution in R language. **Bioinformatics**, v. 20, n. 2, p. 289-290, 2004.

PARK, T.; CASELLA, G. The bayesian lasso. **Journal of the American Statistical Association**, v. 103, n. 482, p. 681-686, 2008.

PATTERSON, H. D.; THOMPSON, R. Recovery of inter-block information when block sizes are unequal. **Biometrika**, v. 58, n. 3, p. 545-554, 1971.

PAULINO, C. D.; TURKMAN, M. A. A.; MURTEIRA, B. Estatística Bayesiana. Lisboa: Fundação Calouste Gulbenkian, 2003. 447p.

PEARSON, T. A.; MANOLIO, T. A. How to interpret a genome-wide association study. **Jama**, v. 299, n. 11, p. 1335-1344, 2008.

PEWSEY, A. Problems of inference for Azzalini's skewnormal distribution. **Journal of applied statistics**, v. 27, n. 7, p. 859-870, 2000.

PÉREZ-BRITO, D. et al. QTL mapping of *Fusarium moniliforme* ear rot resistance in high land maize, Mexico. **Agrociencia**, v. 35, p. 181-196, 2001.

PÉREZ-RODRÍGUEZ, P. et al. A Bayesian genomic regression model with skew normal random errors. **G3: Genes, Genomes, Genetics**, v. 8, n. 5, p. 1771-1785, 2018.

PINTO, N. F. J. de A. Grãos ardidos em milho. **Embrapa Milho e Sorgo-Circular Técnica (INFOTECA-E)**, 2005.

PROCTOR, R. H. et al. Fumonisin production in the maize pathogen *Fusarium verticillioides*: genetic basis of naturally occurring chemical variation. **Journal of Agricultural and Food Chemistry**, v. 54, n. 6, p. 2424-2430, 2006.

R CORE TEAM. R: A language and environment for statistical computing. **R Foundation for Statistical Computing**, Vienna, Austria, 2016.

REID, L. M. et al. Effect of inoculation time and point of entry on disease severity in *Fusarium graminearum*, *Fusarium verticillioides*, or *Fusarium subglutinans* inoculated maize ears1. **Canadian journal of plant pathology**, v. 24, n. 2, p. 162-167, 2002.

REIS, E. M.; CASA, R. T.; BRESOLIN, A. C. R. **Manual de diagnose e controle de doenças do milho**. Lages: Graphel, 44p., 2004.

RESENDE, M. D. V. et al. Seleção Genômica Ampla (GWS) e maximização da eficiência do melhoramento genético. **Pesquisa Florestal Brasileira**, v. 56, p. 63-77, 2008.

ROBERTSON, L. A.; PAYNE, G. A.; HOLLAND, J. B. “Marker assisted breeding for resistance to mycotoxin contamination,” in *Aflatoxin and Food Safety*, ed. H. K. Abbas (New York, NY: Marcel Dekker, Inc.), p. 423–435, 2005.

ROBERTSON-HOYT, L. A. et al. QTL mapping for fusarium ear rot and fumonisin contamination resistance in two maize populations. **Crop Science**, v. 46, n. 4, p. 1734–1743, 2006.

ROSSI, R. M.; DOS SANTOS, L. A. Modelagem Bayesiana para curvas de crescimentos de codornas assumindo assimetria nos erros. **Semina: Ciências Agrárias**, v. 35, n. 3, p. 1637-1647, 2014.

SAITOU, N.; NEI, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. **Molecular Biology and Evolution**, v. 4, n. 4, p. 406-425, 1987.

SANSALONI, C. P. et al. A high-density Diversity Arrays Technology (DARt) microarray for genome-wide genotyping in Eucalyptus. **Plant Methods**, v. 6, n. 1, p. 1-11, 2010.

SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). **Biometrika**, v. 52, n. 3/4, p. 591-611, 1965.

SINGH, B. D.; SINGH, Ashok Kumar. **Marker-assisted plant breeding: principles and practices**. New Delhi: Springer India, 2015.

SOLLERO, B. P. et al. Seleção de Marcadores a partir do Estudo de Associação Genômica Ampla para Resistência ao Carrapato em Bovinos das Raças Hereford e Braford. Simpósio Brasileiro de Melhoramento Animal, 11., 2015, Santa Maria, RS. **Melhoramento animal da academia ao campo: uma parceria em construção: anais**. Santa Maria, RS: SBMA: UFSM, 2015.

STEEL, R. G. D; TORRIE, J. H.; DICKEY, D. A. **Principles and procedures of statistics a biometrical approach**. 3.ed. New York: McGraw-Hill, 1997. 666p.

STORCK, L. et al. **Experimentação vegetal**. Santa Maria: UFSM. 200p. 2011.

TRENTO, S. M.; IRGANG, H. H.; REIS, E. M. Efeito da rotação de culturas, da monocultura e da densidade de plantas na incidência de grãos ardidos em milho. **Fitopatologia Brasileira**, v. 27, n. 4, p. 609-613, 2002.

VANRADEN, P. M. et al. Invited review: Reliability of genomic predictions for North American Holstein bulls. **Journal of dairy science**, v. 92, n. 1, p. 16-24, 2009.

VARONA, L. et al. Bayesian analysis of quantitative traits using skewed distributions. **Genetics Research**, v. 90, n. 2, p. 179-190, 2008.

VIANA, G. Milho: novos sistemas de produção e busca por maiores produtividades provocam aumento da severidade das doenças. **Jornal Eletrônico da Embrapa Milho e Sorgo**, Sete Lagoas, nov. 2009.

VITEZICA, Z. G.; VARONA, L.; LEGARRA, A. On the additive and dominant variance and covariance of individuals within the genomic selection scope. **Genetics**, v. 195, n. 4, p. 1223-1230, 2013.

WENZL, P. et al. Diversity Arrays Technology (DArT) for whole-genome profiling of barley. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v.101, n. 26, p. 9915-9920, June 2004.

WENZL, P. et al. A high-density consensus map of barley linking DArT markers to SSR, RFLP and STS loci and agricultural traits. **BMC genomics**, v. 7, n. 1, p. 1-22, 2006.

YI, N.; XU, S. Bayesian LASSO for quantitative trait loci mapping. **Genetics**, v. 179, n. 2, p. 1045-1055, 2008.

ZAMBOLIM, L.; CASA, R. T.; REIS, E. M. Sistema plantio direto e doenças em plantas. **Fitopatologia Brasileira**, v. 25, n. 4, p. 585-595, 2000.

ZHANG, H. et al. Factors affecting the accuracy of genomic selection for agricultural economic traits in maize, cattle, and pig populations. **Frontiers in genetics**, v. 10, p. 189, 2019.