



UFG

**UNIVERSIDADE FEDERAL DE GOIÁS
ESCOLA DE AGRONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E
MELHORAMENTO DE PLANTAS**

**SELEÇÃO GENÔMICA AMPLA NO MELHORAMENTO
GENÉTICO DE ESPÉCIES POLIPLOIDES**

IVONE DE BEM OLIVEIRA

Orientador:

Prof. Dr. Alexandre Siqueira Guedes Coelho

Co-orientador:

Prof. Dr. Patricio Ricardo Muñoz

**TERMO DE CIÊNCIA E DE AUTORIZAÇÃO PARA DISPONIBILIZAR
VERSÕES ELETRÔNICAS DE TESES E DISSERTAÇÕES
NA BIBLIOTECA DIGITAL DA UFG**

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a Lei nº 9610/98, o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou *download*, a título de divulgação da produção científica brasileira, a partir desta data.

1. Identificação do material bibliográfico: ☐ Dissertação ☒ Tese

2. Identificação da Tese ou Dissertação:

Nome completo do autor: Ivone de Bem Oliveira

Título do trabalho: Seleção Genômica Ampla no Melhoramento Genético de Espécies Poliploides

3. Informações de acesso ao documento:

Concorda com a liberação total do documento ☐ SIM ☒ NÃO¹

Havendo concordância com a disponibilização eletrônica, torna-se imprescindível o envio do(s) arquivo(s) em formato digital PDF da tese ou dissertação.

Ivone de Bem Oliveira

Assinatura do(a) autor(a)²

Ciente e de acordo:

Alexandre Siqueira Mendes Costa

Assinatura do(a) orientador(a)²

Data: 07 /05 / 2018

¹ Neste caso o documento será embargado por até um ano a partir da data de defesa. A extensão deste prazo suscita justificativa junto à coordenação do curso. Os dados do documento não serão disponibilizados durante o período de embargo.

Casos de embargo:

- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro;
- Publicação da dissertação/tese em livro.

² A assinatura deve ser escaneada.

IVONE DE BEM OLIVEIRA

**SELEÇÃO GENÔMICA AMPLA NO MELHORAMENTO
GENÉTICO DE ESPÉCIES POLIPLOIDES**

Tese apresentada ao Programa de Pós-Graduação
em Genética e Melhoramento de Plantas, da
Universidade Federal de Goiás, como requisito
parcial à obtenção do título de Doutor em Genética
e Melhoramento de Plantas.

Orientador:

Prof. Dr. Alexandre Siqueira Guedes Coelho

Co-orientador:

Prof. Dr. Patricio Ricardo Muñoz

Goiânia, GO – Brasil
2018

Ficha de identificação da obra elaborada pelo autor, através do
Programa de Geração Automática do Sistema de Bibliotecas da UFG.

de Bem Oliveira, Ivone
SELEÇÃO GENÔMICA AMPLA NO MELHORAMENTO GENÉTICO
DE ESPÉCIES POLIPLOIDES [manuscrito] / Ivone de Bem Oliveira. -
2018.
CLXXIV, 174 f.: il.

Orientador: Prof. Alexandre Siqueira Guedes Coelho; co
orientador Patrício Ricardo Muñoz.

Tese (Doutorado) - Universidade Federal de Goiás, Escola de
Agronomia e Engenharia de Alimentos (EAEA), Programa de Pós
Graduação em Genética & Melhoramentos de Plantas , Goiânia, 2018.

Bibliografia. Anexos.
Inclui gráfico, tabelas.

1. GWS. 2. aloploidie. 3. autotetraploidie. 4. modelos
bayesianos. 5. cienciometria. I. Coelho, Alexandre Siqueira Guedes,
orient. II. Título.

CDU 633



SERVIÇO PÚBLICO FEDERAL
MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE GOIÁS
ESCOLA DE AGRONOMIA
**PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E
MELHORAMENTO DE PLANTAS**



ATA DE DEFESA PÚBLICA DE TESE DE IVONE DE BEM OLIVEIRA. Aos vinte e um dias do mês de maio do ano de dois mil e dezoito (21.05.2018), às 9h, na sala de reuniões do Setor de Melhoramento de Plantas, da Escola de Agronomia da Universidade Federal de Goiás, reuniram-se os membros da Banca Examinadora: Dr. Alexandre Siqueira Guedes Coelho – Orientador/Presidente, Dr. Lázaro José Chaves, Dr.^a Rosane Garcia Collevatti e Dr.^a Tereza Cristina de Oliveira Borba. Dois membros da banca: Dr. Patricio Ricardo Muñoz – Coorientador e Dr. Márcio Fernando Ribeiro de Resende Júnior e a doutoranda Ivone de Bem Oliveira participaram por videoconferência. Sob a presidência do orientador, em sessão pública, procedeu-se à avaliação da defesa da tese intitulada: **“Seleção genômica ampla no melhoramento genético de espécies poliploides”**, de autoria de **Ivone de Bem Oliveira**, discente do Programa de Pós-Graduação em Genética e Melhoramento de Plantas, no nível de Doutorado, da Universidade Federal de Goiás. A sessão foi aberta pelo presidente da Banca Examinadora, Dr. Alexandre Siqueira Guedes Coelho, que fez a apresentação formal dos membros da Banca. A palavra, a seguir, foi concedida ao autor da tese que, em 40 minutos, apresentou o seu trabalho. Terminada a apresentação, cada membro da Banca arguiu a doutoranda, tendo-se adotado o sistema de diálogo sequencial. Ao final, a banca reunida em separado procedeu à avaliação da defesa. A tese foi considerada APROVADA pela Banca Examinadora, cumprindo integralmente este requisito para fins de obtenção do título de DOUTOR EM GENÉTICA E MELHORAMENTO DE PLANTAS, pela Universidade Federal de Goiás, em conformidade com o estabelecido pela resolução nº 1403/2016, do Conselho de Ensino, Pesquisa, Extensão e Cultura da UFG (CEPEC/UFG), que regulamenta o Programa de Pós-Graduação em Genética e Melhoramento de Plantas. Para fins de publicação eletrônica, a doutoranda poderá efetuar as modificações eventualmente sugeridas pela Banca Examinadora e encaminhar à Secretaria do PGMP, respeitando-se o prazo máximo de 30 dias após a data da Defesa. A conclusão do curso e a emissão do diploma dar-se-ão em conformidade com o estabelecido pela Resolução CEPEC nº 1403/2016. Cumpridas as formalidades de pauta, às 13:40 a presidência da mesa encerrou esta sessão de defesa de Dissertação e, para constar eu, Sérgio Tadeu Sibov, Coordenador do PGMP, lavrei a presente Ata que depois de lida e aprovada, segue assinada pelos membros da Banca Examinadora, em duas vias de igual teor.


Dr. Alexandre Siqueira Guedes Coelho
Orientador/Presidente


Dr. Patricio Ricardo Muñoz
Coorientador


Dr. Lázaro José Chaves
Membro Interno


Dr. Márcio Fernando Ribeiro de Resende Júnior
Membro Externo


Dr.^a Rosane Garcia Collevatti
Membro Externo


Dr.^a Tereza Cristina de Oliveira Borba
Membro Externo

DEDICO este trabalho à minha
família por todo apoio, orientação e amor.

OFEREÇO este trabalho a
Deus e aos meus pais, Magali e Junior, por
serem meus exemplos de vida e de moral.

“Deus nos concede, a cada dia,
uma página de vida nova no livro do tempo.
Aquilo que colocarmos nela, corre por
nossa conta” (Chico Xavier).

AGRADECIMENTOS

Primeiramente a Deus, pela oportunidade, pelos desafios e pela coragem e determinação que me conferiu.

Aos meus pais, Magali e Júnior, por estarem sempre presentes, mesmo a quilômetros de distância, pelas conversas, pelas broncas, pelos mimos, pela confiança irrestrita, por me auxiliarem durante a formação intelectual e moral, por serem meus exemplos de vida e de determinação. Agradeço a vocês pela vida!

Ao meu irmão Gabriel, pelo amor, pela amizade, pelas risadas, pelas danças, pelas caminhadas, pelas conversas, pelos conselhos, mas especialmente por me ensinar o quanto a vida é simples de ser vivida.

À madrinha Nídia, por sempre acreditar, por me amar tanto e por todo o auxílio e carinho dispendidos durante toda a minha vida.

À minha avó Dulce por ser um anjo e por ter tornado o mundo mais doce e as dificuldades mais simples, devido à sua simples existência. Por todo o apoio e principalmente por todo o amor e por todas as ótimas lembranças.

À Teté (Onélia) por ser mais uma mãe, pelo carinho, pelo apoio, pelas conversas, por estar presente em todos os momentos e por ser uma das únicas pessoas que realmente me conhecem e me entendem.

À Ester e à Hizumi, simplesmente por existirem e por deixarem as batalhas da vida um pouco mais fáceis de serem vividas. Amo vocês!

À tia Cideli e às primas Tati e Jéssica, pelas alegrias, pela amizade e pelo amor.

Ao grupo maravilhoso de amigos que a pós me deu: Suzy, Guilherme, Mariane e Sabrina. Como é bom saber que existem pessoas como vocês na terra. Muito obrigada por cada pequeno detalhe que faz de vocês o que são. E obrigada por serem tão maravilhosos e especiais.

À amiga Clistiane que me mostrou que o mundo pode mudar e que basta força de vontade para ser uma pessoa maravilhosa!

Às amigas queridas Stela e Isabela, por todas as conversas e carinho dispendidos. É maravilhoso relembrar todas as experiências que tivemos nestes sete anos de pós-graduação. Obrigada pelas risadas, choros, almoços, festas surpresas, bolos,

conversas. Obrigada por tudo, vou sempre ter vocês no meu coração.

Ao meu amigo amado Rhewter por ser um exemplo tanto no sentido profissional como humano. Teu empenho, força, amizade, carinho e humanidade serão sempre meus guias no âmbito pessoal e profissional. Obrigada por ser sempre tão verdadeiro!

À minha família na Florida tão presentes e tão importantes durante este primeiro ano fora do Brasil: Chase, Diego, Ediene, Rodrigo, Fernanda, Aline e André. Obrigada por todo o auxílio, carinho e companheirismo. Obrigada por estarem presentes nesta fase tão importante da minha vida e por permitirem que eu fizesse parte da vida de vocês nesta existência. Obrigado a todos por serem pessoas maravilhosas. A ausência da família faz muita falta, mas perto de vocês as dificuldades se tornam mais leves.

Ao meu co-orientador Prof. Dr. Patricio R. Muñoz, obrigada por todas as oportunidades e por todo o crescimento profissional que estás despertando em mim. És um exemplo de orientador e de profissional.

Aos professores do Programa de Pós-graduação em Genética e Melhoramento de Plantas da EA-UFG, pelos conhecimentos compartilhados. Em especial para os professoras Dra. Mariana Pires Campos Telles e Dra. Rosane Collevatti. Obrigada por todos os ensinamentos e por serem um exemplo a ser seguido.

Ao Professor Dr. Alexandre Siqueira Guedes Coelho, pela orientação nestes sete anos de pós-graduação. Aprendi muito com o senhor!

A todos os colegas do Laboratório de Genética e Genômica de Plantas da EA-UFG, muito obrigada pela ajuda e por compartilharem desta batalha.

À Escola de Agronomia da Universidade Federal de Goiás, pela oportunidade e incentivo e ao Programa de Pós-Graduação em Genética e Melhoramento de Plantas, por todo auxílio oferecido.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES, pela concessão da Bolsa de Estudos.

SUMÁRIO

1	INTRODUÇÃO.....	14
2	REVISÃO DE LITERATURA.....	18
2.1	CANA-DE-AÇÚCAR E BLUEBERRY	18
2.2	POLIPLOIDIA: UM EXEMPLO EM CANA-DE-AÇÚCAR.....	20
2.3	SEGREGAÇÃO EM POLIPLOIDES: UM EXEMPLO EM CANA-DE-AÇÚCAR.....	22
2.4	USO DE MARCADORES <i>SINGLE</i> E <i>MULTIPLE DOSE</i> EM POLIPLOIDES: UM EXEMPLO EM CANA-DE-AÇÚCAR.....	25
2.5	DOSAGEM ALÉLICA EM POLIPLOIDES: UM EXEMPLO EM CANA-DE-AÇÚCAR.....	28
3	RESEARCH AND PARTNERSHIP IN STUDIES OF SUGARCANE USING MOLECULAR MARKERS: A SCIENTOMETRIC APPROACH	34
3.1	ABSTRACT	34
3.2	INTRODUCTION	35
3.3	MATERIAL AND METHODS.....	36
3.4	RESULTS	39
3.5	DISCUSSION.....	47
3.6	CONCLUSIONS	49
3.7	ACKNOWLEDGMENTS	50
3.8	REFERENCES	50
3.9	SUPPLEMENTAL INFORMATION.....	53
4	SELEÇÃO GENÔMICA AMPLA PARA CANA-DE-AÇÚCAR EM CONDIÇÕES DE ELEVADO DESEQUILÍBRIO DE LIGAÇÃO CONSIDERANDO EFEITOS DE ENDOGAMIA E ESTRUTURAÇÃO.....	83
4.1	RESUMO.....	83
4.2	INTRODUÇÃO	84
4.3	MATERIAL E MÉTODOS	85
4.4	RESULTADOS	91
4.5	DISCUSSÃO	111
4.6	CONSIDERAÇÕES FINAIS	115
4.7	REFERÊNCIAS	116
5	GENOMIC PREDICTION OF AUTOTETRAPLOIDS; THE INFLUENCE OF ALLELE DOSAGE, NON-ADDITIVE EFFECTS AND CONTINUOUS GENOTYPES.....	125
5.1	ABSTRACT	125
5.2	INTRODUCTION	125
5.3	MATERIAL AND METHODS.....	128
5.4	RESULTS	133
5.5	DISCUSSION.....	144
5.6	CONCLUSIONS	150
5.7	FUNDING	151
5.8	REFERENCES	151
5.9	SUPPLEMENTAL INFORMATION.....	159

6	CONCLUSÕES.....	162
7	REFERÊNCIAS BIBLIOGRÁFICAS.....	164

RESUMO

DE BEM OLIVEIRA, I. **Seleção genômica ampla no melhoramento genético de espécies poliploides**. 2018. 174f. Tese (Doutorado em Genética e Melhoramento de Plantas) – Escola de Agronomia, Universidade Federal de Goiás, Goiânia, 2018.¹

A seleção genômica (GS) é usada para predizer os efeitos simultâneos de ampla gama de marcadores moleculares distribuídos pelo genoma. Esta metodologia não pretende identificar os efeitos de locos individuais, mas sim capturar o efeito combinado dos polimorfismos causais envolvidos na variação genética de um caráter quantitativo. A aplicação dessa metodologia proporcionou ganhos genéticos significativos para espécies diploides. Apesar disso, sua aplicação em culturas poliploides continua sendo um desafio por várias razões, incluindo a complexa arquitetura genômica e a indisponibilidade de métodos e softwares. Além do potencial da GS para remodelar programas de melhoramento e obter ganhos de seleção mais rápidos, uma contribuição importante da presente pesquisa é discutir como considerar um modelo preditivo apropriado no contexto dos poliploides. Para isso aplicamos a inserção de diferentes fatores em modelos GS considerando a aplicação em duas espécies: Cana-de-açúcar (*Saccharum* spp.) e blueberry (*Vaccinium corymbosum*).

A cana-de-açúcar apresenta um dos genomas mais complexos entre as espécies cultivadas. As cultivares modernas têm origens interespecíficas, altos, variáveis e desconhecidos níveis de ploidia (frequente aneuploidia), alta heterozigosidade e grandes genomas. Para a blueberry a ploidia é conhecida, as cultivares comerciais (*Southern Highbush*) são autotetraploides, apresentando $2n = 4X = 48$ cromossomos. Entretanto a cultura também apresenta múltiplos alelos por locos, alto número de classes genotípicas, possibilidade de pareamento multi-valente e pouco conhecimento do comportamento cromossômico durante a meiose. Além disso, ambas as espécies têm ausência de genoma de referência.

No estudo de GS envolvendo cana-de-açúcar, alto nível de desequilíbrio de ligação e a correção da estrutura intrapopulacional foram utilizados. Além disso, o uso de seis metodologias foi examinado (incluindo regressão e métodos bayesianos). Para isso, uma população composta por 157 indivíduos derivados de autofecundação (88) e cruzamento biparental (69) foi fenotipada para seis caracteres, relacionados à produção de biomassa e teor de açúcar, utilizando um delineamento de blocos aumentados. Um total de 894 marcadores DArT foram genotipados para cada indivíduo. Os resultados obtidos ilustram o potencial de predição dos modelos GS para populações de melhoramento de cana-de-açúcar. Uma vez que o desequilíbrio de ligação máximo foi induzido, as estimativas de acurácia devem ser consideradas ótimas para condições experimentais similares. Entretanto, as baixas acurácias obtidas para brix e número de internódios refletiram a baixa variância genética para esses caracteres nesta população, o que é preocupante em programas de melhoramento genético, especialmente para brix. Além disso, a similaridade entre os resultados obtidos para as diferentes metodologias sugere uma menor importância desse fator para gerar os modelos de GS. Os resultados também mostraram que a correção dos efeitos da estrutura genética intrapopulacional na predição é fundamental para a obtenção de modelos mais realistas.

Na aplicação da GS para blueberry, o objetivo foi investigar a influência de

¹ Orientador: Prof. Dr. Alexandre Siqueira Guedes Coelho. EA – UFG.
Co-orientador: Prof. Dr. Patricio Ricardo Muñoz. IFAS – UF.

diferentes matrizes de parentesco, informações de ploidia e efeitos não-aditivos na estimativa de componentes de variância genética e na predição de fenótipos utilizando GS. Para isso, foram utilizados dados fenotípicos e genotípicos (~85K), além da informação de pedigree para 1,847 indivíduos. Os resultados obtidos mostraram que matrizes baseadas em marcadores geraram melhores resultados do que o uso do pedigree, para a maioria dos caracteres. As acurácias obtidas são encorajadoras para a aplicação de GS no melhoramento de blueberry, podendo reduzir em três anos a liberação de cultivares com um ganho genético médio esperado de 1.94X em relação ao sistema utilizado (matriz identidade). Além disso, os resultados também indicam que a importância da dosagem em GS irá depender do caráter analisado. E que, a adição de efeitos não-aditivos levou a um melhor ajuste dos modelos, bem como mostrou que estes efeitos representaram mais de 20% da variância total para a maioria das caracteres. Neste estudo também foi proposto o uso de dados contínuos para realizar GS, o que pode reduzir o tempo computacional e evitar problemas associados à classificação errônea dos genótipos.

Com o objetivo de investigar o conhecimento científico relacionado ao uso de marcadores moleculares em estudos genéticos para a cana-de-açúcar uma análise cienciométrica foi realizada. Os resultados obtidos trazem evidências de que a principal diferença entre o uso de marcadores em cana-de-açúcar e outras culturas está na quantidade e qualidade das informações obtidas com essa ferramenta. O uso de genótipos contínuos traz a possibilidade de inserção de dados de marcadores mais informativos no processo de seleção, com o uso de GS. Entretanto, desenvolvimento de ferramentas de análise genética e estatística são necessários para aumentar o conhecimento e a aplicação de dados genéticos em espécies poliploides.

Apesar das dificuldades associadas à complexidade genética das espécies poliploides, este trabalho gerou evidências empíricas de que a seleção genômica pode aumentar o ganho genético para essas espécies. Além disso, foi demonstrado que o uso de modelos que melhor representam a variabilidade genética em poliploides (ou seja, uso de informações de dosagem e inserção de efeitos não aditivos) pode ajudar na obtenção de melhores resultados (*i.e.* acurácia) e de informações (*i.e.* proporções explicadas pelos componentes genéticos). As informações aqui geradas podem auxiliar na compreensão da arquitetura e expressão de caracteres quantitativos, e essa também pode ser usada para maximizar a exploração dos efeitos genéticos no processo de melhoramento. Além disso, a análise das gerações subsequentes das populações pode ser usada para verificar os resultados indicados, permitindo a validação dos modelos.

Estudos futuros, considerando metodologias desenvolvidas especificamente para os poliploides, bem como a análise de ampla gama de variabilidade, podem ajudar na compreensão dos efeitos de dosagem e efeitos não-aditivos em futuras análises de GS para poliploides. Considerando a análise cienciométrica dos dados, o desenvolvimento de metodologias específicas é uma das principais barreiras ao uso da informação genética para poliploides. Nesse sentido, o investimento em recursos humanos e a intensificação de parcerias internacionais poderiam ajudar a desenvolver teorias para essas espécies, a fim de promover melhor interpretação e uso de dados genéticos em poliploides.

Palavras-chave: GWS, alopoliploide, autotetraploide, modelos bayesianos, ciencimetria.

ABSTRACT

DE BEM OLIVEIRA, I. **Genomic selection applied to breeding in polyploid species.** 2018. 174f. Thesis (Doctor of Science in Genetics and Plant Breeding) – Escola de Agronomia, Universidade Federal de Goiás, Goiânia, 2018.²

Genome selection (GS) is used to predict the simultaneous effects of a wide range of molecular markers distributed across the genome. It does not intend to identify individual locus effects, but to capture the combined effect of the causal polymorphisms involved in the genetic variation of a quantitative trait. The advent of next generation sequencing methodologies, concomitantly to the development of genetic and statistical analysis tools, has provided significant genetic gains for diploid species.

Despite the successes, its application in polyploid crops remains a challenge for several reasons, including the complex genomic architecture and unavailability of methods and software. Besides of GS potential to reshape breeding programs and to achieve faster selection gains, one major contribution of the present research is to discuss how to consider an appropriate predictive model in the context of polyploids. For this we applied the insertion of different factors in GS models considering predictions in two polyploid species: Sugarcane (*Saccharum* spp.) and Blueberry (*Vaccinium corymbosum*).

Sugarcane presents one of the most complex genome among crops. Modern cultivars have interspecific origins, high and variable ploidy levels (frequently aneuploidy), high heterozygosity and large genomes, where even the ploidy is unknown. In other hand, the ploidy is known, the cultivated southern highbush blueberry, it is an autotetraploid, presenting $2n = 4X = 48$ chromosomes, however the crop also presenting multiple alleles per locus, high number of genotype classes, possibility of multi-valent pairing, and poor knowledge of chromosome behavior during meiosis, both species have absence of reference genome.

In the GS study involving sugarcane the use of high level of linkage disequilibrium and the correction of population structure were tested. Furthermore, the use of six different methods (including regression and bayesian methods) was examined. For this a population, comprised of 157 individuals derived from selfing (88) and biparental crosses (69), was phenotyped for six traits, related to plant growth, biomass production and sugar content, using an augmented block design. A total of 894 DArT markers were genotyped for each individual. The results obtained illustrate the potential accuracy of GS models for sugarcane breeding populations. Since maximum linkage disequilibrium was induced, the accuracies estimates should be considered optimal for similar experimental conditions. Moreover, the low accuracies obtained for brix and number of internodes (near zero) reflected the low genetic variance for these traits in this population, which is worrying in breeding programs, especially for brix. Furthermore, the similarity among results obtained by the different methods suggests a minor importance of the choice of method used to generate the GS models. Results also showed that the correction for the intrapopulacional genetic structure effects on the accuracy is imperative for obtaining more realistic models.

² Advisor: Orientador: Prof. Dr. Alexandre Siqueira Guedes Coelho. EA – UFG.
Co-Advisor: Prof. Dr. Patricio Ricardo Muñoz. IFAS – UF.

In the application of GS for Blueberry the goal was to investigate the influence of different relationship matrices, ploidy information and non-additive effects in the estimation of genetic variance components, and on prediction of phenotypes using genome-wide selection (GS). For this, phenotypic, pedigree, and high-density marker data (~85K) for 1,847 individuals were used. The results obtained shown that marker-based matrices could lead to better predictions than pedigree for most of the traits. The obtained accuracies are encouraging for application of GS for blueberry breeding. This could reduce three year in the release of cultivars with a mean expected genetic gain representing 1.94X over the conventional blueberry breeding scheme (identity matrix). Furthermore, the results also indicate that the importance of dosage in GS will vary depending on the trait being analyzed. And that the addition of non-additive effects has lead to a better fit of the models, as well as showed that the non-additive effects represented more than 20% of the total variance for most of the traits. Going beyond previous studies for polyploid species here is also proposed the use of continuous data to perform GS. This would reduce the computational time and it avoids problems associated with misclassification of genotypic classes.

A science mapping analysis was also performed in order to investigate the scientific knowledge related to the use of molecular markers in genetic studies for sugarcane. The results obtained bring evidence that the main difference between the use of markers in sugarcane and other crops lays in the quantity and quality of information obtained from this tool. The use of continuous genotypes brings the possibility of insertion of more informative marker data in the process of selection, with the use of GS. But development is required in genetic and statistical analysis tools, to increase the knowledge and application of genetic data in polyploid species.

Despite the difficulties associated with the genetic complexity of polyploid species, this work generated empirical evidence that genomic selection may increase the genetic gain for these species. In addition, it has been shown that the use of models that best represent the genetic variability in polyploids (*i.e.* use of dosage information and insertion of non-additive effects) can help to obtain better results (*i.e.* accuracy) and information (*i.e.* proportions explained by the genetic components). The information generated here can also help in the understanding of quantitative traits architecture and expression. The results obtained can also be used in order to maximize the use of genetic effects in the breeding process (non-additive and additive proportions of variance). The analysis of subsequent generations of the populations can be used to verify the results presented here, allowing the validation of the models.

Future studies considering methodologies developed specifically for polyploids as well as the analysis of wide range of variability may help in the understanding of dosage and non-additive effects in future GS analysis for polyploids. Considering the scientometric analysis of the data results, development of methodologies specifically developed for polyploids is one of the main barriers to the use of genetic information in these cultures. In this sense, investment in human resources and the intensification of international partnerships could help develop theories for these species in order to foster progress in the interpretation and use of polyploid genetic data.

Keywords: GWS, allopolyploid, autotetraploid, bayesian model, scientometric analysis.

1 INTRODUÇÃO

Concebido por Meuwissen et al. (2001) com a intenção de acelerar o processo de melhoramento genético, a seleção genômica ampla (GS) é utilizada na predição dos efeitos simultâneos de grande quantidade de marcadores amplamente espalhados no genoma. Essa metodologia não busca a identificação de efeitos para locos individuais, mas sim a captura do efeito combinado dos polimorfismos causais envolvidos na variação genética dos caracteres quantitativos. Por princípio, a GS se baseia na existência de desequilíbrio de ligação (LD) entre os polimorfismos envolvidos na expressão fenotípica (polimorfismos causais) e os marcadores utilizados na análise (Meuwissen et al., 2001; Endelman et al., 2011; Zhang et al., 2011).

O método de GS faz uso tanto da fenotipagem quanto da genotipagem dos indivíduos e se baseia no desenvolvimento de um modelo de predição que, ao ser utilizado para a seleção, permite a predição dos valores genéticos de cada um dos indivíduos com base apenas no seu genótipo. Desta forma, pode-se fazer a seleção em fases precoces do programa, diminuindo o tempo de obtenção de materiais superiores (Meuwissen et al., 2001; Resende et al., 2010). Esta modelagem pode ser realizada com o uso de diversas metodologias e considerando diferentes fatores na análise (De Los Campos et al., 2013).

O advento das metodologias de sequenciamento de nova geração, em paralelo ao desenvolvimento de ferramentas de análise genética e estatística, tem proporcionado ganhos genéticos significativos para diversas espécies (Desta & Ortiz, 2014). Para espécies diploides a seleção baseada em metodologias estatísticas avançadas envolvendo marcadores moleculares é amplamente utilizada (Wiggans et al., 2011; Resende et al., 2012). Entretanto, a aplicação destas mesmas metodologias continua a ser um desafio para poliploides, devido à inexistência de metodologias e softwares capazes de lidar com a complexa arquitetura genômica inerente a essas espécies (Ramsey et al., 1998; Comai et al., 2005; Garcia et al., 2013; Grandke et al., 2016).

Eventos de poliploidização não são considerados uma exceção em plantas, cerca de 70% das Angiospermas e 95% das Pteridófitas passaram por ao menos um evento de poliploidização (Soltis & Soltis, 1999). Considerando a possibilidade de variação

fenotípica gerada por esses eventos, a poliploidia é considerada uma poderosa força evolutiva (Hieter & Griffiths, 1999). Apesar do importante papel da poliploidização, seus efeitos na herança de caracteres agronômicos ainda são pouco compreendidos em comparação ao desenvolvimento existente para espécies diploides (Dufresne et al., 2014).

Os primeiros estudos envolvendo o uso de GS para poliploides consideram a informação genética de maneira simplificada para contornar a complexidade genética destas espécies. Normalmente são utilizados modelos pseudo-diploides, que desconsideram os possíveis efeitos devidos à dosagem alélica (ou seja, o número de cópias de cada alelo em um determinado loco polimórfico). Apesar disso, resultados promissores foram reportados (Gouy et al., 2013; Annicchiarico et al., 2015; Ashraf et al., 2016; Ashraf et al., 2016b; Slater et al., 2016). O fato de que a maioria dos estudos envolve modelos simplificados para a aplicação de GS em poliploides, abre possibilidades para o desenvolvimento de metodologias mais apropriadas para o endereçamento dos efeitos de dosagem alélica.

A poliploidia pode estar envolvida em sistemas biológicos adaptativos complexos. Este fenômeno pode interferir na expressão fenotípica devido à alteração na dosagem alélica, causada pelo efeito aditivo dos mesmos alelos (efeito de dosagem), ou por interações mais complexas entre alelos e locos diferentes, englobando efeitos não-aditivos, de dominância e de epistasia, uma vez que características complexas são geralmente manifestações de intrincados processos de interação gênica (Osborn et al., 2003). Neste cenário, a inclusão de informações de dosagem alélica pode gerar melhorias nos modelos de GS (ou seja, melhor ajuste e/ou aumento da acurácia), assumindo uma interpretação mais realista dos efeitos para cada uma das possíveis classes genotípicas.

Apesar do conhecimento do efeito da poliploidia na expressão gênica (Guo et al., 1995; Osborn et al., 2003; Birchler et al., 2001; Adams et al., 2003), existem poucos estudos ligando efeitos de dosagem à variação fenotípica em poliploides (*e.g.* Nyine et al., 2018). É interessante ressaltar que a identificação genotípica é uma tarefa árdua para poliploides. Estudos de avaliação da genotipagem em poliploides mostram que ainda não há métodos que funcionem adequadamente (Voorrips et al., 2011; Grandke et al., 2016), a menos que uma alta cobertura de sequenciamento (60X a 80X) seja utilizada (Uitdewilligen et al., 2013).

Entre os efeitos que podem ser considerados nas análises de GS estão também os efeitos não-aditivos. Em certas condições teóricas (*e.g.* sob equilíbrio de Hardy-

Weinberg), os efeitos aditivos e não-aditivos são ortogonais (Cockerham, 1954; Kempthorne, 1954). Entretanto, em populações submetidas aos processos de melhoramento, pode haver confundimento estatístico no processo de estimação, levando à computação errônea dos efeitos não-aditivos como efeitos aditivos (Muñoz et al., 2014). A inclusão de matrizes independentes, capazes de capturar os efeitos aditivos e não-aditivos, pode auxiliar na qualidade de ajuste dos modelos utilizados, otimizando os resultados obtidos no processo de seleção, além de possibilitar a geração de informações sobre o controle genético do caráter na espécie estudada.

A correção dos efeitos da estruturação genética populacional também possui extrema importância na estimação dos componentes genéticos e na acurácia de predição dos modelos de GS. Na existência de estrutura genética populacional, os marcadores utilizados na GS podem capturar, além do efeito de LD, o efeito do parentesco, afetando os valores preditos e, conseqüentemente, as acurácias (Habier et al., 2007). Isto ocorre devido à geração de LD espúrio, em que a mistura de subpopulações, com diferentes composições em termos de frequências alélicas, resulta na alteração das frequências de combinações alélicas de locos distintos (haplótipos). Esse fator pode inflar as medidas de LD e causar associações espúrias entre marcadores e as variações fenotípicas observadas (Guo et al., 2014).

Uma importante contribuição da presente pesquisa é discutir como considerar um modelo preditivo apropriado no contexto de espécies poliploides. Para isso, dados de duas espécies poliploides, cana-de-açúcar (*Saccharum* spp.) e mirtilo (*Vaccinium* spp.; blueberry), foram utilizadas. Neste contexto, os estudos aqui desenvolvidos têm o objetivo de demonstrar a aplicação da seleção genômica em espécies poliploides considerando a influência da inserção de diferentes fatores na predição de fenótipos. Foram consideradas: i) correções para estrutura genética populacional; ii) efeitos da endogamia; iii) efeitos da utilização de diferentes matrizes de parentesco; iv) efeitos do uso de diferentes metodologias estatísticas para obtenção dos modelos de predição (e.g. métodos paramétricos baseados em regressão explícita e métodos Bayesianos); v) efeitos da utilização da informação dos níveis de ploidia; vi) efeitos da inserção dos efeitos não-aditivos nos modelos e vii) uso de dados genotípicos de natureza quantitativa.

O objetivo final foi produzir informação para nortear a otimização dos processos de melhoramento de espécies poliploides, pelo aprimoramento da capacidade de se prever valores de caracteres complexos pelo uso de modelos de seleção genômica

ampla. Os resultados aqui apresentados fornecem a primeira proposta de utilização de modelos de seleção genômica para blueberry. Representam também uma prova de conceito da aplicação de GS para cana-de-açúcar. Além disso, uma nova metodologia de análise é proposta, em que se considera o uso de dados genotípicos de natureza quantitativa, em substituição à adoção de abordagens simplificadas para aplicação da GS em poliploides.

2 REVISÃO DE LITERATURA

2.1 CANA-DE-AÇÚCAR E BLUEBERRY

A cana-de-açúcar é uma *commodity* agrícola de grande importância no mundo. É cultivada em regiões tropicais e subtropicais, sendo 50,2% da produção global de cana-de-açúcar proveniente das Américas, 40,5% da Ásia, 6,5% da África e 2,8% da Oceania. Em 2016, seu plantio atingiu 27 milhões de hectares, produzindo cerca de 1,891 bilhões de toneladas, das quais 71,9% foram produzidas pelo Brasil e pela Índia (FAO, 2018). A cana-de-açúcar assume papel importante não só pelo suprimento de açúcar mundial, mas também como fonte de álcool (Menossi et al., 2008), fator que vêm impulsionando a pesquisa com a cultura devido ao impacto econômico das energias sustentáveis (Cheavegatti-Gianotto et al., 2011). O interesse na cultura tornou-se ainda maior após a descoberta de processos enzimáticos que quando aplicados sobre a celulose e a hemicelulose, presentes nos resíduos produzidos pelas indústrias (bagaço), são capazes de produzir o etanol de segunda geração, tornando o cultivo de cana-de-açúcar ainda mais produtivo e rentável (Sims et al., 2006; Menossi et al., 2008; Piperidis et al., 2010).

Saccharum é considerado um dos gêneros mais complexos entre as plantas cultivadas, possuindo genoma extenso e grande número de cromossomos (D'Hont et al., 1994; D'Hont et al., 1996; Piperidis et al., 2010; Zhang et al., 2012). A ocorrência de múltiplas cópias de um mesmo loco em todos os cromossomos suporta a hipótese de autoploidia (Grivet et al., 1996), mas a espécie continua sem um parente diploide conhecido (D'Hont et al., 1994; Zhang et al., 2012). A complexidade genômica do gênero é acentuada pela coexistência de dois genomas nas cultivares comercialmente utilizadas (híbridos interespecíficos), devido às hibridizações realizadas no decorrer do melhoramento genético da cultura (Cuadrado et al., 2004; D'Hont, 2005; Piperidis et al., 2010).

O mirtilo, referido como blueberry daqui por diante, é um arbusto perene cultivado principalmente por sua fruta comestível. É uma cultura de alto valor agregado, sendo os dez países principais produtores: Estados Unidos, Chile, Canadá, Espanha, China,

Argentina, Polônia, Perú, México e Marrocos (Brazelton et al., 2017). Dados da FAO (2018), considerando o período de 1961 à 2016, apontam que as Américas representam 79,5% da produção total de blueberry, sendo a maior parte desta produção obtida pelos Estados Unidos, Chile e Canadá. Europa, Ásia e Oceania também participam deste mercado, representando 18,7%, 0,3% e 1,4% da produção total. A análise dos dados da série histórica revela a quadruplicação da produção quando são comparados os dados de 1961 à 2016, em que foram obtidos 126 mil toneladas em 1961 e 552 mil toneladas em 2016. Observou-se também um aumento na área plantada para essa cultura, passando de 50 mil hectares em 1961 para 111 mil hectares em 2016 (FAO, 2018; Figura 1).

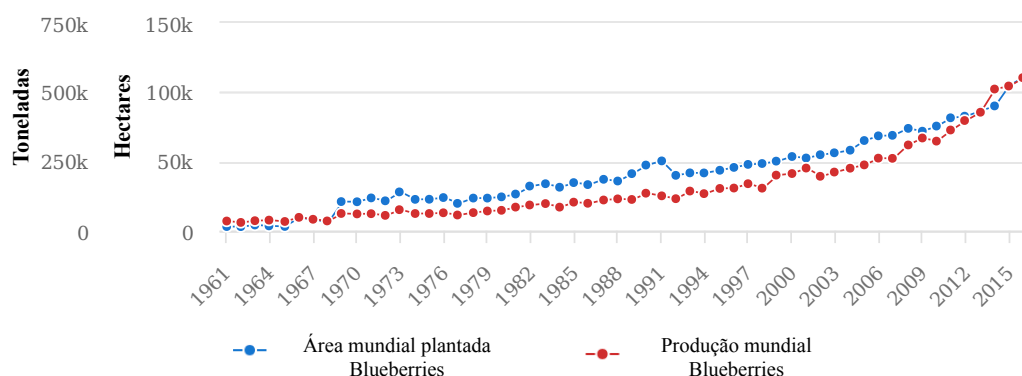


Figura 1. Área plantada e produção mundial de blueberry, no período de 1961 a 2016.
Fonte: Adaptado de FAO (2018).

Blueberries são classificadas sob o gênero *Vaccinium*, que é dividido em diferentes sessões (e.g. *Cyanococcus*, *Oxycoccus*, *Vitis-idaea*, *Myrtillus* e *Vaccinium*), sendo que um dos principais diferenciais entre as seções é o nível de ploidia das espécies (Luby et al., 1991). Entretanto, também são observados múltiplos níveis de ploidia dentro de sessões (Lyrene et al., 2003). Diferentes espécies de blueberry são utilizadas em sistemas de produção comercial. A fonte primária de variabilidade genética para o melhoramento está centrada principalmente em três espécies: Highbush Blueberry (HB; i.e. *V. corymbosum*, $2n = 4x = 48$), Lowbush Blueberry (LB; i.e. *V. angustifolium*, $2n = 4x = 48$) e Rabbiteye Blueberry (i.e. *V. virgatum*, $2n = 6x = 72$). Espécies selvagens da sessão *Cyanococcus* são consideradas como fonte secundária de variabilidade (Brevis et al., 2008).

O grupo *Southern Highbush Blueberry* (SHB) compreende espécies de blueberries (e.g. *V. fuscatum* e *V. darrowi*) e híbridos cultivados que requerem baixa

quantidade de horas de frio (*chilling hours*; horas de frio abaixo de 7°C; e.g. ~300 horas). As variedades de blueberry cultivadas na Flórida possuem baixo requerimento de horas de frio, pertencendo ao grupo SHB, resultante da hibridação interespecífica entre as espécies HB, LB e Rabbiteye (Sharpe & Darrow, 1959; Ballington, 1990; Brevis et al., 2008). Estes cruzamentos permitiram a obtenção de híbridos com alto vigor, resistência a doenças e condições ambientais adversas (*i.e.* calor e humidade), maturação precoce, arquitetura vertical e melhor qualidade do fruto, características condizentes com o requerimento de plantios comerciais (Lyrene, 1997; Lyrene et al., 2003). Apesar da hibridação, a execução de múltiplos retrocruzamentos (desde 1950, no programa de melhoramento de blueberries da Universidade da Flórida; Lyrene, 2008), para a retrogressão ao genótipo de *V. corymbosum*, foram executados de forma que as cultivares comerciais são autotetraploides, apresentando $2n = 4x = 48$ cromossomos (Lyrene et al., 2005). Em particular, a natureza complexa de espécies autoploiploides deve-se à presença de múltiplos alelos por loco, grande número de classes genóticas, possibilidade de pareamento multivalente, pouco conhecimento do comportamento cromossômico durante a meiose (Slater et al., 2013; Dufresne et al., 2014; Mollinari et al., 2015) e ausência de uma sequência de referência para o genoma.

Um dos pontos críticos dos programas de melhoramento genético, responsáveis em boa parte pelos progressos obtidos em produtividade, é a identificação de genótipos superiores. Tanto para cana-de-açúcar quanto para blueberry, uma vez identificados, estes genótipos são tipicamente clonados. Existe, portanto, a possibilidade de fixação dos genótipos favoráveis e o aproveitamento dos componentes aditivos e não-aditivos no melhoramento destas culturas. Neste contexto, além da possibilidade de que os efeitos de dosagem alélica sejam relevantes, os efeitos não-aditivos são também potencialmente importantes, haja vista a possibilidade de uso da propagação vegetativa.

2.2 POLIPLOIDIA: UM EXEMPLO EM CANA-DE-AÇÚCAR

Uma das características básicas do genoma de cana-de-açúcar é a poliploidia, ou seja, cada indivíduo carrega k (sendo $k \geq 4$) conjuntos de cromossomos do mesmo tipo e origem (mesmo grupo de homologia). Desta maneira, como existem k cópias de um mesmo cromossomo carregando os mesmos locos, o número de cópias de um mesmo alelo pode variar. Essa quantidade de cópias de um mesmo alelo é definida como a dosagem deste alelo (Garcia et al., 2013).

As variedades cultivadas de cana-de-açúcar são constituídas por dois genomas, haja vista que são tipicamente híbridos interespecíficos, tornando ainda maior a complexidade do genoma da cultura (D'Hont et al., 2005). Assim, as variedades comerciais de cana-de-açúcar são geralmente tratadas como aloploidos, em que existem genomas de mais de uma espécie e múltiplas cópias de dado cromossomo em um mesmo indivíduo.

Os eventos de poliploidia não são exceção no reino vegetal, cerca de 70% das Angiospermas e aproximadamente 95% das Pteridófitas sofreram pelo menos um evento de poliploidização, sendo este processo uma poderosa força evolutiva (Hieter & Griffiths, 1999; Soltis & Soltis, 1999; Pikaard, 2001), já que nestes eventos novas combinações genótípicas podem ser geradas (Figura 2). Este fator pode ter contribuído para o sucesso dos poliploides na natureza e para o seu uso na agricultura (Osborn et al., 2003).

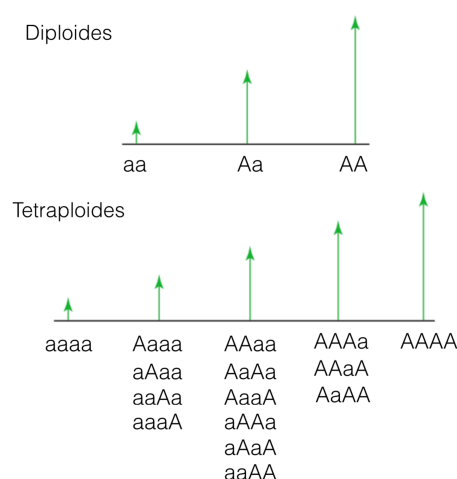


Figura 2. Variações teóricas nos valores genotípicos (eixo vertical) determinados por herança aditiva em diploides e tetraploides.
Fonte: Adaptado de Osborn et al. (2003).

Uma das hipóteses para explicar o sucesso dos poliploides envolve as alterações nos níveis de expressão gênica como consequência de variações nas dosagens alélicas. Uma outra hipótese está relacionada à possibilidade de divergência evolutiva nos genes duplicados, uma vez que sua função inicial pode ser cumprida por apenas uma das cópias existentes no genoma – teoria da Neofuncionalização (Osborn et al., 2003). Em espécies diploides, o efeito de dosagem alélica sobre o fenótipo já é conhecido e estudado para muitos genes, incluindo os efeitos de regulação em diferentes fases do

desenvolvimento das plantas e os efeitos sobre o controle genético de caracteres quantitativos (efeitos aditivos) (Vencovsky & Barriga, 1992; Osborn et al., 2003; Griffiths et al., 2008).

Um exemplo do potencial de variação obtido para poliploides devido à dosagem pode ser visto em milho. O estudo de Guo et al. (1995), envolvendo plantas mono, di, tri e tetraploides, evidenciou a interferência da dosagem na expressão de 18 genes, obtendo correlação positiva entre o aumento da dosagem e o nível de transcritos para a maioria dos genes. Neste estudo foram também observados casos de correlações negativas e de correlações positivas maiores do que as esperadas, evidenciando o quanto a dosagem pode interferir na expressão gênica, mostrando a importância do estudo do efeito deste fator sobre a expressão de caracteres quantitativos.

2.3 SEGREGAÇÃO EM POLIPLOIDES: UM EXEMPLO EM CANA-DE-AÇÚCAR

Haldane (1930) estudando o comportamento genético apresentado por poliploides mostrou a existência de grande variedade de gametas produzidos por autoploiploides, sendo esta tanto maior quanto maior for o nível de ploidia. Isto se deve à existência de uma grande gama de alelos de um mesmo loco segregando em conjunto. Por exemplo, considerando-se o nível de ploidia $2n = 8x$, o genótipo para um determinado loco com dois alelos segregando (A e a) pode ser desde *nulplex* (i.e. *aaaaaaaa*), *simplex* (i.e. *Aaaaaaaaa*), *duplex* (i.e. *AAaaaaaaaa*), *triplex* (i.e. *AAAaaaaaaaa*), ou *multiplex* (i.e. com mais de três cópias do alelo A). O mesmo comportamento pode ser inferido para aloploiploides, como no caso das variedades comerciais de cana-de-açúcar (Figura 3).

Em cana-de-açúcar, um nível adicional de complexidade é apresentado devido à existência, relativamente comum, de aneuploidias e à presença de trivalentes ou de univalentes durante a meiose, mesmo diante da segregação regular geralmente observada durante a Metáfase I. Assim, a complexidade genética da cana-de-açúcar se deve ao comportamento diferenciado apresentado pelo seu genoma, que, mesmo diante da ocorrência de padrões regulares de segregação, apresenta mecanismos de herança que podem variar na dependência dos genitores envolvidos nos cruzamentos (Figura 4).

$2m$	Zygote	Gametes
4	$A^3 a$ $A^2 a^2$ $A a^3$	$1A^3 : 1A a$ $1A^2 : 4A a : 1 a^2$ $1A a : 1 a^2$
6	$A^5 a$ $A^4 a^2$ $A^3 a^3$ $A^2 a^4$ $A a^5$	$1A^3 : 1A^2 a$ $1A^3 : 3A^2 a : 1A a^2$ $1A^3 : 9A^2 a : 9A a^2 : 1a^3$ $1A^2 a : 3A a^2 : 1a^3$ $1A a^2 : 1a^3$
8	$A^7 a$ $A^6 a^2$ $A^5 a^3$ $A^4 a^4$ etc.	$1A^4 : 1A^3 a$ $3A^4 : 8A^3 a : 3A^2 a^2$ $1A^4 : 6A^3 a : 6A^2 a^2 : 1A a^3$ $1A^4 : 16A^3 a : 36A^2 a^2 : 16A a^3 : 1a^4$
10	$A^9 a$ $A^8 a^2$ $A^7 a^3$ $A^6 a^4$ $A^5 a^5$ etc.	$1A^5 : 1A^4 a$ $2A^5 : 5A^4 a : 2A^3 a^2$ $1A^5 : 5A^4 a : 5A^3 a^2 : 1A^2 a^3$ $1A^5 : 10A^4 a : 20A^3 a^2 : 10A^2 a^3 : 1A^4 a$ $1A^5 : 25A^4 a : 100A^3 a^2 : 100A^2 a^3 : 25A^4 a : 1a^5$
12	$A^{11} a$ $A^{10} a^2$ $A^9 a^3$ $A^8 a^4$ $A^7 a^5$ $A^6 a^6$ etc.	$1A^6 : 1A^5 a$ $5A^6 : 12A^5 a : 5A^4 a^2$ $2A^6 : 9A^5 a : 9A^4 a^2 : 2A^3 a^3$ $1A^6 : 8A^5 a : 15A^4 a^2 : 8A^3 a^3 : 1A^2 a^4$ $1A^6 : 15A^5 a : 50A^4 a^2 : 50A^3 a^3 : 15A^2 a^4 : 1A a^5$ $1A^6 : 36A^5 a : 225A^4 a^2 : 400A^3 a^3 : 225A^2 a^4 : 36A a^5 : 1a^6$
16	$A^{15} a$ $A^{14} a^2$ $A^{13} a^3$ $A^{12} a^4$ $A^{11} a^5$ $A^{10} a^6$ $A^9 a^7$ $A^8 a^8$ etc.	$1A^8 : 1A^7 a$ $7A^8 : 16A^7 a : 7A^6 a^2$ $1A^8 : 4A^7 a : 4A^6 a^2 : 1A^5 a^3$ $5A^8 : 32A^7 a : 56A^6 a^2 : 32A^5 a^3 : 5A^4 a^4$ $1A^8 : 10A^7 a : 28A^6 a^2 : 28A^5 a^3 : 10A^4 a^4 : 1A^3 a^5$ $1A^8 : 16A^7 a : 70A^6 a^2 : 112A^5 a^3 : 70A^4 a^4 : 16A^3 a^5 : 1A^2 a^6$ $1A^8 : 28A^7 a : 196A^6 a^2 : 490A^5 a^3 : 490A^4 a^4 : 196A^3 a^5 : 28A^2 a^6 : 1A a^7$ $1A^8 : 64A^7 a : 784A^6 a^2 : 3136A^5 a^3 : 4900A^4 a^4 : 3136A^3 a^5 : 784A^2 a^6 : 64A a^7 : 1a^8$

Figura 3. Conformações gaméticas produzidas por indivíduos poliploides, considerando-se níveis de ploidia variando entre 4x e 16x e possíveis dosagens durante a segregação mendeliana. Os genótipos estão representados por $A^x a^y$, em que x é a dose do alelo A e y a dose do alelo a . Fonte: Haldane (1930).

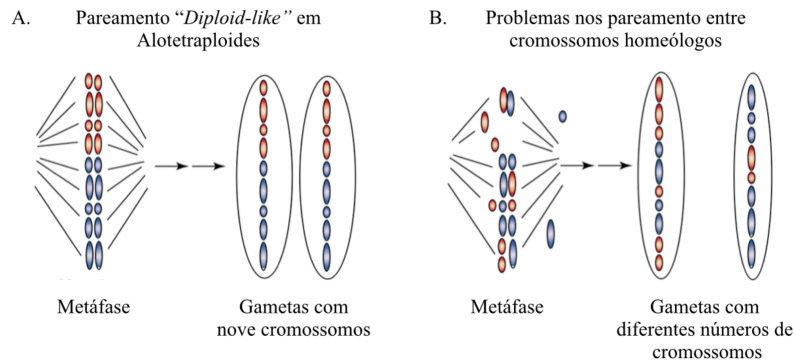


Figura 4. Representação do pareamento entre os cromossomos hom(e)ólogos durante a Metáfase I da meiose em alotetraploides. A) Pareamento regular de cromossomos homólogos em alotetraploides. B) Pareamento irregular de cromossomos hom(e)ólogos e geração de gametas com diferentes números cromossômicos. Fonte: Adaptado de Pikaard (2001).

A primeira consequência da segregação irregular de cromossomos durante a meiose é a obtenção de grupos de hom(e)ologia com diferentes números de cromossomos nos híbridos (Cuadrado et al., 2004). O uso de técnicas de hibridização *in situ* de DNA permitiu a visualização desta disparidade em cana-de-açúcar, além da identificação de grupos de co-segregação de cromossomos e da identificação dos cromossomos homólogos e homeólogos nos genomas dos híbridos (Figura 5) (D'Hont et al., 1994; D'Hont et al., 1996).

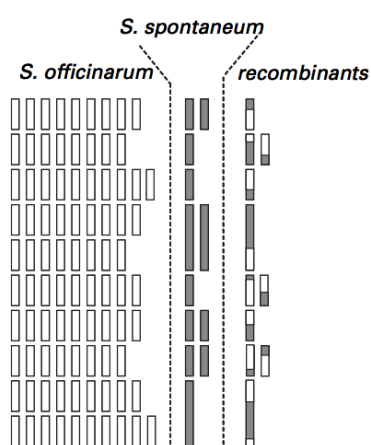


Figura 5. Representação esquemática do genoma das cultivares híbridas de cana-de-açúcar. Cada cromossomo é representado por uma barra. As colorações cinza e branca representam as frações de cada cromossomo herdadas de *S. spontaneum* e *S. officinarum*, respectivamente. Fonte: D'Hont et al. (2005).

Considerando-se conjuntamente a variação das dosagens alélicas e os diferentes tipos de segregação que podem existir em poliploides, algumas metodologias foram propostas para se estimar as probabilidades de ocorrência de todas as classes genotípicas possíveis. Estas probabilidades foram inicialmente inferidas para tetraploides por Muller (1914), em seu estudo com *Prímulas*. Haldane (1930) extrapolou esta abordagem para abranger níveis de ploidia variando entre 1 e 16, o que foi posteriormente corroborado pelos modelos de Ripol et al. (1999). Observa-se nestes três estudos que à medida em que os níveis de ploidia aumentam, as probabilidades esperadas para cada uma das doses se tornam cada vez mais próximas umas das outras (Tabela 1). As probabilidades esperadas na segregação de autopoliploides quando a dosagem é alta ($n > 6$) são tão próximas que a partir da segregação observada, em termos práticos, não é possível sua separação. Este fato, consequentemente, impossibilita a estimação da dosagem pela análise de uma distribuição teórica, uma vez que as frequências das diferentes classes, na prática, acabam por se sobrepor (Ripol et al., 1999).

Do ponto de vista de análise existem ainda outros fatores que podem interferir na identificação da dosagem, como o confundimento durante as atribuições dos alelos aos locos, em decorrência de duplicações cromossômicas e a existência de paralogia (Albino et al., 2006). Assim, é importante ressaltar que diferentes aspectos devem ser considerados para que haja uma correta determinação e classificação da dosagem de diferentes alelos, incluindo a possibilidade tanto de realização de estudos genéticos com diferentes tipos de marcadores genéticos, quanto de estudos dos efeitos da dosagem nos níveis de expressão gênica, além do entendimento básico do genoma da espécie como um todo.

Tabela 1. Relação entre níveis de ploidia e frequências esperadas (%) de um determinado alelo de referência. Fonte: Adaptado de Cordeiro et al. (2006).

Frequência do Alelo de referência	Nível de ploidia						
	8	9	10	11	12	13	14
1	12,50	11,11	10,00	9,09	8,33	7,69	7,14
2	25,00	22,22	20,00	18,18	16,67	15,38	14,29
3	37,50	33,33	30,00	27,27	25,00	23,08	21,43
4	50,00	44,44	40,00	36,36	33,33	30,77	28,57
5	62,50	55,56	50,00	45,45	41,67	38,46	35,71
6	75,00	66,67	60,00	54,55	50,00	46,15	42,86
7	87,50	77,78	70,00	63,64	58,33	53,85	50,00
8	100,00	88,89	80,00	72,73	66,67	61,54	57,14
9		100,00	90,00	81,82	75,00	69,23	64,29
10			100,00	90,91	83,33	76,92	71,43
11				100,00	91,67	84,62	78,57
12					100,00	92,31	85,71
13						100,00	92,86
14							100,00

2.4 USO DE MARCADORES *SINGLE* E *MULTIPLE DOSE* EM POLIPLOIDES: UM EXEMPLO EM CANA-DE-AÇÚCAR

Os estudos utilizando marcadores moleculares com o intuito de se compreender e se desvendar a complexidade genética de cana-de-açúcar se basearam inicialmente na análise de marcadores *single dose* (Moore et al., 1995). Posteriormente, foram incorporadas metodologias de análise de marcas com diferentes segregações (*double single dose*) (Da Silva et al., 1995; Aitken et al., 2005). A configuração denominada de *single dose* ocorre quando o alelo (marcador) está presente em cópia única, em apenas um dos parentais, apresentando segregação mendeliana com razão esperada de 1:1 na progênie. A configuração *double single dose* é aquela em que o alelo (marcador) está presente em cópia única nos dois parentais simultaneamente, apresentando segregação mendeliana esperada

de 3:1 na progênie (Figura 6).

Apesar destes avanços metodológicos, os estudos com marcadores moleculares em indivíduos poliploides se tornam muito complicados quando se considera possibilidade de segregação de três ou mais alelos para um mesmo loco. Nestes casos, devido à grande quantidade de homólogos que podem se parear e se recombinar, um grande número de combinações genótípicas se torna possível. Outro problema se dá pela não ocorrência total de pareamento preferencial, como ocorre nos aloploiploides. Devido a estes fatos, os alelos que aderem à segregação mendeliana simples (*single dose*) foram mais utilizados para interpretação da informação genética em cana-de-açúcar (Ming et al., 2001; Garcia et al., 2006).

Apesar da existência de métodos que permitem a utilização de informações de marcas com diferentes dosagens (Da Silva, 1993; Wu et al., 2002) e da evidência de que a utilização destas marcas possibilitaria o melhor uso da informação genética disponível para cana-de-açúcar (Da Silva et al., 1995; Aitken et al., 2005), poucos trabalhos foram publicados utilizando marcadores com configurações diferentes da *single dose*.

Além do trabalho de Da Silva (1993), diversos estudos têm apontado para a necessidade de inclusão de informações de locos com configurações diferentes da *single dose* (Dufour et al., 1997; Da Silva, 1993a; Da Silva et al., 1995; Aitken et al., 2005). Entretanto, devido à inexistência de métodos estatísticos capazes de determinar a ligação destes locos com diferentes dosagens, eles foram pouco utilizados (Sobral & Honeycutt, 1993). Desta maneira, pode-se dizer que a análise genética da cana-de-açúcar ainda não foi beneficiada com toda a informação que os marcadores moleculares podem efetivamente disponibilizar (Garcia et al., 2013) (Figura 7).

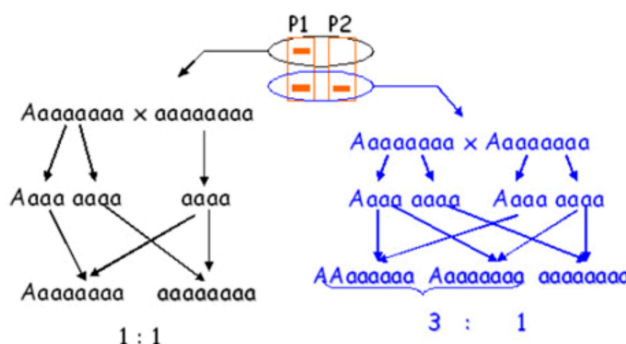


Figura 6. Representação do padrão de segregação esperado para locos *single dose* (1:1) e *double single dose* (3:1) em cruzamentos de clones de cana-de-açúcar, considerando-se o nível de ploidia igual a oito ($2n = 8x$). Fonte: Albino et al. (2006).

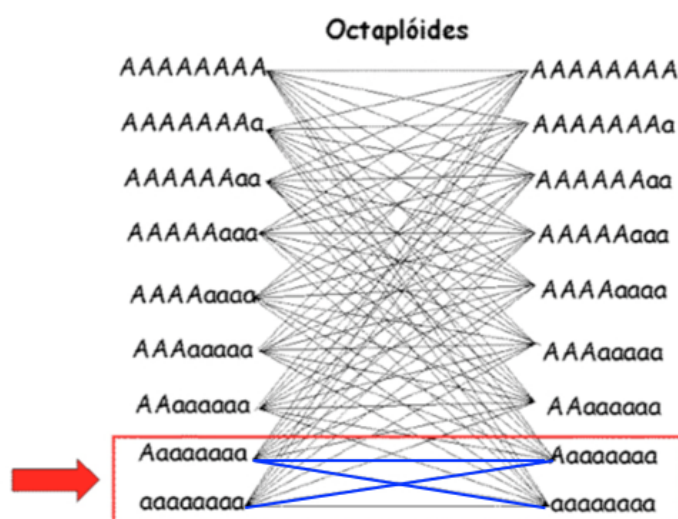


Figura 7. Padrões de segregação possíveis em cruzamentos biparentais entre octaploides ($2n = 8x$) considerando um único loco, evidenciando as três combinações mais utilizadas para estudos em cana-de-açúcar (*single dose* no parental 1, *single dose* no parental 2 e *double single dose*, representadas pelas linhas azuis dentro do quadro em evidência) e a quantidade de informação que se deixa de usar pela inexistência de métodos estatísticos capazes de capturar esta informação. Fonte: Adaptado de Albino et al. (2006).

Outra evidência que aponta para a necessidade de estudos sobre dosagem alélica em poliploides é o alto nível de heterozigosidade apresentado por estas espécies (Soltis & Soltis, 2000). Ming et al. (1998) e Grivet & Arruda (2002) mostraram a existência de correlação positiva entre os níveis de ploidia e a frequência de marcas *multi dose*, evidenciando que estas marcas são muito mais abundantes no genoma do que havia se pensado, já que as marcas *single dose* chegaram a ser consideradas mais abundantes no genoma do que as *multiple dose* (Wu et al., 1992; Da Silva, 1993b; Da Silva et al., 1995; Guimarães et al., 1997; Ming et al., 1998). Pela análise de probabilidades, pode-se demonstrar que a condição *single dose* é menos provável do que a *multiple dose* em indivíduos com elevados níveis de ploidia, na presença de elevados níveis de heterozigosidade. Uma evidência neste sentido pode ser observada no trabalho de Casu et al. (2012), em que para a maior parte dos genes analisados foi verificada a ocorrência de múltiplas cópias no genoma e somente para uma minoria foi detectada uma única cópia.

Mesmo com as marcas *single dose* tendo sido consideradas como ferramentas poderosas para o desenvolvimento de estudos de diversidade, bem como para a identificação de variedades e de espécies (Lu et al., 1994; Lu et al., 1994b), alguns estudos podem ser prejudicados pela sua utilização exclusiva. Exemplo deste fato está na

construção de mapas genéticos, em que o uso exclusivo desta configuração de segregação reduz drasticamente o número de locos mapeáveis, exigindo a análise de um número bem maior de marcadores para que se alcance a saturação do genoma e para o estabelecimento correto dos grupos de ligação (Ritter et al., 1990; Da Silva et al., 1995). Estudos envolvendo a determinação da dosagem poderiam auxiliar no aumento da quantidade de informação disponível para estudos genéticos em cana-de-açúcar.

2.5 DOSAGEM ALÉLICA EM POLIPLOIDES: UM EXEMPLO EM CANA-DE-AÇÚCAR

A partir do reconhecimento da complexidade do genoma de cana-de-açúcar, inúmeros métodos têm sido descritos para possibilitar o uso de dados genéticos moleculares para a espécie. O uso do teste de qui-quadrado (X^2), a partir da segregação de locos dominantes observada em progênies, permitiu a construção de muitos dos estudos genéticos em cana-de-açúcar (Wu et al., 1992; Da Silva, 1993; Wu et al., 2002). Além do uso do X^2 , intervalos de confiança com base em distribuição binomial também foram utilizados.

Uma metodologia descrita para poliploides foi elaborada por Voorrips et al. (2011) para batata, com base em curvas teóricas construídas com base em cinco diferentes distribuições normais, cada uma correspondendo a uma classe genotípica diferente (de 1 a 4 cópias do alelo de referência). O uso desta metodologia para cana-de-açúcar não se mostrou adequado, uma vez que, devido às aneuploidias, ao pareamento irregular durante a meiose e, ainda, ao não conhecimento dos níveis exatos de ploidia de cada grupo de homólogos, o comportamento apresentado para as diferentes classes genotípicas não apresentaram aderência as curvas teóricas, impedindo que pudessem ser realizadas as inferências dos genótipos dos indivíduos.

O mesmo ocorreu para os métodos que fazem uso da aderência à distribuição de X^2 e de intervalos de confiança com base na distribuição binomial. A complexidade do genoma faz com que as curvas teóricas de frequências genotípicas esperadas na segregação sejam menos eficazes para a estimação da dosagem. Estes métodos se revelaram eficazes somente para a determinação de dosagem alélica em locos com níveis de ploidia menores ou iguais a $2n = 4x$ (Baker et al., 2010) (Figura 8).

Conforme já salientado, a complicação com níveis maiores de ploidia se deve à

grande proximidade que existe entre as proporções genotípicas esperadas na segregação (Rippol et al., 1999), como pode ser visto na Tabela 1. Se forem estatisticamente comparadas, as frequências esperadas para quatro e cinco doses, nos níveis de ploidia $2n = 10x$ e $2n = 12x$, respectivamente, elas não permitem a diferenciação das duas classes de ploidia, visto que são muito próximas (40,0% e 41,7%, respectivamente). O mesmo ocorre para as frequências esperadas para seis e sete doses, nestes mesmos níveis de ploidia.

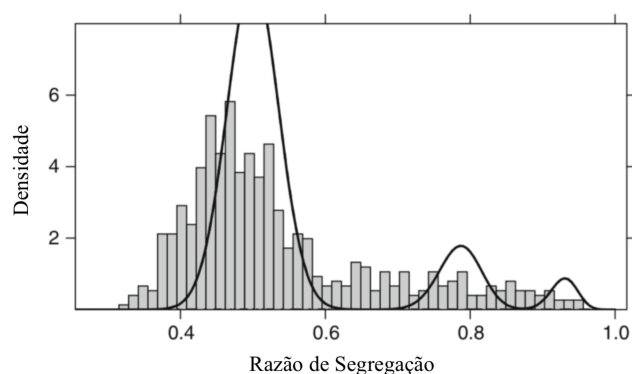


Figura 8. Histograma das razões de segregação observadas para 566 locos AFLP obtidos por Baker et al. (2010). A linha sólida representa a distribuição teórica esperada para uma mistura de locos *single*, *double* e *triple dose*, com as proporções para cada tipo estimadas com o teste de χ^2 . Nota-se um deslocamento para a esquerda que pode ser atribuído à complexidade genética da espécie, principalmente à existência de aneuploidias. Fonte: Adaptado de Baker et al. (2010).

O primeiro trabalho descrito para determinação de dosagem em cana foi o de Baker et al. (2010), recebido para impressão em 2008. Esta metodologia teve o intuito de auxiliar a determinação de dosagem de marcadores dominantes, afim de se promover uma maior qualidade nos mapas construídos e nas análises de QTLs, tentando corrigir os problemas associados ao processo de estimação baseado na distribuição de χ^2 . A metodologia utilizada, baseada em modelos mistos, foi elaborada utilizando-se a abordagem Bayesiana de análise. Também considerando marcadores dominantes e a abordagem Bayesiana, George & Aitken (2010) desenvolveram outra metodologia para a estimação de dosagem alélica em cana-de-açúcar observando, em determinadas condições, resultados superiores àqueles obtidos pelo método de χ^2 .

Tanto para o uso do modelo proposto por Baker et al. (2010), quanto para uso do modelo de George & Aitken (2010), são utilizados conhecimentos específicos sobre os marcadores, como a segregação observada e o conhecimento prévio da dosagem dos locos

nos parentais, informações que, muitas vezes, não estão disponíveis ou não são conhecidas. Além disso, o modelo de George & Aitken (2010) considera pareamento preferencial. Este fenômeno é negligenciado no modelo de Baker et al. (2010).

Desta maneira, estas metodologias são aplicáveis apenas para determinados tipos de dados, quando se tem a proporção de segregação observada e quando as dosagens dos parentais são conhecidas. Estas metodologias podem gerar resultados não acurados para genótipos com dosagens superiores à quatro, devido ao viés gerado pela dificuldade de separação das classes genotípicas, pelo uso apenas de dados de segregação (Ripol et al., 1999).

As ferramentas computacionais que implementaram o uso dos modelos descritos por Baker et al. (2010) e por George & Aitken (2010) foram disponibilizadas de forma livre. A primeira na plataforma R (R Core Team, 2018), representada pelos pacotes *polySegratio* e *polySegratioMM* e a segunda em um *software*, elaborado em linguagem C, chamado *bdose*.

Casu et al. (2012) desenvolveram uma metodologia independente do padrão de segregação e que não exige o conhecimento prévio das dosagens nos parentais. A técnica se baseia na análise de RT-qPCR. Esta aplicação, que se demonstrou eficaz para a dosagem de transgenes em cana-de-açúcar, utiliza como referência dois genes de cópia única no genoma (encontrados com o auxílio de ferramentas de bioinformática). O processo se baseia na comparação das taxas de amplificação por RT-qPCR de locos de cópia única (chamados de referência) e aquelas observadas para os locos de interesse (chamados de teste). A metodologia se mostrou altamente eficaz na determinação da dosagem, porém sua aplicação quando considerada uma grande quantidade de locos, como é o caso da informação gerada pelas novas tecnologias de sequenciamento, se torna inviável.

Garcia et al. (2013) realizaram o estudo quantitativo de SNPs no genoma poliploide de cana, permitindo a obtenção da dosagem alélica para estes polimorfismos, estabelecendo de um novo *framework* de análises de dados para a cana-de-açúcar. Estimativas do número de alelos em locos dos diferentes grupos de segregação foram obtidas com o uso de espectrometria de massa, pelo uso da tecnologia *Sequenom MassArray*, interpretados com o *software SuperMASSA*. Os resultados obtidos permitiram a observação de diferentes números de cromossomos para cada grupo de hom(e)ologia, sendo este número variável para diferentes grupos, mas constante para cada um dos grupos de hom(e)ologia. Os autores apresentaram os resultados obtidos para duas diferentes

populações: a primeira contendo 180 indivíduos obtidos pelo cruzamento entre as cultivares IACSP 95-3018 e IAC93-3046 e a segunda representada por um painel de mapeamento associativo constituído por 142 genótipos, representando cultivares comerciais e ancestrais das cultivares modernas. Foram analisados dados obtidos para 1034 SNPs (dentre os quais, 943 obtidos com base em dados do projeto SUCEST). Este estudo também verificou que poucos locos apresentavam configuração em *single dose*, rompendo com ideias anteriores de que esta configuração apresenta-se com maior frequência no genoma (Wu et al., 1992; Da Silva, 1993a; Da Silva et al., 1995; Guimarães et al., 1997; Ming et al., 1998). O que corrobora a hipótese de que, rotineiramente, somente uma pequena parte da informação genética disponível para cana-de-açúcar é utilizada na maioria dos estudos. Isto já havia sido verificado por Rossi et al. (2003), que mostraram que o uso de marcas com dosagens diferentes de *single dose* pode agregar uma grande quantidade de informação para as análises genéticas em cana-de-açúcar.

Apesar dos modelos desenvolvidos, a identificação completa de genótipos e a dosagem alélica em poliploides ainda é um desafio, mesmo diante do desenvolvimento das modernas tecnologias computacionais e de sequenciamento genômico. Estudos de avaliação dos métodos de identificação completa de genótipos em poliploides demonstram que nenhum dos métodos disponíveis funciona adequadamente (Voorrips et al., 2011; Grandke et al., 2016). O problema se encontra nas distribuições de frequências das intensidades de sinais obtidos no sequenciamento de cada uma das classes genotípicas (Grandke et al., 2016). Os softwares desenvolvidos para a identificação completa de genótipos se baseiam na existência de diferenças de sinal entre as diferentes classes genotípicas (Figura 8A). Entretanto, muitas vezes ocorre a sobreposição das distribuições observadas das intensidades de sinais para as diferentes classes genotípicas (Figura 8B), o que acarreta em viés na classificação dos genótipos (Grandke et al., 2016). Este fator, conjuntamente à sobreposição das probabilidades de observação das classes genotípicas (Ripol et al., 1999) dificulta a identificação correta dos genótipos em poliploides. Uitdewilligen et al. (2017) estimam em que uma cobertura de 60X a 80X é necessária para a obtenção de 98,4% de acurácia na genotipagem de autotetraploides.

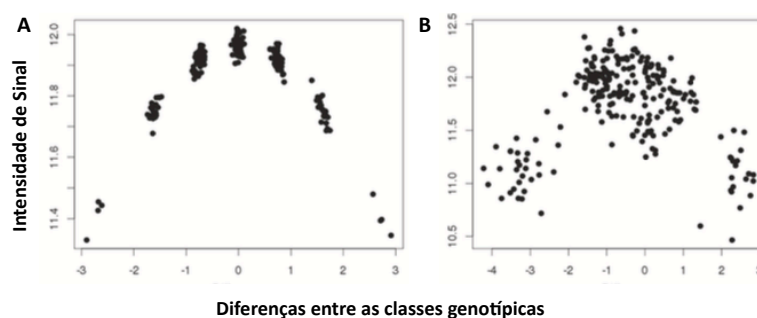


Figura 10. Distribuição das intensidades de sinal obtidas para diferentes genótipos em dados simulados (A) e em dados observados para um hexaploide (crisântemo). Fonte: Adaptado de Grandke et al. (2016).

Grandke et al. (2016) testaram o uso de genótipos contínuos em um estudo de associação genômica ampla (GWAS) para crisântemo (hexaploide). Este estudo mostrou que os erros na atribuição de genótipos resultam em perda de associações significativas entre marcadores e os caracteres analisados. Desta maneira, o uso de genótipos livres de pressuposições quanto aos níveis de ploidia produziu melhores resultados. O uso de genótipos contínuos evitou a inserção de erros no modelo, além de tornar a análise mais simples, por evitar a fase de genotipagem. É interessante salientar que, no caso de cana-de-açúcar, a adoção de um único nível de ploidia geraria viés na análise, dada a presença de aneuploidias e a existência de diferentes números de cromossomos nos diferentes grupos homo(e)ólogos.

É importante destacar que o uso de informações que possibilitam a captação dos efeitos da ploidia pode complementar os tradicionais métodos de análise de dados de marcadores moleculares. Neste sentido, novos métodos estatísticos devem ser desenvolvidos para permitir a análise e a interpretação de dados de espécies poliploides. Esta necessidade é reforçada pela disponibilidade atual de novas tecnologias de genotipagem baseadas em sequenciamento, de forma que o desenvolvimento de novas abordagens de análise irá permitir o melhor uso dos recursos existentes para a análise genética de espécies poliploides.

CAPÍTULO 3

RESEARCH AND PARTNERSHIP IN STUDIES OF SUGARCANE USING MOLECULAR MARKERS: A SCIENTOMETRIC APPROACH¹

Ivone de Bem Oliveira ^{2,3*}; Rhewter Nunes ^{2,4}; Lucia Mattiello ⁵; Stela Barros Ribeiro ²;
Isabela Pavanelli de Souza ²; Alexandre Siqueira Guedes Coelho ²;
Rosane Garcia Collevatti ⁴

¹Capítulo submetido como artigo no periódico científico Plos One;

²Programa de Pós-graduação em Genética e Melhoramento de Plantas (PGMP), Escola de Agronomia, Universidade Federal de Goiás (UFG), Goiânia, Goiás, Brazil.

³Institute of Food and Agricultural Science, University of Florida (UF), Gainesville, Florida, USA.

⁴Laboratório de Genética & Biodiversidade, Instituto de Ciências Biológicas, Universidade Federal de Goiás (UFG), Goiânia, Goiás, Brazil.

⁵Instituto de Biologia, Departamento de Genética, Evolução, Microbiologia e Imunologia, Universidade Estadual de Campinas (UNICAMP), Campinas, São Paulo, Brazil

*Corresponding author

3 RESEARCH AND PARTNERSHIP IN STUDIES OF SUGARCANE USING MOLECULAR MARKERS: A SCIENTOMETRIC APPROACH

3.1 ABSTRACT

In this study, a science mapping framework was applied to investigate the scientific knowledge, the temporal trends and research partnership in genetic studies of sugarcane using molecular markers. For this we performed a search on Web of Science database Thomson Reuters, 344 articles were retrieved considering the time frame from 1974 to 2017. Our findings show that there is still a lot to explore in order to produce knowledge for this crop with the use of molecular markers, because few topics (characterization and association studies) represented most (~82%) of the subjects explored in the articles. Furthermore, marker information is under-exploited for the culture and related knowledge is structured in only a few journals. Dominant markers were more used than markers based in single nucleotide polymorphism (SNPs) and a significantly low number of studies are using high-throughput sequencing technologies. Our results showed high connectivity among researchers and institutions publishing worldwide, but few authors represented a bridge between different research groups. Thus, despite the indication of increase in human resources in the research area, the international partnerships still low. We also found dominance in country publication. Six countries (India, Brazil, USA, Australia, China, and France) were responsible for 80% of the published articles and received most citations. We suggest special reinforcement in building international partnership, since collaborations still occurring mostly within countries. International collaboration could lead to a better exploration and more development of knowledge for sugarcane.

Keywords: *Saccharum* spp., allele dosage; co-dominant molecular markers; dominant molecular markers, genetic markers, network; polyploidy.

3.2 INTRODUCTION

Sugarcane (*Saccharum* spp.) is one of the most important agricultural commodities in the world. It is cultivated in both tropical and subtropical regions. About 50.2% of the world production comes from the Americas, 40.5% from Asia, 6.5% from Africa and 2.8% from Oceania [1]. In 2016, this commodity achieved 27 million (M) hectares encompassing ~1,891 M tones, of which 71.9% was produced solely in Brazil and India [1]. The global production of processed sugar for the marketing year of 2017/2018 is forecasted to increase by 185 M tones [2]. In Brazil, the estimated production is 40.2 M tones and in India, 27.7 M tones, representing 36.7 % of the world total sugar production [2]. The use of sugarcane as a biofuel feedstock has been renewing the interest of researchers for this crop [3, 4].

Due to its high importance worldwide, sugarcane breeding programs have been challenged to release new varieties in a fast and cost-effective manner, considering different traits including: disease resistance, high yield, drought and waterlogging tolerance [4, 5]. The use of molecular markers has been improving the effectiveness of conventional plant breeding achieving a great impact on crop production [6-8]. In diploid species, the use of advanced methods in breeding programs, such as genome wide selection (GS) is already common [9]. However, the lack of both theoretical knowledge and proper analytical methods for polyploid species may hinder the achievement of similar success compared to diploid crops [7, 10, 11]. The use of molecular markers for sugarcane is still behind its possibilities. Most of the studies with sugarcane are based on oversimplified methods, in which only a small part of the data is used (*i.e.* single dose markers), the genotypic information is assumed as pseudo-diploid, or the marker information is simplified to presence and absence of a reference allele in the individual's genotypes [7, 11, 12]. This is not only due to the species' high ploidy levels, which vary for each homeologous groups of chromosomes, but also to the high complexity of the sugarcane genome.

Sugarcane genetics is widely acknowledged as one of the most complex among crops [13-16]. Modern cultivars have interspecific origins, high and variable ploidy levels (frequently aneuploidy), high heterozygosity and large genomes. The estimated number of chromosomes range from $2n = 100$ to $2n = 130$, distributed in 12 homeologous groups, with a total genome size ($2C$) of 10 Gbp (giga base pairs) [13-16].

The analysis of scientific production trend over time may contribute to the understanding of the development level in a specific field and allow the detection of research opportunities (e.g. [17,18]). Therefore, our analysis can help to understand the current status of researches using molecular markers in sugarcane and to unravel trends in the use of genetic information and the development of technology development for this crop.

Here, we use a science mapping analysis framework to summarize the current state of the art in the scientific knowledge of the use of molecular markers in sugarcane. We also identify the temporal trends and patterns of institutional and author partnerships to provide information for science and technology policies and to support decision-making processes for directing new research efforts.

3.3 MATERIAL AND METHODS

3.3.1. Data base survey and variables

We surveyed the Thomson Reuters Web of Science database (WoS) (available at: <http://apps.webofknowledge.com>) to search for publications using molecular markers in genetic studies of sugarcane. We used the most frequent keywords in the subject area: “genetic” and “marker*” or “molecular” and “marker*”, combined with “Saccharum” or “sugarcane”. We also included in our research the specific terms for molecular markers classes (SNP, SSR, AFLP, RAPD, RFLP, isoz* and microsat*). Only articles presented in the WoS core collection were analyzed. The survey was performed based on the availability of publications in the database from 1945 to 2017. We excluded reviews and studies that did not match the search theme.

For each article the following variables were analyzed: i) year; ii) journal; iii) author; iv) citations; v) country; vi) institutions of the authors; vii) WoS areas of research; viii) study subject (fingerprinting or characterization, genetic diversity, mapping, molecular marker development, marker assisted selection (MAS), comparative genomics or evolution, marker-trait association); ix) most common terms in the title and abstract; x) molecular marker used in the study. We also verified the information about the number of polymorphic loci and use of dosage information.

3.3.2. Science mapping analyses

The analyses and classification of the data were carried out using packages and functions implemented in R [19]. We used the package *bibliometrix* [20] to obtain information from the articles and statistics about the authors (H-index). Frequency distributions, regression, and correlation analyses and Student t-tests were performed using functions implemented in *stats* package [19].

We first analyzed the temporal trends in the number of articles published from 1945 to 2017. For this, we normalized the number of articles using molecular markers in sugarcane to the total number of articles using molecular markers in any plant species. To compare the annual increase of papers for sugarcane with the increase for other grass crop with high production worldwide [1], we repeated the same search replacing the sugarcane terms for “oryza” or “rice”. These results were also normalized using the same approach adopted for sugarcane. Such normalization removes the natural effect of the increase in scientific publications and yields unbiased estimates of temporal trends in scientific production. The normalized values were calculated as $z_i = (x_i - x_{min}) / (x_{max} - x_{min})$, where z_i is the normalized value, x_i is the number of articles using molecular markers for the specific topic (i.e. sugarcane or rice terms) and x_{max} and x_{min} are the maximum and minimum values observed for the search considering molecular markers for any plant species.

We then analyzed temporal trends in scientific production using linear regression models (normalized number of papers vs. year of publication) for each variable. The analyses were performed using the *cor.test* and *lm* functions implemented in the *stats* package in R environment [19]. To compare the regression coefficients we performed a t-test considering the angular coefficients (β) and the standard error of each regression (SE) ($t = \beta_1 - \beta_2 / \sqrt{(SE_{12} + SE_{22})}$).

We also analyzed the number of articles per journal and the authors' countries of origin. A bibliographic coupling network analysis was performed using the program VOSViewer v. 1.6.6 [21]. The association among journals was estimated by the number of cross citations. Co-citation sources with at least ten publications with five citations were considered.

We analyzed the distribution of the number of authors per article, as well as the total number of authors involved in the publications and their affiliations. The ranking of the ten most productive authors was also obtained. To analyze the citation and productivity

of each author we obtained the number of citations of each article and the most cited articles. We also obtained the *h-indexes* [22, 23] for the ten most productive authors. This result was obtained by the function *Hindex* implemented in the *bibliometrix* package [20] using the productivity and citation information for each author.

To analyze the temporal trend in collaborations based on the number of co-authors per article, we used a linear regression fitted with the *lm* function implemented in the *stats* package in R [19]. The co-citation and country partnership levels were estimated by counting the number of different countries and institutions involved in each publication and building networks to visualize the relationships using VOSViewer v. 1.6.6 [21]. For author partnership we considered only authors with a minimal number of ten articles and at least five citations. For the country co-authorship analysis, we considered all countries with at least one article, and a network was built only for connected countries.

We also analyzed the number of articles published using different molecular markers, the number of polymorphic loci, and whether the allele dose information was used in the study using the *bibliometrix* package [20].

The articles were assigned to specific subject areas using a two-step procedure. At first, we considered WoS's research area classification. For this, the articles were submitted to the WoS analysis, in which a specific article can be classified in multiple research areas. In the second stage, the articles were manually assigned to seven different areas (fingerprinting or characterization, genetic diversity, mapping, molecular marker development, marker assisted selection, comparative genomics or evolution, and marker-trait association). In this analysis, each article was classified in only one subject area.

A term co-occurrence map was built using VOSViewer v. 1.6.6 [21] to describe the main topics in the articles. This analysis was based on text data obtained from the WoS's search. The full count of terms was extracted from the title and abstract fields. We selected terms that appeared at least 30 times in the set of articles and applied a filter based in the relevance score, keeping the 60% most relevant terms. All unspecific terms were excluded.

3.4 RESULTS

3.4.1. Temporal trends

Our survey retrieved 344 articles reporting the use of molecular markers in sugarcane (S1 Table). A total of 116,091 articles were published using molecular markers for any plant species and 9,734 specifically for rice. The first article in this subject was published in 1974, thus the temporal analysis includes the time frame between 1974 and 2017 (Fig 1).

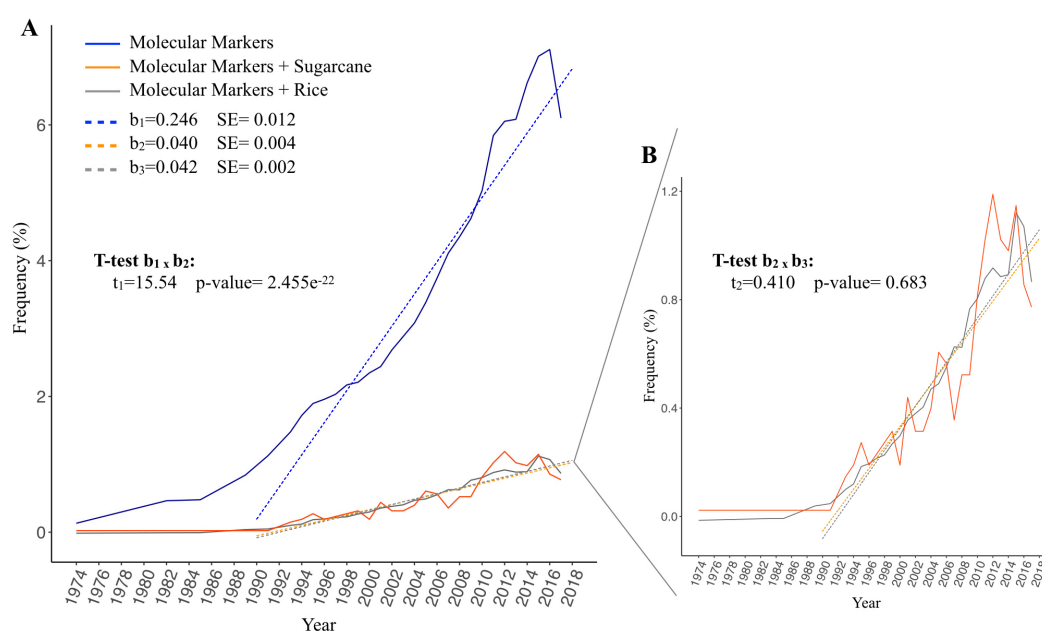


Fig 1. Temporal trends in the number of publications using molecular markers for any plant species, sugarcane and rice. Analyses based on the Thomson Reuters Web of Science database and considering the period from 1974 to 2017. Continuous lines represented the distribution of the number of papers for each search. Dashed lines represented the regression obtained for each set of publications. Student's t-test values represent the comparison of the estimated increase of use of molecular markers for all plant species and sugarcane (t_1), and the comparison between the estimated increase of use of molecular markers for sugarcane and rice (t_2).

The mean annual growth rate in the number of articles was 10.69%, but only 1.16% of the articles were published before 1990. A high increase in the number of articles using molecular markers in sugarcane occurred after 1990, most likely due to the Polymerase Chain Reaction (PCR) development. When we considered the normalized frequencies, although the number of articles published in sugarcane is relatively smaller in

relation to the total number of articles found with any plant species, a significant and strong correlation was observed between the two distributions ($r = 0.94$; $p < 2.23\text{E-}15$, Fig 1). A similar result was observed when we compared the results for sugarcane and for rice ($r = 0.87$; $p = 8.61\text{E-}10$, Fig 1). We also found a significant difference between the growth rate in the number of articles per year for any plant species and sugarcane ($t = 15.51$, $p = 2.455\text{E-}22$), but the comparison between sugarcane and rice was not significantly different ($t = 0.41$, $p = 0.6830$).

3.4.2. Sources

The 344 analyzed articles were published in 102 different journals, but most journals published less than 4 articles (81.37%, S2 Table). The journal Theoretical and Applied Genetics published the highest number of articles (TAG; 39), followed by Sugar Tech (ST; 29), Euphytica (EU; 25) and Molecular Breeding (MB; 19). Ten journals comprised nearly 50% of all publications (Fig 2A) and were considered the main sources of publications using molecular markers in sugarcane.

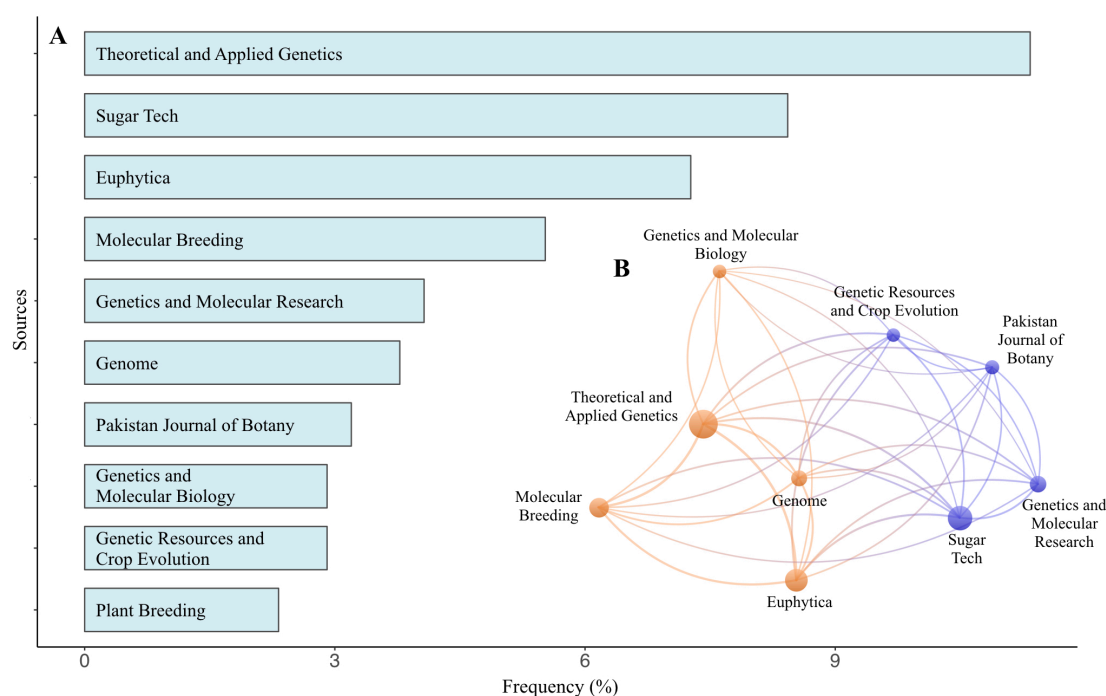


Fig 2. Main sources of publications using molecular markers in sugarcane and their bibliographic coupling relationship. Analyses based on the Thomson Reuters Web of Science database and considering the period from 1974 to 2017. A) Ranking of journals that published at least ten articles; B) Network analysis representing the bibliographic coupling among the 10 journals that published at least ten articles and with at least five citations.

The co-citation analysis indicated two groups of bibliographic coupling (Fig 2B, S3 Table). TAG, EU, MB, Genome (GE), and Genetics and Molecular Biology (GMB) composed the first group, and the second group comprised ST, Genetic Resources and Crop Evolution (GRCE), Genetic and Molecular Research (GMR), Pakistan Journal of Botany (PJB), and GMB (Fig 2B, S3 Table).

Considering the authors' affiliation countries, most of the publications in Theoretical and Applied Genetics were developed in France. Sugar Tech publications were mainly from Indian authors. The publications in Molecular Breeding came mostly from USA and Australia and in Genetics and Molecular Research came mostly from Brazilian authors (S1 Fig).

3.4.3. Authors

We identified 1,019 researchers publishing on the use of molecular markers in genetic studies of sugarcane. Only three articles were single authored and the mean number of co-authors per article was 5.88. The number of authors per article slightly increased since 1974 ($r = 0.22$; $p = 3.89E-05$; Fig 3A). The total number of authors publishing per year, discounting duplicates, increased significantly in the analyzed period ($r = 0.86$; $p = 1.64E-09$; Fig 3B).

Angelique D'Hont (CIRAD, France) published most of the articles (37), followed by Jean C. Glaszmann (25; CIRAD, France), Karen S. Aitken (23; CSIRO, Australia), Cathrine L. McIntyre (21 CSIRO, Australia; Fig 3C), and Yong-Bao Pan (21; USDA-ARS, United States of America) (S4 Table, Figure 3C). The co-authorship analysis showed five groups (Fig 3D). Angelique D'Hont (node 15), Jean C. Glaszmann (node 12), Laurent Grivet (13), Cathrine L. McIntyre (8), and Karen S. Aitken (10) had the five highest total link strengths (Fig 3D). The co-authorship groups also corresponded to the affiliation of the authors. Therefore we had one group comprising Brazil affiliated authors (Fig 3D, nodes 1 to 4), one comprising North America and China affiliated authors (Fig 3D, nodes 5 to 7), one with Australia affiliated authors (Fig 3D, nodes 8 to 10), and one formed by France affiliated authors (Fig 3D, nodes 11 to 15). These countries also had the highest representativeness in publications on the theme (Fig 4A). Although Singh, R.K. from India had more than ten papers and was cited 78 times, he didn't have any connection in the network analysis (total strength = 0; S5 Table). We found a high and significant correlation between the total link strength of the co-authorship network and the H-index

for the ten most productive authors ($r = 0.96$; $p = 1.21\text{E-}05$; S5 Table).

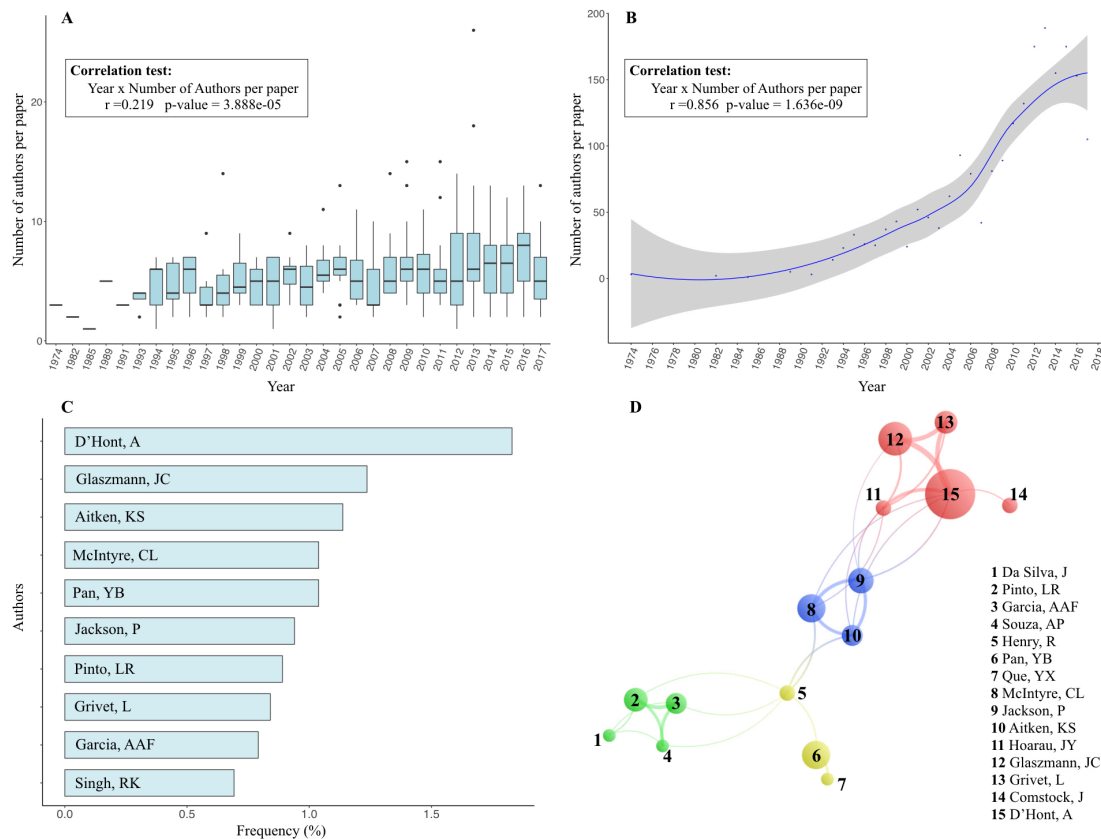


Fig 3. Number of authors per paper, total number of authors per paper, main authors, and network analysis depicting the collaboration between authors. A) Box plot representing the variation in the number of authors per paper in the period analyzed (1974-2017); B) Regression using Loess method portraying the total number of authors within articles per year; C) Network analysis depicting the collaboration between authors that presented at least 10 papers and 5 citations, and at least one collaboration inside the 344 papers analyzed.

3.4.4. Citations

Overall, the 344 articles were cited 5,735 times (S1 Table). Twelve articles showed more than 80 citations and only 85 articles (24.71%) received more than 20 citations (S5 Table). Considering the total citation per author, the five most cited authors were Angelique D'Hont (1,492 citations), followed by Jean C. Glaszmann (1,257), Laurent Grivet (1,015), Cathrine L. McIntyre (818), and Karen S. Aitken (444) (S5 Table). Table S6 describes the most cited papers and their associated information.

3.4.5. Countries and Institutions

More than 80% of the studies were developed by researchers from India, Brazil, USA, Australia, France, and China (Fig 4A). These countries also showed the highest total link strength values in the collaboration analysis (Fig 4B, S7 Table). From the most frequent countries associated with the publications in the area (Fig 4A), USA and France researchers published in almost the entire period analyzed (Fig 4C). India had an increase in publication after 2010 (Figure 4C). Brazil also participated with significant proportion of publications (Figure 4C) mostly after 2010.

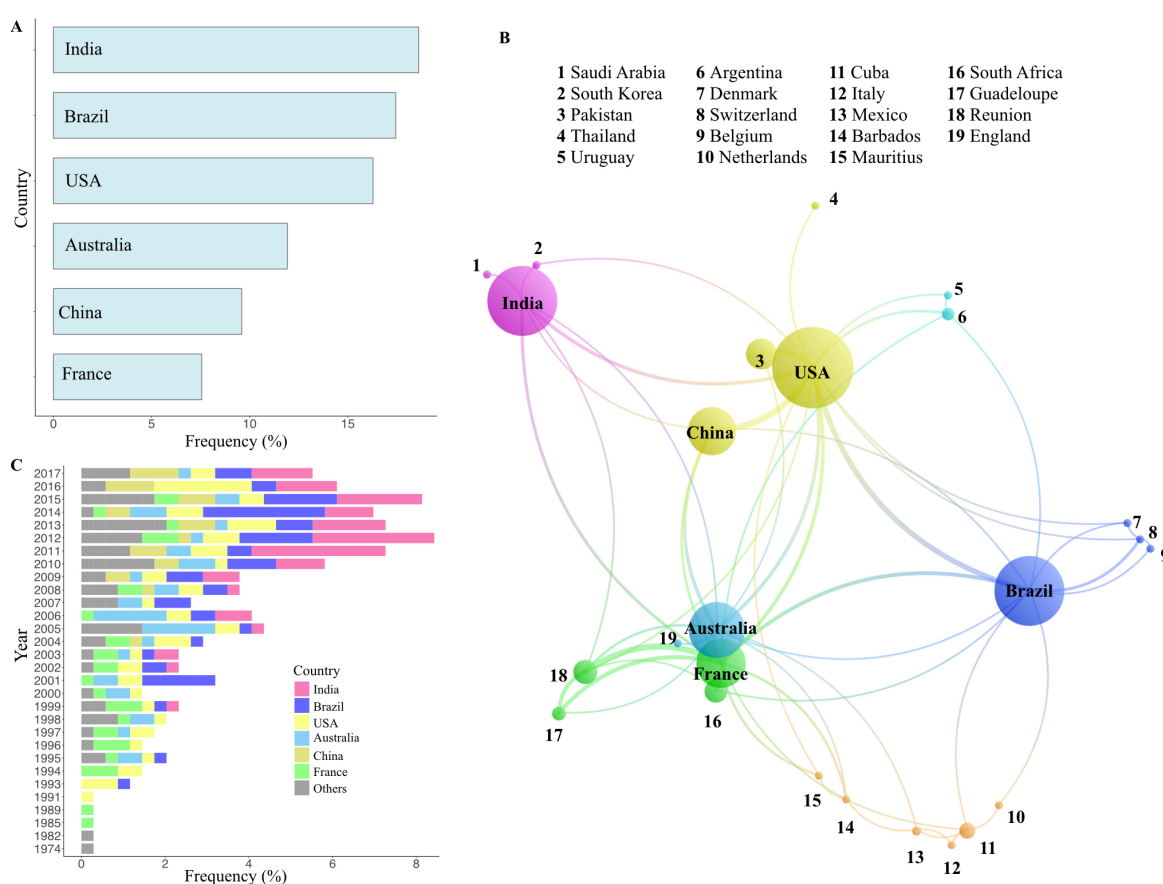


Fig 4. Ranking and partnership among countries in publications with molecular markers in sugarcane. A) Countries that published at least 20 articles; **B)** Network depicting the collaboration among countries that presented at least one connection with each other; **C)** Proportion of country publications per year for the most frequent countries.

Considering the institutions obtained with the WoS enhanced institution analysis, 248 organizations contributed to the production of the 344 articles. Seven institutions produced at least 25 articles: Indian Council of Agricultural Research (ICAR,

110), United States Department of Agriculture (USDA, 49), Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD, 46), and Commonwealth Scientific Industrial Research Organisation (CSIRO, 34), Universidade de São Paulo (USP, 29), Texas A. M. University College Station (TAMU, 26), and Universidade Estadual de Campinas (UNICAMP, 26). These institutions represented 35.1% of all author's affiliations (S8 Table).

3.4.6. Research areas and description of the articles' main topics

The WoS's research area classification attributed more than 50.0% of the classifications to "Plant sciences" (26.0%) and "Agronomy" (23.5%) (S9 Table). Considering the classification of the articles in seven areas (S1 Table). From the 344 articles, 36.05% (124) used molecular markers for fingerprinting of sugarcane varieties or species, 24.13% (83) for genetic diversity, and 21.80% (75) performed marker-trait association studies (S1 Table). These three subjects represented 81.98% of the total number of articles. The remaining 18.02% were classified as mapping analysis (28 articles, 8.14%), development of markers (25, 7.27%), marker-assisted selection (5, 1.45%), and comparative or evolutionary studies (4, 1.16%).

We found 40 main terms in the articles distributed in four groups (Fig 5). One group (Fig 5, red) contains the term "genetic diversity" linked to different molecular markers or terms related to molecular markers (*e.g.* RAPD, SSR, primer, band, fragment), as well as "cultivar" and "hybrids". In the second group (Fig 5, green) the term "species" received the highest score connected to species related terms (*e.g.* *S. spontaneum*, *S. officinarum*), as well as to "sequence" and "genome". The third group (Fig 5, blue) was composed by mapping-associated terms (*e.g.* map, linkage groups, QTL, *i.e.* quantitative trait loci). For the last group (Fig 5, yellow) "resistance" was the main term and the related terms "brown rust" and "bru1" (name of the genetic region associated with the resistance of the brown rust in sugarcane) were also present.

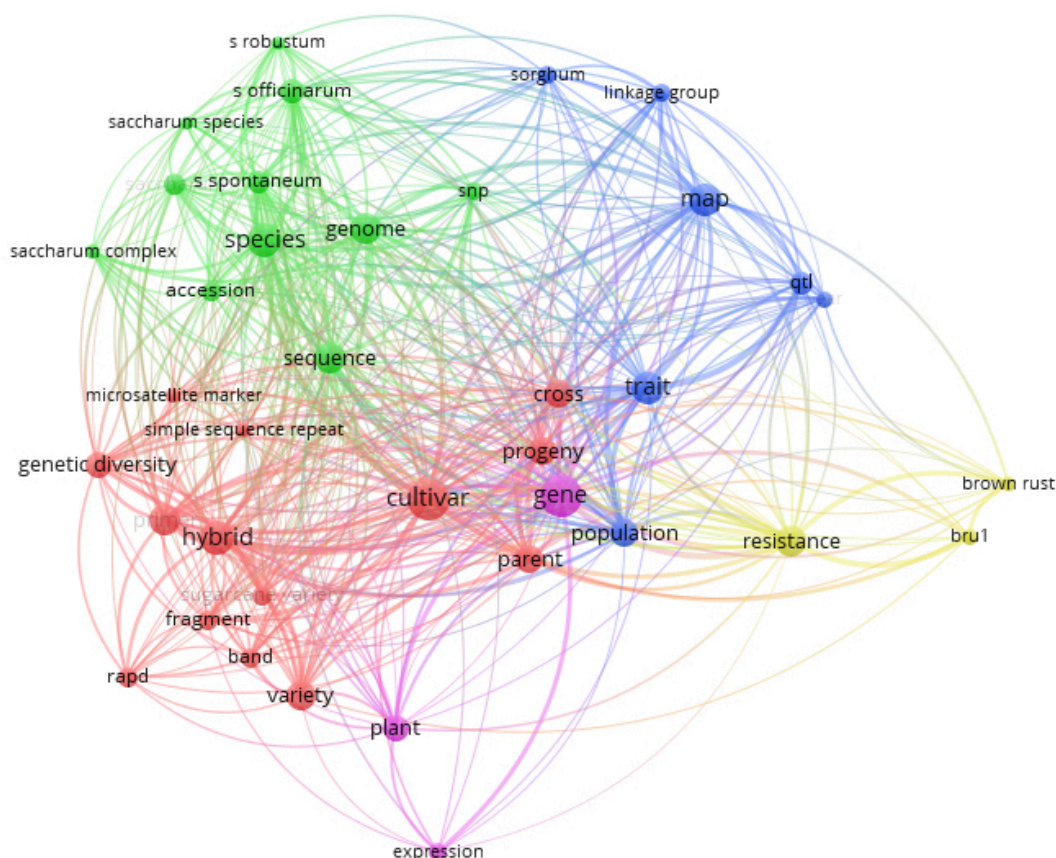


Fig 5. Network representing the relationship among the 40 most relevant terms in publications with molecular markers in sugarcane. Terms co-occurrence map was built using the terms in title and abstract that appeared at least 30 times in the set of articles.

3.4.7. Markers

The articles mention 20 classes of different molecular markers used in studies of sugarcane (S1 Table). Microsatellites or Simple Sequence Repeat (SSR) were the most used molecular markers (34.45%), followed by RAPD (17.07%), AFLP (14.94%), RFLP (14.02%) and SNPs (6.10%) (Fig 6). Some articles reported the use of more than one class of molecular markers (S1 Table). For example, 17 articles reported the use of two classes, 12 articles reported the use of three classes, and two articles reported information from four classes of markers (S1 Table). The use of SSR, AFLP and RAPD showed an increase since 2004 (Fig 7). DArT and SNP markers were used after 2000 and represented 10.06% of the articles considering the main marker classes (Fig 7).

Sugarcane presents multiple copies of a given locus leading to high number of alleles and high heterozygosity. Therefore, a single marker can yield more than one band and the number of bands is normally used to provide information about the number of

“molecular markers” applied in the studies instead of the real number of markers used. Hereafter, the number of bands used in the studies will be referred as “marker”.

In our survey, 41.00% of the articles reported the use of less than 100 markers, 14.00% reported the use of less than 200, and 10.33% less than 300 markers. Only 13.33% of the studies reported the use of more than 1,000 markers (S1 Table). Due to the high ploidy levels of sugarcane, an important information in molecular markers applications is allele dosage. Despite the evidence of dosage effects in expression of important economic traits in polyploids [24], only eight articles reported the use of information of allele dosage. From the eight articles, seven reported the use of information from single and double-single dose markers and only one reported the use of information of more complex dosages (single, double, and triple dose markers; S1 Table).

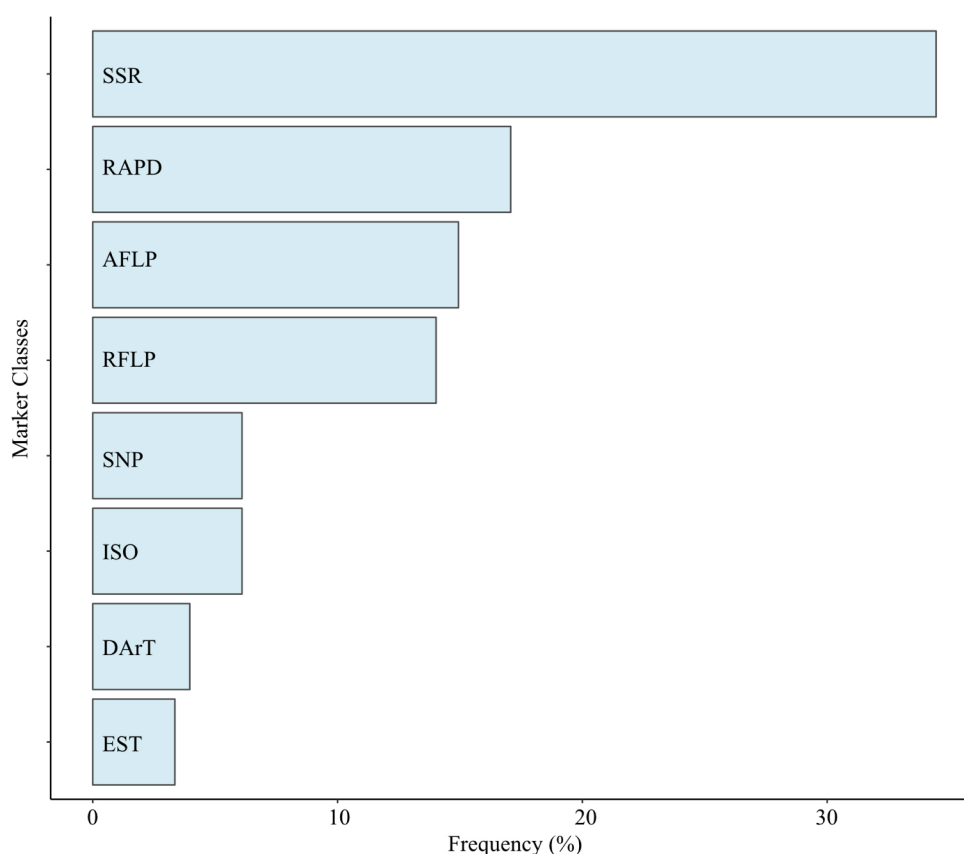


Fig 6. Ranking of molecular markers classes. Ranking of molecular markers classes considering the markers that were used at least 20 times in the set of analyzed articles.

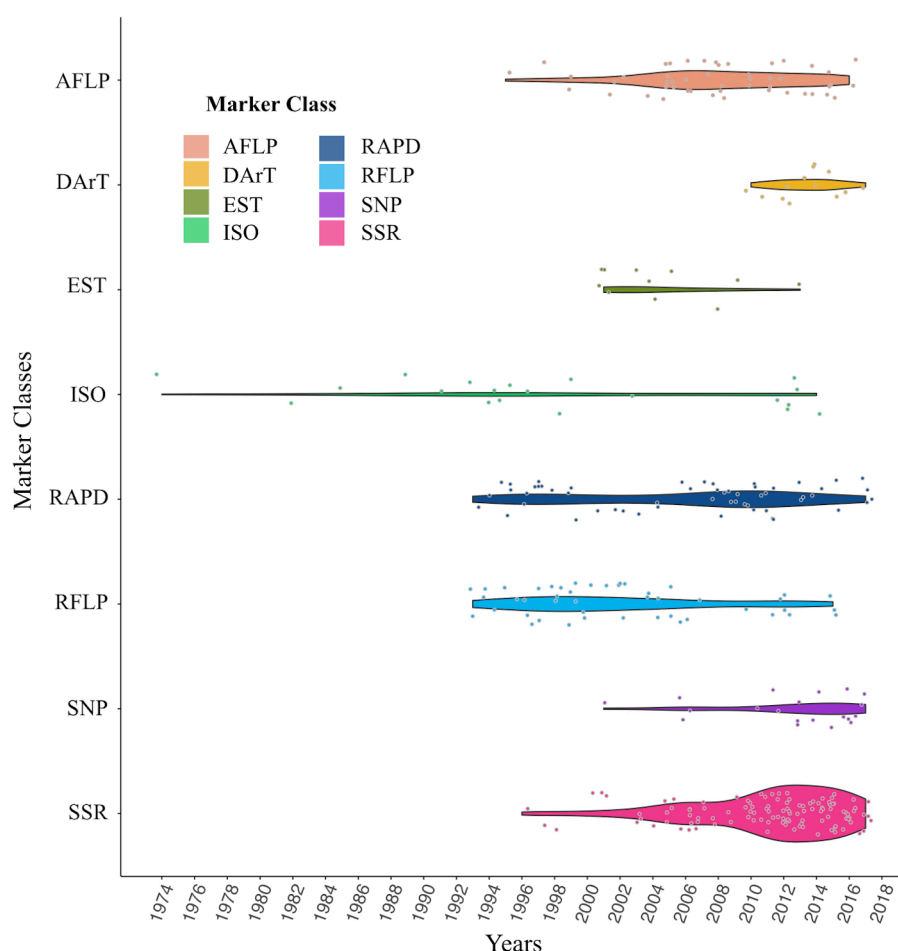


Fig 7. Temporal trends in the use of molecular markers in publications of sugarcane studies. Temporal trends in the use of the eight main molecular markers classes.

3.5 DISCUSSION

3.5.1. Use of molecular marker information in sugarcane genetic studies

In this study, a scientometric analysis was applied to investigate the use of molecular markers in genetic studies for sugarcane. Temporal trends were analyzed and compared with the use of molecular markers in genetic studies of any plant species and also for rice, another important grass crop. Our results point out an increase in publication in the area, but also show that molecular marker data have been underexploited for sugarcane. 102 journals had published in the area but nine journals represent more than 50% of the total publications. The same occur with the institutions involved in the studies 248 were identified but only 7 produced more than 25 articles. The number of authors is increasing with time, nevertheless the international partnership remain low, with few

authors represented a bridge between different research groups, and a only six countries been responsible for 80% of the published articles.

Our results showed a significant increase in publications for studies that reported the use of molecular markers for genetic studies in sugarcane in recent decades. Even though this growth was significantly smaller than the growth in the number of publications for any plants, when we compared the temporal trends estimated for rice and sugarcane, no significant difference was observed (Fig 1).

The main difference between the use of molecular markers in sugarcane and other crops lies in the quantity and quality of information obtained from these tools. The advent of new DNA sequencing technologies allowed the development of a high number of markers throughout the genome of many diploid species. For example, in rice, genetic diversity analysis, association studies, linkage mapping, marker-assisted selection and evolutions studies have been performed with high-throughput markers such as SNP and DArT [25-29]. Meanwhile, only 6.10% of the sugarcane studies reported the use of SNP and only 13.33% of the articles reported the use of more than one thousand markers. Most studies with sugarcane used dominant molecular markers such as RAPD and AFLP (Fig 6). In addition, due to the ploidy levels and genome complexity of sugarcane, the information generated from co-dominant molecular markers is often interpreted only as presence and absence of markers. The study of polyploids using molecular markers is complicated by the possibility of segregation of multiple alleles in populations of these species. Moreover, chromosome pairing in meiosis may also lead to unexpected segregation (*i.e.* double reduction) in polyploids [30]. For these reasons, most studies in sugarcane were developed using makers that fit in an expected Mendelian class by segregation analysis, mostly using single-dose markers (alleles with only one dose and with Mendelian segregation of 1:1 in a biparental cross) [11,31].

Despite evidences that dosage information can improve the amount of information retrieved for sugarcane [32, 33], few studies were developed exploring dosage levels other than single-dose [12, 34]. Because of the lack of methodology and software for genetic analyses of polyploids, the information that molecular markers can generate for sugarcane have been underexplored.

3.5.2. State of the art of studies involving the application of molecular markers in sugarcane

Our results showed high connectivity among researchers and institutions publishing worldwide, but few authors are responsible for the bridge between different research groups (Fig 3D). Despite the increase in the number of co-authors per article, which may indicate increase in human resources in the research area, the network analysis shows that the international partnership is still low. We also found an uneven distribution of publications per country. Six countries (India, Brazil, USA, Australia, China, and France) were responsible for 80% of the published articles and received most of the citations. USA, Australia, Brazil and France, were responsible for the majority of citations in the area (18.29%, 16.71%, 16.05%, and 15.00%, respectively; S7 Table). Although a considerable number of institutions were represented in research affiliation (248), only seven institutions published more than 25 articles (ICAR, USDA, CIRAD, CSIRO, USP, TAMU, and UNICAMP), and 215 institutions published less than three articles. It is important to note that the majority of collaborations also occurred inside research groups and within countries (Fig 3D and 4B).

Although the number of publications has significantly increased over time (Fig 1), as well as the number of authors participating in these publications (Fig 3B), the total number of journals publishing in the area (102) is still low considering the total number of journals available at WoS's core collection. Moreover, the publications in the area are mainly focused in few journals (Fig 2), with 50.00% of the publications coming from nine journals (TAG, ST, EU, MB, GMR, GE, PJB, GRCE, and GMB). The most frequent journals are from subject areas of plant science, agronomy, genetics heredity, and horticulture (S9 Table). Considering the subject classification, the majority of the studies were related to fingerprinting (describing or characterizing species or cultivars), followed by genetic diversity analysis, and association studies. These topics comprised 81.98% of the total number of articles. These results show that there is still much to explore in generating knowledge for sugarcane.

3.6 CONCLUSIONS

Our findings show that the use of molecular markers in sugarcane is still underexplored. The development of new techniques for the interpretation of genetic data

for sugarcane would be a way out of this dilemma, since one of the main obstacles to obtain more genetic information for the crop is the unavailability of methods to deal with the polyploidy and to estimate allele dosage. The development of new methods may complement the existing approaches that support the use of molecular markers data in sugarcane genetics [35]. This new type of data may also contribute in the genetic breeding for the crop. Future prospects are promising since our results findings showed a regular increase in publications reporting sugarcane research results. We suggest that the arrangement of international collaborations and the development of new methods for the interpretation and use of more complex genetic data in sugarcane are strategic landmarks that should be pursued in the near future. Furthermore, investment in human resources development in this area is essential since the number of authors and partnerships is still low, considering the amount of articles that have already been published. It is common to see that the cutting-edge articles involving molecular markers in sugarcane are, most of the time, published by the same groups.

3.7 ACKNOWLEDGMENTS

SBR and IPS was supported by CAPES PhD fellowship. IBO was supported by CAPES Foundation within the Ministry of Education, Brazil (grant n. 88881.131685/2016-01). RN has been supported by INCT-EECBio (grant n. 465610/2014-5). LM was supported by BEPE FAPESP fellowship (grant n. 2017-09817-0), and fellowship FAPESP (grant n. 2015/23789-4). RGC has been supported by productivity grants from CNPq, which we gratefully acknowledge.

3.8 REFERENCES

1. FAOSTAT, FAO. Statistical data. Food and Agriculture Organization of the United Nations, Rome, 2017.
2. USDA, Sugar: World Markets and Trade. 2017. Available in: <https://apps.fas.usda.gov/psdonline/circulars/sugar.pdf>.
3. Menossi, M., Silva-Filho, M. C., Vincentz, M., Van-Sluys, M. A., & Souza, G. M.. Sugarcane functional genomics: gene discovery for agronomic trait development. **International journal of plant genomics**, v. 2008.

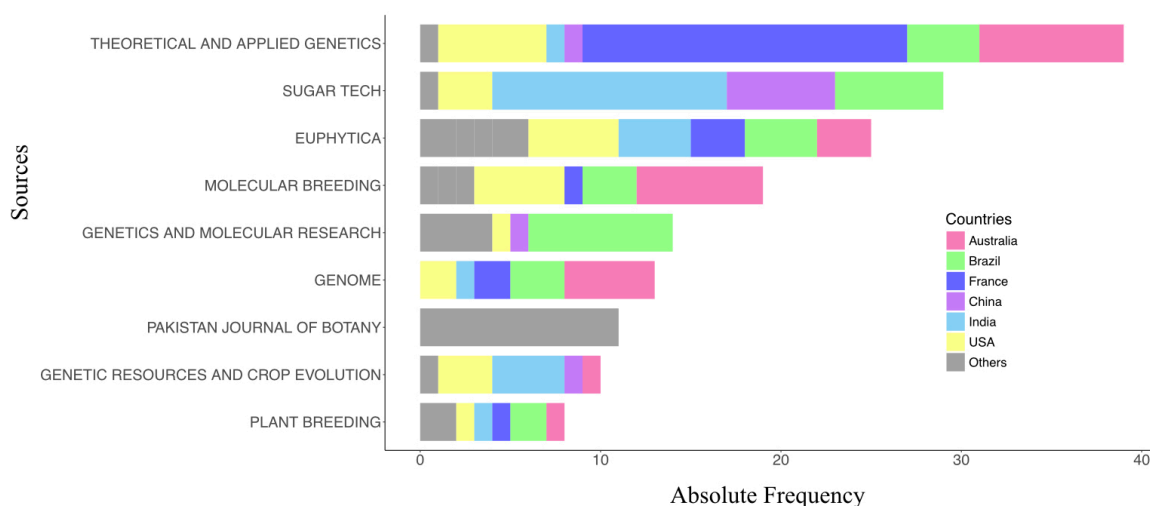
4. Aitken K., Li J, Piperidis G, Qing C, Yuanhong F, Jackson P. Worldwide Genetic Diversity of the Wild Species *Saccharum spontaneum* and Level of Diversity Captured within Sugarcane Breeding Programs. *Crop Science*, 2018;58:218-229. doi:10.2135/cropsci2017.06.0339
5. Li WF, Shan HL, Zhang RY, Pu HC, Wang XY, Cang XY, et al. Identification of field resistance and molecular detection of the brown rust resistance gene *Bru1* in new elite sugarcane varieties in China. *Crop Prot.* Elsevier Ltd; 2018;103: 46–50. doi:10.1016/j.cropro.2017.09.007
6. Moose SP, Mumm RH. Molecular Plant Breeding as the Foundation for 21st Century Crop Improvement. *Plant Physiol.* 2008;147: 969–977. doi:10.1104/pp.108.118232
7. Gouy M, Rousselle Y, Bastianelli D, Lecomte P, Bonnal L, Roques D, et al. Experimental assessment of the accuracy of genomic selection in sugarcane. *Theor Appl Genet.* 2013;126: 2575–2586. doi:10.1007/s00122-013-2156-z
8. Gouy M, Rousselle Y, Thong Chane A, Anglade A, Royaert S, Nibouche S, et al. Genome wide association mapping of agro-morphological and disease resistance traits in sugarcane. *Euphytica.* 2014;202: 269–284. doi:10.1007/s10681-014-1294-y
9. Piepho HP, Möhring J, Melchinger AE, Büchse A. BLUP for phenotypic selection in plant breeding and variety testing. *Euphytica.* 2008;161: 209–228. doi:10.1007/s10681-007-9449-8
10. de Setta N, Monteiro-Vitorello CB, Metcalfe CJ, Cruz GMQ, Del Bem LE, Vicentini R, et al. Building the sugarcane genome for biotechnology and identifying evolutionary trends. *BMC Genomics.* 2014;15. doi:10.1186/1471-2164-15-540
11. Garcia AAF, Mollinari M, Marconi TG, Serang OR, Silva RR, Vieira MLC, et al. SNP genotyping allows an in-depth characterisation of the genome of sugarcane and other complex autopolyploids. *Sci Rep.* 2013;3: 1–10. doi:10.1038/srep03399
12. Silva JAG da, Sorrells ME, Burnquist WL, Tanksley SD. RFLP linkage map and genome analysis of *Saccharum spontaneum*. *Genome.* 1993;36: 782–791. doi:10.1139/g93-103
13. D'Hont a, Lu YH, León DG, Grivet L, Feldmann P, Lanaud C, et al. A molecular approach to unraveling the genetics of sugarcane, a complex polyploid of the *Andropogoneae* tribe. *Genome.* 1994;37: 222–230. doi:10.1139/g94-031
14. D'Hont A, Grivet L, Feldmann P, Glaszmann JC, Rao S, Berding N. Characterisation of the double genome structure of modern sugarcane cultivars (*Saccharum* spp.) by molecular cytogenetics. *MGG Mol Gen Genet.* 1996;250: 405–413. doi:10.1007/BF02174028
15. Piperidis G, Piperidis N, D'Hont A. Molecular cytogenetic investigation of chromosome composition and transmission in sugarcane. *Mol Genet Genomics.* 2010;284: 65–73. doi:10.1007/s00438-010-0546-3

16. Riaño-Pachón DM, Mattiello L. Draft genome sequencing of the sugarcane hybrid SP80-3280. *F1000Research*. 2017;6: 861. doi:10.12688/f1000research.11859.1
17. Tinoco CF, Lima NE, Lima-Ribeiro MS, Collevatti RG. Research and partnerships in studies on population genetics of Neotropical plants: A scientometric evaluation. *Biochem Syst Ecol*. 2015;61: 357–365. doi:10.1016/j.bse.2015.07.002
18. Liu B, Zhang L, Wang X. Scientometric profile of global rice research during 1985-2014. *Curr Sci*. 2017;112: 1003–1011. doi:10.18520/cs/v112/i05/1003-1011
19. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, 2018. Available from: <https://www.R-project.org/>.
20. Aria M, Cuccurullo C. Bibliometrix: An R-tool for comprehensive science mapping analysis, *Journal of Informetrics*. 2017;11:959-975, Elsevier, doi: 10.1016/j.joi.2017.08.007
21. van Eck NJ, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*. 2010;84: 523–538. doi:10.1007/s11192-009-0146-3
22. Hirsch JE. An index to quantify an individual's scientific research output. *Proc Natl Acad Sci*. 2005;46:16569–16572, doi:10.1073/pnas.0507655102
23. Hirsch JE. Does the H index have predictive power? *Proc Natl Acad Sci*. 2007; 49:19193–19198, doi:10.1073/pnas.0507655102
24. Osborn TC, Pires JC, Birchler JA, Auger DL, Chen ZJ, Lee HS, et al. Understanding mechanisms of novel gene expression in polyploids. *Trends in Genetics*, 2003; 3:141-147, doi:10.1016/S0168-9525(03)00015-5
25. Fuchs EJ, Meneses Martínez A, Calvo A, Muñoz M, Arrieta-Espinoza G. Genetic diversity in *Oryza glumaepatula* wild rice populations in Costa Rica and possible gene flow from *O. sativa*. *PeerJ*. 2016;4: e1875. doi:10.7717/peerj.1875
26. Li G, Wang Y, Chen M-S, Edae E, Poland J, Akhunov E, et al. Precisely mapping a major gene conferring resistance to Hessian fly in bread wheat using genotyping-by-sequencing. *BMC Genomics*. 2015;16: 108. doi:10.1186/s12864-015-1297-7
27. Yang H, You A, Yang Z, Zhang F, He R, Zhu L, et al. High-resolution genetic mapping at the Bph15 locus for brown planthopper resistance in rice (*Oryza sativa* L.). *Theor Appl Genet*. 2004;110: 182–191. doi:10.1007/s00122-004-1844-0
28. Morais Júnior OP, Duarte JB, Breseghello F, Coelho ASG, Borba TCO, Aguiar JT, et al. Relevance of additive and non-additive genetic relatedness for genomic prediction in rice population under recurrent selection breeding. *Genet Mol Res*. 2017;16. doi:10.4238/gmr16039849
29. Choi JY, Purugganan MD. Multiple Origin but Single Domestication Led to *Oryza sativa*. *G3: Genes|Genomes|Genetics*. 2018;8: g3.300334.2017. doi:10.1534/g3.117.300334

30. Amadeu RR, Cellon C, Olmstead JW, Garcia AAF, Resende MFR, Muñoz PR. AGHmatrix: R Package to Construct Relationship Matrices for Autotetraploid and Diploid Species: A Blueberry Example. *Plant Genome*. 2016;9: 0. doi:10.3835/plantgenome2016.01.0009
31. Ming R, Wang YW, Draye X, Moore PH, Irvine JE, Paterson AH. Molecular dissection of complex traits in autopolyploids: Mapping QTLs affecting sugar yield and related traits in sugarcane. *Theor Appl Genet*. 2002;105: 332–345. doi:10.1007/s00122-001-0861-5
32. da Silva J, Honeycutt RJ, Burnquist W, Al-Janabi SM, Sorrells ME, Tanksley SD, et al. *Saccharum spontaneum* L. “SES 208” genetic linkage map combining RFLP- and PCR-based markers. *Mol Breed*. 1995;1: 165–179. doi:10.1007/BF01249701
33. Aitken KS, Jackson PA, McIntyre CL. A combination of AFLP and SSR markers provides extensive map coverage and identification of homo (eo) logous linkage groups in a sugarcane cultivar. *Theoretical and Applied Genetics*, 2005; 5: 789-801, 10.1007/s00122-004-1813-7
34. Wu R, Ma CX, Painter I, Zeng ZB. Simultaneous Maximum Likelihood Estimation of Linkage and Linkage Phases in Outcrossing Species. *Theor Popul Biol*. 2002;61: 349–363. doi:10.1006/tpbi.2002.1577
35. Serang O, Mollinari M, Garcia AAF. Efficient exact maximum a posteriori computation for Bayesian SNP genotyping in polyploids. *PLoS One*. 2012;7: 1–13. doi:10.1371/journal.pone.0030906

3.9 SUPPLEMENTAL INFORMATION

S1 Fig. Relationship among countries and source of publication



Relationship among sources of publication and the first authors' affiliation countries.

S1 Table. Complete information for the 344 articles.

Authors	Title	Classification	Marker Class	Number of alleles	Dosage information	Source	Total Citation	Publication Year	DOI
Corderio, GM; Casu, R; McIntyre, CL; Mammers, JM; Henry, RJ Ming, R; Liu, SC; Lin, YR; da Silva, J; Wilson, W; Braga, D; van Deynze, A; Weisblat, TF; Wu, KK; Moore, PH; Burnquist, W; Sorrells, ME; Irvine, JE; Paterson, AH	Microsatellite markers from sugarcane (Saccharum spp.) ESTs cross transferable to erianthus and sorghum Detailed alignment of Saccharum and Sorghum chromosomes: Comparative organization of closely related diploid and polyploid genomes	Marker development Diversity	SSR RFLP	35 428	n n	PLANT SCIENCE GENETICS	257 156	2001 1998	10.1016/S0168-9452(01)00365-X
Grivet, L; DHont, A; Roques, D; Feldmann, P; Lanaud, C; Glaszmann, JC	RFLP mapping in cultivated sugarcane (Saccharum spp): Genome organization in a highly polyploid and aneuploid interspecific hybrid	Mapping	RFLP	408	n	GENETICS	131	1996	
Corderio, GM; Taylor, GO; Henry, RJ	Characterisation of microsatellite markers from sugarcane (Saccharum sp.), a highly polyploid species	Fingerprinting/ Characterization	SSR	100	n	PLANT SCIENCE	125	2000	10.1016/S0168-9452(00)00208-9
ALIANABI, SM; HONEYCUTT, RJ; MCCLELLAND, M; SOBRAL, BWS	A GENETIC-LINKAGE MAP OF SACCHARUM-SPONTANEUM/L SES 208	Mapping	RAPD	208	n	GENETICS	106	1993	
DASILVA, JAG; SORRELLS, ME; BURNOUST, WL; TANKSLEY, SD	RFLP LINKAGE MAP AND GENOME ANALYSIS OF SACCHARUM.SPONTANEUM	Mapping	RFLP	216	n	GENOME	100	1993	
Ming, R; Liu, SC; Moore, PH; Irvine, JE; Paterson, AH	QTL analysis in a complex autopolyploid: Genetic control of sugar content in sugarcane	Association	RFLP	1255	n	GENOME RESEARCH	90	2001	10.1101/gr.198801
DHONT, A; RAO, PS; FELDMANN, P; GRIVET, L; ISLAMFARIDI, N; TAYLOR, P; GLASZMANN, JC	IDENTIFICATION AND CHARACTERIZATION OF SUGARCANE INTERGENERIC HYBRIDS, SACCHARUM-OFFICINARUM X BRIANTHUS-ARUNDINACEUS, WITH MOLECULAR MARKERS AND DNA IN-SITU HYBRIDIZATION	Fingerprinting/ Characterization	RFLP/ Isozyme / STS	52	n	THEORETICAL AND APPLIED GENETICS	89	1995	
Dufour, P; Deu, M; Grivet, L; DHont, A; Paulet, F; Bouet, A; Lanaud, C; Glaszmann, JC; Hannon, P	Construction of a composite sorghum genome map and comparison with sugarcane, a related complex polyploid	Mapping	RFLP	199	n	THEORETICAL AND APPLIED GENETICS	87	1997	10.1007/s001220050430
Lima, ML A; Garcia, AAF; Oliveira, KM; Matsuoka, S; Arizono, H; de Souza, CL; de Souza, AP	Analysis of genetic similarity detected by AFLP and coefficient of parentage among genotypes of sugar cane (Saccharum spp.)	Diversity	AFLP	1121	n	THEORETICAL AND APPLIED GENETICS	86	2002	10.1007/s001220200003
Aitken, K; Jackson, P; McIntyre, C	A combination of AFLP and SSR markers provides extensive map coverage and identification of homo(eo)logous linkage groups in a sugarcane cultivar	Mapping	AFLP/ RAP/ SSR	1365	s+d	THEORETICAL AND APPLIED GENETICS	83	2005	10.1007/s00122-004-1813-7
Houran, JY; Offmann, B; D'Hont, A; Risterucci, AM; Roques, D; Glaszmann, JC; Grivet, L	Genetic dissection of a modern sugarcane cultivar (Saccharum spp.) 1. Genome mapping with AFLP markers	Mapping	AFLP	939	n	THEORETICAL AND APPLIED GENETICS	81	2001	10.1007/s001220000390
Rossi, M; Araujo, PG; Paulet, F; Garsmeur, O; Dias, VM; Chen, H; Van Sluys, MA; D'Hont, A	Genomic distribution and characterization of EST-derived resistance gene analogs (RGAs) in sugarcane	Fingerprinting/ Characterization	EST	88	n	MOLECULAR GENETICS AND GENOMICS	80	2003	10.1007/s00438-003-0849-8
Nair, NV; Nair, S; Sreenivasan, TV; Mohan, M	Analysis of genetic diversity and phylogeny in Saccharum and related genera using RAPD markers	Diversity	RAPD	195	n	GENETIC RESOURCES AND CROP EVOLUTION	76	1999	10.1023/A:1008696808645
Daugrois, JH; Grivet, L; Roques, D; Houran, JY; Lombard, H; Glaszmann, JC; DHont, A	A putative major gene for rust resistance linked with a RFLP marker in sugarcane cultivar R570	Association	RFLP/ Isozyme	505	n	THEORETICAL AND APPLIED GENETICS	76	1996	10.1007/BF00224049
Corderio, GM; Pan, YB; Henry, RJ	Sugarcane microsatellites for the assessment of genetic diversity in sugarcane germplasm	Diversity	SSR	187	n	PLANT SCIENCE	75	2003	10.1016/S0168-9452(03)00157-2
Guimaraes, CT; Sills, GR; Sobral, BWS	Comparative mapping of Andropogoneae: Saccharum L. (sugarcane) and its relation to sorghum and maize	Mapping	RFLP/ RAPD/ AFLP	646	n	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA	70	1997	10.1073/pnas.94.26.14261

S1 Table. (cont)

Authors	Title	Classification	Marker Class	Number of alleles	Dosage information	Source	Total Citation	Publication Year	DOI
DASILVA, J.; HONEYCUTT, R.; BURROUGHS, W.; ALJANABI, SM.; SORELLS, ME.; TANKSLEY, SD.; SORAL, BVS	SACCHARUM-SPONTANEUM L. SES-208 GENETIC-LINKAGE MAP COMBINING RFLP-BASED AND PCR-BASED MARKERS	Mapping	RAPD	527	n	MOLECULAR BREEDING	67	1995	
Tonkins, JP; Yu, Y; Miller-Smith, H; Frisch, DA; Woo, SS; Wing, RA	A bacterial artificial chromosome library for sugarcane RELATIONSHIPS AMONG ANCESTRAL SPECIES OF SUGARCANE REVEALED WITH RFLP USING SINGLE-COPY MAIZE NUCLEAR PROBES	Fingerprinting/ Characterization	RFLP		n	THEORETICAL AND APPLIED GENETICS	66	1999	10.1007/s001220051252
RAO, PS; FELDMANN, P; GLASZMANN, JC	Survey in the sugarcane expressed sequence tag database (SUCEST) for simple sequence repeats	Diversity	RFLP	1110	n	EUPHYTICA	65	1994	
Pinto, LR; Oliveira, KM; Ulian, EC; Garcia, AAF; de Souza, AP	Targeted mapping of a sugarcane rust resistance gene (Brul) using bulked segregant analysis and AFLP markers	Marker development	EST/ SSR	52	n	GENOME	60	2004	10.1139/G04-055
Asmighi, C; Roques, D; Ruffel, S; Kaye, C; Hoarau, JY; Telismart, H; Girard, JC; Raboin, LM; Ristnerci, AM; Grivet, L; D'Hont, A	Targeted mapping of a sugarcane rust resistance gene (Brul) using bulked segregant analysis and AFLP markers	Association	AFLP	2	n	THEORETICAL AND APPLIED GENETICS	56	2004	10.1007/s00122-003-1487-6
Alwalia, S; Suman, A; Arro, JA; Veremis, JC; Kimbeng, CA	Target region amplification polymorphism (TRAP) for assessing genetic diversity in sugarcane germplasm collections	Marker development	TRAP	529	n	CROP SCIENCE	55	2006	10.2133/cropsci.2005.0274
Hoarau, JY; Grivet, L; Offmann, B; Raboin, LM; Dordier, JP; Payet, J; Heilmann, M; D'Hont, A; Glaszmann, JC	Genetic dissection of a modern sugarcane cultivar (Saccharum spp.). II. Detection of QTLs for yield components	Association	AFLP	939	n	THEORETICAL AND APPLIED GENETICS	55	2002	10.1007/s00122-002-1047-5
TAYLOR, PMJ; GEISKES, JR; KO, HL; FRASER, TA; HENRY, RJ; BIRCH, RG	SENSITIVITY OF RANDOM AMPLIFIED POLYMORPHIC DNA ANALYSIS TO DETECT GENETIC CHANGE IN SUGARCANE DURING TISSUE-CULTURE	Fingerprinting/ Characterization	RAPD	112	n	THEORETICAL AND APPLIED GENETICS	54	1995	
Parida, SK; Kalita, SK; Kaul, S; Dalal, V; Hemrajabha, G; Sethi, A; Pandit, A; Singh, A; Gaikwad, K; Sharma, TR; Srivastava, PS; Singh, NK; Mohapatra, T	Informative genomic microsatellite markers for efficient genotyping applications in sugarcane	Marker development	SSR	1315	n	THEORETICAL AND APPLIED GENETICS	53	2009	10.1007/s00122-008-0902-4
Arenchba, AD; Carnona, ER; Conide, MT; Castiglione, S; O'Reilly, J; China, A; Oramas, P; Sala, F	Somaclonal variation in insect-resistant transgenic sugarcane (Saccharum hybrid) plants produced by cell electroporation	Fingerprinting/ Characterization	D/ AFLP/ Rf	51	n	TRANSGENIC RESEARCH	52	1999	10.1023/A:1008900230144
Iannoo, N; Grivet, L; Seguin, M; Paullet, F; Domingue, R; Rao, PS; Dookun, A; D'Hont, A; Glaszmann, JC	Molecular investigation of the genetic base of sugarcane cultivars	Diversity	RFLP	386	n	THEORETICAL AND APPLIED GENETICS	50	1999	10.1007/s001220051222
Ming, R; Wang, YW; Draye, X; Moore, PH; Irvine, JE; Paterson, AH	Molecular dissection of complex traits in autopolyploids: mapping QTLs affecting sugar yield and related traits in sugarcane	Association	RFLP	735	n	THEORETICAL AND APPLIED GENETICS	50	2002	10.1007/s00122-001-0081-5
Garcia, AAF; Kido, EA; Meza, AN; Souza, HMB; Pinto, LR; Pestina, MM; Leite, CS; da Silva, JAG; Ulian, EC; Figueira, A; Souza, AP	Development of an integrated genetic map of a sugarcane (Saccharum spp.) commercial cross, based on a maximum-likelihood approach for estimation of linkage and linkage phases	Mapping	LP/ SSR/ AF	1118	n	THEORETICAL AND APPLIED GENETICS	49	2006	10.1007/s00122-005-0129-6
D'HONT, A; LU, YH; DELLEON, DG; GRIVET, L; FELDMANN, P; LANAUD, C; GLASZMANN, JC	A MOLECULAR APPROACH TO UNRAVELING THE GENETICS OF SUGARCANE, A COMPLEX POLYPLIOD OF THE ANDROPOGONEAE TRIBE	Mapping	sszyme/ RFL	348	n	GENOME	48	1994	10.1139/g94-031
Papini-Terzi, FS; Rocha, FR; Vencio, RZN; Felix, JM; Branco, DS; Wajsbomovsky, AI; Del Bem, LEV; Lealhe, CG; Costa, MDL; Nishiyama, MY; Vicentini, R; Vincenz, MGA; Ulian, EC; Memossi, M; Souza, GM	Sugarcane genes associated with sucrose content	Association	sequence	117	n	BMC GENOMICS	48	2009	10.1186/1471-2164-10-120
Mudge, J; Andersen, WR; Kehrer, RL; Fairbanks, DJ	A RAPD genetic map of Saccharum officinarum	Mapping	RAPD	279	n	CROP SCIENCE	47	1996	
Cia, MC; Guimarães, ACR; Medici, LO; Chabregas, SM; Azevedo, RA	Antioxidant responses to water deficit by drought-tolerant and -sensitive sugarcane varieties	Fingerprinting/ Characterization	Isozyme	5	n	ANNALS OF APPLIED BIOLOGY	46	2012	10.1111/j.1744-7348.2012.00575.x

S1 Table. (cont)

Authors	Title	Classification	Marker Class	Number of alleles	Dosage information	Source	Total Citation	Publication Year	DOI
LU, YH; DHONT, A; PAULET, F; GRIVET, L; ARNAUD, M; GLAZSMANN, JC	MOLECULAR DIVERSITY AND GENOME STRUCTURE IN MODERN SUGARCANE VARIETIES	Diversity	RFLP	411	n	EUPHYTICA	45	1994	10.1007/BF00027520
Raboin, LM; Oliveira, KM; Lecunff, L; Telismart, H; Roques, D; Butterfield, M; Hoarau, Y; DHont, A	Genetic mapping in sugarcane, a high polyploid, using bi-parental progeny: identification of a gene controlling stalk colour and a new rust resistance gene	Association	SSR/ RFLP	1666	n	THEORETICAL AND APPLIED GENETICS	45	2006	10.1007/s00122-006-0240-3
Janno, N; Grivet, L; Dookun, A; DHont, A; Glazsmann, JC	Linkage disequilibrium among modern sugarcane cultivars	Diversity	RFLP	33	n	THEORETICAL AND APPLIED GENETICS	44	1999	10.1007/s001220051414
Pinto, LR; Oliveira, KM; Marconi, T; Garcia, AAF; Ulian, EC; de Souza, AP	Characterization of novel sugarcane expressed sequence tag microsatellites and their comparison with genomic SSRs	Mapping	SSR	519	n	PLANT BREEDING	44	2006	10.1111/j.1439-0523.2006.01227.x
GLAZSMANN, JC; FAUTRET, A; NOYER, JT; FELDMANN, P; LANAUD, C	BIOCHEMICAL GENETIC-MARKERS IN SUGARCANE	Fingerprinting/ Characterization	Isozyme	9	n	THEORETICAL AND APPLIED GENETICS	43	1989	10.1007/BF00290839
Piperidis, G; Christopher, M; Carroll, BJ; Berding, N; DHont, A	Molecular contribution to selection of intergeneric hybrids between sugarcane and the wild species <i>Eriandrus arundinaceus</i>	Fingerprinting/ Characterization	RFLP	2	n	GENOME	43	2000	10.1139/gen-43-6-1033
Besse, P; Taylor, G; Carroll, B; Berding, N; Burner, D; McIntyre, CL	Assessing genetic diversity in a sugarcane germplasm collection using an automated AFLP analysis	Diversity	RFLP	276	n	GENETICA	42	1998	10.1023/A:1003436403678
Selvi, A; Nair, NV; Balasundaram, N; Mohapatra, T	Evaluation of maize microsatellite markers for genetic diversity analysis and fingerprinting in sugarcane	Fingerprinting/ Characterization	SSR	95	n	GENOME	41	2003	10.1139/G03-018
Aitken, KS; Jackson, PA; McIntyre, CL	Quantitative trait loci identified for sugar related traits in a sugarcane (Saccharum spp.) cultivar x Saccharum officinarum population	Association	SSR	1000	n	THEORETICAL AND APPLIED GENETICS	41	2006	10.1007/s00122-006-0233-2
Grivet, L; Glazsmann, JC; Vincenz, M; da Silva, F; Arruda, P	ESTs as a source for sequence polymorphism discovery in sugarcane: example of the Adh genes	Marker development Fingerprinting/ Characterization	sequence SSR	88	n	THEORETICAL AND APPLIED GENETICS	40	2003	10.1007/s00122-002-1075-1
Alix, K; Baurens, FC; Paulet, F; Glazsmann, JC; DHont, A	Isolation and characterization of a satellite DNA family in the Saccharum complex	Diversity	RAPD/ SSR	351	n	EUPHYTICA	39	1998	10.1139/gen-41-6-854
Harvey, M; Botha, FC	Use of PCR-based methodologies for the determination of DNA diversity between Saccharum varieties	Diversity	RAPD	167	n	EUPHYTICA	39	1996	10.1007/BF00034614
Nair, NV; Selvi, A; Sreenivasan, TV; Pushpalatha, KN	Molecular diversity in Indian sugarcane cultivars as revealed by randomly amplified DNA polymorphisms	Diversity	RAPD	1209	n	THEORETICAL AND APPLIED GENETICS	39	2002	10.1023/A:1020234428681
Wei, XM; Jackson, PA; McIntyre, CL; Aitken, KS; Croft, B	Associations between DNA markers and resistance to diseases in sugarcane and effects of population substructure	Association	AFLP/ SSR	1209	n	THEORETICAL AND APPLIED GENETICS	39	2006	10.1007/s00122-006-0418-8
Calas, T; Figueira, A	Serial analysis of gene expression in sugarcane (Saccharum spp.) leaves revealed alternative C-4 metabolism and putative antisense transcripts	Fingerprinting/ Characterization	sequence		n	PLANT MOLECULAR BIOLOGY	38	2007	10.1007/s11103-006-9121-z
Oliveira, KM; Pinto, LR; Marconi, TG; Mollinari, M; Ulian, EC; Chabregas, SM; Falco, MC; Burnquist, W; Garcia, AAF; Souza, AP	Characterization of new polymorphic functional markers for sugarcane	Marker development	SSR/ EST	1456	n	GENOME	38	2009	10.1139/G08-105
Dufour, P; Grivet, L; DHont, A; Deu, M; Trouche, G; Glazsmann, JC; Hamon, P	Comparative genetic mapping between duplicated segments on maize chromosomes 3 and 8 and homologous regions in sorghum and sugarcane	Mapping	RFLP	49	n	THEORETICAL AND APPLIED GENETICS	38	1996	

S1 Table. (cont)

Authors	Title	Classification	Marker Class	Number of alleles	Dosage information	Source	Total Citation	Publication Year	DOI
Asmighi, C; Paulet, F; Kaye, C; Grivet, L; Deu, M; Glaszmann, JC; D'Hont, A	Application of synteny across Poaceae to determine the map location of a sugarcane rust resistance gene	Mapping	RFLP	1059	s	THEORETICAL AND APPLIED GENETICS	38	2000	10.1007/s001220051568
Thiebaut, F; Rojas, CA; Almeida, KL; Grativolt, C; Domiciano, GC; Lamb, CRC; Engler, JD; Hemerly, AS; Ferreira, PCG	Regulation of miR319 during cold stress in sugarcane	Association	sequence	1	n	PLANT CELL AND ENVIRONMENT	38	2012	10.1111/j.1365-3040.2011.02430.x
Jannoo, N; Grivet, L; David, J; D'Hont, A; Glaszmann, JC	Differential chromosome pairing affinities at meiosis in polyploid sugarcane revealed by molecular markers	Fingerprinting/ Characterization	RFLP	11	n	HEREDITY	36	2004	10.1038/sj.hdy.6800524
BUCCYNSKI, SR; THOM, M; CHOUREY, P; MARETZKI, A	TSISLE DISTRIBUTION AND CHARACTERIZATION OF SUCROSE SYNTHASE Isozymes IN SUGARCANE	Fingerprinting/ Characterization	Isozyme	2	n	JOURNAL OF PLANT PHYSIOLOGY	35	1993	
WILLIAMS, CA; HARBORNE, JB; SMITH, P	TAXONOMIC SIGNIFICANCE OF LEAF FLAVONOIDS IN SACCCHARUM AND RELATED GENERA	Marker development	Isozyme	9	n	PHYTOCHEMISTRY	34	1974	10.1016/0031-9422(74)80088-9
Oliveira, KM; Pinao, LR; Marconi, TG; Margarido, GRA; Pastina, MM; Teixeira, LHM; Figueira, AV; Ulian, EC; Garcia, AAF; Souza, AP	Functional integrated genetic linkage map based on EST-markers for a sugarcane (Saccharum spp.) commercial cross	Mapping	SSR/ RFLP	664	n	MOLECULAR BREEDING	34	2007	10.1007/s11032-007-9082-1
McIntyre, CL; Casu, RE; Drenth, J; Knight, D; Whan, VA; Croft, BJ; Jordan, DR; Manners, JM	Resistance gene analogues in sugarcane and sorghum and their association with quantitative trait loci for rust resistance	Association	EST	54	n	GENOME	34	2005	10.1139/G05-006
D'Hont, A; Paulet, F; Glaszmann, JC	Oligoclonal interspecific origin of 'North Indian' and 'Chinese' sugarcanes	Fingerprinting/ Characterization	RFLP	193	n	CHROMOSOME RESEARCH	33	2002	10.1023/A:1015204424287
CHOWDHURY, MKU; VASIL, JK	MOLECULAR ANALYSIS OF PLANTS REGENERATED FROM EMBRYOGENIC CULTURES OF HYBRID SUGARCANE CULTIVARS (SACCHARUM SPP)	Fingerprinting/ Characterization	RFLP		n	THEORETICAL AND APPLIED GENETICS	32	1993	
Besse, P; McIntyre, CL; Berding, N	Characterization of Erianthus sect Rhipidum and Saccharum germplasm (Andropogoneae-Saccharinae) using RFLP markers	Fingerprinting/ Characterization	RAPD/ SSR	351	n	EUPHYTICA	32	1997	10.1023/A:1002940701171
Cordero, GM; Elliot, F; McIntyre, CL; Casu, RE; Henry, RJ	Characterisation of single nucleotide polymorphisms in sugarcane ESTs	Fingerprinting/ Characterization	SNP	69	n	THEORETICAL AND APPLIED GENETICS	30	2006	10.1007/s00122-006-0300-8
Pan, YB; Burner, DM; Legendre, BL	An assessment of the phylogenetic relationship among sugarcane and related taxa based on the nucleotide sequence of 5S rRNA intergenic spacers	Diversity	sequence	1	n	GENETICA	30	2000	10.1023/A:1004191625603
Le Cunff, L; Garsmeur, O; Raboin, LM; Pauquet, J; Telesmat, H; Selvi, A; Grivet, L; Philippe, R; Begum, D; Deu, M; Costet, L; Winge, R; Glaszmann, JC; D'Hont, A	Diploid/polyploid syntenic shuttle mapping and haplotype-specific chromosome walking toward a rust resistance gene (Brr1) in highly polyploid sugarcane (2n similar to 12x similar to 115)	Mapping	AFLP	3	n	GENETICS	30	2008	10.1534/genetics.108.091355
Mudge, SK; Osabe, K; Casu, RE; Bonnet, GD; Manners, JM; Birch, RG	Efficient silencing of reporter transgenes coupled to known functional promoters in sugarcane, a highly polyploid crop species	Fingerprinting/ Characterization	sequence	8	n	PLANTA	29	2009	10.1007/s00425-008-0852-8
Ma, HM; Schultze, S; Lee, S; Yang, M; Mifkoy, E; Irvine, J; Moore, P; Paterson, A	An EST survey of the sugarcane transcriptome	Marker development	EST	3401	n	THEORETICAL AND APPLIED GENETICS	29	2004	10.1007/s00122-003-1510-y
Aiken, KS; Hermann, S; Kanno, K; Bonnet, GD; McIntyre, LC; Jackson, PA	Genetic control of yield related stalk traits in sugarcane	Association	AFLP/ SSR	2439	n	THEORETICAL AND APPLIED GENETICS	29	2008	10.1007/s00122-008-0856-6
Selvi, A; Nair, NV; Noye, JL; Singh, NK; Balasundaram, N; Bansal, KC; Koundal, KR; Mohapatra, T	AFLP analysis of the phenetic organization and genetic diversity in the sugarcane complex, Saccharum and Erianthus	Diversity	AFLP	1122	n	GENETIC RESOURCES AND CROP EVOLUTION	28	2006	10.1007/s10722-004-6376-6

S1 Table. (cont)

Authors	Title	Classification	Marker Class	Number of alleles	Dosage information	Source	Total Citation	Publication Year	DOI
Parida, SK; Pandit, A; Gaikwad, K; Sharma, TR; Srivastava, PS; Singh, NK; Mohapatra, T	Functionally relevant microsatellites in sugarcane unigenes	Marker development	SSR	726	n	BMC PLANT BIOLOGY	28	2010	10.1186/1471-2229-10-251
Cai, Q; Aiken, K; Deng, HH; Chen, XW; Fu, C; Jackson, PA; McIntyre, CL	Verification of the introgression of <i>Erianthus arundinaceus</i> germplasm into sugarcane using molecular markers	Fingerprinting/Characterization	SSR	2	n	PLANT BREEDING	27	2005	10.1111/j.1439-0523.2005.01099.x
Ming, R; Liu, SC; Bowers, JE; Moore, PH; Irvine, JE; Peterson, AH	Construction of a <i>Saccharum</i> consensus genetic map from two interspecific crosses	Mapping	REL P	982	n	CROP SCIENCE	27	2002	
Rossi, M; Araujo, PG; Van Sluys, MA	Survey of transposable elements in sugarcane expressed sequence tags (ESTs)	Fingerprinting/Characterization	sequence	276	n	GENETICS AND MOLECULAR BIOLOGY	26	2001	
Zucchi, MI; Arizono, H; Morais, VA; Fungaro, MHP; Vieira, MLC	Genetic instability of sugarcane plants derived from meiosis cultures	Fingerprinting/Characterization	RAPD	98	n	GENETICS AND MOLECULAR BIOLOGY	25	2002	10.1590/S1415-47572002000100017
RefRAY, N; Jackson, PA; Aiken, KS; Hoarau, JY; D'Hont, A; Besse, P; McIntyre, CL	Characterisation of genome regions incorporated from an important wild relative into Australian sugarcane	Fingerprinting/Characterization	AFLP/SSR	400	n	MOLECULAR BREEDING	25	2005	10.1007/s11032-004-7981-y
Crete, S; Landell, MGA; Mazzafra, P; Azevedo, RA	Water stress reveals differential antioxidant responses of tolerant and non-tolerant sugarcane genotypes	Fingerprinting/Characterization	Isozyme	3	n	PLANT PHYSIOLOGY AND BIOCHEMISTRY	25	2014	10.1016/j.plaphy.2013.11.016
Pastina, MM; Maloescu, M; Garafiti, R; Mollinari, M; Margarido, GRA; Oliveira, KM; Pinto, LR; Souza, AP; van Eeuwijk, FA; Garcia, AAF	A mixed model QTL analysis for sugarcane multiple-harvest-location trial data	Association	REL P/SSR	741	n	THEORETICAL AND APPLIED GENETICS	25	2012	10.1007/s00122-011-1748-8
Selvi, A; Nair, NV; Noyer, JL; Singh, NK; Balasundaram, N; Bansal, KC; Koundal, KR; Mohapatra, T	Genomic constitution and genetic relationship among the tropical and subtropical Indian sugarcane cultivars revealed by AFLP	Diversity	AFLP	621	n	CROP SCIENCE	24	2005	10.2135/cropsci2004.0528
Raboin, LM; Pauquet, J; Butterfield, M; D'Hont, A; Glaszmann, JC	Analysis of genome-wide linkage disequilibrium in the highly polyploid sugarcane	Diversity	AFLP	1537	n	THEORETICAL AND APPLIED GENETICS	24	2008	10.1007/s00122-007-0703-1
Cai, Q; Aiken, KS; Fan, YH; Piperidis, G; Jackson, P; McIntyre, CL	A preliminary assessment of the genetic relationship between <i>Erianthus rockii</i> and the "Saccharum complex" using microsatellite (SSR) and AFLP markers	Diversity	SSR/ AFLP	414	n	PLANT SCIENCE	23	2005	10.1016/j.plantsci.2005.07.002
Hellet-Liszynska, K; Liszynski, G; Huttner, E; Evers, M; Carlig, J; Cai, V; Aiken, K; Jackson, P; Piperidis, G; Cox, M; Ghmour, R; D'Hont, A; Butterfield, M; Glaszmann, JC; Kilian, A	Diversity Arrays Technology effectively reveals DNA polymorphism in a large and complex genome of sugarcane	Marker development	DART	667	n	MOLECULAR BREEDING	22	2011	10.1007/s11032-010-9460-y
ALIANABI, SM; HONEYCUTT, RJ; SOBRAL, BWS	CHROMOSOME ASSORTMENT IN SACCHARUM	Mapping	RAPD	1458	n	THEORETICAL AND APPLIED GENETICS	22	1994	
Andru, S; Pan, YB; Thongthawee, S; Burner, DM; Kimbeng, CA	Genetic analysis of the sugarcane (<i>Saccharum</i> spp.) cultivar T1CP 85-384, 1. Linkage mapping using AFLP, SSR, and TRAP markers	Mapping	FLP/ SSR/ TRAP	955	n	THEORETICAL AND APPLIED GENETICS	22	2011	10.1007/s00122-011-1568-x
Takahashi, S; Furukawa, T; Asano, T; Terajima, Y; Shimada, H; Sugimoto, A; Kadowaki, K	Very close relationship of the chloroplast genomes among <i>Saccharum</i> species	Diversity	sequence	18	n	THEORETICAL AND APPLIED GENETICS	21	2005	10.1007/s00122-005-1990-z
Alwaia, S; Kimbeng, CA; Veremis, JC; Gravos, KA	Linkage mapping and genome analysis in a <i>Saccharum</i> interspecific cross using AFLP, SRAP and TRAP markers	Mapping	LP/ TRAP/ SR	650	n	EUPHYTICA	21	2008	10.1007/s10681-007-9634-9
Costet, L; Le Cunff, L; Royacert, S; Raboin, LM; Hervouet, C; Toubi, L; Teismann, H; Garnier, O; Roussel, Y; Pauquet, J; Nibouche, S; Glaszmann, JC; Hoarau, JY; D'Hont, A	Haplotype structure around Br1 reveals a narrow genetic basis for brown rust resistance in modern sugarcane cultivars	Mapping	FLP/ SSR/ CAI	38	n	THEORETICAL AND APPLIED GENETICS	21	2012	10.1007/s00122-012-1875-x

S1 Table. (cont)

Authors	Title	Classification	Marker Class	Number of alleles	Dosage information	Source	Total Citation	Publication Year	DOI
Edne, S.J, Glynn, NG, Comstock, JC	Genetic segregation of microsatellite markers in <i>Saccharum officinarum</i> and <i>S. spontaneum</i>	Fingerprinting/ Characterization	SSR	193	n	HEREDITY	20	2006	10.1038/sj.hdy.6800888
Pan, YB; Bumer, DM; Legendre, BL; Grisham, MP; White, WH	An assessment of the genetic diversity within a collection of <i>Saccharum spontaneum</i> L. with RAPD-PCR	Diversity	RAPD	157	n	GENETIC RESOURCES AND CROP EVOLUTION	20	2004	
Swarna, M; Sivaranj, K.; Sharma, RK; Singh, NK; Mohapatra, T	Single-Strand Conformational Polymorphism of EST-SSRs: A Potential Tool for Diversity Analysis and Varietal Identification in Sugarcane	Marker development	SSR	118	n	PLANT MOLECULAR BIOLOGY REPORTER	20	2011	10.1007/s11105-010-0254-5
SILL, S, GR, BRIDGES, W, ALIANABI, SM, SOBRAL, BWS	GENETIC-ANALYSIS OF AGRONOMIC TRAITS IN A CROSS BETWEEN SUGARCANE (<i>SACCHARUM-OFFICINARUM</i> L) AND ITS PRESUMED PROGENITOR (<i>S-ROBUSTUM</i> BRANDES AND JESW EX GRASSI)	Association	AFLP	83	n	MOLECULAR BREEDING	20	1995	10.1007/BF01248413
Grivet, L.; Glaszmann, JC; Arruda, P	Sequence polymorphism from EST data in sugarcane: a fine analysis of 6-phosphogluconate dehydrogenase genes	Fingerprinting/ Characterization	SNP/ sequence/ EST	40	n	GENETICS AND MOLECULAR BIOLOGY	19	2001	10.1590/S1415-47572001000100022
Pan, YB; Cordeiro, GM; Richard, EP; Henry, RJ	Molecular genotyping of sugarcane clones with microsatellite DNA markers	Fingerprinting/ Characterization	SSR	52	n	MAVADICA	19	2003	
Saccaro, NL; Kajihara, D; Massa, R; de Felix, JM; Drummond, RD; Falco, MC; Chabregas, SM; Ulian, EC; Menossi, M; Van Sluys, MA	Transcriptionally active transposable elements in recent hybrid sugarcane	Fingerprinting/ Characterization	sequence		n	PLANT JOURNAL	19	2005	10.1111/j.1365-313X.2005.02579.x
Garcia, AAF; Mollinari, M; Marconi, TG; Serang, OR; Silva, RR; Vieira, MLC; Vicentini, R; Costa, EA; Mancini, MC; Garcia, MOS; Pastina, MM; Gazaffi, R; Martins, ERF; Dahmer, N; Storca, DA; Silva, CBC; Bundoek, P; Henry, RJ; Souza, GM; van Sluys, MA; Landell, MGA; Carneiro, MS; Vincenz, MAG; Pinto, LR; Vencovsky, R; Souza, AP	SNP genotyping allows an in-depth characterisation of the genome of sugarcane and other complex autopolyploids	Fingerprinting/ Characterization	SNP	249	n	SCIENTIFIC REPORTS	19	2013	10.1038/srep03399
Aitken, KS; Jackson, PA; McInyre, CL	Construction of a genetic linkage map for <i>Saccharum officinarum</i> incorporating both simplex and duplex markers to increase genome coverage	Mapping	AFLP/ SSR	595	n	GENOME	19	2007	10.1139/G07-056
Jordan, DR; Casu, RE; Besse, P; Carroll, BC; Bedding, N; McInyre, CL	Markers associated with stalk number and suckering in sugarcane collocate with tillering and rhizomatousness QTLs in sorghum	Association	RFLP	258	n	GENOME	19	2004	10.1139/G04-040
OROPEZA, M; GUEVARA, P; DEGARCIA, E; RAMIREZ, JL	IDENTIFICATION OF SOMACLONAL VARIANTS OF SUGARCANE (<i>SACCHARUM SPP</i>) RESISTANT TO SUGARCANE MOSAIC-VIRUS VIA RAPD MARKERS	Fingerprinting/ Characterization	RAPD		n	PLANT MOLECULAR BIOLOGY REPORTER	18	1995	10.1007/BF02668790
Gilbert, RA.; Glynn, NC.; Comstock, JC; Davis, MJ	Agonomic performance and genetic characterization of sugarcane transformed for resistance to sugarcane yellow leaf virus	Fingerprinting/ Characterization	SSR	66	n	FIELD CROPS RESEARCH	18	2009	10.1016/j.fcr.2008.10.009
Singh, RK; Mishra, SK; Singh, SP; Mishra, N; Sharma, ML	Evaluation of microsatellite markers for genetic diversity analysis among sugarcane species and commercial hybrids	Diversity	SSR	263	n	AUSTRALIAN JOURNAL OF CROP SCIENCE	18	2010	
Alix, K.; Paulet, F.; Glaszmann, JC; Dhont, A	Inter-Alu-like species-specific sequences in the <i>Saccharum</i> complex	Marker development	sequence	18	n	THEORETICAL AND APPLIED GENETICS	18	1999	10.1007/s001220051403
de Cesare, M; Hodgkinson, TR; Barth, S Lao, M; Arencibia, AD; Garmoa, ER; Acevedo, R; Rodriguez, E; Leon, O; Santana, I	Differential expression analysis by cDNA-AFLP of <i>Saccharum spp.</i> after inoculation with the host pathogen <i>Sporisorium scitamineum</i>	Marker development	SSR, SNP		n	MOLECULAR BREEDING	18	2010	10.1007/s11032-010-9451-z
Pinto, LR; Garcia, AAF; Pastina, MM; Teixeira, LHM; Bressiani, JA; Ulian, EC; Biotto, MAP; Souza, AP	A analysis of genomic and functional RFLP derived markers associated with sucrose content, fiber and yield QTLs in a sugarcane (<i>Saccharum spp.</i>) commercial cross	Association	AFLP	64	n	PLANT CELL REPORTS	18	2008	10.1007/s00299-008-0524-y
		Association	RFLP	222	n	EUPHYTICA	18	2010	10.1007/s10681-009-9988-2

S1 Table. (cont)

Authors	Title	Classification	Marker Class	Number of alleles	Dosage information	Source	Total Citation	Publication Year	DOI
Cardoso-Silva, CB; Costa, EA; Mancini, MC; Bastalote, TWA; Canesin, LEC; Pinto, LR; Carneiro, MS; Garcia, AAF; de Souza, AP; Vicentini, R	De Novo Assembly and Transcriptome Analysis of Contrasting Sugarcane Varieties	Fingerprinting/ Characterization	sequence	14	n	PLOS ONE	17	2014	10.1371/journal.pone.0088462
Gouy, M; Roussele, Y; Bastianelli, D; Leconte, P; Bonnal, L; Roques, D; Efile, JC; Rocher, S; Daugrois, J; Toubi, L; Nabeneza, S; Hervouet, C; Teissant, H; Denis, M; Thong-Chane, A; Glaszmann, JC; Hourau, JY; Ribouche, S; Costet, L	Experimental assessment of the accuracy of genomic selection in sugarcane	Marker Assisted Selection	DART	1499	n	THEORETICAL AND APPLIED GENETICS	17	2013	10.1007/s00122-013-2156-z
Borras-Hidalgo, O; Thonma, BPHJ; Camarona, E; Borroto, CP; Pujol, M; Arencibia, A; Lopez, J	Identification of sugarcane genes induced in disease-resistant somaclones upon inoculation with <i>Ustilago scitaminea</i> or <i>Biopolaris sacchari</i>	Association	AFLP	62	n	PLANT PHYSIOLOGY AND BIOCHEMISTRY	17	2005	10.1016/j.plaphy.2005.07.011
Melnyte, CL; Jackson, M; Cordeiro, GM; Amouyal, O; Hermann, S; Aitken, KS; Elliott, F; Henry, RJ; Casu, RE; Bonnet, GD	The identification and characterisation of alleles of sucrose phosphatase synthase gene family III in sugarcane	Association	SNP	10	n	MOLECULAR BREEDING	17	2006	10.1007/s11032-006-9012-7
Aitken, KS; Li, JC; Jackson, P; Piperidis, G; Melnyte, CL	AFLP analysis of genetic diversity within <i>Saccharum officinarum</i> and comparison with sugarcane cultivars	Diversity	AFLP	614	n	AUSTRALIAN JOURNAL OF AGRICULTURAL RESEARCH	16	2006	10.1071/AR05391
Mary, S; Nair, NV; Chaturvedi, PK; Selvi, A	Analysis of genetic diversity among <i>Saccharum spontaneum</i> L. from four geographical regions of India, using molecular markers	Diversity	ISSR, RAPD	412	n	GENETIC RESOURCES AND CROP EVOLUTION	16	2006	10.1007/s10722-005-2433-z
Zambrano, AY; Deney, JR; Fuchs, M; Gonzalez, V; Rea, R; De Sousa, O; Gutierrez, Z	Selection of sugarcane plants resistant to SCMV	Marker Assisted Selection	RAPD	1	n	PLANT SCIENCE	16	2003	10.1016/j.plphys.2006.03.013
Fusaro, A; Mangon, A; Jaqueira, RM; Roeha, CAB; Coutinho, TC; Margis, R; Sacchetto-Martins, G	Classification, expression pattern and comparative analysis of sugarcane expressed sequence's tags (ESTs) encoding glycine-rich proteins (GRPs)	Association	EST/ sequence	150	n	GENETICS AND MOLECULAR BIOLOGY	16	2001	10.1590/S1415-47572001000100035
Borges, JC; Cagliari, TC; Ramos, CHI	Expression and variability of molecular chaperones in the sugarcane exprosome	Fingerprinting/ Characterization	sequence		n	JOURNAL OF PLANT PHYSIOLOGY	15	2007	
Khan, IA; Dabot, MU; Seema, N; Yasmin, S; Bibi, S; Raza, S; Khatt, A	GENETIC VARIABILITY IN SUGARCANE PLANTLETS DEVELOPED THROUGH IN VITRO MUTAGENESIS	Fingerprinting/ Characterization	RAPD	17	n	PAKISTAN JOURNAL OF BOTANY	15	2009	
da Silva, JAG	Preliminary analysis of microsatellite markers derived from sugarcane expressed sequence tags (ESTs)	Marker development	SSR	20	n	GENETICS AND MOLECULAR BIOLOGY	15	2001	
Singh, RK; Jena, SN; Khan, S; Yadav, S; Banarjee, N; Raghuvaran, S; Bhardwaj, V; Datamajunder, SK; Kapur, R; Solomon, S; Swarna, M; Srivastava, S; Tyagi, AK	Development, cross-species/genera transferability of novel EST-SSR markers and their utility in revealing population structure and genetic diversity in sugarcane	Marker development	SSR	954	n	GENE	15	2013	10.1016/j.gene.2013.03.125
Wei, XM; Jackson, PA; Hermann, S; Kilian, A; Heller-Uzyska, K; Deomano, E	Simultaneously accounting for population structure, genotype by environment interaction, and spatial variation in marker-trait associations in sugarcane	Association	DART	1531	n	GENOME	15	2010	10.1139/G10-050
Coto, O; Cornile, MT; Calvo, D; Canales, E; Dhont, A; de Prada, F	Genetic diversity among wild sugarcane germplasm from Laos revealed with markers	Diversity	RFLP	168	n	EUPHYTICA	14	2002	10.1023/A:101447902930
Besse, P; Hourau, JY; Courtois, B; Aitken, KS; Melnyte, CL	Comparative genetics in sugarcane enables structured map enhancement and validation of marker-trait associations	Mapping	SSR, AFLP	1000	n	MOLECULAR BREEDING	14	2008	10.1007/s11032-007-9124-8
da Silva, JA; Bressiani, JA	Sucrose synthase molecular marker associated with sugar content in elite sugarcane progeny	Association	RFLP	29	n	GENETICS AND MOLECULAR BIOLOGY	14	2005	
Alwala, S; Kimberg, CA; Veremis, JC; Gravois, KA	Identification of molecular markers associated with sugar-related traits in a <i>Saccharum</i> interspecific cross	Association	SRAP/ TRAP	267	n	EUPHYTICA	14	2009	10.1007/s10681-008-9869-0

SI Table. (cont)

Authors	Title	Classification	Marker Class	Number of alleles	Dosage information	Source	Total Citation	Publication Year	DOI
Aitken, K.I., Li, J., Wang, L., Qmg, C., Fan, Y.H., Jackson, P	Characterization of intergeneric hybrids of <i>Erianthus rockii</i> and <i>Saccharum</i> using molecular markers	Fingerprinting/ Characterization	AFIP	391	n	GENETIC RESOURCES AND CROP EVOLUTION	13	2007	10.1007/s10722-006-9124-2
Glynn, NC; Comstock, JC; Sood, SG; Dang, PM; Chaparro, JX	Isolation of nucleotide binding site-leucine rich repeat and kinase resistance gene analogues from sugarcane (<i>Saccharum</i> spp.)	Fingerprinting/ Characterization	sequence		n	PEST MANAGEMENT SCIENCE	13	2008	10.1002/ps.1469
Watt, MP; Banasiak, M.; Reddy, D; Albertse, EH; Smyrnan, SJ	In vitro minimal growth storage of <i>Saccharum</i> spp. hybrid (genotype 88H0019) at two stages of direct somatic embryogenic regeneration	Fingerprinting/ Characterization	AFIP	580	n	PLANT CELL TISSUE AND ORGAN CULTURE	13	2009	10.1007/s11240-008-9483-9
Liu, PW; Que, YX; Pan, YB	Highly Polymorphic Microsatellite DNA Markers for Sugarcane Germplasm Evaluation and Variety Identity Testing	Fingerprinting/ Characterization	SSR	199	n	SUGAR TECH	13	2011	10.1007/s12355-011-0077-1
Burner, DM; Pan, YB; Webster, RD	Genetic diversity of North American and Old World <i>Saccharum</i> assessed by RAPD analysis	Diversity	RAPD	283	n	GENETIC RESOURCES AND CROP EVOLUTION	13	1997	10.1023/A:1008631731506
Kawar, PQ; Devarumath, RM; Nerkar, Y	Use of RAPD markers for assessment of genetic diversity in sugarcane cultivars	Diversity	RAPD	134	n	INDIAN JOURNAL OF BIOTECHNOLOGY	13	2009	
Carmona, E; Vargas, D; Borroto, CI; Lopez, J; Fernandez, AI; Arencibia, A; Borrás-Hidalgo, O	cDNA-AFLP analysis of differential gene expression during the interaction between sugarcane and <i>Puccinia melanocephala</i>	Association	sequence	34	n	PLANT BREEDING	13	2004	10.1111/j.1439-0523.2004.00935.x
Aljanabi, SM; Parmessur, Y; Kross, H; Dhayan, S; Samtally, S; Ramdoyal, K; Autey, LC; Dookun-Samtally, A	Identification of a major quantitative trait locus (QTL) for yellow spot (<i>Mycovellosiella koepkei</i>) disease resistance in sugarcane	Association	SSR	557	n	MOLECULAR BREEDING	13	2007	10.1007/s11032-006-9008-3
Srivastava, S; Pathak, AD; Gupta, PS; Shrivastava, AK; Shrivastava, AK	Hydrogen peroxide-scavenging enzymes impart tolerance to high temperature induced oxidative stress in sugarcane	Association	Isozyme	10	n	JOURNAL OF ENVIRONMENTAL BIOLOGY	12	2012	
Besnard, G; Pincon, G; D'Hont, A; Hoarau, JY; Cadet, F; Offmann, B	Characterisation of the phosphoenolpyruvate carboxylase gene family in sugarcane (<i>Saccharum</i> spp.)	Fingerprinting/ Characterization	sequence	22	n	THEORETICAL AND APPLIED GENETICS	12	2003	10.1007/s00122-003-1268-2
Khan, IA; Bibi, S; Yasmin, S; Khatri, A; Seema, N; Aigjan, S	GENETIC VARIABILITY IN MUTATED POPULATION OF SUGARCANE CLONE NIA-98 THROUGH MOLECULAR MARKERS (RAPD AND TRAP)	Fingerprinting/ Characterization	RAPD	65	n	PAKISTAN JOURNAL OF BOTANY	12	2010	
Glynn, NC; Laborde, C; Davidson, RW; Irey, MS; Glaz, B; D'Hont, A; Comstock, JC	Utilization of a major brown rust resistance gene in sugarcane breeding	Marker Assisted Selection	sequence	1	n	MOLECULAR BREEDING	12	2013	10.1007/s11032-012-9792-x
Singh, RK; Singh, RB; Singh, SP; Sharma, ML	Identification of sugarcane microsatellites associated to sugar content in sugarcane and transferability to other cereal genomes	Association	SSR	388	n	EUPHYTICA	12	2011	10.1007/s10681-011-0484-0
NAGAI, C; AHLLOOMALLA, BS; TEW, TL	SOMATOCLONAL VARIANTS FROM AN INTERGENERIC HYBRID - SACCCHARUM SPP HYBRID X ERLANTHUS-ARUNDINACEUM	Fingerprinting/ Characterization	Isozyme		n	EUPHYTICA	11	1991	10.1007/BF00023271
McIntyre, CL; Jackson, PA	Low level of selfing found in a sample of crosses in Australian sugarcane breeding programs	Fingerprinting/ Characterization	RAPD	8	n	EUPHYTICA	11	2001	10.1023/A:1026527617892
Schlogl, PS; Nogueira, FTS; Drummond, R; Felix, JM; De Rosa, Y; Vicentini, R; Leite, A; Ulian, EC; Memosini, M	Identification of new ABA- and MEA-activated sugarcane bZIP genes by data mining in the SUCEST database	Fingerprinting/ Characterization	EST/ sequence	96	n	PLANT CELL REPORTS	11	2008	10.1007/s00299-007-0468-7
Zhang, HY; Li, FS; He, IL; Zhong, HQ; Yang, QH; He, SC	Identification of sugarcane interspecies hybrids with RAPDs	Fingerprinting/ Characterization	RAPD	280	n	AFRICAN JOURNAL OF BIOTECHNOLOGY	11	2008	

S1 Table. (cont)

Authors	Title	Classification	Marker Class	Number of alleles	Dosage information	Source	Total Citation	Publication Year	DOI
Zhang, JS; Arro, J; Chen, YQ; Ming, R	Haplotype analysis of sucrose synthase gene family in three <i>Saccharum</i> species	Fingerprinting/ Characterization	sequence/ SNP		n	BMC GENOMICS	11	2013	10.1186/1471-2164-14-314
Perera, MF; Arias, ME; Costilla, D; Luque, AC; Garcia, MB; Romero, CD; Racedo, J; Ostipone, S; Filippone, MP; Cuenya, ML; Castagnaro, AP	Genetic diversity assessment and genotype identification in sugarcane based on DNA markers and morphological traits	Diversity	SSR	294	n	EUPHYTICA	11	2012	10.1007/s10681-012-0661-9
Pallares, AC; Rodrigues-Morais, TB; Van Sluys, MA; Domingues, DS; Maccheroni, W; Jordao, H; Souza, AP; Marconi, TG; Mollinari, M; Gazzaffi, R; Garcia, AAF; Vieira, MLC	A novel linkage map of sugarcane with evidence for clustering of retrotransposon-based markers	Mapping	SSR	730	n	BMC GENETICS	11	2012	10.1186/1471-2156-13-51
Drummond, RD; Guimarães, CT; Felix, J; Ninamango-Cardenas, FE; Carneiro, NP; Parva, E; Menossi, M	Prospecting sugarcane genes involved in aluminum tolerance	Association	3ST/ sequence	55	n	GENETICS AND MOLECULAR BIOLOGY	11	2001	10.1590/S1415-47572001000100029
McIntyre, CL; Whan, VA; Croft, B; Magarey, R; Smith, GR	Identification and validation of molecular markers associated with <i>paedynetra</i> root rot and brown rust resistance in sugarcane using map- and association-based approaches	Association	R / AFLP/ RF	1000	n	MOLECULAR BREEDING	11	2005	10.1007/s11032-005-7492-5
Schafer, WE; Rohwer, JM; Botha, FC	Partial purification and characterisation of sucrose synthase in sugarcane	Fingerprinting/ Characterization	sequence	3	n	JOURNAL OF PLANT PHYSIOLOGY	10	2005	10.1016/j.jpnp.2004.04.010
Brown, JS; Schnell, RJ; Power, EJ; Douglas, SL; Kuhn, DN	Analysis of clonal germplasm from five <i>Saccharum</i> species: <i>S. barberi</i> , <i>S. robustum</i> , <i>S. officinarum</i> , <i>S. sinense</i> and <i>S. spontaneum</i> . A study of inter- and intra species relationships using microsatellite markers	Diversity	SSR	468	n	GENETIC RESOURCES AND CROP EVOLUTION	10	2007	10.1007/s10722-006-0035-z
Guimaraes, CT; Honeycutt, RJ; Sills, GR; Sobral, BWS	Genetic maps of <i>Saccharum officinarum</i> L. and <i>Saccharum robustum</i> Brandes & Lew. Ex Grassl	Mapping	PD/ RFLP/ AI	341	n	GENETICS AND MOLECULAR BIOLOGY	10	1999	10.1590/S1415-47571999000100024
Aitken, KS; McNeil, MD; Hermann, S; Bundock, PC; Kilian, A; Heller-Ursynska, K; Henry, RJ; Li, JC	A comprehensive genetic map of sugarcane that provides enhanced map coverage and integrates high-throughput Diversity Array Technology (DArT) markers	Mapping	/ AFLP/ SSR	2267	n	BMC GENOMICS	10	2014	10.1186/1471-2164-15-152
Butterfield, MK; Rutherford, RS; Carson, DL; Hackett, BI	Application of gene discovery to varietal improvement in sugarcane	Association	RFLP	275	n	SOUTH AFRICAN JOURNAL OF BOTANY	10	2004	
Vargas, L; Brigida, ABS; Mota, JP; de Carvalho, TG; Rojals, CA; Vanechoute, D; Van Bel, M; Faminelli, L; Ferreira, PCG; Vandepoel, K; Hemery, AS	Drought Tolerance Conferred to Sugarcane by Association with <i>Glucanase</i> <i>diastrophicus</i> : A Transcriptomic View of Hormone Pathways	Association	sequence		n	PLOS ONE	10	2014	10.1371/journal.pone.0114744
GSD, de Brito, JZ; Donato, VMTS; da Silva, MV	Microsatellite characterization of the sugarcane cultivars obtained by issr markers	Fingerprinting/ Characterization	ISSR	53	n	CIENCIA E AGROTECNOLOGIA	9	2009	
Tew, TL; Pan, YB	Microsatellite (Simple Sequence Repeat) Marker-based Paternity Analysis of a Seven-Parent Sugarcane Poly-cross	Fingerprinting/ Characterization	SSR	57	n	CROP SCIENCE	9	2010	10.2135/cropsci.2009.10.0579
Melotto-Passarin, DM; Tambarussi, EV; Dressano, K; De Martin, VF; Carer, H	Characterization of chloroplast DNA microsatellites from <i>Saccharum</i> spp and related species	Fingerprinting/ Characterization	SSR	204	n	GENETICS AND MOLECULAR RESEARCH	9	2011	10.4238/vol10-3gm1019
Kim, C; Lee, TH; Compton, RO; Robertson, JS; Pierce, GJ; Paterson, AH	A genome-wide BAC end-sequence survey of sugarcane elucidates genome composition, and identifies BACs covering much of the euchromatin	Fingerprinting/ Characterization	sequence		n	PLANT MOLECULAR BIOLOGY	9	2013	10.1007/s11103-012-9987-x
Alvi, AK; Iqbal, J; Shah, AH; Pan, YB	DNA BASED GENETIC VARIATION FOR RED ROT RESISTANCE IN SUGARCANE	Diversity	RAPD	156	n	PAKISTAN JOURNAL OF BOTANY	9	2008	
Duarte, LSC; Silva, PP; Santos, JM; Barbosa, GVS; Ramalho-Neio, CE; Soares, L; Andrade, JCF; Almeida, C	Genetic similarity among genotypes of sugarcane estimated by SSR and coefficient of parentage	Diversity	SSR	57	n	SUGAR TECH	9	2010	

SI Table. (cont)

Authors	Title	Classification	Marker Class	Number of alleles	Dosage information	Source	Total Citation	Publication Year	DOI
Mumtaz, AS; Dure-Nayab; Iqbal, MJ; Shinnawi, ZK	PROBING GENETIC DIVERSITY TO CHARACTERIZE RED ROT RESISTANCE IN SUGARCANE	Diversity	RAPD	136	n	PAKISTAN JOURNAL OF BOTANY	9	2011	
Sindhu, R; Govindaraj, P; Balamurugan, A; Appunni, C	Genetic diversity in sugarcane hybrids (Saccharum spp complex) grown in tropical India based on STMS markers	Diversity	STMS	100	n	JOURNAL OF PLANT BIOCHEMISTRY AND BIOTECHNOLOGY	9	2011	10.1007/s13562-011-0036-7
dos Santos, JM; Filho, LSC; Soriano, MI; da Silva, PP; Nascimento, VX; Barbosa, GVD; Todaro, AR; Neto, CER; Almeida, C	Genetic diversity of the main progenitors of sugarcane from the RIDEISA germplasm bank using SSR markers	Diversity	SSR	124	n	INDUSTRIAL CROPS AND PRODUCTS	9	2012	10.1016/j.indcrop.2012.03.005
James, BT; Chen, CX; Rudolph, A; Swaminathan, K; Murray, JE; Na, JK; Spence, AK; Smith, B; Hudson, ME; Moose, SP; Ming, R	Development of microsatellite markers in autopolyploid sugarcane and comparative analysis of conserved microsatellites in sorghum and sugarcane	Marker development	SSR	1203	n	MOLECULAR BREEDING	9	2012	10.1007/s11032-011-9651-1
Que, YX; Lin, JW; Song, XX; Xu, LP; Chen, RK	Differential Gene Expression in Sugarcane in Response to Challenge by Fungal Pathogen Ustilago scitaminea Revealed by cDNA-AFLP	Association	AFLP	136	n	JOURNAL OF BIOMEDICINE AND BIOTECHNOLOGY	9	2011	10.1155/2011/160934
Wu, QB; Xu, LP; Guo, JL; Su, YC; Que, YX	Transcriptome Profile Analysis of Sugarcane Responses to Sportsorium scitaminea Infection Using Solexa Sequencing Technology	Association	sequence	2015	n	BIOMED RESEARCH INTERNATIONAL	9	2013	10.1155/2013/298920
Su, YC; Xu, LP; Fu, ZW; Yang, YT; Guo, JL; Wang, SS; Que, YX	ScCh1, Encoding an Acidic Class III Chitinase of Sugarcane, Confers Positive Responses to Biotic and Abiotic Stresses in Sugarcane	Association	sequence		n	INTERNATIONAL JOURNAL OF MOLECULAR SCIENCES	9	2014	10.3390/ijms15022738
Huckett, BJ; Botha, FC	Stability and potential use of RAPD markers in a sugarcane genealogy	Fingerprinting/Characterization	RAPD	10	n	EUPHYTICA	8	1995	10.1007/BF00022017
Moyle, RL; Birch, RG	Diversity of sequences and expression patterns among alleles of a sugarcane loading stem gene	Fingerprinting/Characterization	sequence	7	n	THEORETICAL AND APPLIED GENETICS	8	2013	10.1007/s00122-013-2091-z
Scherck, S; Crepeau, MW; Wu, KK; Moore, PH; Yu, Q; Ming, R	Genetic diversity and relationships in native Hawaiian Saccharum officinarum sugarcane	Diversity	RFLP	228	n	JOURNAL OF HEREDITY	8	2004	10.1093/jhered/est052
Creste, S; Accoroni, KAG; Pinto, LR; Vencovsky, R; Gimenes, MA; Xavier, MA; Landell, MGA	Genetic variability among sugarcane genotypes based on polymorphisms in sucrose metabolism and drought tolerance genes	Diversity	TRAP	340	n	EUPHYTICA	8	2010	10.1007/s10681-009-0078-2
Devaramath, RM; Kalwade, SB; Kawat, PG; Sashtir, KV	Assessment of Genetic Diversity in Sugarcane Germplasm Using ISSR and SSR Markers	Diversity	SSR/ISSR	139	n	SUGAR TECH	8	2012	10.1007/s12355-012-0168-7
Welker, CAD; Souza-Chies, TT; Longhi-Wagner, HM; Pachoto, MC; McKain, MR; Kellogg, EA	PHYLOGENETIC ANALYSIS OF SACCHARUM S.L. (POACEAE: ANDROPOGONEAE), WITH EMPHASIS ON THE CIRCUMSCRIPTION OF THE SOUTH AMERICAN SPECIES	Diversity	sequence	5	n	AMERICAN JOURNAL OF BOTANY	8	2015	10.3732/ajb.1400397
Cammoa, ER; Arencibia, AD; Lopez, J; Simpson, J; Vargas, D; Sola, F	Analysis of genomic variability in transgenic sugarcane plants produced by Agrobacterium tumefaciens infection	Fingerprinting/Characterization	AFLP	335	n	PLANT BREEDING	7	2005	10.1111/j.1439-0523.2004.01021.x
Baker, P; Jackson, P; Aitken, K	Bayesian estimation of marker dosage in sugarcane and other autopolyploids	Characterization	AFLP	566	s	THEORETICAL AND APPLIED GENETICS	7	2010	10.1007/s00122-010-1283-z
Creste, S; Sansoli, DM; Tardiani, ACS; Silva, DN; Goncalves, FK; Favero, TM; Medeiros, CNF; Festucci, CS; Carlini-Garcia, LA; Landell, MGA; Pinto, LR	Comparison of AFLP, TRAP and SSRs in the estimation of genetic relationships in sugarcane	Diversity	LP/TRAP/SSR	410	n	SUGAR TECH	7	2010	
Govindaraj, P; Sindhu, R; Balamurugan, A; Appunni, C	Molecular Diversity in Sugarcane Hybrids (Saccharum spp. Complex) Grown in Peninsular and East Coast Zones of Tropical India	Diversity	RAPD	69	n	SUGAR TECH	7	2011	10.1007/s12355-011-0095-z

S1 Table. (cont)

65

Authors	Title	Classification	Marker Class	Number of alleles	Dosage information	Source	Total Citation	Publication Year	DOI
Suman, A., Ali, K.; Arto, J.; Parco, AS; Kimbeng, CA.; Bainsakh, N	Molecular Diversity Among Members of the Saccharum Complex Assessed Using TRAP Markers Based on Lignin-Related Genes	Diversity	TRAP	939	n	BIOENERGY RESEARCH	7	2012	10.1007/s12155-011-9123-9
Khan, IA.; Dahot, MU.; Khatri, A	Study of genetic variability in sugarcane induced through mutation breeding	Fingerprinting/ Characterization	RAPD	42	n	PAKISTAN JOURNAL OF BOTANY	6	2007	
Devarumath, RM.; Kalwade, SB.; Bundoock, P.; Eliot, FG.; Henry, R	Independent target region amplification polymorphism and single-nucleotide polymorphism marker utility in genetic evaluation of sugarcane genotypes	Diversity	SNP	852	n	PLANT BREEDING	6	2013	10.1111/pbr.12092
McNeil, MD.; Hermann, S.; Jackson, PA.; Aitken, KS	Conversion of AFLP markers to high-throughput markers in a complex polyploid, sugarcane	Marker development	AFLP/ SNP	454	n	MOLECULAR BREEDING	6	2011	10.1007/s11032-010-9441-1
Aitken, KS; McNeil, MD.; Berkman, PJ.; Hermann, S.; Kilian, A.; Bundoock, PC.; Li, J.C	Comparative mapping in the Poaceae family reveals translocations in the complex polyploid genome of sugarcane	Mapping	DAKT	7846	n	BMC PLANT BIOLOGY	6	2014	10.1186/s12870-014-0190-x
Wan, A.; Robinson, N.; Lakshmanan, P.; Schmidt, S.; Aitken, K	A quantitative genetics approach to nitrogen use efficiency in sugarcane	Association	AFLP/ SSR	1365	n	FUNCTIONAL PLANT BIOLOGY ACTA	6	2010	10.1071/FP09260
Pagariya, MC.; Hankrishnan, M.; Kulkarni, PA.; Devarumath, RM.; Kavar, PG	Physio-biochemical analysis and transcript profiling of Saccharum officinarum L. submitted to salt stress	Association	RAPD	335	n	PHYSIOLOGIAE PLANTARUM	6	2011	10.1007/s11738-010-0676-6
Singh, RK.; Singh, SP.; Tiwari, DK.; Srivastava, S.; Singh, SB.; Sharma, ML.; Singh, R.; Mohapatra, T.; Singh, NK	Genetic mapping and QTL analysis for sugar yield-related traits in sugarcane	Association	SSR	208	n	EUPHYTICA	6	2013	10.1007/s10681-012-0841-7
Kalwade, SB.; Devarumath, RM	Functional Analysis of the Potential Enzymes Involved in Sugar Modulation in High and Low Sugarcane Cultivars	Association	sequence	7	n	APPLIED BIOCHEMISTRY AND BIOTECHNOLOGY	5	2014	10.1007/s12010-013-0622-3
Lee, DI.; Berding, N.; Jackes, BR.; Biehl, LM	Isozyme markers in Saccharum spp. hybrids and Erianthus arundinaceus (Retz.) Jeswiet	Fingerprinting/ Characterization	Isozyme	16	n	AUSTRALIAN JOURNAL OF AGRICULTURAL RESEARCH	5	1998	10.1071/A96163
Manjunatha, BR.; Virupakshi, S.; Naik, GR	Peroxidase Isozyme polymorphism in popular sugarcane cultivars	Fingerprinting/ Characterization	Isozyme	7	n	CURRENT SCIENCE	5	2003	
Nair, NV.; Selvi, A.; Sreenivasan, TV.; Pushpalatha, KN.; Mary, S	Characterization of intergeneric hybrids of Saccharum using molecular markers	Fingerprinting/ Characterization	RAPD	107	n	GENETIC RESOURCES AND CROP EVOLUTION	5	2006	10.1007/s10722-004-1810-3
Da Silva, CM.; Mangolin, CA.; Mott, AS.; Machado, MFPS	Genetic diversity associated with in vitro and conventional bud propagation of Saccharum varieties using RAPD analysis	Fingerprinting/ Characterization	RAPD	144	n	PLANT BREEDING	5	2008	10.1111/j.1439-0523.2007.01438.x
Doule, RB.; Kavar, PG.; Devarumath, RM.; Nekar, YS	Field performance and RAPD analysis for assessment of genetic variation in sugarcane somaclones	Fingerprinting/ Characterization	RAPD		n	INDIAN JOURNAL OF GENETICS AND PLANT BREEDING	5	2008	
Khan, IA.; Dahot, MU.; Seema, N.; Bibi, S.; Khatri, A	Genetic variability in plantlets derived from callus culture in sugarcane	Fingerprinting/ Characterization	RAPD	21	n	PAKISTAN JOURNAL OF BOTANY	5	2008	
Wang, XH.; Yang, QH.; Li, FS.; He, LL.; He, SC	Molecular Identification of Saccharum spp. X Erianthus fulvus Hybrids Using Sequence-Characterized Amplified Region Markers	Fingerprinting/ Characterization	APD/ SCAR	14	n	CROP SCIENCE	5	2009	10.2135/cropsci.2008.04.0241
Tabasum, S.; Khan, FA.; Nawaz, S.; Iqbal, MZ.; Saeed, A	DNA profiling of sugarcane genotypes using randomly amplified polymorphic DNA	Fingerprinting/ Characterization	RAPD	219	n	GENETICS AND MOLECULAR RESEARCH	5	2010	10.4238/v019-1-gm709

S1 Table. (cont)

Authors	Title	Classification	Marker Class	Number of alleles	Dosage information	Source	Total Citation	Publication Year	DOI
Hameed, U.; Pan, YB; Muhammad, K; Afghani, S; Iqbal, J	Use of simple sequence repeat markers for DNA fingerprinting and diversity analysis of sugarcane (<i>Saccharum</i> spp.) cultivars resistant and susceptible to red rot	Fingerprinting/ Characterization	SSR	144	n	GENETICS AND MOLECULAR RESEARCH	5	2012	10.4238/2012.May.8.1
Gratiotol, C; Regulski, M; Beralian, M; McCombie, WR; da Silva, FR; Neto, AZ; Vicentini, R; Farnelli, L; Henery, AS; Mariensen, RA; Ferreira, PCG	Sugarcane genome sequencing by methylation filtration provides tools for genomic research in the genus <i>Saccharum</i>	Fingerprinting/ Characterization	sequence		n	PLANT JOURNAL	5	2014	10.1111/tpj.12539
Sandhu, HS; Gilbert, RA; Comstock, JC; Gordon, VS; Kondorfer, P; Arundale, RA; El-Hout, N	Registration of UHCP 82-1655 Sugarcane	Fingerprinting/ Characterization	SSR	?	n	JOURNAL OF PLANT REGISTRATIONS	5	2016	10.3198/jpr2014.10.0074src
Que, YX; Pan, YB; Lu, YH; Yang, C; Yang, YT; Huang, N; Xu, LP	Genetic Analysis of Diversity within a Chinese Local Sugarcane Germplasm Based on Start Codon Targeted Polymorphism	Diversity	SNP	163	n	BIOMED RESEARCH INTERNATIONAL	5	2014	10.1155/2014.468375
Khan, MS; Yadav, S; Srivastava, S; Swapna, M; Chandra, A; Singh, RK	Development and utilisation of conserved-intron scanning marker in sugarcane	Marker development	CISP	88	n	AUSTRALIAN JOURNAL OF BOTANY	5	2011	10.1071/BT10188
Comide, MT; Coto, O; Calvo, D; Canales, E; Esquivel, FD; Oramas, GP	Molecular markers for the identification and assisted management of genetic resources for sugarcane breeding	Marker Assisted Selection	RFLP		n	PLANT VARIETIES AND SEEDS	5	2000	
Costet, L; Raboin, LM; Payet, M; D'Hont, A; Nibouche, S	A major quantitative trait allele for resistance to the Sugarcane yellow leaf virus (Luteoviridae)	Association	SSR/ RFLP/ DART	2822	n	PLANT BREEDING	5	2012	10.1111/j.1439-0523.2012.02003.x
Debiakos, S; Rocher, S; Gansmeur, O; Toubi, L; Roques, D; D'Hont, A; Hoarau, JY; Daugrois, JH	Prospecting sugarcane resistance to Sugarcane yellow leaf virus by genome-wide association	Association	DART/ AFLP/	3949	n	THEORETICAL AND APPLIED GENETICS	5	2014	10.1007/s00122-014-2334-7
FELDMAN, P	VARIETAL IDENTIFICATION OF SUGAR-CANE (SACCHARUM SP) BY Isozyme ELECTROPHORESIS	Fingerprinting/ Characterization	Isozyme		n	AGRONOMIE TROPICALE PHYTON- JOURNAL OF EXPERIMENTAL BOTANY	4	1985	
Oropesa, M; deGarcia, E	Use of molecular markers for identification of sugar cane varieties (<i>Saccharum</i> sp)	Fingerprinting/ Characterization	RAPD		n	INTERNATIONAL JOURNAL OF EXPERIMENTAL BOTANY	4	1997	
Suprasanna, P; Manjunatha, BR; Palade, VY; Desai, NS; Bapat, VA	Profiling of culture-induced variation in sugarcane plants regenerated via direct and indirect somatic embryogenesis by using transposon-insertion polymorphism	Fingerprinting/ Characterization	sequence	254	n	SUGAR TECH	4	2010	
Snyman, SJ; Nkwananya, PD; Watt, MP	Alleivation of hyperhydricity of sugarcane plantlets produced in RTA (R) vessels and genotypic and phenotypic characterization of acclimated plants	Fingerprinting/ Characterization	AFLP	932	n	SOUTH AFRICAN JOURNAL OF BOTANY	4	2011	10.1016/j.sagb.2011.03.004
You, Q; Xu, LP; Zheng, YF; Que, YX	Genetic Diversity Analysis of Sugarcane Parents in Chinese Breeding Programmes Using gSSR Markers	Diversity	SSR	88	n	SCIENTIFIC WORLD JOURNAL	4	2013	10.1155/2013/613062
Berkman, PJ; Burdlock, PC; Casu, RE; Henry, RJ; Rae, AL; Aitken, KS	A Survey Sequence Comparison of <i>Saccharum</i> Genotypes Reveals Allelic Diversity Differences	Marker development	sequence		n	TROPICAL PLANT BIOLOGY	4	2014	10.1007/s12042-014-9139-3
RAMAGOPAL, S	PROTEIN VARIATION ACCOMPANIES LEAF DIFFERENTIATION IN SUGARCANE (SACCHARUM-OFFICINARUM) AND IS INFLUENCED BY GENOTYPE	Association	Isozyme	2	n	PLANT CELL REPORTS	4	1994	
Ramos, RLB; Iovar, FI; Junqueira, RM; Lino, FB; Sachetto-Martins, G	Sugarcane expressed sequences tags (ESTs) encoding enzymes involved in lignin biosynthesis pathways	Association	EST/ sequence	6	n	GENETICS AND MOLECULAR BIOLOGY	4	2001	10.1590/S1415-47572001000100031
Nibouche, S; Raboin, LM; Hoarau, JY; D'Hont, A; Costet, L	Quantitative trait loci for sugarcane resistance to the spotted stem borer <i>Chilo sacchariphagus</i>	Association	AFLP/ RFLP/ SSR	1405	n	MOLECULAR BREEDING	4	2012	10.1007/s11032-010-9531-0

Authors	Title	Classification	Marker Class	Number of alleles	Dosage information	Source	Total Citation	Publication Year	DOI
GALLACHER, DJ; LEE, DJ; BERDING, N	USE OF ISOZYME PHENOTYPES FOR RAPID DISCRIMINATION AMONG SUGARCANE CLONES	Fingerprinting/ Characterization	Isozyme		n	AUSTRALIAN JOURNAL OF AGRICULTURAL RESEARCH	3	1995	10.1071/A/R9950601
Deng, ZH; Zhang, MQ; Lin, WL; Cheng, F; Zhang, CM; Li, YC; Lai, LP; Lin, YQ; Chen, RK	Analysis of Disequilibrium Hybridization in Hybrid and Backcross Progenies of Saccharum officinarum x Erianthus arundinaceus	Fingerprinting/ Characterization	ITS	2	n	AGRICULTURAL SCIENCES IN CHINA	3	2010	10.1016/S1671-2927(09)60216-9
Khan, IA; Bibi, S; Yasmeen, S; Seema, N; Khatri, A; Siddiqui, MA; Nizami, GS; Aghlan, S	IDENTIFICATION OF ELITE SUGARCANE CLONES THROUGH TRAP	Fingerprinting/ Characterization	sequence	2	n	PAKISTAN JOURNAL OF BOTANY	3	2011	
Que, YX; Lin, JX; Xu, LP; Guo, JR; Xu, JS; Chen, RK	Molecular cloning and expression analysis of a zeta-class glutathione S-transferase gene in sugarcane	Fingerprinting/ Characterization	sequence	1	n	AFRICAN JOURNAL OF BIOTECHNOLOGY	3	2011	
Sahoo, DP; Samantrai, D; Rout, GR	Rapid clonal propagation of Saccharum officinarum L. Vars. CO-6907 and CO-86249 and to assess the genetic uniformity through molecular markers	Fingerprinting/ Characterization	RAPD	20	n	PLANT BIOSYSTEMS	3	2011	10.1080/11263504.2011.566314
Gomathi, R; Manohari, G; Rakkiyappan, P	Antioxidant Enzymes on Cell Membrane Integrity of Sugarcane Varieties Differing in Flooding Tolerance	Fingerprinting/ Characterization	Isozyme	3	n	SUGAR TECH	3	2012	10.1007/s12355-012-0156-y
Parade, VV; Bhargava, S; Suprasanna, P	Halopriming mediated salt and iso-osmotic PEG stress tolerance and gene expression profiling in sugarcane (Saccharum officinarum L.)	Fingerprinting/ Characterization	sequence	6	n	MOLECULAR BIOLOGY REPORTS	3	2012	10.1007/s11033-012-1821-7
Parco, AS; Avelandea, MC; Hale, AH; Hoy, JW; Kimbeng, CA; Pontif, MJ; Gravois, KA; Barsakh, N	Frequency and distribution of the brown rust resistance gene Br1 and implications for the Louisiana sugarcane breeding programme	Fingerprinting/ Characterization	sequence	2	n	PLANT BREEDING	3	2014	10.1111/ptr.12186
Bacci, M; Miranda, VFO; Martins, VO; Figuereira, AVO; Lemos, MV; Pereira, JO; Martini, CL	A search for markers of sugarcane evolution	Diversity	ITS	17	n	GENETICS AND MOLECULAR BIOLOGY	3	2001	10.1590/S1415-47572001000100023
Shahid, MTH; Khan, FA; Saeed, A; Fareed, I	Variability of red rot-resistant somaclones of sugarcane genotype S97US297 assessed by RAPD and SSR	Diversity	RAPD/ SSR	64	n	GENETICS AND MOLECULAR RESEARCH	3	2011	10.4238/vol10-3gmt1122
Dutra, JA; Resende, LV; Bastos, GO; Neto, DES; Machado, PR	Use of molecular markers RAPD, and ESTs SSR to study genetic variability in sugarcane	Diversity	RAPD/ EST/ SSR	92	n	REVISTA CIENCIA AGRONOMICA	3	2013	
Silva, DC; Souza, MCP; Duarte, LSC; dos Santos, JM; Barbosa, GVD; Almeida, C	New Polymorphic EST-SSR Markers in Sugarcane	Marker development	SSR	576	n	SUGAR TECH	3	2012	10.1007/s12355-012-0184-7
Racedo, J; Pereira, MF; Bertani, R; Funes, C; Gonzalez, V; Cuenya, MI; D'Hont, A; Wein, B; Castagnaro, AP	Br1 gene and potential alternative sources of resistance to sugarcane brown rust disease	Marker Assisted Selection	AFLP	2	n	EUPHYTICA	3	2013	10.1007/s10681-013-0905-3
Gouy, M; Luquet, D; Rouan, L; Martine, JF; Thong-Chane, A; Costet, L; Nibouché, S; Goze, E	Site and Saccharum spontaneum introgression level drive sugarcane yield component traits and their impact on sucrose yield in contrasted radiation and thermal conditions in La Réunion	Association	DART	419	n	FIELD CROPS RESEARCH	3	2015	10.1016/j.fcr.2014.11.002
Li, WF; Wang, XY; Huang, YK; Zhang, RY; Shan, HL; Yin, J; Luo, ZM	Molecular Detection of Br1 Gene and Identification of Brown Rust Resistance in Chinese Sugarcane Germplasm	Fingerprinting/ Characterization	sequence	1	n	SUGAR TECH	2	2017	10.1007/s12355-016-0453-y
Thorat, AS; Sonone, NA; Choudhari, VV; Devanurath, RM; Babu, KH	Plant regeneration from cell suspension culture in Saccharum officinarum L. and ascertaining of genetic fidelity through RAPD and ISSR markers	Fingerprinting/ Characterization	RAPD/ ISSR	15	n	3 BIOTECH	2	2017	10.1007/s12305-016-0579-3
Arenobia, AD; Carmona, E; Perez, G; Vinagre, F; Hemeryk, AS; Santana, I	Identification and characterization of hypervariable sequences within the Saccharum complex	Fingerprinting/ Characterization	AFLP	14	n	PLANT SCIENCE	2	2005	10.1016/j.plantsci.2005.03.017

Authors	Title	Classification	Marker Class	Number of alleles	Dosage information	Source	Total Citation	Publication Year	DOI
Bibi, S.; Khan, J.A.; Khatri, A.; Yasmin, S.; Secma, N.; Aftgan, S.; Anan, MA	SCREENING OF MUTATED POPULATION OF SUGARCANE THROUGH RAPD	Fingerprinting/ Characterization	SSR	55	n	PAKISTAN JOURNAL OF BOTANY	2	2010	
Kim, C.; Robertson, J.S.; Paterson, AH	Inference of subgenomic origin of BACs in an interspecific hybrid sugarcane cultivar by overlapping oligonucleotide hybridizations	Fingerprinting/ Characterization	sequence		n	GENOME	2	2011	10.1139/G11-038
Li, Y.; He, YM; Zu, YQ; Zhan, FD	Identification and cloning of molecular markers for UV-B tolerant gene in wild sugarcane (<i>Saccharum spontaneum</i> L.)	Fingerprinting/ Characterization	RAPD	95	n	JOURNAL OF PHOTOCHEMISTRY AND PHOTOBIOLOGY B-BIOLOGY	2	2011	10.1016/j.jphotobiol.2011.07.002
Tiwari, AK; Mishra, N.; Tripathi, S.; Lal, M.; Singh, RK; Sharma, ML	Assessment of Genetic Stability in Micro propagated Population of Sugarcane Variety CoS 07250 Through SSR Markers	Fingerprinting/ Characterization	SSR	121	n	VEGETOS	2	2011	
Vioia, VR; Lekshmi, M.; Premachandran, MN	Differentiation of cytoplasm of <i>Saccharum</i> and <i>Erianthus</i> species by mitochondrial DNA polymorphism	Fingerprinting/ Characterization	RFLP		n	INDIAN JOURNAL OF GENETICS AND PLANT BREEDING	2	2011	
Hermann, SR; Aitken, KS; Jackson, PA; George, AW; Piperidis, N.; Wei, X.; Kilian, A.; Detering, F	Evidence for second division restitution as the basis for 2n+tn maternal chromosome transmission in a sugarcane cross	Fingerprinting/ Characterization	SSR/ AFLP/ DART	8079	n	EUPHYTICA	2	2012	10.1007/s10681-012-0698-9
Fukuhara, S.; Terajima, Y.; Irei, S.; Sakaguchi, T.; Ujihara, K.; Sugimoto, A.; Matsuda, M	Identification and characterization of intergeneric hybrid of commercial sugarcane (<i>Saccharum</i> spp. hybrid) and <i>Erianthus arundinaceus</i> (Retz.) Jeswiet	Fingerprinting/ Characterization	sequence		n	EUPHYTICA	2	2013	10.1007/s10681-012-0748-3
dos Santos, JM; Barbosa, GVD; Neto, CER; Almeida, C	Efficiency of biparental crossing in sugarcane analyzed by SSR markers	Fingerprinting/ Characterization	SSR	47	n	CROP BREEDING AND APPLIED BIOTECHNOLOGY GENETICS AND MOLECULAR RESEARCH	2	2014	10.1590/1984-70332014v14n2a18
Lu, X.; Zhou, H.; Pan, YB.; Chen, CY; Zhu, JR; Chen, PH; Li, YR; Cai, Q.; Chen, RK	Segregation analysis of microsatellite (SSR) markers in sugarcane polyploids	Fingerprinting/ Characterization	SSR	2392	s+dd+tt	MOLECULAR RESEARCH	2	2015	10.4238/2015.December.23.26
Hermann, S.; Brumbley, S.; McIntyre, CL	Analysing diversity in sugarcane resistance gene analogues	Diversity	sequence/ SNP		n	AUSTRALASIAN PLANT PATHOLOGY	2	2006	10.1071/AP06066
Zhang, HY; He, LL; Zhong, HQ; Li, FS; Heand, SC; Yang, QH	Identification of intergeneric hybrids between <i>Saccharum</i> spp. and <i>Erianthus fulvus</i> with ITS	Diversity	ITS	79	n	AFRICAN JOURNAL OF BIOTECHNOLOGY	2	2009	
Nawaz, S.; Khan, FA.; Tabasum, S.; Zakria, M.; Saeed, A.; Iqbal, MZ	Phylogenetic relationships among <i>Saccharum</i> clones in Pakistan revealed by RAPD markers	Diversity	RAPD	219	n	GENETICS AND MOLECULAR RESEARCH	2	2010	10.4238/vol9-3gm903
Zhang, GM; Li, YR.; He, WZ.; He, H.; Liu, XH.; Song, HZ.; Liu, HB.; Zhu, RC.; Fang, WK	Analysis of the genetic diversity in <i>Saccharum spontaneum</i> L. accessions from Guangxi Province of China with RAPD-PCR	Diversity	RAPD	87	n	SUGAR TECH	2	2010	
Singh, RK.; Khan, MS.; Singh, R.; Pandey, DK.; Kumar, S.; Lal, S	Analysis of Genetic Differentiation and Phylogenetic Relationships among Sugarcane Genotypes Differing in Response to Red Rot	Diversity	SSR	90	n	SUGAR TECH	2	2011	10.1007/s12355-011-0076-2
Srivastava, S.; Gupta, PS.; Singh, PK.; Singh, J.; Jain, R	Genetic diversity analysis of <i>Saccharum spontaneum</i> germplasm using SSR-SSCP and RAPD markers	Diversity	SSR/ RAPD	1089	n	INDIAN JOURNAL OF AGRICULTURAL SCIENCES	2	2011	
Qi, YW.; Pan, YB.; Lao, FY.; Zhang, CM.; Fan, LN.; He, HY.; Liu, R.; Wang, QN.; Liu, SM.; Liu, FY.; Li, QW.; Deng, HH	Genetic Structure and Diversity of Parental Cultivars Involved in China Mainland Sugarcane Breeding Programs as Inferred from DNA Microsatellites	Diversity	SSR	303	n	JOURNAL OF INTEGRATIVE AGRICULTURE	2	2012	
Khan, JA.; Bibi, S.; Yasmin, S.; Khatri, A.; Secma, N	PHENOTYPIC AND GENOTYPIC DIVERSITY INVESTIGATIONS IN SUGARCANE FOR DROUGHT TOLERANCE AND SUCROSE CONTENT	Diversity	RAPD/ TRAP/ STS	48	n	PAKISTAN JOURNAL OF BOTANY	2	2013	

S1 Table. (cont)

Authors	Title	Classification	Marker Class	Number of alleles	Dosage information	Source	Total Citation	Publication Year	DOI
Chandra, A.; Grisham, MP.; Pan, YB	Allelic divergence and cultivar-specific SSR alleles revealed by capillary electrophoresis using fluorescence-labeled SSR markers in sugarcane	Diversity	SSR	159	n	GENOME	2	2014	10.1139/gen-2014-0072
Nayak, SN; Song, J; Villa, A; Pathak, B; Ayala-Silva, T; Yang, XP; Todd, J; Glynn, NC; Kulm, DN; Glaz, B; Gilbert, RA; Comstock, JC; Wang, JP	Promoting Utilization of Saccharum spp. Genetic Resources through Genetic Diversity Analysis and Core Collection Construction	Diversity	SSR	209	n	PLOS ONE	2	2014	10.1371/journal.pone.0110856
Park, JW; Solis-Gracia, N; Trevino, C; da Silva, JA	Exploitation of conserved intron scanning as a tool for molecular marker development in the Saccharum complex	Marker development	RFLP/ CAPS/ SSR /SNP	564	n	MOLECULAR BREEDING	2	2012	10.1007/s11032-011-9683-6
Hodkinson, TR; de Cesare, M; Barth, S	NUCLEAR SSR MARKERS FOR MISCANTHUS, SACCCHARUM, AND RELATED GRASSES (SACCHARINAE, POACEAE)	Marker development	SSR	29	n	APPLICATIONS IN PLANT SCIENCES	2	2013	10.3732/apps.1300042
Banerjee, N; Siratee, A; Yadav, S; Kumar, S; Singh, J; Kumar, S; Pandey, DK; Singh, RK	Marker-trait association study for sucrose and yield contributing traits in sugarcane (Saccharum spp. hybrid)	Association	SSR	989	n	EUPHYTICA	2	2015	10.1007/s10681-015-1422-3
Gouy, M; Rousselle, Y; Chane, AT; Anglade, A; Royovert, S; Nibouche, S; Costet, L	Genome wide association mapping of agro-morphological and disease resistance traits in sugarcane	Association	AFLP/ DART/ SSR	3327	n	EUPHYTICA	2	2015	10.1007/s10681-014-1294-y
Liu, Y; Yao, YL; Hu, XW; Xing, SL; Xu, L	Cloning and allelic variation of two novel catalase genes (SacAT-1 and SacCAT-1) in Saccharum officinarum L. and Saccharum spontaneum L.	Association	sequence		n	BIOTECHNOLOGY & BIOTECHNOLOGICAL EQUIPMENT	2	2015	10.1080/13102818.2015.1018839
Pan, YB; Liu, PW; Que, YX	Independently Segregating Simple Sequence Repeats (SSR) Alleles in Polyploid Sugarcane	Fingerprinting/ Characterization	SSR	7	n	SUGAR TECH	1	2015	10.1007/s12355-014-0330-5
Ahmed, MS; Gardezi, SDA	MOLECULAR CHARACTERIZATION OF LOCALLY ADOPTED SUGARCANE (Saccharum officinarum L.) VARIETIES USING MICROSATELLITE MARKERS	Fingerprinting/ Characterization	SSR	380	n	JOURNAL OF ANIMAL AND PLANT SCIENCES	1	2017	
BARRETO, A; SIMON, JP	THE USE OF ISOENZYMES AS GENETIC TRACERS IN SACCCHARUM	Fingerprinting/ Characterization	Isozyme		n	TURKALBA	1	1982	
Govindaraj, P; Balamurugan, A; Nararajan, US	Identification of intergeneric hybrids between Erianthus arundinaceus and Saccharum spontaneum through STMS markers	Fingerprinting/ Characterization	STMS	117	n	INTERNATIONAL SUGAR JOURNAL	1	2012	
Nakayama, S	Inter-MITE polymorphisms of a newly identified MITE show relationships among sugarcane (Saccharum) species	Fingerprinting/ Characterization	sequence		n	GENETIC RESOURCES AND CROP EVOLUTION	1	2012	10.1007/s10722-011-9766-6
Adriano, RC; Azania, CAM; Pinto, LR; Azania, AADMF; Perecin, D	Phenotypic and biochemical responses of sugarcane cultivars to glyphosate application	Fingerprinting/ Characterization	Isozyme	2	n	SUGAR TECH	1	2013	10.1007/s12355-013-0209-x
Baer, OT; Lalusin, AG	Molecular Markers Associated to Downy Mildew [Peronosclerospora philippinensis (W. Weston) CG Shaw] Resistance in Sugarcane (Saccharum officinarum L.) Hybrids (CP 57-604 X PHIL 84-77)	Fingerprinting/ Characterization	SSR	26	n	PHILIPPINE JOURNAL OF CROP SCIENCE	1	2013	
Gomathi, R; Yukashini, K; Shiyamala, S; Vasantha, S; Suganya, A; Rakkayappan, P	Induced Response of Sugarcane Variety Co 86032 for Thermotolerance	Fingerprinting/ Characterization	Isozyme	5	n	SUGAR TECH	1	2013	10.1007/s12355-012-0192-7
Khan, SA; Hanif, Z; Ishaq, U; Ahmad, R; Yasin, M; Chandhary, MF; Afroz, A; Javed, MT; Rashid, U; Rashid, H	Genetic Transformation of Sugarcane Variety HSF-240 with Marker Gene GUS	Fingerprinting/ Characterization	sequence	1	n	INTERNATIONAL JOURNAL OF AGRICULTURE AND BIOLOGY GENETICS AND MOLECULAR RESEARCH	1	2013	
Meloni, MLG; Scarpati, MS; Pinto, LR; Perecin, D; Xavier, MA; Landell, MGA	Selfing rate estimation in sugarcane under unfavorable natural conditions of crossing by using microsatellite markers	Fingerprinting/ Characterization	SSR	64	n		1	2014	10.4238/2014.March.31.8

SI Table. (cont)

Authors	Title	Classification	Marker Class	Number of alleles	Dosage information	Source	Total Citation	Publication Year	DOI
Xavier, MA; Pinto, LR; Favero, TM; Pereira, D; Carlini-Garcia, LA; Landell, MGA	Paternity identification in sugarcane polycrosses by using microsatellite markers	Fingerprinting/ Characterization	SSR	453	n	GENETICS AND MOLECULAR RESEARCH	1	2014	10.4238/2014.March.31.7
Singh, RK; Singh, RB; Singh, SP; Sharma, ML	Genes tagging and molecular diversity of red rot susceptible/tolerant sugarcane hybrids using c-DNA and unigene derived markers	Diversity	SSR	105	n	WORLD JOURNAL OF MICROBIOLOGY & BIOTECHNOLOGY	1	2012	10.1007/s11274-011-0974-1
Fan, LN; Deng, HH; Luo, QW; He, HY; Li, Y; Wang, QN; Huang, ZX; Wu, JT; Li, QW; Liu, SM; Qi, YW	Genetic diversity of Saccharum spontaneum from geographical regions of China assessed by simple sequence repeats	Diversity	SSR	239	n	GENETICS AND MOLECULAR RESEARCH	1	2013	10.4238/2013.November.26.1
Govindaraj, P; Sindhu, P; Appunu, C; Parthiban, S; Senthilkumar, S	Genetic diversity analysis among interspecific and intergeneric hybrids of Saccharum spp. using STMS markers	Diversity	STMS	192	n	RESEARCH ON CROPS	1	2013	
Muhammad, K; Afghani, S; Pan, YB; Iqbal, J	GENETIC VARIABILITY AMONG THE BROWN RUST RESISTANT AND SUSCEPTIBLE GENOTYPES OF SUGARCANE BY RAPD TECHNIQUE	Diversity	RAPD	164	n	PAKISTAN JOURNAL OF BOTANY	1	2013	
Saravakumar, K; Govindaraj, P; Appunu, C; Senthilkumar, S; Kumar, R	Analysis of genetic diversity in high biomass producing sugarcane hybrids (Saccharum spp. complex) using RAPD and STMS markers	Diversity	RAPD/ STMS	162	n	INDIAN JOURNAL OF BIOTECHNOLOGY	1	2014	
Seena, N; Khan, IA; Raza, S; Yasmeen, S; Bibi, S; Nizamat, GS	ASSESSMENT OF GENETIC VARIABILITY IN SONACLONAL POPULATION OF SUGARCANE	Diversity	RAPD	52	n	PAKISTAN JOURNAL OF BOTANY	1	2014	
Raj, P; Selvi, A; Prathima, PT; Nair, NV	Analysis of Genetic Diversity of Saccharum Complex Using Chloroplast Microsatellite Markers	Diversity	SSR	54	n	SUGAR TECH	1	2016	10.1007/s12355-015-0382-1
Manjbas, NL; Villegas, LC	Molecular markers for improving selection of sugarcane varieties with downy mildew resistance	Association	SSR	147	n	PHILIPPINE JOURNAL OF CROP SCIENCE	1	2007	
Margarido, GRA; Pastina, MM; Souza, AP; Garcia, AAF	Multi-trait multi-environment quantitative trait loci mapping for a sugarcane commercial cross provides insights on the inheritance of important traits	Association	RFLP/ SSR	317	s+d	MOLECULAR BREEDING	1	2015	10.1007/s11032-015-0366-6
Santos, FRC; Pinto, LR; Carlini-Garcia, LA; Garzañi, R; Mancini, MC; Gonçalves, BS; Medeiros, CNF; Pereira, D; Garcia, AAF; Souza, AP; Zucchi, MI	Marker-trait association and epistasis for brown rust resistance in sugarcane	Association	SSR/ AFLP	488	n	EUPHYTTICA	1	2015	10.1007/s10681-014-1257-3
Vantini, JS; Dedero, GC; Gimenez, DEJ; Fonseca, LFS; Tezza, RID; Mutton, MA; Ferro, JA; Ferro, MIT	Differential gene expression in drought-tolerant sugarcane roots	Association	AFLP	173	n	GENETICS AND MOLECULAR RESEARCH	1	2015	10.4238/2015.June.29.13
Liu, PW; Chandra, A; Que, YX; Chen, PH; Gristham, MP; White, WH; Dailey, CD; Tew, TL; Pan, YB	Identification of quantitative trait loci controlling sucrose content based on an enriched genetic linkage map of sugarcane (Saccharum spp. hybrids) cultivar 'LCP 85-384'	Association	SSR/ TRAP/ AFLP	967	s+d	EUPHYTTICA	1	2016	10.1007/s10681-015-1538-5
McCORD, PH; Migeault, AJ	Genotyping Sugarcane for the Brown Rust Resistance Locus Brul Using Unlabeled Probe Melting	Association	SNP	2	n	SUGAR TECH	1	2016	10.1007/s12355-015-0390-1
Hsie, BS; Brito, JZ; Nova, MXV; Borges-Paluch, LR; Silva, MV; Donato, VMST	Determining the genetic stability of micropropagated sugarcane using inter-simple sequence repeat markers	Fingerprinting/ Characterization	ISSR	108	n	GENETICS AND MOLECULAR RESEARCH	0	2015	10.4238/2015.December.21.38
Ali, A; Wang, JD; Pan, YB; Deng, ZH; Chen, ZW; Chen, RK; Gao, SJ	Molecular Identification and Genetic Diversity Analysis of Chinese Sugarcane (Saccharum spp. Hybrids) Varieties using SSR Markers	Diversity	SSR	117	n	TROPICAL PLANT BIOLOGY	0	2017	10.1007/s12042-017-9195-6
Augusto, R; Maranhão, RC; Mangolin, CA; Bespalhok, IC; Machado, MFRS	Changes on microsatellites of expressed sequence tag of sugarcane (Saccharum spp) during vegetative propagation	Diversity	SSR	27	n	GENETICS AND MOLECULAR RESEARCH	0	2017	10.4238/gmr16019519

S1 Table. (cont)

Authors	Title	Classification	Marker Class	Number of alleles	Dosage information	Source	Total Citation	Publication Year	DOI
Chen, S.; Shen, WK.; Xu, GH.; Wu, XM.; Deng, QQ.; Dou, ZM	Assessment of Genetic Relationship and Diversity among Chinese Sugarcane Parental Clones using SCot and ISSR Markers	Diversity	ISSR	200	n	INTERNATIONAL JOURNAL OF AGRICULTURE AND BIOLOGY	0	2017	10.17957/IJAB/15.0279
dos Santos, FRC.; Zucchi, MI.; Park, JW.; Benatti, TR.; da Silva, JA.; Souza, GM.; Landell, MGA.; Pinto, LR	New Sugarcane Microsatellites and Target Region Amplification Polymorphism Primers Designed from Candidate Genes Related to Disease Resistance	Marker development	SSR	10	n	SUGAR TECH	0	2017	10.1007/s12355-016-0457-7
Forough, JF.; Avinash, T.; Rachayya, D	Analysis of Molecular Assortment in Sugarcane Varieties using RAPD and ISSR markers	Diversity	ISSR/ RAPD	269	n	RESEARCH JOURNAL OF BIOTECHNOLOGY IN VITRO CELLULAR & DEVELOPMENTAL BIOLOGY- PLANT	0	2017	
Martinez-Estrada, E.; Camal-Velazquez, JH.; Salinas-Ruiz, J.; Bello-Bello, JJ	Assessment of somaclonal variation during sugarcane micropropagation in temporary immersion bioreactors by intersimple sequence repeat (ISSR) markers	Fingerprinting/ Characterization	ISSR	38	n		0	2017	10.1007/s11627-017-9852-3
Sharma, R.; Tanta, S	Red rot resistant gene characterization using RGAP markers among sugarcane cultivars resistant and susceptible to the red rot disease	Association	sequence	28	n	3 BIOTECH	0	2017	10.1007/s13205-017-0941-0
Singh, MP.; Comstock, JC.; Davidson, W.; Gordon, V.; Sandhu, HS.; McCord, P.; Zhao, DL.; Sood, S.; Balazsar, M.; McCorkle, K	Registration of 'CP 06-2425', 'CP 06-2495', 'CP 06-2964', 'CP 06-3103', and 'CP 07-1313' Sugarcane for Sand Soils in Florida	Fingerprinting/ Characterization	SSR	26	n	JOURNAL OF PLANT REGISTRATIONS	0	2017	10.3198/jp2016.09.0045csc
Singh, P.; Singh, SP.; Tiwari, AK.; Sharma, BL	Genetic diversity of sugarcane hybrid cultivars by RAPD markers	Diversity	RAPD	120	n	3 BIOTECH	0	2017	10.1007/s13205-017-0855-x
Sirace, A.; Banerjee, N.; Kumar, S.; Khan, MS.; Singh, PK.; Kumar, S.; Sharma, S.; Singh, RK.; Singh, J	Identification of marker-trait associations for morphological descriptors and yield component traits in sugarcane	Association	SSR	1546	n	PHYSIOLOGY AND MOLECULAR BIOLOGY OF PLANTS	0	2017	10.1007/s12298-016-0403-x
Wang, HB.; Chen, PH.; Yang, YQ.; D'Hont, A.; Lu, YH	Molecular insights into the origin of the brown rust resistance gene Br1 among Saccharum species	Fingerprinting/ Characterization	sequence	1	n	THEORETICAL AND APPLIED GENETICS	0	2017	10.1007/s00122-017-2968-3
Wei, XM.; Jackson, P	Addressing slow rates of long-term genetic gain in sugarcane	Diversity	DART	384	n	INTERNATIONAL SUGAR JOURNAL	0	2017	
Yang, XP.; Sood, S.; Glynn, N.; Islam, MS.; Comstock, J.; Wang, JP	Constructing high-density genetic maps for polyploid sugarcane (Saccharum spp.) and identifying quantitative trait loci controlling brown rust resistance	Association	SNP	17358	single	MOLECULAR BREEDING	0	2017	10.1007/s11032-017-0716-7
Feldmann, P.; Daugrois, J.; Davis, MJ.; Chantrel, M.; Kott, P	First report of leaf scald disease and ratoon stunting disease of sugarcane in French Guyana	Fingerprinting/ Characterization	RFLP	54	n	PLANT DISEASE	0	1997	10.1094/PDIS.1997.81.6.696A
Barnes, JM.; Bolka, FC	Progress towards identifying a marker for rust resistance in sugarcane variety NCo376	Fingerprinting/ Characterization	RAPD/ SCAR	173	n	PROCEEDINGS OF THE SEVENTY-SECOND ANNUAL CONGRESS OF THE SOUTH AFRICAN SUGAR TECHNOLOGISTS ASSOCIATION	0	1998	
Harvey, M.; D'Hont, A.; Alix, K.; Huckett, BI	Use of PCR-based markers for identification of Erianthus genetic material in putative intergeneric hybrids (Saccharum x Erianthus)	Fingerprinting/ Characterization	RFLP		n	PROCEEDINGS OF THE SEVENTY-SECOND ANNUAL CONGRESS OF THE SOUTH AFRICAN SUGAR TECHNOLOGISTS ASSOCIATION	0	1998	
Harvey, M.; Huckett, B	Preliminary genetic analysis of the sugarcane population AA40	Fingerprinting/ Characterization	RFLP		n	PROCEEDINGS OF THE SEVENTY-SECOND ANNUAL CONGRESS OF THE SOUTH AFRICAN SUGAR TECHNOLOGISTS ASSOCIATION	0	1998	
Shoda, M.; Nagamine, T.; Terauchi, T.; Akamine, F.; Sugimoto, A	Isozyme application for variety identification and progeny hybridity in Japanese sugarcane	Fingerprinting/ Characterization	Isozyme	50	n	BREEDING SCIENCE	0	1999	
Liang, C.; Nan, L.; Zhen, H.; Pin, W.; Chanbei, L.; Mingshan, S	Molecular assay of novel pasturage sugarcane-grass 94-42, using RAPD markers	Fingerprinting/ Characterization	RAPD	116	n	GENETIC RESOURCES AND CROP EVOLUTION	0	2004	023/B/GRES.0000023463.42275

Authors	Title	Classification	Marker Class	Number of alleles	Dosage information	Source	Total Citation	Publication Year	DOI
Scott, AW; Da Silva, JA; Rozzel, N; Hernandez, E; Legendre, BL; Miller, JD	Registration of 'TTCP89-3505' sugarcane	Fingerprinting/ Characterization			n	CROP SCIENCE	0	2005	10.2135/cropps2003.0639
Rani, K; Sandhu, SK; Gosal, SS	Genetic Augmentation of Sugarcane Through Direct Gene Transformation with Osgly II Gene Construct	Fingerprinting/ Characterization	sequence	1	n	SUGAR TECH	0	2012	10.1007/s12355-012-0149-x
Glynn, NC; Zhao, DL; Comstock, JC; Mhligan, SB; Glaz, B; Edme, SJ; Davidson, RW; Gilbert, RA; Sood, S; Dhont, A; Hu, CJ	Registration of 'CPCL 02-0926' Sugarcane	Fingerprinting/ Characterization	SSR	29	n	JOURNAL OF PLANT REGISTRATIONS	0	2013	10.3198/jpr2012.12.0054erc
Costa, PMA; Almeida, CF; Silveira, G; Soares, B; Barfi, DCF; Peternelli, LA; Bhering, LL; Barbosa, MHP	Selfing confirmation in sugarcane by using simple sequence repeat markers: an individual reciprocal recurrent selection scheme	Fingerprinting/ Characterization	SSR	62	n	GENETICS AND MOLECULAR RESEARCH	0	2014	10.4238/2014.October.31.11
Priji, PJ; Hemaprabha, G	Sugarcane-specific drought responsive candidate genes belonging to ABA-independent pathway identified from tolerant and susceptible clones of Saccharum and Erianthus species	Fingerprinting/ Characterization	sequence	14	n	INDIAN JOURNAL OF GENETICS AND PLANT BREEDING	0	2014	10.5958/j.0975-6906.74.1.009
Berkman, PJ; Casu, RE; Stiller, J; Rae, AL; Aitken, KS	Towards the sugarcane genome sequence and an understanding of polyploidy	Fingerprinting/ Characterization	sequence		n	INTERNATIONAL SUGAR JOURNAL	0	2015	
Gao, YJ; Liu, XH; Zhang, RH; Zhou, H; Liao, JX; Dian, WX; Zhang, GM	Verification of Progeny from Crosses Between Sugarcane (Saccharum spp.) and an Intergenic Hybrid (Erianthus arundinaceus x Saccharum spontaneum) with Molecular Markers	Fingerprinting/ Characterization	SSR/ SRAP	223	n	SUGAR TECH	0	2015	10.1007/s12355-014-0310-9
Kumar, R; Appunni, C; Durai, AA; Premchandran, MN; Viola, VR; Ram, B; Mahadevaiah, C; Meena, MR; Manjunatha, T	Genetic Confirmation and Field Performance Comparison for Yield and Quality Among Advanced Generations of Erianthus arundinaceus, E. bengalensis and Saccharum spontaneum Cyto-Nuclear Genome Introgressed Sugarcane Intergenic Hybrids	Fingerprinting/ Characterization	RFLP	?	n	SUGAR TECH	0	2015	10.1007/s12355-014-0333-2
Kumar, R; Premchandran, MN; Appunni, C; Durai, AA; Viola, VR; Mahadevaiah, C; Meena, MR	Molecular identification and genetic improvement vis-a-vis comparison of yield and quality in different generations of Erianthus arundinaceus, E. bengalensis and Saccharum spontaneum cyto-nuclear genome introgressed sugarcane	Fingerprinting/ Characterization	RFLP	4	n	INDIAN JOURNAL OF BIOTECHNOLOGY	0	2015	
Nogera, A; Enrique, R; Petera, MF; Ostengo, S; Racedo, J; Costilla, D; Zossi, S; Cuenya, MT; Filippone, MP; Weilin, B; Castagnaro, AP	Genetic characterization and field evaluation to recover parental phenotype in transgenic sugarcane: a step toward commercial release	Fingerprinting/ Characterization	TRAP	212	n	MOLECULAR BREEDING	0	2015	10.1007/s11032-015-0300-y
Gordon, VS; Comstock, JC; Sandhu, HS; Gilbert, RA; Sood, S; Kondorfer, P; El-Hout, N; Arundale, R	Registration of 'UF-CP 84-1047' Sugarcane for Use as a Biofuel Feedstock	Fingerprinting/ Characterization	SSR	?	n	Journal of Plant Registrations	0	2016	10.3198/jpr2015.03.0021erc
Gordon, VS; Comstock, JC; Sandhu, HS; Gilbert, RA; Sood, SG; Kondorfer, P; El-Hout, N; Arundale, R	Registration of 'UF-CP 87-0053' Sugarcane for Use as a Biofuel Feedstock	Fingerprinting/ Characterization	SSR	?	n	Journal of Plant Registrations	0	2016	10.3198/jpr2015.03.0022erc
Sandhu, JS; Kaur, M; Kaur, A; Kalra, A	Single step direct transgenic plant regeneration from adventive embryos of agro-injected sugarcane (Saccharum spp.) spindle leaf roll segments with assured genetic fidelity	Fingerprinting/ Characterization	SSR	21	n	PLANT CELL, TISSUE AND ORGAN CULTURE	0	2016	10.1007/s11240-015-0936-7
Saipute, SA; Suprasanna, P; Patade, VY; Kulkarni, UG; Bapat, VA	MOLECULAR DIVERSITY IN SUGARCANE (Saccharum officinarum L.) CLONES AS REVEALED BY DIFFERENT MOLECULAR MARKERS	Diversity	RAPD	158	n	SABRAO JOURNAL OF BREEDING AND GENETICS	0	2009	
Lavanya, DL; Hemaprabha, G	Genetic diversity within sucrose rich parental pool of sugarcane and its application in sugarcane breeding through hybridization and selection	Diversity	STMS		n	INDIAN JOURNAL OF GENETICS AND PLANT BREEDING PHYTON-INTERNATIONAL JOURNAL OF EXPERIMENTAL BOTANY	0	2010	
Rosas, JAA; Tovar, MEG; Espinosa, HEL; Gomez, OGA	Genetic variability in 22 sugarcane hybrid varieties (Saccharum spp. Hybrid)	Diversity	RAPD		n	INTERNATIONAL JOURNAL OF EXPERIMENTAL BOTANY	0	2010	
Caraballoso, V; Jorge, H; Garcia, H; Gonzalez, A; Bernal, N; Cespedes, A; Rodriguez, R; Puchades, Y; Arencibia, AD	Management Flowering Ability to Increase Efficiency in the Sugarcane Breeding Program	Diversity	ISTR	2	n	SUGAR TECH	0	2012	10.1007/s12355-011-0121-1

S1 Table. (cont)

Authors	Title	Classification	Marker Class	Number of alleles	Dosage information	Source	Total Citation	Publication Year	DOI
Hemaphysa, G.; Simon, S	Genetic Diversity and Selection Among Drought Tolerant Genotypes of Sugarcane Using Microsatellite Markers	Diversity	SSR	214	n	SUGAR TECH	0	2012	10.1007/s12355-012-0164-y
Singh, RK; Singh, RB; Singh, SP; Mishra, N; Kasioglu, J; Sharma, ML; Kumar, A	Genetic Diversity Among Saccharum spontaneum Clones and Commercial Hybrids Through SSR Markers	Diversity	SSR	2489	n	SUGAR TECH	0	2013	10.1007/s12355-013-0203-3
Kalivade, SB; Devanurath, RM	Single strand conformation polymorphism of genomic and EST-SSRs marker and its utility in genetic evaluation of sugarcane	Diversity	SSR	99	n	PHYSIOLOGY AND MOLECULAR BIOLOGY OF PLANTS	0	2014	10.1007/s12298-014-0231-9
Maranho, RC; Augusto, R; Mangolin, CA; Machado, MPPS	Use of differential levels of mean observed heterozygosity in microsatellite loci of commercial varieties of sugarcane (Saccharum spp)	Diversity	SSR	70	n	GENETICS AND MOLECULAR RESEARCH	0	2014	10.4238/2014.December.4.7
Augusto, R; Maranho, RC; Mangolin, CA; Machado, MDDP	High Polymorphism in Est-SSR Loci for Cellulose Synthase and beta-Amylase of Sugarcane Varieties (Saccharum spp.) Used by the Industrial Sector for Ethanol Production	Diversity	SSR	14	n	BIOCHEMISTRY AND BIOTECHNOLOGY INDIAN JOURNAL OF GENETICS AND PLANT BREEDING	0	2015	10.1007/s12010-014-1340-1
Chaudhary, CP; Singh, S; Rao, VP; Sengur, RS	Assessment of genetic variability in sugarcane (Saccharum officinarum L.) under salinity conditions	Diversity	SSR	54	n	INDIAN JOURNAL OF GENETICS AND PLANT BREEDING	0	2015	10.5958/0975-6906.2015.00021.8
Hapsoro, D; Waganegara, HA; Utomo, SD; Stryani, N; Yusnita	GENETIC DIVERSITY AMONG SUGARCANE (Saccharum officinarum L.) GENOTYPES AS SHOWN BY RANDOMLY AMPLIFIED POLYMORPHIC DNA (RAPD)	Diversity	RAPD	29	n	AGRIVITA	0	2015	10.17503/Agriwita-2015-37-3-p247-257
Pocovi, MI; Mariotti, JA	A bayesian approach to inferring the genetic population structure of sugarcane accessions from INTA (Argentina)	Diversity	SSR/ AFLP	243	n	CHILEAN JOURNAL OF AGRICULTURAL RESEARCH REVISTA DE LA FACULTAD DE AGRONOMIA DE LA UNIVERSIDAD DEL ZULIA	0	2015	10.4067/S0718-58392015000200003
Sosa, D; Rea, R; Latorre, C; Molina, S; Demey, J; Briceno, R; De Sousa-Vieira, O	Genetic diversity in sugarcane germplasm using trap markers based on sucrose-related genes	Diversity	TRAP	1303	n	INDIAN JOURNAL OF GENETICS AND PLANT BREEDING	0	2015	10.5958/0975-6906.2015.00013.9
Lin, XL; Li, Xi; Xu, CH; Lin, XQ; Deng, ZH	Genetic Diversity of Populations of Saccharum spontaneum with Different Ploidy Levels Using SSR Molecular Markers	Diversity	SSR	491	n	SUGAR TECH	0	2016	10.1007/s12355-015-0399-5
Ul-Haq, S; Kumar, P; Singh, RK; Verma, KS; Bhatt, R; Sharma, M; Kachhwaha, S; Kohari, SL	Assessment of Functional EST-SSR Markers (Sugarcane) in Cross-Species Transferability: Genetic Diversity among Poaceae Plants, and Bulk Segregation Analysis	Diversity	SSR	519	n	GENETICS RESEARCH INTERNATIONAL	0	2016	10.1155/2016/7052323
You, Q; Pan, YB; Xu, LP; Gao, SW; Wang, QN; Su, YC; Yang, YQ; Wu, QB; Zhou, DG; Que, YX	Genetic Diversity Analysis of Sugarcane Germplasm Based on Fluorescence-Labeled Simple Sequence Repeat Markers and a Capillary Electrophoresis-based Genotyping Platform	Diversity	SSR	205	n	SUGAR TECH	0	2016	10.1007/s12355-015-0395-9
Hemaphysa, G; Lavanya, DL	Sucrose specific TRAP markers as genus and species specific markers in Saccharum and Erianthus spp.	Marker development	TRAP	256	n	INDIAN JOURNAL OF GENETICS AND PLANT BREEDING	0	2015	10.5958/0975-6906.2015.00013.9
Huang, DL; Gao, YJ; Gu, YY; Chen, ZL; Qin, CX; Wang, M; Liao, Q; Yang, LT; Li, YR	Transcriptome of High-Sucrose Sugarcane Variety GT135	Marker development	sequence		n	SUGAR TECH	0	2016	10.1007/s12355-015-0420-z
Chen, JW; Liao, FY; Chen, XW; Deng, HH; Liu, R; He, HY; Fu, C; Chen, YS; Liu, FY; Li, QW; Jackson, P; Aitken, K	DNA Marker Transmission and Linkage Analysis in Populations Derived from a Sugarcane (Saccharum spp.) x Erianthus arundinaceus Hybrid	Mapping	SSR/ AFLP	756	s+r+dd	PLOS ONE	0	2015	10.1371/journal.pone.0128865
Arro, J; Park, JW; Wai, CM; VanBuren, R; Pan, YB; Nagai, C; da Silva, J; Ming, R	Balancing selection contributed to domestication of autopolyploid sugarcane (Saccharum officinarum L.)	Comparative Genomics or Evolution	SNP	1741380	n	EUPHYTICA	0	2016	10.1007/s10681-016-1672-8
Lin, XL; Li, Xi; Lin, HB; Xu, CH; Lin, XQ; Li, C; Deng, ZH	Phylogenetic Analysis of Different Ploidy Saccharum spontaneum Based on rDNA-ITS Sequences	Comparative Genomics or Evolution	ITS		n	PLOS ONE	0	2016	10.1371/journal.pone.0151524

S1 Table. (cont)

Authors	Title	Classification	Marker Class	Number of alleles	Dosage information	Source	Total Citation	Publication Year	DOI
Song, J, Yang, XP, Resende, MFR, Neves, LG, Todd, J, Zhang, JS, Comstock, JC, Wang, JP	Natural Allelic Variations in Highly Polyploidy Saccharum Complex	Comparative Genomics or Evolution	SNP	34397	n	FRONTIERS IN PLANT SCIENCE	0	2016	10.3389/fpls.2016.00804
Tsuruta, S, Ebina, M, Kobayashi, M, Takahashi, W	Complete Chloroplast Genomes of Erianthus arundinaceus and Miscanthus sinensis: Comparative Genomics and Evolution of the Saccharum Complex	Comparative Genomics or Evolution	sequence		n	PLOS ONE	0	2017	10.1371/journal.pone.0169992
Pinto, LR, Leite, DC, Favero, TM, Pastina, MM, Garcia, AAF, Perecin, D, Goncalves, BS, Cresle, S, Xavier, MA, Bidola, MAP, Landell, MGA, de Souza, AP	Identification of microsatellites markers associated with yield components and quality parameters in sugarcane	Association	SSR	215	n	INTERNATIONAL SUGAR JOURNAL	0	2011	
Lavanya, DL, Hemaprabha, G	Identification of a STMS marker for sucrose content in sugarcane (Saccharum spp.)	Association	STMS	42	n	INDIAN JOURNAL OF GENETICS AND PLANT BREEDING	0	2012	
Simon, S, Hemaprabha, G	Sugarcane specific drought responsive candidate genes identified through differential expression in resistant and susceptible genotypes of sugarcane (Saccharum sp.)	Association	sequence	17	n	INTERNATIONAL SUGAR JOURNAL	0	2012	
Singels, A, Smit, MA, Butterfield, MK, Van Heerden, PDR, Van Den Berg, M, Diola, V, Barbosa, MHP, Veiga, CFM, Fernandes, EC	Identifying quantitative trait alleles for physiological traits in sugarcane: An exploratory study	Association	DART	2054	n	INTERNATIONAL SUGAR JOURNAL	0	2012	
Bilal, M, Saeed, M, Nasir, IA, Tabassum, B, Zameer, M, Khan, A, Tariq, M, Javed, MA, Husnain, T	Molecular Markers EST-SSRs for Genotype-Phenotype Association in Sugarcane	Association	SSR	115	n	SUGAR TECH	0	2014	10.1007/s12355-013-0268-z
Hameed, U, Pan, YB, Iqbal, J	Association mapping of cane weight and tillers per plant in sugarcane	Association	SSR	114	n	BIOTECHNOLOGY & BIOTECHNOLOGICAL EQUIPMENT JOURNAL OF PHYTOPATHOLOGY	0	2015	10.1080/13102818.2015.1008203
Kumar, SS, Govindaraj, P, Appunni, C	Genetic Analysis of Resistance Gene Analogues from a Sugarcane Cultivar Resistant to Red Rot Disease	Association	NBS-LRR		n		0	2015	10.1111/jph.12372
McIntyre, CL, Goode, ML, Cordeiro, G, Bandook, P, Eliot, F, Henry, RF, Casu, RE, Bonnett, GD, Aitken, KS	Morphological and Molecular Characterization of High Biomass IGH, ISH and Saccharum Hybrids	Association	STMS/ RAPD	348	n	SUGAR TECH	0	2015	10.1007/s12355-014-0327-0
	Characterisation of alleles of the sucrose phosphate synthase gene family in sugarcane and their association with sugar-related traits	Association	SNP	22	s+dd	MOLECULAR BREEDING	0	2015	10.1007/s11032-015-0286-5
Priji, PJ, Hemaprabha, G	Sugarcane Specific Drought Responsive Candidate Genes Belonging to ABA Dependent Pathway Identified from Basic Species Clones of Saccharum sp and Erianthus sp.	Association	sequence		n	SUGAR TECH	0	2015	10.1007/s12355-014-0313-6
Brigida, ABS, Rojas, CA, GratiVol, C, de Armas, EM, Entenza, JOP, Thiebaut, F, Lima, MD, Farninelli, L, Hemery, AS, Litschitz, S, Ferreira, PCG	Sugarcane transcriptome analysis in response to infection caused by Acidovorax avenae subsp avenae	Association	sequence	798	n	PLOS ONE	0	2016	10.1371/journal.pone.0166473
Costa, EA, Anoni, CO, Mancini, MC, Santos, FRC, Marconi, TG, Gazaffi, R, Pastina, MM, Perecin, D, Mollinari, M, Xavier, MA, Pinto, LR, Souza, AP, Garcia, AAF	QTL mapping including codominant SNP markers with ploidy level information in a sugarcane progeny	Association	AFLP/ SSR/ SNP	668	s+dd	EU/PHYTICA	0	2016	10.1007/s10681-016-1746-7
Evans, DL, Joshi, SV	Elucidating modes of activation and herbicide resistance by sequence assembly and molecular modelling of the Acetolactate synthase complex in sugarcane	Association	sequence		n	JOURNAL OF THEORETICAL BIOLOGY	0	2016	10.1016/j.jtbi.2016.07.025
Jung, JH, Kannan, B, Dermawan, H, Moxley, GW, Altpeter, F	Precision breeding for RNAi suppression of a major 4-coumarate:coenzyme A ligase gene improves cell wall saccharification from field grown sugarcane	Association	sequence		n	PLANT MOLECULAR BIOLOGY	0	2016	10.1007/s11103-016-0527-y
Pardai, SK, Kalra, S, Pandit, A, Nayak, P, Singh, RK, Galkwad, K, Srivastava, PS, Singh, NK, Mohapatra, T	Single nucleotide polymorphism in sugar pathway and disease resistance genes in sugarcane	Association	SNP/ CAPS	26	n	PLANT CELL REPORTS	0	2016	10.1007/s00299-016-1978-y

S1 Table. (cont)

Authors	Title	Classification	Marker Class	Number of alleles	Dosage information	Source	Total Citation	Publication Year	DOI
Racedo, J; Gutierrez, L; Pereira, MF; Ostengo, S; Pardo, EM; Cuenya, MI; Weirin, B; Castagnaro, AP	Genome-wide association mapping of quantitative traits in a breeding population of sugarcane	Association	DART/ TRAP	1745	n	BMC PLANT BIOLOGY	0	2016	10.1186/s12870-016-0829-x
Singh, RK; Banerjee, N; Khan, MS; Yadav, S; Kumar, S; Duttaajunder, SK; Lal, RJ; Patel, JD; Guo, H; Zhang, D; Paterson, AH	Identification of putative candidate genes for red rot resistance in sugarcane (Saccharum species hybrid) using LD-based association mapping	Association	SSR	994	s+d	MOLECULAR GENETICS AND GENOMICS	0	2016	10.1007/s00438-016-1190-3
Balsalobre, TWA; Pereira, GD; Margardo, GRA; Gazaffi, R; Barreto, FZ; Anoni, CO; Cardoso-Silva, CB; Costa, EA; Mancini, MC; Hofmann, HP; de Souza, AP; Garcia, AAF; Carneiro, MS	GBS-based single dosage markers for linkage and QTL mapping allow gene mining for yield-related traits in sugarcane	Association	SNP	993	single	BMC GENOMICS	0	2017	10.1186/s12864-016-3383-x
Khan, M; Pan, YB; Iqbal, J	Development of an RAPD-based SCAR marker for smut disease resistance in commercial sugarcane cultivars of Pakistan	Association	RAPD/ SCAR	84	n	CROP PROTECTION	0	2017	10.1016/j.cropro.2016.12.024

Complete information for the set of 344 articles obtained in the search for studies with molecular markers for sugarcane at the Thomson Reuters Web of Science database.

S2 Table. Journals ranking.

SOURCE	ABSOLUTE FREQUENCY	RELATIVE FREQUENCY	SOURCE	ABSOLUTE FREQUENCY	RELATIVE FREQUENCY
THEORETICAL AND APPLIED GENETICS	39	11.33721	AUSTRALASIAN PLANT PATHOLOGY	1	0.2907
SUGAR TECH	29	8.43023	AUSTRALIAN JOURNAL OF BOTANY	1	0.2907
EUPHYTICA	25	7.26744	AUSTRALIAN JOURNAL OF CROP SCIENCE	1	0.2907
MOLECULAR BREEDING	19	5.52326	BIOENERGY RESEARCH	1	0.2907
GENETICS AND MOLECULAR RESEARCH	14	4.06977	BMC GENETICS	1	0.2907
GENOME	13	3.77907	BREEDING SCIENCE	1	0.2907
PAKISTAN JOURNAL OF BOTANY	11	3.19767	CHILEAN JOURNAL OF AGRICULTURAL RESEARCH	1	0.2907
GENETIC RESOURCES AND CROP EVOLUTION	10	2.90698	CHROMOSOME RESEARCH	1	0.2907
GENETICS AND MOLECULAR BIOLOGY	10	2.90698	CIENCIA E AGROTECNOLOGIA	1	0.2907
PLANT BREEDING	8	2.32558	CROP BREEDING AND APPLIED BIOTECHNOLOGY	1	0.2907
CROP SCIENCE	7	2.03488	CROP PROTECTION	1	0.2907
INDIAN JOURNAL OF GENETICS AND PLANT BREEDING	7	2.03488	CURRENT SCIENCE	1	0.2907
PLOS ONE	7	2.03488	FRONTIERS IN PLANT SCIENCE	1	0.2907
INTERNATIONAL SUGAR JOURNAL	6	1.74419	FUNCTIONAL PLANT BIOLOGY	1	0.2907
PLANT SCIENCE	6	1.74419	GENE	1	0.2907
JOURNAL OF PLANT REGISTRATIONS	5	1.45349	GENETICS RESEARCH INTERNATIONAL	1	0.2907
BMC GENOMICS	4	1.16279	GENOME RESEARCH	1	0.2907
GENETICS	4	1.16279	IN VITRO CELLULAR & DEVELOPMENTAL BIOLOGY-PLANT	1	0.2907
PLANT CELL REPORTS	4	1.16279	INDIAN JOURNAL OF AGRICULTURAL SCIENCES	1	0.2907
3 BIOTECH	3	0.87209	INDUSTRIAL CROPS AND PRODUCTS	1	0.2907
AFRICAN JOURNAL OF BIOTECHNOLOGY	3	0.87209	INTERNATIONAL JOURNAL OF MOLECULAR SCIENCES	1	0.2907
AUSTRALIAN JOURNAL OF AGRICULTURAL RESEARCH	3	0.87209	JOURNAL OF ANIMAL AND PLANT SCIENCES	1	0.2907
BMC PLANT BIOLOGY	3	0.87209	JOURNAL OF BIOMEDICINE AND BIOTECHNOLOGY	1	0.2907
INDIAN JOURNAL OF BIOTECHNOLOGY	3	0.87209	JOURNAL OF ENVIRONMENTAL BIOLOGY	1	0.2907
JOURNAL OF PLANT PHYSIOLOGY	3	0.87209	JOURNAL OF HEREDITY	1	0.2907
PLANT MOLECULAR BIOLOGY	3	0.87209	JOURNAL OF INTEGRATIVE AGRICULTURE	1	0.2907
PROCEEDINGS OF THE SEVENTY- SECOND ANNUAL CONGRESS OF THE SOUTH AFRICAN SUGAR TECHNOLOGISTS' ASSOCIATION	3	0.87209	JOURNAL OF PHOTOCHEMISTRY AND PHOTOBIOLOGY B- BIOLOGY	1	0.2907
APPLIED BIOCHEMISTRY AND BIOTECHNOLOGY	2	0.5814	JOURNAL OF PHYTOPATHOLOGY	1	0.2907
BIOMED RESEARCH INTERNATIONAL	2	0.5814	JOURNAL OF PLANT BIOCHEMISTRY AND BIOTECHNOLOGY	1	0.2907
BIOTECHNOLOGY \& BIOTECHNOLOGICAL EQUIPMENT	2	0.5814	JOURNAL OF THEORETICAL BIOLOGY	1	0.2907
FIELD CROPS RESEARCH	2	0.5814	MAYDICA	1	0.2907
GENETICA	2	0.5814	MOLECULAR BIOLOGY REPORTS	1	0.2907
HEREDITY	2	0.5814	PEST MANAGEMENT SCIENCE	1	0.2907
INTERNATIONAL JOURNAL OF AGRICULTURE AND BIOLOGY	2	0.5814	PHYTOCHEMISTRY	1	0.2907
MOLECULAR GENETICS AND GENOMICS	2	0.5814	PLANT BIOSYSTEMS	1	0.2907
PHILIPPINE JOURNAL OF CROP SCIENCE	2	0.5814	PLANT CELL AND ENVIRONMENT	1	0.2907

S2 Table. Journals ranking (cont.)

SOURCE	ABSOLUTE FREQUENCY	RELATIVE FREQUENCY	SOURCE	ABSOLUTE FREQUENCY	RELATIVE FREQUENCY
PHYSIOLOGY AND MOLECULAR BIOLOGY OF PLANTS	2	0.5814	PLANT DISEASE	1	0.2907
PHYTON-INTERNATIONAL JOURNAL OF EXPERIMENTAL BOTANY	2	0.5814	PLANT VARIETIES AND SEEDS	1	0.2907
PLANT CELL TISSUE AND ORGAN CULTURE	2	0.5814	PLANTA	1	0.2907
PLANT JOURNAL	2	0.5814	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA	1	0.2907
PLANT MOLECULAR BIOLOGY REPORTER	2	0.5814	RESEARCH JOURNAL OF BIOTECHNOLOGY	1	0.2907
PLANT PHYSIOLOGY AND BIOCHEMISTRY	2	0.5814	RESEARCH ON CROPS	1	0.2907
SOUTH AFRICAN JOURNAL OF BOTANY	2	0.5814	REVISTA CIENCIA AGRONOMICA	1	0.2907
TROPICAL PLANT BIOLOGY	2	0.5814	REVISTA DE LA FACULTAD DE AGRONOMIA DE LA UNIVERSIDAD DEL ZULIA	1	0.2907
ACTA PHYSIOLOGIAE PLANTARUM	1	0.2907	SABRAO JOURNAL OF BREEDING AND GENETICS	1	0.2907
AGRICULTURAL SCIENCES IN CHINA	1	0.2907	SCIENTIFIC REPORTS	1	0.2907
AGRIVITA	1	0.2907	SCIENTIFIC WORLD JOURNAL	1	0.2907
AGRONOMIE TROPICALE	1	0.2907	TRANSGENIC RESEARCH	1	0.2907
AMERICAN JOURNAL OF BOTANY	1	0.2907	TURRIALBA	1	0.2907
ANNALS OF APPLIED BIOLOGY	1	0.2907	VEGETOS	1	0.2907
APPLICATIONS IN PLANT SCIENCES	1	0.2907	WORLD JOURNAL OF MICROBIOLOGY & BIOTECHNOLOGY	1	0.2907

Journals in which publications for sugarcane with the use of molecular markers were found. Absolute and relative frequencies are presented.

S3 Table. Main journals and co-citation analysis results.

Source	Absolute Frequency	Citations	Total link Strenght
Theoretical and Applied Genetics	39	1678	6960
Euphytica	25	406	5517
Molecular Breeding	19	305	4033
Genome	13	493	3062
Sugar Tech	29	87	2952
Genetic Resources and Crop Evolution	10	198	1704
Genetics and Molecular Research	14	44	1670
Pakistan Journal of Botany	11	76	1087
Genetics and Molecular Biology	10	157	865

Rank and co-citation analysis results for the main journals that published studies related to the use of molecular markers in sugarcane.

S4 Table. Number of publications per author.

Authors			Authors			Authors			Authors			Authors		
Authors	Absolute Frequency	Relative Frequency	Authors	Absolute Frequency	Relative Frequency	Authors	Absolute Frequency	Relative Frequency	Authors	Absolute Frequency	Relative Frequency	Authors	Absolute Frequency	Relative Frequency
DHONT, A	37	1.82887	SINGH, NK	7	0.34619	BURNQUIST, W	4	0.19782	CARLINI-GARCIA, LA	3	0.14837	OFFMANN, B	3	0.14837
GLAZSMANN, JC	25	1.2364	XU, LP	7	0.34619	CASTAGNARO, AP	4	0.19782	CARNEIRO, MS	3	0.14837	OSTENGO, S	3	0.14837
AITKEN, K	23	1.13749	BERDING, N	6	0.29674	COSTA, EA	4	0.19782	CHARBEGAS, SM	3	0.14837	PARIDA, SK	3	0.14837
MCINTYRE, C	21	1.03858	BUNDOCK, P	6	0.29674	CRESTE, S	4	0.19782	CHANDRA, A	3	0.14837	PARK, JW	3	0.14837
PAN, YB	21	1.03858	CARMONA, E	6	0.29674	CUENYA, MI	4	0.19782	CHEN, PH	3	0.14837	PATDAE, VY	3	0.14837
JACKSON, P	19	0.93966	DE, SOUZA, AP	6	0.29674	DAUGROS, J	4	0.19782	CORINDE, MT	3	0.14837	PAUQUET, J	3	0.14837
PINTO, LR	18	0.89021	GAZAFIL, R	6	0.29674	DE, SOUZA, BARBOSA, GV	4	0.19782	DAHOT, MU	3	0.14837	PREMACHANDRA N, MN	3	0.14837
GRIVET, L	17	0.84075	GILBERT, RA	6	0.29674	DENG, HH	4	0.19782	DOS, SANTOS, JM	3	0.14837	RAO, PS	3	0.14837
GARCIA, AAF	16	0.7913	GLYNN, N	6	0.29674	DENG, ZH	4	0.19782	EL-HOUT, N	3	0.14837	ROMAS, CA	3	0.14837
SINGH, RK	14	0.69238	GOVINDARAJ, P	6	0.29674	DEU, M	4	0.19782	FARINELLI, L	3	0.14837	ROSSI, M	3	0.14837
COMSTOCK, J	12	0.59347	KIMBENG, CA	6	0.29674	ELIOTT, F	4	0.19782	FAVERO, TM	3	0.14837	ROUSSELLE, Y	3	0.14837
HENRY, R	12	0.59347	KUMAR, S	6	0.29674	FERREIRA, PCG	4	0.19782	FELIX, J	3	0.14837	SAEED, A	3	0.14837
HOARAU, JY	12	0.59347	MOORE, P	6	0.29674	GARSMEIR, O	4	0.19782	GAIKWAD, K	3	0.14837	SELVI, AIR, NV	3	0.14837
DA, SILVA, J	10	0.49456	NIBOLCHE, S	6	0.29674	GORDON, V	4	0.19782	GLAZ, B	3	0.14837	SILLS, GR	3	0.14837
QUE, YX	10	0.49456	PERECIN, D	6	0.29674	HONEYCUTT, RU	4	0.19782	GOUY, M	3	0.14837	SINGH, J	3	0.14837
SOUZA, AP	10	0.49456	SELI, A	6	0.29674	HUCKETT, B	4	0.19782	GRATVOU, C	3	0.14837	SINGH, P	3	0.14837
CASU, R	9	0.4451	SINGH, SP	6	0.29674	KAIWADE, SB	4	0.19782	GRAVOIS, KA	3	0.14837	SINGH, RB	3	0.14837
ULIAN, EC	9	0.4451	SOBRAL, BWS	6	0.29674	KAWAR, PG	4	0.19782	GRISHAM, MP	3	0.14837	SORELLIS, ME	3	0.14837
DEVARUMATHI, RM	8	0.39565	SOOD, S	6	0.29674	KHAN, MS	4	0.19782	GUIMARAES, CT	3	0.14837	SOUZA, GM	3	0.14837
HEMAPRAHMA, G	8	0.39565	AFGHAN, S	5	0.24728	LANAID, C	4	0.19782	HARVEY, M	3	0.14837	SREENIVASAN, TV	3	0.14837
HERMANN, S	8	0.39565	ALLANABI, SM	5	0.24728	MANGOLIN, CA	4	0.19782	HE, HY	3	0.14837	SRIVASTAVA, PS	3	0.14837
KHAN, IA	8	0.39565	ALMEIDA, C	5	0.24728	MARGARIDO, GRA	4	0.19782	HELLER-USZYNSKA, K	3	0.14837	SU, YC	3	0.14837
LANDELL, MGA	8	0.39565	BESSE, P	5	0.24728	MENOSSI, M	4	0.19782	JANNOO, N	3	0.14837	SUGIMOTO, A	3	0.14837
MING, R	8	0.39565	BURNER, D	5	0.24728	PERERA, MF	4	0.19782	KORNDORFER, P	3	0.14837	SUPRASANNA, P	3	0.14837
MOHAPATRA, T	8	0.39565	BUTTERFIELD, M	5	0.24728	PIERIDS, G	4	0.19782	KUMAR, R	3	0.14837	SWAPNA, M	3	0.14837
PASTINA, MM	8	0.39565	CHEN, RK	5	0.24728	RACEDO, J	4	0.19782	LAVANYA, DL	3	0.14837	TEW, TL	3	0.14837
PATERSON, A	8	0.39565	HEMERLY, AS	5	0.24728	SANDHU, HS	4	0.19782	LEGENDRE, BL	3	0.14837	VEREMIS, JC	3	0.14837
PAULET, F	8	0.39565	IOBAL, J	5	0.24728	VADAV, S	4	0.19782	LI, FS	3	0.14837	VIEIRA, MLC	3	0.14837
RABON, LM	8	0.39565	IRVINE, J	5	0.24728	VASMIN, S	4	0.19782	LI, JC	3	0.14837	VIOLA, YR	3	0.14837
APRINU, C	7	0.34619	KILIAN, A	5	0.24728	ALIX, K	3	0.14837	LI, QW	3	0.14837	WANG, JP	3	0.14837
ARENCEBIA, A	7	0.34619	LIU, YH	5	0.24728	ALWALA, S	3	0.14837	LI, YR	3	0.14837	WANG, ON	3	0.14837
BIBI, S	7	0.34619	MANGINI, MC	5	0.24728	ARRO, J	3	0.14837	LIU, PW	3	0.14837	WEI, XM	3	0.14837
CORDEIRO, G	7	0.34619	MARCONI, TG	5	0.24728	ARDUNDALE, R	3	0.14837	LIU, SC	3	0.14837	WEILIN, B	3	0.14837
COSTET, L	7	0.34619	MOLLINARI, M	5	0.24728	ALIGUSTO, R	3	0.14837	LOPEZ, J	3	0.14837	YANG, QH	3	0.14837
FELDMANN, P	7	0.34619	SRIVASTAVA, S	5	0.24728	BALASUNDARAM, N	3	0.14837	MACHADO, MFPs	3	0.14837	YANG, XP	3	0.14837
KHATRI, A	7	0.34619	TELSMART, H	5	0.24728	BANERJEE, N	3	0.14837	MANNERS, JM	3	0.14837	ZUCCHI, MI	3	0.14837
OLIVEIRA, KM	7	0.34619	VAN, SLUYS, MA	5	0.24728	BERKMAN, PI	3	0.14837	MARANHO, RC	3	0.14837	Authors that published only two papers	2	1404522
ROQUES, D	7	0.34619	VICENTINI, R	5	0.24728	BIRCH, RG	3	0.14837	MCNEIL, MD	3	0.14837	Authors that published only one paper	1	33.68226
SEEMA, N	7	0.34619	XAVIER, MA	5	0.24728	BOTHA, FC	3	0.14837	MISHRA, N	3	0.14837			
SHARMA, M	7	0.34619	BONNETT, GD	4	0.19782	CAL, Q	3	0.14837	NOYER, JL	3	0.14837			

Absolute and relative frequencies for the total number of papers published per author.

S5 Table. Author's ranking and network results.

Author	Absolute Frequency	Relative Frequency	H-Index	Total Citation	Total Link Strenght*
Angelique D'Hont	37	182.987	24	1492	56
Jean C. Glaszmann	25	12.364	22	1257	43
Karen S. Aitken	23	113.749	13	444	28
Cathrine L. McIntyre	21	103.858	16	818	29
Yong-Bao Pan	21	103.858	9	264	7
Phillip Jackson	19	0.93966	13	418	27
Luciana R. Pinto	18	0.89021	10	363	23
Laurent Grivet	17	0.84075	17	1015	35
Antônio Augusto F. Garcia	16	0.7913	11	453	24
R. K. Singh	14	0.69238	5	78	0

*Author coupling network analysis

H-index values based on production measures and local citation obtained for the 10 most productive authors in the data set, accompanied by the numbers referring to their position in the network, and the values assigned to them due to the number of connections they presented in the analysis.

S7 Table. Countries ranking considering network results.

Country	Total number of Documents	Citations	Total Link Strenght	Country	Total number of Documents	Citations	Total Link Strenght
USA	80	1617	50	Mexico	4	102	4
France	40	1326	46	Denmark	1	10	3
Brazil	64	1419	26	Italy	2	60	3
Reunion	15	592	23	Belgium	1	16	2
Australia	48	1477	22	Netherlands	2	48	2
China	39	307	21	South Korea	2	1	2
Guadaloupe	7	255	12	Uruguay	1	7	2
India	65	594	12	England	1	34	1
Pakistan	21	111	7	Saudi Arabia	1	19	1
Cuba	9	140	6	Thailand	1	24	1
South Africa	14	191	6	Indonesia	1	0	0
Switzerland	3	26	6	Ireland	3	32	0
Argentina	6	41	5	Japan	5	28	0
Barbados	3	213	5	Philippines	2	2	0
Mauritius	3	111	4	Venezuela	3	37	0

Absolute and relative frequencies, citation and network analysis results for all countries that contribute to production of the 344 articles analyzed.

S6 Table. Most cited articles.

Authors	Title	Source	Total Citation	DOI and/or ISI identification
Cordeiro, GM; Casu, R; McIntyre, CL; Manners, JM; Henry, RJ	Microsatellite markers from sugarcane (<i>Saccharum</i> spp.) ESTs cross transferable to <i>erianthus</i> and sorghum	PLANT SCIENCE	257	10.1016/S0168-9452(01)00365-X and ISI000168778200004
Ming, R; Liu, SC; Lin, YR; da Silva, J; Wilson, W; Braga, D; van Deynze, A; Wenslaiff, TF; Wu, KK; Moore, PH; Burnquist, W; Sorrells, ME; Irvine, JE; Paterson, AH	Detailed alignment of <i>Saccharum</i> and <i>Sorghum</i> chromosomes: Comparative organization of closely related diploid and polyploid genomes	GENETICS	156	ISI000077401300029
Grivet, L; Dhont, A; Roques, D; Feldmann, P; Lanaud, C; Glaszmann, JC	RFLP mapping in cultivated sugarcane (<i>Saccharum</i> spp): Genome organization in a highly polyploid and aneuploid interspecific hybrid	GENETICS	131	ISIA1996TX45500029
Cordeiro, GM; Taylor, GO; Henry, RJ	Characterisation of microsatellite markers from sugarcane (<i>Saccharum</i> sp.), a highly polyploid species	PLANT SCIENCE	125	10.1016/S0168-9452(00)00208-9 and ISI000087823800006
ALJANABI, SM; HONEYCUTT, RJ; MCCLELLAND, M; SOBRAL, BWS	A GENETIC-LINKAGE MAP OF <i>SACCHARUM-SPONTANEUM</i> L SES 208	GENETICS	106	ISIA1993LP87300026
DASILVA, JAG; SORRELLS, ME; BURNQUIST, WL; TANKSLEY, SD	RFLP LINKAGE MAP AND GENOME ANALYSIS OF <i>SACCHARUM-SPONTANEUM</i>	GENOME	100	ISIA1993LU81200020
Ming, R; Liu, SC; Moore, PH; Irvine, JE; Paterson, AH	QTL analysis in a complex autopolyploid: Genetic control of sugar content in sugarcane	GENOME RESEARCH	90	10.1101/gr.198801 and ISI000172475600014
DHONT, A; RAO, PS; FELDMANN, P; GRIVET, L; ISLAMFARIDI, N; TAYLOR, P; GLASZMANN, JC	IDENTIFICATION AND CHARACTERIZATION OF SUGARCANE INTERGENERIC HYBRIDS, <i>SACCHARUM-OFFICINARUM</i> X <i>ERIANTHUS-ARUNDINACEUS</i> , WITH MOLECULAR MARKERS AND DNA IN-SITU HYBRIDIZATION	THEORETICAL AND APPLIED GENETICS	89	ISIA1995RL40300019
Dufour, P; Deu, M; Grivet, L; Dhont, A; Paulet, F; Bouet, A; Lanaud, C; Glaszmann, JC; Hamon, P	Construction of a composite sorghum genome map and comparison with sugarcane, a related complex polyploid	THEORETICAL AND APPLIED GENETICS	87	10.1007/s001220050430
Lima, ML A; Garcia, AAF; Oliveira, KM; Matsuoaka, S; Arizono, H; de Souza, CL; de Souza, AP	Analysis of genetic similarity detected by AFLP and coefficient of parentage among genotypes of sugar cane (<i>Saccharum</i> spp.)	THEORETICAL AND APPLIED GENETICS	86	10.1007/s001220200003

Ten most cited papers in the set of 344 articles obtained in the search for studies with molecular markers for sugarcane at the Thomson

Reuters Web of Science database

S8 Table. Institutions involved in the articles production.

Institution	Absolute Frequency	Relative Frequency	Organization	Absolute Frequency	Relative Frequency
INDIAN COUNCIL OF AGRICULTURAL RESEARCH ICAR	110	12.061	CHINESE ACADEMY OF AGRICULTURAL SCIENCES	5	0.548
UNITED STATES DEPARTMENT OF AGRICULTURE USDA	49	5.373	CTR TECNOL CANAVIEIRA CTC	5	0.548
CIRAD	46	5.044	AYUB AGR RES INST	4	0.439
COMMONWEALTH SCIENTIFIC INDUSTRIAL RESEARCH ORGANISATION CSIRO	34	3.728	CALIF INST BIOL RES	4	0.439
UNIVERSIDADE DE SAO PAULO	29	3.180	COUNCIL OF SCIENTIFIC INDUSTRIAL RESEARCH CSIR INDIA	4	0.439
UNIVERSIDADE ESTADUAL DE CAMPINAS	26	2.851	NATIONAL AGRICULTURE FOOD RESEARCH ORGANIZATION JAPAN	4	0.439
TEXAS A M UNIVERSITY COLLEGE STATION	26	2.851	SASRI	4	0.439
UNIVERSITY OF QUEENSLAND	22	2.412	UNIVERSIDADE ESTADUAL DE MARINGA	4	0.439
FUJIAN UNIVERSITY	21	2.303	UP COUNCIL SUGARCANE RES	4	0.439
BSES LTD	17	1.864	DIVERS ARRAYS TECHNOL PTY LTD	4	0.439
STATE UNIVERSITY SYSTEM OF FLORIDA	16	1.754	CTR AVANCAO PESQUISA TECNOL	4	0.439
UNIVERSITY OF FLORIDA	15	1.645	AGRONEGOCIO CANA IAC	3	0.329
LOUISIANA STATE UNIVERSITY	14	1.535	BHABHA ATOMIC RESEARCH CENTER	3	0.329
INSTITUTO AGRONOMICO DE CAMPINAS IAC	13	1.425	CENTRO DE INGENIERIA GENETICA Y BIOTECNOLOGIA	3	0.329
SOUTHERN CROSS UNIVERSITY	13	1.425	CLEMSON UNIVERSITY	3	0.329
YUNNAN AGRICULTURAL UNIVERSITY	11	1.206	CORNELL UNIVERSITY	3	0.329
UNIVERSIDADE ESTADUAL PAULISTA	9	0.987	ERCANE	3	0.329
VASANTDADA SUGAR INST	9	0.987	FASTERIS SA	3	0.329
S AFRICAN SUGAR ASSOC	9	0.987	FLORIDA SUGAR CANE LEAGUE INC	3	0.329
NUCLEAR INSTITUTE FOR AGRICULTURE BIOLOGY PAKISTAN	8	0.877	INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE INRA	3	0.329
GUANGXI UNIVERSITY	8	0.877	JAMIA HAMDARD UNIVERSITY	3	0.329
UNIVERSIDADE FEDERAL DO RIO DE JANEIRO	7	0.768	MAURITIUS SUGAR IND RES INST	3	0.329
UNIVERSITY OF GEORGIA	7	0.768	SHIVAJI UNIVERSITY	3	0.329
UNIVERSITY OF PUNJAB	7	0.768	TEAGASC	3	0.329
UNIVERSITY SYSTEM OF GEORGIA	7	0.768	UNIVERSIDADE FEDERAL DA INTEGRACAO LATINO AMERICANA	3	0.329
BUR SUGAR EXPT STN	6	0.658	UNIVERSIDADE FEDERAL DE VICOSA	3	0.329
CERF	6	0.658	UNIVERSITY OF AGRICULTURE FAISALABAD	3	0.329
GUANGZHOU SUGARCANE IND RES INST	6	0.658	UNIVERSITY OF ILLINOIS SYSTEM	3	0.329
NATL INST SUGARCANE RES	6	0.658	UNIVERSITY OF ILLINOIS URBANA	3	0.329
UNIVERSITY OF LA REUNION	6	0.658	CHAMPAIGN	3	0.329
COPERSUCAR	6	0.658	UNIVERSITY OF SINDH	3	0.329
CONSEJO NACIONAL DE INVESTIGACIONES CIENTIFICAS Y TECNICAS CONICET	5	0.548	FLORIDA CRYSTALS CORP	3	0.329
EMPRESA BRASILEIRA DE PESQUISA AGROPECUARIA EMBRAPA	5	0.548	SHAKARGANJ SUGAR RES INST	3	0.329
UNIVERSIDADE FEDERAL DE ALAGOAS	5	0.548	BP BIOFUELS NORTH AMER	3	0.329
UNIVERSIDADE FEDERAL DE SAO CARLOS	5	0.548	INST NACL INVEST AGR	3	0.329
UNIVERSITY OF KWAZULU NATAL	5	0.548	Intituiton that published only two papers		5.92105263
			Intituiton that published only one paper		16.6666667

Institutions identified for the WoS organization improved analysis.

S9 Table. WoS category classification.

Area - WoS	Absolute Frequency	Relative Frequency
PLANT SCIENCES	217	26.019
AGRONOMY	196	23.501
GENETICS HEREDITY	133	15.947
HORTICULTURE	92	11.031
BIOTECHNOLOGY APPLIED MICROBIOLOGY	59	7.074
BIOCHEMISTRY MOLECULAR BIOLOGY	53	6.355
AGRICULTURE MULTIDISCIPLINARY	17	2.038
MULTIDISCIPLINARY SCIENCES	16	1.918
FOOD SCIENCE TECHNOLOGY	13	1.559
BIOLOGY	5	0.600
BIOCHEMICAL RESEARCH METHODS	4	0.480
ENVIRONMENTAL SCIENCES	4	0.480
CELL BIOLOGY	3	0.360
DEVELOPMENTAL BIOLOGY	3	0.360
ENERGY FUELS	3	0.360
EVOLUTIONARY BIOLOGY	3	0.360
MEDICINE RESEARCH EXPERIMENTAL	3	0.360
ECOLOGY	2	0.240
AGRICULTURAL ENGINEERING	1	0.120
BIOPHYSICS	1	0.120
CHEMISTRY MULTIDISCIPLINARY	1	0.120
ENTOMOLOGY	1	0.120
MATHEMATICAL COMPUTATIONAL BIOLOGY	1	0.120
MICROBIOLOGY	1	0.120
SOIL SCIENCE	1	0.120
VETERINARY SCIENCES	1	0.120

Classification of the 344 articles performed by the WoS organizations analysis.

Considering that a specific article can be assigned to multiple research areas.

4 SELEÇÃO GENÔMICA AMPLA PARA CANA-DE-AÇÚCAR EM CONDIÇÕES DE ELEVADO DESEQUILÍBRIO DE LIGAÇÃO CONSIDERANDO EFEITOS DE ENDOGAMIA E ESTRUTURAÇÃO

4.1 RESUMO

A cana-de-açúcar é considerada uma das culturas mais produtivas e eficientes na conversão de energia devido ao seu metabolismo C4, à sua natureza perene e à sua alta adaptabilidade. A cultura assume grande importância no suprimento mundial de açúcar e etanol. Aumentos de produtividade e ganhos genéticos para a cana-de-açúcar são esperados com a incorporação de novos métodos de seleção, como a Seleção Genômica Ampla (GS). Ainda que resultados promissores tenham sido obtidos é necessário que sejam avaliados os efeitos da inserção de diferentes fatores nos modelos estatísticos utilizados. Neste estudo, uma população, constituída por dois conjuntos de indivíduos provenientes de um cruzamento biparental e de uma geração de autofecundação, foi avaliada para fins de construção de modelos de GS. Considera-se que o presente trabalho represente uma prova de conceito para a aplicação da GS em cana-de-açúcar já que a população avaliada apresentou um alto desequilíbrio de ligação (LD) e uma forte estrutura intrapopulacional. Foram analisados seis caracteres relacionados à produtividade de biomassa e teor de açúcar, fazendo uso de seis diferentes metodologias de análise. Tanto a regressão estratificada quanto a correção para os efeitos de endogamia foram aplicados na tentativa de se realizar a correção dos efeitos espúrios decorrentes da presença de estruturação genética. Estimativas de acurácias acima de 0.5 foram obtidas para a maior parte dos caracteres avaliados. As acurácias estimadas podem ser consideradas como valores potenciais a serem obtidos, devido ao alto LD presente na população. Nossos resultados mostram que escolha dos modelos não influencia de maneira determinante os resultados obtidos para a cultura. Os resultados obtidos evidenciam um forte potencial de aplicação da estratégia de GS em cana-de-açúcar.

4.2 INTRODUÇÃO

Saccharum é considerado um dos gêneros geneticamente mais complexos entre as plantas cultivadas, possuindo genoma extenso e grande número de cromossomos (D'Hont et al., 1994; D'Hont et al., 1996; Piperidis et al., 2010; Zhang et al., 2012). A complexidade genômica da espécie é acentuada pela coexistência de dois genomas nas cultivares comercialmente utilizadas, devido às hibridizações realizadas no decorrer do melhoramento genético da cultura (Cuadrado et al., 2004; D'Hont, 2005; Piperidis et al., 2010). A cana-de-açúcar é uma das *commodities* agrícolas mais importantes do mundo. É cultivada em regiões tropicais e subtropicais (FAO, 2018). A cana-de-açúcar assume papel importante não só pelo suprimento de açúcar mundial, mas também como fonte de álcool (Menossi et al., 2008), fator que vêm impulsionando a pesquisa para a cultura devido ao impacto econômico das energias sustentáveis (Cheavegatti-Gianotto et al., 2011).

O advento das metodologias de sequenciamento de DNA de nova geração, em paralelo ao desenvolvimento de ferramentas de análise genética e estatística, tem proporcionado ganhos genéticos significativos para diversas espécies (Desta & Ortiz, 2014). Neste contexto, um dos métodos mais aclamados é o método de Seleção Genômica Ampla (GS). Esta metodologia faz uso tanto de fenotipagem quanto de genotipagem de indivíduos e se baseia no desenvolvimento de um modelo de predição que, ao ser utilizado para a seleção, permite a predição dos valores genéticos de cada um dos indivíduos com base apenas no seu genótipo.

A GS não tem como objetivo indicar locos individuais para a seleção e sim capitalizar, simultaneamente, os efeitos de milhares de locos dispersos no genoma, de maneira a capturar seus efeitos conjuntos. O objetivo neste caso é explicar toda ou grande parte da variação genética associada a determinado caráter, ou seja, detectar a maior parte dos polimorfismos causais envolvidos na variação genética de um caráter quantitativo (Meuwissen et al., 2001; Resende et al., 2010; Endelman et al., 2011; Lorenz, 2013). Neste sentido, a GS permite a inferência, com determinada acurácia, do efeito de vários locos responsáveis pela variação fenotípica de determinado caractere, com base no comportamento dos marcadores utilizados. Trata-se, portanto, de uma expansão dos fundamentos teorizados por Thoday (1961).

Os modelos de GS podem ser obtidos pelo uso de diversas metodologias, considerando diferentes fatores na análise (De Los Campos et al., 2013; Muñoz et al.,

2014). O método se fundamenta na premissa de existência de desequilíbrio de ligação (LD) entre os marcadores moleculares utilizados e os locos que participam do controle genético do caráter quantitativo de interesse (Meuwissen et al., 2001; Zhang et al., 2011). Além disso, o desenvolvimento de modelos estatísticos vem apontando que além do LD, o parentesco apresenta alta influência nas acurácias e na perpetuação dos modelos ao longo das gerações (Habier et al., 2007; Gianola et al., 2009; Habier et al., 2010). Apenas um estudo avaliou o potencial de aplicação da GS para cana-de-açúcar (Gouy et al., 2013). Entretanto este potencial ainda não foi demonstrado para populações de melhoramento da cultura.

No sentido de se investigar o potencial de aplicação da seleção genômica ampla para cana-de-açúcar, considerando-se diferentes fatores para a predição dos valores genéticos genômicos (VGG), o objetivo deste trabalho foi comparar além da eficácia da aplicação de diferentes modelos de GS para cana-de-açúcar, o uso de abordagens metodológicas destinadas a realizar as correções para os efeitos de estruturação genética intrapopulacional e de níveis variáveis de endogamia existentes nas populações de melhoramento. Foram avaliados seis caracteres de interesse agrônomo relacionados à produtividade de biomassa e teor de açúcar e obtidas acurácias de predição para seis diferentes modelos de GS.

4.3 MATERIAL E MÉTODOS

4.3.1. Material vegetal

Os dados fenotípicos e genotípicos utilizados neste trabalho foram obtidos de uma população de melhoramento de cana-de-açúcar da Rede Interuniversitária para o Desenvolvimento do Setor Sucroenergético (RIDESA). Um total de 172 indivíduos compõem esta população, incluindo 91 indivíduos resultantes do cruzamento entre os clones elite RB97327 e RB72454 e 81 indivíduos obtidos pela autofecundação do clone RB97327, além de duas testemunhas experimentais (clones RB99395 e RB98710).

Os cruzamentos para obtenção destes clones foram realizados na Estação Experimental da RIDESA, localizada na Serra do Ouro, em Murici (AL). As mudas, geradas a partir de sementes, foram alocadas em casa de vegetação na Escola de Agronomia, da Universidade Federal de Goiás (UFG) e, posteriormente, transplantadas para uma área experimental da Usina Centro-Álcool, localizada no município de Inhumas

(GO) (16°20'50"S, 49°29'2"W).

O experimento de campo foi realizado seguindo o delineamento de blocos aumentados de Federer (Federer, 1956) e foram adotadas práticas de cultivo padrão para a cultura. A área foi dividida em nove linhas de 67,5m, subdivididas em 135 blocos, dos quais 40 foram aqui considerados por incluírem os indivíduos genotipados. Cada bloco foi composto por 8 parcelas (touceiras) das quais duas eram testemunhas comuns a todos os blocos, aleatoriamente distribuídas em cada bloco. Os espaçamentos utilizados entre parcelas dentro de linhas e entre linhas foi de 0,5 m e 1,5 m, respectivamente.

4.3.2. Material genético e genotipagem

O DNA genômico foi obtido pelo uso do protocolo de Aljanabi et al. (1999) modificado para extração em microtubos de 1,5mL. Os dados de genotipagem foram obtidos pela metodologia DArT, realizada pela empresa Diversity Arrays Technology Pty Ltd (DArT P/L, Yarralumla, Austrália), baseada na tecnologia de hibridização por microarranjos proposta por Jaccoud et al. (2001), gerando dados para 15360 locos, dos quais 894 locos foram aqui considerados por apresentarem dados faltantes inferior a 10% e padrões de segregação 1:1 ou 3:1, em um dos dois conjuntos de indivíduos analisados.

4.3.3. Caracterização fenotípica

Os caracteres morfo-agronômicos analisados foram: produtividade, expressa em toneladas de colmos por hectare (TSH), diâmetro médio de colmos (MSD), comprimento médio de colmos (MSL), número médio de entrenós (MNI) e número de colmos por parcela (NSP). Com base nos dados obtidos para o teor de sólidos solúveis (°Brix) de cada parcela, foram estimadas as produtividades de sólidos solúveis, expressas em toneladas de sólidos solúveis por hectare (TBH). Todos os caracteres fenotípicos foram obtidos no Centro de Excelência em Melhoramento Genético de Cana-de-Açúcar no Cerrado, da UFG. Os valores associados à MSD, MSL, MNI e BRIX resultaram da média de avaliação de cinco colmos retirados de maneira aleatória de cada parcela. Para TSH e NSP, todos os colmos da parcela foram utilizados para mensuração.

O diâmetro de cada colmo foi avaliado no centro do entrenó localizado na porção mediana do colmo. O comprimento de cada colmo foi obtido com base na medida tomada a partir da base do colmo até o último entrenó. O número de colmos por parcela e o número de entrenós foram obtidos por contagem direta. Para a mensuração do teor de

sólidos solúveis totais um refratômetro manual foi utilizado. O caldo utilizado para esta análise foi retirado da porção mediana dos colmos, para obtenção do valor de BRIX, utilizado para o cômputo do valor de TBH.

4.3.4. Análise estatística dos dados fenotípicos

Após o descarte de indivíduos provenientes de pólen contaminante, 159 indivíduos foram utilizados nas análises (157 genótipos oriundos de cruzamentos controlados + duas testemunhas). Modelos lineares mistos foram aplicados para a predição dos valores genotípicos (eBLUPs), permitindo também a estimação dos componentes de variância e de herdabilidade para cada caráter. Este processo foi realizado pelo uso do pacote *lme4* v.1.1-8 (Bates et al., 2015), seguindo o modelo (1):

$$y = X_1\beta_1 + X_2\beta_2 + Zu + e \quad (1)$$

Em que y é o vetor de fenótipos observados, β_1 é o vetor dos efeitos fixos de blocos, associados à matriz de incidência X_1 , β_2 é o vetor de efeitos fixos de testemunhas, associados à matriz de incidência X_2 , u é o vetor de efeitos aleatórios, composto pelo efeitos de populações e genótipos dentro de populações, sendo os efeitos de populações considerados como normalmente distribuídos, com média zero e variância dada por $I\sigma_p^2$. Os efeitos de genótipos dentro de populações foram considerados normalmente distribuídos, com média zero e variância dada por $I\sigma_g^2$. O vetor e é o vetor aleatório de resíduos do modelo, normalmente distribuídos com média zero e variância $I\sigma_e^2$, sendo I a matriz identidade.

As estimativas de componentes de variância foram obtidas pelo método REML (*Restricted Maximum Likelihood*). Para cada caráter, os coeficientes de herdabilidade, no sentido amplo, foram estimados de duas formas: a primeira pela razão entre a variância associada ao efeito de genótipos (σ_g^2) e a variância fenotípica total ($\sigma_p^2 + \sigma_g^2 + \sigma_e^2$) e a segunda considerando a razão entre a soma da variância associada aos efeitos de genótipos (σ_g^2) e àquela devida aos efeitos de populações (σ_p^2) e a variância fenotípica total ($\sigma_p^2 + \sigma_g^2 + \sigma_e^2$). A comparação entre estes valores permitiu a realização de inferências sobre os efeitos da estruturação genética sobre estes parâmetros.

4.3.5. Análise de estruturação genética intrapopulacional

A fim de se comprovar a existência de estruturação genética intrapopulacional

a matriz $(1 - G)$, sendo G a matriz de parentesco obtida pelo estimador de VanRaden (2008), obtida a partir dos dados de 894 locos DArT, foi submetida a uma análise de coordenadas principais, pelo uso do pacote *ape* (Paradis et al., 2004). Segundo VanRaden (2008), os valores da matriz G são análogos aos coeficientes de parentesco de Wright (1922). Além da análise de coordenadas principais, os dados foram submetidos a uma análise de agrupamento hierárquico (UPGMA), permitindo a construção de um dendrograma, fazendo uso de funções e pacotes implementados no ambiente R (R Core Team, 2018): *stats* (R Core Team, 2018), *ggdendro* (de Vries & Ripley, 2016), *plyr* (Wickham, 2011) e *zoo* (Zeileis & Grothendieck, 2005).

4.3.6. Desequilíbrio de ligação

Devido às dificuldades de determinação das frequências alélicas e haplotípicas inerentes à complexidade do genoma de cana-de-açúcar e à natureza dominante dos marcadores utilizados, as medidas usuais de desequilíbrio de ligação (LD) não puderam ser obtidas. Uma alternativa para acessar este valor é o uso de testes de significância, no qual a verificação estatística confirma a existência, ou não, da associação significativa entre os alelos dos diferentes locos permitindo inferências sobre o LD (Slatkin, 1994). Neste sentido, para a obtenção dos valores de LD o teste Exato de Fisher foi utilizado. Para cada par de marcas uma tabela de contingência 2x2 foi construída e a probabilidade (P) associada foi obtida, seguindo a metodologia descrita por Raboin et al. (2008). Assim, quanto menor a probabilidade obtida, menor a independência entre os alelos e maior o LD entre eles.

A avaliação da significância estatística das associações par-a-par foi realizada de duas maneiras. A primeira estabelecendo um nível crítico com o uso do FDR (False Discovery Rate), em que foram declarados significativos os valores de associação par-a-par com p -valores inferiores a 0,05. O segundo se deu pelo estabelecimento do limite empírico de $p = 1,6 \times 10^{-5}$ ($-\log p = 4,79588$), o mesmo utilizado por Raboin et al. (2008), valor para o qual estes autores constatarem a presença de associações verdadeiras (dentro do mesmo grupo de hom(e)ologia), descartando as associações espúrias (devidas a outros fatores que não a ligação física entre os locos).

Para a avaliação do LD as duas subpopulações foram consideradas separadamente. Além disso, estas subpopulações foram subdivididas em grupos, de acordo com a configuração observada dos locos nos parentais (Tabela 1). Estes grupos foram

identificados como single dose (com presença do marcador em apenas um dos parentais) e *double single-dose* (com presença do marcador nos dois parentais). Para a população oriunda de autofecundação apenas a configuração *double single-dose* foi analisada. Foram utilizados dados de presença e ausência para 850 locos DArT nas análises da subpopulação derivada de cruzamento biparental e 470 locos DArT para a população obtida por autofecundação.

Tabela 1. Subpopulações e grupos dentro de subpopulações utilizados para a análise do LD utilizando o modelo de Raboin et al. (2008).

População	Grupo	Segregação esperada
Cruzamento biparental	1x0	1:1
	0x1	1:1
	1x1	3:1
Autofecundação	1x1	3:1

4.3.7. SELEÇÃO GENÔMICA AMPLA

Foram utilizadas seis metodologias para a obtenção dos modelos de GS: *ridge regression* (RR-BLUP) (Hoerl & Kennard, 1970), BLUP genômico (GBLUP) (VanRaden, 2008), Bayes A, Bayes B (Meuwissen et al., 2001), Bayes C (Habier et al., 2011) e Bayesian LASSO (Park & Casella, 2008). Os modelos foram construídos com o uso do pacote BGLR (Pérez & de los Campos, 2014), implementado na plataforma R (R Core Team, 2018).

Os dados genotípicos faltantes foram substituídos pelas frequências médias de presenças estimadas para cada loco, pois com dados faltantes não é possível se obter uma matriz de genótipos sobre a qual o sistema de equações matriciais seja resolvido, inviabilizando a obtenção dos parâmetros necessários à construção do modelo preditivo. Além disso, locos que apresentaram frequências menores que 0,05 ou maiores que 0,95, ou com quantidade de dados perdidos acima de 10% foram desconsiderados na análise. Desta maneira, foram utilizados dados de 894 locos DArT, compondo a matriz de genótipos do modelo para 157 indivíduos da população original. Os seis caracteres fenotípicos foram analisados independentemente.

As análises foram realizadas considerando o conjunto total de indivíduos como uma única população (2) e também fazendo uso das matrizes relativas aos efeitos de marcadores de maneira estratificada (3), como tentativa para a correção do efeito de estruturação intrapopulacional gerado pela união das duas populações (de los Campos et

al., comunicação pessoal).

i. Regressão homogênea:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \beta_0 + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} \quad (2)$$

ii. Regressão estratificada

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 \\ 0 \end{bmatrix} \beta_1 + \begin{bmatrix} 0 \\ X_2 \end{bmatrix} \beta_2 + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} \quad (3)$$

Em que as matrizes de incidência contendo as informações de marcadores (X_1 e X_2) interagem de maneira conjunta (regressão homogênea) ou separadamente (regressão estratificada) para a predição dos efeitos dos marcadores ($\beta_0, \beta_1, \beta_2$), nas análises desenvolvidas para a obtenção dos valores genéticos genômicos (y_1, y_2). A correção para o efeito de endogamia foi realizada pela inserção dos coeficientes de endogamia como uma covariável no modelo. O efeito de endogamia foi obtido pelo uso dos valores da diagonal da matriz G menos um, ou seja ($G_{jj} - 1$) (VanRaden, 2008).

As seguintes combinações de modelagem foram aplicadas: i) Análise homogênea; ii) Análise homogênea com correção para os efeitos de endogamia; iii) Análise estratificada; iv) Análise estratificada com correção para os efeitos de endogamia. Foi realizada também a análise das estatísticas descritivas dos eBLUPs e dos VGG obtidos com cada um dos modelos de GS.

4.3.8. Acurácia preditiva dos modelos

A validação cruzada foi utilizada para a avaliação da acurácia preditiva dos modelos de GS obtidos. Para isso fez-se uso de *jack-knife* considerando-se a metodologia *10-fold-validation*. Assim, o conjunto de dados foi dividido aleatoriamente em 10 partes iguais e a predição dos VGG foi realizada para um dos grupos a partir dos valores gerados com o conjunto de treinamento composto pelos demais 9 grupos. As distribuições *a posteriori* para cada um dos parâmetros de interesse foram amostradas utilizando-se o algoritmo MCMC com 1,000,000 de passos, com descarte (*burnin*) de 100,000 passos, e um valor de *thinning* igual a cinco (*default*), evitando a autocorrelação entre os pontos amostrados. A avaliação das Cadeias de Markov produzidas foi realizada graficamente para os componentes de variância genética e residual.

Os valores de acurácia preditiva foram obtidos a partir da correlação entre os valores genéticos genômicos preditos por cada modelo e os eBLUPs obtidos para cada um dos indivíduos pelo uso de modelos lineares mistos.

4.3.9. Análise comparativa entre os modelos

Elaborado por Spiegelhalter et al. (2002), o DIC (Deviance Information Criterion) foi construído relacionando uma medida de adequação do modelo (D) à complexidade assumida por esse modelo (pD), sendo essa complexidade função do número efetivo de parâmetros utilizados. O DIC pode ser utilizado para se comparar modelos, sendo que quanto menor o valor de DIC, melhor a relação entre ajuste e complexidade, ou seja, melhor o modelo obtido.

Os valores de DIC foram calculados pela média dos resultados obtidos por 30 análises independentes realizadas para cada uma das abordagens realizadas neste trabalho, para cada um dos caracteres avaliados. A complexidade dos modelos dada pelo número de parâmetros efetivos na análise (pD) também foi considerada como critério de comparação entre os modelos. Os valores de DIC e pD foram obtidos com o uso do pacote BGLR.

4.4 RESULTADOS

4.4.1. Dados fenotípicos

Em termos absolutos, as menores variâncias foram encontradas para MSL. A herdabilidade apresentada para esse caráter foi alta. O caráter MNI apresentou a menor herdabilidade (0,238). Todos os demais caracteres analisados apresentaram herdabilidade superior à 0,600. Uma grande proporção das estimativas de herdabilidade obtidas para MSL, MNI e TBH são atribuíveis à variância genotípica dentro de subpopulações, sendo que para os dois últimos caracteres praticamente a totalidade da herdabilidade estimada (>98%) foi ditada pela variância dos efeitos de genótipos dentro de subpopulações. Para os demais caracteres (MSD e TSH), uma grande proporção das estimativas de herdabilidade obtidas se deve à variância dos efeitos de populações, sendo estas proporções superiores a 80%. O caráter NSP mostrou uma influência equilibrada dos efeitos genotípicos dentro e entre subpopulações (Tabela 2). Os histogramas obtidos para todos os caracteres analisados indicam distribuição próxima à Gaussiana (Figura 1), bem como os resíduos dos modelos após a análise de cada caráter pelo uso do pacote *lme4* (Figura 2).

Tabela 2. Estimativas de componentes de variância e de coeficientes de herdabilidade obtidas pela análise de modelos lineares mistos para os seis caracteres fenotípicos avaliados, considerando-se os componentes genéticos dentro de subpopulações (G), entre subpopulações (P) e residuais (E).

Caráter		Variâncias Estimadas	Herdabilidade (Genética+Populacional)	Herdabilidade (Genética)	Herdabilidade (Populacional)
MSL	G	0.0721	0.7184	0.3167 (44.08%)	0.4017 (55.92%)
	P	0.0914			
	E	0.0641			
MSD	G	4.0931	0.6947	0.1364 (19.03%)	0.5583 (80.37%)
	P	16.7581			
	E	9.1647			
NSP	G	9.223	0.8763	0.4679 (53.39%)	0.4085 (46.61%)
	P	8.0525			
	E	2.4379			
MNI	G	8.022	0.2384	0.2384 (100%)	0 (0%)
	P	0			
	E	25.6261			
TSH	G	500.0876	0.6582	0.1286 (19.54%)	0.5296 (80.46%)
	P	2058.5617			
	E	1328.5894			
TBH	G	737.5937	0.6504	0.6419 (98.7%)	0.0085 (1.3%)
	P	9.7119			
	E	401.7063			

4.4.2. Análise de estruturação genética intrapopulacional

A análise baseada na utilização dos dados de indivíduos oriundos do cruzamento biparental e de autofecundação como uma única população revelou, como esperado, um alto grau de estruturação genética. O dendrograma obtido mostra a divisão clara do conjunto de dados em dois grupos de indivíduos, correspondentes a estas subpopulações (Figura 3). A Análise de Coordenadas Principais corrobora este resultado, sendo que o primeiro componente os separa completamente, explicando 33,02% da variação (Figura 3). Verifica-se, portanto, a disjunção dos dois conjuntos de indivíduos resultante da organização diferencial da diversidade genética nos dois tipos de cruzamentos utilizados.

4.4.3. Desequilíbrio de ligação

A partir da aplicação da metodologia de Raboin et al. (2008) foram obtidos gráficos relacionando as estimativas de LD ($-\log p$) à distância genética estimada entre os locos (cM). A significância do desequilíbrio em si e sua extensão no genoma foi analisada, considerando-se qual seria o tamanho dos blocos de ligação nas populações em estudo

(Figura 4).

Considerando-se a análise da população de cruzamento biparental, a quantidade de associações significativas representou em média 80% do total de comparações realizadas dentro de cada um dos grupos. Para a análise do desequilíbrio no grupo 1x0 desta população foram realizadas 23.005 testes de associação par-a-par. Deste total 18.651 se apresentaram significativos ou seja aproximadamente 81% das comparações geraram resultado significativo, mesmo após a correção por FDR. Já para o segundo grupo desta população (0x1), foram realizadas 72.010 comparações par-a-par entre os locos e aproximadamente 80% delas (57.520) envolviam locos com associações significativas. Para os indivíduos do grupo 1x1, 32.385 comparações foram realizadas, sendo 17.931 significativas (55,4%).

Além da correção por FDR, foi considerado como nível crítico de significância do LD o limite empírico de $-\log p = 4,79588$ (equivalente a um valor de $r^2 = 0,2$) indicado nos gráficos como uma linha paralela à abcissa. Este valor é indicado por Raboin et al. (2008) como o limite que contém as associações que se devem à ligação física, ou seja, que contém os locos que estão em um mesmo grupo de hom(e)ologia. Tomando este limite como referência, o desequilíbrio de ligação encontrado para os grupos da população de cruzamento biparental se estendem até uma distância entre 25 e 30 cM.

Considerando-se os indivíduos resultantes de autofecundação da variedade RB97327 (subgrupo 1x1), a relação entre o desequilíbrio de ligação e a distância genética (cM) existente foi similar à observada na população do cruzamento biparental. A significância considerando-se a correção FDR aponta que das 65.535 associações par-a-par avaliadas, 47,1% mostraram-se significativas (30.869). O uso deste limite empírico também aponta para uma extensão do LD de cerca de 25 cM. Os padrões de decaimento do LD foram obtidos para cada grupo de subpopulações analisadas e, a partir do limite empírico obtido por Raboin et al. (2008), foi traçada uma linha paralela à abcissa, indicando as associações declaradas como significativas.

4.4.4. Modelos de seleção genômica ampla

4.4.4.1. Acurácia preditiva dos modelos

Apesar das metodologias diferirem quanto ao modo com que os efeitos genéticos associados aos marcadores são modelados, as estimativas de acurácia preditiva obtidas por cada uma das seis metodologias utilizadas foram similares, tanto quando

realizada a análise estratificada, quanto quando considerados os efeitos de endogamia (Figura 5, Tabela 3). Em geral, a acurácia preditiva dos modelos atingiu valores superiores a 0.5 para a maioria dos modelos, sendo que valores não significativos foram obtidos para MNI e TBH (valores inferiores a 0,1 ou não diferindo significativamente de 0,0). Os maiores valores de acurácia preditiva foram obtidos para TSH ($\cong 0,95$), MSD ($\cong 0,94$) e MSL ($\cong 0,73$) (Figura 5, Tabela 3).

Gráficos evidenciando a relação entre os VGG preditos para os indivíduos e os valores de eBLUPs foram construídos (Figuras 6 a 11). Para exemplificar os valores de acurácia obtidos foram utilizados os gráficos de dispersão obtidos para TBH e TSH, caracteres que apresentaram as menores e as maiores acurácias, respectivamente (Figuras 10 e 11). Os efeitos de estruturação puderam ser verificados nestes gráficos pela formação de dois grupos distintos de indivíduos. Este efeito permaneceu evidente mesmo quando realizada a correção pelo modelo estratificado de análise, ou pela correção pela modelagem dos efeitos de endogamia, sendo este efeito tanto mais evidente quanto maior a estimativa de acurácia obtida (Figura 5). As estimativas de acurácia preditiva também mostraram uma forte relação com a porcentagem da variação genética explicada pelo fator subpopulação (Tabela 2). Quanto maior a variância explicada por este fator, consequentemente maior a porção da herdabilidade total por ele explicada, mais definidos os grupos se apresentaram, mostrando o forte efeito deste fator sobre as estimativas de acurácias obtidas. Para melhor interpretação, os valores dos gráficos foram coloridos de acordo com a subpopulação de origem.

4.4.4.2. Valores de DIC dos modelos

O ajuste dos diferentes modelos em cada uma das abordagens de análise foi avaliado pelo uso dos valores de DIC, ilustrados pelo uso de boxplots (Figuras 12 e 13). As médias dos valores de DIC e as medidas de complexidade dos modelos foram obtidas, permitindo a comparação entre modelos e entre abordagens para cada um dos caracteres sob análise (Tabela 4). As diferenças apresentadas entre os DIC foram consideradas para fins de comparação. Diferenças maiores que 5 unidades mostram diferenças substanciais entre os modelos (<http://www.mrc-bsu.cam.ac.uk/software/bugs/the-bugs-project-dic/>).

A estratégia com melhor ajuste foi a regressão estratificada, apresentando os menores valores de DIC, para todos os caracteres analisados, apesar deste modelo apresentar maior complexidade. Considerando os modelos dentro desta estratégia, os

modelos Bayes B e Bayes A apresentaram os menores valores de DIC para a maioria dos caracteres, apenas para TSH e MNI o modelo RRBLUP, apresentou melhor relação ajuste-complexidade. Neste contexto, na análise realizada para este conjunto de dados, para todos os seis caracteres recomenda-se a estratégia de análise estratificada unida aos seguintes modelos: i) Para MNI e TSH pode-se recomendar o RRBLUP ; ii) Para MSD e MSL recomenda-se o modelo Bayes B e iii) Para NSP e TBH recomenda-se a aplicação do modelo Bayes A.

Tabela 3. Acurácias obtidas para os seis modelos de seleção genômica ampla testados para cana-de-açúcar em condições de alto desequilíbrio de ligação, para os seis caracteres analisados, considerando a inserção ou não dos efeitos de endogamia e da estratificação da análise.

Modelo	MNI			MSD				
	Homogênea	Homogênea + Endogamia	Estratificada + Endogamia	Homogênea	Homogênea + Endogamia	Estratificada + Endogamia		
GBLUP	-0.089	-0.0263	-0.1099	-0.0721	0.9189	0.9217	0.919	0.9177
RRBLUP	-0.1429	-0.1162	-0.145	-0.1128	0.919	0.9173	0.9187	0.9176
Bayes A	-0.1933	-0.2137	-0.105	-0.2241	0.9244	0.9178	0.9341	0.9213
Bayes B	-0.0872	-0.1767	-0.2057	-0.106	0.9207	0.92	0.9357	0.9272
Bayes C	-0.0972	-0.1188	-0.0073	-0.0642	0.9216	0.9149	0.922	0.9251
Bayes LASSO	-0.0505	-0.136	0.1549	-0.1361	0.9229	0.9176	0.9275	0.9138
GBLUP	MSL			NSP				
	0.6969	0.7239	0.7168	0.7057	0.5366	0.5274	0.5368	0.54
RRBLUP	0.7044	0.7258	0.6897	0.6842	0.5511	0.5513	0.502	0.5268
Bayes A	0.7223	0.7173	0.7099	0.7092	0.5094	0.5313	0.5079	0.4884
Bayes B	0.723	0.718	0.6983	0.7177	0.5189	0.528	0.5357	0.5055
Bayes C	0.7061	0.7201	0.7106	0.7143	0.5314	0.5273	0.5244	0.526
Bayes LASSO	0.7212	0.7252	0.7366	0.7215	0.5443	0.5337	0.7118	0.5417
GBLUP	TBH			TSH				
	0.0635	0.0387	0.0232	-0.0077	0.9343	0.9321	0.9341	0.9329
RRBLUP	0.017	0.0737	0.01	0.0307	0.9334	0.9326	0.932	0.9315
Bayes A	-0.0767	0.0834	-0.0034	0.06	0.9359	0.931	0.9448	0.929
Bayes B	-0.0176	0.1086	0.0385	0.0951	0.9377	0.9307	0.9451	0.9342
Bayes C	0.0481	0.0633	-0.016	0.0261	0.9366	0.9345	0.9392	0.9366
Bayes LASSO	0.0045	0.0411	0.0546	0.0569	0.9357	0.9323	0.9563	0.9789

Tabela 4. Estatísticas descritivas dos eBLUPs e VGG obtidos para os seis modelos utilizados, considerando a estratégia de análise estratificada (selecionada pelo DIC).

Caráter	Descritivas						Modelo	Caráter	Descritivas					
	Média	IC -Média (95%)	Variancia	Desvio padrão	Amplitude	Mediana			Média	IC -Média (95%)	Variancia	Desvio padrão	Amplitude	Mediana
MNI	31.378	31.177 - 31.579	1.612	1.27	6.531	31.303	eBLUP	MSD	22.872	22.389 - 23.355	9.335	3.055	10.666	24.12
	31.384	31.348 - 31.42	0.052	0.228	1.469	31.372	GBLUP		22.862	22.428 - 23.295	7.559	2.749	7.837	24.519
	31.386	31.333 - 31.439	0.112	0.335	2.326	31.379	RRBLUP		22.918	22.482 - 23.354	7.649	2.766	7.597	24.874
	31.378	31.343 - 31.413	0.049	0.222	1.301	31.371	Bayes A		22.881	22.434 - 23.329	8.063	2.839	6.621	25.275
	31.376	31.335 - 31.417	0.069	0.262	2.143	31.374	Bayes B		22.905	22.456 - 23.353	8.105	2.847	7.166	25.258
	31.384	31.357 - 31.41	0.028	0.166	1.025	31.385	Bayes C		22.863	22.429 - 23.298	7.601	2.757	7.399	24.5
MSL	31.319	31.274 - 31.363	0.007	0.083	0.351	31.32	Bayes LASSO	NSP	23.017	22.323 - 23.711	7.597	2.756	9.252	25.129
	1.842	1.798 - 1.887	0.078	0.28	1.372	1.89	eBLUP		4.832	4.256 - 5.407	10.223	3.197	15.365	4.175
	1.839	1.807 - 1.871	0.041	0.203	0.687	1.912	GBLUP		4.753	4.442 - 5.064	3.896	1.974	8.487	5.01
	1.844	1.812 - 1.877	0.042	0.204	0.741	1.943	RRBLUP		4.801	4.454 - 5.149	4.864	2.205	8.87	5.129
	1.847	1.814 - 1.879	0.043	0.207	0.769	1.983	Bayes A		4.712	4.37 - 5.054	4.703	2.169	9.039	5.115
	1.848	1.816 - 1.881	0.043	0.206	0.798	1.981	Bayes B		4.767	4.417 - 5.118	4.946	2.224	9.426	5.242
TBH	1.842	1.81 - 1.874	0.041	0.202	0.701	1.931	Bayes C	HSH	4.739	4.422 - 5.055	4.027	2.007	8.297	5.05
	1.85	1.812 - 1.887	0.04	0.2	0.692	1.991	Bayes LASSO		4.548	3.705 - 5.391	5.286	2.299	7.271	5.025
	264.654	261.132 - 268.176	415.265	20.378	110.753	268.436	eBLUP		56.552	51.118 - 61.986	1126.648	33.566	112.545	69.885
	264.518	263.741 - 265.296	24.334	4.933	21.797	264.179	GBLUP		56.35	51.54 - 61.161	931.198	30.516	85.34	73.855
	264.514	263.49 - 265.538	42.188	6.495	30.207	263.442	RRBLUP		56.796	51.898 - 61.695	965.584	31.074	91.388	76.305
	263.834	262.827 - 264.84	40.741	6.383	30.921	263.571	Bayes A		58.512	52.939 - 64.085	1249.565	35.349	173.259	81.692
	264.369	263.437 - 265.301	34.972	5.914	30.474	264.23	Bayes B	57.984	52.752 - 63.216	1101.345	33.187	124.652	81.672	
	264.576	263.854 - 265.298	20.998	4.582	20.823	264.161	Bayes C	56.296	51.445 - 61.147	946.811	30.77	83.726	73.374	
	265.133	264.06 - 266.205	13.643	3.694	18.039	265.631	Bayes LASSO	49.402	32.194 - 66.609	1042.865	32.293	75.309	26.306	

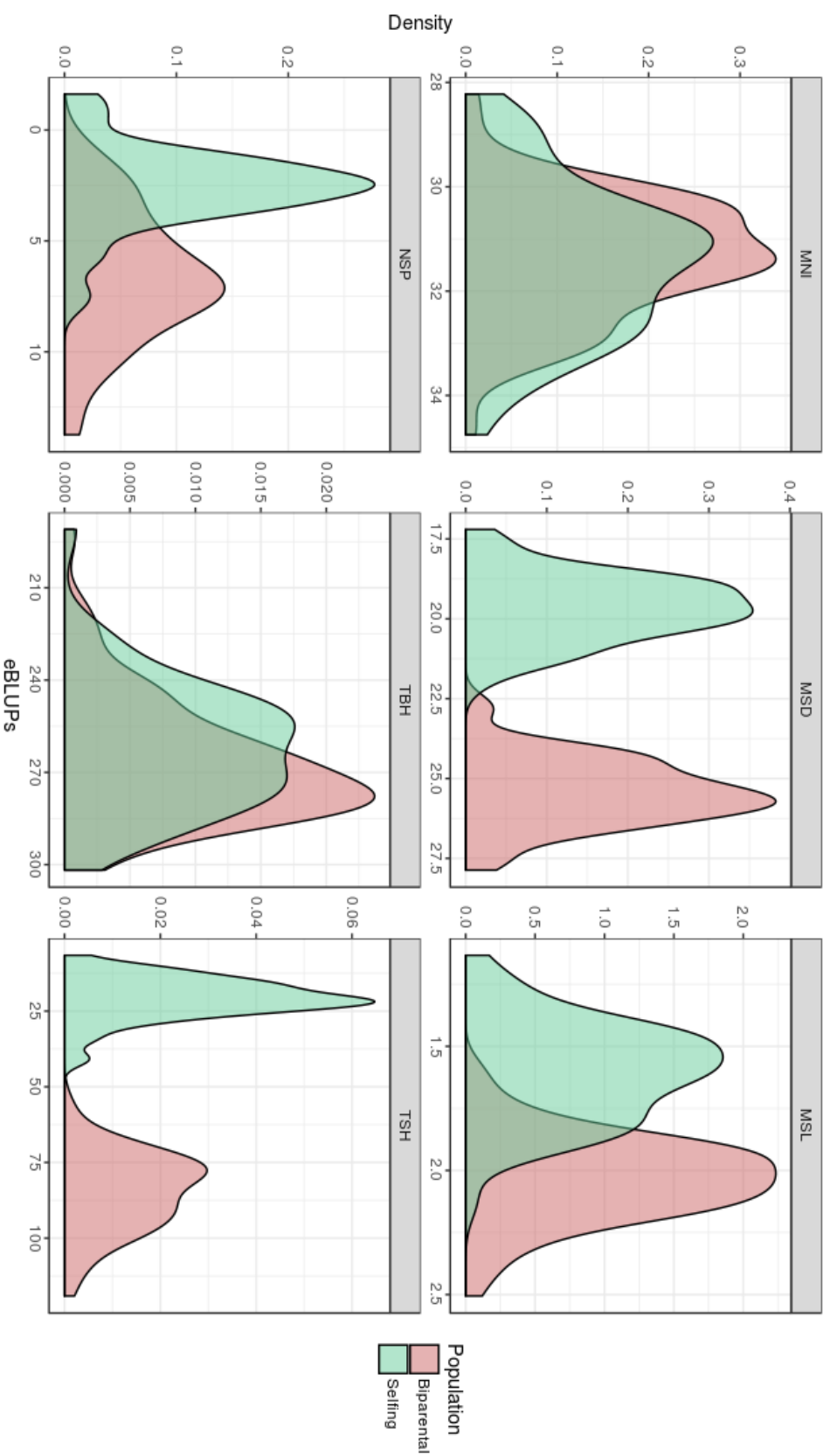


Figura 1. Distribuição de densidade dos eBLUPs obtidos na análise de modelos mistos realizada com o uso do pacote lme4 v.1.1-8 (Bates et al., 2015) para 159 indivíduos referentes aos seis caracteres analisados.

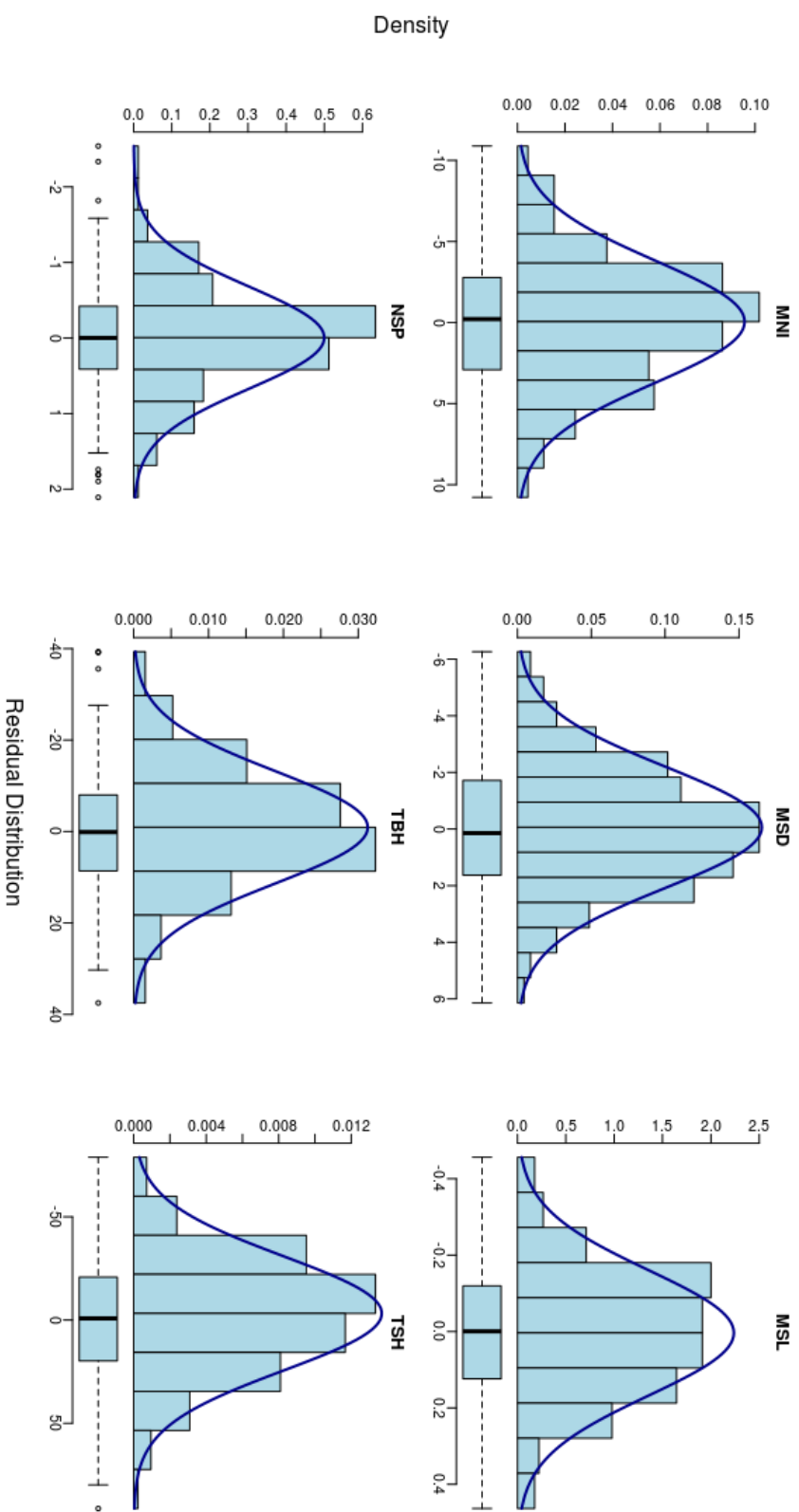


Figura 2. Distribuição de densidade, histograma e boxplot dos resíduos obtidos com a análise dos eBLUPs com o uso do pacote lme4 v.1.1-8 (Bates et al., 2015) para 159 indivíduos referentes aos seis caracteres analisados.

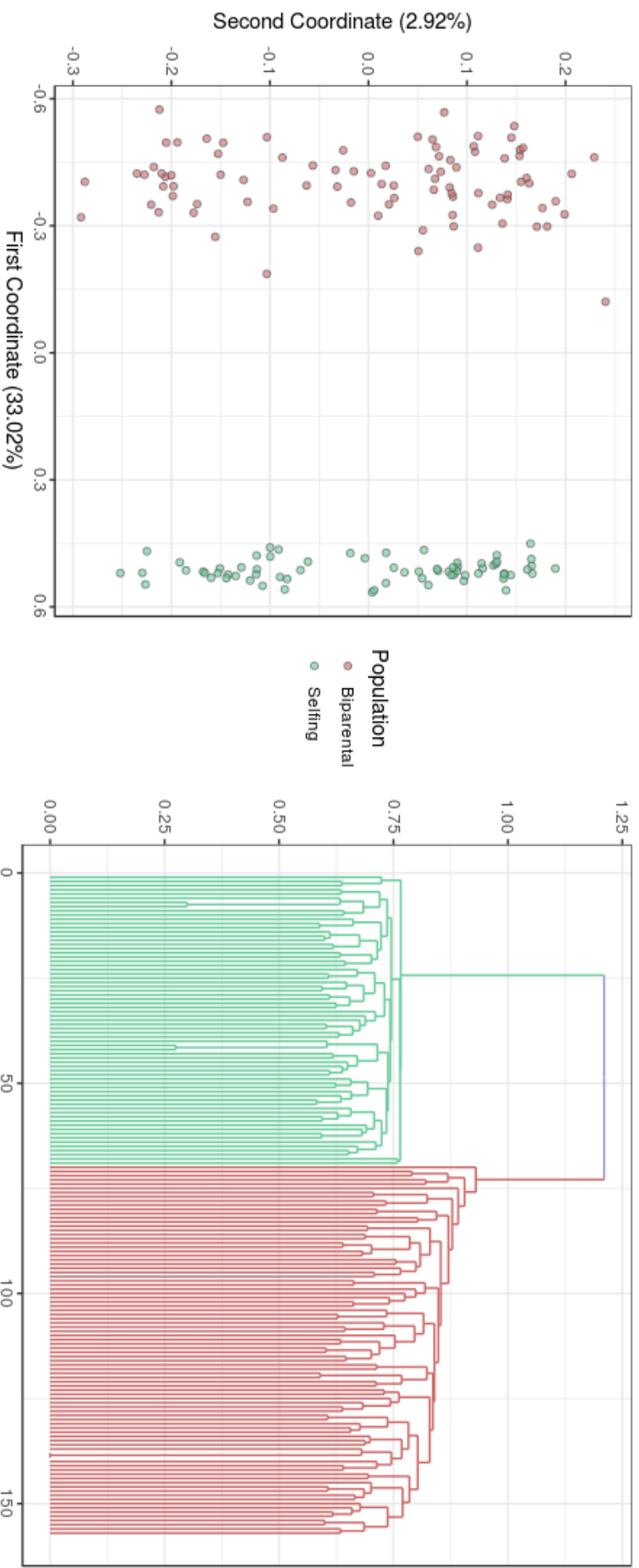


Figura 3. Análise de estruturação intrapopulacional existente na união das populações de clones derivados de cruzamento biparental e de autofecundação em uma única população. À esquerda análise de coordenadas principais. À direita análise hierárquica de clusterização. Resultados obtidos com o uso dos dados de 894 DArT para os 157 indivíduos..

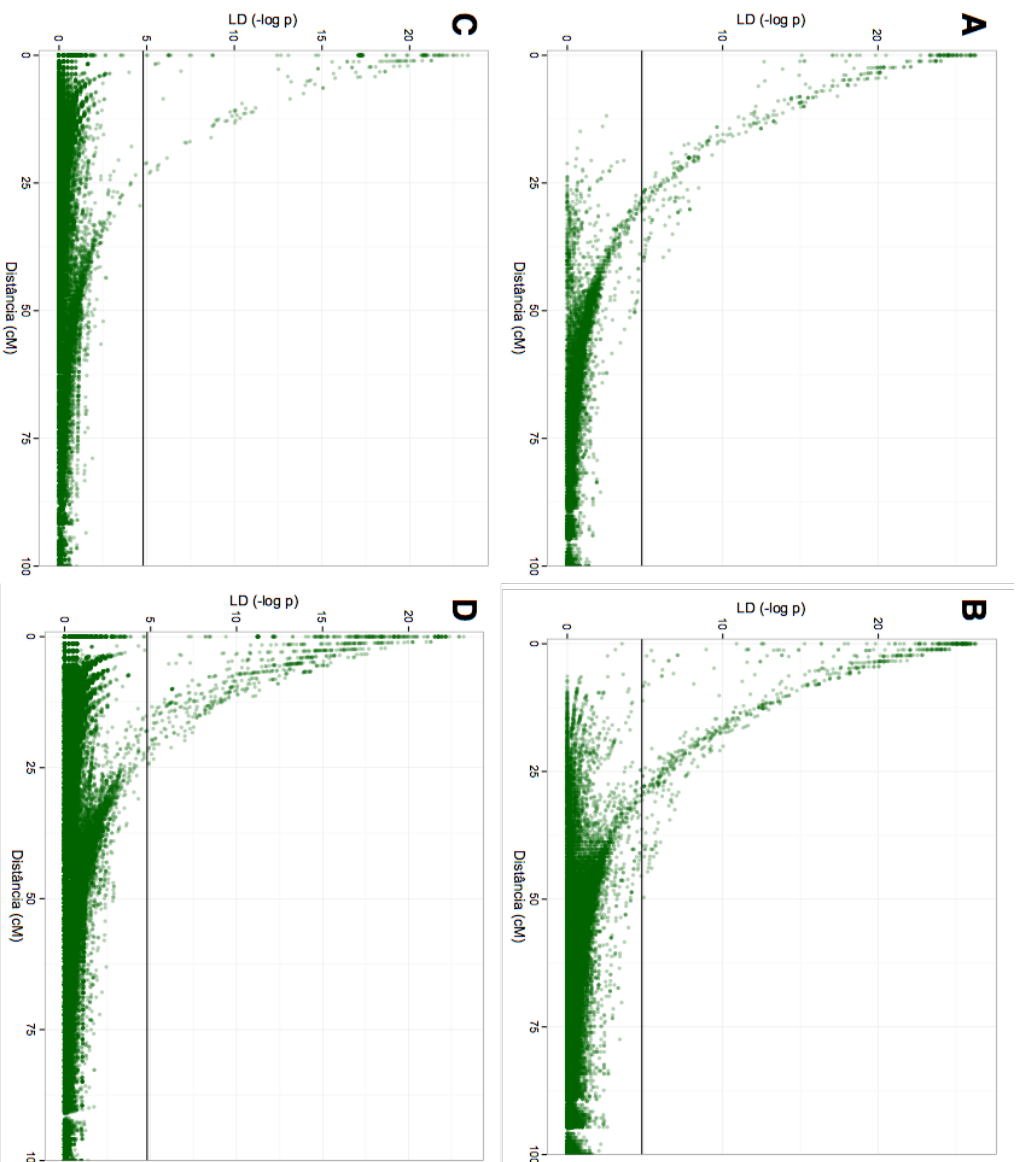


Figura 4. Relação entre o desequilíbrio de ligação e a distância, em cM, entre locos, nas quatro composições possíveis; A) Indivíduos oriundos do cruzamento entre os clones-élite RB97327 e RB72454 (marcadores *single-dose* presentes no genitor feminino); B) Indivíduos oriundos do cruzamento entre os clones-élite RB97327 e RB72454 (marcadores *single-dose* presentes no genitor masculino); C) Indivíduos oriundos do cruzamento entre os clones-élite RB97327 e RB72454 (marcadores *double single-dose* presentes nos dois genitores) e D) Indivíduos oriundos da autofecundação do clone RB97327 (marcadores *double single-dose* presentes no genitor autofecundado).

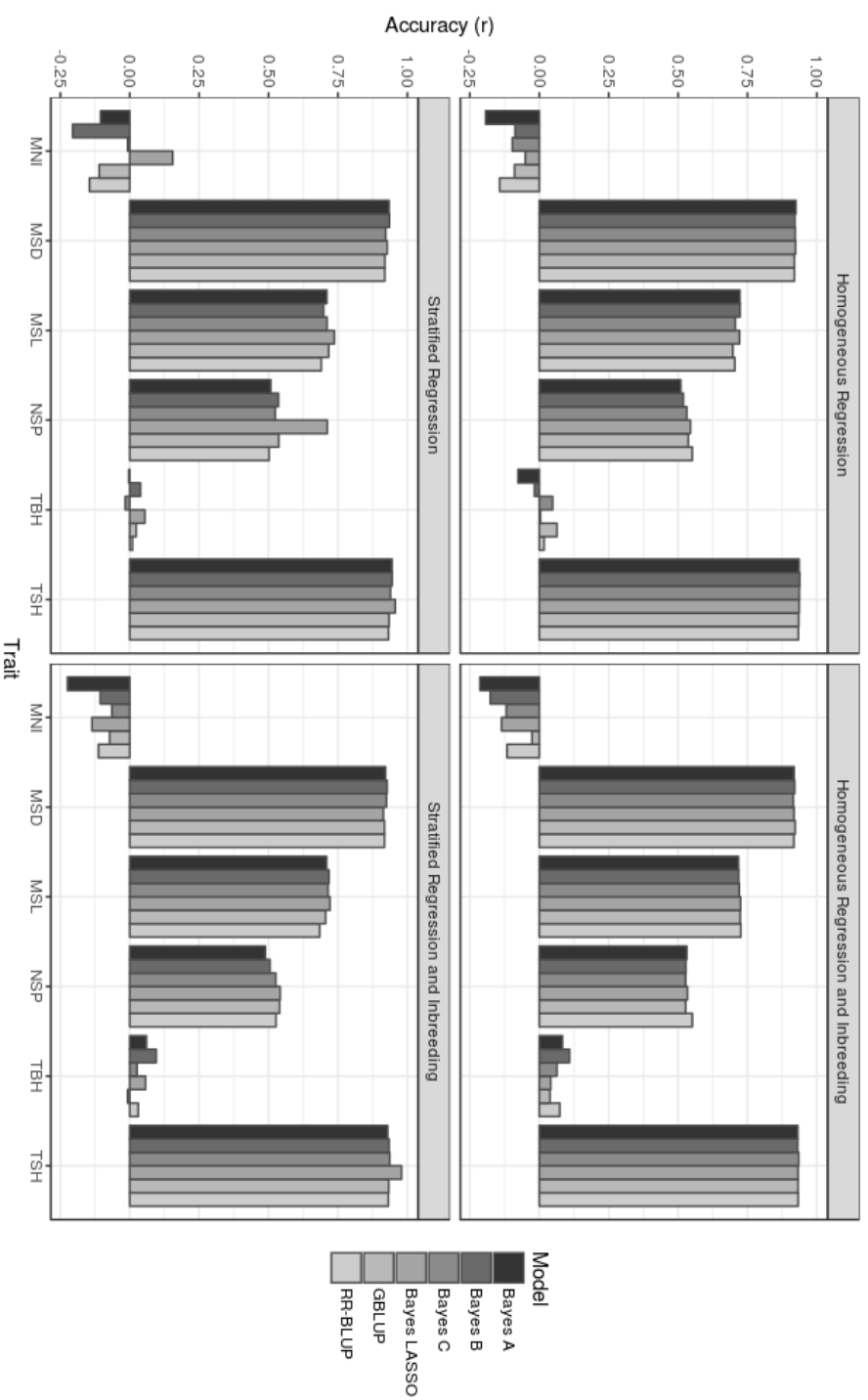


Figura 5. Valores de acurácia obtidos pela correlação entre os valores genéticos obtidos com o uso de modelos mistos (eBLUPs) e os valores genéticos genômicos (VGG) obtidos por validação cruzada para as seis diferentes metodologias de GS testadas. Considerando os seis caracteres analisados: número médio de entrenós (MNI), diâmetro médio do colmo (MSD), comprimento médio de colmo (MSL), número de colmos por parcela (NSP), tonelada de BRX por hectare (TBH) e tonelada de colmos por hectare (TSH).

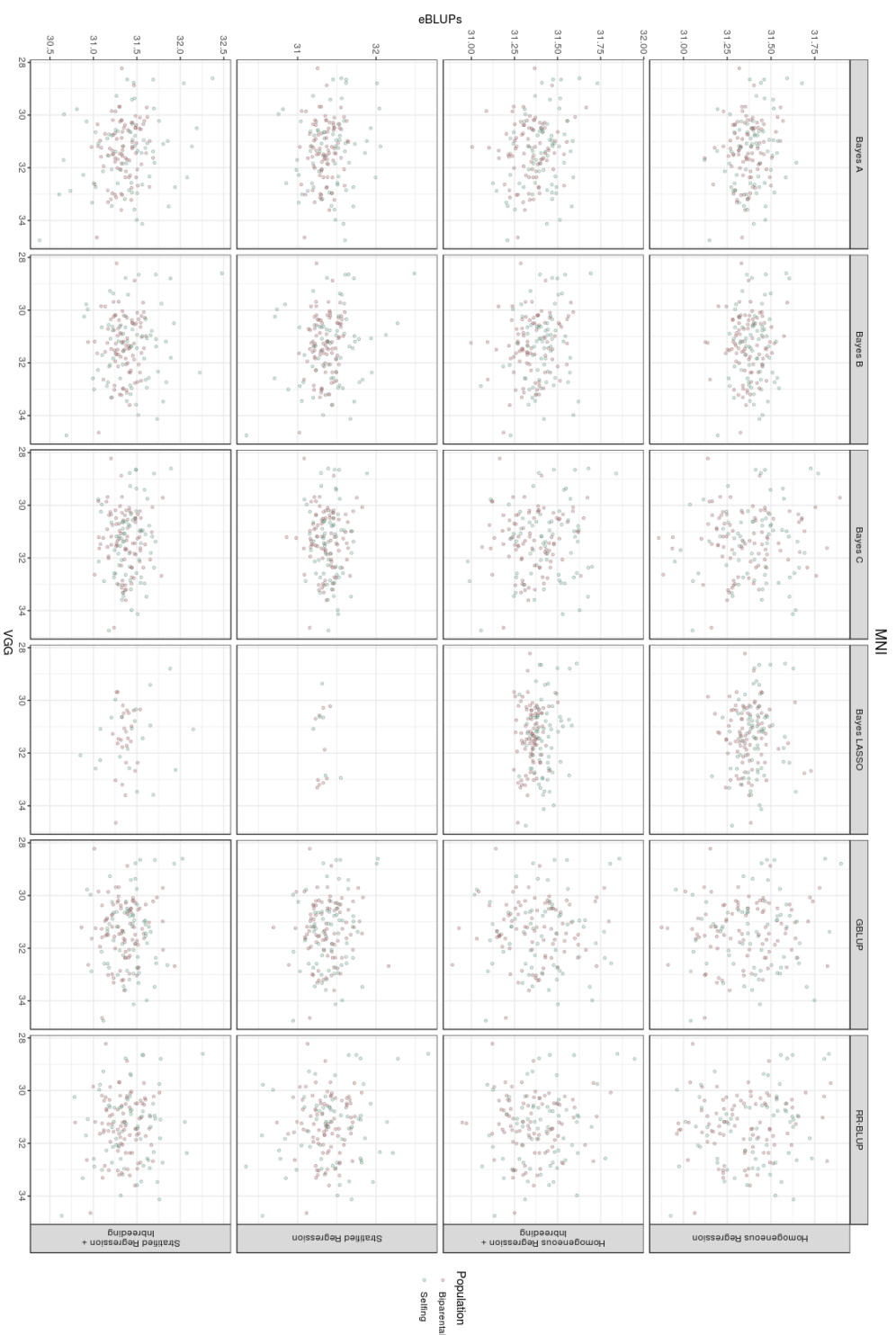


Figura 6. Gráficos de dispersão obtidos para MNI, considerando-se as seis metodologias de GS e as quatro estratégias de modelagem na análise dos dados genotípicos e fenotípicos dos 157 clones de cana-de-açúcar pertencentes a duas subpopulações (indivíduos oriundos do cruzamento biparental, em vermelho, e de autofecundação, em verde).

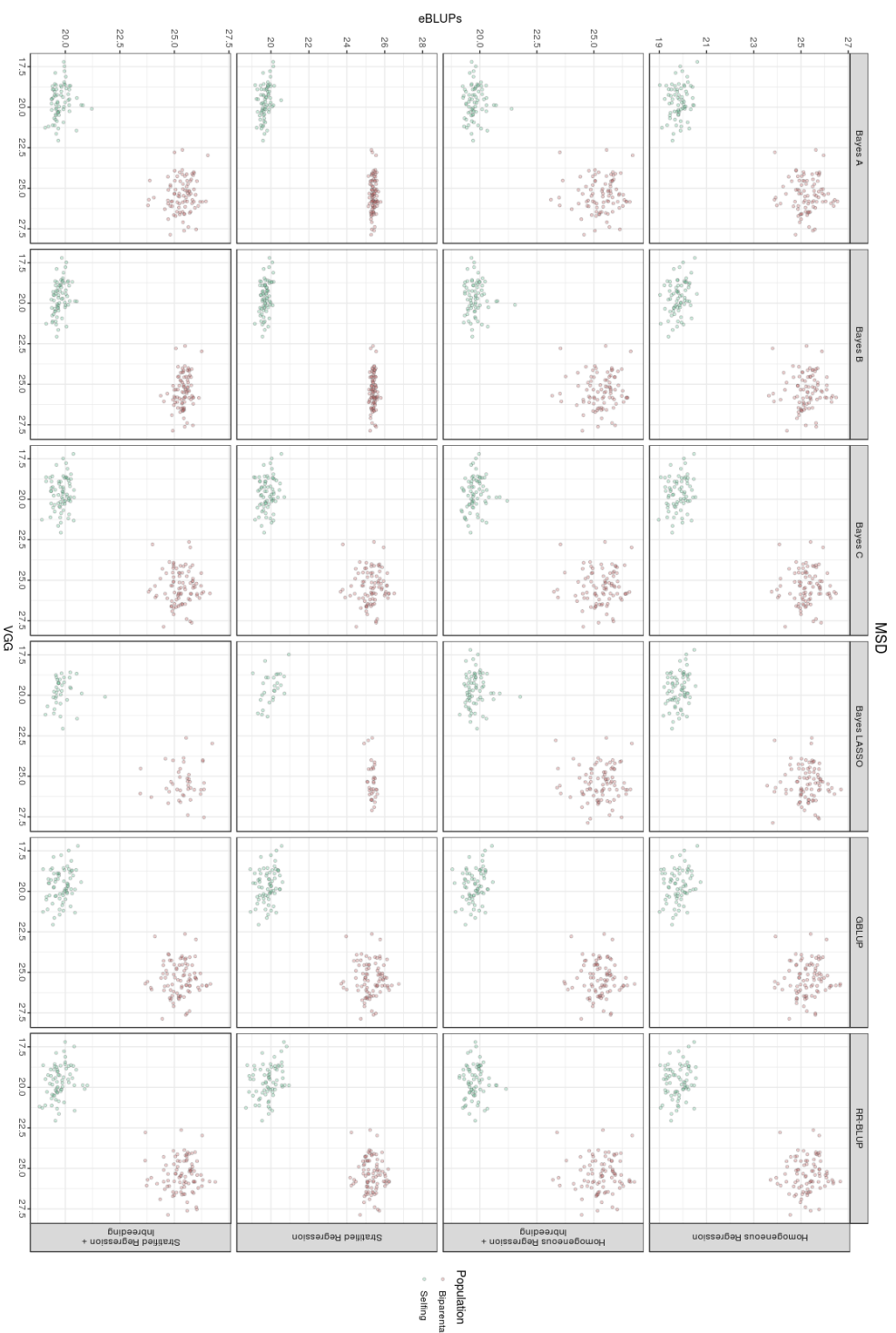


Figura 7. Gráficos de dispersão obtidos para MSD, considerando-se as seis metodologias de GS e as quatro estratégias de modelagem na análise dos dados genotípicos e fenotípicos dos 157 clones de cana-de-açúcar pertencentes a duas subpopulações (indivíduos oriundos do cruzamento biparental, em vermelho, e de autofecundação, em verde).

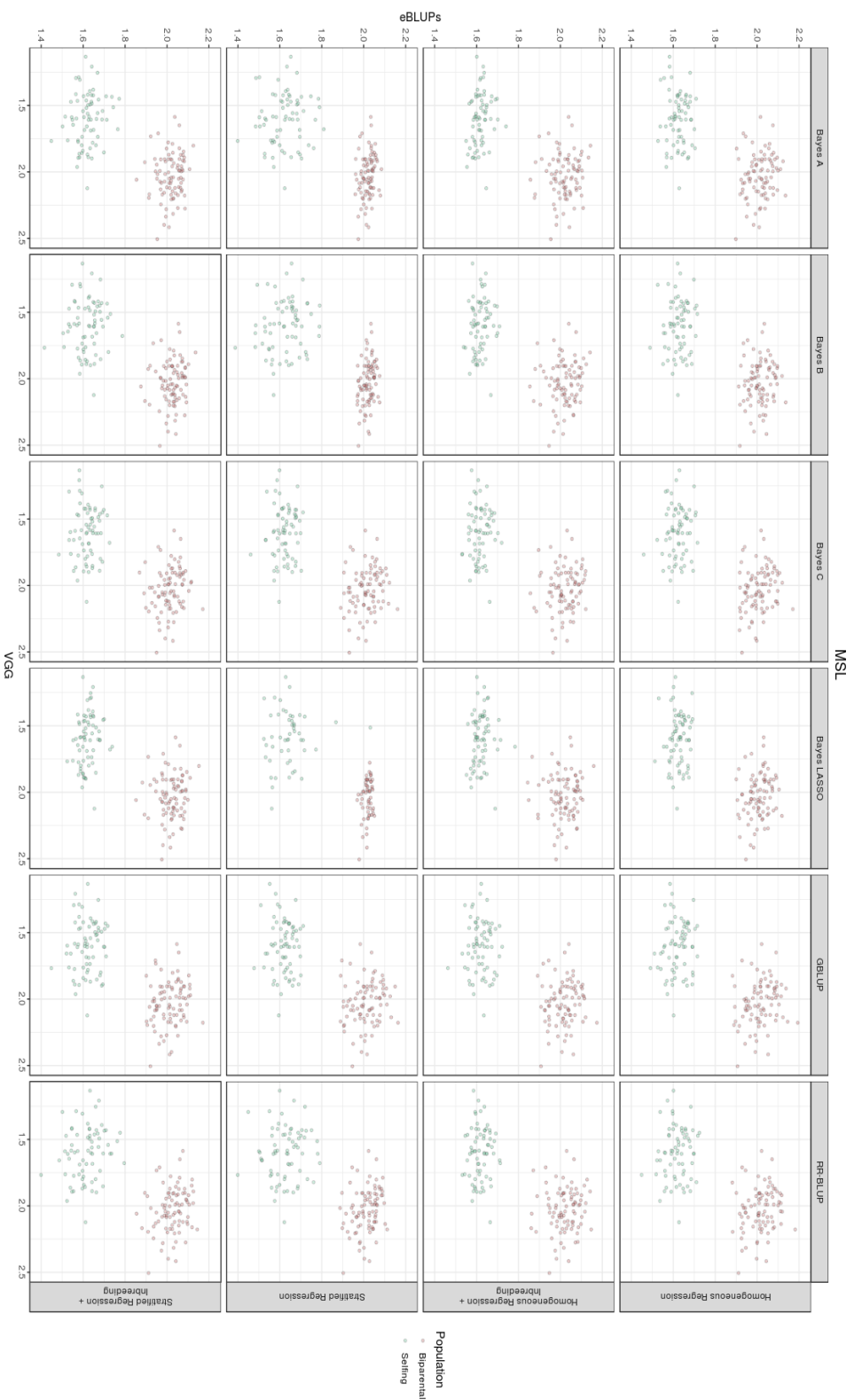


Figura 8. Gráficos de dispersão obtidos para MSL, considerando-se as seis metodologias de GS e as quatro estratégias de modelagem na análise dos dados genotípicos e fenotípicos dos 157 clones de cana-de-açúcar pertencentes a duas subpopulações (indivíduos oriundos do cruzamento biparental, em vermelho, e de autofecundação, em verde).

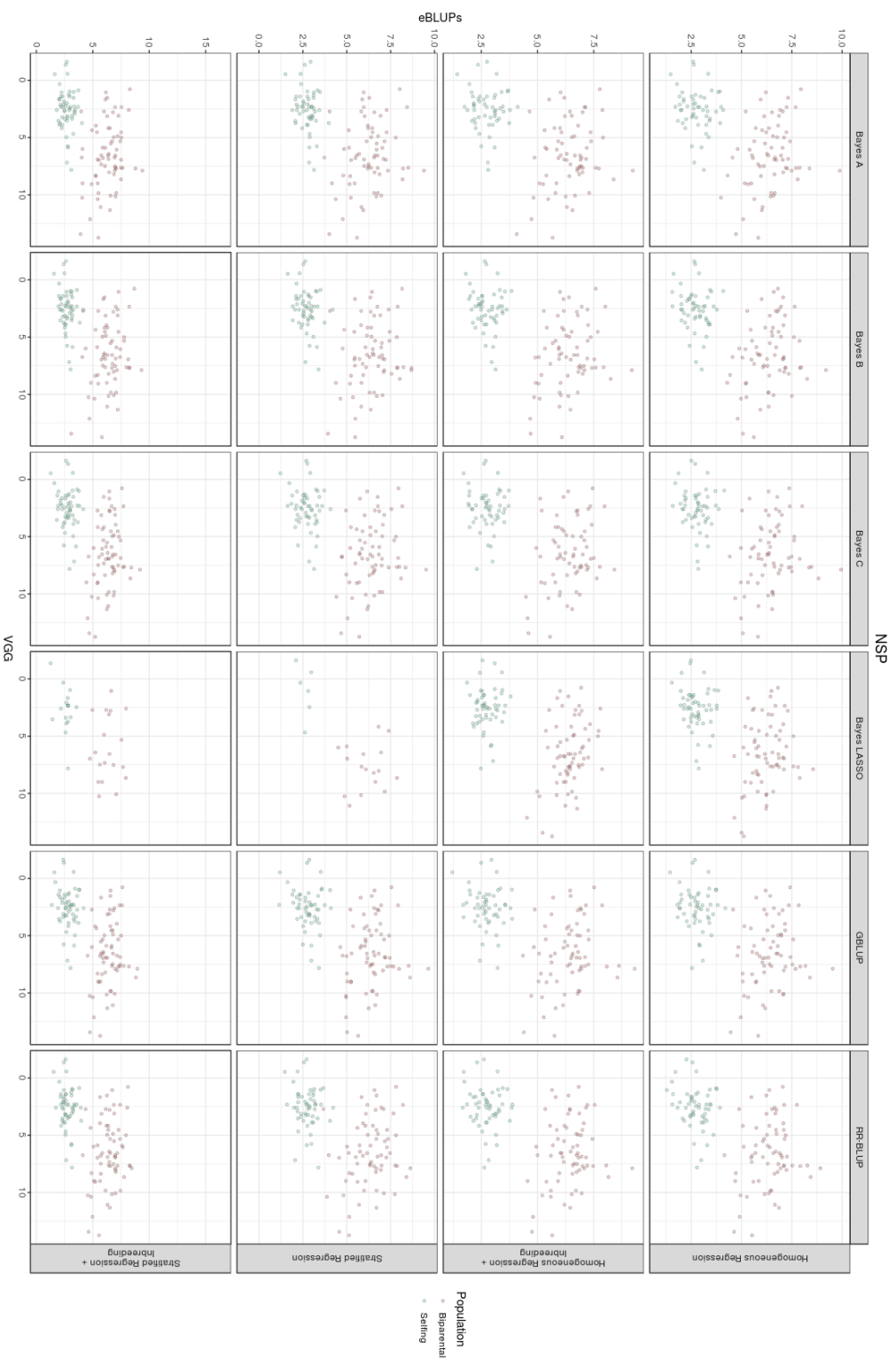


Figura 9. Gráficos de dispersão obtidos para NSP, considerando-se as seis metodologias de GS e as quatro estratégias de modelagem na análise dos dados genotípicos e fenotípicos dos 157 clones de cana-de-açúcar pertencentes a duas subpopulações (indivíduos oriundos do cruzamento biparental, em vermelho, e de autofecundação, em verde).

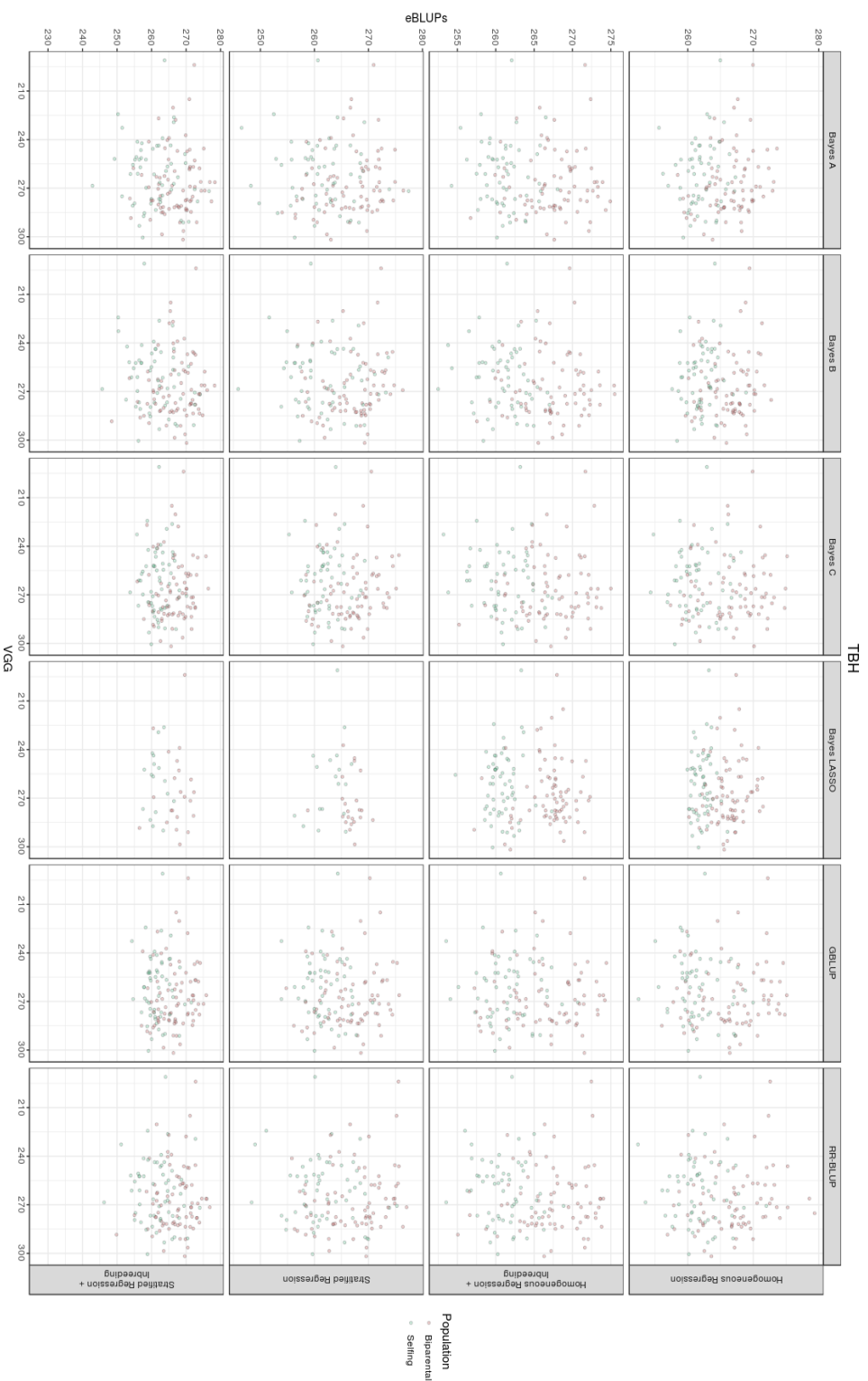


Figura 10. Gráficos de dispersão obtidos para TBH, considerando-se as seis metodologias de GS e as quatro estratégias de modelagem na análise dos dados genotípicos e fenotípicos dos 157 clones de cana-de-açúcar pertencentes a duas subpopulações (indivíduos oriundos do cruzamento biparental, em vermelho, e de autofecundação, em verde).

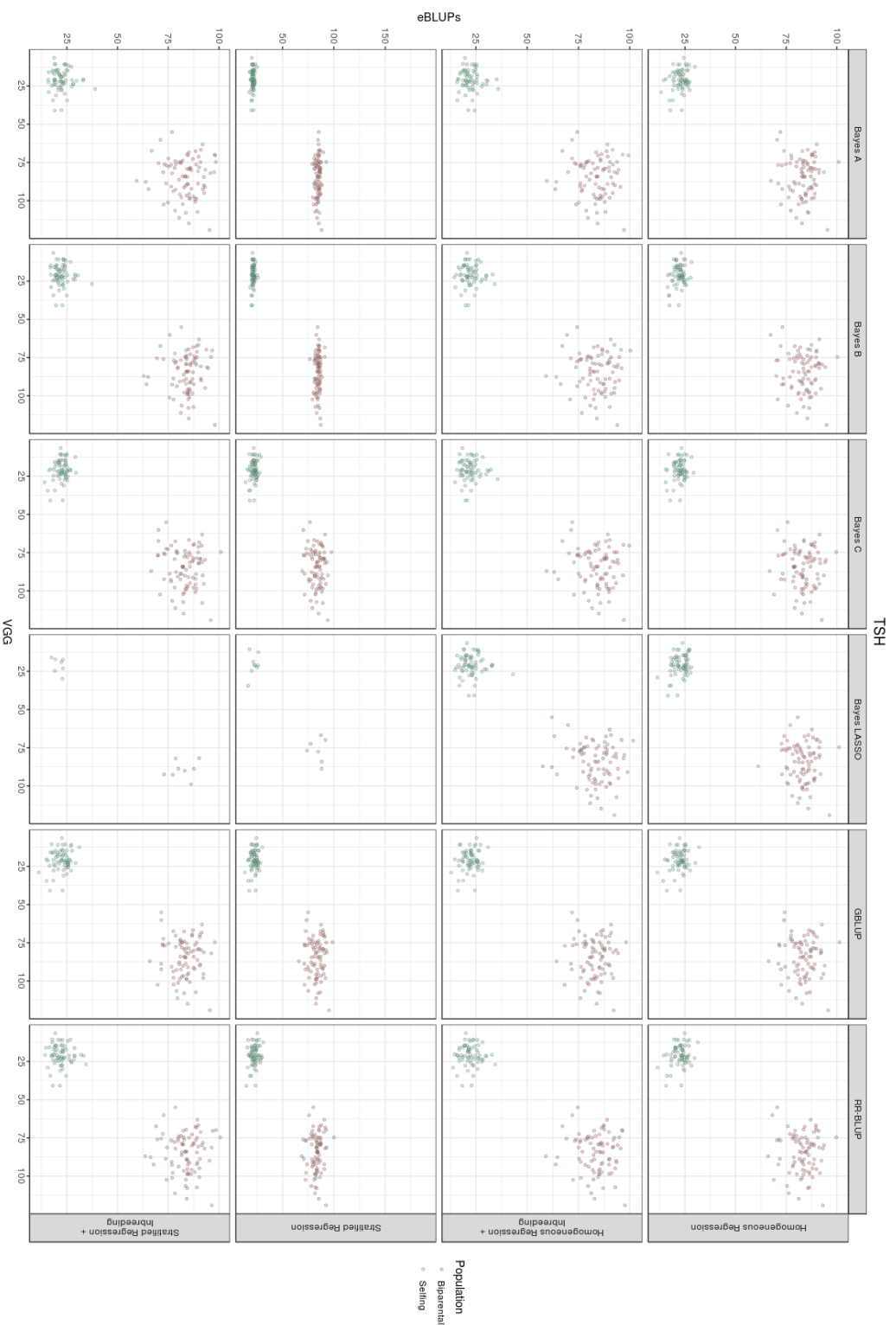


Figura 11. Gráficos de dispersão obtidos para TSH, considerando-se as seis metodologias de GS e as quatro estratégias de modelagem na análise dos dados genotípicos e fenotípicos dos 157 clones de cana-de-açúcar pertencentes a duas subpopulações (indivíduos oriundos do cruzamento biparental, em vermelho, e de autofecundação, em verde).

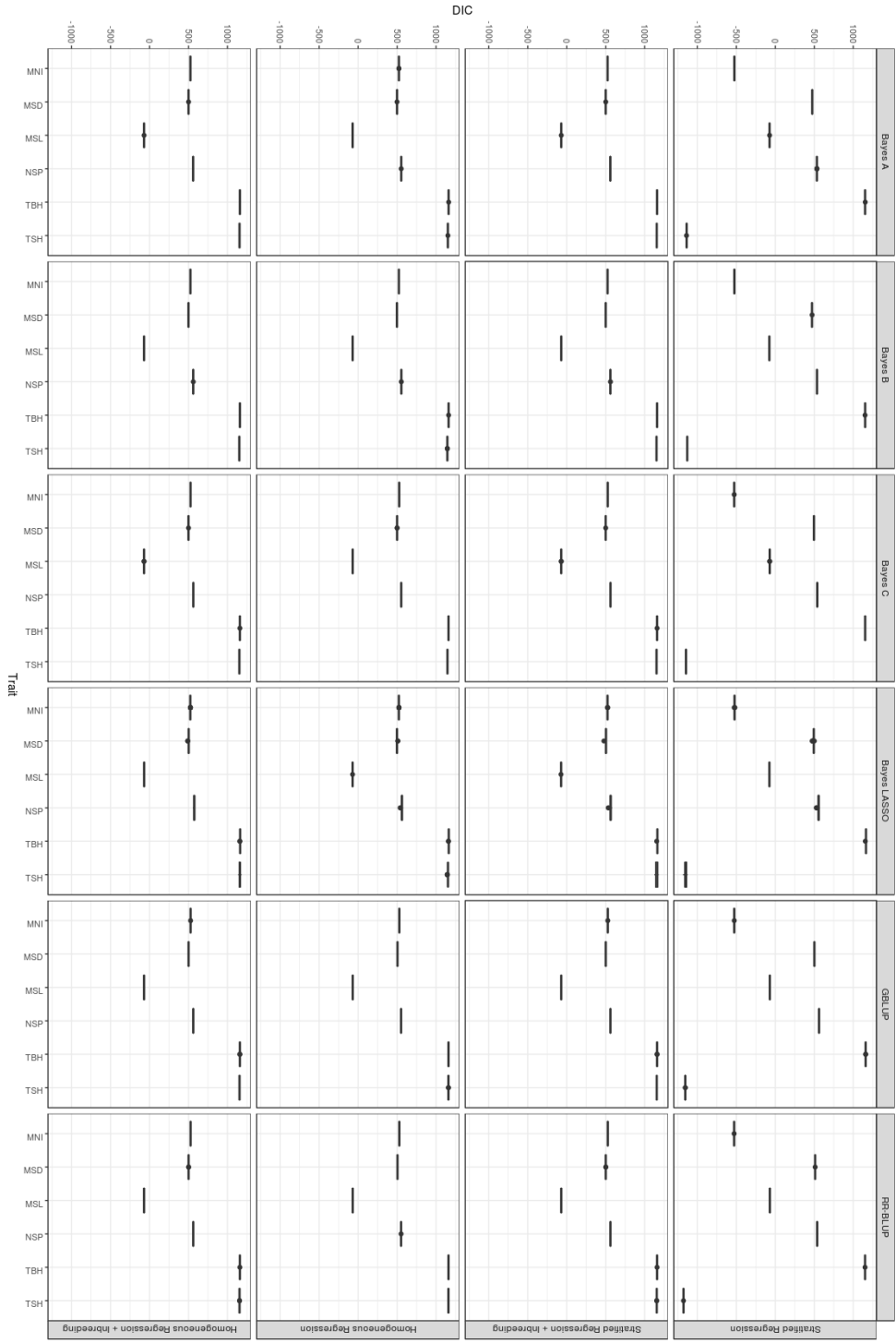


Figura 12. Boxplots representando a distribuição dos valores de DIC para os seis caracteres analisados: número médio de entrenós (MNI), diâmetro médio de colmos (MSD), comprimento médio de colmos (MSL), número de colmos por parcela (NSP), toneladas de colmos por hectare (TBH) e toneladas de colmos por hectare (TSH), considerando-se as 6 metodologias e as 4 abordagens de análise.

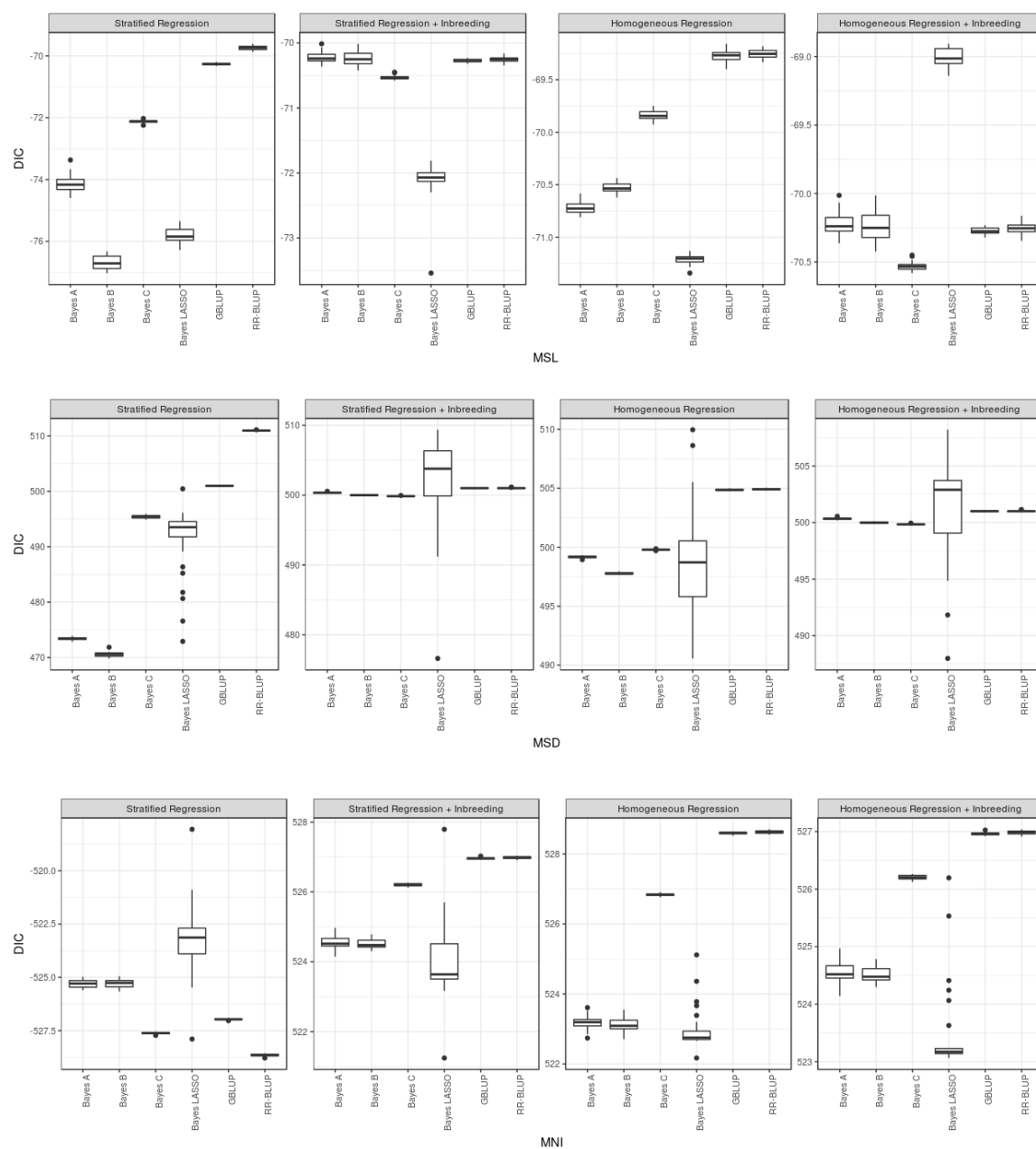


Figura 13. Boxplots por carácter, representando a distribuição dos valores de DIC para os três dos seis caracteres analisados: número médio de entrenós (MNI), diâmetro médio de colmos (MSD), comprimento médio de colmos (MSL), considerando-se as 6 metodologias e as 4 abordagens de análise.

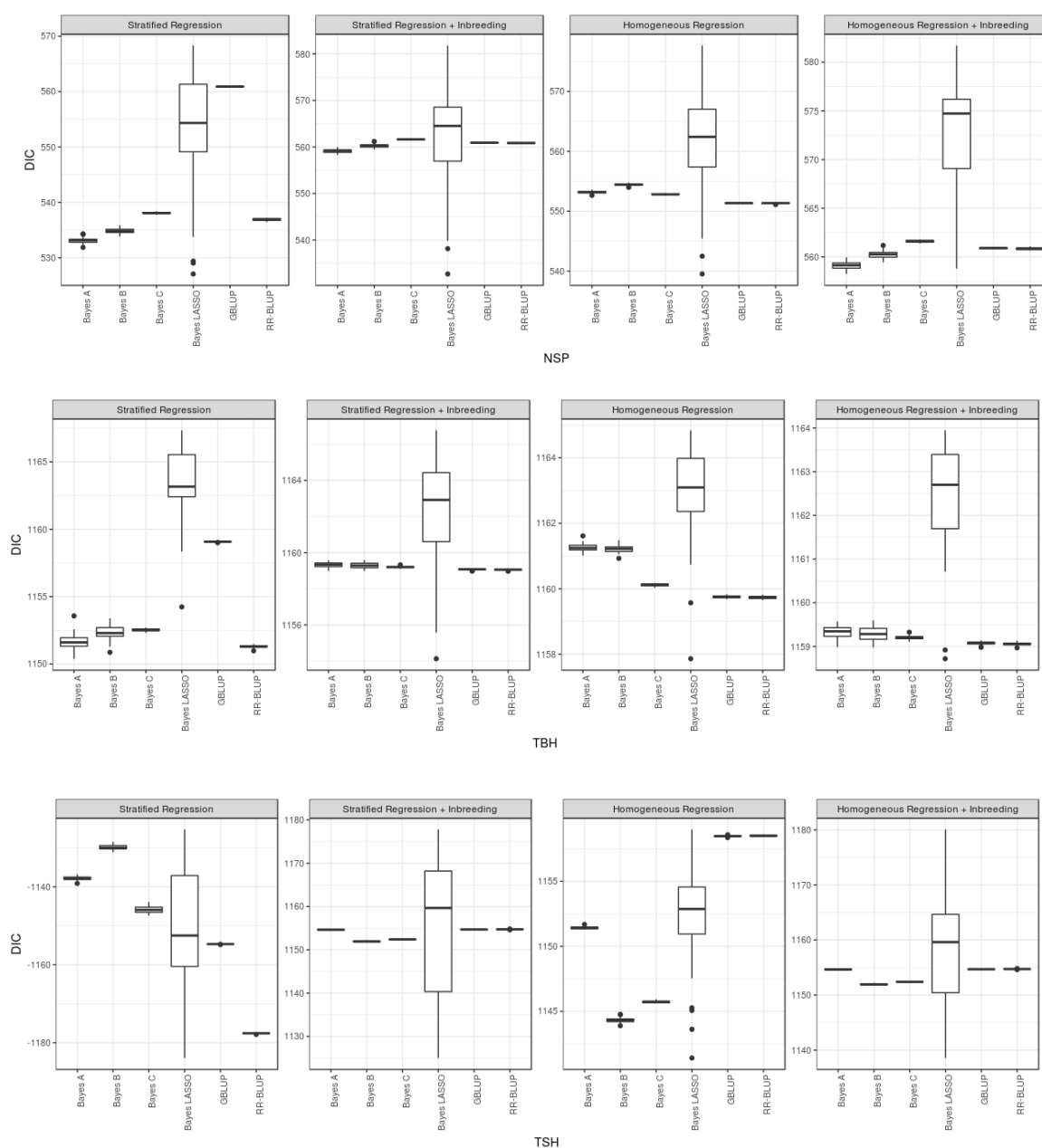


Figura 14. Boxplots por carácter, representando a distribuição dos valores de DIC para três dos seis caracteres analisados: número de colmos por parcela (NSP), toneladas de BRIX por hectare (TBH) e toneladas de colmos por hectare (TSH), considerando-se as 6 metodologias e as 4 abordagens de análise.

4.5 DISCUSSÃO

Nossos resultados ilustram o potencial de aplicação da estratégia de seleção genômica ampla como ferramenta de análise para o melhoramento genético de cana-de-açúcar. Verifica-se que o uso de marcadores DArT foi eficiente para a predição de VGGs

para os caracteres analisados, mesmo diante dos entraves decorrentes da complexidade genética da cultura. É importante ressaltar que os valores de acurácia preditiva obtidos devem ser considerados como potenciais para populações de base genética mais ampla, já que na população avaliada o LD apresentado foi bastante elevado.

O alto LD apresentado na população também justifica a obtenção de valores de acurácia relativamente altos, apesar do número de marcadores utilizados na análise, que não pode ser considerado como elevado. Quanto maior a extensão do LD, maior o tamanho dos blocos de ligação observados na população. Neste sentido, um menor o número de marcadores é necessário para se capturar os efeitos dos polimorfismos causais envolvidos na expressão dos caracteres quantitativos. Como não é intuito da GS estimar efeitos para locos individuais (Meuwissen et al., 2001), evitando a possibilidade de superestimação de efeitos (*i.e.* efeito Beavis) (Xu et al., 2005), pode-se inferir que o número de marcadores utilizados foi suficiente para a obtenção de altas acurácias de seleção para a maioria dos caracteres analisados. Os baixos valores de acurácia obtidos para MNI se devem à baixa variância genética apresentada para este caráter (Tabela 2). No caso de TBH a baixa variância genética apresentada para BRIX nesta população corrobora os resultados obtidos anteriormente (de Bem Oliveira, 2014).

Considerando-se como referência uma acurácia teórica de 0,7 (herdabilidade ao nível de médias igual a 0,5) para a predição de valores genotípicos exclusivamente com base em fenótipos, mesmo com uso de informações de genealogia (Goddart, 2009), os resultados obtidos sugerem que valores superiores aos praticados nas tradicionais estratégias de melhoramento podem ser obtidos para cana-de-açúcar com a aplicação de GS. Cumpre destacar que valores tão ou mais altos de herdabilidade são difíceis de se obter com o uso apenas de seleção fenotípica, sem informações de marcadores, dado que grande parte dos caracteres de importância agrônômica possuem caráter quantitativo e sofrem alta influência do ambiente.

Os resultados obtidos também sugerem que valores similares de acurácia foram obtidos, independente do modelo escolhido para análise (Tabela 3), corroborando resultados existentes na literatura (Heslot et al., 2012; Resende et al., 2012; Gouy et al., 2013). Resultados como estes são esperados quando a distribuição dos efeitos dos marcadores se dá de maneira equitativa, ou seja, quando não são encontradas grandes diferenças nos efeitos dos marcadores para a explicação das variações observadas para o caráter. Modelos Bayesianos (atribuindo efeitos de magnitudes diferentes aos marcadores

utilizados, *i.e.* Bayes A, Bayes B; Bayesian LASSO) se adaptam melhor quando os efeitos de alguns marcadores são mais pronunciados do que outros, ou quando um menor número de locos controla a expressão dos caracteres (Pérez & de Los Campos; 2014). Como os caracteres avaliados neste trabalho mostraram comportamento próximo ao Gaussiano, a aplicação de modelos como RRBLUP e GBLUP mostraram-se tão eficientes quanto os modelos Bayesianos.

Os efeitos de *shrinkage* também foram similares, independente do modelo aplicado na predição dos VGG (Tabela 3). Entretanto, este efeito foi mais evidente nos modelos de estimativas mais baixas de acurácia, como foram os casos dos modelos obtidos para os caracteres MNI e TBH. O efeito de *shrinkage* pode ser observado nos gráficos de dispersão entre os eBLUPs e os VGG, e também na baixa amplitude observada para os valores preditos em relação aos valores observados (Tabela 4).

Um outro aspecto que merece destaque é a inserção do fator de penalização, explícito no caso do RRBLUP, que não só viabiliza a realização da análise mas sobretudo acarreta a melhoria das acurácias dos modelos, pois apesar da inserção de viés para a obtenção dos valores de VGG, minimiza-se a variância do estimador, com redução final nos erros quadráticos médios (de Los Campos et al., 2013). Sem esse procedimento o aumento da variância cresceria proporcionalmente ao aumento do número de parâmetros a serem estimados, dado que se tem um tamanho de amostra fixo e muito menor do que o número de parâmetros de interesse. Assim, a inserção da penalização auxilia no controle dos problemas gerados quando se tem grande quantidades de parâmetros a serem estimados (p) e pequeno tamanho amostral (n).

A baixa variabilidade dos caracteres MNI e BRIX na população avaliada (de Bem Oliveira, 2014) fez com que as acurácias obtidas para este caráter não atingissem valores elevados, inviabilizando a aplicação de GS nestes casos. Este resultado acentua a importância da manutenção da variabilidade genética nos programas de melhoramento, sem a qual a aplicação da seleção fica desprovida de sentido.

Apesar do notório efeito da reduzida variância genética observada, em tese, os fortes efeitos de *shrinkage* e os reduzidos valores de acurácia obtidos nos modelos para os caracteres TBH e MNI poderiam ser também atribuídos ao baixo número de marcadores utilizados ou ao pequeno tamanho amostral, dado que o aumento na densidade de marcadores proporciona ganhos na acurácia (Solberg et al., 2008), bem como o aumento do tamanho das amostras utilizadas para obtenção e validação do modelo (Desta & Ortiz,

2014). Entretanto, outros trabalhos foram desenvolvidos considerando-se números ainda menores tanto de indivíduos como de marcadores, apresentando resultados semelhantes aos obtidos para cana-de-açúcar, também com acurácias elevadas (Fritsche-Neto et al., 2012; Cavalcanti et al., 2012). Estes relatos demonstram que o método pode ser eficiente, mesmo com baixo número de marcas e principalmente com baixo número de indivíduos, dada a estrutura genética apresentada pelos indivíduos sob análise.

Considerando-se a proximidade dos valores de acurácia obtidos, a comparação dos modelos foi realizada pela definição de medidas de ajuste (*goodness-of-fit*) e de complexidade dos modelos, ou seja, a quantidade de parâmetros envolvidos no ajuste do modelo. Pela comparação dos valores de DIC pôde-se apontar estratégias e modelos mais eficientes para cada um dos caracteres analisados, sendo que o uso da regressão estratificada foi a estratégia que gerou os menores valores de DIC para todos os caracteres. Fazendo uso de dados de simulação, Pérez & de Los Campos (2014) mostraram o favorecimento dos modelos Bayes A e Bayes B sobre o GBLUP e RRBLUP, sugerindo que este efeito se deve à simplicidade de arquitetura genética gerada para os caracteres analisados. No presente trabalho (Figuras 12 a 14,) os modelos Bayesianos apresentaram melhor relação ajuste-complexidade para MSD, MSL, NSP e TBH, o que pode sugerir uma arquitetura genética mais simples para esses caracteres. Já para MNI e TSH os modelos apontados com melhor relação ajuste-complexidade foram baseados na metodologia RRBLUP, sugerindo uma arquitetura genética mais complexa para estes caracteres, uma distribuição dos efeitos dos locos sobre o caráter mais próxima à Gaussiana (maior número de locos contribuindo para a expressão do caráter).

Outro resultado relevante obtido neste trabalho trata da influência da estrutura genética intrapopulacional nas acurácias obtidas, corroborando resultados existentes na literatura (Habier et al., 2007, Gianola et al., 2009; Habier et al., 2010, Guo et al., 2014). Assim, mesmo para espécies poliploides e de alta complexidade genômica como a cana-de-açúcar, os efeitos de parentesco influenciam grandemente os resultados obtidos. Este efeito pode ser visualizado na análise dos gráficos de dispersão, que retratam a relação entre os valores preditos (VGG) e os valores observados (eBLUPs), em que se observa a presença de dois diferentes aglomerados de indivíduos. Verificou-se que quanto maior a acurácia mais coesos e determinados estes grupos se encontram, como pode ser visto claramente nos gráficos que retratam a dispersão dos valores para MSL e MSD. Para os demais caracteres, este efeito foi tão menos evidente quanto menor a acurácia do modelo.

Além disso, quando consideradas as estimativas dos componentes de variância, vê-se que a conformação dos grupos está diretamente relacionada à proporção da variabilidade genética explicada pelo fator subpopulação – quanto menor a porção da herdabilidade atribuída a este fator, menor a separação entre os grupos. Assim, não se vê a formação de grupos distintos para TBH e MNI (Figuras 6 e 10), que apresentaram baixa proporção da herdabilidade correspondente ao efeito de população, 2,33% e 0% respectivamente (Tabela 2).

Apesar da conhecida importância do LD nas acurácias da seleção genômica (Mewissen et al., 2001), os resultados obtidos neste trabalho evidenciam a força do efeito de parentesco nas estimativas obtidas de acurácia. Como reportado por Habier et al. (2007), os marcadores utilizados na seleção genômica podem capturar além dos efeitos resultantes do LD entre os marcadores e os polimorfismos causais, os efeitos de associação entre estes locos resultantes do parentesco. Assim, nestes modelos a existência de parentesco entre os indivíduos sob análise afeta os VGG preditos e, conseqüentemente, as estimativas de acurácia. Conforme salientam Habier et al. (2007), este efeito pode ser tal que, mesmo na inexistência de LD, a acurácia obtida na comparação entre os valores observados e os valores preditos com os modelos de GS pode ser diferente de zero.

Dada a quantidade de marcadores utilizados e o principalmente devido ao desequilíbrio de ligação induzido na população sobre análise, pode-se considerar alta a probabilidade de que parte dessas marcas se encontrem em forte desequilíbrio de ligação com os polimorfismos causais associados aos caracteres analisados. Ao mesmo tempo, dada a forte relação de parentesco entre os indivíduos analisados, admite-se que o efeito do parentesco tenha, provavelmente, contribuído para o incremento das estimativas de acurácia preditiva obtidas com os modelos de GS. É importante destacar que o efeito de parentesco sobre as acurácias dos modelos não impede a sua aplicação em programas de seleção genômica, mas reduz a longevidade dos modelos obtidos, uma vez que os efeitos de parentesco se dissipam com maior rapidez, quando comparados com os efeitos decorrentes do desequilíbrio de ligação.

4.6 CONSIDERAÇÕES FINAIS

Os valores de acurácia aqui apresentados servem como prova de conceito para a aplicação de GS em cana-de-açúcar, fornecendo evidências empíricas de que a estratégia de seleção genômica pode aumentar o ganho genético para a cultura, dado aos altos valores

de acurácia observados, além de permitir a diminuição do tempo e do custo envolvidos no melhoramento genético (Lorenzana & Bernardo, 2009; Heffner et al., 2011).

Este trabalho mostra também a importância da avaliação da existência de estruturação genética populacional, antes da aplicação de GS, visto que esse efeito tem grande influência nos resultados obtidos. Estratégias para o controle destes efeitos de estruturação genética populacional devem ser aprimorados, permitindo a aplicação de GS em populações com alta estruturação intrapopulacional.

Destaca-se ainda a importância de desenvolvimento de modelos que incluam os efeitos genéticos não-aditivos para espécies poliploides como a cana-de-açúcar, em que os efeitos de dominância e de epistasia poderiam impactar favoravelmente o melhoramento baseado na obtenção de clones superiores.

4.7 REFERÊNCIAS

- ALJANABI, S. M.; FORGET, L.; DOOKUN, A. An improved and rapid protocol for the isolation of polysaccharide-and polyphenol-free sugarcane DNA. *Plant Molecular Biology Reporter*, Amsterdam, v. 17, n. 3, p. 281-281, 1999.
- DE VRIES, A. e RIPLEY, B. D. gg dendro: Create Dendrograms and Tree Diagrams Using 'ggplot2'. R package version 0.1-20. <https://CRAN.R-project.org/package=ggdendro>. 2016.
- ASNAGHI, C.; PAULET, F.; KAYE, C.; GRIVET, L.; DEU, M.; GLASZMANN, J. C.; D'HONT, A. Application of synteny across Poaceae to determine the map location of a sugarcane rust resistance gene. *Theoretical and Applied Genetics*, Heidelberg, v. 101, n. 5-6, p. 962-969, 2000.
- BATES, D.; MAECHLER, M.; BOLKER, B.; WALKER, S. lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-8, <http://CRAN.R-project.org/package=lme4>.
- BRASIL, MINISTÉRIO DA AGRICULTURA, PECUÁRIA E ABASTECIMENTO (MAPA). Produção e produtividade médias de lavouras temporárias e permanentes – Cana-de-Açúcar. 2012. Disponível em: <http://www.agricultura.gov.br>. Acesso em: 23 jul. 2012.
- BREMER, G. Problems in breeding and cytology of sugar cane. *Euphytica*, Wageningen, v. 10, n. 1, p. 59-78, 1961.
- BURR, G. O.; HARTT, C. E.; BRODIE, H. W.; TANIMOTO, T.; KORTSCHAK, H. P.;

- TAKAHASHI, D.; ASHTON, F. M.; COLEMAN, R. E. The sugarcane plant. Annual Review of Plant Physiology, Califórnia, v. 8, n. 1, p. 275-308, 1957.
- CARNEIRO, M. S. & VIEIRA, M. L. C. Mapas genéticos em plantas. Bragantia, Campinas - SP, v. 61, n. 2, p. 89-100, ago. 2002.
- CAVALCANTI, J. J. V.; RESENDE, M. D. V.; SANTOS, F. H. C.; PINHEIRO, C. R.. Simultaneous prediction of the effects of molecular markers and genome wide selection in cashew. Revista Brasileira de Fruticultura, v. 34, n. 3, p. 840-846, 2012.
- CHEAVEGATTI-GIANOTTO, A.; DE ABREU, H. M. C.; ARRUDA, P.; BESPALHOK FILHO, J. C.; BURNQUIST, W. L.; CRESTE, S.; DI CIERO, L.; FERRO, J. A.; FIGUEIRA, A. V. O.; FILGUEIRAS, T. S.; GROSSI-DE-SÁ, M. F.; GUZZO, E. C.; HOFFMANN, H. P.; LANDELL, M. G. A.; MACEDO, N.; MATSUOKA, S.; REINACH, F. C.; ROMANO, E.; SILVA, W. J.; SILVA FILHO, M. C.; ULIAN, E. C. Sugarcane (*Saccharum X officinarum*): a reference study for the regulation of genetically modified cultivars in Brazil. Tropical Plant Biology, Aiea, Hawaii, v. 4, n. 1, p. 62-89, 2011.
- CONAB, Companhia Nacional de Abastecimento. Acompanhamento de safra brasileira : Cana-de-açúcar, quarto levantamento, abril/2013 - Companhia Nacional de Abastecimento. Brasília: Conab 2013. Disponível em:< <http://www.conab.gov.br>>. Acesso em: 20 nov. 2012.
- CUADRADO, A., ACEVEDO, R., DE LA ESPINA, S. M. D., JOUVE, N., & DE LA TORRE, C. Genome remodelling in three modern *S. officinarum* × *S. spontaneum* sugarcane cultivars. Journal of experimental botany, Lancaster, v. 55, n. 398, p. 847-854, 2004.
- D'HONT, A.; GRIVET, L.; FELDMANN, P.; GLASZMANN, J. C.; RAO, S.; BERDING, N. Characterisation of the double genome structure of modern sugarcane cultivars (*Saccharum* spp.) by molecular cytogenetics. Molecular and General Genetics, Heidelberg, v. 250, n. 4, p. 405-413, 1996.
- D'HONT, A.; LU, Y. H.; LEÓN, D. G. D.; GRIVET, L.; FELDMANN, P.; LANAUD, C.; GLASZMANN, J. C. A molecular approach to unraveling the genetics of sugarcane, a complex polyploid of the Andropogoneae tribe. Genome, Birmingham, v. 37, n. 2, p. 222-230, 1994.
- D'HONT, A. Unraveling the genome structure of polyploids using FISH and GISH;

examples of sugarcane and banana. Cytogenetic and genome research, Basel, v. 109, n. 1-3, p. 27-33, 2005.

DAUGROIS, J. H.; GRIVET, L.; ROQUES, D.; HOARAU, J. Y.; LOMBARD, H.; GLASZMANN, J. C.; & D'HONT, A. A putative major gene for rust resistance linked with a RFLP marker in sugarcane cultivar 'R570'. Theoretical and Applied Genetics, Heidelberg, v. 92, n. 8, p. 1059-1064, 1996.

DE BEM OLIVEIRA, I. Desequilíbrio de ligação e análise de seleção genômica em cana-de-açúcar. 85f.

DE LOS CAMPOS, G.; HICKEY, J. M.; PONG-WONG, R.; DAETWYLER, H. D.; CALUS, M. P. L. Whole-genome regression and prediction methods applied to plant and animal breeding. Genetics, Pittsburg, v. 193, n. 2, p. 327-345, 2013.

DESTA, Z. A.; ORTIZ, R. Genomic selection: genome-wide prediction in plant improvement. Trends in plant science, v. 19, n. 9, p. 592-601, 2014.

DURÃES, F. O. M.; SANTOS, M. X. dos; GAMA, E. G.; MAGALHÃES, P. C.; ALBUQUERQUE, P. E. P.; GUIMARÃES, C. T. 2004. Fenotipagem associada à tolerância a seca em milho para uso em melhoramento, estudos genômicos e seleção assistida por marcadores. Sete Lagoas: Embrapa Milho e Sorgo, 17 f. Disponível em: <http://www.cnpms.embrapa.br/publicacoes/publica/2004/circular/Circ_39.pdf>. Acesso em: 10 ago. 2012.

FALEIRO, F. G. Marcadores Genético-Moleculares aplicados a programas de conservação e uso de recursos genéticos. Planaltina: Embrapa Cerrados, 2007. 102 p.

FAOSTAT (2015) http://faostat3.fao.org/browse/Q/*/E. Acessado em 4 Julho 2015.

FEDERER, Walter T.; RAGHAVARAO, D. On augmented designs. Biometrics, p. 29-35, 1975.

FERNANDO, R.L.; GROSSMAN, M. Marker Assisted Selection using Best Linear Unbiased Prediction. Genetics Selection Evolution, Londres, n. 21, p. 467-477, 1989.

FERREIRA, M.E.; GRATTAPAGLIA, D. Introdução ao uso de marcadores moleculares em análise genética. Brasília: Embrapa Recursos Genéticos e Biotecnologia, 1998. 220 p.

FRITSCHÉ-NETO, R.; RESENDE, M. D. V.; MIRANDA, G. V.; DO VALE, J. C. Seleção genômica ampla e novos métodos de melhoramento do milho. Revista Ceres,

Viçosa – MG, v. 59, n. 6, 2012.

GIANOLA, D.; DE LOS CAMPOS, G.; HILL, W. G.; MANFREDI, E.; FERNANDO, R. Additive genetic variability and the Bayesian alphabet. *Genetics*, v. 183, n. 1, p. 347-363, 2009.

GODDARD, M. E.; HAYES, B. J. Genomic selection. *Journal of Animal Breeding and Genetics*, v. 124, n. 6, p. 323-330, 2007.

GOUY, M.; ROUSSELLE, A.; THONG CHANE, A.; ANGLADE, A.; ROYAERT, S.; NIBOUCHE, S.; COSTET, L. Genome wide association mapping of agro-morphological and disease resistance traits in sugarcane. *Euphytica*, v. 202, n. 2, p. 269-284, 2015.

GOUY, M.; RUSSELL, Y.; BASTIANELLI, D.; LECOMTE, P.; BONNAL, L.; ROQUES, D.; EFILE, J.-C.; ROCHER, S.; DAUGROIS, J.; TOUBI, L.; NABENEZA, S.; HERVOUET, C.; TELISMART, H.; DENIS, M.; THONG-CHANE, A.; GLASZMANN, J. C.; HOARAU, J.-Y.; NIBOUCHE, S.; COSTET, L. Experimental assessment of the accuracy of genomic selection in sugarcane. *Theoretical and Applied Genetics*, Nova York: NY, v. 126, n. 10, p. 2575-2586, 2013.

GUO, Z.; TUCKER, D. M.; BASTEN, C. J.; GANDHI, H.; ERSOZ, E.; GUO, B.; XU, Z.; WANG, D.; GAY, G. The impact of population structure on genomic prediction in stratified populations. *Theoretical and Applied Genetics*, p. 1-14, 2014.

HABIER, D.; FERNANDO, R. L.; KIZILKAYA, K.; GARRICK, D. Extension of the Bayesian alphabet for genomic selection. *BMC Bioinformatics* 12(1), 186, 2011.

HABIER, D.; FERNANDO, R. L.; DEKKERS, J. C. M. The impact of genetic relationship information on genome-assisted breeding values. *Genetics*, v. 177, n. 4, p. 2389-2397, 2007.

HABIER, D.; TETENS, J.; SEEFRIED, F. R.; LICHTNER, P.; THALLER, G. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. *Genetics Selection Evolution*, v. 42, n. 1, p. 1, 2010.

HEDRICK, P. W. *Genetics of Population*. 4a ed. USA: Jones and Bartlett Publishers, Inc., 2011. 675 p.

HEFFNER, E. L.; JANNINK, J.; SORRELLS, M. E. Genomic selection accuracy using multifamily prediction models in a wheat breeding program. *The Plant Genome*, Madison, v. 4, n. 1, p. 65-75, 2011.

- HOARAU, J. Y.; GRIVET, L.; OFFMANN, B.; RABOIN, L.M.; DIOFLAR, J.P.; PAYET, J.; HELLMANN, M.; D'HONT, A.; GLASZMANN, J.C. Genetic dissection of a modern sugarcane cultivar (*Saccharum* spp.). II. Detection of QTLs for yield components. *Theoretical and Applied Genetics*, Heidelberg, v. 105, n. 6-7, p. 1027-1037, 2002.
- HOERL, A. E. AND R. W. KENNARD (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 42(1), 80–86.
- JACCOUD, D.; PENG, KAIMAN; FEINSTEIN, D.; KILIAN, A. Diversity arrays: a solid state technology for sequence information independent genotyping. *Nucleic acids research*, Oxford, v. 29, n. 4, p. e25-e25, 2001.
- JANNOO, N.; GRIVET, L.; DOOKUN, A.; D'HONT, A.; GLAZMANN, J. C. Linkage disequilibrium among modern sugarcane cultivars. *Theoretical and Applied Genetics*, Heidelberg, v. 99, n. 6, p. 1053-1060, 1999b.
- JANNOO, N.; GRIVET, L.; SEGUIN, M.; PAULET, F.; DOMAINGUE, R.; RAO, P. S.; DOOKUN, A.; D'HONT, A.; GLASZMANN, J. C. Molecular investigation of the genetic base of sugarcane cultivars. *Theoretical and applied genetics*, Heidelberg, v. 99, n. 1-2, p. 171-184, 1999.
- LORENZANA, R. E.; BERNARDO, R. Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. *Theoretical and applied genetics*, Heidelberg, v. 120, n. 1, p. 151-161, 2009.
- MENOSSE, M., SILVA-FILHO, M. C., VINCENTZ, M., VAN-SLUYS, M. A., SOUZA, G. M. Sugarcane functional genomics: gene discovery for agronomic trait development. *International Journal of Plant Genomics*, New York, v. 2008, n. 458732, 11 p. 2008.
- MEUWISSEN, T. H. E.; HAYES, B. J.; GODDART, M. E. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, Heidelberg, v. 157, n. 4, p. 1819-1829, 2001.
- MING, R.; S.C.; L.; MORRE, P. H.; IRVINE, J. E.; PATERSON, A. H. QTL analysis in a complex autopolyploid: genetic control of sugar content in sugarcane. *Genome Research*, Nova York, v. 11, n. 12, p. 2075-2084, 2001.
- MUÑOZ, P. R.; RESENDE, M. F. R.; GEZAN, S. A.; RESENDE, M. D. V.; de los CAMPOS, G.; KIRST, M.; HUBER, D.; PETER, G. F. Unraveling Additive from Nonadditive Effects Using Genomic Relationship Matrices. *Genetics*, v. 198, n. 4, p. 1759-

1768, 2014.

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

PARK, T.; CASELLA, G. The Bayesian LASSO. *Journal of the American Statistical Association* 103(482), 681–686, 2008.

PASTINA, M. M.; MALOSETTI, M.; GAZAFFI, R.; MOLLINARI, M.; MARGARIDO, G. R.; OLIVEIRA, K. M.; PINTO, L. R.; SOUZA, A. P.; van EEUWIJK, F. A.; GARCIA, A. A. F. A mixed model QTL analysis for sugarcane multiple-harvest-location trial data. *Theoretical and Applied Genetics*, v. 124, n. 5, p. 835-849, 2012.

PÉREZ, P.; DE LOS CAMPOS, G. Genome-wide regression & prediction with the BGLR statistical package. *Genetics*, p. genetics. 114.164442, 2014.

PIPERIDIS, G.; PIPERIDIS, N.; D'HONT, A. Molecular cytogenetic investigation of chromosome composition and transmission in sugarcane. *Molecular Genetics and Genomics*, New York, v. 284, n. 1, p. 65-73, 2010.

POWELL, J. E.; HENDERS, A. K.; McRAE, A. F.; KIM, J.; HEMANI, G.; MARTIN, N. G.; DERMITZAKIS, E. T.; GIBSON, G.; MONTGOMERY, G. W.; VISSCHER, P. M. Congruence of additive and non-additive effects on gene expression estimated from pedigree and SNP data. *PLoS genetics*, v. 9, n. 5, p. e1003502, 2013.

R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>, 2018.

RABOIN, L. M.; OLIVEIRA, K. M.; LECUNFF, L.; TELISMART, H.; ROQUES, D.; BUTTERFIELD, M.; HOARAU, J. Y.; D'HONT, A. Genetic mapping in sugarcane, a high polyploid, using bi-parental progeny: identification of a gene controlling stalk colour and a new rust resistance gene. *Theoretical and Applied Genetics*, Heidelberg, v. 112, n. 7, p. 1382-1391, 2006.

RABOIN, L. M.; PAUQUET, J.; BUTTERFIELD, M.; D'HONT, A.; GLASZMANN, J. C. Analysis of genome-wide linkage disequilibrium in the highly polyploidy sugarcane. *Theoretical and Applied Genetics*, Heidelberg, v. 116, n. 5, p. 701-714, jan. 2008.

RESENDE, M. D. V. *Genética Quantitativa e Seleção no Melhoramento de Plantas Perenes e Animais*. Colombo: Embrapa Florestas, 330 p., 2008.

RESENDE, M. D. V.; LOPES, P. S.; SILVA, R. L.; PIRES, I. E. Seleção genômica ampla

(GS) e maximização da eficiência do melhoramento genético. Pesquisa Florestal Brasileira, Colombo –PR, n. 56, p. 63, 2008.

RESENDE, M. D.V.; RESENDE JÚNIOR, M. F. R.; AGUIAR, A. M.; ABAD, J. I. M.; MISSIAGGIA, A. A.; SANSALONI, C.; PETROLI, C.; GRATTAPAGLIA, D. Computação da seleção genômica ampla (GS). 1a ed. Colombo: Embrapa Florestas 2010. 79p.

SCHAEFFER, L. R. Strategy for applying genome - wide selection in dairy cattle. Journal of Animal Breeding and Genetics, Massachusetts, v. 123, n. 4, p. 218-223, 2006.

SIMS, R. E., HASTINGS, A., SCHLAMADINGER, B., TAYLOR, G., & SMITH, P. Energy crops: current status and future prospects. Global Change Biology, Massachusetts, v. 12, n. 11, p. 2054-2076, 2006.

SLATKIN, M. Linkage disequilibrium in growing and stable populations. Genetics, Heidelberg, v. 137, n. 1, p. 331-336, 1994.

SOLBERG, T. R.; SONESSON, A. K.; WOOLLIAMS, J. A.; MEUWISSEN, T. H.. Genomic selection using different marker types and densities. Journal of Animal Science, v. 86, n. 10, p. 2447-2454, 2008.

SPIEGELHALTER, D. J.; BEST, N. G.; CARLIN, B. P.; VAN DER LINDE, A. Bayesian measures of model complexity and fit. J. R. Stat. Soc. Series B Stat. Methodol. 64(4): 583–639, 2002.

THODAY, J. M. Location of polygenes. Nature, v. 191, p. 368-370, 1961.

VANRADEN, P. M. Efficient Methods to Compute Genomic Predictions. Journal of dairy science, vol. 91, n.11, p. 4414-4423, 2008.

WICKHAM, H. The Split-Apply-Combine Strategy for Data Analysis. Journal of Statistical Software, 40(1), 1-29. URL <http://www.jstatsoft.org/v40/i01/>, 2011.

ZEILEIS, A.; GROTHENDIECK, G. zoo: S3 Infrastructure for Regular and Irregular Time Series. Journal of Statistical Software, 14(6), 1-27. doi:10.18637/jss.v014.i06, 2005.

ZHANG, J.; NAGAI, C.; YU, Q.; PAN, Y. B.; AYALA-SILVA, T.; SCHNELL, R. J.; COMSTOCK, J.C.; ARUMUGANATHAN, A. K.; MING, R. Genome size variation in three *Saccharum* species. Euphytica, Wageningen, v. 185, n. 3, p. 511-519, 2012.

ZHANG, Z.; DING, X.; LIU, J.; KONING, D-J.; ZHANG, Q. ZHANG, Genomic selection

for QTL-MAS data using a trait-specific relationship matrix. In: BMC proceedings. BioMed Central, p. S15, 2011.

CAPÍTULO 5

GENOMIC PREDICTION OF AUTOTETRAPLOIDS; THE INFLUENCE OF ALLELE DOSAGE, NON-ADDITIVE EFFECTS AND CONTINUOUS GENOTYPES¹

Ivone de Bem Oliveira^{2,3}, Marcio F. R. Resende⁴, Luis Felipe Ventorin Ferrão², Rodrigo R. Amadeu², Jeffrey Endelman⁵, James Olmstead^{2,6}, Matias Kirst⁷, Alexandre Siqueira Guedes Coelho³, and Patricio R. Munoz^{2*}

¹Capítulo elaborado conforme as normas do periódico científico Genetics;

² Blueberry Breeding and Genomics Lab, Horticultural Sciences Department, University of Florida, Gainesville, FL, USA

³ Plant Genetics and Genomics Lab, Agronomy College, Federal University of Goiás, GO, Brazil

⁴ Sweet Corn Genomics and Breeding, Horticultural Sciences Department, University of Florida, Gainesville, FL, USA

⁵ Horticulture Department, University of Wisconsin, Madison, WI, USA

⁶ Driscoll's, Watsonville, CA, USA

⁷ Forest Genomics Laboratory, School of Forest Resources and Conservation, University of Florida, Gainesville, FL, USA

*Corresponding author

5 GENOMIC PREDICTION OF AUTOTETRAPLOIDS; THE INFLUENCE OF ALLELE DOSAGE, NON-ADDITIVE EFFECTS AND CONTINUOUS GENOTYPES

5.1 ABSTRACT

In this study our goal was to investigate the influence of different relationship matrices, ploidy information and non-additive effects in the estimation of genetic variance components, and on prediction of phenotypes using genome-wide selection (GS) for autotetraploid species. For this nine linear mixed models were applied to predict breeding values in a real blueberry breeding (BB) population based on phenotypic, pedigree, and high-density marker data for 1,847 individuals. Our results show that marker-based matrices could lead to better predictions than pedigree for most of the traits. The obtained accuracies are encouraging for application of GS for BB. With our proposal, the application of GS would reduce three year in the release of cultivars with a mean expected genetic gain representing 1.94X over the conventional BB scheme (identity matrix). Besides it, our results indicate that the importance of dosage in GS will vary depending on the trait being analyzed. And that the addition of non-additive effects has lead to a better fit of the models, as well as showed that the non-additive effects represented more than 20% of the total variance for most of the traits. Going beyond previous studies for polyploid species we also proposed the use of continuous data to perform GS, instead of assuming standardization. This would reduce the computational time and it avoids problems associated with misclassification of genotypic classes.

5.2 INTRODUCTION

Conceived by Meuwissen et al. (2001) to increase the efficiency and speed up the selection process in breeding programs, genome selection (GS) is used to predict the simultaneous effects of a wide range of molecular markers distributed across the genome. It does not intend to identify individual locus effects, but to capture the combined effect of the causal polymorphisms involved in the genetic variation of a quantitative trait. It is

based in the premise of linkage disequilibrium existence between causal polymorphisms and the markers used in the analysis (Meuwissen et al., 2001; Endelman et al., 2011; Zhang et al., 2011; De Los Campos et al., 2013).

The advent of next generation sequencing methodologies, concomitantly to the development of genetic and statistical analysis tools, has provided significant genetic gains for diploid species (i.e., Desta & Ortiz, 2014). Despite the successes, its application in polyploid crops remains a challenge for several reasons, including the complex genomic architecture (Ramsey et al., 1998; Comai et al., 2005; Grandke et al., 2016) and unavailability of methods and software to handle polyploids.

Polyploidy events are not an exception in plants, about 70% of the Angiosperms and about 95% of the Pteridophytes underwent at least one polyploidization event. Since specification via polyploidy can generate new phenotypic variations, this phenomena is considered a powerful evolutionary source (Hieter and Griffiths, 1999). Despite the important role of polyploidization, its effects on inheritance of many agronomic traits and population genetics are still poorly understood compared with diploid species (Dufresne et al., 2014). This especially holds true for autopolyploids, originated via doubling of homologous chromosome set. In particular, the complex nature of autopolyploid data is due to the presence of multiple alleles per locus, larger number of genotypic classes, possibility of non bivalent pairing and poor knowledge of chromosome behavior during meiosis (Slater et al., 2013; Dufresne et al., 2014; Mollinari et al., 2015).

The first GS studies addressing polyploids considered diploid genetic models to circumvent the complexity of allelic dosage definition (i.e., the number of copies of each allele at a given polymorphic locus). Promising results have been reported (Gouy et al., 2013; Annicchiarico et al., 2015; Ashraf et al., 2016; Ashraf et al., 2016b; Slater et al., 2016), even though several simplified assumptions were mostly used for the genetic and statistical inferences (Garcia et al., 2013). The fact that mostly GS research addressing polyploidy have considered oversimplified concepts opens possibilities for more appropriate methods.

Polyploidy can be enrolled in complex adaptive biological systems. It could interfere in the phenotypic expression due to alteration in allelic dosage, caused by the additive effect of the same alleles (dosage effect), or for more complex interactions between loci encompassing non-additive effects, since complex traits are manifestations of intricate gene interaction networks (Osborn et al., 2003). In this scenario, the inclusion of

allelic dosage information may improve GS models results (i.e., better fit, increase of accuracy), by assuming more realistic interpretation of the effects for each of the possible genotypic classes. Although the evidence of dosage effects in the expression of important economic traits exist (Guo et al., 1995; Birchler et al., 2001; Adams et al., 2003; Osborn et al., 2003), few studies linking dosage effects to phenotype variation have been reported in autopolyploid species (Nyine et al., 2018). It is interesting to note that genotype classification is one of the major challenges for polyploids. Studies about genotyping calling evaluation for polyploids showed that no method works properly (Voorrips et al., 2011; Grandke et al., 2016). Unless high sequencing coverage is used (60x to 80x; Uitdewilligen et al. (2013).

Among the factors that may influence GS analysis are the non-additive effects. Under theoretical conditions, the additive and non-additive effects are orthogonal (Cockerham, 1954; Kempthorne, 1954). However, in breeding populations these effects may interact, preventing its separation, and resulting in a misclassification of non-additive effects as additive (Lynch & Walsh, 1998; Hill, 2010; Muñoz et al., 2014). Modeling additive and non-additive effects independently in the analysis might improve the model fitting and variance estimations, resulting in higher predictive performance (De Almeida Filho et al., 2016, Muñoz et al., 2014).

Besides of GS potential to reshape breeding programs and to achieve faster selection gains, one major contribution of the present research is to discuss how to consider an appropriate predictive model in the context of autopolyploid, using *Vaccinium corymbosum* (southern highbush blueberry, SHB) as a model. The cultivated southern highbush blueberry is an autotetraploid, presenting $2n = 4X = 48$ chromosomes (Lyrene et al., 2002). Inbreeding depression is strong in SHB and population improvements have been achieved by long-term recurrent phenotypic selection alongside with long testing phase and slow genetic gain per time (Lyrene, 2008).

Here, our goal is to investigate the influence of pedigree and genomic relationship matrices, ploidy information, and non-additive effects on prediction of SHB phenotypes. In doing so, this will accelerate autotetraploid breeding by improving the capacity to predict complex traits with genome selection models. Our results provide the first proposition of genomic selection in blueberry. In addition, a new methodology that uses the continuous genotypes (i.e. ploidy assumption-free) instead of a standardization model is proposed.

5.3 MATERIAL AND METHODS

5.3.1. Data

The field data used in this study comprises one cycle of a recurrent selection population of 1,847 Southern Highbush blueberries (SHB) developed by the UF breeding program (UF BB). This population was originated from 124 controlled crossing of 146 selected parents. Phenotypic data of eight yield and fruit quality related traits were collected during two production seasons (2014 and 2015). Yield (1-9 rating scale), weight (g), firmness (g mm⁻¹ of compression force), scar diameter (mm), yield, fruit diameter (mm), flower bud density (count), soluble solids content (°Brix) and pH were evaluated. The last three traits were phenotyped only for one year, in which soluble solids content and pH were phenotyped in 2014, and flower bud in 2015.

Genomic DNA was extracted and genotyped using capture-seq protocol by Rapid Genomics[®] (Gainesville, FL, USA). Genotypes were called using FreeBayes v.1.0.1 (Garrison and Marth, 2012) considering diploid (-p 2) and tetraploid (-p 4) options. The genotypes for the diploid calling were coded as 0 (AA), 1 (AB), or 2 (BB). The genotypes for the tetraploid calling were coded as 0 (AAAA), 1 (AAAB), 2 (AABB), 3 (ABBB), and 4 (BBBB). A third parameterization (assumption-free) which considered the ratio of the reference allele to the alternative allele using the raw reads was also calculated; these varied continuously between 0-1. SNPs were filtered considering: i) minimum sequencing depth of 40; ii) minimum genotype quality score of 10; iii) only biallelic locus; iv) maximum missing data of 0.5; v) minor allele frequency of 0.05. After filter steps all the matrices were filtered to present the same SNP position for the same individuals. A total of 85,973 SNP were considered in the genomic selection analysis. Further information regarding population constitution and genotyping protocol are described in Amadeu et al. (2016), Cellon et al., (2018) and Ferrão et al., (2018).

5.3.2. Models

Nine linear single-trait Bayesian linear models (one-step) were applied to predict breeding values for each individual in the population (i.e., 1 or 2; Supplementary material).

$$y = 1\mu + Xb + Z_1c + Z_2r + Z_3a + Z_4bxa + e \quad (1)$$

Where y is a vector of the phenotypic values of the trait being analyzed, μ is the population overall mean, b is the fixed effect of year, c is the random effect of i th

column position in the field $\sim N(0, I\sigma_c^2)$, r is the random effect of the i th row position in the field $\sim N(0, I\sigma_r^2)$, a is the random effect of genotype $\sim N(0, G_a\sigma_a^2)$, where G_a was replaced by the different additive relationship matrices as presented in the next section. The bxa is the random effect of the year by genotype interaction $\sim N(0, I\sigma_{bxa}^2)$, and e is the random residual effect $\sim N(0, I\sigma_e^2)$. Row and column effects were considered nested within year only for the traits evaluated in two years. For traits measured a single year, the same equation 1 was used without the year effect or the interactions with year. Non-additive effects (NAD or nad) were added to the model as an extra factor ($Z_5 nad$) in the model to account for these effects, considering $nad \sim N(0, G_{nad}\sigma_{nad}^2)$, where G_{nad} was replaced by the different non-additive relationship matrices as presented in the next section. The random variables variance components were: additive (σ_a^2), non-additive (σ_{nad}^2), column (σ_c^2), row (σ_r^2), year-by-genotype interaction (σ_{bxa}^2), and residual (σ_e^2). X , Z_1 , Z_2 , Z_3 , and Z_4 were incidence matrices for year, column, row, genotype, and year by genotype interaction, respectively. The narrow and broad-sense heritabilities were estimated as follow:

$$h^2_{two\ years\ data} = \frac{\sigma_a^2\ (or\ \sigma_a^2 + \sigma_{nad}^2)}{\sigma_a^2\ (or\ \sigma_a^2 + \sigma_{nad}^2) + \sigma_c^2 + \sigma_r^2 + \sigma_{bxa}^2 + \sigma_e^2} \quad (3)$$

$$h^2_{one\ year\ data} = \frac{\sigma_a^2\ (or\ \sigma_a^2 + \sigma_{nad}^2)}{\sigma_a^2\ (or\ \sigma_a^2 + \sigma_{nad}^2) + \sigma_c^2 + \sigma_r^2 + \sigma_e^2} \quad (4)$$

5.3.3. Relationship matrices

In order to analyze the influence of relationship matrices, ploidy, and non-additive effect in the predictive ability, we performed analyses considering different approaches to modeling the genotypic values in autotetraploid species (Table 1). The factors tested were, i) For the construction of the relationship matrices: pedigree, genomic information, or no relationship information (identity matrix); ii) To compare ploidy effects: diploid, tetraploid, and assumption-free (continuous data) were used, and iii) To use the information of non-additive effects; we considered either including or excluding the non-additive factor in the model.

The methods used to obtain the relationship matrices are shown in the Table 1. The pedigree-based relationship matrices (A) were built considering a diploid model (Henderson, 1976) and an autotetraploid model without double-reduction (Kerr et al.

2012). The marker-based relationship matrices (G) were based on the incidence matrices of markers effects (X) according VanRaden (2008), the G4 additive matrix was computed using the VanRaden (2008) methodology adapted for Ashraf et al. (2016). Different assumption can be made regarding the marker allele dosage in autotetraploids (Table 2). We built the X matrices under three assumptions regarding the additive marker allele dosage effect: i) a pseudo-diploid model, where all the heterozygous genotypes are merged into one effect (data coded as 0, 1, and 2); ii) an additive autotetraploid model, where each genotype had a different value, and the additive allelic effect is considered (data coded as 0, 1, 2, 3, and 4); and iii) an a ploidy assumption-free parameterization based on the ratio of reference and alternative allele count (continuous parameterization between 0-1).

Two assumptions were made regarding the non-additive marker effects: i) a pseudo-diploid model, where only effects for the heterozygotes were assumed and all the classes presented an equal effect (data coded as 0, and 1; Vitezica et al., 2013); ii) a full autotetraploid model, where five different effects are possible for each marker (Slater et al., 2016). Note that in this last methodology the additive and non-additive effects are confounded, and thus its separation is not possible (Table 1). For the construction of the relationship matrices based on marker data the missing genotypes were substituted by the mean. To obtain the relationship matrices the R package AGHmatrix v. 0.0.3003 (Amadeu et al., 2016) was used.

Table 1. Models used to compare the influence of relationship matrices, ploidy and dominance effect in the prediction of breeding values for blueberry

Relationship matrix	Model	Additive effects		Non-additive effects	
		Ploidy	Methodology	Ploidy	Methodology
Identity matrix	I	no	no	no	no
Pedigree-based	A2 _a	2	Henderson (1976)	no	no
	A2 _{a+nad}	2	Henderson (1976)	2	Cockerham (1954)
	A4 _a	4	Kerr et al. (2012)	no	no
Marker-based	G2 _a	2	VanRaden (2008)	no	no
	G2 _{a+nad}	2	VanRaden (2008)	2	Vitezica et al (2013)
	G4 _a	4	VanRaden (2008)	no	no
	G4 _{a+nad}	4	Slater et al (2016)	4	Slater et al (2016)*
	Gr _a	ratio	VanRaden (2008)	no	no

* Both additive and non-additive effects are considered in the same matrix

Table 2. Genotype codes for marker-allele dosage effects assuming different assumptions. Adapted from Slater et al. (2016).

Genotype	Additive Pseudo Diploid	Additive Autotetraploid	Quantitative*	Non-additive Pseudo Diploid	Full Autotetraploid
AAAA	0	0	value+se	0	1 0 0 0 0
AAAB	1	1	value+se	1	0 1 0 0 0
AABB	1	2	value+se	1	0 0 1 0 0
ABBB	1	3	value+se	1	0 0 0 1 0
BBBB	2	4	value+se	0	0 0 0 0 1

Adapted from Slater et al. (2016)

* continuous values with no determination of a standardization class, the values+standard error (se) are the correspondent distribution of continuous values considering the classes in the additive tetraploid model

5.3.4. Model Implementation

The nine models described above were fitted using the R package (R Core Team, 2018) BGLR v. 1.0.5. (de los Campos and Pérez-Rodríguez, 2016). The predictions were based on 30,000 iterations of the Gibbs sampler, in which 5,000 were taken as burn-in, and a thinning of 5 was used to avoid autocorrelation among the sampled points. The number of iterations, burn-in, and thinning interval parameters were evaluated to define the final values used in the analysis (Supplementary Figure 1). The Bayesian Ridge Regression (BRR) prior was used for all analyses of I, A, and G matrices, leading to perform BLUP, P-BLUP, and G-BLUP methodologies, respectively. Default hyper-parameters were assumed (see Perez & de los Campos, 2014). A convergence analysis was performed in order to check if the number of iterations and burning were effective to properly estimate the genetic parameters.

The expected genetic gain was obtained for each trait and model combination. For this we used the following formula: $\Delta G = (PA \cdot \sigma_a \cdot i) / L$. Where PA is the phenotypic predictive ability, σ_a is the square root of additive genetic variance in the population, i is the selection intensity, and L is the breeding cycle length. The selection intensity (i) was considered constant for all methods. We considered a three years reduction in the breeding cycle (L) for the marker-based methodologies, going from 12 (Cellon et al., 2018) to nine years in total.

5.3.5. Validation

A 10-fold cross validation scheme was applied to compute the predictive abilities of the models. Since each validation group might have a different mean (Resende et al., 2012), the phenotypic predictive ability (PA) was obtained as the Pearson correlation coefficient between the best linear unbiased estimation values (eBLUEs) obtained by considering all the variables in the equations 1 and 2 as fixed (i.e., Least Square means estimations; LSMeans) and the cross-validated breeding values (BV) predicted by the GS models. The genotypic predictive ability values were obtained as the ratio between the phenotypic predictive ability and the heritability. The goodness of fit measures was used to compare the models.

5.3.6. Population Genetics Analysis

Linkage disequilibrium (LD) decay was computed considering all marker-based scenarios ($G2_a$, $G4_a$, and $G2_r$). Pearson correlation tests (r^2) were performed for pairwise LD estimation within scaffolds. All scaffolds were pooled to plot a genome-wide LD decay. The LD decay over genetic distance was determined as the mean distance associated with an empirical LD threshold of $r^2 = 0.2$. The diversity captured from each relationship matrix was compared by principal component analysis (PCA). PCA analyses were performed using the R package *ade4* v. 1.3-1 (Jombart & Ahmed, 2011).

In order to compare the information present in the marker matrices we also evaluated the gene diversity in the population. For this, we obtained the ratio between the number of heterozygotes genotypes (i.e., code 1 for the diploid matrix and codes 1, 2 and 3 in the tetraploid matrix) and the total number of loci. To estimate the ratio for the continuous matrix, a density plot was obtained and empirical limits were established. For this we obtained the distribution of the raw value considering the classes of the tetraploid standardization model, the mean and the confidence intervals were used to determine the empirical values. We also compared the attribution of classes in the diploid and tetraploid assumptions with the continuous data.

5.4 RESULTS

5.4.1. Population Genetics Analysis

The pairwise marker associations' probabilities exhibited a similar pattern for all the genomic matrices. The LD decay estimated in the blueberry breeding population presented significant correlation between markers 88.3 Kb apart and 92.6 Kb apart for the diploid and tetraploid standardizations, respectively, and 98.2 Kb apart for continuous model (Figure 1A-C). No significantly difference was observed among the distribution of distances considering the empirical threshold of $r^2 = 0.2$ for the diploid and tetraploid standardizations as well for the continuous genotypes (Figure 1I).

Considering the diversity of the matrices, genotypes assumed as homozygotes in the diploid standardization were classified as one of the possible heterozygote classes in the tetraploid standardization. As a result of this process, the tetraploid standardization (43.61% of heterozygotes) presented 34.47% higher heterozygosity than the diploid standardization (32.43% of heterozygotes; Figure 1D-F). In order to compare the diversity values observed for the diploid and tetraploid standardizations, empirical thresholds were established to determine the proportion of “heterozygotes” in the continuous genotypes using the mean and the confidence interval of the raw values considering the classes in the tetraploid matrix. For this, values equal or below 0.058 and equal or above 0.908 were established as “homozygotes” classes. Considering these limits 61.59% of the genotypes were considered heterozygotes, presenting 89.92% and 41.23% more heterozygotes than the diploid and the tetraploid standardization, respectively.

Similar results were obtained for the PCA analysis of the marker-based matrices (Supplementary Figure 1). The first two PCA components for the marker-based (G) matrices were very consistent across all matrices, explaining approximately 20% of the variation (i.e. G2 matrix captured 20.60% of the variability, while G4 captured 21.71% and GR captured 23.36%, Supplementary Figure 1A-C, respectively). The PCA results were consistent between pedigree methodologies as well. Based on pedigree, approximately 85% of the variation was explained (i.e., 85.61% of the variability was explained for the A2 matrix and 84.49% was explained for the A4 matrix, Supplementary Figure 1E-G, respectively).

The results obtained in the PCA analysis did not justify a stratified sampling to compose the cross-validation populations, since no evidence of sub-population structure was detected for any of the relationship matrices. The first two axis of the PCA analysis for

the pedigree-based matrix explained 16.21% of the variance (Supplementary Figure 1G). For the marker-based non-additive relationship matrix (G_{2a+nad}) the first two PCA axes captured a big part of the variation (95.06%, Supplementary Figure 1H). For the marker-based additive plus non-additive tetraploid relationship matrix (G_{4a+nad}) only 4.22% of the variance was explained by the first two axes in the PCA analysis.

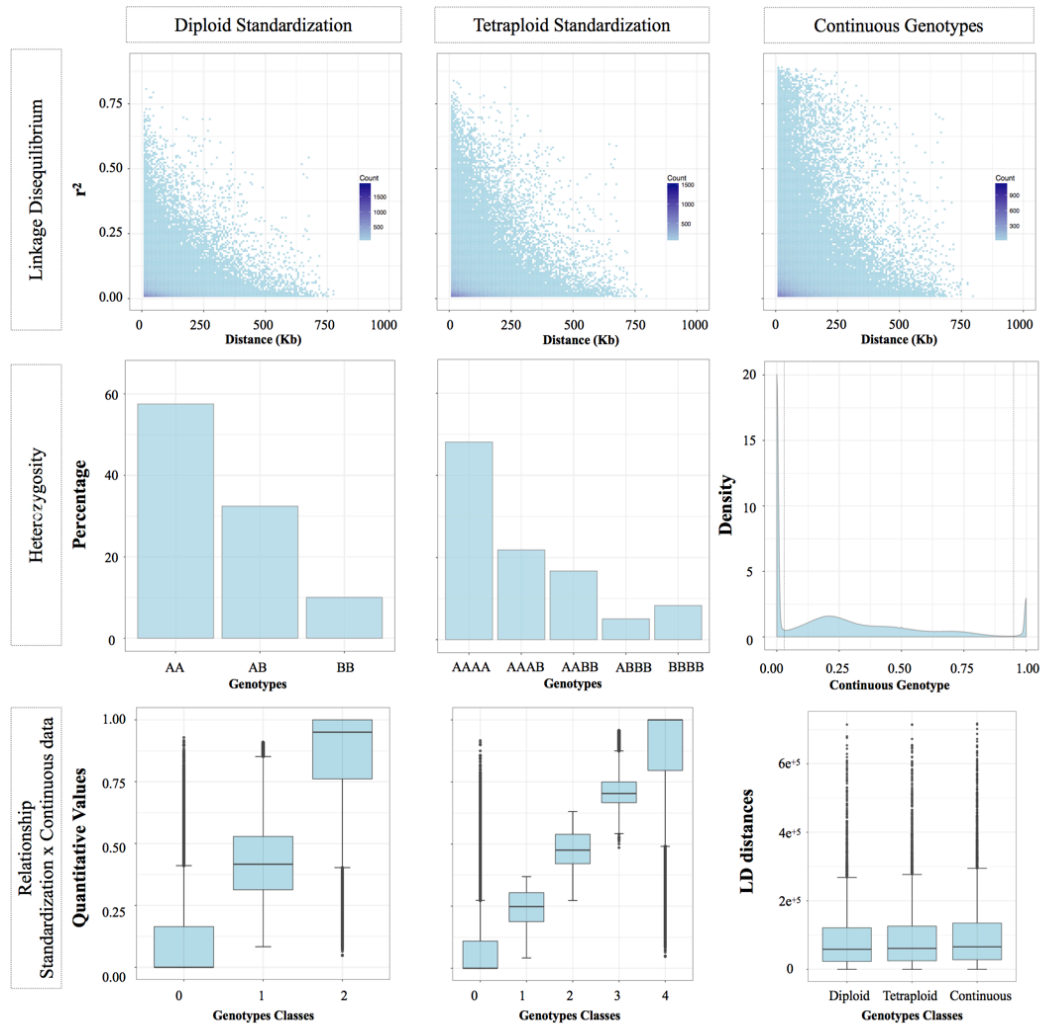


Figure 1. Plot of the genome-wide LD decay considering LD the estimation within scaffolds for the diploid (A), tetraploid (B) and continuous genotypes (C); Heterozygosity observed in the diploid standardization (D), tetraploid standardization (E), and the heterozygosity empirically established for the continuous genotypes, assuming the limits of $0.058 \geq X \leq 0.908$ (F); Relationship between continuous genotypes and the classes presented in the diploid standardization (G), Relationship between continuous genotypes and the classes assumed in the tetraploid standardization (H); Boxplot of the distance distributions of linkage disequilibrium considering the empirical threshold of $r^2 = 0.2$ for the diploid and tetraploid standardizations and for the continuous genotypes (I).

5.4.2. Estimates

The estimate values obtained for the genetic parameters are summarized in Table 3. A wide range of variance was observed within a given parameter for the different methodologies, most of the values were significantly different from each other (Table 3, Supplementary Table 1). All the traits presented additive genetic variance significantly higher than zero. The non-additive variance estimated values were also significantly higher than zero for all the traits. The pedigree methods estimated higher non-additive variance values, except for brix. The narrow-sense heritability values range from 0.07 (size) to 0.57 (weight). After including the non-additive effect, an average decrease in the $\bar{h}^2$ of 49% was observed for the pedigree-based model ($A2_{a+nad}$), and an average decrease of 17% was observed for the marker-based model ($G2_{a+nad}$).

The highest estimated of narrow-sense heritability ($\bar{h}^2$) ranges were obtained for weight ($A2_{a+nad} = 0.104$ to $A2_a = 0.574$) and yield ($A2_{a+nad} = 0.068$ to $A2_a = 0.477$). The decrease in the narrow-sense heritability was associated with the estimated values for the additive variance. This is, in the $\bar{h}^2$ for these traits smaller values attributed to the additive variance were compensated by the high values estimated for the non-additive variance (weight: $A2_{a+nad}$: $\bar{\sigma}_a^2 = 0.062$, $\bar{\sigma}_{nad}^2 = 0.238$; yield: $A2_{a+nad}$: $\bar{\sigma}_a^2 = 0.073$, $\bar{\sigma}_{nad}^2 = 0.340$).

Total genetic variance also decreased with the addition of the matrix to compute the non-additive effects (weight: $A2_a$: $\bar{\sigma}_g^2 = 0.403$ versus $A2_{a+nad}$: $\bar{\sigma}_g^2 = 0.300$; yield: $A2_a$: $\bar{\sigma}_g^2 = 0.549$ versus $A2_{a+nad}$: $\bar{\sigma}_g^2 = 0.413$). A decrease in the total genetic variance also happened for the marker-based method, but with a smaller effect than in the pedigree-based method, i.e. weight: $G2_a$: $\bar{\sigma}_g^2 = 0.344$ versus $G2_{a+nad}$: $\bar{\sigma}_g^2 = 0.338$, and yield: $G2_a$: $\bar{\sigma}_g^2 = 0.470$ versus $G2_{a+nad}$: $\bar{\sigma}_g^2 = 0.459$.

5.4.3. Identity, Pedigree-based, and Marker-based Matrices

The use of pedigree-based relationship matrices generated higher prediction accuracies values for all the traits when compared with models that assumed absence of genetic relationship across individuals (Identity matrix). Unlike the identity matrix, the use of pedigree-based matrix assumes that there is relationship among individuals. This helps in the capture of genetic variance and improves breeding value predictions (Figure 2, Table 3, Supplementary Table 1). The phenotypic predictive ability obtained for the pedigree

methods in 2014 yielded values from 0.20 (flower bud) to 0.49 (fruit firmness), the smaller values were observed for the traits with lower heritability (i.e. pH, brix, and flower bud). The differences can be better observed with the data from 2015 (Figure 2B). Since blueberry presents a juvenility period (Lyrene, 1981), better predictions are expected as plants age. For 2015, the PA results for the models that assumed absence of genetic correlation values were 0.36, 0.38, and 0.42, for fruit weight, fruit scar, and fruit firmness, respectively. The PA values obtained for the same traits with pedigree information were 0.40, 0.45, and 0.49, respectively.

However, better results were obtained for most of the traits when a realized relationship matrix (marker-based) was used. The use of molecular data yielded phenotypic predictive ability values from 0.19 (flower bud) to a 0.51 (fruit firmness). As it occurred with pedigree methods, lower values were also obtained for traits that presented lower heritability and better results were observed for the second year. When compared to the pedigree methods, an increase of 16.20% in the PA values was observed for firmness in 2014 and an increase of more than 11% was observed for fruit diameter and yield in 2015 when marker data was used.

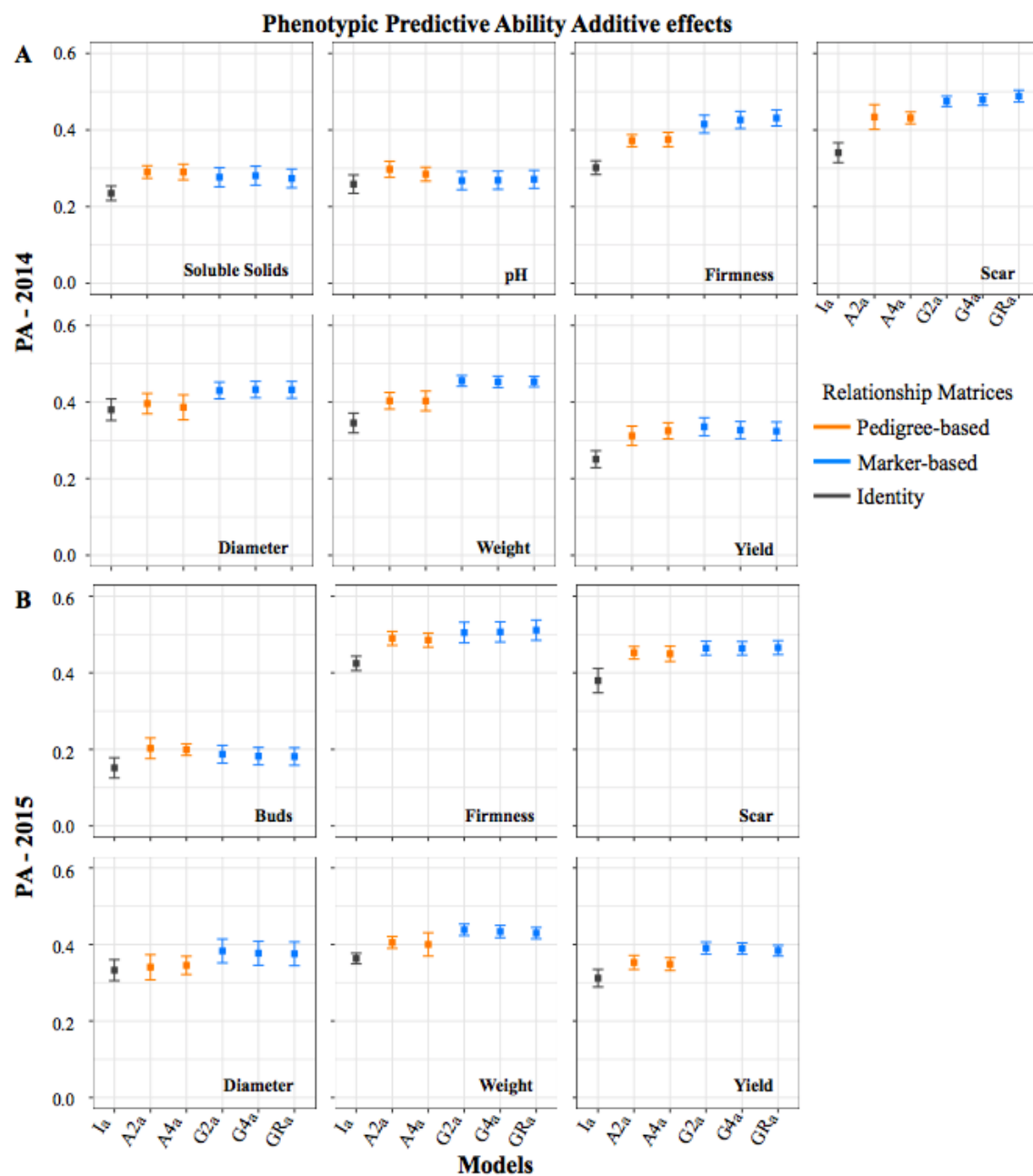


Figure 2. Predictive abilities for the single trait Bayesian analyses considering the addition of ploidy information and continuous genotypes in the prediction of breeding values for 1,847 blueberry individuals

Table 3. Genetic parameters estimated for eight yield and fruit-related traits analyzed with nine linear mixed models, considering the use of ploidy information, non-additive effects and continuous genotypes

Trait	Relationship matrix	Additive Variance	Non-additive Variance	Narrow-sense Heritability	Broad-sense Heritability	Expected Genetic Gain 2014	Expected Genetic Gain 2015	Posterior Mean of the Log Likelihood
Soluble Solid (°Brix)	I	0.806 (0.014) c	-	0.257 (0.005) b	0.257 (0.005) d	0.018 (0.001) b	-	-2976.310 (7.608) c
	A2	0.777 (0.003) d	-	0.239 (0.001) c	0.239 (0.001) e	0.021 (0.001) ab	-	-3149.850 (0.759) f
	A2a+nad	0.533 (0.008) g	0.484 (0.011) b	0.166 (0.002) f	0.317 (0.002) c	0.018 (0.001) b	-	-3046.270 (2.587) d
	A4	0.764 (0.003) d	-	0.236 (0.001) c	0.236 (0.001) e	0.021 (0.002) ab	-	-3148.067 (0.690) f
	G2	0.848 (0.002) b	-	0.262 (0.001) b	0.262 (0.001) d	0.028 (0.003) a	-	-3106.762 (0.518) e
	G2a+nad	0.644 (0.006) f	0.683 (0.007) a	0.197 (0.002) e	0.405 (0.001) b	0.026 (0.002) ab	-	-2917.861 (1.455) b
	G4	0.673 (0.002) e	-	0.215 (0.001) d	0.215 (0.001) f	0.026 (0.002) ab	-	-3142.493 (0.536) f
	G4a+nad	0.931 (0.003) a	-	0.428 (0.002) a	0.428 (0.002) a	0.029 (0.002) a	-	-2332.596 (6.185) a
	Gr	0.546 (0.001) g	-	0.174 (0) f	0.174 (0) g	0.022 (0.002) ab	-	-3195.892 (0.246) g
	I	0.053 (0.001) b	-	0.253 (0.006) b	0.253 (0.006) d	0.005 (0) bc	-	-592.837 (10.332) b
pH	A2	0.052 (0) b	-	0.241 (0.001) c	0.241 (0.001) e	0.006 (0) abc	-	-764.731 (0.984) d
	A2a+nad	0.035 (0) e	0.035 (0) a	0.163 (0.002) e	0.325 (0.001) c	0.005 (0) c	-	-650.910 (1.684) c
	A4	0.052 (0) b	-	0.238 (0.002) c	0.238 (0.002) e	0.005 (0) abc	-	-762.886 (1.265) d
	G2	0.052 (0) b	-	0.241 (0.001) c	0.241 (0.001) e	0.007 (0.001) ab	-	-769.534 (0.507) d
	G2a+nad	0.043 (0) c	0.031 (0) b	0.196 (0.001) d	0.340 (0.002) b	0.006 (0.001) abc	-	-654.735 (1.948) c
	G4	0.040 (0) d	-	0.191 (0) d	0.191 (0) f	0.006 (0.001) abc	-	-805.822 (0.341) e
	G4a+nad	0.063 (0) a	-	0.432 (0.002) a	0.432 (0.002) a	0.007 (0) a	-	11.602 (5.970) a
	Gr	0.035 (0) e	-	0.165 (0) e	0.165 (0) g	0.006 (0) abc	-	-840.501 (0.245) f

Table 3. (cont.)

Trait	Relationship matrix	Additive Variance	Non-additive Variance	Narrow-sense Heritability	Broad-sense Heritability	Expected Genetic Gain 2014	Expected Genetic Gain 2015	Posterior Mean of the Log Likelihood
Flower Bud	I	2.133 (0.036) a	-	0.270 (0.005) b	0.270 (0.005) c	-	0.018 (0.003) ab	-4024.785 (7.671) b
	A2	1.247 (0.006) f	-	0.153 (0.001) fg	0.153 (0.001) gh	-	0.019 (0.002) ab	-4275.597 (0.446) f
	A2a+nad	0.916 (0.012) g	0.900 (0.01) a	0.113 (0.001) h	0.225 (0.001) e	-	0.016 (0.002) b	-4195.294 (0.605) d
	A4	1.232 (0.008) f	-	0.152 (0.001) g	0.152 (0.001) h	-	0.018 (0.001) ab	-4274.008 (0.682) f
	G2	2.106 (0.012) ab	-	0.251 (0.001) c	0.251 (0.001) d	-	0.003 (0.004) a	-4192.561 (1.202) d
	G2a+nad	1.798 (0.007) c	0.798 (0.009) b	0.214 (0.001) d	0.309 (0.001) b	-	0.027 (0.004) ab	-4128.736 (0.981) c
	G4	1.526 (0.007) d	-	0.188 (0.001) e	0.188 (0.001) f	-	0.025 (0.003) ab	-4244.712 (0.696) e
	G4a+nad	2.050 (0.015) b	-	0.342 (0.003) a	0.342 (0.003) a	-	0.021 (0.003) ab	-3660.190 (8.966) a
	Gr	1.315 (0.006) e	-	0.161 (0.001) f	0.161 (0.001) g	-	0.023 (0.003) ab	-4281.097 (0.404) f
	I	509.180 (0.247) h	-	0.275 (0) g	0.275 (0) h	0.567 (0.034) c	0.798 (0.036) e	-16181.026 (10.472) bc
Fruit Firmness	A2	806.908 (0.451) a	-	0.401 (0) a	0.401 (0) a	0.881 (0.037) b	1.160 (0.043) bc	-16183.975 (16.556) bc
	A2a+nad	580.448 (7.122) g	166.752 (5.416) a	0.299 (0.003) f	0.385 (0.001) c	0.747 (0.031) bc	0.983 (0.035) cde	-16235.048 (16.705) cd
	A4	786.601 (0.515) b	-	0.395 (0) b	0.395 (0) b	0.877 (0.044) b	1.135 (0.043) cd	-16189.390 (8.429) bc
	G2	725.192 (0.377) c	-	0.376 (0) c	0.376 (0) e	1.243 (0.070) a	1.511 (0.081) a	-16171.694 (8.989) b
	G2a+nad	640.490 (0.847) f	86.101 (0.819) b	0.335 (0) e	0.380 (0) d	1.162 (0.064) a	1.421 (0.076) ab	-16172.953 (17.285) b
	G4	659.584 (0.528) e	-	0.351 (0) d	0.351 (0) g	1.217 (0.064) a	1.446 (0.076) a	-16207.163 (8.355) bc
	G4a+nad	288.721 (0.131) i	-	0.192 (0) h	0.192 (0) i	0.683 (0.042) bc	0.889 (0.036) de	-16064.662 (8.104) a
	Gr	687.685 (0.444) d	-	0.354 (0) d	0.354 (0) f	1.257 (0.060) a	1.490 (0.077) a	-16283.119 (8.655) d
	I	0.086 (0) g	-	0.381 (0) e	0.381 (0) g	0.008 (0.001) d	0.009 (0.001) c	-351.218 (15.029) ab
	A2	0.139 (0) a	-	0.528 (0) a	0.528 (0) a	0.013 (0.001) b	0.014 (0.001) b	-395.541 (15.801) bcd
Fruit Scar	A2a+nad	0.081 (0.002) h	0.040 (0.001) a	0.333 (0.006) f	0.498 (0.001) d	0.010 (0.001) d	0.011 (0) c	-445.000 (13.182) de
	A4	0.135 (0) b	-	0.522 (0) a	0.522 (0) b	0.013 (0) bc	0.014 (0.001) b	-414.241 (8.501) cde
	G2	0.123 (0) d	-	0.500 (0) b	0.500 (0) c	0.018 (0.001) a	0.018 (0.001) a	-392.265 (8.084) bc
	G2a+nad	0.108 (0) f	0.012 (0) b	0.448 (0.001) d	0.499 (0) c	0.017 (0) a	0.017 (0.001) a	-455.980 (13.013) e
	G4	0.115 (0) e	-	0.479 (0) c	0.479 (0) f	0.018 (0.001) a	0.017 (0.001) a	-451.945 (9.439) e
	G4a+nad	0.046 (0) i	-	0.272 (0) g	0.272 (0) h	0.01 (0) cd	0.011 (0) c	-321.501 (9.031) a
	Gr	0.126 (0) c	-	0.494 (0) b	0.494 (0) e	0.019 (0.001) a	0.018 (0.001) a	-538.444 (8.049) f

Table 3. (cont.)

Trait	Relationship matrix	Additive Variance	Non-additive Variance	Narrow-sense Heritability	Broad-sense Heritability	Expected Genetic Gain 2014	Expected Genetic Gain 2015	Posterior Mean of the Log Likelihood
Fruit Diameter	I	2.236 (0.003) g	-	0.162 (0) g	0.162 (0) g	0.047 (0.003) cd	0.041 (0.003) de	-8142.071 (11.876) ab
	A2	3.647 (0.009) a	-	0.250 (0.001) a	0.250 (0.001) a	0.063 (0.004) bc	0.054 (0.005) bcd	-8147.133 (26.630) ab
	A2a+nad	0.980 (0.019) i	2.214 (0.018) a	0.071 (0.001) i	0.231 (0) d	0.033 (0.002) d	0.028 (0.003) e	-8155.193 (11.653) b
	A4	3.581 (0.007) b	-	0.247 (0) b	0.247 (0) b	0.061 (0.005) c	0.054 (0.004) bcd	-8141.580 (23.214) ab
	G2	3.428 (0.005) c	-	0.242 (0) c	0.242 (0) c	0.088 (0.004) a	0.079 (0.006) a	-8139.628 (13.531) ab
	G2a+nad	2.763 (0.006) f	0.736 (0.005) b	0.196 (0) f	0.248 (0) b	0.079 (0.004) ab	0.07 (0.006) abc	-8146.803 (12.573) ab
	G4	2.992 (0.004) d	-	0.216 (0) d	0.216 (0) e	0.083 (0.004) a	0.072 (0.006) ab	-8178.698 (13.929) bc
	G4a+nad	1.356 (0.001) h	-	0.111 (0) h	0.111 (0) h	0.054 (0.003) c	0.049 (0.004) cde	-8076.080 (13.800) a
	Gr	2.910 (0.004) e	-	0.207 (0) e	0.207 (0) f	0.082 (0.004) a	0.071 (0.006) ab	-8243.890 (13.647) c
	I	0.217 (0) g	-	0.374 (0) f	0.374 (0) g	0.013 (0.001) c	0.014 (0.001) c	-2215.277 (5.208) d
Fruit Weight	A2	0.403 (0) a	-	0.574 (0) a	0.574 (0) a	0.021 (0.001) b	0.021 (0.001) b	-2157.103 (6.119) bc
	A2a+nad	0.062 (0.003) i	0.238 (0.002) a	0.104 (0.005) h	0.505 (0.001) f	0.008 (0.001) d	0.008 (0) d	-2184.623 (7.475) cd
	A4	0.393 (0) b	-	0.568 (0) a	0.568 (0) b	0.021 (0.001) b	0.021 (0.002) b	-2136.886 (8.369) b
	G2	0.344 (0) d	-	0.535 (0) b	0.535 (0) c	0.030 (0.001) a	0.029 (0.001) a	-2152.251 (11.368) bc
	G2a+nad	0.297 (0.001) f	0.041 (0.001) b	0.469 (0.001) e	0.534 (0) c	0.027 (0.001) a	0.027 (0.001) a	-2178.658 (11.158) bcd
	G4	0.323 (0) e	-	0.513 (0) d	0.513 (0) e	0.029 (0.001) a	0.027 (0.001) a	-2220.552 (11.056) d
	G4a+nad	0.127 (0) h	-	0.296 (0) g	0.296 (0) h	0.017 (0.001) c	0.017 (0.001) c	-2041.896 (13.607) a
	Gr	0.352 (0) c	-	0.522 (0) c	0.522 (0) d	0.030 (0.001) a	0.028 (0.001) a	-2349.684 (10.547) e
	I	0.326 (0) g	-	0.310 (0) g	0.310 (0) g	0.012 (0.001) cd	0.015 (0.001) e	-3683.364 (12.452) b
	A2	0.549 (0) a	-	0.447 (0) a	0.447 (0) a	0.019 (0.002) ab	0.022 (0.001) bc	-3668.232 (16.919) b
Yield	A2a+nad	0.073 (0.003) i	0.340 (0.002) a	0.068 (0.002) i	0.382 (0) d	0.007 (0.001) d	0.008 (0) f	-3691.165 (11.919) b
	A4	0.536 (0) b	-	0.441 (0) b	0.441 (0) b	0.020 (0.001) ab	0.021 (0.001) cd	-3667.668 (15.766) b
	G2	0.470 (0) c	-	0.407 (0) c	0.407 (0) c	0.026 (0.002) a	0.030 (0.001) a	-3662.062 (18.679) b
	G2a+nad	0.363 (0.001) f	0.096 (0.001) b	0.321 (0.001) f	0.406 (0) c	0.023 (0.002) a	0.026 (0.001) ab	-3704.370 (25.635) b
	G4	0.421 (0) d	-	0.374 (0) d	0.374 (0) e	0.024 (0.002) a	0.028 (0.001) a	-3730.163 (18.397) b
	G4a+nad	0.179 (0) h	-	0.209 (0) h	0.209 (0) h	0.014 (0.001) bc	0.017 (0.001) de	-3566.362 (19.459) a
	Gr	0.411 (0) e	-	0.356 (0) e	0.356 (0) f	0.023 (0.002) a	0.027 (0.001) a	-3864.783 (18.584) c

5.4.4. Use of dosage information and continuous Genotypes

Our results indicate that the importance of dosage in GS will vary depending on the trait being analyzed. For example, in 2014 the PA results for fruit firmness, fruit scar, and fruit diameter were slightly better when the tetraploid standardization was applied as opposed to the diploid standardization (Figure 2, Supplementary Table 1). The use of relationship matrices derived of continuous genotype data (without assuming ploidy standardization) performed as well or better than the best marker-based models. This model yielded slightly better results in the BV prediction for fruit firmness in 2014 and 2015, and for scar in 2014 (Figure 2, Supplemental table 1).

5.4.5. Non-additive effects

In our study the addition of non-additive effects have not significantly increased the predictive abilities (Figure 3). However the insertion of these matrices allowed a better separation and estimation of the additive and non-additive effects, except for the full tetraploid model (G_{4a+nad}). When the pedigree matrix was used the estimated NAD effects represented over 50% of the total genetic variance for brix, flower buds and pH. For the other traits, higher values were estimated, for example: yield (83.32%), weight (79.33%), and fruit size (69.32%). Fruit firmness (22.32%) fruit scar (33.06%).

However the results for NAD estimated with the marker-based matrix also indicated that the non-additive effects represent an important portion of the genetic variance (Table 3). The estimated values for this component correspond to ~30% for the total genetic variation of flower bud, 42% for pH, approximately 20% for size and yield, and the lower values were observed for fruit weight, fruit firmness, and fruit scar (12%, 12% and 10%, respectively).

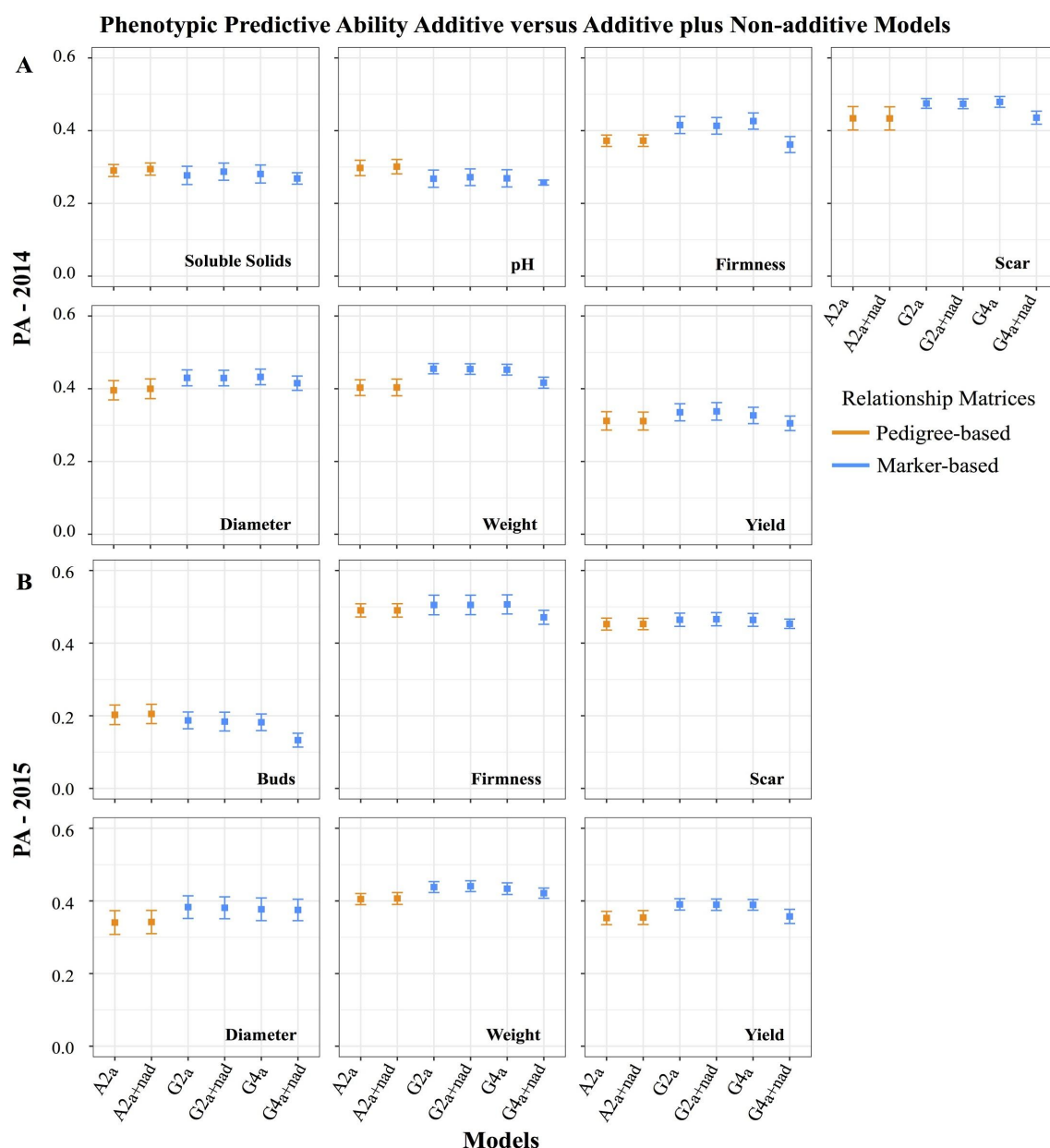


Figure 3. Predictive abilities for the single trait Bayesian analyses considering the addition of ploidy information, non-additive effects, and continuous genotypes in the prediction of breeding values for 1,847 blueberry individuals

5.4.6. Model's Goodness-of-fit and Stability

The goodness-of-fit for the different models was evaluated with measures of the posterior mean of the log likelihood obtained from the full data set. The model with the lowest value for this parameter presented the best fit for the data. The estimated values obtained for the genetic parameters and the goodness-of-fit statistics for each model are summarized in Table 4. The full tetraploid model ($G4_{a+mad}$) presented significantly higher values (poorest fit) for all traits analyses. The relationship matrix obtained from the

continuous genotype data presented significantly the best goodness-of-fit for all traits (Table 3), followed by the tetraploid standardization using marker-based data. The addition of non-additive effects in the marker-based models, while increasing the complexity, either improved or did not affected the model fitting, except for the full tetraploid model (G_{4a+nad}). To compare the predictive stability for each model associates mean square error (MSE) were obtained for each trait and year. Non-significant differences between the models' MSE were observe (Table 3).

5.4.7. Expected Genetic Gain

In this study we compared the expected genetic gain considering different selection methods: i) phenotypic BLUP, ii) pedigree-based BLUP, and iii) marker-based BLUP as well as the addition of ploidy information and non-additive effects. The results obtained for the expected genetic gain are summarized in Table 3.

Considering that GS offers the possibility to accelerate genetic improvement by decreasing the breeding cycle and selecting superior individuals in earlier phases of the breeding program (Munoz et al., 2014b; Resende et al., 2012), we propose that routine genomic selection cycle could be implemented in the second year of the UF Blueberry breeding program (Stage II; Figure 4). This would reduce the breeding cycle by three years, going from 12 (Cellon et al., 2018) to nine years in total.

The reduction in the cycle time associated with the higher PA for the marker-based methods generated a higher expected genetic gain (EGG) for all traits (Table 3). For example, fruit firmness, which presented higher EGG, had a mean value obtained for the marker-based methods (disregarding the method G_{4a+nad}) 2.15X and 1.46X superior to the use of the Identity matrix (I), and the pedigree-based matrices (A), respectively (Table 3). Even for brix, that presented one of the lowest EGG, the ratio between the mean value obtained for the marker-based methods (disregarding the method G_{4a+nad}) was 1.45X, and 1.27X superior to the use of the Identity matrix (I), and the pedigree-based matrices (A), respectively. The use of matrices that account for ploidy information generated significantly better results for the expected genetic gains, for almost all traits (Table 3). The use of continuous data to build a relationship matrix generate significantly higher EGG or values not significantly different of the best models for all traits, which also happened for the tetraploid standardization (Table 3).

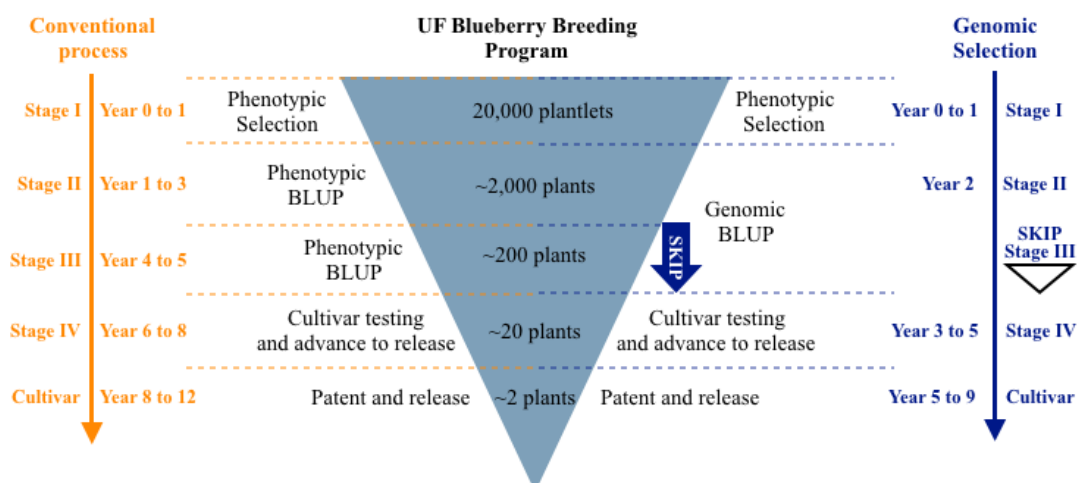


Figure 4. Routine of testing and selection in the UF blueberry breeding program stages and times (conventional process) and comparison with the proposed scheme of genomic selection implementation.

5.5 DISCUSSION

5.5.1. Relationship matrices

In this study nine linear mixed models were applied to predict breeding values in a real blueberry breeding population based on phenotypic, pedigree, and high-density marker data for 1,847 individuals. We compared the expected genetic gain, the stability, and the ability of models considering different relationship matrices: (BLUP, P-BLUP, G-BLUP). Going beyond previous studies using polyploid species we also explored models accounting for ploidy information and the addition of non-additive effects. We proposed the use of continuous data to perform GS, instead of assuming diploid or tetraploid standardization.

The relationship matrix built with pedigree information was able to increase the predictive ability when compared with the phenotypic-BLUP analysis (based on the identity matrix; Table 3). In other words, the insertion of information based on the expected genetic merit of the individuals (e.g. pedigree-based method; Henderson, 1976), yielded better PA results. Pedigree-based models also generated higher estimates for the genetic variance components when compared with all models, as well as high heritability values (Table 3). However, pedigree-based relationship matrices present low efficiency to capture and estimate genetic relationship (Muñoz et al., 2014), which can bias the estimations of genetic variance components. Marker-based methods have proven to

accurately estimate coancestry (Muñoz et al., 2014) between individuals, which is associated with accurate estimation of variance components. In our research the incorporation of molecular data reduced the overestimation of variance components, it also allowed a better separation and estimation of the additive and non-additive effects in accordance with the literature (Muñoz et al., 2014; Resende et al., 2017).

The PA results from pedigree-based matrices were not always significantly different of the marker-based methods. However, the use of pedigree-based relationship matrix is also associated with the overestimation of the reliability of selection (squared correlation between the predicted BV and the phenotypic breeding values: r^2), which is direct associated with the overestimation of predictive abilities, since $PA = \sqrt{r^2}$. This bias is associated with the use of the base additive variance without accounting for the loss of genetic variability caused by inbreeding in the selection process, which is natural in breeding programs (Bulmer, 1971). Marker-based methods also overestimate the reliability of selection, but in a much lower degree than pedigree-based methods (Gorjanc et al., 2015). As we evaluate the PA for a blueberry population submitted to selection since 1908, some overestimation should have occurred in the PA estimation, with bigger effect for pedigree-based methods.

In our study the results indicate that marker-based matrices could lead to better predictions than pedigree for most of the traits. The predictive abilities obtained with marker-based matrices were either superior or not significantly different of the ones obtained with pedigree information. The non-significant difference between methodologies can be an effect of the bias in pedigree-based estimations. Nevertheless, the addition of genome-wide markers information better estimate genetic merit. The use of molecular markers captures Mendelian segregation and also the genetic similarity among common ancestors that are unknown in the pedigree. For example, our population was composed of 117 full-sibs families. In this context, pedigree-based methods have no power to distinguish variance within families, which does not happen in the marker-based methods.

The PA values obtained with GS are encouraging for blueberry, even for traits with low heritability, and it is also an evidence to the potential of GS for autotetraploids breeding. We found accuracies values from 0.13 (flower bud, model: $G4_{a+nad}$) to 0.51 (fruit firmness, models: Gr and G4). As expected lower PA values were associated with lower heritabilities (Supplementary Table 1). The use of GS becomes even more promising when the expected genetic gains are compared. The whole selection process in the University of

Florida blueberry breeding program, takes from 12-15 years for a new cultivar be released (Lyrene et al., 2008; Cellon et al., 2018). Shortening the time required to complete each cycle of selection will maximize the rate of progress and can bring economic benefits to producers (Figure 4). With the application of high-density marker methods we expect a decrease of three year in the release of cultivars, which significantly increased the expected genetic gain (Table 3), supporting the prospects of GS application as a tool for blueberry breeding. The results revealed that GS presented an increase of 1.94X in the average genetic gain over conventional breeding method (Identity matrix), and 1.4X over pedigree-based methods (Table 3).

5.5.2. Allele dosage

Regarding the allele dosage effect, our results indicate that the importance of dosage in GS will vary depending on the trait being analyzed, as reported for banana in Nyine et al. (2018). The results obtained for both models that assume the existence of more than 3 genotypic classes (G4a and Gr) reiterates the importance of dosage in the breeding values prediction (Table 3, Supplementary Table 1). Slightly improvement was verified in the PA for fruit firmness, fruit scar, and fruit diameter when this factor was considered in the models. In addition, the model fitting was significantly better for methods that could add the models dosage information. Showing that allele dosage could have importance in the prediction of breeding values (Table 3).

Even though autopolyploids present relatively more uniform genomes than allopolyploids, dosage-related differences are observed (Guo et al., 1996; Yu et al., 2010). There are two types of dosage regulation, first when a dosage effect is verified (i.e., a increase or a decrease is significantly associated with the gene copy number). And second, when dosage compensation occurs and no differential expression is observed, regardless the ploidy. This is also characterized as a buffering effect, where the existence of multiple genomes masks a mutation effect (Soltis et al., 2010). The small effect in the PAs verified with the addition of ploidy information could be associated with the buffering effect, were a potential mutation is masked. It also could be associated with an inefficiency of the additive models to capture the difference between genotypic classes. If the difference in dosage expression were totally explained by an additive model, polyploidy would increase the potential of variation associated with the number of genotypic classes.

For example, if we observe a 4.0g range in the difference of fruit weight, assuming a full additive ploidy-dependent model, when a pseudo-diploid model is assumed, three genotypic classes are observed, (i.e., AA, AB, and BB) in which AA = 0g and BB = 4g, in this context the addition of each “B” allele would represent an increase of 2g in the phenotype. However, if we assume a full tetraploid model, we would have five genotypic classes instead of three, and the range of a 4g would now be divide among this five classes, now the addition of one “B” allele would represent an increase of 1g in the fruit weigh. In this scenario polyploidy would not be associated with an expansion of the phenotype range, but with the presence of a higher number of intermediate phenotypic classes (Osborn et al., 2003). Depending of the overlay among continuous distribution of the phenotypes, no significantly association can be detected among the genotype classes. Further studies comparing the performance of diploid individuals with the application of different models as well as different estimators could help verify the possible effect of dosage in the autopolyploid blueberry.

5.5.3. Non-additive effects

The clonal propagation widely used in blueberry production makes not only the additive effect important, but also the non-additive effects, because they are fixed in the selected clonal individual. The additive formulation of the relationship matrix G can be extended to include a matrix accounting for non-additive effects, since SNPs are able to capture the Mendelian sampling term within families (Resende et al., 2017). The predictive abilities in our study did not increase by the addition of non-additive effects in the models (Table 3). Besides the possible problem associated with the ploidy for the NAD estimation, which could lead to the non significant effect observed in the PA, the opposition in the direction between additive and non-additive effects can also lead to a lack of improvement in the PA when NAD are included in the models (Li et al., 2016).

Although PA differences were not observed with the addition of non-additive effects in the models, the addition of non-additive effects in the marker-based models, while making the models more complex, either improved or did not affected the model fitting. As a consequence, the addition of NAD in the models allowed better estimation of the variance components. Our results have showed that the non-additive effects represented a significant proportion of the total genetic variance, representing more than 20% for most of the traits (Table 3). This is in agreement with studies in rice (Morais Junior et al., 2017),

forest trees (Muñoz et al., 2014; Resende et al., 2017), and livestock (Li et al., 2017). Knowledge about the proportion of the variance explained for each factor (i.e. additive and non-additive variance components) can help to understand the trait architecture and expression and also be used guide crosses in order to maximize the additive and non-additive effect in the offspring (Slater et al., 2016). A desirable property for modeling is the separation between the variance components between additive and non-additive effects. The full tetraploid model proposed by Slater et al., (2016) does not allow the separation between the additive and non-additive variance components, it also presents the poorest goodness-of-fit and generate the lowest predictive ability values.

In less complex models (i.e., without addition of NAD), the estimates of additive variance were inflated by non-additive effects (Table 3). Confirmation of this can be seen in the decrease in narrow-sense heritability (h^2) caused by the addition of NAD effects. Which was expected in theory, since additive and non-additive effects could be confounded in breeding populations (Muñoz et al., 2014) leading to an overestimation of the additive component, heritability and expected gains. The existence of selection or inbreeding leads to the addition of covariance between alleles. In such a scenario, no guarantee of orthogonality between these effects can be assumed (Vieira et al., 2017). Because of the complex genetic structure in polyploids specific estimators could lead to, with a specific scheme would improve the estimation of variance component and if this could have a favorable effect in the predictive ability of GS models (e.g., Endelman et al., 2018).

5.5.4. Continuous data

In our study, the use of VanRaden's estimator in the construction of the relationship matrix led to better results. This methodology accounts for the identity-by-state of the alleles between individuals. In this case, no genetic assumption is made. The standardization assuming a given ploidy (i.e. diploid or tetraploid as we tested here) is used in the models in order to reduce noise. However, in polyploids the assignment of genotypes into classes is a challenging task with high risk of misclassification (Grandke et al., 2016). Studies of the evaluation of genotyping calling for polyploids show that no method works properly (Voorrips et al., 2011; Grandke et al., 2016). The problem for polyploids is that the expected distribution of signal intensity classes increases as the ploidy increases. For a given level of ploidy, n , the expected number of genotypic classes is $2n+1$. The number of

genotypic classes should correspond with the signal intensity clusters, but what happens is that with a higher number of classes the differences between signals of clusters is not enough to differentiate between them. The distribution of genotypes values approximates a continuous distribution, where no clear separation is observed, mainly for heterozygotes classes (Grandke et al., 2016). Under this scenario a ploidy assumption-free method could have advantages because it does not try to fit the dosage into classes with a high risk of missclassification.

Besides the problems associated with complex genome arrangement in polyploids, additional problems in the genotype identification can be associated with the quality of the genome and also with sequence depth. These problems can generate miscalling for the SNPs. A high coverage is required to correctly differentiate the five possible classes in the autotetraploids. Uitdewilligen et al. (2017) in their study with autotetraploid potato estimated that 60x to 80x of coverage is estimated to achieve 98.4% accuracy in the genotyping calling. The accurate identification of alleles is needed to identify events such double reduction, which is common in autotetraploids. The miscalling can interfere in the prediction accuracy and the effect of this issue in GS needs to be evaluated (Slater et al., 2016).

In this paper we demonstrate that we are able to use continuous genotypes to obtain moderate to high predictive abilities for eight fruit-quality and yield-related traits. We compared the estimates obtained for each model as well as models stability and ability of prediction. In all comparisons the use of continuous data performed as well as the best models and even generated slightly better predictive abilities for some of the traits (i.e., fruit firmness, fruit scar, and fruit diameter; Table 3). The lack of a dosage parameterization did not affect the models negatively, but significantly improved the model fitting for all traits (Table 3). The use of continuous data has been proved as efficient in association studies for polyploids. Grandke et al. (2016) obtained higher number of associations when continuous genotypes were used. They also used simulation to confirm their results. In their study the estimation of dosage resulted in missing association due to errors in genotype calling. Our study showed empirical evidences that the use of continuous data can also be applied in genomic-selection models for polyploid species. The use of continuous genotypes simplifies the analysis by reducing the computational time and also avoids problems associated with misclassification of genotypic classes.

5.6 CONCLUSIONS

Most autopolyploid breeding programs use recurrent phenotypic selection; this scheme involves screening of a large number of individuals, and the selection of the best individuals to be used as parents for the next generation. Simultaneously, the best individuals could be clonally propagated to become cultivars. This method involves not only large amount of resources but also a long time. Additionally, it is well known that it does not work well for low heritability traits. Here we presented results that consider different strategies to model phenotypic and genotypic data that could lead to higher selection accuracy and genetic gains. Specifically, GS can also accelerate the process of developing a new cultivar.

For example, In the case of the blueberry program, if GS is applied in the Stage II population, as we are proposing in this study (Figure 4), not just the breeding cycle time would be improved but there would also be savings to the program with the possibility of smaller field trials, as well as reduction in phenotyping evaluations, that are time and space consuming. The reduction in costs in these processes could allow to breakeven the genotype costs, which would lead to a cost-effective application of GS (Slater et al., 2016). However, phenotyping all selection traits will be necessary, which could be challenging.

The use of genomic data associated with application of quantitative genetics lead to a transformation in breeding (Resende et al., 2017). Our research indicates the potential of genomic selection for autotetraploid species, i.e., the UF blueberry breeding program. The adoption of GS as a tool for the breeding program would lead to an expected genetic gain with average of 1.94X, comparing with the actual methodology used, considering the analyzed traits.

Our results also shed light on the importance of allele dosage and non-additive effects. Despite the fact that these factors have not shown a significant effect in the predictive abilities, there was strong evidence that the importance of dosage in GS will vary depending on the trait being analyzed and the addition of this effect significantly generated the best model fitting. This study also indicated that non-additive effects represent a significant proportion of the total genetic variance, which could have importance to understand the trait architecture and expression, and could maximize the exploration of additive and non-additive effect in the offspring (Slater et al., 2016).

Furthermore, our study demonstrated for the first time that continuous data could be used in GS for prediction of breeding values. The use of continuous genotypes (ploidy assumption-free) improved the model fit for all traits and lead to slightly better predictive ability, with significant effects in the expected genetic gain for almost all traits. Besides it, the use of continuous genotypes simplifies the analysis reducing the computational time and it also avoids problems associated with misclassification of genotypic classes.

This novel result improves our current understanding about the genetic control of quantitative traits in autotetraploid species. Further studies may also generate gains not only to improve selection, but also generating information about the genome of the crop, bringing more information about the influence of different factors in the phenotypic expression of economically important traits.

5.7 FUNDING

USDA - Agriculture and Food Research Initiative Grant no. 2014-67013-22418 to Patricio R. Munoz, James W. Olmstead and Jeffrey B. Endelman from the USDA National Institute of Food and Agriculture. IBO was funded by the CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) [PSDE scholarship: 88881.131685/2016-01].

5.8 REFERENCES

- ADAMS, K. L.; CRONN, R.; PERCIFIELD, R.; WENDEL, J. F.. Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proceedings of the National Academy of sciences*, v. 100, n. 8, p. 4649-4654, 2003.
- AMADEU, R. R. CELLON, C.; OLMSTEAD, J. W.; GARCIA, A. A.; RESENDE, M. F.; MUÑOZ, P. R.. AGHmatrix: R Package to Construct Relationship Matrices for Autotetraploid and Diploid Species: A Blueberry Example. *The plant genome*, v. 9, n. 3, 2016.

ANNICCHIARICO, P.; NAZZICARI, N.; LI, X.; WEI, Y.; PECETTI, L.; BRUMMER, E. C. Accuracy of genomic selection for alfalfa biomass yield in different reference populations. *BMC genomics*, v. 16, n. 1, p. 1020, 2015.

ASHRAF, B. H.; EDRISS, V.; AKDEMIR, D.; AUTRIQUE, E.; BONNETT, D.; CROSSA, J.; JANSSE, L.; SINGH, R.; JANNINK, J. L. Genomic prediction using phenotypes from pedigreed lines with no marker data. *Crop Science*, v. 56, n. 3, p. 957-964, 2016.

ASHRAF, B. H.; BYRNE, S.; FÉ, D.; CZABAN, A.; ASP, T.; PEDERSEN, M. G.; LENK, I.; ROULUND, N.; DIDION, T.; JENSEN, C. S.; JENSEN, J.; JANSSE, L. L. Estimating genomic heritabilities at the level of family-pool samples of perennial ryegrass using genotyping-by-sequencing. *Theoretical and Applied Genetics*, v. 129, n. 1, p. 45-52, 2016.

BIAN, Y.; BALLINGTON, J.; RAJA, A.; BROUWER, C.; REID, R.; BURKE, M.; WANG, X.; ROWLAND, N. B.; BROWN, A. Patterns of simple sequence repeats in cultivated blueberries (*Vaccinium* section *Cyanococcus* spp.) and their use in revealing genetic diversity and population structure. *Mol. Breed.* 34, 675–689, 2014.

BIRCHLER, J. A.; BHADRA, U.; BHADRA, M. P.; AUGER, D. L. Dosage-dependent gene regulation in multicellular eukaryotes: implications for dosage compensation, aneuploid syndromes, and quantitative traits. *Developmental biology*, v. 234, n. 2, p. 275-288, 2001.

BROOKS, R.M., AND H.P. OLMO. Register of new fruit and nut varieties, 1920–1950. Univ. of California Press, Berkeley and Los Angeles, CA, 1952.

BROOKS, R.M., AND H.P. OLMO. Brooks and Olmo register of fruit & nut varieties. 3rd ed. ASHS Press, Alexandria, VA, 1997.

BULMER, MG. The effect of selection on genetic variability. *The American Naturalist*, v. 105, n. 943, p. 201-211, 1971.

CALUS, M. P. L. MEUWISSEN T. H. E.; DE ROOS, A. P. W.; VEERKAMP, R. F. Accuracy of genomic selection using different methods to define haplotypes. *Genetics*, v. 178, n. 1, p. 553-561, 2008.

CELLON, C.; AMADEU, R. R.; OLMSTEAD, J. W.; MATTIA, M. R.; FERRÃO, L. F. V.; MUÑOZ, P. R. Estimation of genetic parameters and prediction of breeding values in

an autotetraploid blueberry breeding population with extensive pedigree data. *Euphytica*, v. 214, p. 1-13, 2018.

COCKERHAM, C. C. An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. *Genetics*, v. 39, n. 6, p. 859-882, 1954.

COMAI, LUCA. The advantages and disadvantages of being polyploid. *Nature reviews genetics*, v. 6, n. 11, p. 836, 2005.

COSTICH, D. E.; ORTIZ, R.; MEAGHER, T. R.; BRUEDERLE, L. P.; VORSA, N.. Determination of ploidy level and nuclear DNA content in blueberry by flow cytometry. *Theoretical and Applied Genetics*, v. 86, n. 8, p. 1001-1006, 1993.

DE ALMEIDA FILHO, J. E.; GUIMARÃES, J. F. R.; E SILVA, F. F.; DE RESENDE, M. D. V.; MUÑOZ, P.; KIRST, M.; RESENDE JR, M. F. R.. The contribution of dominance to phenotype prediction in a pine breeding and simulated population. *Heredity*, v. 117, n. 1, p. 33, 2016.

DE LOS CAMPOS, G.; PEREZ RODRIGUEZ, P. BGLR: Bayesian Generalized Linear Regression. R package version 1.0.5. <https://CRAN.R-project.org/package=BGLR>, 2016.

DE LOS CAMPOS, G.; HICKEY, J. M.; PONG-WONG, R.; DAETWYLER, H. D.; CALUS, M. P. Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics*, v. 193, n. 2, p. 327-345, 2013.

DESTA, ZERATSION ABERA; ORTIZ, RODOMIRO. Genomic selection: genome-wide prediction in plant improvement. *Trends in plant science*, v. 19, n. 9, p. 592-601, 2014.

DUFRESNE, FRANCE; STIFT, MARC; ERGILINO, ROLAND V.; MABLE, BARBARA K. Recent progress and challenges in population genetics of polyploid organisms: an overview of current state-of-the-art molecular and statistical tools. *Molecular ecology*, v. 23, n. 1, p. 40-69, 2014.

ENDELMAN, Jeffrey B.; CARLEY, C. A. S.; BETHKE, P. C.; COOMBS, J. J.; CLOUGH, M. E.; DA SILVA, W. L.; DE JONG, W. S.; DOUCHES, D. S.; FREDERICK, C. M.; HAYNES, K. G.; HOLM, D. G.; MILLER, J. C.; MUÑOZ, P. R.; NAVARRO, F. M.; NOVY, R. G.; PALTA, J. P.; POLTER, G. A.; RAK, K. T.; SATHUVALLI, V. R.; THOMPSON, A. L.; CRAIG YENCHO, G. Genetic variance partitioning and genome-wide prediction with allele dosage information in autotetraploid potato. **Genetics**, v. 209,

n. 1, p. 77-87, 2018.

ENDELMAN, JEFFREY B. Ridge regression and other kernels for genomic selection with R package rrBLUP. *The Plant Genome*, v. 4, n. 3, p. 250-255, 2011.

GARCIA, A. A.; MOLLINARI, M.; MARCONI, T. G.; SERANG, O. R.; SILVA, R. R.; VIEIRA, M. L. C.; VICENTINI, R.; COSTA, E. A.; MANCINI, M. C.; GARCIA, M. O. S.; PASTINA, M. M.; GAZAFFI, R.; MARTINS, E. R. F.; DAHMER, N.; SFORÇA, D. A.; SILVA, C. B. C.; BUNDOCK, P.; HENRY, R. J.; SOUZA, G. M.; SLUYS, M.-A.; LANDELL, M. G. A.; CARNEIRO, M. S.; VICENTZ, M. A. G.; PINTO, L. R.; VENCOSKY, R.; SOUZA, A. P. SNP genotyping allows an in-depth characterisation of the genome of sugarcane and other complex autoployploids. *Scientific reports*, v. 3, n. 3399, Dez. 2013.

GARRISON, E.; MARTH, G. Haplotype-based variant detection from short-read sequencing. *arXiv Prepr. arXiv1207.3907*, 2012.

GORJANC, GREGOR; BIJMA, PITER; HICKEY, JOHN M. Reliability of pedigree-based and genomic evaluations in selected populations. *Genetics Selection Evolution*, v. 47, n. 1, p. 65, 2015.

GOUY, M.; RUSSELL, Y.; BASTIANELLI, D.; LECOMTE, P.; BONNAL, L.; ROQUES, D.; EFILE, J.-C.; ROCHER, S.; DAUGROIS, J.; TOUBI, L.; NABENEZA, S.; HERVOUET, C.; TELISMART, H.; DENIS, M.; THONG-CHANE, A.; GLASZMANN, J. C.; HOARAU, J.-Y.; NIBOUCHE, S.; COSTET, L. Experimental assessment of the accuracy of genomic selection in sugarcane. *Theoretical and Applied Genetics*, Nova York: NY, v. 126, n. 10, p. 2575-2586, 2013.

GRANDKE, F.; SINGH, P.; HEUVEN, H. C.; DE HAAN, J. R.; METZLER, D. Advantages of continuous genotype values over genotype classes for GWAS in higher polyploids: a comparative study in hexaploid chrysanthemum. *BMC genomics*, v. 17, n. 1, p. 672, 2016.

GUO, MEI; DAVIS, DOUG; BIRCHLER, JAMES A. Dosage effects on gene expression in a maize ploidy series. *Genetics*, v. 142, n. 4, p. 1349-1355, 1996.

GUPTA, V.; ESTRADA, A. D.; BLAKLEY, I.; REID, R.; PATEL, K.; MEYER, M. D.; ANDERSEN, S. U.; BROWN, A. F.; LILA, M. A.; LORAIN, A. E. RNA-Seq analysis and annotation of a draft blueberry genome assembly identifies candidate genes involved

in fruit ripening, biosynthesis of bioactive compounds, and stage-specific alternative splicing. *Gigascience* v. 4, n. 1, p. 5, 2015.

HABIER, D.; TETENS, J.; SEEFRIED, F. R.; LICHTNER, P.; THALLER, G. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. *Genetics Selection Evolution*, v. 42, n. 1, p. 5, 2010.

HENDERSON, CHARLES R. A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics*, p. 69-83, 1976.

HIETER, PHILIP; GRIFFITHS, TONY. Polyploidy--more is more or less. *Science*, v. 285, n. 5425, p. 210-211, 1999.

HILL, WILLIAM G. Understanding and using quantitative genetic variation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, v. 365, n. 1537, p. 73-85, 2010.

JOMBART, T.; AHMED, I. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. v. 27, n. 21, p. 3070-3071, 2011.

KEMPTHORNE, OSCAR. The correlation between relatives in a random mating population. *Proc. R. Soc. Lond. B. The Royal Society*. p. 103-113, 1954.

KERR, RICHARD J.; LI, L.; TIER, B.; DUTKOWSKI, G. W.; MCRAE, T. A.. Use of the numerator relationship matrix in genetic analysis of autopolyploid species. *Theoretical and applied genetics*, v. 124, n. 7, p. 1271-1282, 2012.

LI, Y.; HAWKEN, R.; SAPP, R.; GEORGE, A.; LEHNERT, S. A.; HENSHALL, J. M.; REVERTER, A. Evaluation of non-additive genetic variation in feed-related traits of broiler chickens. *Poultry science*, v. 96, n. 3, p. 754-763, 2017.

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

LYRENE, P.M. 1981. Juvenility and production of fast-rooting cuttings from blueberry shoot cultures. **Journal American Society for Horticultural Science**, n. 106, p. 396-398, 1998.

LYRENE, PAUL. Breeding southern highbush blueberries. *Plant Breeding Reviews*, v. 30, p. 353-414, 2008.

LYRENE, PAUL. Development of highbush blueberry cultivars adapted to Florida. *Journal of the American Pomological Society*, v. 56, n. 2, p. 79, 2002.

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

MOLLINARI, M.; SERANG, O. Quantitative SNP genotyping of polyploids with MassARRAY and other platforms. In: *Plant Genotyping*. Humana Press, New York, NY, p. 215-241, 2015.

MORAIS JÚNIOR, O. P.; DUARTE, J. B.; BRESEGHELLO, F.; COELHO, A. S. G.; BORBA, T. C. O.; AGUIAR, J. T.; NEVES, P. C. F.; MORAIS, O. P. Relevance of additive and non-additive genetic relatedness for genomic prediction in rice population under recurrent selection breeding. *Genetics and Molecular Research*, v. 16, n. 4, gmr16039849, 2017.

MUNOZ, P.R.; RESENDE, M. F.; HUBER, D. A.; QUESADA, T.; RESENDE, M. D.; NEALE, D. B.; WEGRZYN, J. L.; KIRST, MATIAS; PETER, G. F.. Genomic relationship matrix for correcting pedigree errors in breeding populations: impact on genetic parameters and genomic selection accuracy. *Crop Science*, v. 54, n. 3, p. 1115-1123, 2014b.

MUÑOZ, P. R.; RESENDE, M. F.; GEZAN, S. A.; RESENDE, M. D. V.; DE LOS CAMPOS, G.; KIRST, M.; HUBER, D., PETER, G. F.. Unraveling additive from nonadditive effects using genomic relationship matrices. *Genetics*, v. 198, n. 4, p. 1759-1768, 2014.

National Center for Genome Resources—Corvallis. 2012. NCGR-Corvallis *Vaccinium* catalog: Blueberry cultivars and selections. USDA–NRCS. <http://www.ars-grin.gov/cor/catalogs/vactc.html> (accessed day Mo. year).

NYINE, MOSES; UWIMANA, BRIGITTE; BLAVET, NICOLAS; HIBOVÁ, EVA; VANRESPAILLE, HELENA; BATTE, MICHAEL; AKECH, VIOLET; BROWN, ALLAN; LORENZEN, JIM; SWENNEN; DALEZEL, JAROSLAV. Genomic Prediction in a Multiploid Crop: Genotype by Environment Interaction and Allele Dosage Effects on Predictive Ability in Banana. *The Plant Genome*, 2018.

OSBORN, T. C.; PIRES, J. C.; BIRCHLER, J. A.; AUGER, D. L.; CHEN, Z. J.; LEE, H.-S.; COMAI, L.; MADLUNG, A.; DOERGE, R. W.; COLOT, V.; MARTIENSSEN, R. A.

Understanding mechanisms of novel gene expression in polyploids. *Trends in genetics*, Cambridge, MA, v. 19, n. 3, p. 141-147, 2003.

PÉREZ, P.; DE LOS CAMPOS, G. Genome-wide regression & prediction with the BGLR statistical package. *Genetics*, p. genetics. 114.164442, 2014.

R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, 2018. Available from: <https://www.R-project.org/>.

RAMSEY, JUSTIN; SCHEMSKE, DOUGLAS W. Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annual Review of Ecology and Systematics*, v. 29, n. 1, p. 467-501, 1998.

RESENDE, M. F. R.; MUÑOZ, P. R.; ACOSTA, J. J.; PETER, G. F., DAVIS, J. M.; GRATTAPAGLIA, D.; KIRST, M. Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments. *New Phytologist*, v. 193, n. 3, p. 617-624, 2012.

RESENDE, M. F. R., MUÑOZ, P. R.; RESENDE, M. D.; GARRICK, D. J.; FERNANDO, R. L.; DAVIS, J. M.; JOKELA, E. J.; MARTIN, T. A.; PETER, G. F.; KIRST, M. Accuracy of genomic selection methods in a standard data set of loblolly pine (*Pinus taeda* L.). *Genetics*, v. 190, n. 4, p. 1503-1510, 2012.

RESENDE, R. T.; RESENDE, M. D. V.; SILVA, F. F.; AZEVEDO, C. F.; TAKAHASHI, E. K.; SILVA-JUNIOR, O. B.; GRATTAPAGLIA, D. Assessing the expected response to genomic selection of individuals and families in *Eucalyptus* breeding with an additive-dominant model. *Heredity*, v. 119, n. 4, p. 245, 2017.

SLATER, A. T., G. M. WILSON, N. O. COGAN, J. W. FORSTER, AND B. J. HAYES, 2013 Improving the analysis of low heritability complex traits for enhanced genetic gain in potato. *Theoretical and applied genetics* 127: 809–820.

SLATER, ANTHONY T.; COGAN, N. O.; FORSTER, J. W.; HAYES, B. J.; DAETWYLER, H. D. Improving genetic gain with genomic selection in autotetraploid potato. *The plant genome*, v. 9, n. 3, 2016.

SOLTIS, DOUGLAS E.; VISGER, CLAYTON J.; SOLTIS, PAMELA S. The polyploidy revolution then... and now: Stebbins revisited. *American Journal of Botany*, v. 101, n. 7, p. 1057-1078, 2014.

UITDEWILLIGEN, J. G.; WOLTERS, A. M. A.; BJORN, B.; BORM, T. J.; VISSER, R. G.; VAN ECK, H. J. A next-generation sequencing method for genotyping-by-sequencing of highly heterozygous autotetraploid potato. *PloS one*, v. 8, n. 5, p. e62355, 2013.

VANRADEN, PAUL M. Efficient methods to compute genomic predictions. *Journal of dairy science*, v. 91, n. 11, p. 4414-4423, 2008.

VIEIRA, I. C.; SANTOS, J.P. R. DOS; PIRES, L.P.M.; LIMA, B.M.; GONÇALVES, F.M.A.; BALESTRE, M. Assessing non-additive effects in GBLUP model. *Genetics and molecular research: GMR*, v. 16, n. 2, 2017.

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

VOORRIPS RE, GORT G, VOSMAN B. Genotype calling in tetraploid species from bi-allelic marker data using mixture models. *BMC bioinformatics*, v. 12, n. 1, p. 172, 2011.

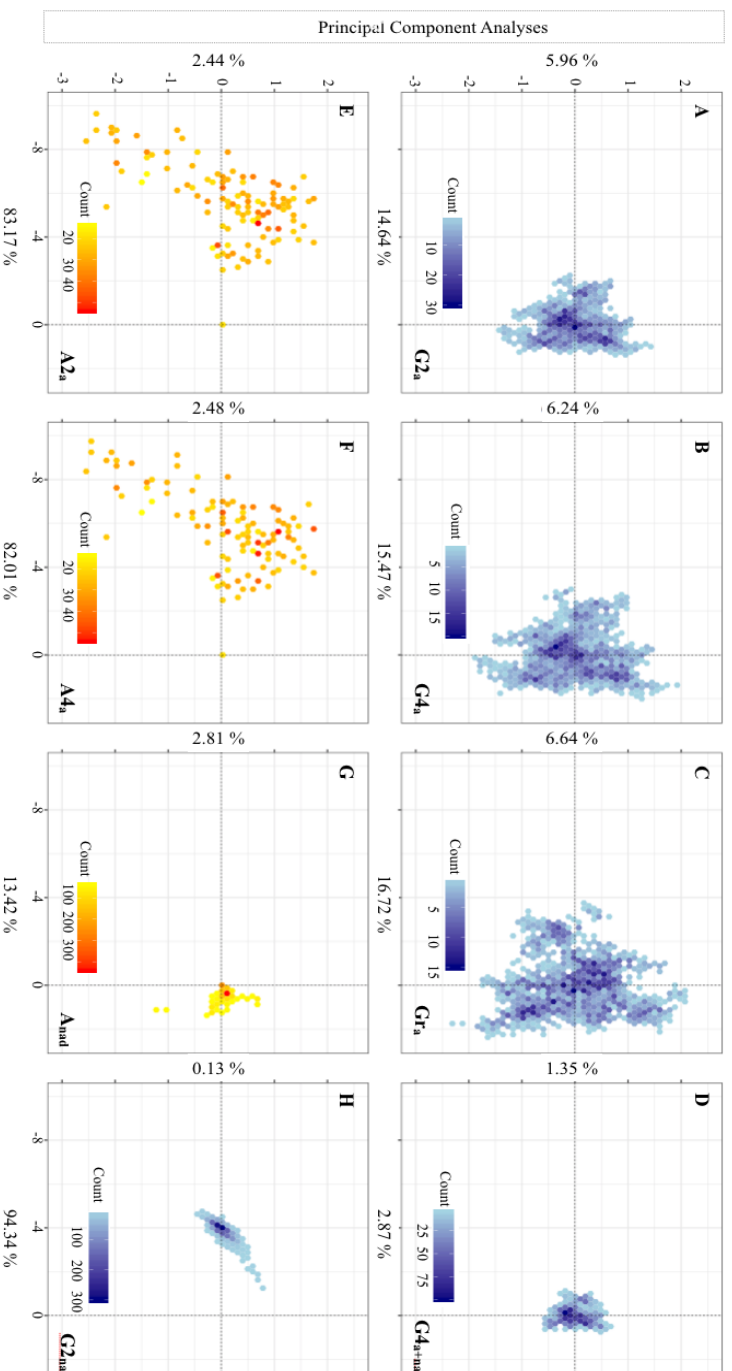
WIGGANS, G. R.; VANRADEN, P. M.; COOPER, T. A. The genomic evaluation system in the United States: Past, present, future. *Journal of dairy science*, v. 94, n. 6, p. 3202–3211, 2011.

YU, ZHENG; HABERER, GEORG; MATTHES, MICHAELA; RATTEI, THOMAS; MAYER, KLAUS F. X.; GIERL, ALFONS; TORRES-RUIZ, RAMON A.. Impact of natural genetic variation on the transcriptome of autotetraploid *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences*, v. 107, n. 41, p. 17809-17814, 2010.

ZHANG, Z.; DING, X.; LIU, J.; KONING, D-J.; ZHANG, Q. Genomic selection for QTL-MAS data using a trait-specific relationship matrix. In: *BMC proceedings*. BioMed Central Ltd, 2011. p. S15.

5.9

SUPPLEMENTAL INFORMATION



Supplementary Figure 1. Principal components heat plots for all the relationship matrices used in the study. In blue results obtained for the marker-based relationship matrices, in orange results obtained for the pedigree-based matrices. **A)** PCA for the diploid relationship matrix using the VanRaden (2008) methodology; **B)** PCA for the tetraploid relationship matrix using the VanRaden (2008) methodology; **C)** PCA for the continuous relationship matrix using the Slater et al. (2016) methodology; **D)** PCA for the additive plus non-additive effects relationship matrix using the Henderson et al. (1976) methodology; **E)** PCA for the additive effects relationship matrix using the Kerr et al. (2012) methodology; **F)** PCA for the non-additive effects relationship matrix using the Vitezica et al. (2013) methodology; **G)** PCA for the non-additive effects relationship matrix using the Vitezica et al. (2013) methodology; **H)** PCA for the non-additive effects relationship matrix using the Vitezica et al. (2013) methodology.

Table S1. Genetic parameters estimated for eight yield and fruit-related traits analyzed with nine linear mixed models, considering the use of ploidy information, non-additive effects and continuous genotypes

Trait	Relationship matrix	Residual Variance	MSE (2014) a	MSE (2015) a	Phenotypic predictive ability 2014	Phenotypic predictive ability 2015
Soluble Solid (°Brix)	I	1.794 (0.014) e	5.795 (2.113) a	-	0.235 (0.019) a	-
	A2	2.129 (0.002) b	4.926 (2.052) a	-	0.290 (0.016) a	-
	A2a+nad	1.897 (0.005) d	4.857 (2.011) a	-	0.294 (0.017) a	-
	A4	2.125 (0.002) b	4.531 (1.750) a	-	0.290 (0.020) a	-
	G2	2.026 (0.001) c	2.238 (1.108) a	-	0.277 (0.025) a	-
	G2a+nad	1.642 (0.003) f	2.143 (1.089) a	-	0.287 (0.024) a	-
	G4	2.109 (0.001) b	2.197 (1.111) a	-	0.281 (0.025) a	-
	G4a+nad	0.868 (0.006) g	3.263 (1.272) a	-	0.269 (0.016) a	-
pH	Gr	2.241 (0.001) a	2.216 (1.084) a	-	0.274 (0.024) a	-
	I	0.118 (0.001) e	0.244 (0.078) a	-	0.259 (0.024) a	-
	A2	0.140 (0) c	0.257 (0.074) a	-	0.297 (0.021) a	-
	A2a+nad	0.124 (0) d	0.264 (0.077) a	-	0.301 (0.020) a	-
	A4	0.140 (0) c	0.126 (0.042) a	-	0.285 (0.018) a	-
	G2	0.141 (0) c	0.253 (0.112) a	-	0.268 (0.024) a	-
	G2a+nad	0.124 (0) d	0.252 (0.111) a	-	0.272 (0.023) a	-
	G4	0.147 (0) b	0.259 (0.116) a	-	0.269 (0.024) a	-
Flower Bud	G4a+nad	0.060 (0) f	0.102 (0.040) a	-	0.257 (0.007) a	-
	Gr	0.153 (0) a	0.256 (0.116) a	-	0.271 (0.023) a	-
	I	4.752 (0.036) e	-	9.261 (4.198) a	-	0.152 (0.026) a
	A2	6.080 (0.003) a	-	8.280 (3.574) a	-	0.203 (0.027) a
	A2a+nad	5.580 (0.003) c	-	8.336 (3.572) a	-	0.205 (0.027) a
	A4	6.070 (0.004) a	-	8.039 (3.368) a	-	0.199 (0.015) a
	G2	5.562 (0.007) c	-	11.469 (7.792) a	-	0.187 (0.023) a
	G2a+nad	5.195 (0.006) d	-	11.286 (7.578) a	-	0.184 (0.026) a
Fruit Firmness	G4	5.881 (0.005) b	-	11.718 (8.030) a	-	0.182 (0.023) a
	G4a+nad	3.224 (0.029) f	-	9.558 (4.703) a	-	0.133 (0.019) a
	Gr	6.115 (0.003) a	-	11.86 (8.177) a	-	0.181 (0.023) a
	I	737.735 (4.313) c	2725.511 (1069.769) a	1109.026 (335.222) a	0.302 (0.018) b	0.425 (0.019) a
	A2	741.089 (6.679) bc	2146.811 (661.845) a	734.108 (336.184) a	0.372 (0.016) ab	0.490 (0.018) a
	A2a+nad	761.548 (6.872) b	2159.564 (672.504) a	758.789 (349.321) a	0.372 (0.016) ab	0.490 (0.018) a
	A4	742.547 (3.461) bc	778.212 (334.476) a	1490.294 (468.136) a	0.375 (0.019) ab	0.486 (0.018) a
	G2	734.332 (3.542) c	1240.152 (542.983) a	2083.417 (779.055) a	0.415 (0.023) a	0.505 (0.027) a
Fruit Scar	G2a+nad	735.150 (6.808) c	1217.937 (502.972) a	2017.248 (739.398) a	0.413 (0.023) a	0.505 (0.027) a
	G4	749.865 (3.374) bc	1245.192 (549.244) a	2085.359 (799.180) a	0.426 (0.022) a	0.507 (0.026) a
	G4a+nad	691.353 (2.961) d	1348.013 (500.904) a	1354.417 (638.530) a	0.362 (0.022) ab	0.471 (0.019) a
	Gr	783.729 (3.617) a	1250.673 (569.916) a	2124.308 (780.860) a	0.431 (0.021) a	0.511 (0.026) a
	I	0.073 (0.001) ef	0.236 (0.165) a	0.424 (0.198) a	0.341 (0.026) b	0.380 (0.032) a
	A2	0.075 (0.001) cde	0.113 (0.033) a	0.416 (0.123) a	0.434 (0.032) ab	0.452 (0.016) a
	A2a+nad	0.077 (0.001) bc	0.115 (0.034) a	0.407 (0.121) a	0.434 (0.032) ab	0.453 (0.016) a
	A4	0.075 (0) bcd	0.112 (0.041) a	0.181 (0.054) a	0.432 (0.016) ab	0.450 (0.020) a
Fruit Scar	G2	0.075 (0) de	0.160 (0.084) a	0.197 (0.123) a	0.475 (0.013) a	0.464 (0.018) a
	G2a+nad	0.077 (0.001) b	0.159 (0.085) a	0.191 (0.122) a	0.474 (0.013) a	0.466 (0.018) a
	G4	0.077 (0) b	0.160 (0.081) a	0.199 (0.130) a	0.479 (0.015) a	0.464 (0.018) a
	G4a+nad	0.072 (0) f	0.118 (0.028) a	0.073 (0.031) a	0.436 (0.018) ab	0.453 (0.013) a
	Gr	0.081 (0) a	0.171 (0.091) a	0.184 (0.122) a	0.488 (0.015) a	0.466 (0.018) a

Table S1. (cont.)

Trait	Relationship matrix	Residual Variance	MSE (2014) a	MSE (2015) a	Phenotypic predictive ability 2014	Phenotypic predictive ability 2015
Fruit Diameter	I	6.804 (0.044) bc	8.827 (3.751) a	8.918 (2.630) a	0.380 (0.028) a	0.333 (0.028) a
	A2	6.854 (0.094) b	8.072 (2.952) a	8.638 (2.614) a	0.396 (0.027) a	0.340 (0.033) a
	A2a+nad	6.865 (0.043) b	7.754 (2.877) a	8.490 (2.526) a	0.400 (0.027) a	0.342 (0.032) a
	A4	6.825 (0.085) b	9.134 (3.254) a	3.409 (1.177) a	0.386 (0.033) a	0.345 (0.024) a
	G2	6.799 (0.047) bc	19.796 (6.812) a	6.339 (2.941) a	0.430 (0.022) a	0.383 (0.031) a
	G2a+nad	6.825 (0.046) b	19.048 (6.303) a	6.173 (2.991) a	0.429 (0.021) a	0.381 (0.030) a
	G4	6.954 (0.050) ab	19.423 (6.577) a	6.484 (3.022) a	0.432 (0.021) a	0.377 (0.031) a
	G4a+nad	6.553 (0.048) c	18.100 (5.304) a	5.595 (2.751) a	0.415 (0.020) a	0.375 (0.030) a
	Gr	7.219 (0.052) a	19.718 (7.238) a	7.241 (3.467) a	0.432 (0.022) a	0.376 (0.031) a
Fruit Weight	I	0.214 (0.001) bc	0.502 (0.171) a	0.337 (0.134) a	0.345 (0.026) b	0.363 (0.014) a
	A2	0.207 (0.001) de	0.529 (0.188) a	0.391 (0.194) a	0.403 (0.022) ab	0.405 (0.015) a
	A2a+nad	0.210 (0.001) bcd	0.527 (0.190) a	0.386 (0.199) a	0.404 (0.023) ab	0.407 (0.016) a
	A4	0.205 (0.001) e	0.404 (0.226) a	1.100 (0.349) a	0.403 (0.026) ab	0.400 (0.030) a
	G2	0.206 (0.001) de	1.030 (0.290) a	0.615 (0.177) a	0.455 (0.014) a	0.438 (0.015) a
	G2a+nad	0.210 (0.001) cde	1.041 (0.291) a	0.613 (0.177) a	0.454 (0.015) a	0.441 (0.015) a
	G4	0.215 (0.001) b	1.059 (0.289) a	0.603 (0.174) a	0.453 (0.015) a	0.434 (0.016) a
	G4a+nad	0.194 (0.001) f	1.069 (0.278) a	0.573 (0.208) a	0.417 (0.015) ab	0.421 (0.014) a
	Gr	0.231 (0.001) a	1.027 (0.303) a	0.702 (0.190) a	0.453 (0.014) a	0.430 (0.015) a
Yield	I	0.444 (0.003) b	0.881 (0.297) a	1.221 (0.447) a	0.251 (0.022) a	0.312 (0.023) a
	A2	0.442 (0.004) b	0.661 (0.207) a	1.214 (0.424) a	0.312 (0.025) a	0.353 (0.018) a
	A2a+nad	0.447 (0.003) b	0.704 (0.227) a	1.184 (0.417) a	0.311 (0.025) a	0.354 (0.019) a
	A4	0.442 (0.004) b	0.638 (0.171) a	1.730 (0.563) a	0.325 (0.021) a	0.348 (0.017) a
	G2	0.441 (0.004) b	1.191 (0.427) a	2.343 (0.723) a	0.335 (0.024) a	0.390 (0.016) a
	G2a+nad	0.450 (0.006) b	1.156 (0.406) a	2.363 (0.722) a	0.338 (0.024) a	0.390 (0.016) a
	G4	0.458 (0.004) b	1.263 (0.449) a	2.308 (0.704) a	0.327 (0.022) a	0.389 (0.015) a
	G4a+nad	0.419 (0.004) c	1.169 (0.446) a	2.705 (0.841) a	0.305 (0.020) a	0.357 (0.019) a
	Gr	0.493 (0.005) a	1.297 (0.446) a	2.294 (0.683) a	0.324 (0.024) a	0.384 (0.014) a

Apesar das dificuldades associada à complexidade genética das espécies poliploides, este trabalho gerou evidências empíricas de que a seleção genômica pode aumentar o ganho genético para estas espécies. Além disso, foi demonstrado que o uso de modelos que melhor representam a variabilidade genética destas espécies (i.e., uso de informação de dosagem e inserção de efeitos não-aditivos) pode auxiliar na obtenção de resultados (i.e., acurácia) e informação (i.e., proporções explicadas pelos componentes genéticos). As informações aqui geradas podem auxiliar na compreensão da arquitetura e expressão dos caracteres quantitativos e podem ser utilizadas para maximizar a exploração de efeitos genéticos no processo de melhoramento. A análise de gerações subsequentes das populações pode ser utilizado para a verificação dos resultados aqui apontados, permitindo a validação dos modelos aqui obtidos.

Além disso, este trabalho demonstrou pela primeira vez o uso de genótipos contínuos em análises de seleção genômica ampla. O uso de genótipos contínuos melhorou o ajuste do modelo para todas os caracteres e levou a uma habilidade preditiva ligeiramente melhor, com efeitos significativos no ganho genético para a maioria dos caracteres, quando aplicado aos dados de blueberry. Além disso, o uso de genótipos contínuos simplifica a análise reduzindo o tempo computacional e também evita problemas associados à classificação errônea de classes genotípicas. O uso de dados quantitativos permite a transposição da barreira gerada para a identificação de genótipos, entretanto, maior desenvolvimento teórico é necessário, permitindo a melhor exploração dos dados genéticos para poliploides.

Estudos futuros considerando metodologias desenvolvidas especificamente desenvolvidas para poliploides, bem como a análise de vasta gama de variabilidade podem auxiliar no entendimento dos efeitos de dosagem e efeitos não-aditivos em futuras nas análise de GS para poliploides. Considerando a análise cienciométrica dos dados, vê-se que o desenvolvimento de metodologias especificamente voltadas para poliploides é uma das

principais barreiras para o uso de informação genética nestas culturas. Neste sentido, o investimento em recursos humanos e a intensificação de parcerias internacionais poderiam auxiliar o desenvolvimento de teorias para estas espécies, de maneira a fomentar o progresso na interpretação e uso dos dados genéticos poliploides.

7 REFERÊNCIAS BIBLIOGRÁFICAS

ADAMS, K. L.; CRONN, R.; PERCIFIELD, R.; WENDEL, J. F. Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. **Proceedings of the National Academy of sciences**, v. 100, n. 8, . 4649-4654, 2003.

AITKEN, K. S.; JACKSON, P. A.; MCINTYRE, C. L. A combination of AFLP and SSR markers provides extensive map coverage and identification of homo (eo) logous linkage groups in a sugarcane cultivar. **Theoretical and Applied Genetics**, v. 110, n. 5, p. 789-801, 2005.

ALBINIO, J. C.; CRESTE, S.; FIGUEIRA, A. Uso de marcadores moleculares no mapeamento genético visando o melhoramento da cana-de-açúcar. **Biotecnologia, Ciência e Desenvolvimento**, v. 36, p. 82-91, 2006.

ANNICCHIARICO, P.; NAZZICARI, N.; LI, X.; WEI, Y.; PECETTI, L.; BRUMMER, E. C. Accuracy of genomic selection for alfalfa biomass yield in different reference populations. **BMC genomics**, v. 16, n. 1, p. 1020, 2015.

ASHRAF, B. H.; EDRISS, V.; AKDEMIR, D.; AUTRIQUE, E.; BONNETT, D.; CROSSA, J.; JANSSE, L.; SINGH, R.; JANNINK, J. L. Genomic prediction using phenotypes from pedigreed lines with no marker data. **Crop Science**, v. 56, n. 3, p. 957-964, 2016.

BAKER, P.; JACKSON, P.; AITKEN, K. Bayesian estimation of marker dosage in sugarcane and other autopolyploids. **Theoretical and applied genetics**, Heidelberg, v. 120, n. 8, p. 1653-1672, 2010.

BALLINGTON, J. R. Germplasm resources available to meet future needs for blueberry cultivar improvement. **Fruit Varieties Journal**, v. 44, n. 2, p. 54-62, 1990.

BOCHES, Peter; BASSIL, Nahla V.; ROWLAND, Lisa. Genetic diversity in the highbush blueberry evaluated with microsatellite markers. **Journal of the American Society for Horticultural Science**, v. 131, n. 5, p. 674-686, 2006.

- BIRCHLER, J. A.; BHADRA, U.; BHADRA, M. P.; AUGER, D. L. Dosage-dependent gene regulation in multicellular eukaryotes: implications for dosage compensation, aneuploid syndromes, and quantitative traits. **Developmental biology**, v. 234, n. 2, p. 275-288, 2001.
- BREVIS, P. A.; BASSIL, N. V.; BALLINGTON, J. R.; HANCOCK, J. F. Impact of wide hybridization on highbush blueberry breeding. **Journal of the American Society for Horticultural Science**, v. 133, n. 3, p. 427-437, 2008.
- BRAZELTON, C.; YOUNG, K.; BAUER, N. Global blueberry statistics and intelligence report. **International Blueberry Organization**. v. 1 84p., 2017.
- CASU, R. E.; SELIVANOVA, A.; PERROUX, J. M. High-throughput assessment of transgene copy number in sugarcane using real-time quantitative PCR. **Plant cell reports**, Heidelberg, v. 31, n. 1, p. 167-177, 2012.
- CHEAVEGATTI-GIANOTTO, A. et al. Sugarcane (*Saccharum X officinarum*): a reference study for the regulation of genetically modified cultivars in Brazil. **Tropical Plant Biology**, Aiea, Hawaii, v. 4, n. 1, p. 62-89, 2011.
- COCKERHAM, C. C. An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. **Genetics**, v. 39, n. 6, p. 859-882, 1954.
- COMAI, LUCA. The advantages and disadvantages of being polyploid. **Nature reviews genetics**, v. 6, n. 11, p. 836, 2005.
- CORDEIRO, G. M.; ELLIOTT, F. G.; KENNEDY, B. G.; HENRY, R. J. SNP analysis tools for functional analysis of sugarcane genes. **In: Plant and Animal Genomes Conference XIV**, San Diego, California, USA, 14-18 January. 2006.
- CUADRADO, A.; ACEVEDO, R.; MORENO DIAZ DE LA ESPINA, S.; JOUVE, N.; DE LA TORRE, C. Genome remodelling in three modern *S. officinarum* × *S. spontaneum* sugarcane cultivars. **Journal of experimental botany**, v. 55, n. 398, p. 847-854, 2004.
- D'HONT, A.; LU, Y. H.; LEÓN, D. G. D.; GRIVET, L.; FELDMANN, P.; LANAUD, C.; GLASZMANN, J. C. A molecular approach to unraveling the genetics of sugarcane, a complex polyploid of the *Andropogoneae* tribe. **Genome**, Birmingham, v. 37, n. 2, p. 222-230, 1994.
- D'HONT, A.; GRIVET, L.; FELDMANN, P.; GLASZMANN, J. C.; RAO, S.;

BERDING, N. Characterisation of the double genome structure of modern sugarcane cultivars (*Saccharum* spp.) by molecular cytogenetics. **Molecular and General Genetics**, Heidelberg, v. 250, n. 4, p. 405-413, 1996.

D'HONT, A. Unraveling the genome structure of polyploids using FISH and GISH; examples of sugarcane and banana. **Cytogenetic and genome research**, v. 109, n. 1-3, p. 27-33, 2005.

DA SILVA, J. A. D. **A methodology for genome mapping of auto- polyploids and its application to sugarcane (*Saccharum* spp.)**. Ph.D. dissertation, Cornell University, Ithaca, Nova York, 1993a.

DA SILVA, J. A. D.; SORRELLS, M. E.; BURNQUIST, W. L.; TANKSLEY, S. D. RFLP linkage map and genome analysis of *Saccharum spontaneum*. **Genome**, Birmingham, v. 36, n. 4, p. 782-791, 1993b.

DA SILVA, J.; HONEYCUTT, R. J.; BURNQUIST, W.; AL-JANABI, S. M.; SORRELLS, M. E.; TANKSLEY, S. D.; SOBRAL, B. W. *Saccharum spontaneum* L. 'SES 208' genetic linkage map combining RFLP-and PCR-based markers. **Molecular Breeding**, New York, v. 1, n. 2, p. 165-179, 1995.

DE BEM OLIVEIRA, I. **Desequilíbrio de ligação e análise de seleção genômica em cana-de-açúcar**. 85f. Tese de mestrado, Universidade Federal de Goiânia, GO, Brazil, 2012.

DE LOS CAMPOS, G.; HICKEY, J. M.; PONG-WONG, R.; DAETWYLER, H. D.; CALUS, M. P. Whole-genome regression and prediction methods applied to plant and animal breeding. **Genetics**, v. 193, n. 2, p. 327-345, 2013.

DESTA, ZERATSION ABERA; ORTIZ, RODOMIRO. Genomic selection: genome-wide prediction in plant improvement. **Trends in plant science**, v. 19, n. 9, p. 592-601, 2014.

DUFOUR, P.; DEU, M.; GRIVET, L.; D'HONT, A.; PAULET, F.; BOUET, M.; LANAUD, C.; GLAZMANN, J. C.; HAMON, P. Construction of a composite sorghum genome map and comparison with sugarcane, a related complex polyploid. **Theoretical and Applied Genetics**, Heidelberg, v. 94, n. 3-4, p. 409-418, 1997.

DUFRESNE, FRANCE; STIFT, MARC; ERGILINO, ROLAND V.; MABLE, BARBARA K. Recent progress and challenges in population genetics of polyploid

organisms: an overview of current state-of-the-art molecular and statistical tools. **Molecular ecology**, v. 23, n. 1, p. 40-69, 2014.

ENDELMAN, JEFFREY B. Ridge regression and other kernels for genomic selection with R package rrBLUP. **The Plant Genome**, v. 4, n. 3, p. 250-255, 2011.

FAOSTAT, FAO. Statistical data. Food and Agriculture Organization of the United Nations, Rome, 2018.

GARCIA, A. A. F.; KIDO, E. A.; MEZA, A. N.; SOUZA, H. M. B.; PINTO, L. R.; PASTINA, M. M.; LEITE, C. S.; DA SILVA, J. A. G.; ULIAN, E. C.; FIGUEIRA, A.; SOUZA, A. P. Development of an integrated genetic map of a sugarcane (*Saccharum* spp.) commercial cross, based on a maximum-likelihood approach for estimation of linkage and linkage phases. **Theoretical and Applied Genetics**, Heidelberg, v. 112, n. 2, p. 298-314, 2006.

GARCIA, A. A.; MOLLINARI, M.; MARCONI, T. G.; SERANG, O. R.; SILVA, R. R.; VIEIRA, M. L. C.; VICENTINI, R.; COSTA, E. A.; MANCINI, M. C.; GARCIA, M. O. S.; PASTINA, M. M.; GAZAFFI, R.; MARTINS, E. R. F.; DAHMER, N.; SFORÇA, D. A.; SILVA, C. B. C.; BUNDOCK, P.; HENRY, R. J.; SOUZA, G. M.; SLUYS, M.-A.; LANDELL, M. G. A.; CARNEIRO, M. S.; VICENTZ, M. A. G.; PINTO, L. R.; VENCOVSKY, R.; SOUZA, A. P. SNP genotyping allows an in-depth characterisation of the genome of sugarcane and other complex autopolyploids. **Scientific reports**, v. 3, n. 3399, Dez. 2013.

GEORGE, A. W.; AITKEN, K. A new approach for copy number estimation in polyploids. **Journal of heredity**, Inglaterra, v. 101, n. 4, p. 521-524, 2010.

GOUY, M.; RUSSELL, Y.; BASTIANELLI, D.; LECOMTE, P.; BONNAL, L.; ROQUES, D.; EFILE, J.-C.; ROCHER, S.; DAUGROIS, J.; TOUBI, L.; NABENEZA, S.; HERVOUET, C.; TELISMART, H.; DENIS, M.; THONG-CHANE, A.; GLASZMANN, J. C.; HOARAU, J.-Y.; NIBOUCHE, S.; COSTET, L. Experimental assessment of the accuracy of genomic selection in sugarcane. **Theoretical and Applied Genetics**, Nova York: NY, v. 126, n. 10, p. 2575-2586, 2013.

GRANDKE, F.; SINGH, P.; HEUVEN, H. C.; DE HAAN, J. R.; METZLER, D. Advantages of continuous genotype values over genotype classes for GWAS in higher

polyploids: a comparative study in hexaploid chrysanthemum. **BMC genomics**, v. 17, n. 1, p. 672, 2016.

GRIFFITHS, A. J. F.; WESSLER, S. R.; LEWONTIN, R. C.; GELBART, W. M.; SUZUKI, D. T. **Introdução à genética**. Tradução de Paulo Armando Motta. 9. ed. Rio de Janeiro: G. Koogan, 2008. 726 p.

GRIVET, L.; ARRUDA, P. Sugarcane genomics: depicting the complex genome of an important tropical crop. **Current Opinion in Plant Biology**, Nova York. v. 5, n. 2, p. 122-127, 2002.

GRIVET, L.; D'HONT, A.; ROQUES, D.; FELDMANN, P.; LANAUD, C.; GLASZMANN, J. C. RFLP mapping in cultivated sugarcane (*Saccharum* spp.): genome organization in a highly polyploid and aneuploid interspecific hybrid. **Genetics**, Pittsburg, v. 142, n. 3, p. 987-1000, 1996.

GUIMARÃES, C. T.; SILLS, G. R.; SOBRAL, B. W. S. Comparative mapping of Andropogoneae: *Saccharum* L.(sugarcane) and its relation to sorghum and maize. **Proceedings of the National Academy of Sciences**, Washington, v. 94, n. 26, p. 14261-14266, 1997.

GUO, M.; DAVIS, D.; BIRCHLER, J. A. Dosage effects on gene expression in a maize ploidy series. **Genetics**, Pittsburg , v. 142, n. 4, p. 1349-1355, 1995.

GUO, Z.; TUCKER, D .M.; BASTEN, C. J.; GANDHI, H.; ERSOZ, E.; GUO, B.; XU, Z.; WANG, D.; GAY, G. The impact of population structure on genomic prediction in stratified populations. **Theoretical and Applied Genetics**, p. 1-14, 2014.

HABIER, D.; FERNANDO, R. L.; DEKKERS, J. C. M. The impact of genetic relationship information on genome-assisted breeding values. **Genetics**, v. 177, n. 4, p. 2389-2397, 2007.

HALDANE, J. BS. Theoretical genetics of autopolyploids. **Journal of genetics**, v. 22, n. 3, p. 359-372, 1930.

HIETER, PHILIP; GRIFFITHS, TONY. Polyploidy more is more or less. **Science**, v. 285, n. 5425, p. 210-211, 1999.

KEMPTHORNE, OSCAR. The correlation between relatives in a random mating population. **Proc. R. Soc. Lond. B. The Royal Society**, 1954. p. 103-113.

- LU, Y. H.; D'HONT, A.; PAULET, F.; GRIVET, L.; ARNAUD, M.; GLASZMANN, J. Molecular diversity and genome structure in modern sugarcane varieties. **Euphytica**, Holanda, v. 78, n. 3, p. 217-226, 1994a.
- LU, Y. H.; D'HONT, A.; WALKER, D. I. T.; RAO, P. S.; FELDMANN, P.; GLASZMANN, J. C. Relationships among ancestral species of sugarcane revealed with RFLP using single copy maize nuclear probes. **Euphytica**, Wageningen, v. 78, n. 1-2, p. 7-18, 1994b.
- LUBY, James J. et al. Blueberries and cranberries (*Vaccinium*). **Genetic Resources of Temperate Fruit and Nut Crops** **290**, p. 393-458, 1991.
- LYRENE, P. M.; VORSA, N.; BALLINGTON, J. R. Polyploidy and sexual polyploidization in the genus *Vaccinium*. **Euphytica**, v. 133, n. 1, p. 27-36, 2003.
- LYRENE, P. M. Breeding low-chill blueberries and peaches for subtropical areas. **HortScience**, v. 40, n. 7, p. 1947-1949, 2005.
- LYRENE, Paul M. Value of various taxa in breeding tetraploid blueberries in Florida. **Euphytica**, v. 94, n. 1, p. 15-22, 1997.
- MENOSSE, M.; SILVA-FILHO, M. C.; VINCENTZ, M.; VAN-SLUYS, M. A.; SOUZA, G. M. Sugarcane functional genomics: gene discovery for agronomic trait development. **International Journal of Plant Genomics**, New York, v. 2008, n. 458732, 11 p. 2008.
- MEUWISSEN, T. H. E.; HAYES, B. J.; GODDART, M. E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, Heidelberg, v. 157, n. 4, p. 1819-1829, 2001.
- MING, R.; LIU, S.C.; LIN, Y.R.; DA SILVA, J.; WILSON, W.; BRAGA, D.; VAN DEYNZE, A.; WENSLAFF, T. F.; WU, K. K.; MOORE, P. H.; BURNQUIST, W.; SORRELLS, M. E.; IRVINE, J. E.; PATERSON, H. Detailed alignment of *Saccharum* and *Sorghum* chromosomes: comparative organization of closely related diploid and polyploid genomes. **Genetics**, Pittsburg, v. 150, n. 4, p. 1663-1682, 1998.
- MING, R.; S.C.; L.; MORRE, P. H.; IRVINE, J. E.; PATERSON, A. H. QTL analysis in a complex autopolyploid: genetic control of sugar content in sugarcane. **Genome Research**, Nova York, v. 11, n. 12, p. 2075-2084, 2001.

- MOLLINARI, M.; SERANG, O. Quantitative SNP genotyping of polyploids with MassARRAY and other platforms. **Plant Genotyping Methods Protoc.**, p. 215–241, 2015.
- MOORE, G.; DEVOS, K. M.; WANG, Z.; GALE, M. D.. Cereal genome evolution: grasses, line up and form a circle. **Current Biology**, Massachusetts, v. 5, n. 7, p. 737-739, 1995.
- MULLER, Hermann J. A new mode of segregation in Gregory's tetraploid Primulas. **American Naturalist**, p. 508-512, 1914.
- MUÑOZ, PATRICIO R.; RESENDE, M. F.; GEZAN, S. A.; RESENDE, M. D. V.; DE LOS CAMPOS, G.; KIRST, M.; HUBER, D., PETER, G. F.. Unraveling additive from nonadditive effects using genomic relationship matrices. **Genetics**, v. 198, n. 4, p. 1759-1768, 2014.
- NYINE, MOSES; UWIMANA, BRIGITTE; BLAVET, NICOLAS; HIBOVÁ, EVA; VANRESPAILLE, HELENA; BATTE, MICHAEL; AKECH, VIOLET; BROWN, ALLAN; LORENZEN, JIM; SWENNEN; DALEZEL, JAROSLAV. Genomic Prediction in a Multiploid Crop: Genotype by Environment Interaction and Allele Dosage Effects on Predictive Ability in Banana. **The Plant Genome**, 2018.
- OSBORN, T. C.; PIRES, J. C.; BIRCHLER, J. A.; AUGER, D. L.; CHEN, Z. J.; LEE, H.-S.; COMAI, L.; MADLUNG, A.; DOERGE, R. W.; COLOT, V.; MARTIENSSEN, R. A. Understanding mechanisms of novel gene expression in polyploids. **Trends in genetics**, Cambridge, MA, v. 19, n. 3, p. 141-147, 2003.
- PIKAARD, Craig S. Genomic change and gene silencing in polyploids. **Trends in Genetics**, Cambridge, MA, v. 17, n. 12, p. 675-677, 2001.
- PIPERIDIS, G.; PIPERIDIS, N.; D'HONT, A. Molecular cytogenetic investigation of chromosome composition and transmission in sugarcane. **Molecular Genetics and Genomics**, New York, v. 284, n. 1, p. 65-73, 2010.
- R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Available from: <https://www.R-project.org/>, 2018.
- RAMSEY, JUSTIN; SCHEMSKE, DOUGLAS W. Pathways, mechanisms, and rates of polyploid formation in flowering plants. **Annual Review of Ecology and Systematics**, v. 29, n. 1, p. 467-501, 1998.

DE RESENDE, M. D. V.; RESENDE JUNIOR, M. F. R.; AGUIAR, A. M.; ABAD, J. I. M.; MISSIAGGIA, A. A.; SANSALONI, C. P.; PETROLI, C.; GRATTAPAGLIA, D. **Computação da seleção genômica ampla (GS)**. 1ª ed. Colombo: Embrapa Florestas 2010. 79p.

RESENDE, M. F. R.; MUÑOZ, P. R.; RESENDE, M. D.; GARRICK, D. J.; FERNANDO, R. L.; DAVIS, J. M.; JOKELA, E. J.; MARTIN, T. A.; PETER, G. F.; KIRST, M. Accuracy of genomic selection methods in a standard data set of loblolly pine (*Pinus taeda* L.). **Genetics**, v. 190, n. 4, p. 1503-1510, 2012.

RIPOL, M. I.; CHURCHILL, G. A.; DA SILVA, J. A.; SORRELLS, M. Statistical aspects of genetic mapping in autopolyploids. **Gene**, v. 235, n. 1, p. 31-41, 1999.

RITTER, E.; GEBHARDT, C.; SALAMINI, F. Estimation of recombination frequencies and construction of RFLP linkage maps in plants from crosses between heterozygous parents. **Genetics**, Pittsburg, v. 125, n. 3, p. 645-654, 1990.

ROSSI, M.; ARAUJO, P. G.; PAULET, F.; GARSMEUR, O.; DIAS, V. M.; CHEN, H.; VAN SLUYS, M. -A.; D'HONT, A. Genomic distribution and characterization of EST-derived resistance gene analogs (RGAs) in sugarcane. **Molecular Genetics and Genomics**, Alemanha, v. 269, n. 3, p. 406-419, 2003.

SHARPE, R. H.; DARROW, G.M. Breeding blueberries for the Florida climate. In: **Proc. Fla. State Hort. Soc.** 1959. p. 308-311.

SIMS, R. E.; HASTINGS, A.; SCHLAMADINGER, B.; TAYLOR, G.; SMITH, P. Energy crops: current status and future prospects. **Global Change Biology**, Massachusetts, v. 12, n. 11, p. 2054-2076, 2006.

SLATER, A. T.; WILSON, G. M.; COGAN, N. O.; FORSTER, J. W.; HAYES, B. J. Improving the analysis of low heritability complex traits for enhanced genetic gain in potato. **Theoretical and applied genetics**, v. 127, n. 4, p. 809-820, 2013.

SLATER, ANTHONY T.; COGAN, N. O.; FORSTER, J. W.; HAYES, B. J.; DAETWYLER, H. D. Improving genetic gain with genomic selection in autotetraploid potato. **The plant genome**, v. 9, n. 3, 2016.

SOBRAL, B. W. S.; HONEYCUTT, R. J. High output genetic mapping of polyploids using PCR-generated markers. **Theoretical and Applied Genetics**, Heidelberg, v. 86, n. 1, p. 105-112, 1993.

SOLTIS, D. E.; SOLTIS, P. S. Polyploidy: recurrent formation and genome evolution. **Trends in Ecology & Evolution**, v. 14, n. 9, p. 348-352, 1999.

SOLTIS, P. S.; SOLTIS, D. E. The role of genetic and genomic attributes in the success of polyploids. **Proceedings of the National Academy of Sciences**, v. 97, n. 13, p. 7051-7057, 2000.

UITDEWILLIGEN, JAN GAML WOLTERS, A. M. A.; BJORN, B.; BORM, T. J.; VISSER, R. G.; VAN ECK, H. J. A next-generation sequencing method for genotyping-by-sequencing of highly heterozygous autotetraploid potato. **PloS one**, v. 8, n. 5, p. e62355, 2013.

VENCOVSKY, R.; BARRIGA, P. Genética biométrica no fitomelhoramento. **Revista Brasileira de Genética**, Ribeirão Preto, 496p., 1992.

VOORRIPS RE, GORT G, VOSMAN B. Genotype calling in tetraploid species from bi-allelic marker data using mixture models. **BMC Bioinformatics**, v. 12, n. 1, p. 172, 2011.

WIGGANS, G. R.; VANRADEN, P. M.; COOPER, T. A. The genomic evaluation system in the United States: Past, present, future. **Journal of dairy science**, v. 94, n. 6, p. 3202-3211, 2011.

WU, K. K.; BURNQUIST, W.; SORRELS, M. E.; TEW, T. L.; MOORE, P. H.; TANKSLEY, S. D. The detection and estimation of linkage in polyploids using single-dose restriction fragments. **Theoretical and Applied Genetics**, Heidelberg, v. 83, n. 3, p. 294-300, 1992.

WU, R.; MA, C. X.; PAINTER, I.; ZENG, Z. B. Simultaneous maximum likelihood estimation of linkage and linkage phases in outcrossing species. **Theoretical Population Biology**, Philadelphia, v. 61, n. 3, p. 349-363, 2002.

ZHANG, J.; NAGAI, C.; YU, Q.; PAN, Y. B.; AYALA-SILVA, T.; SCHNELL, R. J.; COMSTOCK, J.C.; ARUMUGANATHAN, A. K.; MING, R. Genome size variation in three *Saccharum* species. **Euphytica**, Wageningen, v. 185, n. 3, p. 511-519, 2012.

ZHANG, Z.; DING, X.; LIU, J.; DE KONING, D. J.; ZHANG, Q. Genomic selection for QTL-MAS data using a trait-specific relationship matrix. In: **BMC proceedings**. BioMed Central, Londres, 2011. p. S15.