

ITAMARA BOMFIM GOIS

SELEÇÃO GENÔMICA (GWS) E FENÔMICA NO MELHORAMENTO DE Citrus

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Fitotecnia, para obtenção do título de Doctor Science.

VIÇOSA
MINAS GERAIS – BRASIL
2016

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

T

G616s
2016

Gois, Itamara Bomfim, 1986-
Seleção Genômica (GWS) e fenômica no melhoramento de
Citrus / Itamara Bomfim Gois. – Viçosa, MG, 2016.
xi, 55f : il. ; 29 cm.

Orientador: Aluizio Borém de Oliveira.
Tese (doutorado) - Universidade Federal de Viçosa.
Inclui bibliografia.

1. Citrus - Seleção. 2. Melhoramento genético.
3. Diversidade genética. I. Universidade Federal de Viçosa.
Departamento de Fitotecnia. Programa de Pós-graduação em
Fitotecnia. II. Título.

CDD 22. ed. 641.34304

ITAMARA BOMFIM GOIS

SELEÇÃO GENÔMICA (GWS) E FENÔMICA NO MELHORAMENTO DE *Citrus*

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

Aprovada: 01 de julho de 2016



Marcos Deon Vilela de Resende
(Coorientador)



Eveline Teixeira Caixeta



Felipe Lopes da Silva



Antônio Carlos Baião de Oliveira



Aluizio Borém de Oliveira
(Orientador)

"I learned that courage was not the absence of fear, but the triumph over it. The brave man is not he who does not feel afraid, but he who conquers that fear."

Nelson Mandela S2

Aos meus pais, João e Maria, e aos meus irmãos, Camila e Itamar, por todo amor.

Dedico

AGRADECIMENTOS

A Deus, por ter me dado forças para concluir mais uma etapa e por ter me presenteado com pessoas maravilhosas nesta fase da minha vida.

Aos meus pais, João e Maria, por todo amor, incentivo, esforço, ensinamentos, confiança,... Devo tudo a vocês!

Aos meus irmãos, Camila e Itamar, por todo amor e apoio, mesmo com a distância física.

Ao meu orientador, Aluizio Borém, por ter aceitado me orientar, pela confiança, palavras de incentivo, ensinamentos e por ter sido sempre tão cortês.

Ao meu coorientador, Marcos Deon Vilela de Resende, por ter aceitado me orientar, pela amizade, confiança, dedicação e valiosas contribuições para a execução desta pesquisa. Seus conhecimentos foram inovadores e de grande importância para o meu aprendizado e desenvolvimento profissional.

À minha coorientadora, Mariângela Cristofani-Yaly, pela confiança, disponibilização dos dados e por ter sido sempre tão solícita em todas as etapas desta pesquisa.

Aos membros da banca de defesa, Aluizio Borém, Marcos Deon Vilela de Resende, Eveline Teixeira Caixeta, Antônio Carlos Baião de Oliveira e Felipe Lopes da Silva, por toda a contribuição para a melhoria deste trabalho.

À professora Camila Azevedo pela disponibilidade e ajuda para o desenvolvimento desta pesquisa.

Às secretarias do programa de pós-graduação em Fitotecnia, Tati e Rafaela, que sempre foram tão solícitas e acolhedoras.

A amiga-irmã que Viçosa me presenteou, Adriana Uchôa, por todas as gargalhadas, conversas amigas (às vezes um monólogo – só eu falando, óbvio), conselhos e, por me ajudar sempre que precisei. Com você os obstáculos tornaram-se pequenos.

Aos amigos que conquistei em Viçosa e que me acolheram tão bem: Adriana, Adalgisa, Karla, Michele, Karine, Silvaninha, Haroldo e Mayana.

Aos amigos que mesmo distantes se fizeram tão presentes: Lícia, Thalyta, Sheila, Andrea Freire, Susi, Andreza, Júlio, Mariola e Glauber.

Aos professores/pesquisadores Alessandra de Jesus Boari, Robério Anastácio Ferreira e Maria de Fátima Arrigoni-Blank por todo incentivo, amizade e pelo exemplo de pesquisadores.

À Universidade Federal de Viçosa, principalmente ao Programa de Pós-graduação em Fitotecnia e aos professores com os quais tive a oportunidade de adquirir valiosos conhecimentos.

Ao Centro de Citricultura APTA Citros Sylvio Moreira/IAC, Cordeirópolis-SP, pela disponibilização dos dados para que esta pesquisa pudesse ser realizada.

Ao CNPq e CAPES pela concessão das bolsas de estudo.

Muito obrigada!

SUMÁRIO

LISTA DE TABELAS	viii
LISTA DE FIGURAS	ix
RESUMO	x
ABSTRACT	xi
INTRODUÇÃO GERAL	1
REFERÊNCIAS BIBLIOGRÁFICAS	5

CAPÍTULO 1

Avaliação genética e ganhos com a seleção em uma população híbrida de Citrus.....	8
RESUMO	9
ABSTRACT	10
INTRODUÇÃO.....	11
MATERIAL E MÉTODOS.....	13
Caracterização da população e avaliação fenotípica	13
Estimativa dos parâmetros genéticos a partir dos dados fenotípicos.....	14
Correlação genética	15
Índice de seleção – Mulamba e Mock (1978)	16
Diversidade genética.....	16
Eficiência da seleção considerando estrutura de subpopulação	16
RESULTADOS E DISCUSSÃO	18
CONCLUSÕES	28
REFERÊNCIAS BIBLIOGRÁFICAS	29

CAPÍTULO 2

Genome wide selection increases efficiency of Citrus breeding.....	32
RESUMO	33
ABSTRACT	34
INTRODUCTION	35
MATERIAL AND METHODS.....	37
Characterization of the population and phenotypic traits.....	37
Statistical model for the analysis of phenotypic data	37
Genotyping and quality control of the markers	38
Predicting genetic effects of markers using RR-BLUP.....	39
Predictive ability, prediction bias and accuracy of GWS.....	40

Efficiency of GWS	41
RESULTS AND DISCUSSION.....	42
CONCLUSION	52
REFERENCES	53

LISTA DE TABELAS

CAPÍTULO 1

Tabela 1. Análise de deviance (ANADEV) para as variáveis avaliadas na população híbrida de Citrus.....	18
Tabela 2. Estimativas de parâmetros genéticos para as variáveis relacionadas à qualidade de fruto e resistência ao CVC em uma população híbrida de Citrus.....	19
Tabela 3. Correlação genética entre variáveis relacionadas à qualidade de frutos em uma população híbrida de Citrus.....	23
Tabela 4. Classificação dos 10 melhores clones de Citrus com base no índice de Mulamba-Mock (1978).....	24
Tabela 5. Agrupamento dos 180 clones de Citrus pelo método de Tocher, com base na distância genética euclidiana média.....	25
Tabela 6. Análise de deviance (ANADEV) para as variáveis avaliadas na população de Citrus considerando cinco subpopulações.....	26
Tabela 7. Eficiência da seleção com base na estrutura de subpopulação com relação à seleção baseada em informações de clone.....	27

CAPÍTULO 2

Table 1. Estimates of genomic heritability (h_g^2), predictive ability (r_{yg}), prediction bias (b), accuracy of GWS (r_{gg}), QTL number estimates (n_{QTL}) and genetic variance explained by the markers (r_{mq}^2) accounted by the markers when selecting markers based in the magnitudes of their effects in a population of Citrus breeding.....	43
Table 2. Gains in efficiency of GWS regarding selection based solely on phenotypic data in a breeding population of Citrus.....	50

LISTA DE FIGURAS

CAPÍTULO 2

- Figure 1.** Expected proportion of genetic variance explained by markers (r_{mq}^2) with increasing number of markers.....42
- Figure 2.** The accuracy of Genome Wide Selection (GWS) and expected accuracy due to the increase in the number of markers for traits evaluated in a population of Citrus.47

RESUMO

GOIS, Itamara Bomfim, D.Sc., Universidade Federal de Viçosa, Julho, 2016. **Seleção genômica (GWS) e fenômica no melhoramento de Citrus.** Orientador: Aluizio Borém de Oliveira. Coorientadores: Marcos Deon Vilela de Resende e Mariângela Cristofani-Yaly.

O Brasil se destaca como um dos maiores produtores de frutas cítricas do mundo e, emprega o melhoramento genético com o intuito de aumentar a qualidade e produtividade dos produtos. O processo de seleção corresponde a uma das principais etapas de um programa de melhoramento, portanto, a adoção de procedimentos que proporcionem maior acurácia seletiva é de suma importância. A seleção pode ser realizada com base apenas em fenótipos ou com base nos efeitos de marcadores moleculares sobre os fenótipos. Este trabalho foi realizado com o objetivo de estimar parâmetros genéticos, ganhos com a seleção e diversidade genética de genótipos provenientes de uma população de Citrus utilizando a metodologia de modelos mistos (REML/BLUP); aplicar a seleção genômica ampla (GWS) no melhoramento de Citrus e verificar a sua eficiência sobre a seleção baseada apenas em fenótipos. Neste estudo foi analisada uma população híbrida de Citrus composta por 180 indivíduos. Estes, foram fenotipados para 15 variáveis relacionadas à qualidade de frutos e resistência à clorose variegada dos citros (CVC), e genotipados com 5.287 marcadores moleculares DArT_seqTM. Com base nos resultados, observou-se que a população de Citrus avaliada apresenta potencial para o melhoramento genético para a maioria das variáveis avaliadas (11) e, a introdução de novos genótipos deve ser realizada para aumentar o potencial de seleção da população para as variáveis que não apresentaram variabilidade genética. A seleção genômica ampla (GWS) apresenta superioridade à seleção com base apenas em dados fenotípicos, em termos de ganho genético por unidade de tempo. Portanto, GWS é uma ferramenta promissora para o melhoramento genético do Citrus, uma vez que, proporciona acuradas predições dos fenótipos em estádios iniciais do desenvolvimento da planta.

ABSTRACT

GOIS, Itamara Bomfim, D.Sc., Universidade Federal de Viçosa, July, 2016. **Genome Selection (GWS) and phenomics in Citrus breeding.** Adviser: Aluizio Borém de Oliveira. Co-advisers: Marcos Deon Vilela de Resende and Mariângela Cristofani-Yaly.

Brazil stands out as one of the largest producers of Citrus fruits in the world, and employs breeding in order to increase the quality and productivity of products. The selection process is one of the main phases of a breeding program; therefore, the adoption of procedures that provide greater selective accuracy is very important. The selection can be performed based only on phenotypic effects or based on molecular markers effects on phenotypes. This work was carried out in order to estimate genetic parameters, gains from genetic selection and diversity of genotypes from a population of Citrus using the mixed model methodology (REML/BLUP); apply Genome Wide Selection (GWS) in the Citrus breeding and verify its effectiveness on the selection based only on phenotypes. In this study was evaluated a hybrid population of Citrus composed of 180 individuals. The individuals were phenotyped for 15 variables related to fruit quality and resistance to Citrus variegated chlorosis (CVC), and genotyped with 5,287 molecular markers DArT_seqTM. Based on the results, it was observed that the population of Citrus evaluated presents potential for genetic improvement for the majority of variables (11), and the introduction of new genotypes should be performed to increase the selection potential of the population for variables that did not present genetic variability. The GWS shows superiority to the selection based only on phenotypic data, in terms of genetic gain per unit time. Therefore, GWS is a promising tool in Citrus breeding, since it provides accurate predictions of phenotypes in the early stages of plant development.

INTRODUÇÃO GERAL

O gênero *Citrus* representa uma das principais culturas de importância econômica em nível mundial. O Brasil se destaca como o segundo maior produtor de frutas cítricas, sendo responsável por aproximadamente 15,5% da produção mundial, e como maior produtor de laranja, sendo responsável por aproximadamente 26,4% (FAO, 2016). Devido ao seu potencial, pesquisas em melhoramento genético têm sido realizadas com o intuito de aumentar a produtividade e a qualidade dos produtos.

Como apresenta alta plasticidade ecológica, o gênero *Citrus* é cultivado em regiões tropicais e subtropicais, sendo frequentemente encontrado em regiões que apresentam severos estresses abióticos, tais como déficit hídrico e salinidade. Além disso, a cultura apresenta susceptibilidade a diversas doenças e pragas que afetam consideravelmente o crescimento e o rendimento da cultura. A utilização de genótipos geneticamente superiores é, sem dúvida, o melhor método para contornar estas situações, no entanto, a utilização de técnicas tradicionais de melhoramento em *Citrus* apresenta uma série de dificuldades devido à combinação interespecífica, a baixa diversidade biológica presente nas populações de seleção e o tempo demandado para o melhoramento genético (Teroll et al., 2007; Teroll et al., 2008).

A fase inicial de um programa de melhoramento genético compreende as etapas de avaliação e seleção de genótipos superiores (Correa et al., 2015). Para tanto, é necessário realizar a caracterização morfológica/ agrônômica dos candidatos à seleção, bem como os estudos de diversidade genética e estimação dos valores genéticos dos indivíduos para, assim, selecionar aqueles portadores de alelos desejáveis (Farias Neto et al., 2013). No caso do *Citrus*, deve-se fazer a seleção com base nos valores genotípicos totais dos indivíduos, uma vez que é realizada a propagação clonal. Tais etapas tendem a ser morosas e trabalhosas quando se está trabalhando com espécies perenes como o *Citrus*, o qual apresenta um longo período juvenil. Nesse sentido, a estimação acurada dos parâmetros genéticos e dos valores genotípicos dos indivíduos em avaliação é primordial para o sucesso do programa de melhoramento e, consequentemente, a utilização de métodos estatísticos mais eficientes torna-se indispensável (Alves e Resende, 2008).

Atualmente, o procedimento estatístico que tem sido recomendado para estimação/predição no melhoramento de plantas perenes é o REML/BLUP, sendo os componentes de variância estimados por máxima verossimilhança restrita (REML) e os

valores genéticos preditos por meio do melhor preditor linear não viesado (BLUP) (Farias Neto e Resende, 2001; Resende et al., 2001). Diversos autores tem constatado a eficiência desta abordagem no melhoramento de plantas perenes devido à maximização dos ganhos com a seleção (Farias Neto et al., 2009; Arantes et al., 2010; Costa et al., 2010; Pompeu Junior, 2013; Maia et al., 2014; Oliveira et al., 2014). Sendo que em Citrus, apenas um artigo foi publicado (Pompeu Junior et al., 2013).

As principais vantagens relatadas para o uso da metodologia de modelos mistos (REML/BLUP) são: pode ser aplicado a dados desbalanceados; não exige dados obtidos sob estruturas rígidas de experimentação; permite utilizar simultaneamente um grande número de informações provenientes de vários experimentos, gerando estimativas mais precisas em relação aos métodos baseados em quadrados mínimos; corrige os dados para efeitos ambientais e prediz de maneira precisa e não viesada os valores genotípicos, conduzindo à maximização do ganho genético com a seleção (Resende e Rosa-Perez, 2001; Resende et al., 2001; Resende, 2007)

A utilização de métodos estatísticos aprimorados é uma das formas de aumentar a eficiência da seleção fenotípica, no entanto, o processo de seleção em plantas perenes permanece demorado e oneroso, principalmente se as características avaliadas são expressas em estádios tardios do desenvolvimento da planta (Resende Junior et al. 2012). A utilização de ferramentas biotecnológicas, como os marcadores moleculares, foi proposta como forma de minimizar estes entraves.

A seleção assistida por marcadores (SAM) foi proposta por Lande e Thompson (1990) como forma de aumentar a eficiência da seleção baseada apenas em dados fenotípicos. A SAM utiliza simultaneamente dados fenotípicos e dados de marcadores moleculares que estejam em ligação gênica com alguns locos controladores de características quantitativas (QTL). Estes marcadores são eleitos ou não como determinantes dos efeitos de QTL após a modelagem estatística associada ao erro do tipo II, o qual se refere à probabilidade de aceitar uma hipótese falsa de ausência de efeitos (Resende et al., 2012). A justificativa para a utilização da genética molecular no melhoramento de plantas é que o uso de informação em nível de DNA permitirá ganhos mais rápidos e maior eficiência seletiva quando comparado à seleção baseada apenas em dados fenotípicos (Meuwissen, 2001; Resende et al., 2012).

A utilização de marcadores baseados em DNA, como RAPD (Random Amplified Polymorphic DNA), AFLP (Amplified fragment length polymorphism), SSR (Simple Sequence Repeats), TRAP (Target Region Amplification Polymorphism) e SNP (Single

Nucleotide Polymorphism), possibilitou a construção de mapas genéticos saturados na cultura do Citrus. A construção destes mapas é a única forma de cobrir o genoma sem o uso de sequenciamento e é de suma importância para o entendimento dos mecanismos de herança para resistência ou susceptibilidade a fatores bióticos e abióticos e para a localização de genes ligados às características que apresentam importância agrônômica (Machado et al., 2011).

Atualmente, a disponibilidade de marcadores moleculares que cobrem o genoma de maneira densa propiciou a introdução da seleção genômica ampla (GWS) no melhoramento genético. A GWS foi proposta por Meuwissen et al. (2001) que, usando dados simulados, mostraram que com o uso de milhares de marcadores é possível estimar os valores genéticos genômicos dos indivíduos sem informações fenotípicas (Moser et al., 2009), depois que os efeitos dos marcadores forem estimados em uma população de treinamento. Portanto, a GWS foi proposta como uma alternativa para aumentar a eficiência dos programas de melhoramento genético, e consiste na seleção simultânea de centenas ou milhares de marcadores de forma que todos os genes de um caráter quantitativo e qualitativo estejam em desequilíbrio de ligação com pelo menos uma parte dos marcadores, não sendo necessário o estabelecimento de fases de ligação em cada família (Meuwissen et al., 2001; Meuwissen, 2007; Resende, 2007; Resende, 2008; Resende et al., 2008).

A GWS enfatiza a predição simultânea, sem o uso de testes de significância para marcas individuais, dos efeitos genéticos dos marcadores de DNA dispersos no genoma, de forma a capturar os efeitos dos locos e explicar toda a variação genética de um caráter quantitativo ou qualitativo. Para tal, é necessário que haja desequilíbrio de ligação, em nível populacional, entre alelos dos marcadores e alelos dos genes que controlam o caráter (Resende et al., 2013). Esses marcadores em desequilíbrio de ligação com os QTL explicarão quase/ ou a totalidade da variação genética de um caráter quantitativo ou qualitativo. Além disso, a predição por meio de marcadores tende a ser mais acurada, uma vez que, considera o real parentesco genético dos indivíduos em avaliação, em detrimento do parentesco médio esperado matematicamente (Resende, 2007; Resende, 2008).

As propriedades que tornam a GWS promissora para o uso nos programas de melhoramento genético são: a não necessidade de mapeamento de QTL; a utilização de uma alta densidade de marcadores que permite obter o real parentesco genético entre os indivíduos, como também o desequilíbrio de ligação no genoma que é suficiente para

informar sobre os efeitos dos QTL; possibilitar, de forma eficiente, a seleção de caracteres que apresentam baixa herdabilidade; permitir a melhor utilização dos recursos genéticos disponíveis por meio da seleção de cruzamentos de acordo com as informações genéticas; e, aumentar o ganho genético por unidade de tempo, pois propicia a seleção precoce direta (Daetwyler et al., 2013; Legarra et al., 2008; Resende et al., 2012).

Diante do exposto, este trabalho foi realizado com os seguintes objetivos: (i) estimar os parâmetros genéticos, correlação genética, ganhos com a seleção e diversidade genética de genótipos provenientes de uma população híbrida de Citrus utilizando a metodologia de modelos mistos (REML/BLUP); (ii) aplicar de forma pioneira a seleção genômica ampla (GWS) no melhoramento de Citrus e verificar sua eficiência sobre a seleção baseada apenas em fenótipos.

REFERÊNCIAS BIBLIOGRÁFICAS

- ALVES, R.M.; RESENDE, M.D.V. Avaliação genética de indivíduos e progênes de cupuaçuzeiro no estado do Pará e estimativas de parâmetros genéticos. **Rev. Bras. Frutic.** 30(3): 696-701, 2008.
- ARANTES, F.C.; GONÇALVES, P.S.; SCALOPPI JUNIOR, E.J.; MORAES, M.L.T.; RESENDE, M.D.V. Ganho genético com base no tamanho efetivo populacional de progênes de seringueira. **Pesq. agropec. bras.** 45(12): 1419-1424, 2010.
- CORREA, E.R.; MEDEIROS, G.C.R.; BARROS, W.S.; BIANCHI, V.J. Evaluation and ranking of Macadamia genotypes using mixed models. **Afr. J. Agric. Res.** 10(38): 3696-3703, 2015.
- COSTA, R.B.; RESENDE, M.D.V.; GONÇALVES, P.S.; ROA, R.A.R.; FEITOSA, K.C.O. Predição de parâmetros e valores genéticos para caracteres de crescimento e produção de látex em progênes de seringueira. **Bragantia** 69(1): 49-56, 2010.
- DAETWYLER, A.D.; CALUS, M.P.L.; PONG-WONG, R.; DE LOS CAMPOS, G.; HICKEY, J.M. Genomic Prediction in Animals and Plants: Simulation of Data, Validation, Reporting, and Benchmarking. **Genetics** 193: 347-365, 2013.
- FAO - **FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED.** Disponível em: <http://faostat.fao.org/>; Acessado em: 10 de junho de 2016.
- FARIAS NETO, J.T.; CLEMENT, C.R.; RESENDE, M.D.V. Estimativas de parâmetros genéticos e ganho de seleção para produção de frutos em progênes de polinização aberta de pupunheira no Estado do Pará, Brasil. **Bragantia** 72(2): 122-126, 2013.
- FARIAS NETO, J.T.; LINS, P.M.P.; RESENDE, M.D.V.; MULLER, A.A. Seleção genética em progênes híbridas de coqueiro. **Rev. Bras. Frutic.** 31(1): 190-196, 2009.
- FARIAS NETO, J.T.; RESENDE, M.D.V. Aplicação da metodologia de modelos mistos (REML/BLUP) na estimação de componentes de variância e predição de valores genéticos em pupunheira (*Bactris gasipaes*). **Rev. Bras. Frutic.** 23(2): 320-324, 2001.
- LANDE, R.; THOMPSON, R. Efficiency of marker-assisted selection in the improvement of quantitative traits. **Genetics** 124: 743-756, 1990.
- LEGARRA, A.; ROBERT, GRANIÉ, C.; MANFREDI, E.; ELSSEN, J.M. Performance of genomic selection in mice. **Genetics** 180: 611-618, 2008.
- MACHADO, M.A.; CRISTOFANI-YALY, M.; BASTIANEL, M. Breeding, genetic and genomic of Citrus for disease resistance. **Revista Brasileira de Fruticultura**, Volume especial, p. 158-172. 2011.
- MAIA, M.C.C.; RESENDE, M.D.V.; OLIVEIRA, L.C.; VASCONCELOS, L.F.L.; LIMA NETO, J.F.P. Análise genética em genótipos de manga rosa via REML/BLUP. **Revista Agrotecnologia** 5(1): 1-16, 2014.

MEUWISSEN, T.H.E.; HAYES, B.J.; GODDARD, M.E. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. **Genetics** n. 157, p. 1819–1829, 2001.

MEUWISSEN, T.H.E. Genomic selection: marker assisted selection on a genome wide scale. **J. Anim. Breed. Genet.** 124: 321–322, 2007.

MOSER, G.; TIER, B. CRUMP, R.E.; KHATKAR, M.S.; RAADSMA, H.W. A comparison of five methods to predict genomic breeding values of dairy bulls from genome-wide SNP markers. **Genetics Selection Evolution** 41(56): 1-16, 2009.

OLIVEIRA, E.J.; SANTANA, F.A.; OLIVEIRA, L.A.; SANTOS, V.S. Genetic parameters and prediction of genotypic values for root quality traits in cassava using REML/BLUP. **Genetics and Molecular Research** 13(3): 6683-6700, 2014.

POMPEU JUNIOR, J.; BLUMER, S.; RESENDE, M.D.V. Avaliação genética de seleções e híbridos de limões cravo, volkameriano e rugoso como porta-enxertos para laranjeiras Valência na presença da morte súbita dos citros. **Rev. Bras. Frutic.** 35(1): 199-209, 2013.

RESENDE, M.D.V. **Seleção Genômica Ampla (GWS) e Modelos Lineares Mistos.** In: RESENDE, M.D.V. Matemática e Estatística na Análise de Experimentos. Colombo: Embrapa Florestas, 2007. 362p. Cap. 11. p. 517-533.

RESENDE, M.D.V. **Genômica quantitativa e seleção no melhoramento de plantas perenes e animais.** Colombo: Embrapa Florestas, 2008, 330p.

RESENDE, M.D.V.; FURLANI-JUNIOR, E.; MORAES, M.L.T.; FAZUOLI, L.C. Estimativas de parâmetros genéticos e predição de valores genotípicos no melhoramento do cafeeiro pelo procedimento REML/BLUP. **Bragantia** 60(3): 185-193, 2001.

RESENDE, M.D.V.; LOPES, P.S.; SILVA, R.L.; PIRES, I.E. Seleção genômica ampla (GWS) e maximização da eficiência do melhoramento genético. **Pesquisa Florestal Brasileira** 56: 63-77, 2008.

RESENDE, M. D. V.; ROSA-PEREZ, J. R. H. **Genética e Melhoramento de Ovinos.** Curitiba. Ed. UFPR, 2001. 185p.

RESENDE, M.D.V.; SILVA, F.F.; LOPES, P.S.; AZEVEDO, C.F. **Seleção Genômica Ampla (GWS) via Modelos Mistos (REML/BLUP), Inferência Bayesiana (MCMC), Regressão Aleatória Multivariada e Estatística Espacial.** Viçosa: universidade Federal de Viçosa/ Departamento de Estatística. 2012. 291p.

RESENDE, M.D.V.; SILVA, F.F.; RESENDE JUNIOR, M.F.R. **Seleção Genômica Ampla.** In: Biotecnologia aplicada ao melhoramento de plantas. BORÉM, A.; FRITSCHÉ-NETO, R. Suprema, 2013, 336p.

TEROL, J.; CONESA, A; COLMENERO, J.M.; CERCOS, M.; TADEO, F.; AGUSTÍ, J.; ALÓS, E.; ANDRES, F.; SOLER, G.; BRUMOS, J.; IGLESIAS, D.J.; GOTZ, S.; LEGAZ, F.; ADOURT, X.; COURTOIS, B.; OLLITRAULT, P.; DOSSAT, C.;

WINCKER, P.; MORILLON, R.; TALON, M. Analysis of 13000 unique Citrus clusters associated with fruit quality, production and salinity tolerance. **BMC Genomics**, n. 8, v. 31, p. 1-22, 2007.

TEROL, J.; NARANJO, M.A.; OLLITRAULT, P.; TALON, M. Development of genomic resources for Citrus clementina: Characterization of three deep-coverage BAC libraries and analysis of 46,000 BAC end sequences. **BMC Genomics**, n. 9, v. 423, p. 1-12, 2008.

CAPÍTULO 1

**Avaliação genética e ganhos com a seleção em uma população
híbrida de Citrus**

Viçosa – MG

Julho, 2016

RESUMO

GOIS, Itamara Bomfim, D.Sc., Universidade Federal de Viçosa, Julho, 2016. **Avaliação genética e ganhos com a seleção em uma população híbrida de Citrus.** Orientador: Aluizio Borém de Oliveira. Coorientadores: Marcos Deon Vilela de Resende e Mariângela Cristofani-Yaly.

A estimação acurada dos parâmetros genéticos de uma população de melhoramento proporciona a maximização do ganho genético com a seleção. Este trabalho foi realizado com o objetivo de estimar a herdabilidade, correlação genética, ganhos com a seleção e diversidade genética de genótipos provenientes de uma população híbrida de Citrus utilizando a metodologia de modelos mistos (REML/BLUP). O experimento foi instalado em delineamento inteiramente ao acaso, com três repetições clonais, e, 15 variáveis relacionadas à qualidade de frutos e resistência à clorose variegada dos citros (CVC) foram avaliadas. As variáveis espessura da casca, facilidade de descasque, sólidos solúveis e resistência ao CVC não apresentaram variabilidade genética, indicando que a população não apresenta potencial para seleção para estas variáveis. As estimativas de herdabilidade, para as demais variáveis, assumiram valores entre 0,11 e 0,59, demonstrando a possibilidade de ganhos efetivos com a seleção, resultados que estão de acordo com as estimativas de acurácia. Ganhos superiores a 50%, com base no índice de Mulamba e Mock, foram observados com uma intensidade de seleção de 5,5% (10 clones). Na análise de diversidade genética, os genótipos foram divididos em cinco grupos, sendo que a maioria foi agrupada no grupo 1. A utilização de estrutura de subpopulação no modelo estatístico proporcionou eficiência seletiva superior, variando de 2% a 85%, ao modelo que considera apenas informações de clones. A introdução de novos genótipos deve ser realizada para aumentar o potencial de seleção da população para as características que não apresentaram variabilidade genética.

Palavras-chave: modelos mistos; acurácia; eficiência seletiva.

ABSTRACT

GOIS, Itamara Bomfim, D.Sc., Universidade Federal de Viçosa, July, 2016. **Genetic assessment and selection gain in Citrus** Adviser: Aluizio Borém de Oliveira. Co-advisers: Marcos Deon Vilela de Resende e Mariângela Cristofani-Yaly.

The accurate estimation of genetic parameters of a breeding population provides maximizing genetic gain with selection. This study was carried out to estimate heritability, genetic correlation, gains from genetic selection and diversity of genotypes from a hybrid population of Citrus using the mixed model methodology (REML/BLUP). The experiment was conducted in an entirely randomized design with three clonal replications, and, 15 variables related to fruit quality and resistance to Citrus variegated chlorosis (CVC) were evaluated. The variable thickness of the shell, easiness of peeling, soluble solids and resistance to CVC not exhibited genetic variability, indicating that the population has no potential for selection of these variables. The heritability estimates for the other variables, assumed values between 0.11 and 0.59, demonstrating the possibility of effective gains with the selection, results that are consistent with the accuracy's estimates. Gains above 50%, based on Mulamba and Mock index, were observed with an intensity of selection 5.5% (10 hybrids). In genetic diversity analysis, the genotypes were divided into five groups, and most were grouped in group 1. The use of subpopulation structure provided efficiency higher than model considers only clones information, varying of 2 to 85%. The introduction of new genotypes should be performed to increase the population's potential selection for the characteristics that showed no genetic variability.

Key-words: mixed models; accuracy; selective efficiency.

INTRODUÇÃO

O gênero *Citrus* representa uma das principais culturas de importância econômica do mundo, sendo o Brasil o maior produtor de laranja e o segundo maior produtor de frutas cítricas (FAO, 2016). Desta forma, com o intuito de manter a produtividade e qualidade dos produtos produzidos no Brasil, diversas instituições de pesquisa tem utilizado o melhoramento genético como forma de alcançar tais objetivos.

A avaliação e seleção de genótipos superiores compreendem as principais fases de um programa de melhoramento genético (Correa et al., 2015). Portanto, as caracterizações morfológicas e/ou agrônômicas além dos estudos de diversidade entre os genótipos e as estimativas de seus valores genéticos constituem etapas imprescindíveis para o processo de seleção e cruzamentos de genótipos superiores (Resende et al., 2014). Tais etapas tendem a ser morosas e trabalhosas quando se está trabalhando com gêneros perenes como o *Citrus*, o qual apresenta um longo período juvenil. Assim, a adoção de métodos estatísticos mais eficientes para a seleção é de suma importância para o melhoramento genético de espécies perenes devido ao longo prazo e altos custos requeridos para a condução dos programas de melhoramento. Além disso, a seleção baseada na utilização de método estatístico inadequado pode distinguir incorretamente os efeitos genotípicos e ambientais (Alves e Resende, 2008).

Atualmente, o método REML/BLUP (Máxima verossimilhança residual/ Melhor predição linear não viesada), conhecido também como metodologia de modelos mistos (Henderson, 1973), é considerado o procedimento padrão para os estudos de genética quantitativa e para a prática de seleção em espécies perenes (Resende, 2007; Scarpinati et al., 2009; Carias et al., 2014). Desta forma, os componentes de variância, a partir dos quais são estimados os parâmetros genéticos, são obtidos por meio da metodologia REML e os valores genéticos preditos por meio do BLUP (Resende, 2002).

As principais vantagens listadas para o uso da metodologia de modelos mistos (REML/BLUP) são: pode ser aplicado a dados desbalanceados; não exige dados obtidos sob estruturas rígidas de experimentação; permite o uso simultâneo das informações do indivíduo, da família e de medidas repetidas no tempo, além de um grande número de informações provenientes de vários experimentos; corrige os dados para efeitos ambientais e prediz de maneira precisa e não viesada os valores genotípicos, conduzindo à maximização do ganho genético com a seleção (Resende e Rosa-Perez, 2001; Resende et al., 2001; Resende, 2002; Resende, 2007). Diversos autores relataram

a eficiência deste método no melhoramento de plantas perenes devido à maximização dos ganhos com a seleção (Farias Neto et al., 2009; Arantes et al., 2010; Costa et al., 2010; Maia et al., 2014; Oliveira et al., 2014). A utilização da metodologia de modelos mistos (REML/BLUP) não é comum no melhoramento genético de Citrus, sendo citado até o momento apenas no trabalho de Pompeu Junior et al. (2013).

As estimativas dos parâmetros genéticos tais como a herdabilidade e correlações genéticas são fundamentais para o delineamento de eficientes estratégias de melhoramento. Informações a esse respeito são praticamente inexistentes em Citrus. Assim, este trabalho foi realizado com o objetivo de estimar a herdabilidade, correlação genética, ganhos com a seleção e diversidade genética de genótipos provenientes de uma população híbrida de Citrus, utilizando a metodologia de modelos mistos (REML/BLUP), para gerar informações essenciais aos programas de melhoramento genético da cultura.

MATERIAL E MÉTODOS

Caracterização da população e avaliação fenotípica

As análises, com base nos dados fenotípicos, foram realizadas para uma população híbrida de Citrus, utilizada como copa, constituída pelo cruzamento entre laranja Pera (*Citrus sinensis* Osbeck) e tangor Murcott (*Citrus sinensis* Osbeck X *Citrus reticulata* Blanco). Para a formação da população, composta por 180 indivíduos (híbridos), foram utilizados dois genitores.

O experimento foi instalado em 2004, sem irrigação, no Centro de Citricultura Sylvio Moreira do Instituto Agrônomo (IAC), em Cordeirópolis – SP, que apresenta as seguintes coordenadas geográficas: 22°32' S e 47°27' O, altitude de 639 m e clima do tipo Cwa, segundo a classificação de Köppen. O solo é do tipo Latossolo Vermelho distrófico típico. A média de precipitação pluvial anual é de 1.375,3 mm, e a média anual da umidade relativa do ar de 74,2%. A temperatura média anual é de 20,2 °C, sendo a média das máximas igual a 27,5 °C, e a das mínimas 14,5 °C. O delineamento experimental utilizado foi o inteiramente ao acaso (DIC), com três repetições clonais, e espaçamento 6 x 3 m.

A avaliação fenotípica foi realizada no ano de 2012. Uma amostra de cada repetição (clone), composta por cinco frutos, foi utilizada para a obtenção das avaliações fenotípicas, em que: a massa do fruto (g) foi obtida em uma balança Filizola com capacidade para 15 kg e precisão de 5 g; os diâmetros longitudinal e transversal dos frutos (cm) foram obtidos por leitura direta com o auxílio de uma régua graduada; a espessura da casca (mm) foi medida com auxílio de um paquímetro; a facilidade de descasque foi avaliada com notas de 1 a 3 (1 = casca fortemente aderida, 2 = casca medianamente aderida e 3 = casca solta); o número de gomos e de sementes por meio da contagem direta; o número de frutos/caixa foi obtido dividindo-se 40,8 Kg (capacidade da caixa) pelo peso médio dos frutos; o rendimento de suco (mL) determinado após esmagamento do fruto em extratora OIC modelo OTTO 1800 e calculado através da relação massa do suco/massa do fruto; o teor de sólidos solúveis (°Brix) determinado por leitura direta em refratômetro B&S modelo RFM 330; a acidez foi obtida por titulação com uma solução padronizada de 0,3125 N de NaOH, usando-se fenolftaleína como indicadora; o ratio calculado através da relação teor de sólidos solúveis/ acidez, o qual indica o estágio de maturação dos frutos; e a resistência à clorose variegada dos citros (CVC) foi avaliada através de notas de 0 a 3 (0 = sem

sintomas, 1 = um ramo sintomático, 2 = 25% da planta com sintomas e 3 = 50% ou mais da planta com sintomas). As análises foram realizadas no Laboratório de Qualidade e Pós-Colheita do Centro APTA Citros Sylvio Moreira/IAC, Cordeirópolis-SP.

Estimativa dos parâmetros genéticos a partir dos dados fenotípicos

A análise dos dados fenotípicos foi realizada a partir da metodologia REML/BLUP, também denominada metodologia de modelos mistos (Henderson, 1973). O modelo 83 do software Selegen REML/BLUP (Resende, 2002) foi utilizado para a realização da análise, o qual equivale a: $y = Xu + Zg + e$, em que, y é o vetor de dados; u é o escalar referente à média geral (efeito fixo); g é o vetor dos efeitos genotípicos (assumidos como aleatórios); e é o vetor de resíduos (aleatórios). As letras maiúsculas representam as matrizes de incidência para os referidos efeitos.

As estruturas de médias e variâncias associadas ao modelo são: $y|u, V \sim N(Xu, V)$; $g|\sigma_g^2 \sim N(0, I\sigma_g^2)$; $e|\sigma_e^2 \sim N(0, I\sigma_e^2)$; $Cov(g, e) = 0$.

Assim,

$$E \begin{bmatrix} y \\ g \\ e \end{bmatrix} = \begin{bmatrix} Xu \\ 0 \\ 0 \end{bmatrix} \text{ e } Var. \begin{bmatrix} y \\ g \\ e \end{bmatrix} = \begin{bmatrix} V & ZG & R \\ GZ' & G & 0 \\ R & 0 & R \end{bmatrix}$$

$G = I\sigma_g^2$ = Matriz de variância – covariância dos efeitos aleatórios;

$R = I\sigma_e^2$ = Matriz de variância – covariância dos erros aleatórios;

I = Matriz Identidade;

V = Variância fenotípica e equivale a: $Z\sigma_g^2Z' + I\sigma_e^2 = ZGZ' + R$.

σ_g^2 e σ_e^2 são as variâncias genotípica e residual, respectivamente.

As equações de modelos mistos para o BLUP (Best Linear Unbiased Predictor) dos valores genéticos individuais equivalem a:

$$\begin{bmatrix} \hat{u} \\ \hat{g} \end{bmatrix} = \begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + I\lambda \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

Em que:

$$\lambda = \sigma_e^2 / \sigma_g^2 = \frac{1-h_g^2}{h_g^2},$$

$$h_g^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2} : \text{herdabilidade individual no sentido amplo.}$$

Os componentes de variância foram estimados por meio do procedimento REML (Restricted Maximum Likelihood - Patterson e Thompson, 1971), a partir de iterações nas equações de modelos mistos, cujos estimadores são:

$$\sigma_g^2 = [\hat{g}'\hat{g} + \sigma_e^2 \text{tr}(C^{22})]/q: \text{variância genotípica};$$

$$\sigma_e^2 = [y'y - \hat{u}'X'y - \hat{g}'Z'y]/[N - r(X)]: \text{variância residual}.$$

C^{22} provém da inversa generalizada da matriz dos coeficientes das equações de modelos mistos;

tr: operador traço matricial;

r(X): posto da matriz X;

N-r(X): número de graus de liberdade do erro;

N e q: número total de dados e de indivíduos, respectivamente.

A herdabilidade ao nível de média foi calculada a partir da herdabilidade individual por meio da expressão (Resende et al., 2001): $h_{gm}^2 = \frac{n_i h_g^2}{1 + (n_i - 1)h_g^2}$, em que n_i é o número de repetições. A acurácia da predição dos valores genotípicos dos indivíduos avaliados foi calculada a partir da inversa do lado esquerdo das equações de modelos mistos (Resende et al., 2002).

O coeficiente de variação genética (CV_g) foi calculado por meio da fórmula:

$$CV_g = \frac{\sqrt{\sigma_g^2}}{\mu} \times 100 \text{ e, o coeficiente de variação residual por: } CV_e = \frac{\sqrt{\sigma_e^2}}{\mu} \times 100.$$

Correlação genética

A correlação genética entre as variáveis avaliadas foi estimada de acordo com Falconer e Mackay (1996): $r_g = \frac{\hat{\sigma}_{gxy}}{\sqrt{\hat{\sigma}_{gx}^2 \hat{\sigma}_{gy}^2}}$, em que $\hat{\sigma}_{gxy}$ é o estimador da covariância

genética entre as variáveis x e y ; e $\hat{\sigma}_{gx}^2$ e $\hat{\sigma}_{gy}^2$ são os estimadores da variância genética das variáveis x e y , respectivamente.

Índice de seleção – Mulamba e Mock (1978)

O índice de Mulamba e Mock (1978) é baseado na soma de ordens e equivale ao teste não paramétrico de Wilcoxon (1945), conforme relatado por Resende et al. (2014). A identificação dos clones superiores na população analisada foi realizada pelo índice de rank médio, adaptado de Mulamba e Mock (1978). Consiste em classificar os genótipos em ordem favorável ao melhoramento para cada variável analisada e, após a classificação, é calculada a média dos ranks (ordem do genótipo para cada variável analisada) para cada genótipo.

As estimativas de ganhos com a seleção dos melhores genótipos, de acordo com a classificação obtida no índice de seleção baseado no rank médio, também foram estimadas.

Diversidade genética

Para o estudo de diversidade genética considerou-se a distância euclidiana média como medida de dissimilaridade e a determinação dos grupos foi realizada pelo método de agrupamento de Tocher (Rao, 1952), o qual se baseia na formação de grupos cujas distâncias dentro dos grupos sejam menores que as distâncias entre grupos.

Eficiência da seleção considerando estrutura de subpopulação

Os grupos formados com base no método de agrupamento de Tocher foram considerados como subpopulações para a análise considerando estrutura de subpopulação. O modelo 122 do software Selegen REML/BLUP (Resende, 2002) foi utilizado para a realização da análise, o qual equivale a: $y = Xu + Zg + Ws + e$, em que, y é o vetor de dados; u é o escalar referente à média geral (efeito fixo); g é o vetor dos efeitos genotípicos (assumidos como aleatórios); s é o vetor dos efeitos de subpopulações (assumidos como aleatórios); e é o vetor de resíduos (aleatórios). As letras maiúsculas representam as matrizes de incidência para os referidos efeitos.

As estruturas de médias e variâncias associadas ao modelo são: $y|u, V \sim N(Xu, V)$; $g|\sigma_g^2 \sim N(0, I\sigma_g^2)$; $s|\sigma_s^2 \sim N(0, I\sigma_{sp}^2)$ e $e|\sigma_e^2 \sim N(0, I\sigma_e^2)$; $Cov(s, e) = 0$; $Cov(g, e) = 0$; $Cov(s, g) = 0$.

Assim,

$$E \begin{bmatrix} y \\ g \\ s \\ e \end{bmatrix} = \begin{bmatrix} Xu \\ 0 \\ 0 \\ 0 \end{bmatrix} \text{ e } Var. \begin{bmatrix} y \\ g \\ s \\ e \end{bmatrix} = \begin{bmatrix} V & ZG & WS & R \\ GZ' & A & 0 & 0 \\ SW' & 0 & S & 0 \\ R & 0 & 0 & R \end{bmatrix}$$

$G = I\sigma_g^2 =$ Matriz de variância – covariância dos efeitos genotípicos aleatórios;

$S = I\sigma_s^2 =$ Matriz de variância – covariância dos efeitos de subpopulações aleatórios;

$V =$ Variância fenotípica e equivale a: $Z\sigma_g^2Z' + W\sigma_s^2W' + I\sigma_e^2 = ZGZ' + WSW' + R$.

σ_g^2 , σ_s^2 e σ_e^2 são as variâncias genotípica, de subpopulações e residual, respectivamente.

A eficiência da seleção do modelo estatístico que considera estrutura de subpopulações sobre o modelo que considera apenas os valores genotípicos dos clones foi calculada por meio da expressão: $efic. = r_{total}/r_{clone}$, em que r_{total} é a acurácia da seleção com base no modelo considerando estrutura de subpopulações; e r_{clone} é a acurácia da seleção com base no modelo que considera apenas efeito genético (clone).

A acurácia da seleção que considera estrutura de subpopulações foi calculada por meio da fórmula: $r_{total} = \sqrt{3\sigma_{total}^2/(1 + 2\sigma_{total}^2)}$; em que σ_{total}^2 é a variância genotípica total e é calculada por meio da fórmula: $\sigma_{total}^2 = \sigma_{sub}^2 + \sigma_{cl/sub}^2$; em que σ_{sub}^2 é a variância genotípica entre as subpopulações; e $\sigma_{cl/sub}^2$ é a variância genotípica de clones dentro das subpopulações.

RESULTADOS E DISCUSSÃO

A análise de deviance (ANADEV) para as variáveis relacionadas à qualidade de fruto e resistência ao CVC da população híbrida de Citrus avaliada é apresentada na Tabela 1. Observa-se que as variáveis: espessura da casca, facilidade de descasque, sólidos solúveis e resistência ao CVC não apresentaram resultados significativos pelo teste da máxima verossimilhança (LRT) para a fonte de variação genética (clone), indicando que a população não apresenta variabilidade genética para estas variáveis. As demais variáveis apresentaram resultados significativos, mesmo com a ausência de controle local, confirmando a existência de variabilidade genética entre os híbridos e, conseqüentemente, a possibilidade de ganhos com a seleção.

Tabela 1. Análise de deviance (ANADEV) e teste da máxima verossimilhança (LRT) para as variáveis avaliadas na população híbrida de Citrus.

Caracteres	Deviance – Clones	Deviance – Modelo completo	LRT (X ²)
MF	4402,06	4224,18	177,88**
DL	175,15	35,14	140,01**
DT	3,81	-124,87	128,68**
L/T	-2085,44	-2110,90	25,46**
EC	526,36	526,29	0,07 ^{NS}
NG	1032,83	973,56	59,27**
NS	2608,37	2448,86	159,51**
FD	2083,31	2083,15	0,16 ^{NS}
RS	2676,95	2604,08	72,87**
Ac	386,01	380,51	5,5*
°B	888,78	761,67	127,11**
Rt	1671,76	1501,21	170,55**
SS	3871,96	3871,91	0,05 ^{NS}
F/C	5109,97	5030,99	78,98**
CVC	3454,02	3453,98	0,04 ^{NS}

** e * significativo a 1% e 5% de probabilidade, respectivamente, e ^{NS} não-significativo pelo teste qui-quadrado com 1 grau de liberdade. MF – massa do fruto; DL – diâmetro longitudinal; DT – diâmetro transversal; L/T – relação entre o diâmetro longitudinal e o diâmetro transversal; EC – Espessura da casca; NG – número de gomos; NS – número de sementes; FD – Facilidade de descasque; RS – rendimento de suco; Ac – Acidez; °B - °Brix corrigido; Rt – relação entre sólidos solúveis e acidez; SS – sólidos solúveis; F/C – frutos por caixa; CVC – resistência à clorose variegada do citros.

Na Tabela 2 são apresentados os resultados das estimativas dos parâmetros genéticos para as variáveis analisadas. Estas estimativas são de suma importância para avaliar o potencial da população analisada para o melhoramento genético.

Tabela 2. Estimativas de parâmetros genéticos e ambiental para as variáveis relacionadas à qualidade de fruto e resistência ao CVC em uma população híbrida de Citrus.

Parâmetros genéticos	MF	DL	DT	L/T	EC	NG	NS	FD	RS	Ac	°B	Rt	SS	F/C	CVC
σ_g^2	816,72	0,27	0,19	0,0017	0,01	0,89	27,35	0,33	22,12	0,08	0,98	4,87	5,44	2319,24	2,35
σ_e^2	567,71	0,24	0,18	0,0056	0,96	1,63	20,98	17,66	33,05	0,67	0,94	3,49	507,14	3378,03	232,00
σ_f^2	1384,42	0,51	0,37	0,0073	0,97	2,52	48,33	17,99	55,17	0,75	1,92	8,36	512,58	5697,27	234,35
h_g^2	0,59	0,53	0,51	0,24	0,0137	0,35	0,57	0,02	0,40	0,11	0,51	0,58	0,01	0,41	0,01
h_{gm}^2	0,81	0,77	0,76	0,49	0,04	0,62	0,80	0,06	0,67	0,27	0,76	0,81	0,03	0,68	0,03
\bar{X}	157,45	6,28	6,96	0,90	0,49	11,06	16,85	2,18	44,97	1,27	10,61	8,96	4,16	274,06	1,87
$CV_{gi}(\%)$	18,15	8,26	6,22	4,64	23,48	8,53	31,03	26,37	10,46	22,65	9,35	24,64	56,12	17,57	81,83
$CV_e(\%)$	15,13	7,76	6,10	8,30	199,03	11,54	27,18	192,73	12,78	64,41	9,15	20,86	541,83	21,21	812,90
$CV_{relativo}$	1,20	1,06	1,02	0,56	0,12	0,74	1,14	0,14	0,82	0,35	1,02	1,18	0,10	0,83	0,10
r_{gg}	0,89	0,87	0,87	0,69	0,20	0,78	0,89	0,23	0,81	0,52	0,87	0,89	0,17	0,81	0,17

σ_g^2 : Variância genética; σ_e^2 : Variância residual; σ_f^2 : Variância fenotípica; h_g^2 : Herdabilidade individual no sentido amplo (efeitos genotípicos totais); h_{gm}^2 : herdabilidade genotípica ao nível de média; \bar{X} : Média geral; $CV_{gi}(\%)$: Coeficiente de variação genotípico; $CV_e(\%)$: Coeficiente de variação residual; $CV_{relativo} = CV_{gi}(\%)/CV_e(\%)$; r_{gg} : acurácia seletiva. MF – massa do fruto; DL – diâmetro longitudinal; DT – diâmetro transversal; L/T – relação entre o diâmetro longitudinal e o diâmetro transversal; EC – Espessura da casca; NG – número de gomos; NS – número de sementes; FD – Facilidade de descasque; RS – rendimento de suco; Ac – Acidez; °B - °Brix corrigido; Rt – relação entre sólidos solúveis e acidez; SS – sólidos solúveis; F/C – frutos por caixa; CVC – resistência à clorose variegada do citros (CVC).

As estimativas das variâncias genóticas (σ_g^2) das variáveis massa do fruto, diâmetro longitudinal, número de sementes e °Brix corrigido foram superiores às suas variâncias residuais (σ_e^2), situação observada em experimentos que apresentam maior acurácia seletiva (Resende e Duarte, 2007) e maior possibilidade de identificação de genótipos superiores na população (Vencovsky, 1987).

As estimativas de herdabilidade no sentido amplo, ou seja, considerando os efeitos aditivos e de dominância, para as variáveis avaliadas assumiram valores entre 0,01 (CVC) e 0,59 (Massa do fruto) (Tabela 2). Este parâmetro expressa a proporção da variância fenotípica que é de natureza genética, ou seja, expressa a correlação entre o valor fenotípico e o valor genético auxiliando, portanto, na escolha dos métodos que deverão ser utilizados nos programas de melhoramento (Falconer, 1987). Resende (2002) classifica a herdabilidade no sentido restrito como de baixa magnitude quando $h_a^2 < 0,15$; de média magnitude quando h_a^2 assume valores entre 0,15 e 50; e de alta magnitude quando $h_a^2 > 0,50$.

As variáveis espessura da casca, facilidade de descasque, sólidos solúveis e CVC apresentaram valores de herdabilidade individual muito baixos (inferiores a 0,05), revelando que a população não apresenta potencial de seleção para estas variáveis, corroborando com os resultados obtidos na análise de deviance (Tabela 1). As demais variáveis apresentaram valores superiores a 0,1 indicando a possibilidade de seleção com ganhos genéticos significativos, uma vez que, o ganho esperado com a seleção depende diretamente da herdabilidade e da intensidade de seleção (Resende, 2007; Scarpinati et al. 2009; Santos et al., 2014). Além disso, altos valores de herdabilidade estão associados a maior variância genética e/ou menor influência ambiental (Fehr, 1987; Pompeu Junior et al., 2013).

A baixa magnitude da herdabilidade observada para algumas variáveis pode ser contornada por meio do controle das condições ambientais e da introdução de novos materiais genéticos nos próximos ciclos do programa de melhoramento. Este processo visa o aumento da variabilidade genética, condição essencial para a obtenção de ganhos com a seleção (Carias et al., 2016; Falconer e Mackay, 1996).

As estimativas de herdabilidade baseadas na média foram maiores do que as estimativas em nível individual para todas as variáveis, demonstrando que a seleção é mais efetiva com base em informações em nível de média (Costa et al., 2008; Costa et al., 2010). Além disso, segundo Vencovsky e Barriga (1992), estimativas realizadas a

partir da média apresentam maior acurácia devido à redução da influência do erro experimental.

O coeficiente de variação residual (CV_e) expressa a magnitude das variações causadas pelo ambiente e é um indicativo de precisão experimental. Este parâmetro variou de 6,10 (diâmetro longitudinal) a 812,90 (CVC). Pimentel Gomes (2009) considera CV_e baixos, quando inferiores a 10%; médios quando de 10 a 20%; altos, quando de 20 a 30%; e muito altos, quando superiores a 30%. De acordo com esta classificação, as variáveis número de sementes, facilidade de descasque, acidez, ratio, sólidos solúveis, frutos/ caixa e resistência ao CVC apresentaram CV_e altos ou muito altos. Amaral et al. (1997) propuseram faixas de CV_e para as principais variáveis estudadas na cultura do Citrus e, de acordo com as faixas apresentadas pelos autores, verificou-se que os valores observados de CV_e para massa do fruto ($> 11,9$), diâmetro longitudinal ($>4,3$), rendimento de suco ($>6,9$), acidez ($>16,5$), °Brix (>7), ratio ($>10,2$) são considerados muito altos.

O CV_{gi} (%) expressa, em porcentagem da média geral, a magnitude das variações causadas pelos efeitos genéticos, portanto, permite inferir sobre a magnitude da variabilidade genética presente nas populações (Morales et al., 2014; Resende, 2002). Para a população estudada, retirando a variável CVC, o CV_{gi} (%) variou de 4,64 (relação entre o diâmetro longitudinal e o diâmetro transversal) a 56,12 (SS), valores que revelam uma situação favorável para o processo de seleção para as variáveis massa de fruto, diâmetro longitudinal e transversal, relação diâmetro longitudinal e transversal, número de gomos, número de sementes, rendimento de suco, °Brix corrigido, ratio e frutos por caixa.

O coeficiente de variação relativo (CVr) refere-se à razão entre o coeficiente de variação genético e o coeficiente de variação residual e, para o caso em que se têm duas ou três repetições, quando assume valores próximos ou maiores do que a unidade indica que a população apresenta condições favoráveis para o melhoramento genético (Vencovsky, 1987), garantindo ganhos genéticos expressivos. Esta situação foi observada para as variáveis massa do fruto, diâmetro longitudinal, diâmetro transversal, número de sementes, °Brix corrigido e ratio. De acordo com Vencovsky e Barriga (1992), maiores valores de CVr sugerem que as variáveis apresentam maior controle genético e são menos influenciadas por fatores ambientais, o que favorece a seleção.

Com CV_g/CV_e menor que 1, maiores números de repetições são necessários para se ter alta acurácia na seleção (Resende e Duarte, 2007).

Os valores de acurácia (r_{gg}) variaram de 0,17 a 0,89 (Tabela 2) e, de acordo com Resende e Duarte (2007), é a estatística mais adequada para avaliação da qualidade experimental. A acurácia congrega o coeficiente de variação genético, o coeficiente de variação residual e o número de repetições e, deve ser preferido para a avaliação da qualidade experimental devido à facilidade de interpretação (Resende, 2002). A acurácia pode ser classificada em muito alta (> 90%), alta (entre 70% e 90%), moderada (entre 50% e 70%) e baixa (< 50%) (Resende e Duarte, 2007). As variáveis espessura da casca, facilidade de descasque, sólidos solúveis e CVC apresentaram baixos valores de acurácia, enquanto as demais variáveis apresentaram acurácias classificadas em moderadas a altas. Portanto, para as variáveis com valores de acurácia considerados altos, as predições dos valores genéticos avaliados apresentam maior confiabilidade. Vale ressaltar que estimativas de acurácia superiores a 0,7 são consideradas suficientes para inferir sobre o valor genético dos indivíduos em avaliação (Resende, 2002; Brasileiro et al., 2013).

As variáveis que não apresentaram efeitos genéticos significativos, de acordo com a análise de deviance (Tabela 1), não foram incluídas nas análises subsequentes (correlação genética, índice de Mulamba-Mock, agrupamento de Tocher e ANADEV com base em estrutura de populações), visto que, não é possível realizar seleção nas mesmas.

A correlação genotípica representa o grau de associação de natureza genética entre duas variáveis e, correlações genéticas altas indicam que, a alteração em um caráter, via seleção, promove alterações nos caracteres correlacionados (Resende, 2002). Portanto, o conhecimento desta medida é de suma importância para os trabalhos de melhoramento genético, principalmente, se a seleção em alguma variável apresenta dificuldades, em razão da baixa herdabilidade e/ou problemas de medição (Cruz et al., 2012), e/ou para a adoção da seleção precoce em culturas perenes, via seleção indireta (Verardi et al., 2011). Na Tabela 3 são apresentados os resultados de correlação genotípica entre as variáveis analisadas.

Tabela 3. Correlação genética entre variáveis relacionadas à qualidade de frutos em uma população híbrida de Citrus.

	MF	DL	DT	L/T	NG	NS	RS	Ac	°B	Rt	F/C
MF	1	0,84	0,89	0,17	0,08	-0,12	0,09	-0,04	-0,35	0,13	-0,89
DL		1	0,78	0,44	0,01	-0,14	-0,07	0,13	-0,40	0,00	-0,79
DT			1	0,01	0,09	-0,16	0,10	-0,04	-0,35	0,18	-0,86
L/T				1	-0,11	0,11	0,08	-0,29	-0,10	-0,07	-0,16
NG					1	0,11	-0,08	0,05	-0,17	0,06	-0,02
NS						1	0,12	-0,15	0,11	0,05	0,08
RS							1	-0,35	0,14	0,11	-0,20
Ac								1	-0,11	-0,60	0,05
°B									1	0,27	0,34
Rt										1	-0,12
F/C											1

MF – massa do fruto; DL – diâmetro longitudinal; DT – diâmetro transversal; L/T – relação entre o diâmetro longitudinal e o diâmetro transversal; NG – número de gomos; NS – número de sementes; RS – rendimento de suco; Ac – Acidez; °B - °Brix corrigido; Rt – relação entre sólidos solúveis e acidez; F/C – frutos por caixa.

Observa-se que as correlações genotípicas foram altas e positivas entre as variáveis: massa do fruto e diâmetro longitudinal (0,84), massa do fruto e diâmetro transversal (0,89), e diâmetro longitudinal e transversal (0,78). A principal causa de correlação genética é o pleiotropismo, sendo as ligações gênicas consideradas uma causa transitória, especialmente em populações derivadas de cruzamentos entre indivíduos divergentes (Falconer e Mackay, 1996). Assim, com correlação genética alta e positiva entre duas variáveis é possível obter ganho genético em uma por meio da seleção na outra (seleção indireta). Por outro lado, com correlação genética negativa a seleção em uma variável irá contribuir negativamente para a outra, o que será observado se a seleção for praticada entre as variáveis: massa do fruto e frutos por caixa (-0,89), diâmetro longitudinal e frutos por caixa (-0,79), e diâmetro transversal e frutos por caixa (-0,86). Constata-se então, que dificuldades com a seleção são observadas quando duas variáveis apresentam correlação positiva e uma delas é indesejável, ou quando as duas variáveis são desejáveis, mas a correlação é negativa. As correlações observadas entre as demais variáveis são consideradas baixas.

Na população de Citrus avaliada, as dificuldades com a seleção serão observadas quando esta for realizada nas variáveis massa do fruto, diâmetro longitudinal e transversal e frutos por caixa. Este fato é decorrente das correlações genotípicas observadas na Tabela 3 entre essas variáveis, visto que o objetivo do programa de

melhoramento para a população analisada é aumentar a massa do fruto, aumentar o diâmetro longitudinal, reduzir o diâmetro transversal e reduzir o número de frutos por caixa. Para contornar esta situação, os índices de seleção tem sido utilizados no melhoramento genético de plantas como uma forma de escolher materiais superiores com base na seleção simultânea para vários caracteres. Na Tabela 4 são apresentados os 10 melhores clones com base no índice de Mulamba e Mock (1978). Os valores da tabela representam a posição (rank) que o genótipo assumiu para cada variável analisada.

Tabela 4. Classificação dos 10 melhores clones de Citrus com base no índice de Mulamba-Mock (1978).

Ordem	Híbrido	MF	DL	DT	L/T	NG	NS	RS	Ac	°B	Rt	F/C	Rank-Medio	%GS
1	116	15	16	131	20	143	9	9	13	80	5	16	41,55	117,83
2	227	3	5	177	16	22	141	92	7	32	2	3	45,45	108,05
3	188	7	8	173	28	135	4	107	26	45	32	6	51,91	95,45
4	124	25	36	179	171	37	44	28	11	35	8	20	54,00	87,65
5	285	9	15	168	101	64	33	13	43	128	60	7	58,27	80,15
6	329	61	30	113	19	100	89	136	27	41	16	63	63,18	72,73
7	16	1	2	178	24	75	150	61	17	159	28	1	63,27	67,75
8	192	10	10	155	14	39	159	32	51	148	67	11	63,27	64,21
9	239	17	12	156	31	111	104	77	22	119	40	18	64,27	61,23
10	302	16	68	171	164	25	146	6	1	99	1	14	64,64	58,82

MF – massa do fruto; DL – diâmetro longitudinal; DT – diâmetro transversal; L/T – relação entre o diâmetro longitudinal e o diâmetro transversal; NG – número de gomos; NS – número de sementes; RS – rendimento de suco; Ac – Acidez; °B - °Brix corrigido; Rt – relação entre sólidos solúveis e acidez; F/C – frutos por caixa.

Observa-se que com uma intensidade de seleção de 5,5% (10 clones) espera-se um ganho de 58,82% com relação à população original. Assim, constata-se que é possível aumentar a concentração de alelos favoráveis na população com a seleção simultânea dos caracteres. A eficiência do uso do Índice de Mulamba-Mock tem sido confirmada em programas de melhoramento de diversas culturas (Carias et al., 2016; Farias Neto et al., 2012; Teixeira et al., 2012; Resende et al., 2014).

A análise de agrupamento multivariado dos genótipos de Citrus pelo método de Tocher, com base na distância genética euclidiana média, é apresentada na Tabela 5.

Tabela 5. Agrupamento dos 180 clones de Citrus pelo método de Tocher, com base na distância genética euclidiana média.

Grupo	Clones
1	2, 4, 5, 9, 10, 11, 12, 13, 14, 15, 19, 20, 23, 24, 29, 31, 34, 42, 44, 46, 47, 48, 50, 52, 53, 55, 60, 61, 66, 71, 72, 74, 75, 79, 80, 84, 86, 87, 88, 89, 96, 97, 98, 100, 101, 102, 103, 109, 110, 111, 112, 116, 118, 119, 124, 125, 129, 141, 142, 143, 147, 153, 155, 156, 157, 162, 163, 166, 167, 170, 171, 175, 180, 181, 183, 184, 185, 188, 191, 192, 196, 199, 202, 203, 204, 206, 207, 208, 210, 212, 213, 214, 216, 217, 219, 222, 225, 226, 227, 228, 230, 233, 234, 235, 239, 240, 241, 242, 243, 244, 247, 249, 250, 252, 256, 260, 261, 264, 265, 268, 272, 273, 275, 276, 280, 281, 283, 285, 286, 288, 289, 290, 291, 292, 294, 296, 305, 307, 308, 309, 313, 314, 315, 317, 319, 323, 324, 326, 327, 328, 329, 330, 331, 332, 345, 347, 351, 352, 362, 364, 366, 369, 371, 372, 377, 378, 379 e 380.
2	90, 105, 107, 115 e 117.
3	37, 160 e 194.
4	16 e 169.
5	182 e 302.

Verificou-se a formação de cinco grupos com base na divergência genética entre os clones. O grupo número 1 foi composto pela maioria dos clones (168); o grupo número 2 por cinco clones, o grupo 3 por três, e os grupos 4 e 5 por dois. Com base no resultado, pode-se observar que, mesmo havendo divergência genética na população para as variáveis consideradas na análise de agrupamento, a maioria dos clones é considerada semelhante. Esta informação é de relevante interesse no delineamento dos programas de melhoramento genético, principalmente para a seleção de genitores para a formação de novas combinações híbridas.

Dentre os 10 clones que apresentaram maiores ganhos genéticos com a seleção simultânea, com base no índice de seleção de Mulamba-Mock (Tabela 5), oito foram alocadas no grupo 1, um no grupo 4 e um no grupo 5. Com base nesses dados, é possível delinear cruzamentos entre os melhores genitores pertencentes a distintos grupos. Por exemplo, o genótipo 16, pertencente ao grupo 4, que apresenta a melhor performance para massa de fruto e frutos por caixa; e o genótipo 302, pertencente ao grupo 5, que apresenta a melhor performance para as variáveis acidez e ratio, podem ser

utilizados como genitores em cruzamentos com os genótipos do grupo 1 que apresentem bons desempenhos para as outras variáveis. Assim, o cruzamento entre os melhores genótipos pertencentes a distintos grupos aumentará a probabilidade de obtenção de genótipos com características superiores.

Com base no agrupamento de Tocher, uma nova análise considerando estrutura de subpopulação (cada grupo formado no método de agrupamento de Tocher foi considerado como uma subpopulação, ou seja, a análise foi realizada com 5 subpopulações) foi realizada e os resultados da análise de deviance (ANADEV) são apresentados na Tabela 6.

Tabela 6. Análise de deviance (ANADEV) para as variáveis avaliadas na população híbrida de Citrus considerando a presença de cinco subpopulações.

Variáveis	Deviance – Clones	Deviance – Subpopulação	Deviance – Modelo completo	LRT – Clones/ subp	LRT - Subpopulação
MF	4224.18	4278.99	4172.33	51.85**	106.66**
DL	35.14	81.01	-6.67	41.81**	87.68**
DT	-124.87	-77.07	-160.70	35.83**	83.63**
L/T	-2110.90	-2125.82	-2138.77	27.87**	12.95**
NG	973.56	1000.96	958.24	15.32**	42.72**
NS	2448.86	2600.80	2446.68	2.18 ^{NS}	154.12**
RS	2604.08	2602.54	2560.71	43.37**	41.83**
Ac	380.51	203.75	203.74	176.77**	0.01 ^{NS}
°B	761.67	859.77	750.20	11.47**	109.57**
Rt	1501.21	1607.95	1477.23	23.98**	130.72**
F/C	5030.99	4981.65	4958.02	72.97**	23.63**

** e * significativo a 1% e 5% de probabilidade, respectivamente, e ^{NS} não-significativo pelo teste qui-quadrado. MF – massa do fruto; DL – diâmetro longitudinal; DT – diâmetro transversal; L/T – relação entre o diâmetro longitudinal e o diâmetro transversal; NG – número de gomos; NS – número de sementes; RS – rendimento de suco; Ac – Acidez; °B - °Brix corrigido; Rt – relação entre sólidos solúveis e acidez; F/C – frutos por caixa.

Com base na análise de deviance (ANADEV) observa-se que a maioria das variáveis apresentou resultados significativos para o fator de variação genético (clones) e para o fator de variação subpopulação, o que indica que há variabilidade genética dentro das subpopulações e entre as subpopulações analisadas. No entanto, a variável número de sementes apresentou resultado não significativo para o fator de variação genético (clones), ou seja, não há variabilidade genética dentro da subpopulação, apenas entre subpopulações; e a variável acidez apresentou resultado não significativo para o fator de variação subpopulação, indicando que não há variância genética entre as mesmas.

Na Tabela 7 é apresentado o ganho em eficiência da seleção considerando estrutura de subpopulação com relação à seleção baseada apenas em informações de clones.

Tabela 7. Ganhos em eficiência da seleção com base na estrutura de subpopulação com relação à seleção baseada em informações de clone.

	h_{sub}^2	$h_{cl/sub}^2$	h_{total}^2	r_{total}	r_{clone}	<i>Efic.</i>
MF	0,82	0,08	0,91	0,98	0,89	0,09
DL	0,75	0,11	0,85	0,97	0,87	0,11
DT	0,75	0,10	0,85	0,97	0,87	0,12
L/T	0,53	0,08	0,61	0,91	0,69	0,31
NG	0,49	0,15	0,65	0,92	0,78	0,17
NS	0,12	0,49	0,61	0,91	0,89	0,02
RS	0,48	0,16	0,63	0,92	0,81	0,12
°B	0,27	0,35	0,62	0,91	0,87	0,05
Rt	0,71	0,15	0,86	0,97	0,89	0,08
F/C	0,83	0,04	0,87	0,98	0,81	0,19

h_{sub}^2 : herdabilidade genotípica entre subpopulações; $h_{cl/sub}^2$: herdabilidade genotípica de clones dentro de subpopulações; r_{total} : acurácia da seleção com base no modelo que considera estrutura de subpopulações; r_{clone} : acurácia da seleção com base em informações de clone; *Efic.*: Eficiência seletiva do modelo usando informações de subpopulação + clone dentro de subpopulação sobre o modelo utilizando apenas informações de clones; MF – massa do fruto; DL – diâmetro longitudinal; DT – diâmetro transversal; L/T – relação entre o diâmetro longitudinal e o diâmetro transversal; NG – número de gomos; NS – número de sementes; RS – rendimento de suco; Ac – Acidez; °B - °Brix corrigido; Rt – relação entre sólidos solúveis e acidez; F/C – frutos por caixa.

O modelo utilizando estrutura de subpopulação apresentou superioridade variando de 2% a 31% quando comparado com o modelo que considera apenas informações de clones para as variáveis analisadas. Segundo Resende et al. (2016), este fato ocorre devido ao maior número de repetições que é utilizado quando considera-se estrutura de subpopulações (5 subpopulações x 3 repetições), ao contrário do uso apenas das informações dos clones (3 repetições).

CONCLUSÕES

- A população de Citrus avaliada apresenta potencial para o melhoramento genético para a maioria das variáveis avaliadas;
- A introdução de novos genótipos é necessária para aumentar a diversidade genética e o potencial de seleção da população;
- A utilização de estrutura de subpopulações nas análises aumenta a eficiência do processo de seleção.

REFERÊNCIAS BIBLIOGRÁFICAS

ALVES, R. M.; RESENDE, M. D. V. Avaliação genética de indivíduos e progênes de cupuaçuzeiro no Estado do Pará e estimativas de parâmetros genéticos. *Revista Brasileira de Fruticultura*, v. 30, n. 3, 696-701, 2008.

AMARAL, A.M.; MUNIZ, J.A.; SOUZA, M. Avaliação do coeficiente de variação como medida da precisão na experimentação com citros. **Pesq. agropec. bra.**, v.32, n. 12, p.1221-1225, 1997.

ARANTES, F.C.; GONÇALVES, P.S.; SCALOPPI JUNIOR, E.J.; MORAES, M.L.T.; RESENDE, M.D.V. Ganho genético com base no tamanho efetivo populacional de progênes de seringueira. **Pesq. agropec. bras.** 45(12): 1419-1424, 2010.

BRASILEIRO, B.P.; SILVA, S.A.; SOUZA, D.R.; SANTOS, P.A. OLIVEIRA, R.S.; LYRA, D.H. Genetic diversity and selection gain in the physic nut (*Jatropha curcas*). **Genetics and Molecular Research** 12(3): 2341-2350, 2013.

CARIAS, C.M.O.M.; TOMAZ, M.A.; FERRÃO, M.A.G.; FONSECA, A.F.A.; FERRÃO, R.G.; GONÇALVES, L.S.A. Produtividade de grãos de cafeeiro conilon de diferentes grupos de maturação pelo procedimento REML/BLUP. **Semina: Ciências Agrárias** 35(2): 707-718, 2014.

CARIAS, C.M.O.M.; GRAVINA, G.A.; FERRÃO, M.A.G.; FONSECA, A.F.A.; FERRÃO, R.G.; VIVAS, M.; VIANA, A.P. Predição de ganhos genéticos via modelos mistos em progênes de café conilon. **Coffee Science** 11(1): 39 - 45, 2016.

CORREA, E.R.; MEDEIROS, G.C.R.; BARROS, W.S.; BIANCHI, V.J. Evaluation and ranking of Macadamia genotypes using mixed models. **Afr. J. Agric. Res.** 10(38): 3696-3703, 2015.

COSTA, R.B.; RESENDE, M.D.V.; GONÇALVES, P.S.; CHICHORRO, J.F.; ROA, R.A.R. Variabilidade genética e seleção para caracteres de crescimento da seringueira. **Bragantia** 67(2): 299-305, 2008.

COSTA, R.B.; RESENDE, M.D.V.; GONÇALVES, P.S.; ROA, R.A.R.; FEITOSA, K.C.O. Predição de parâmetros e valores genéticos para caracteres de crescimento e produção de látex em progênes de seringueira. **Bragantia** 69(1): 49-56, 2010.

CRUZ, C.D.; REGAZZI, A.J.; CARNEIRO, P.C.S. Modelos biométricos aplicados ao melhoramento genéticos. Viçosa, Editora UFV. 2012. 515p.

FALCONER, D.S.; MACKAY, T.F.C. **Introduction to Quantitative Genetics**. 4.ed. Edinburgh : Longman Group Limited, 1996. 464p.

FALCONER, D. S. **Introdução à genética quantitativa**. Viçosa: UFV, 1987. 279 p.

FAO - **Food and Agriculture Organization of the United**. Disponível em: <http://faostat.fao.org/>; Acessado em: 10 de junho de 2016.

FARIAS NETO, J.T.; LINS, P.M.P.; RESENDE, M.D.V.; MULLER, A.A. Seleção genética em progênies híbridas de coqueiro. **Rev. Bras. Frutic.** 31(1): 190-196, 2009.

FARIAS NETO, J.T.; OLIVEIRA, M.S.P.; RESENDE, M.D.V.; RODRIGUES, J.C. Parâmetros genéticos e ganhos com a seleção de progênies de Euterpe oleracea na fase juvenil. **Cerne** 18(3): 515-521, 2012.

FEHR, W. R. Principles of cultivars development. New York: Macmillan, 1987. 536 p.

HENDERSON, C.R. Maximum likelihood estimation of variance components. **In press.**, 1973.

MAIA, M.C.C.; RESENDE, M.D.V.; OLIVEIRA, L.C.; VASCONCELOS, L.F.L.; LIMA NETO, J.F.P. Análise genética em genótipos de manga rosa via REML/BLUP. **Revista Agrotecnologia** 5(1): 1-16, 2014.

MORAES, C.B.; FREITAS, T.C.M.; PIERONI, G.B.; RESENDE, M.D.V.; ZIMBACK, L.; MORI, E.S. Estimativas de parâmetros genéticos para seleção precoce de clones de Eucalyptus para região com ocorrência de geadas. **Sci. For.** 42(102): 219-227, 2014.

MULAMBA, N.N.; MOCK, J.J. Improvement of yield potential of the Eto Blanco maize (*Zea mays* L.) population by breeding for plant traits. **Egyptian Journal of Genetics and Cytology** 7: 40-51, 1978.

OLIVEIRA, E.J.; SANTANA, F.A.; OLIVEIRA, L.A.; SANTOS, V.S. Genetic parameters and prediction of genotypic values for root quality traits in cassava using REML/BLUP. **Genetics and Molecular Research** 13(3): 6683-6700, 2014.

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

PIMENTEL-GOMES, F. **Curso de Estatística Experimental**. Piracicaba: FEALQ, 2009, 451p.

POMPEU JUNIOR, J.; BLUMER, S.; RESENDE, M. D. V. Avaliação genética de seleções e híbridos de limões cravo, volkameriano e rugoso como porta-enxertos para laranjeiras Valência na presença da morte súbita dos citros. **Rev. Bras. Frutic.**, v.35, n.1, 2013.

RAO, R.C. **Advanced statistical methods in biometric research**. New York: John Wiley and Sons, 1952. 390p.

RESENDE, M.A.V.; FREITAS, J.A.; LANZA, M.A.; RESENDE, M.D.V.; AZEVEDO, C.F. Divergência genética e índice de seleção via BLUP em acessos de algodoeiro para características tecnológicas da fibra. **Pesquisa agropecuária tropical** 44(3): 334-340, 2014.

RESENDE, M. D. V. **Genética biométrica e estatística no melhoramento de plantas perenes**. Brasília: Embrapa Informação Tecnológica, 2002. 975p.

RESENDE, M.D.V. **Matemática e Estatística na Análise de Experimentos e no Melhoramento Genético**. 1. ed. Colombo: Embrapa, 2007. 561p.

RESENDE, M.D.V.; DUARTE, J.B. Precisão e controle de qualidade em experimentos de avaliação de cultivares. **Pesq. Agropec.Trop.** 37: 182-194, 2007.

RESENDE, M. D. V.; ROSA-PEREZ, J. R. H. **Genética e Melhoramento de Ovinos**. Curitiba: Ed. UFPR, 2001. 185p.

RESENDE, M. D. V.; SILVA, F. F.; AZEVEDO, C. F. **Estatística matemática, biométrica e computacional: modelos mistos, multivariados, categóricos e generalizados (REML/BLUP), inferência bayesiana, regressão, aleatória, seleção genômica, QTL-GWAS, estatística espacial e temporal, competição, sobrevivência**. Editora: Suprema, Visconde do Rio Branco. 2014. 881p.

RESENDE, M.D.V.; RAMALHO, M.A.P.; CARNEIRO, P.C.S.; CARNEIRO, J.E.S.; BATISTA, L.G.; GOIS, I.B. Selection Index with Parents, Populations, Progenies, and Generations Effects in Autogamous Plant Breeding. **Crop Science**, 2016.

RESENDE, M.D.V.; FURLANI-JUNIOR, E.; MORAES, M.L.T.; FAZUOLI, L.C. Estimativas de parâmetros genéticos e predição de valores genotípicos no melhoramento do cafeeiro pelo procedimento REML/BLUP. **Bragantia** 60(3): 185-193, 2001.

SANTOS, A.M.; ROSADO, S.C.S.; OLIVEIRA, A.N. Estimation of genetic parameters and verification of early selection efficiency in baru (*Dipteryx alata*). **Crop Breeding and Applied Biotechnology** 14: 238-243, 2014.

SCARPINATI, E.A.; PERECIN, D.; PAULA, R.C.; BONINI, C.A.V.; PAVAN, B.E.; CANDIDO, L.S. Influência do modelo de análise estatística e da forma das parcelas experimentais na seleção de clones de *Eucalyptus* spp. **R. Árvore**, v.33, n.4, p.769-776, 2009.

TEIXEIRA, D.H.L.; OLIVEIRA, M.S.P.; GONCALVES, F.M.A.; NUNES, J.A.R. Índices de seleção no aprimoramento simultâneo dos componentes da produção de frutos em açaizeiro. **Pesquisa agropecuária brasileira** 47(2): 237-243, 2012.

VENCOVSKY, R. Herança quantitativa. In: PATERNIANI, E.; VIEGAS, G. P. (Ed.). **Melhoramento e produção de milho**. Campinas: Editora Fundação Cargil, 1987.

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

VERARDI, C.K.; SCALOPPI JUNIOR, E.J.; GONÇALVES, P.S. Prediction of direct and indirect genetic gains and genotypic correlations in rubber tree progenies. **Pesq. agropec. bras.** 46(9): 1076-1081, 2011.

WILCOXON, F. Individual Comparisons by Ranking Methods. **Biometrics Bulletin** 1(6): 80-83, 1945.

CAPÍTULO 2

Genome wide selection increases efficiency of Citrus breeding

Viçosa – MG

July, 2016

RESUMO

GOIS, Itamara Bomfim, D.Sc., Universidade Federal de Viçosa, Julho, 2016. **Seleção genômica ampla aumenta a eficiência no melhoramento de Citrus.** Orientador: Aluizio Borém de Oliveira. Coorientadores: Marcos Deon Vilela de Resende e Mariângela Cristofani-Yaly.

Seleção genômica ampla (GWS) é considerada uma ferramenta essencial no melhoramento genético de espécies perenes, como Citrus, devido ao seu potencial em aumentar o ganho por unidade de tempo e possibilitar a seleção, de forma mais eficiente, de características que apresentam baixa herdabilidade. Este trabalho foi realizado com o objetivo de verificar a eficiência da GWS em uma população de Citrus e compará-la com a seleção baseada apenas em dados fenotípicos. Foram avaliados 180 indivíduos de uma população proveniente do cruzamento entre laranja Pera (*Citrus sinensis* Osbeck) e tangor Murcott (*Citrus sinensis* Osbeck X *Citrus reticulata* Blanco) para 10 características relacionadas à qualidade de frutos. Os clones foram genotipados com 5.287 marcadores moleculares DArT_seqTM e os efeitos destes nos fenótipos preditos a partir do método RR-BLUP. A capacidade preditiva, o viés de predição e a acurácia da GWS foram estimados para verificar sua eficácia em prever os fenótipos. A proporção da variância genética explicada pelos marcadores também foi computada. A herdabilidade das características, determinada com base nos marcadores, variou de 16% a 28%. A capacidade preditiva variou de 0,53 a 0,64, e o coeficiente de regressão entre os fenótipos preditos e os observados apresentou valores próximos à unidade, o que indica ausência de viés. A proporção da variância genética explicada pelos marcadores foi superior a 35%. As estimativas da acurácia obtidas a partir da GWS foram inferiores às aquelas obtidas a partir das análises fenotípicas, no entanto, foi possível observar a superioridade da GWS, em termos de ganho genético por unidade de tempo. Portanto, GWS é uma ferramenta promissora no melhoramento genético do Citrus, uma vez que, proporciona acuradas predições dos fenótipos em estádios iniciais do desenvolvimento da planta, permitindo, assim, a redução da duração do ciclo de seleção.

Palavras-chave: marcadores moleculares; DArT_seqTM; desequilíbrio de ligação; seleção precoce; acurácia seletiva.

ABSTRACT

GOIS, Itamara Bomfim, D.Sc., Universidade Federal de Viçosa, July, 2016. **Genome Wide Selection increases efficiency in Citrus breeding.** Adviser: Aluizio Borém de Oliveira. Co-advisers: Marcos Deon Vilela de Resende e Mariângela Cristofani-Yaly.

Genome wide selection (GWS) is an essential tool for genetic improvement of perennial species such as Citrus because of its ability to potentially increase gain per unit time and enable efficient selection of characteristics with low heritability. This work assessed GWS efficiency in a population of Citrus and compared it with selection based on phenotypic data only. A total of 180 individual trees from a population derived from a cross between Pera sweet orange (*Citrus sinensis* Osbeck) and Murcott tangor (*Citrus sinensis* Osbeck x *Citrus reticulata* Blanco) were evaluated for 10 characteristics related to fruit quality. The clones were genotyped using 5,287 DArT_seqTM molecular markers and their effects on phenotypes were predicted using the RR-BLUP method. The predictive ability, prediction bias, and accuracy of GWS were estimated to verify its effectiveness in predicting phenotypes. The proportion of genetic variance explained by the markers was also computed. The heritability of the traits, as determined by markers, ranged from 16% to 28%. The predictive ability of these markers ranged from 0.53 to 0.64, and the regression coefficient values between predicted and observed phenotypes were close to unity (which means no bias). The proportion of genetic variance accounted by the markers was greater than 35%. Estimates of accuracy with GWS were lower than those obtained using phenotypic analysis, however, GWS was superior in terms of genetic gain per unit time. Thus, GWS can be useful to Citrus breeding as it can predict well and early the phenotypes and reduces the length of selection cycle. This paper seems to be the first attempt to demonstrate the feasibility of genomic selection in Citrus.

Keywords: molecular markers; DArT_seqTM; linkage disequilibrium; early selection; selective accuracy.

INTRODUCTION

The genus *Citrus* is a major food and economic crop worldwide and many strategies have been adopted to increase its productivity and quality. Breeding is the preferred strategy to address these issues. When referring to *Citrus* breeding it is important to note that the *Citrus* plant itself consists of two parts, i.e., the canopy and the rootstock, often from two different species, thus requiring targeted studies to improve their individual quality and the interaction between them. In general phenotypic selection is applied and it is costly and time consuming, especially when the desired traits are expressed at later stages.

Marker-assisted selection (MAS) was proposed by Lande and Thompson (1990) to obtain faster gains and increase selection efficiency compared to selection based on phenotypic data only. MAS uses both phenotypic data and molecular markers in genetic linkage with certain loci controlling desired quantitative traits (QTL). QTL markers are selected after appropriate statistical modeling and are subjected to type II error, i.e., the probability of accepting a false hypothesis (Resende et al., 2014).

In *Citrus* breeding, MAS strategy has been adopted for selecting several traits of economic importance such as resistance to biotic and abiotic stress (Ito et al., 2014) and physicochemical characteristics of fruit. However, the identification of a QTL associated with such characteristics can explain only a small fraction of the observed genetic variation in each character (Weber et al., 2003; Siviero et al. 2002; Siviero et al. 2006; Gussen et al. 2011; Asins et al. 2012), thus making the selection process based on these markers unfeasible. The rationale behind the nonuse of MAS is that traits under selection are governed by many loci with small effects that do not reach statistical significance when used in traditional QTL analysis, and this in turn leads to the identification of only a small number of QTL with major effects (Kemper and Goddard 2012; Resende et al., 2014).

First theorized by Meuwissen et al. (2001), genome wide selection (GWS) is a form of genetic selection assisted by molecular markers that simultaneously predict the genetic effects of hundreds or thousands of DNA markers on phenotypes (Meuwissen, 2007; Resende et al., 2008). These markers, which are in linkage disequilibrium with the QTL, can have both large and small effects and can account for nearly all of the genetic variation in a quantitative trait (Resende et al., 2008). Additionally, the whole genome prediction tends to be more accurate because it better accommodates the

variation created by Mendelian segregation during gamete formation (Resende et al., 2008; Zapata-Velenzuela et al., 2013).

The increasing availability and decreasing costs of molecular markers spanning the whole genome has led to the implementation of large-scale genomic selection in breeding programs for different species (Daetwyler et al., 2010; Grattapaglia and Resende, 2011; Misztal et al., 2009; Resende et al., 2012; Resende Junior et al., 2012; Zhong et al., 2009). The properties that make GWS promising for use in breeding programs include efficient selection of traits with low heritability, permitting the best use of available genetic resources by selecting appropriate genetic crosses, and increasing genetic gain per unit time as it enables early identification of the best individuals for selection (Daetwyler et al., 2013; Legarra et al., 2008; Resende et al., 2014).

In Citrus species, the juvenile period is considerably long with a range of 1-20 years; although flowering and fruiting may occur within 3-7 years depending on the species. Therefore, breeding programs suffer a delay in the hybridization process and in the selection of desired characteristics, apart from requiring large areas for the development of hybrids which not only increases the cost of plant maintenance in the field but also limits the number of families and the number of individuals per family that can be evaluated (Talon and Gmitter Junior, 2008; Gmitter Junior et al., 2012). GWS can minimize these undesirable consequences as selection can be performed in juvenile plants, because this reduces generation interval, increases intensity of selection and, therefore, the gain per unit time and cost (Resende et al., 2008; Wong and Bernardo, 2008; Heffner et al., 2009; Grattapaglia and Resende 2011; Kemper and Goddard, 2012).

Applications of the GWS in fruit trees has been scarce, being restricted to apples (Kumar et al., 2012), pear (Iwata et al., 2013) and grapes (Viana et al., 2016). The present paper seems to be the first attempt to study the feasibility of genomic selection in Citrus. This research was carried out to apply GWS in a biparental hybrid population of Citrus using DArT_seqTM markers and phenotypes of 10 traits, assess its selection efficiency and compare it to selection based on phenotypic data.

MATERIAL AND METHODS

Characterization of the population and phenotypic traits

The genetic effects of the markers were predicted in a hybrid population of Citrus which is used as canopy. The population composed by 180 individuals was generated by crossing Pera sweet orange (*Citrus sinensis* Osbeck) and Murcott tangor (*Citrus sinensis* Osbeck x *Citrus reticulata* Blanco).

The experiment was established in 2004 at Cordeirópolis, SP, in a completely randomized design with three clonal replicates. Phenotypic evaluation was performed in 2012 for the following characteristics: fruit mass (g), longitudinal fruit diameter (cm), transverse fruit diameter (cm), ratio of longitudinal fruit diameter (cm)/transverse fruit diameter (cm), number of segments per fruit, number of seeds per fruit, juice yield (mL), soluble solids (°Brix), ratio of soluble solids / acidity and number of fruits per box.

A sample of each replication, consisting of five fruits each, was used to obtain phenotypic data. Fruit mass was measured using a digital balance with a capacity of 15 kg and accuracy of 5 g; the longitudinal and transverse fruit diameters were measured using a graduated ruler; the number of segments per fruit and seed number were directly counted; the number of fruits per box was obtained by dividing capacity of the box (40.8 kg) by average weight of the fruit; juice yield was measured after crush extraction on an OIC model OTTO 1800 and the juice/fruit ratio was calculated; the soluble solids content (°Brix) was determined by direct reading on a refractometer (B & S model RFM 330); and acidity was determined by titrating 25 mL of juice against normal sodium hydroxide solution with phenolphthalein as indicator. The soluble solids/ acidity ratio, which indicates degree of fruit ripening, was also calculated. All analyses were performed in the Quality and Postharvest Laboratory of the Center APTA Citrus Sylvio Moreira/ IAC, Cordeirópolis-SP.

Statistical model for the analysis of phenotypic data

The analyses of phenotypic data for the estimation of genetic variance and heritability were carried out using the REML / BLUP method, also called the mixed model methodology (Henderson, 1973), using the Selegen REML/BLUP software (Resende, 2002). The following model was fit: $y = Xu + Zg + e$; where y is the vector of observations; u is the scalar related to overall mean (fixed effect); g is the vector of

genotypic effects (assumed to be random); and e is the residual vector (random). The capital letters (X , Z) represent incidence matrices for these effects. The structures of means and variances associated with the model are: $y|u, V \sim N(Xu, V)$; $\sigma_g^2 \sim N(0, I\sigma_g^2)$; $e|\sigma_e^2 \sim N(0, I\sigma_e^2)$; and $Cov(g, e) = 0$; where, V is the phenotypic covariance matrix; I is an identity matrix; σ_g^2 and σ_e^2 are the genotypic and residual variances, respectively. The mixed model equations for the BLUP method (Best Linear Unbiased Predictor) to obtain individual breeding values are:

$$\begin{bmatrix} \hat{u} \\ \hat{g} \end{bmatrix} = \begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + I\lambda \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ Z'y \end{bmatrix}, \text{ where:}$$

$\lambda = \sigma_e^2 / \sigma_g^2 = \frac{1-h^2}{h^2}$ is a penalization parameter; and

$h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2}$ is the individual broad sense heritability.

Variance components were estimated by REML (Restricted Maximum Likelihood; Patterson and Thompson, 1971) iterating in the mixed model equations. Prediction accuracy of genotypic values for clone selection was calculated from the inverse of the left hand side of the mixed model equations (Resende, 2002).

Genotyping and quality control of the markers

Molecular assessment of the 180 individuals in the hybrid Citrus population was performed using the DArT Method (Diversity Arrays Technology; Jaccoud et al., 2001), which reduces genome complexity by using a combination of restriction enzymes and next generation sequencing (NGS) technique called DArT_seqTM. Total DNA was extracted, quantified on a Nanodrop-8000 spectrophotometer, normalized to 100 ng/ μ L, distributed into plates, and submitted to genotyping on the Diversity Arrays Technology platform (DArT P/ L, Australia).

The method for obtaining the DArT_seqTM markers involves reducing genome complexity using the restriction enzymes PstI and TaqI, and followed by sequencing. PstI and adapters specific for 96 different bar codes were linked to the restriction fragments. The resulting products were amplified and their quality was checked. Samples were sequenced on an Illumina Hiseq2000 machine. The PstI adapters included a sequencing primer such that the sequences were always read from the PstI

restriction sites. The resulting sequences (FASTQ files) were filtered for quality, with a confidence cut-off level of 90%. Sequences were aligned with the Clementine tangerine reference genome available in <https://phytozome.jgi.doe.gov>. This procedure yielded 27,960 markers DArT_seqTM which were coded as "0", if absent, or "1", if present. All markers were analyzed in F1 progeny hybrid seeds for expected Mendelian segregation patterns.

The quality control of the markers was performed by eliminating low polymorphic loci (MAF - minor allele frequency $\leq 5\%$) and/or loci with low call rate in the genotyped individuals (call rate $\leq 95\%$). MAF was calculated using the equation: $MAF = \min[p, (1-p)]$, where p is the allele frequency in a locus; and Call Rate was calculated using the expression: $Call\ Rate = 1 - M/P$, where M refers to individuals with missing genotypes and P refers to individuals with present genotypes (not missing).

Predicting genetic effects of markers using RR-BLUP

The random regression (RR-BLUP) method (Meuwissen et al., 2001) was used for predicting the genetic effects of the markers. This method fits the random effects of markers as covariate with effects on phenotypes (Resende et al., 2012).

The estimators associated with this random regression procedure promote shrinkage on the markers effects as a function of the penalty parameter λ , which makes it possible to estimate a higher number of parameters than the number of data points. The penalty parameter is given by: $\lambda = \frac{\sigma_e^2}{\sigma_{gi}^2} = \frac{\sigma_e^2}{\sigma_g^2/n_Q}$, where σ_{gi}^2 is the genetic variance associated with the marker locus i ; σ_g^2 is the trait genetic variance; σ_e^2 is the residual variance; and n_Q is associated to the number of marker loci and, for DArT dominant markers, is given by $n_Q = \sum p_i(1 - p_i)$ (Resende et al., 2012).

The following general linear mixed model was designed to predict the effects of the markers: $y = Xu + Wm_a + Sm_d + e$; where y is the vector of phenotypic observations; u is the vector of fixed effects (overall means); m_a is the vector of additive genetic markers effects; m_d is the vector of dominance genetic markers effects; e is the vector of random errors; X , W and S are the incidence matrices for u , m_a and m_d , respectively.

Parameterizations of additive and dominance marker effects with dominant markers such DArT were performed according to Viana and Resende (2014). In an allogamous population that is in Hardy-Weinberg equilibrium (HWE), the appropriate

parameterization in the W matrix, associated with additive effects, is ‘0’ for the absence of the marker (mm type) and 1.78 for the presence of the marker (MM or Mm types); while in the S matrix along with the effects of dominance, parameterization is: ‘0’ for the absence of the marker (mm) and 0.89 for the presence of marker (MM or Mm). Assuming HWE, the additive genetic variation of the character in the population is given by: $\sigma_a^2 = \sum_{i=1}^n p_i q_i \alpha_i^2$; where p_i is the frequency G_{MM+Mm} , α_i is the additive effect at locus i ; and dominance genetic variation is given by: $\sigma_d^2 = \sum_{i=1}^n (p_i^2 q_i^2 d_i^2)$; where d_i is the dominant effect for locus i .

The structure of means and variances is defined as: $m_a \sim N(0, WW'\sigma_{ma}^2)$; $m_d \sim N(0, SS'\sigma_{md}^2)$; $e \sim N(0, I\sigma_e^2)$, in that σ_{ma}^2 is the marker additive genetic variance given by $\sigma_{ma}^2 = \frac{\sigma_a^2}{\sum_{i=1}^n (p_i q_i)}$; σ_{md}^2 is the marker variance due to dominance deviations and defined as $\sigma_{md}^2 = \frac{\sigma_d^2}{\sum_{i=1}^n (p_i q_i)^2}$; σ_e^2 are residual variances.

The mixed model equations to predict m_a and m_d using the RR-BLUP method are:

$$\begin{bmatrix} X'X & X'W & X'S \\ W'X & W'W + I \frac{\sigma_e^2}{\sigma_{ma}^2} & W'S \\ S'X & S'W & S'S + I \frac{\sigma_e^2}{\sigma_{md}^2} \end{bmatrix} \begin{bmatrix} \hat{u} \\ \hat{m}_a \\ \hat{m}_d \end{bmatrix} = \begin{bmatrix} X'Y \\ W'Y \\ S'Y \end{bmatrix}$$

The genomic values (GV) of individuals were estimated using the equation: $GV = Wm_a + Sm_d$. All analyses were performed on R software, version 3.1.2 (R Development Core Team 2012) using the rr-BLUP package (Endelman, 2011).

Predictive ability, prediction bias and accuracy of GWS

The predictive ability (r_{yg}) was obtained by correlating the corrected phenotypic values with predicted genomic values and was determined by cross-validation through a Jackknife procedure. Prediction bias (b) was calculated as the regression coefficient of phenotypic values on the predicted genomic values, wherein ‘b’ values close to 1 indicate that predictions are not biased and are effective in predicting the actual magnitude of difference between the individuals being assessed (Resende et al., 2012).

The experimental accuracy of GWS was obtained by the estimator: $r_{gg} = r_{yg}/h$, which r_{gg} corresponds to the accuracy of GWS; r_{yg} the predictive ability; and h the square root of the heritability of the character being studied (Legarra et al., 2008).

The expected accuracy was calculated using the formula proposed by Resende et al. (2008): $r_{ggE} = \sqrt{\frac{r_{mq}^2(Nh_m^2)}{1+[(N-1)h_m^2]}}$, wherein $r_{mq}^2 = \frac{n_m}{n_m+M_e}$ is the proportion of variance explained by the markers; N is the number of genotyped and phenotyped individuals; n_m is the number of markers; h_m^2 is the individual heritability of a locus given by $h_m^2 = \frac{r_{mq}^2 h^2}{M_e}$, and h^2 is the heritability of the character. For calculating M_e (effective number of chromosome segments) the approach proposed by Goddard et al. (2011) was used: $M_e = 2NeL$, where Ne is the effective population number (2, in the present paper) and L is the length of the Citrus genome in Morgans.

The estimated number of QTL controlling each character was calculated by the expression: $n_{QTL} = \frac{Nr_{mq}^2 h^2 (r_{mq}^2 - r_{gg}^2)}{r_{gg}^2}$. These expressions were obtained using accuracy formulae proposed by Resende et al. (2008), wherein the term ' r_{mq}^2 ' was estimated from the number of markers that maximized the accuracy of GWS for each character.

Efficiency of GWS

The superiority of GWS, compared to selection based only on phenotypic data on a perennial species, is mainly related to the gain per unit time (Resende et al., 2012). To verify this fact, the gain in selective efficiency of GWS, compared to the selection based only on phenotypes, was calculated using the expression: $G = \frac{r_{gg} t_f}{r_{yy} t_{GWS}} - 1$, where r_{gg} is the accuracy of GWS; r_{yy} is the accuracy of selection based on phenotypes means; t_f is the average time required for the completion of the selection cycle based on phenotypes means; and t_{GWS} is the average time required to complete the selection cycle based on the GWS (Resende et al., 2012).

RESULTS AND DISCUSSION

After quality control 5,287 markers were found to be useful and had their effects estimated. The estimated genome size of the sweet orange is 367 MB (Xu et al., 2013) and, according to Jarrell et al. (1992), it is 1,700 cM. Therefore, the density of markers used in the analyses was ~3 markers per cM, which corresponds to a distance of 0.32 cM between markers. This value was used to calculate the proportion of the variance that could be explained by the markers (r_{mq}^2), which, in turn, is related to the expected magnitude of linkage disequilibrium (r^2) (Resende et al., 2012). Figure 1 shows the expected proportion of variance accounted for by the markers (r_{mq}^2) depending on the number of markers used. As can be seen, r_{mq}^2 increases as the number of markers increases as this value depends on recombination frequency, which is in turn, a function of the distance between the markers (Grattapaglia and Resende, 2011). Assuming that LD (linkage disequilibrium) between pairs of markers is approximately equal to LD between markers and the QTL (Zhong et al., 2009), the number of markers used in this study yield high LD.

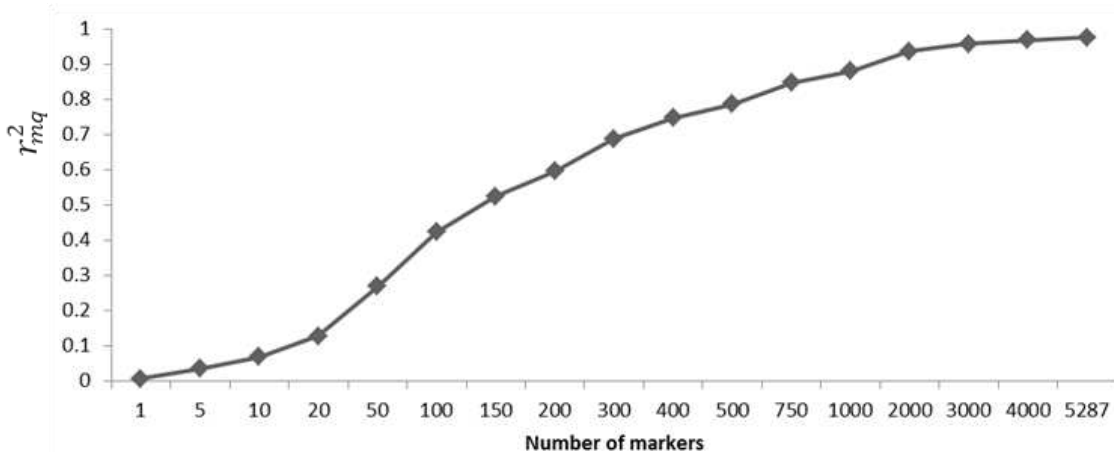


Figure 1. Expected proportion of genetic variance explained by markers (r_{mq}^2) with increasing number of markers.

Estimates of genomic heritability (h_G^2), predictive ability (r_{yg}), prediction bias (b), accuracy of GWS (r_{gg}), QTL number estimates (n_{QTL}) and proportion of genetic variance explained by the markers (r_{mq}^2) when selecting them based on the magnitudes of their effects are presented in Table 1.

Table 1. Estimates of genomic heritability (h_G^2), predictive ability (r_{yg}), prediction bias (b), accuracy of GWS (r_{gg}), QTL number estimates (n_{QTL}) and genetic variance explained by the markers (r_{mq}^2) accounted by the markers when selecting markers based in the magnitudes of their effects in a population of Citrus breeding.

Traits	h^2	h_G^2	w	n_m	B	r_{yg}	r_{yy}	r_{gge}	r_{ggo}	r_{mq}^2	n_{QTL}
DL	0.77	0.24	0.31	1000	0.95	0.64	0.88	0.79	0.73	0.54	99
DT	0.76	0.25	0.33	750	0.90	0.63	0.87	0.77	0.72	0.53	97
DL/ DT	0.49	0.26	0.53	1000	0.90	0.63	0.70	0.72	0.91	0.84	11
FM	0.81	0.28	0.35	300	0.88	0.64	0.90	0.72	0.72	0.53	68
F/B	0.68	0.22	0.32	400	0.98	0.64	0.82	0.72	0.77	0.61	47
NG	0.62	0.27	0.44	200	0.81	0.59	0.79	0.74	0.73	0.54	34
NS	0.80	0.22	0.28	2000	0.95	0.62	0.89	0.81	0.69	0.49	131
JY	0.67	0.21	0.31	1000	0.94	0.56	0.82	0.77	0.69	0.49	110
SS	0.76	0.27	0.36	100	0.74	0.53	0.87	0.63	0.61	0.38	49
Rt	0.81	0.16	0.20	750	1.02	0.53	0.90	0.78	0.59	0.36	219

h_2 : Heritability based on phenotypes; h_G^2 : Heritability based markers; $w = h_G^2 / h_2$: Proportion of heritability based on phenotypes captured by markers; n_m : Number of markers that maximized the selective accuracy of GWS; b: bias of prediction; r_{yg} : Predictive ability; r_{yy} : Accuracy of phenotypic selection based on REML / BLUP method; r_{gge} : expected accuracy of GWS; r_{ggo} : Observed accuracy of GWS; r_{mq}^2 : genetic variance explained for the markers; n_{QTL} : Estimated number of QTL that controls the trait in the population; DL: Longitudinal fruit diameter (cm); DT: Transverse fruit diameter (cm); DL/DT: Ratio of longitudinal fruit diameter/ transverse fruit diameter; FM: Fruit mass (g); F/B: number of fruits per box; NG: Number of segments per fruit; NS: Number of seeds per fruit; JY: juice yield (mL); SS: soluble solids content; Rt: ratio of soluble solids / acidity.

The fraction of the trait heritability attributable to the markers (ω) varied from 20% (ratio of soluble solids/ acidity) to 53% (ratio of longitudinal/ transverse fruit diameter). The number of markers necessary to provide these fractions ranged from 100 for soluble solids content to 2,000 for number of seeds (Table 1). As the numbers of markers increased, it was observed that the accuracy values described by these markers tend to remain constant after reaching a maximum value.

The observed missing heritability, defined as the fraction of the heritability not accounted for by markers, is probably due to incomplete LD among the alleles causing genetic variation and the markers used for heritability estimation. A loss in LD may occur when the alleles causing genetic variation have lower MAF values than the marker (Kemper and Goddard, 2012).

The predictive ability of GWS, which reflects the ability of molecular information to consistently predict a phenotype (Cavalcanti et al., 2012), ranged from 0.53, for ratio of soluble solids/ acidity and soluble solids traits, to 0.64, for the traits longitudinal diameter of the fruit, fruit mass and number of fruits per box (Table 1). These magnitudes are similar to the ones reported in forest trees species (Resende et al., 2012, Resende Junior et al., 2012) and dairy cattle breeding (Hayes et al., 2009).

The regression coefficient (b), a measure of bias of prediction, showed values close to unity, implying no bias of prediction. However, the trait soluble solids had a 'b' value of 0.74, indicating overestimation of genetic variance (Resende et al., 2012).

The estimates of GWS accuracy values ranged from 0.59 to 0.91 implying moderate to high accuracy, according to the classification proposed by Resende and Duarte (2007). On the other hand, in this work, estimates of accuracy based on phenotypic selection ranged from 0.70 to 0.90, implying a high accuracy. A comparison these values shows that accuracy estimates based on phenotypic selection were superior to those obtained by GWS for all characteristics, except for ratio of longitudinal/transverse fruit diameter trait.

The experimental accuracy values were closer to the expected accuracy values (Table 1). The formula used for obtaining expected accuracy values is important in the design of GWS studies as it can be used to calculate the number of individuals and markers required to obtain a desired accuracy. Using the formula for expected accuracy, it was observed that for a trait with lowest heritability (ratio of longitudinal/transverse diameter; 0.49), at least 250 individuals and 4,000 markers in LD would be necessary to obtain an accuracy greater than 0.70. Contrarily, for traits with high heritability, such as fruit mass and ratio of soluble solids/ acidity the use of 180 individuals and 1000 markers yielded a much higher accuracy of 0.81. According to Grattapaglia and Resende (2011), accuracy calculated using the deterministic approach is directly proportional to the product of heritability and the ratio between the number of individuals used and the number of QTLs governing the trait. In a simulation study, those authors observed that the decrease in the accuracy with decreasing heritability

could be compensated for by using a larger number of individuals in the training population.

The proportion of variance explained by the markers ranged from 36% (ratio of soluble solids / acidity) to 84% (ratio of longitudinal fruit diameter / transverse fruit diameter). The fraction of the genetic variance not explained by these markers is, therefore, probably due to incomplete LD between these markers and the alleles causing the genetic variation in the character. GWS could explain a greater proportion of the observed variance in traits compared to methods based on the identification of individual QTL. This was observed by Viana et al. (2016) who compared MAS and genomic selection in a grapes breeding population constituted of 143 individuals obtained from a cross between (*Vitis rupestris* × *Vitis arizonica/girdiana*) × *Vitis vinifera*. The efficiency of genomic selection was between 1.7 and 11.6 times higher than the approach by traditional QTL analysis.

In the Citrus crop, the use of MAS also allowed the identification of only a few QTL with major effects on traits of interest. Siviero et al. (2002) found only one QTL associated with the quantitative inheritance of traits (number of fruits per plant and number of seeds per fruit) in a F1 progeny derived from a cross between *Citrus sunki* and *Poncirus trifoliata* 'Rubidoux'. A study on the root rot resistance caused by *Phytophthora* in progeny derived from a cross between *Citrus sunki* and *Poncirus trifoliata* by Siviero et al. (2006) provided two QTL maps for the species *P. trifoliata* which could explain 16-24% of the genetic variation, and only one QTL mapped for the species *C. sunki* which could account for only 14% of the variation. Asins et al. (2012) performed a QTL analysis in a segregating population of *C. grandis* and *C. clementine* to evaluate resistance to CTV (*Citrus tristeza virus*) and found a QTL with a big effect that contributed to about 24% of the total genetic variance.

Considering the five markers with big effects for each trait, only 21.5 to 31.36% of the variance was explained by these five markers. These values account for only a part of the genetic variation and further confirm the quantitative nature of these characteristics which are governed by many genes of small effects and, probably, by some genes with moderately large effects (Kemper and Goddard, 2012).

The estimated number of QTL controlling a trait varied from 11 (ratio of longitudinal fruit diameter / transverse fruit diameter) to 219 (ratio of soluble solids/ acidity) (Table 1). There are no other reports in literature on the number of loci that control these traits, as the studies conducted so far with MAS allowed the detection of

only one or more loci with major effects. These results only serve to confirm the quantitative nature of the traits studied and reiterate the fact that the accuracy of GWS is inversely proportional to the number of QTL identified, as observed by Wong and Bernardo (2008). In a study conducted by Resende et al. (2012) in two populations of Eucalyptus, 200 markers with large effects were found to be sufficient to account for more than 80% of the heritability of the traits being analyzed.

An increase in the number of markers did not result in a linear increase in the accuracy of GWS (Figure 2).

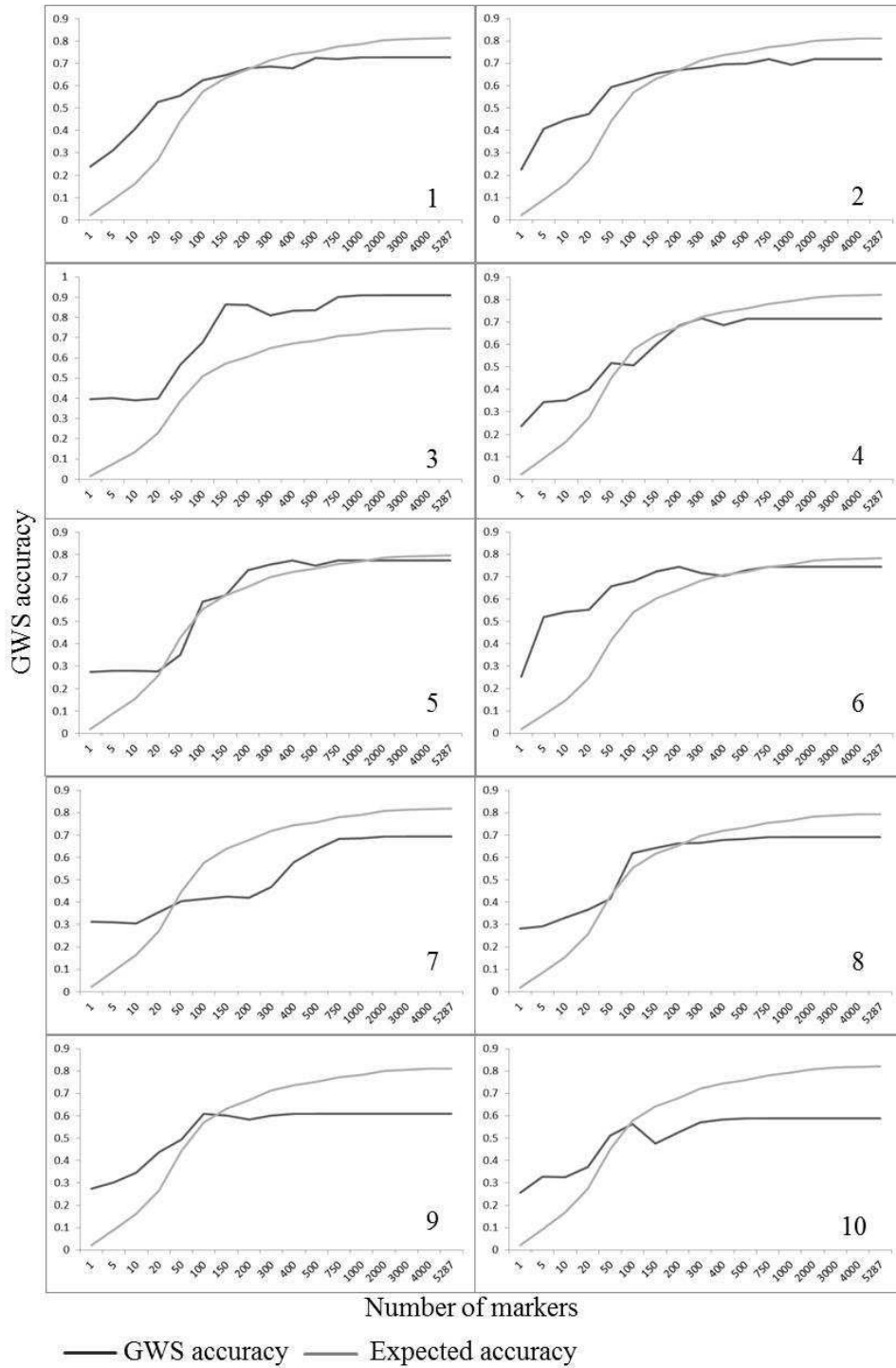


Figure 2. The accuracy of Genome Wide Selection (GWS) and expected accuracy due to the increase in the number of markers for traits evaluated in a population of Citrus. 1. Longitudinal fruit diameter; 2. Transverse fruit diameter; 3. Ratio of longitudinal fruit diameter/ transverse fruit diameter; 4. Fruit mass; 5. Number of fruits per box; 6. Number of segments per fruit; 7. Number of seeds; 8. Juice yield; 9. Soluble solids content; 10. Ratio of soluble solids/ acidity.

The accuracy of genome-wide selection depends on the proportion of the variance that can be explained by the markers and the accuracy of predicting the effects of the markers that are in linkage disequilibrium with the QTL. Thus, the main factors that modify the accuracy of GWS are the heritability of the trait, the number of loci that control this trait and the distribution of its effects; the number of individuals in the training population; and the extent of LD between the markers and the QTL. The LD depends on the effective population size and the number of markers used for the analyses (Grattapaglia and Resende, 2011).

The importance of the number of individuals used to estimate the effect of the loci can be seen from the formula of accuracy. Therefore, the higher the number of individuals used, the more accurate are the estimates of genomic values of the individuals (Hayes et al., 2009). Thus, the use of only 180 individuals to predict the effects of the markers probably contributed to the moderate accuracy. It is important to state here that the number of individuals used in this study is acceptable for the selection of superior clones in full-sib families since this number represents 99.45% of the genetic variability of the family. This fraction is given by $N_{ef}/2$, where $N_{ef} = 2n/(n+1)$ is the effective population size of a full-sib family and n is the number of individuals per family.

According to Resende et al. (2012), when more individuals per family and more families are analyzed, the power of detection increases asymptotically and more QTL are revealed. Zhao et al. (2013) analyzed accuracy values estimated based on the predicted effects of genetic markers in the entire maize population consisting of 11 families or based within each family. They reported that the estimates of accuracy were substantially reduced when individual GV (Genomic Value) were predicted in a within family basis in comparison with the predictions in the entire population. However, when the estimation of genetic effects and the prediction of GV were performed within each family (for large families), the accuracy was satisfactory, as seen by the results presented here. Cavalcanti et al. (2012) confirmed that within families GWS is an interesting alternative for increasing the breeding efficiency of cashew. In the study, for a characteristic with high heritability, the authors obtained a high accuracy value (0.86) with only 74 individuals and a limited number of markers (238).

One of the assumptions of GWS is that the phases of LD between markers and QTL are identical in the populations used for training and selection (Hayes et al., 2009). As genotyping and phenotyping of the individuals belonging to only one family was

performed, it probably did not cover the extent and pattern of the LD present in all Citrus populations used in the breeding program, and therefore, using these results in other families or populations is not recommended.

The small effective size of the Citrus population used in this study ($N_e = 2$) and the large number of markers used have favorably contributed to the accuracy values reported here. The impact of the effective size and number of markers used for predicting GWS accuracy has been demonstrated by Grattapaglia and Resende (2011) using simulated data. Resende et al. (2012) also applied GWS to two Eucalyptus breeding populations and found that higher accuracy estimates were observed in the population that had the lowest N_e . This is because an increase in effective size reduces the accuracy of genomic predictions due to an increase in the number of chromosome segments and because LD between the marker and the causative mutation remain consistent only if the two are present very close to each other on the chromosome (Kemper and Goddard, 2012).

GWS requires a high density of markers in order to maximize the number of QTL that are in LD with at least one marker (Heffner et al, 2009; Resende et al, 2014). Resende et al. (2008) demonstrated the importance of using a high density of markers to account for both large and small effects markers, and thereby to increase the probability of explaining all of the genetic variation in the trait of interest. The results of this study corroborate these claims since the proportion of variance explained by the markers was higher than those of the studies that used a smaller number of markers.

Table 2 compares the efficiency of GWS with selection based solely on phenotypic data.

Table 2. Gain in efficiency of GWS regarding selection based solely on phenotypic data in a breeding population of Citrus.

Traits	Reduction in the selection cycle		
	25%	50%	75%
DL	0.11	0.66	2.32
DT	0.10	0.66	2.31
DL/ DT	0.73	1.60	4.20
FM	0.07	0.60	2.20
F/B	0.25	0.88	2.76
NG	0.23	0.85	2.70
NS	0.03	0.55	2.10
JY	0.12	0.68	2.37
SS	-0.07	0.40	1.80
Rt	-0.13	0.31	1.62

DL: Longitudinal fruit diameter (cm); DT: Transverse fruit diameter (cm); DL/DT: Ratio of longitudinal fruit diameter / transverse fruit diameter; FM: Fruit mass (g); F/B: number of fruits per box; NG: Number of segments per fruit; NS: Number of seeds per fruit; JY: juice yield (mL); SS: soluble solids content; Rt: ratio of soluble solids / acidity.

When reducing the duration of the selection cycle by 25%, genetic gain with GWS was higher than those of solely phenotypic selection for all traits, except for soluble solids and ratio of soluble solids / acidity (Table 2). When time was reduced by 50%, GWS was superior (range from 31% to 160%) for all traits. When time was reduced by 75% GWS was better (range from 162%-420%). The higher values of gain in efficiency obtained for the ratio of longitudinal fruit diameter/ transverse fruit diameter with GWS further confirms the importance of GWS in selecting characteristics that exhibit low heritability.

Heffner et al. (2009) stated that even when only moderate accuracy is achieved using GWS, it is possible to obtain a genetic gain that is superior to that obtained using phenotypic selection as GWS reduces the duration of the selection cycle. This was clearly observed in this study when duration of the selection cycle was reduced. Wong and Bernard (2008) reported that GWS implementation reduced selection cycle of oil palm from 19 to 6 years, and this result, according to the authors, can be extrapolated to other perennial species as well. The referred study used a population of 50 individuals, which is the typical size for an oil palm breeding population. GWS yielded higher gains

per unit time and reduced the costs even with this small number of individuals. Importantly, GWS has been shown to be a promising tool for genetic improvement when applied to traits of economic interest that are difficult to measure and/ or have a high cost of evaluation (tolerance to cold, drought, disease resistance) (Grattapaglia and Resende, 2011). These kinds of traits should be approached by genomic selection in the Citrus breeding as well.

CONCLUSION

Results presented here suggest that GWS can be useful to Citrus breeding as it can predict well and early the phenotypes and reduces the length of the selection cycle. Thus, GWS is a promising tool for improving Citrus culture. This paper seems to be the first to demonstrate the feasibility of genomic selection in Citrus.

REFERENCES

- Asins MJ, Fernandez-Ribacoba J, Bernet GP, Gadea J, et al. (2012). The position of the major QTL for Citrus tristeza virus resistance is conserved among *Citrus grandis*, *C. aurantium* and *Poncirus trifoliata*. *Mol. Breeding* 29:575–587.
- Cavalcanti JJV, Resende MDV, Santos FHC, Pinheiro CR (2012). Predição simultânea dos efeitos de marcadores moleculares e seleção genômica ampla em cajueiro. *Rev. Bras. Frutic.*34, 3: 840-846.
- Daetwyler HD, Pong-Wong R, Villanueva B, Woolliams JA (2010). The impact of genetic architecture on genome-wide evaluation methods. *Genetics* 185: 1021-1031.
- Daetwyler AD, Calus MPL, Pong-Wong R, De Los Campos G, et al. (2013). Genomic Prediction in Animals and Plants: Simulation of Data, Validation, Reporting, and Benchmarking. *Genetics* 193: 347–365.
- Endelman JB (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome* 4:250-255.
- Gmitter Junior FG, Chen C, Machado MA, Souza AA, et al. (2012). Citrus genomics. *Tree Genetics & Genomes* 8: 611–626.
- Goddard ME, Hayes BJ, Meuwissen THE (2011). Using the genomic relationship matrix to predict the accuracy of genomic selection. *J. anim. Breed. Genet* 128(6): 409-421.
- Grattapaglia D, Resende MDV (2011). Genomic selection in forest tree breeding. *Tree Genetics & Genomes* 7: 241-255.
- Gussen O, Uzun A, Seday U, Kafa G (2011). QTL analysis and regression model for estimating fruit setting in young Citrus trees based on molecular markers. *Scientia Horticulturae* 130(2): 418-424.
- Hayes BJ, Chamberlain AJ, Goddard ME (2009) Genomic selection in dairy cattle: Progress and Challenges. *J.Dairy Sci.*92(2): 433-443.
- Heffner EL, Sorrells ME, Jannink JL (2009). Genomic Selection for Crop Improvement. *Crop Science* 49:1-12.
- Henderson CR (1973) Maximum likelihood estimation of variance components. In press.
- Ito TM, Polido PB, Rampim MC, Kaschuk, G, et al. (2014) Genome-wide identification and phylogenetic analysis of the AP2/ERF gene superfamily in sweet orange (*Citrus sinensis*). *Genet. Mol. Res.*13: 7839-7851.
- Iwata H, Hayashi T, Terakami S, Takada N, et al. (2013). Potential assessment of genome-wide association study and genomic selection in Japanese pear *Pyrus pyrifolia*. *Breeding Science* 63: 125–140.

- Jaccoud D, Peng K, Feinstein D, Kilian A (2001). Diversity arrays: a solid state technology for sequence information independent genotyping. *Nucleic Acids Res.* 29(4):1-7.
- Jarrell DC, Roose ML, Traugh SN, Kupper RS (1992). A genetic map of citrus based on the segregation of isozymes and RFLPs in an intergeneric cross. *Theor. Appl. Genet.* 84:49–56.
- Kemper KE, Goddard ME (2012). Understanding and predicting complex traits: knowledge from cattle. *Hum. Mol. Genet.* 21:45-51.
- Kumar S, Bink MCAM, Volz RK, Bus VGM, et al. (2012). Towards genomic selection in apple (*Malus × domestica* Borkh.) breeding programmes: Prospects, challenges and strategies. *Tree Genetics & Genomes* 8:1–14.
- Lande R, Thompson R (1990). Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124: 743-756.
- Legarra A, Granié RC, Manfredi E, Elsen JM (2008) Performance of genomic selection in mice. *Genetics* 180: 611-618.
- Meuwissen THE (2007) Genomic selection: marker assisted selection on a genome wide scale. *J. Anim. Breed. Genet.* 124: 321–322.
- Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics* 157:1819–1829.
- Misztal I, Legarra A, AGUILAR I (2009). Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *J. Dairy Sci.* 92: 4648–4655.
- Patterson HD, Thompson R (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika* 58: 545-554.
- Resende MDV (2002). *Genética Biométrica e estatística no melhoramento de plantas perenes*. Brasília: Embrapa Informação tecnológica.
- Resende MDV, Duarte JB (2007). Precisão e controle de qualidade em experimentos de avaliação de cultivares. *Pesq. Agropec.Trop.* 37: 182-194.
- Resende MDV, Lopes OS, Silva RL, Pires IE (2008). Seleção genômica ampla (GWS) e maximização da eficiência do melhoramento genético. *Pesq. Flor. Bra.* 56: 63-77.
- Resende MDV, Resende Junior MFR, Sansaloni CP, Petrolí CD, et al. (2012) Genomic selection for growth and wood quality in *Eucalyptus*: capturing the missing heritability and accelerating breeding for complex traits in forest trees. *New Phytologist* 194: 116-128.

- Resende MDV, Silva FF, Resende Junior MFR, Azevedo CF (2014). Genome-wide selection. In: *Biotechnology and Plant Breeding* (Borem A, Fritsche-Neto R, eds.). Elsevier.
- Resende Junior MFR, Muñoz P, Resende MDV, Garrick DJ, et al. (2012). Accuracy of Genomic Selection Methods in a Standard Data Set of Loblolly Pine (*Pinus taeda* L.). *Genetics* 190: 1503–1510.
- Siviero A, Cristofani M, Boava LP, Machado MA (2002) Mapeamento de QTLs associados à produção de frutos e sementes em híbridos de *Citrus sunki* vs. *Poncirus trifoliata*. *Rev. Bras. Frutic.* 24: 741-743.
- Siviero A, Cristofani M, Furtado EL, Garcia AAF, et al. (2006) Identification of QTLs associated with citrus resistance to *Phitophthora gummosis*. *J. Appl. Genet.* 47(1): 23-28.
- Talon M, Gmitter Junior FG (2008). Citrus genomics. *Intern. J. Plant Genomics*: 1-17.
- Viana AP, Resende MDV (2014). Seleção Genômica Ampla (GWS). In: *Genética Quantitativa no Melhoramento de Fruteiras* (Viana AP, Resende MDV, eds.). Interciência, Rio de Janeiro.
- Viana AP, Resende MDV, Riaz S, Walker MA (2016) Genome selection in fruit breeding: application to table grapes. *Sci. Agric.* 73(2): 142-149.
- Weber CA, Moore GA, Deng Z, Gmitter Junior FG (2003) Mapping freeze tolerance quantitative trait loci in a *Citrus grandis* x *Poncirus trifoliata* F1 pseudo-testcross using molecular markers. *J. Amer. Soc. Hort. Sci.* 128(4): 508-514.
- Wong CK, Bernardo R (2008). Genome wide selection in oil palm: increasing selection gain per unit time and cost with small populations. *Theor. Appl. Genet.* 116:815–824.
- Xu Q, Chen LL, Ruan X, Chen D, et al. (2013). The draft genome of sweet orange (*Citrus sinensis*). *Nat. Genet.* 45(1): 59-68.
- Zapata-Velenzuela J, Whetten RW, Neale D, Mckeand S, et al. (2013). Genomic Estimated Breeding Values Using Genomic Relationship Matrices in a Cloned Population of Loblolly Pine. *Genes, genome, Genetics* 3: 909-916.
- Zhao Y, Gowda M, Liu W, Wurschum T, et al. (2013). Choice of shrinkage parameter and prediction of genomic breeding values in elite maize breeding populations. *Plant Breeding* 132:99–106.
- Zhong S, Dekkers JCM, Fernando RL, Jannink JL (2009) Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: A barley cases study. *Genetics* 182: 355–364.