



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

**SELEÇÃO GENÔMICA AMPLA (GWS) APLICADA  
AO MELHORAMENTO POPULACIONAL**

Odilon Peixoto de Moraes Júnior  
Orientador: Prof. Dr. João Batista Duarte  
Co-orientadores: Dr. Flávio Breseghello  
Dr. Alexandre Siqueira Guedes Coelho

Revisão bibliográfica apresentada à Coordenação do Programa de Pós-Graduação em Genética e Melhoramento de Plantas, da Universidade Federal de Goiás, como parte integrante da disciplina Seminários III.

Goiânia  
2013

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> .....	<b>3</b>
<b>2</b>	<b>MARCADORES MOLECULARES NO MELHORAMENTO DE PLANTAS</b> ....	<b>6</b>
<b>3</b>	<b>SELEÇÃO ASSISTIDA POR MARCADORES (MAS)</b> .....	<b>8</b>
<b>4</b>	<b>SELEÇÃO GENÔMICA AMPLA (GWS)</b> .....	<b>11</b>
4.1	ASPECTOS GERAIS.....	14
4.2	EFICIÊNCIA SELETIVA E ACELERAÇÃO DOS GANHOS DE SELEÇÃO.....	14
4.3	TIPO E DENSIDADE DE MARCADORES.....	17
4.4	CARACTERÍSTICAS DA POPULAÇÃO DE ESTIMAÇÃO.....	19
4.5	MÉTODOS ESTATÍSTICOS DE PREDIÇÃO GENÔMICA.....	20
<b>5</b>	<b>SELEÇÃO RECORRENTE NA ERA DA GENÔMICA</b> .....	<b>25</b>
5.1	SELEÇÃO RECORRENTE ASSISTIDA POR MARCADORES (MARS).....	25
5.2	SELEÇÃO RECORRENTE GENÔMICA (GERS).....	26
<b>6</b>	<b>CONSIDERAÇÕES FINAIS</b> .....	<b>30</b>
<b>7</b>	<b>REFERÊNCIAS</b> .....	<b>32</b>

## 1 INTRODUÇÃO

Na agricultura moderna novos desafios têm surgido e os programas de melhoramento genético têm buscado, ao longo dos anos, estratégias eficientes para seleção de genótipos que apresentem desempenho superior. A dificuldade de melhorar populações para os caracteres quantitativos como produção de grãos é reconhecida e tem exigido o uso de métodos de melhoramento que permitam a manutenção da variabilidade genética, para contínuos ganhos de seleção e aumento da probabilidade de seleção de genótipos com maior número de alelos favoráveis (Bernardo, 2002; Hallauer et al., 2010).

Os melhoristas de plantas têm à sua disposição diferentes métodos de melhoramento, desde os mais simples e rápidos como a seleção massal, até aqueles mais elaborados e dinâmicos baseados no melhoramento populacional. Esses métodos é que garantem o sucesso de esquemas como a seleção recorrente, em que os genótipos com desempenho superior são selecionados e recombinados dentro de sua respectiva população. Portanto, o melhoramento populacional torna-se útil aos programas de melhoramento, por possibilitar a manutenção da variabilidade genética nas populações e o deslocamento da média do caráter no sentido desejado, devido ao contínuo aumento da frequência de alelos favoráveis na população (Ramalho et al., 2012).

Ganhos genéticos significativos têm sido obtidos no melhoramento populacional de várias espécies por meio de seleção recorrente (Ramalho et al., 2012; Morais Júnior, 2013). Porém, o tempo necessário para condução das populações elevam os custos de avaliação, limitando o processo de seleção e impactado nas estratégias adotadas, principalmente ao se considerar o elevado número de genótipos, a baixa pressão de seleção e o reduzido número de repetições, típicos dos programas de melhoramento populacional. Isso se deve, sobretudo à baixa eficiência da seleção fenotípica para caracteres quantitativos (Briggws & Shebeski, 1970; Cutrim et al., 1997), até mesmo quando se lança mão de seleção baseada na melhor predição linear não viesada (BLUP) (Henderson, 1973) dos efeitos aleatórios de genótipos. A baixa acurácia do método BLUP, em alguns casos, pode ser detectada quando se utilizam informações fenotípicas para fazer inferências sobre os efeitos dos genótipos dos indivíduos, e não informações genotípicas (genótipos para os alelos

marcadores) visando às inferências sobre os valores fenotípicos futuros (ou valores genéticos genômicos preditos) dos indivíduos (Resende et al., 2008; Crossa et al., 2010).

Os caracteres quantitativos de interesse agrônômico são, em sua maioria, resultantes da ação conjunta de vários genes, o que implica em forte influência ambiental na expressão fenotípica. A denominação QTL (*Quantitative Trait Loci*) é utilizada para nomear as regiões cromossômicas que contêm genes (ou locos) que controlam esses caracteres quantitativos (Falconer & Mackay, 1996). Assim, como ferramenta-suporte aos programas de melhoramento genético, as técnicas baseadas na utilização de marcadores moleculares surgiram para agregar informações à etapa de seleção de genótipos superiores, permitindo estudar com maior facilidade as regiões que influenciam a expressão desses caracteres e respectivos locos (QTL).

Lande & Thompson (1990) estão entre os primeiros a utilizarem marcadores moleculares como ferramenta para a seleção indireta no melhoramento genético. Nessa abordagem são usados dados de marcadores moleculares juntamente aos dados fenotípicos para a identificação de marcadores ligados a locos para genes específicos ou controladores de características quantitativas (QTL), caracterizando a chamada seleção assistida por marcadores. Na implantação desta estratégia de seleção deve-se considerar que cada gene/QTL que controla o caráter de interesse esteja identificado ou mapeado, e que existam marcadores disponíveis para monitorar a presença dos alelos favoráveis. Deste modo, o seu sucesso depende da precisão na detecção do QTL e da proporção da variância genética explicada por eles (Ferreira & Grattapaglia, 1998; Borém & Caixeta, 2009). Os marcadores moleculares surgiram, então, como ferramenta adicional importante para se melhorar a eficiência dos processos de seleção.

Nos últimos anos, por meio das tecnologias de sequenciamento de DNA de nova geração, houve redução drástica no custo da genotipagem, o que tem permitindo a produção de grande número de dados genômicos. Essa nova realidade surgiu como mudança de paradigmas, modificando, sobretudo, a escala das análises. Ademais, surgiu como abordagem que viabiliza a análise de genomas completos, sem necessidade de identificação ou mapeamento de QTL, mas também sem restringir o número de QTL considerado na expressão do caráter quantitativo sob seleção genômica ampla (GWS) (Rex Bernardo & Yu, 2007).

Nesse sentido, GWS ou seleção genômica (GS), proposta por Meuwissen et al. (2001), é uma análise baseada na avaliação de grande número de marcadores amplamente

distribuídos ao longo do genoma. Obtidos estes marcadores, são estimados os seus efeitos com base em dados fenotípicos de uma população de estimação. Estimados os efeitos, são obtidos os modelos de predição dos valores genéticos genômicos (VGG). As acurácias dos modelos são obtidas nas populações de validação e, em seguida, eles são aplicados em populações de seleção (Crossa et al., 2011).

No melhoramento genético, uma vez preditos os valores genéticos genômicos, estes poderão ser utilizados para a avaliação dos valores genéticos de indivíduos em fases precoces de seleção, com a predição dos fenótipos futuros de indivíduos genotipados em etapas preliminares do programa de melhoramento. Tal estratégia acelera a seleção, tendo em vista que torna possível, em determinados ciclos de seleção, eleger genótipos de interesse sem necessidade de fenotipagem (Resende et al., 2008).

Os maiores ganhos genéticos obtidos com a abordagem da seleção genômica ampla, em relação aos métodos tradicionais de seleção, são devido à redução do intervalo entre ciclos e maior acurácia de seleção. Esses ganhos têm justificado a sua utilização em programas de melhoramento animal (Schaeffer, 2006; Goddard & Hayes 2007; Meuwissen et al., 2013) e vegetal (Resende Júnior, 2010; Crossa et al., 2011; Fritsche Neto, 2011; Hayes et al., 2013). No caso do melhoramento vegetal, podemos observar como exemplo os resultados obtidos por Cavalcanti et al. (2012), em cajueiro, que obtiveram 86% de acurácia na seleção sobre o caráter peso de amêndoas. Heslot et al. (2012), em soja e trigo, também obtiveram elevadas acurácias (83% e 99%, respectivamente), entre outros resultados bastante promissores.

O objetivo deste trabalho é apresentar, na forma de revisão bibliográfica, as características, princípios, vantagens e limitações do método de seleção genômica ampla aplicado ao melhoramento genético populacional. Deste modo, busca contribuir como referencial teórico no assunto e, concomitantemente, como fonte de consulta para pesquisadores, professores, técnicos e estudantes de graduação ou pós-graduação.

## 2 MARCADORES MOLECULARES NO MELHORAMENTO DE PLANTAS

Qualquer forma alélica originária de um genoma pode ser empregada como marcador genético, podendo ser um dado fenotípico, uma proteína ou um fragmento de DNA que codifique ou não um gene, e que possua uma sequência repetida ou única no genoma. Muitos genes e sequências de DNA não codificadoras estão representados em uma espécie por duas ou mais formas alélicas diferentes, correspondentes ao mesmo loco cromossômico, o que caracteriza o fenômeno denominado polimorfismo genético. Por conveniência, um loco é considerado polimórfico quando a frequência de seu alelo mais comum é igual ou inferior a 99% na população. Logo, teoricamente qualquer fragmento de DNA pode ser empregado como marcador molecular, desde que revele polimorfismo entre indivíduos. Em geral, o número destes fragmentos é limitado e eles devem ser desprovidos de efeito epistático ou pleiotrópico (Ferreira & Grattapaglia, 1998; Borém & Caixeta, 2009).

Desde a sugestão feita por Botstein et al. (1980) para utilização de marcadores moleculares no mapeamento genético, inúmeras outras aplicações surgiram (Lee, 1995). Atualmente, com os avanços no desenvolvimento de tecnologias de genotipagem em larga escala, a partir do início do século XXI (Jenkins & Gibson, 2002; Wenzl et al., 2004), novos sistemas de marcadores moleculares como SNP (*Single Nucleotide Polymorphism*) e DArT (*Diversity Array Technology*) têm permitido eficiente avaliação genética em nível molecular. Tais tecnologias proporcionaram a redução do preço por *data point* e permitiram que grande número de marcadores fosse usado para várias culturas (Jehan & Lakhanpaul, 2006). Uma vez gerado grande número de marcadores espalhados por todo o genoma de um indivíduo, alguns destes marcadores, além de estarem muito pertos, estarão em desequilíbrio de ligação (LD) com QTLs (Hastbacka et al., 1994). O conceito de desequilíbrio de ligação refere-se à associação não aleatória entre dois genes ou entre um QTL e um loco marcador.

O desenvolvimento da tecnologia de SNP *chips* promoveu aumento exponencial no número de marcadores disponíveis para seres humanos, animais e vegetais, e permitiu estudos genômicos com alta densidade de marcadores. Existem SNP *chips* com 4 milhões de SNPs para estudos em seres humanos. Em bovinos, os SNP *chips* são formados por mais de 777 mil SNPs, o que barateou consideravelmente o custo da genotipagem por marcadores

moleculares. Atualmente, enquanto a genotipagem para um marcador microssatélite (SSR - *Simple Sequence Repeat*) custa cerca de U\$2,50/animal/marcador, a genotipagem por SNP é menos de U\$0,0005/animal/marcador; isso porque o custo de genotipagem para SNP *chip* de 777 Kb já pode ser realizado por menos de U\$400,00/animal (Illumina, 2011). Para vegetais, trabalhos relacionados à genotipagem com SNP em alta resolução também já têm sido desenvolvidos. Em algumas regiões do genoma do milho foi detectado um SNP por 70 pares de bases (pb); em trigo, um SNP por 20 pb (Jehan & Lakhanpaul, 2006); e em arroz, um SNP por 10 pb (McCouch et al., 2010).

### 3 SELEÇÃO ASSISTIDA POR MARCADORES (MAS)

Em todos os programas de melhoramento, a predição de valores genéticos de indivíduos candidatos à seleção tem sido uma estratégia muito útil para obtenção de ganhos com a seleção, seja por meio de métodos que se baseiam em dados fenotípicos, como o BLUP (Piepho et al., 2008), seja por dados genotípicos como os marcadores moleculares (Moose & Mumm, 2008). Muitos estudos já destacaram a importância dos marcadores moleculares como ferramenta para predição de valores genéticos (William et al., 2007).

Para emprego da seleção assistida por marcadores (MAS) deve-se utilizar simultaneamente dados fenotípicos e de marcadores moleculares em ligação gênica próxima com alguns QTLs. De forma geral, os dados de marcadores podem ser utilizados como covariáveis, de efeito fixo, na explicação dos valores fenotípicos dos genótipos candidatos à seleção, ou como efeitos aleatórios incorporados no modelo associado à fenotipagem (Resende, 2008).

O uso de MAS no melhoramento de plantas tem sido rotina em programas de melhoramento visando aumento dos ganhos com a seleção por unidade de tempo (Eathington et al., 2007). Há duas estratégias básicas para uso de MAS em programas de melhoramento: i) retrocruzamento de alelos favoráveis em germoplasma elite, por meio de retrocruzamento assistido por marcadores (MABC), e ii) predição de valores genéticos de indivíduos candidatos à seleção, por meio de seleção recorrente assistida por marcadores (MARS) (Bernardo, 2008).

O método MABC tem sido uma importante ferramenta de aplicação na piramidação de alelos de resistência em genótipos, haja vista que tal processo é extremamente difícil e trabalhoso, principalmente pela dificuldade em identificar, de modo preciso, os sintomas de resistência após múltiplas inoculações. O uso de marcadores moleculares, particularmente os que flanqueiam o alelo de resistência, pode diminuir consideravelmente o efeito do arraste de genes indesejáveis (*linkage drag*) durante o processo de transferência de alelos de resistência. Outra importante vantagem do método consiste em facilitar e acelerar a recuperação do genoma do genitor recorrente, por meio da definição de “*fingerprint*” molecular desse genitor, para contraste com aqueles obtidos dos



genótipos desenvolvidos ao longo do processo de melhoramento. Assim, apenas os genótipos portadores do alelo de resistência, ou seja, que possuam “*fingerprint*” semelhante ao do genitor recorrente, são selecionados para os próximos ciclos de retrocruzamento (Alzate-Marin et al., 2005).

De forma geral, a estratégia MARS possui como base o melhoramento de populações  $F_2$  por ciclo sucessivos de seleção assistida por marcadores; isto é, baseada em dados fenotípicos e moleculares em  $F_2$ , seguido por três ou mais ciclos de seleção baseada exclusivamente em marcadores. Resultados de simulação mostram que esta estratégia contribui com respostas superiores às obtidas pela seleção fenotípica, quanto ao acúmulo de alelos favoráveis em genótipos ao longo dos ciclos de seleção, apesar de ser mais indicada para caracteres de média a alta herdabilidade (Van Berloo & Stam, 1998, 2001). Por se enquadrar nos métodos de seleção assistida por marcadores, assim como o MABC, o método MARS utiliza apenas QTLs com efeitos significativos associados aos caracteres sob seleção.

Para emprego de MAS como estratégia de melhoramento, há necessidade da determinação do número, posição e efeitos dos QTLs marcados. Normalmente, QTLs de pequenos e raros efeitos não são detectados, resultando na capitação de apenas parte da variância genética devida aos marcadores/QTLs, o que leva à subestimação de tais efeitos (Goddard & Hayes, 2007). Dessa forma, seleções genotípicas realizadas com base em tais informações apresentam baixa precisão, pois sabe-se que os caracteres quantitativos, além de efeitos de QTLs, ainda poder apresentar efeitos de interações interalélicas (epistasia) e de interação de genótipos com ambientes (Podlich et al., 2004). Isso acaba restringindo a seleção genotípica para populações de distintos grupos gênicos, para diferentes ambientes ou por mais de um ciclo de seleção (Burgueño et al., 2012).

Têm-se verificado que MAS apresenta superioridade em relação à seleção fenotípica apenas quando o tamanho da população genotipada é muito grande, da ordem de 500 genótipos ou mais (Resende, 2008). Sabe-se, também, que a vantagem de MAS sobre a seleção fenotípica é proporcional à porcentagem da variância genética explicada pelo marcador. A questão chave, portanto, é determinar quantos marcadores/QTLs são responsáveis pela variação no caráter quantitativo e quantos são necessários para explicar a maior parte dessa variância genética. Na prática, a maioria dos marcadores explica muito pouco da variância genética e, portanto, o ganho em utilizá-los geralmente é muito pequeno. Desse modo, MAS tem sido mais utilizada em programas de melhoramento apenas para casos específicos, como introdução de alelos de herança monogênica em germoplasma,

seleção de plantas dentro de progênies e avanço de gerações em casa de vegetação, para caracteres de média ou alta herdabilidade (Hospital et al., 1997; Holland, 2004).

## **4 SELEÇÃO GENÔMICA AMPLA (GWS)**

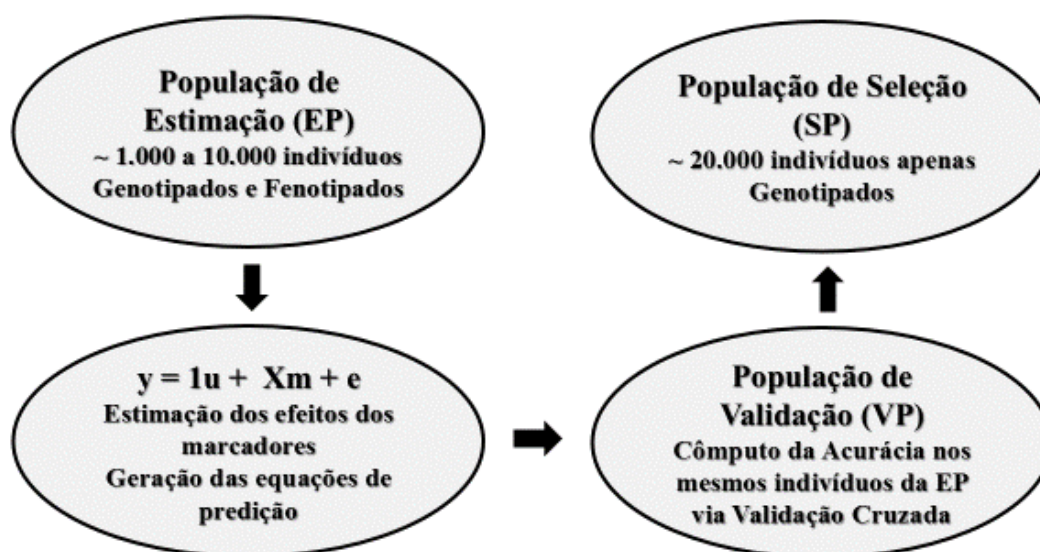
### **4.1 ASPECTOS GERAIS**

A seleção genômica ampla (GWS) é um método que integra as ferramentas da genética quantitativa e as tecnologias genômicas em uso no melhoramento, propiciando grande salto qualitativo nos sistemas de avaliação genética. Esta nova abordagem experimental vem rapidamente mudando os paradigmas do melhoramento genético de plantas, causando verdadeira revolução na capacidade preditiva e, com isso, aumento da acurácia seletiva em idade precoce, de forma a maximizar o ganho genético por unidade de tempo. Para isto, GWS enfatiza a predição simultânea (sem uso de testes de significância para marcadores individuais) dos efeitos genéticos de milhares de marcadores de DNA (SNP, DArT ou SSR) dispersos no genoma de um organismo, possibilitando capturar os efeitos da maioria dos locos (tanto de pequenos quanto de grandes efeitos) e explicar quase a totalidade da variação genética do caráter quantitativo. O método envolve três etapas: estimação e validação do modelo de predição; predição dos valores genéticos dos candidatos à seleção; e seleção dos genótipos baseada nas predições (Meuwissen et al., 2001; Resende, 2008).

Três populações podem ser definidas para a prática de GWS: população de treinamento ou de estimação (EP); população de validação (VP); e população de seleção (SP). Com base nelas, três estratégias populacionais podem ser empregadas, dependendo da natureza e exequibilidade do programa de melhoramento: (i) três populações fisicamente diferentes; (ii) uma só população usada para estimação e validação; e (iii) uma só população usada para estimação, validação e seleção. Têm-se verificado que, normalmente, que as estratégias (i) e (ii) são mais empregadas (Resende et al., 2010). A Figura 1 ilustra a estratégia (ii).

Na estimação do modelo de GWS, a população EP é fenotipada em condições normais de cultivo da espécie, isto é, em nível de campo, e genotipada para estimação dos efeitos dos marcadores ou dos haplótipos. As estimativas dos valores genéticos genômicos associadas a cada marcador ou alelo são utilizadas para cálculo do valor genético genômico

estimado (GEBV) de cada candidato à seleção. A estimativa GEBV corresponde à soma de todos os efeitos de marcadores ou de haplótipos incluídos no modelo, para determinado genótipo. A seleção dos genótipos candidatos à seleção é, então, realizada usando GEBV como critério de seleção. Desse modo, a seleção (GWS) atua capturando a variância genética aditiva, decorrente dos efeitos dos genes/QTLs que afetam o caráter quantitativo em análise (Heffner et al., 2009; Jannink et al., 2010).



**Figura 1.** Esquema de aplicação da seleção genômica ampla em programa de melhoramento genético (Resende et al., 2010).

Os efeitos de QTLs estimados no contexto de GWS são específicos para cada condição experimental, sendo necessário estimá-los para cada população e ambiente. Desse modo, modelos de estimação que incluem interação “genótipo x ambiente” podem ser utilizados com o objetivo de verificar a possibilidade de se conseguir estimativas para cada grupo de ambientes, dependendo da magnitude dessa interação (Resende et al., 2012). Por esta abordagem há, ainda, a possibilidade e o interesse de se identificar QTLs estáveis ao longo de vários ambientes.

Para estimação dos efeitos de QTLs em GWS pode-se utilizar de três estratégias: (i) uso de apenas dos marcadores moleculares; (ii) uso de haplótipos (intervalos definidos por dois marcadores); e (iii) uso de haplótipos definidos por mais de dois marcadores, com a inclusão de covariância entre haplótipos devida à ligação. Trabalhos têm demonstrado que para caracteres de baixa herdabilidade não existem diferenças significativas entre as três

estratégias (Callus et al., 2008). Segundo Solberg et al. (2006), é possível realizar GWS eficientemente com uso apenas dos marcadores, com a vantagem de não se necessitar da estimação das fases de ligação entre os marcadores, propensa a erros.

Em síntese, GWS contrasta com a estratégia MAS, pois não se restringe apenas aos marcadores identificados como de efeito significativo para predição e seleção, isto é, no uso de pequena quantidade de marcadores. Contrariamente, GWS pode usar os marcadores de pequenos efeitos, não significativos estatisticamente, para auxiliar na predição do valor genético. Este é um fator importante a ser considerado, haja vista que caracteres quantitativos são expressos por grande número de genes com pequeno efeito individual, ou seja, com distribuição infinitesimal (Resende, 2008, Nakaya & Isobe, 2012).

A GWS tem sido empregada com maior frequência na estratégia em que a mesma população é utilizada para estimação e validação. Neste caso, geralmente com aplicação de validação cruzada por Jackknife para estimação das acurácias seletivas. A validação cruzada é um método empregado para avaliar a capacidade de generalização de um modelo preditivo, a partir de um conjunto de dados. Na aplicação do método, o conjunto de dados é particionado em subconjuntos mutuamente exclusivos. Posteriormente, alguns subconjuntos são utilizados para a estimação dos parâmetros do modelo (dados de estimação), e os demais (dados de validação ou de teste), na sua validação. Existem diversos métodos para particionamento dos dados, sendo os mais utilizados: *holdout* (retenção), *k-fold* (k duplicado) e *leave-one-out* (deixar um fora) (Kohavi, 1995).

A acurácia do modelo de GWS é determinada por meio da correlação entre as estimativas GEBV (valor genético genômico estimado) e os valores genéticos (TBV) obtidos via fenotipagem. A abordagem da validação cruzada é geralmente empregada para avaliar a acurácia da seleção genômica, por meio de uma população com dados fenotipados e genotipados. Nesta abordagem, um subconjunto da população de estimação (EP) ou de validação (VP), dependendo da estratégia populacional empregada, é removido do modelo de GWS (Lee et al., 2008). Como o subconjunto de validação não foi envolvido na predição dos efeitos dos alelos ou haplótipos, os erros em TBV e GEBV são independentes, sendo toda correlação entre estes valores, de natureza genética e equivale à capacidade preditiva ( $r_{y\hat{y}}$ ) de GWS em estimar os fenótipos, podendo ser estimada pela própria acurácia seletiva ( $r_{g\hat{g}}$ ) multiplicada pela raiz quadrada da herdabilidade individual ( $h^2$ ), isto é;  $r_{y\hat{y}} = r_{g\hat{g}} \cdot h$ . Desde modo, na estimação da própria acurácia a seguinte expressão deve ser utilizada:  $r_{g\hat{g}} = r_{y\hat{y}} / h$ , quando são usados valores fenotípicos brutos para cálculo da correlação (Resende,

2008; Nakaya & Isobe, 2012). Na população de seleção (SP), por sua vez, é realizada apenas genotipagem nos candidatos à seleção, não necessitando de fenotipagem. As equações de predição oriundas de EP são, assim, utilizadas para predição de TBV, ou melhor, dos fenótipos futuros dos candidatos à seleção, com acurácia seletiva correspondente àquela calculada em VP (Bernardo & Yu, 2007; Resende, 2008).

Estudos de simulação e baseados em dados reais sobre GWS para vários caracteres em gado leiteiro têm detectado acurácias de GEBV maiores que 0,85, sendo que, na maioria dos casos, elas excederam àquelas estimadas com base em dados fenotípicos (Schaeffer, 2006; VanRaden et al., 2009; Calus, 2010; Meuwissen & Goddard 2010; Meuwissen et al., 2013). Em plantas também o método tem apresentado excelentes resultados, seja com dados simulados (Bernardo & Yu, 2007; Wong & Bernardo, 2008; Resende et al., 2008; Heffner et al., 2009; Grattapaglia et al., 2011; Hayes et al., 2013), ou baseados em dados reais; sejam em plantas anuais (Albrecht et al., 2011; Crossa et al., 2011; Fritsche Neto, 2011; Môro, 2011; Heslot et al., 2012) ou perenes (Resende Júnior, 2010; Cavalcanti et al., 2012).

#### 4.2 EFICIÊNCIA SELETIVA E ACELERAÇÃO DOS GANHOS DE SELEÇÃO

O uso de estratégias seletivas baseadas em marcadores moleculares para aumentar ganhos genéticos por unidade de tempo não é um conceito novo, principalmente para caracteres de difícil fenotipagem e natureza qualitativa (Eathington et al., 2007). Estudos de simulação (Wong & Bernardo, 2008; Zhong et al., 2009) e empíricos (Lorenzana & Bernardo, 2009) demonstraram que a seleção genômica ampla (GWS) aplicada em populações de plantas permite taxas de ganhos por unidade de tempo maiores que a seleção fenotípica e a seleção assistida por marcadores (MAS).

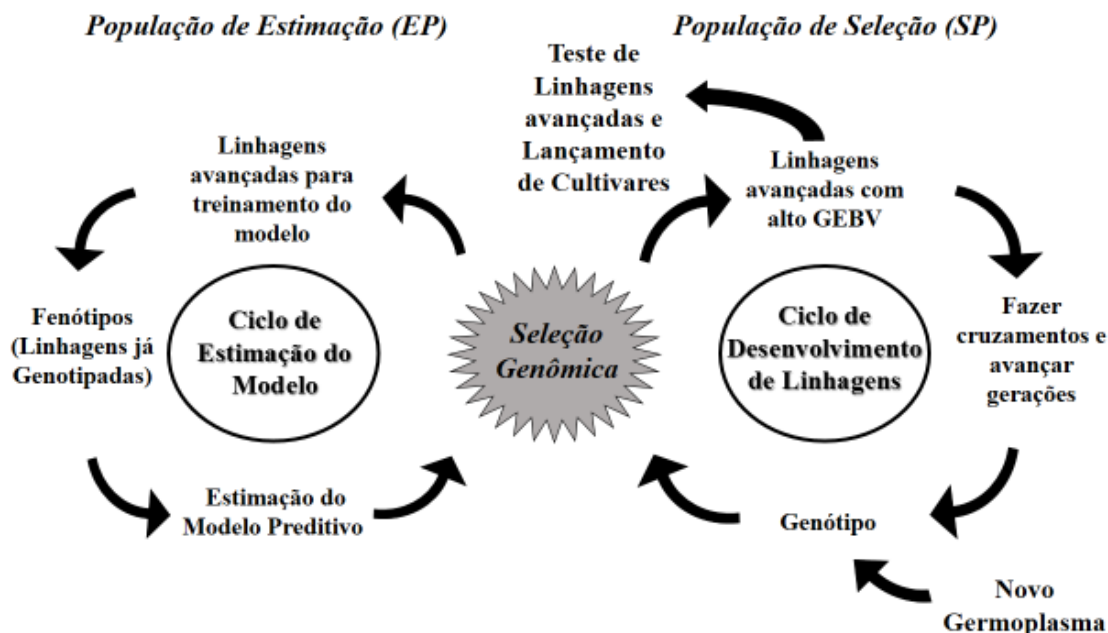
Simulações comparando GWS e MAS, em programa de melhoramento de milho (*Zea mays* L.), identificaram maior eficiência de GWS em elevar a resposta à seleção, principalmente para caracteres de baixa herdabilidade (Bernardo & Yu, 2007). Wong & Bernardo (2008) também observaram maiores ganhos proporcionados por GWS em relação à MAS em dendê (*Elaeis guineensis*), mesmo com populações de baixo tamanho efetivo, o que é normalmente empregado em espécies perenes. Heffner et al. (2010), comparando os custos de GWS e MAS, concluíram que acurácias de GEBV de apenas 0,50, obtidas por GWS, permitiram ganhos anuais duas vezes maiores que aqueles obtidos por MAS, em

programa de melhoramento de trigo (*Triticum aestivum* L.) com baixo investimento, e três vezes maiores em programa de melhoramento de milho com alto investimento. Em eucalipto (*Eucalyptus* sp.), Grattapaglia & Resende (2011) verificaram que GWS pode ser mais eficiente que a seleção fenotípica baseada em BLUP. Quando GWS é usada para encurtar o ciclo de melhoramento por seleção precoce de indivíduos, o método mostra forte vantagem sobre a seleção baseada em BLUP. O mesmo já havia sido relatado por Meuwissen et al. (2001) e Muir (2007), o que se justifica, pois a eficiência da seleção aumenta exponencialmente com a redução do tempo necessário para completar o ciclo de melhoramento.

Em GWS, de forma geral, o aumento da eficiência seletiva pode ser alcançado pela alteração nos quatro componentes envolvidos na expressão do ganho genético:  $\Delta G = \frac{i r_{g\hat{g}} \sigma_g}{L}$ , em que  $i$  é a intensidade de seleção;  $r_{g\hat{g}}$  é a acurácia na estimação de GEBV;  $\sigma_g$  é o desvio padrão genético; e  $L$  é o intervalo de tempo necessário para completar cada ciclo seletivo. É evidente na expressão que GWS possibilita: prática de maior  $i$ , por meio da avaliação de grande número de genótipos; maior  $r_{g\hat{g}}$ , devido ao baixo vício e variância do erro de predição, resultando em alta precisão; detecção de maior variância genética entre genótipos, em função da genotipagem com alta densidade de marcadores; e da redução de  $L$ , por meio da diminuição do número de gerações por ciclo, o que é possível devido à possibilidade de seleção precoce (Resende et al., 2012; Hayes et al., 2013) (Figura 2). Assim, considerando-se a redução do ciclo seletivo à metade, mesmo com  $r_{g\hat{g}}$  em torno de 40%, os esquemas seletivos baseados em GWS já possibilitariam ganhos genéticos superiores aos obtidos em esquemas baseados em seleção fenotípica; se, diferentemente, considerarmos valores de  $r_{g\hat{g}}$  de aproximadamente 60%, GWS poderia aumentar esses ganhos em até 50% (Hayes et al., 2013).

Cada espécie vegetal possui sua peculiaridade intrínseca à forma e fase de aplicação de GWS. Em espécies perenes, GWS possibilita a elevação de  $r_{g\hat{g}}$  e redução em  $L$  (tempo). O aumento em  $r_{g\hat{g}}$  decorre do uso da matriz de parentesco real de cada caráter, determinada pelos marcadores moleculares (Resende, 2007), ao invés, por exemplo, de parentescos obtidos por genealogia. Sabe-se, ainda, que o aumento de  $r_{g\hat{g}}$  depende do tamanho da população de estimação e da densidade de marcadores. O componente  $L$  é fortemente reduzido via GWS, haja vista que a predição genômica e a seleção podem ser realizadas precocemente nas plântulas. Este fator possibilita maiores ganhos com uso de

GWS em relação à seleção fenotípica, mesmo com  $r_{g\hat{g}}$  semelhantes para os dois tipos de seleção (Resende, 2008; Resende et al., 2011).



**Figura 2.** Esquema de seleção genômica ampla aplicada em programa de melhoramento genético, refletindo a possibilidade de redução do tempo e custo pela diminuição da frequência de fenotipagem (Heffner et al., 2009).

Em espécies vegetais alógamas e anuais, o aumento da eficiência seletiva proporcionado por GWS ocorre devido a três fatores: elevação de  $i$  e de  $r_{g\hat{g}}$ , e redução de  $L$ . A elevação em  $r_{g\hat{g}}$  é determinada pelo uso da matriz de parentesco real (Resende, 2007), e pela exploração de toda a variância genética populacional, não apenas aquela entre progênies. Sendo a seleção por GWS praticada em fase precoce, antes do florescimento, torna-se possível a seleção em nível de indivíduo e em ambos os sexos, sem a necessidade de duas estações de plantio: uma para a avaliação de progênies e outra para recombinação, no método de seleção recorrente. Desse modo, a redução em  $L$  ocorre em função da coincidência entre as unidades de seleção e de recombinação. Outro fator bastante favorável consiste na maximização da herdabilidade do método de seleção, devido à exploração adicional de 0,50 ou 0,75 da variância genética aditiva presente dentro de progênies. O aumento em  $i$  também pode ser alcançado pela seleção praticada em nível de indivíduo, refletindo na elevação da eficiência seletiva propiciada por GWS (Resende, 2008; Resende et al., 2011).



Em espécies vegetais autógamas e anuais, o uso de GWS também pode aumentar consideravelmente a eficiência seletiva. Quando duplo-haploides são utilizados no processo de desenvolvimento direto de linhagens, o benefício de GWS é verificado pelo aumento de  $i$ , de  $r_{g\hat{g}}$ , e da variância genética aditiva, além de redução em  $L$ . Neste caso, pode-se explorar máxima variância genética aditiva observada em linhagens homozigotas. Uma vez empregado o método genealógico de melhoramento, GWS não apresenta grandes benefícios ao ser aplicada na geração  $F_2$ , pois os genótipos em avaliação devem ser direcionados à homozigose, para a seleção final, não possibilitando redução de tempo ( $L$ ). Porém, com uso de GWS, alelos favoráveis podem ser identificados na geração  $F_2$ , com consequente direcionamento dos cruzamentos entre as melhores plantas e possibilidade de emprego da seleção recorrente intrapopulacional. Desse modo, GWS permite aumentar  $r_{g\hat{g}}$  e a variância genética, conferindo maior ganho genético ao programa, além do aumento em  $i$  devido ao maior número de plantas  $F_2$  avaliadas (Resende, 2008; Resende et al., 2011).

#### 4.3 TIPO E DENSIDADE DE MARCADORES

Com a rápida redução dos custos para genotipagem de indivíduos, por meio do uso de tecnologias de genotipagem em larga escala, será possível realizar GWS em todas as espécies vegetais. Para a obtenção de genomas efetivamente cobertos com marcadores, no mínimo um marcador deve estar em desequilíbrio de ligação (LD) com cada QTL, para que sejam detectados e expliquem a variância genética da população. O número mínimo de marcadores necessários para alcançar alta cobertura do genoma depende, portanto, da taxa de LD, que varia amplamente entre espécies, populações e genomas, devido às forças de mutação, recombinação, tamanho populacional e tipo de cruzamento (Flint-Garcia et al., 2003). De forma geral, a densidade de marcadores deve aumentar com a elevação de  $Ne*c$ , em que  $Ne$  é o tamanho efetivo populacional e  $c$  é a taxa de recombinação entre locos (Hill & Robertson, 1968). Em populações divergentes que passaram por muitas gerações de recombinação,  $Ne$  e  $c$  são grandes, e, assim, a densidade de marcadores será alta, com possibilidade de se alcançar forte LD entre marcadores e QTLs, com consequente melhor genotipagem.

O desequilíbrio de ligação (LD) entre marcadores e QTLs deve ser conservado entre a população de estimação (EP) e a população de seleção (SP), para que altas acurácias sejam alcançadas (Goddard, 2009). Por exemplo, se LD apresentar maior redução em SP do

que em EP, as acurácias irão declinar, pois os marcadores estarão segregando dos QTLs que foram estimados em EP. Este fato pode ocorrer se os genótipos em SP forem geneticamente divergentes de EP, ou se EP e SP forem separadas por múltiplas gerações de seleção. Portanto, EP deveria ter iguais ou maiores taxas de redução de LD que SP, e alta densidade de marcadores para reduzir a perda de acurácia em função da divergência genética ou múltiplas gerações de seleção (Muir, 2007; Zhong et al., 2009).

O aumento da densidade de marcadores tem contribuído para o aumento da acurácia (Solberg et al., 2008), assim como o aumento do número de fenótipos, que também eleva as estimativas de acurácia dos efeitos de marcadores ou de haplótipos (Bernardo & Yu, 2007; Hayes et al., 2009). Para qualquer herdabilidade e número de QTL, as respostas à GWS são tanto maiores quanto maior for o número de marcadores utilizado. Para obtenção de acurácias satisfatórias, Bernardo & Yu (2007) determinaram que na faixa de 128-256 marcadores polimórficos devem ser utilizados em GWS aplicada em população de milho duplo-haplóides.

A plataforma de marcadores utilizada para GWS também pode afetar a densidade requerida. A maioria dos estudos sobre GWS têm utilizado marcadores genéticos (SNP, DArT ou SSR) para genotipagem, ou seja, para estimar os efeitos de QTL. Resultados baseados em outros tipos de marcadores com sistema de genotipagem em alta densidade, como RAD (*restriction site-associated*) (Miller et al., 2007) e GBS (*genotyping by sequencing*), também têm sido obtidos (Elshire et al., 2011). Os marcadores DArT, RAD e GBS identificam polimorfismo por hibridização ou sequências de DNA digeridas por meio de enzimas de restrição, sendo ainda, marcadores dominantes (Nakaya & Isobe, 2012).

O marcador genético mais empregado atualmente em estudos genômicos é o SNP. Os dados relatados para SNP podem ser estendidos para outros tipos de marcadores bi-alélicos. Solberg et al. (2008) realizaram estudo visando avaliar o tipo e a densidade de marcadores para GWS. Os autores concluíram que SNP de genótipos possibilitam maiores acurácias de GEBV que SNP de haplótipos. Quando compararam SNP com SSR, foram detectadas acurácias semelhantes quando a densidade dos marcadores SNP foi 2-3 vezes maior que as de SSR. A respeito do requerimento de alta densidade, os marcadores SNP são ainda preferidos, devido à facilidade de automação e abundância no genoma. Marcadores do tipo DArT, assim como SNP, são também bialélicos e adequados à GWS, pois também são abundantes e podem ser obtidos com alta velocidade e rendimento. Porém, tais marcadores

são dominantes, sendo esta a desvantagem principal em relação a SNP, que são codominantes.

#### 4.4 CARACTERÍSTICAS DA POPULAÇÃO DE ESTIMAÇÃO

Estimativas acuradas de efeitos de QTL na população de seleção (SP) podem ser conseguidas por meio de aumento do número de informações obtidas na população de estimação (EP) (Zhong et al., 2009), e, também, pela re-estimação dos efeitos de QTL em populações em que a frequência alélica e a estrutura de LD muda devido à seleção (Goddard, 2009). A re-estimação dos efeitos de QTL pode ser obtida pela atualização do modelo de GWS via inclusão de genótipos candidatos à seleção em EP. Alguns estudos têm identificado efeitos significativos de interação QTL x base genética, em EP e SP, contribuindo para a redução das acurácias dos efeitos estimados de QTL. Considerando que tais efeitos são conservados ao longo das populações, se forem utilizadas altas densidades de marcadores e grandes populações de estimação, seria possível desenvolver modelos de GWS capazes de prever acuradamente GEBV de indivíduos distantemente relacionados com a EP (Meuwissen, 2009). Bernardo & Yu (2007) observaram que quantidade mínima de fenotipagem é necessária para GWS aplicada em milho duplo-haplóide; assim, indicaram a avaliação de aproximadamente 100 duplo-haplóides no ciclo 0. Caso este número seja reduzido à metade, a resposta à GWS poderá ser severamente limitada, indiferentemente do tamanho da população usada nos ciclos subsequentes para seleção baseada em marcadores.

Em plantas, muitos tipos de populações de estimação (EP) podem ser formadas. Recentes estudos avaliaram EPs formadas por cruzamentos bi-parentais (Lorenzana & Bernardo, 2009), duplo-haplóides (Bernardo & Yu, 2007) e intercruzamentos de linhagens endogâmicas (Zhong et al., 2009). A situação ideal pode ser a formação de uma nova EP para cada progênie derivada de cada cruzamento bi-parental, o que permitiria altas acurácias, pois as populações de seleção (SP) estariam diretamente relacionadas com a EP em relação a: desequilíbrio de ligação, frequência alélica, efeitos de QTL e base genética. Porém, para que a situação ideal seja atendida, também é necessário que a EP de cada cruzamento seja fenotipada em um conjunto de ambientes para o desenvolvimento do modelo de GWS. Desse modo, uma alternativa seria a formação de EPs constituídas por amostra representativa do programa de melhoramento. Tal procedimento permitiria a predição acurada de GEBV para genótipos candidatos à seleção, mesmo em curtos ciclos de seleção.

Estudos com bovinos têm indicado que caracteres com baixa herdabilidade requerem grandes populações de estimação para se atingir altas acurácias (Hayes et al., 2009). Goddard (2009) relatou que acurácias de GEBV estão relacionadas com o número de informações fenotípicas e com a herdabilidade do caráter, haja vista que a diminuição da herdabilidade contribui para baixas acurácias. Embora as acurácias de GEBV sejam maiores quando a herdabilidade do caráter é alta, neste caso a acurácia da seleção fenotípica ( $h$ ) também será alta, de modo que as acurácias de GEBV ( $r_{g\hat{g}}$ ) não excedem  $h$ . Assim, GWS é indicada para caracteres de baixa herdabilidade, sendo que  $r_{g\hat{g}}$  supera  $h$  quando houver, no mínimo, duas gerações incluídas na população de estimação (Muir, 2007). Para avaliar tais efeitos, Bernardo & Yu (2007) realizaram estudo de simulação com GWS em milho, sob diferentes cenários de herdabilidade:  $h^2 = 0,2, 0,5$  e  $0,8$ . Para os três níveis de  $h^2$  GWS mostrou maior resposta à seleção em relação às seleções fenotípica e MAS, sendo mais favorável à primeira quando  $h^2 = 0,2$  e o número de QTL = 100. As acurácias de GEBV reduziram com a diminuição de  $h^2$ . Deste modo, caracteres de baixa  $h^2$  apresentam um problema para ambas as seleções, fenotípica e genômica ampla, pois resultam em GEBV de acurácias reduzidas. Porém, isto pode ser revertido com grande investimento inicial na população de estimação.

#### 4.5 MÉTODOS ESTATÍSTICOS DE PREDIÇÃO GENÔMICA

Para a aplicação prática de GWS em programas de melhoramento, os efeitos de marcadores ou de haplótipos devem ser estimados, para, então, serem obtidos os GEBVs por meio de métodos de predição genômica. Desse modo, na implementação de GWS existem alguns desafios estatísticos computacionais, como a definição de métodos de predição genômica ampla que possibilitem melhor tratamento dos dados genômicos, considerando a sua dimensionalidade, a colinearidade entre marcadores e a complexidade dos caracteres quantitativos. A escolha de métodos estatísticos para predição dos efeitos de marcadores também pode afetar a acurácia de GEBV. Muitos métodos para predição genômica estão disponíveis atualmente (Callus, 2010; Jannink et al., 2010), seja paramétricos (Meuwissen et al., 2001) ou semi-paramétricos (Gianola et al., 2006; Gianola & Van Kaam, 2008).

Em modelos genéticos padrões, os dados fenotípicos,  $y_i$  ( $i = 1, \dots, n$ ), correspondem à soma de um valor genético,  $g_i$ , e do efeito residual,  $e_i$ , isto é:  $y_i = g_i + e_i$ . Em modelos paramétricos de GWS,  $g_i$  corresponde à regressão dos marcadores moleculares,

empregados como covariáveis  $x_{ij}$  ( $j = 1, \dots, p$  marcadores moleculares), ou seja,  $g_i = \sum_{j=1}^p x_{ij}\beta_j + \varepsilon_i$ , (ou  $y = X\beta + \varepsilon$ , em notação matricial), em que  $\beta_j$  é a regressão dos  $y_i$  sobre as referidas covariáveis  $x_{ij}$ . A estimação de  $\beta_j$  via regressão linear múltipla, pelo método de quadrados mínimos ordinários (OLS) ou generalizados (GLS), não é possível quando  $p > n$ , ou seja, o número de parâmetros a serem estimados (marcadores) é maior que o número de fenótipos avaliados (fenômeno denominado ‘grande  $p$ , pequeno  $n$ ’). Esta impossibilidade ocorre devido ao número insuficiente de graus de liberdade para ajustar todos esses efeitos de marcadores simultaneamente. Em OLS, as estimativas são obtidas visando maximizar o ajustamento do modelo para o conjunto de dados, sem considerar a complexidade do modelo. Quando o número de marcadores moleculares é elevado, como em estudos de seleção genômica ampla, normalmente o quadrado médio do erro das estimativas dos efeitos dos marcadores também é alto, gerando baixa capacidade preditiva (Resende et al., 2008; Resende et al., 2012).

Métodos adequados para predição genômica devem proporcionar a maior precisão possível na explicação da variância devida aos genes de pequenos e grandes efeitos, além de suas distribuições; minimizar a superparametrização do modelo para o grande número de marcadores no conjunto de dados, por meio de estimadores do tipo *shrinkage*; e basear-se, tanto quanto possível, nos efeitos oriundos da associação entre marcadores e QTLs, em vez do parentesco entre os genótipos. Os métodos também devem ser de fácil implementação para diferentes caracteres e conjuntos de dados, além de computacionalmente eficientes (Habier et al., 2007; Resende et al., 2012).

Vários métodos estatísticos de predição genômica têm sido propostos para GWS (Callus, 2010; Jannink et al., 2010). Os principais métodos podem ser divididos em três classes: regressão explícita, implícita e com redução dimensional. Os métodos da classe de regressão explícita podem ser divididos em dois grupos: (i) métodos de estimação penalizada, como RR-BLUP (Random “Ridge” Regression-Best Linear Unbiased Prediction) (Whittaker et al., 2000; Meuwissen et al., 2001), LASSO (Least Absolute Shrinkage and Selection Operator) (Tibshirani, 1996), Rede Elástica (Elastic Net) (Zou & Hastie, 2005) e RR-BLUP-Het (RR-BLUP com heterogeneidade de variância genética); e (ii) métodos de estimação bayesiana, tais como BayesA, BayesB (Meuwissen et al., 2001), Fast BayesB (Meuwissen, 2009), Bayes C, BayesC $\pi$ , BayesD, BayesD $\pi$  (Habier et al., 2011), LASSO Bayesiano, IBLASSO (LASSO Bayesiano melhorado) (Park & Casella, 2008; Campos et al., 2009) e outros. Na classe de regressão implícita, destacam-se as redes

neurais artificiais (RKHS – Reproducing Kernel Hilbert Spaces) (Gianola & Campos, 2009), que é um método semi-paramétrico, e a regressão Kernel não-paramétrica, via modelos aditivos generalizados (Gianola et al., 2006). Os métodos de regressão com redução dimensional, por sua vez, compreendem os de componentes independentes, quadrados mínimos parciais e de componentes principais (Solberg et al., 2009).

Os modelos estatísticos fornecem diferentes suposições sobre o número e os efeitos de QTL, possivelmente, afetando as respectivas acurácias de GEBV. De forma geral, os métodos apresentam abordagens que diferem na suposição sobre o modelo genético associado ao caráter quantitativo. O método RR-BLUP permite estimar todos os efeitos alélicos simultaneamente, pois assume o modelo infinitesimal com muitos locos de pequenos efeitos, ou seja, considera que os efeitos de QTL apresentam distribuição normal com variância constante ao longo dos segmentos cromossômicos (Resende et al., 2010). Por outro lado, os métodos bayesianos, com destaque para BayesB, assumem a existência de poucos genes de grandes efeitos e muitos genes com pequenos efeitos. Neste método muitos efeitos de marcadores são assumidos como zero, *a priori*, reduzindo o tamanho do genoma, o que permite o enfoque em regiões do genoma onde realmente existem QTLs. Quando os efeitos de QTL apresentam distribuição exponencial, sendo poucos desses efeitos com valor zero, o melhor estimador dos efeitos alélicos é o método LASSO. Desse modo, observa-se que o melhor método de predição genômica é aquele que reflete melhor a natureza biológica do caráter quantitativo sob análise, em termos de efeitos gênicos (Resende et al., 2008).

O procedimento RR-BLUP é semelhante ao BLUP tradicional, no entanto, na predição dos efeitos aleatórios, não é necessário o uso da matriz de parentesco (Schaeffer, 2006). A matriz de parentesco baseada em “pedigree”, usada no BLUP tradicional, é substituída pela matriz de parentesco estimada pelos marcadores. Deste modo, o método RR-BLUP é superior ao BLUP tradicional, pois efetivamente captura a matriz de parentesco realizada e não a matriz de parentesco médio associada às genealogias dos indivíduos (Resende et al., 2011).

Os métodos Bayesianos maximizam a distribuição *a posteriori* do parâmetro ou distribuição condicional do parâmetro dadas as observações, de acordo com o princípio de probabilidade condicional, e não mais frequentista, como em métodos de mínimos quadrados. Desde modo, a função de verossimilhança conecta a distribuição *a priori* do parâmetro à *posteriori*, usando para isto os dados experimentais (amostras). De forma geral,

na estimação Bayesiana a distribuição *a posteriori* possui o conhecimento prévio sobre o parâmetro e as informações adicionais propiciadas pelo experimento (Resende et al., 2012).

O método BayesA é semelhante ao método BLUP quando as variâncias dos efeitos de QTL são heterogêneas. Isto por que as variâncias se diferem para cada segmento cromossômico e são estimadas por meio do método BayesA considerando a informação combinada dos dados e da distribuição *a priori* para as variâncias. Os métodos Bayesianos normalmente proporcionam acurácias mais altas, pois induzem muitos efeitos de seguimentos cromossômicos a valores próximos a zero (BayesA) ou a zero (BayesB), e as estimativas dos efeitos dos demais seguimentos são regressadas de acordo com uma quantidade determinada pelas distribuições *a priori* dos efeitos de QTL (Resende, 2008; Resende et al., 2012).

Comparações entre métodos de predição genômica têm sido realizadas. Meuwissen et al. (2001) detectaram ligeira superioridade teórica do método BayesB, em relação a RR-BLUP. Habier et al. (2011) relataram que o método BayesA mostrou-se superior na maioria das situações, porém, nenhum dos métodos bayesianos foi claramente superiores em todas as situações. Mrode et al. (2010) também detectaram a superioridade do método BayesA e de Fast BayesB sobre o método BayesB. O método Fast BayesB (Meuwissen et al., 2009) foi desenvolvido visando reduzir o tempo de computação do método BayesB, originalmente implementado via simulação estocástica por meio de procedimento Monte Carlo via Cadeia de Markov (MCMC).

Os principais métodos de regressão com redução dimensional são aqueles de regressão via quadrados mínimos parciais (PLSR) e de regressão via componentes principais (PCR). Solberg et al. (2009) avaliaram estes métodos e verificaram que eles são mais simples e rápidos computacionalmente, embora menos acurados que o BayesB. Enquanto os primeiros atingiram acurácias da ordem de 0,68 (PLSR e PCR) e o último resultou em 0,84 (BayesB).

Heffner et al. (2011) avaliaram quatro modelos de predição genômica (RR-BLUP, BayesA, BayesB e BayesC) via acurácia de GEBV, em trigo. Para isso, analisaram 374 genótipos, cultivados em três locais por dois anos, genotipados com 1158 marcadores DArT e fenotipados para treze caracteres. Os autores detectaram acurácias que foram 28% superiores que as obtidas por MAS e 5% menos acuradas que a seleção fenotípica. Crossa et al. (2010) também avaliaram três modelos de predição genômica (RR-BLUP, LASSO bayesiano e RKHS) em trigo. Neste caso analisaram 599 genótipos, cultivados em quatro

ambientes, genotipados com 1279 marcadores DArT e fenotipados para produtividade de grãos. Os autores observaram acurácias para GEBV aproximadamente semelhantes nos três modelos.



## 5 SELEÇÃO RECORRENTE NA ERA GENÔMICA

### 5.1 SELEÇÃO RECORRENTE ASSISTIDA POR MARCADORES (MARS)

Esquemas seletivo baseados em seleção recorrente assistida por marcadores (MARS) visam desenvolver populações com potencial para extração de genótipos superiores, por meio do aumento da frequência de alelos favoráveis e manutenção da variabilidade genética, utilizando-se de informações relacionadas à identificação de QTL e seus efeitos. Desse modo, os alelos dos QTLs são acumulados ao longo dos sucessivos cruzamentos, por meio de seleção genotípica, sem necessidade de avaliações exclusivamente fenotípicas. De maneira geral, em MARS o mérito genético de cada indivíduo é predito após a identificação de QTLs de maior efeito para os caracteres sob seleção, a fim de melhorar o desempenho global da população através da seleção recorrente (Bernardo & Charcosset, 2006; Bernardo, 2008).

Basicamente, na aplicação de MARS utiliza-se um índice de seleção formado por dados fenotípicos do caráter y de interesse e dados dos marcadores moleculares (Lande & Thompson, 1990). Este índice pode ser descrito pela seguinte expressão:  $I_i = b_p P_i + b_m M_i$ , em que  $P_i$  é a informação fenotípica do indivíduo i;  $M_i = \sum b_j X_{ij}$  é a informação molecular ou score molecular associado ao indivíduo i;  $X_{ij}$  é o número observado (0 = ausência, 1 = presença) de um dos alelos do loco marcador j no indivíduo i;  $b_j$  é o coeficiente de regressão parcial associado ao loco j;  $b_p$  e  $b_m$  são os pesos dados às informações fenotípicas e moleculares, respectivamente. Deste modo, se dez QTLs são identificados com efeito significativo, por exemplo, cada um com algum efeito sobre o fenótipo, então um conjunto de genótipos será marcado com base nos dez alelos e o valor genético multiplicado pelo efeito de cada QTL na expressão fenotípica. A soma dos dez locos fornecerá a pontuação para cada um dos genótipos a serem avaliados, possibilitando o uso da seleção truncada para seleção dos melhores; por exemplo, 10% para recombinação (Bernardo, 2008).

Existem várias limitações no método MARS associadas aos QTLs. Primeiramente, os QTLs mapeados em uma população não expressam a mesma segregação

em outra população e, mesmo se isto ocorrer, os efeitos deles podem ser diferentes (até mesmo opostos) nas duas populações. Acrescenta-se que o mapeamento de QTL não identificará, em geral, todos os locos que contribuem para o caráter quantitativo, e, como consequência, os marcadores não irão explicar toda a variância genética, mas apenas uma parte dela. Ainda, mesmo se QTLs forem mapeados na população de interesse, seus efeitos irão mudar em ambientes diferentes. Neste caso, os marcadores conhecidos e associados com um caráter, nas condições em que o mapeamento foi realizado, podem não se associar ao caráter em ambientes onde será realizada a seleção. Ademais, as relações epistáticas entre os locos podem mudar quando a população é selecionada para locos específicos, e isto pode afetar negativamente o ganho genético. Além de tudo isso, estudos de QTL são limitados pelo fato de tipicamente envolverem apenas uma ou poucas famílias segregantes e número restrito de marcadores. Como resultado, somente um pequeno número de QTLs pode ser detectado, e seus efeitos são fortemente superestimados devido ao conhecido efeito “Beavis” (Beavis, 1998). O efeito “Beavis” causa um viés importante ao subestimar o número total de QTLs controlando o caráter, justamente pelo fato da análise não conseguir detectar QTLs de efeitos médios e pequenos (Bearzoti & Vencovsky, 2002; Bernardo & Charcosset, 2006; Bernardo, 2008; Gimelfarb & Lande, 2009).

Ultimamente, para a maioria dos caracteres quantitativos estudados, têm-se verificado que a seleção baseada em marcadores moleculares é menos efetiva que a seleção fenotípica, se o objetivo é obter ganhos genéticos. No entanto, a vantagem dos marcadores se relaciona, por exemplo, à possibilidade de seleção precoce de plantas em casa de vegetação, com a utilização do índice de seleção dos marcadores como critério de seleção, na ausência de quaisquer dados fenotípicos. Desse modo, a seleção pode ser bastante eficaz para caracteres pouco influenciados pelo ambiente e de natureza menos complexa, reduzindo os custos e podendo ser até mais rentável que a avaliação exclusivamente fenotípica (Bearzoti & Vencovsky, 2002; Bernardo & Charcosset, 2006; Bernardo, 2008; Gimelfarb & Lande, 2009).

## 5.2 SELEÇÃO RECORRENTE GENÔMICA (GERS)

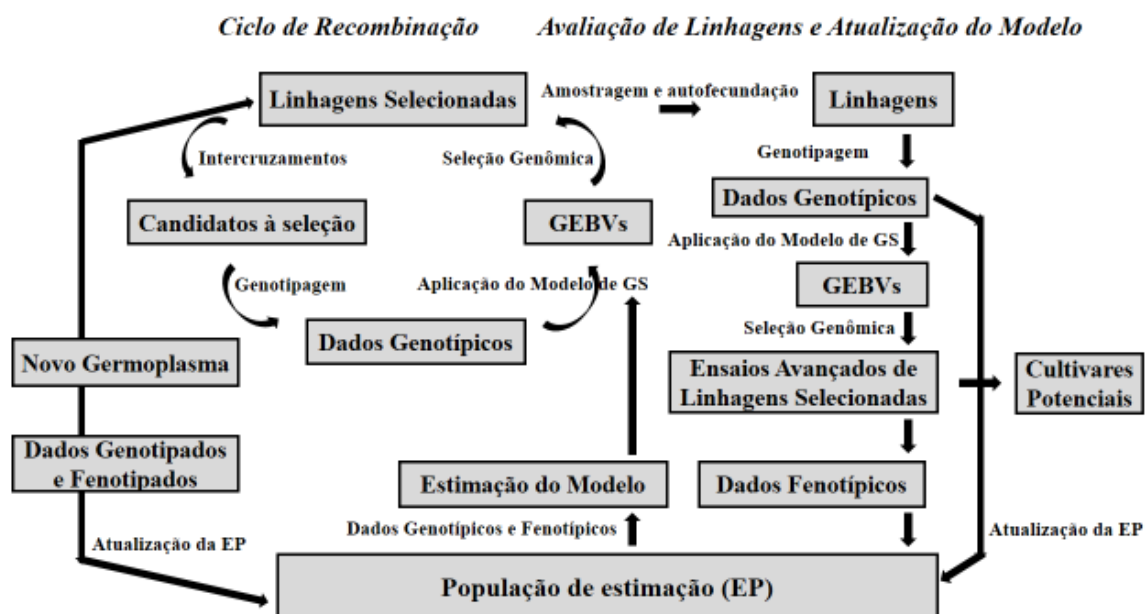
Além das vantagens já elencadas em seções anteriores para a seleção genômica ampla (GWS), outra baseia-se na facilidade de sua implementação no âmbito do método de seleção recorrente. Isso é especialmente de interesse em espécies de difícil exploração desse

método para melhoramento populacional, como em plantas autógamas. O sucesso da seleção recorrente têm sido relatado, tanto em espécies alógamas (Hallauer & Darrah, 1985), como em autógamas (Morais Júnior, 2013). Todavia, embora a seleção recorrente tenha mostrado sucesso também em autógamas, este método ainda não tem sido empregado em programas de melhoramento populacional como estratégia rotineira. O principal impedimento disso consiste no problema da determinação do valor genético de indivíduos, devido à dificuldade na produção de sementes, em adequada quantidade para realização de seleções baseadas em experimentos de teste de progênes com repetições. Desse modo, GWS possibilita a determinação direta do valor genético dos genótipos avaliados, por meio de sua genotipagem, ao invés da fenotipagem. Portanto, a seleção recorrente genômica (GERS), em plantas autógamas, requer pequena quantidade de sementes para seleções após intercruzamentos (Rutkoski et al., 2011).

Maiores ganhos de seleção são esperados com essa estratégia de melhoramento (GERS), quando comparada à seleção recorrente fenotípica e à seleção recorrente assistida por marcadores (MARS). A razão principal é devido à maior oportunidade de recombinação em função da redução dos ciclos de seleção, o que facilita a combinação de alelos favoráveis em um mesmo genótipo (Rutkoski et al., 2011). Na Figura 3 está apresentado um esquema de GERS que permite três ciclos de recombinação por ano, além de possível sistema de avaliação de linhagens recombinantes e de atualização da população de estimação, visando assegurar acurácias satisfatórias e estáveis ao longo dos ciclos de seleção. Nesse sistema, novo germoplasma pode ser introduzido na população de estimação e no ciclo de recombinação, ao mesmo tempo em que se mantém a variabilidade genética para possibilitar ganhos contínuos de seleção.

O esquema de GERS ilustrado na Figura 3 apresenta as seguintes características: o ciclo de recombinação consiste de rodadas de intercruzamento e seleção baseada em GEBV; para espécies autógamas, três a quatro ciclos de recombinação por ano são possíveis; a avaliação de linhagens e a atualização do modelo são realizadas simultaneamente; após pelo menos um ciclo de recombinação, as linhagens selecionadas são autofecundadas e selecionadas novamente com base em GEBV ou possíveis fenótipos, se necessário, para avaliação das linhagens e atualização da população de estimação; dados fenotípicos e genotípicos de linhagens selecionadas são usadas para atualizar o modelo; a diversidade genética pode ser mantida no programa de melhoramento pela introdução de novo germoplasma em cada ciclo de recombinação; e dados fenotípicos e genotípicos do novo

germoplasma devem ser introduzidos na população de estimação para atualizar o modelo de predição genômica (Rutkoski et al., 2011).



**Figura 3.** Esquema de seleção recorrente genômica (GERS) aplicada em espécies autógamas (Rutkoski et al., 2011).

Estudos têm sido realizados visando avaliar a potencialidade da estratégia GERS. Bernardo & Yu (2007) detectaram superioridade de GERS em relação à MARS, por meio de estudos de simulação. Os autores verificaram que o uso de GWS para seleção de genótipos em programas de seleção recorrente permite maiores respostas à seleção do que sob MARS. Este resultado era esperado, pois a GWS permite estimar os valores genéticos dos indivíduos na população por meio da soma de todos os efeitos dos marcadores. Muitas limitações relacionadas com a detecção de associações significativas entre marcadores e QTL são contornadas com uso de GWS, pois os marcadores presentes em alta densidade no genoma podem capturar melhor a variância genética total para os caracteres de interesse (Rutkoski et al., 2011).

A produção de sementes por meio de cruzamentos manuais é, conforme já apontado, um fator limitante para uso da seleção recorrente em espécies autógamas. Para contornar esta situação, Bernardo (2010) propôs um esquema de seleção recorrente modificado, denominado *select-recombine-self* (selecionar, recombinar e autofecundar), que faz uso de GWS entre indivíduos  $F_2$ , em espécies em que a insuficiente produção de sementes

inibe a seleção e recombinação na geração  $F_1$ . Os autores verificaram que a estratégia GERS aplicada em indivíduos  $F_2$  resultou em respostas à seleção ligeiramente inferiores à GWS aplicada em indivíduos  $F_1$ , pois a geração adicional reduziu os ganhos anuais. Deste modo, o esquema *select-recombine-self* pode ser uma alternativa interessante para espécies em que a produção de sementes  $F_1$  limita a capacidade de produzir populações  $F_1$  suficientemente grandes para aumento da eficiência da GWS.

## 6 CONSIDERAÇÕES FINAIS

A seleção genômica ampla (GWS), pelas potencialidades descritas na presente revisão, representa uma estratégia muito promissora aos programas de melhoramento de plantas. Em especial deve-se destacar que GWS garante melhoria nos ganhos em seleção por unidade de tempo e custo, principalmente para caracteres quantitativos que expressam baixa herdabilidade, e em espécies perenes ou ciclo longo. Para espécies autógamas, em que a seleção recorrente em geral não é aplicada rotineiramente devido à dificuldade de realização de cruzamentos e consequente limitação de sementes para uma seleção baseada em experimentos com repetições, GWS também apresenta-se como bastante promissora, pois facilita a aplicação do método. A GWS pode ser implementada por meio do uso das tecnologias de marcadores moleculares atuais e uma população de estimação (EP) com indivíduos cuidadosamente selecionados e associados aos candidatos à seleção. Porém, com o constante melhoramento e redução dos custos das tecnologias de genotipagem, é possível incluir facilmente maior diversidade de indivíduos na EP, condicionando maiores acurácias aos modelos de predição genômica. Resultados de simulação e com dados reais têm revelado o grande potencial dessa estratégia em aumentar a eficiência do melhoramento. Porém, para alcançar tal benefício é necessário: alto grau de desequilíbrio de ligação envolvendo marcadores moleculares (SNPs, DArTs ou SSR) proximamente espaçados; elevada acurácia seletiva, determinada pela estimação dos efeitos genéticos dos marcadores nos caracteres sob melhoramento, a partir dos fenótipos; uso dos efeitos genéticos dos marcadores na própria população e ambientes em que forem estimados; e aquisição de experiência prática com GWS, por parte dos melhoristas, visando inferir sobre a sua efetividade e potencial de aplicação.

As plataformas de genotipagem vêm se tornando cada vez mais viáveis e financeiramente acessíveis aos programas de melhoramento de plantas, de forma a possibilitarem a aplicação de GWS na maioria das espécies de importância econômica, pelo menos em médio prazo. Como os custos de genotipagem continuam a reduzir, a seleção fenotípica tornará, comparativamente, mais onerosa para amplo número de caracteres de

interesse. Mesmo com os atuais custos em genotipagem e capacidade de aplicação limitada, a estratégia de GWS pode aumentar os ganhos em seleção por unidade de tempo e custo para muitas espécies, principalmente em plantas perenes e espécies autógamas.

## 7 REFERÊNCIAS

ALBRECHT, T.; WIMMER, V.; AUINGER, H.; ERBE, M.; KNAAK, C.; OUZUNOVA, M.; SIMIANER, H.; SCHÖN, C. C. Genome-based prediction of testcross values in maize. **Theoretical and applied genetics**, v. 123, n. 2, p. 339–350, 2011.

ALZATE-MARIN, A. L.; CERVIGNI, G. D. L.; MOREIRA, M. A.; BARROS, E. G. Seleção assistida por marcadores moleculares visando ao desenvolvimento de plantas resistentes a doenças, com ênfase em feijoeiro e soja. **Fitopatologia brasileira**, v. 30, n. 4, p. 333–342, 2005.

BEARZOTI, E.; VENCOVSKY, R. Simulation of marker-assisted recurrent selection in autogamous species. **Crop Breeding and Applied Biotechnology**. v. 2, n. 1, p. 1–10, 2002.

BEAVIS, W. D. QTL analyses: power, precision and accuracy. In: PATERSON, A. H. **Molecular analysis of complex traits**. Boca Raton: FL, 1998. p. 145–16.

BERNARDO, R. **Breeding for quantitative traits in plants**. Stemma Press, Woodbury, MN. 2002. 369 p.

BERNARDO, R.; CHARCOSSET, A. Usefulness of gene information in marker-assisted recurrent selection: a simulation appraisal. **Crop Science**, v. 46, n. 2, p. 614–621, 2006.

BERNARDO, R.; YU, J. Prospects for genomewide selection for quantitative traits in maize. **Crop Science**, v. 47, n. 3, p. 1082-1090, 2007.

BERNARDO, R. Molecular markers and selection for complex traits in plants: learning from the last 20 years. **Crop Science**, v. 48, n. 5, p. 1649-1664, 2008.

BERNARDO, R. Genomewide selection with minimal crossing in self-pollinated crops. **Crop Science**, v. 50, n. 2, p. 624-627, 2010.

BORÉM, A.; CAIXETA, E. T. **Marcadores moleculares**. 2 ed. Viçosa, MG: Editora UFV. 2009.

BOTSTEIN, D. R. L.; WHITE, M. S.; DAVIS, R W. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. **American journal of human genetics**, v. 32, n. 3, p. 314–31, 1980.



- BRIGGWS, K. G.; SHEBESKI, L. H. Visual selection for yielding ability of F<sub>3</sub> lines in a hard red spring wheat breeding program. **Crop Science**, Madison, v. 10, n. 4, p. 400-402, 1970.
- BURGUEÑO, J.; CAMPOS, G. L.; WEIGEL, K.; CROSSA, J. Genomic prediction of breeding values when modeling genotype × environment interaction using pedigree and dense molecular markers. **Crop Science**, v. 52, n. 2, p. 707-719. 2012.
- CALLUS, M. P. L. MEUWISSEN, T. H. E.; ROOS, A. P. W.; VEERKAMP, R. F. Accuracy of genomic selection using different methods to define haplotypes. **Genetics**, v. 178, n. 1, p. 553-561, 2008.
- CALLUS, M. P. L. Genomic breeding value prediction: methods and procedures. **Animal**, v. 4, n. 1, p. 157–164, 2010.
- CAMPOS, G. DE LOS.; NAYA, H.; GIANOLA, D.; CROSSA, J.; LEGARRA, A.; MANFREDI, E.; WEIGEL, K. Predicting quantitative traits with regression models for dense molecular markers. **Genetics**, v. 182, n. 1, p. 375-385, 2009.
- CAVALCANTI, J. J. V.; RESENDE, M. D. V.; SANTOS, F. H. C.; PINHEIRO, C. R. Predição simultânea dos efeitos de marcadores moleculares e seleção genômica ampla em cajueiro. **Revista Brasileira de Fruticultura**. v. 34, n. 3, p. 840–846, 2012.
- COTES, J. M. Predicting quantitative traits with regression models for dense molecular markers. **Genetics**, v. 182, n. 1, p. 375–385, 2009.
- CROSSA, J.; CAMPOS, G. DE LOS.; PÉREZ, P.; GIANOLA, D.; BURGUEÑO, J.; ARAUS, J. L.; MAKUMBI, D.; SINGH, R. P.; DREISIGACKER, S.; YAN, J.; ARIEF, V.; BANZIGER, M.; BRAUN, H. J.; YAN, I. Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. **Genetics**. v. 186, n. 2, p. 713–24, 2010.
- CROSSA, J.; PÉREZ, P.; CAMPOS, G. DE LOS.; MAHUKU, G.; DREISIGACKER, S.; MAGOROKOSHO, C. Genomic Selection and Prediction in Plant Breeding. **Journal of Crop Improvement**, v. 25, n. 3, p. 239–261, 2011.
- CUTRIM, V. A.; RAMALHO, M. A. P.; CARVALHO, A. M. Eficiência da seleção visual na produtividade de grãos de arroz (*Oryza sativa* L.) irrigado. **Pesquisa Agropecuária Brasileira**, v. 32, n. 6, p. 601–606, 1997.
- EATHINGTON, S. R.; CROSBIE, T. M.; EDWARDS, M. D.; REITER, R.; BULL, J. K. Molecular markers in a commercial breeding program. **Crop Science**, v. 47, n. 1, p. 154–163, 2007.
- FALCONER, D. S.; MACKAY, T. F. C. **Introduction to quantitative genetics**. 4. ed. Edinburgh: Longman Group Limited, 1996. 464 p.

FERREIRA, M. E.; GRATTAPAGLIA, D. **Introdução ao uso de marcadores moleculares em análise genética**. 3. ed. Brasília: EMBRAPA-CENARGEN, 1998. 220p.

FLINT-GARCIA, S.; THORNSBERRY, J. M.; BUCKLER, E. S. Structure of linkage disequilibrium in plants. **Annual Review of Plant Biology**, v. 54, n. 1, p. 357–374, 2003.

FRITSCH NETO, R. **Seleção genômica ampla e novos métodos de melhoramento do milho**. 2011. 39 f. Tese (Doutorado em Genética e Melhoramento de Plantas) – Universidade Federal de Viçosa, Viçosa, 2011.

GIANOLA, D.; FERNANDO, R. L.; STELLA, A. Genomic-assisted prediction of genetic value with semiparametric procedures. **Genetics**, v. 173, n. 1, p. 1761–1776, 2006.

GIANOLA, D.; CAMPOS, G. DE LOS. Inferring genetic values for quantitative traits non-parametrically. **Genetical research**, v. 90, n. 1, p. 525–540, 2009.

GIMELFARB, A.; LANDE, R. Simulation of marker assisted selection in hybrid populations. **Genetical Research**, v. 63, n. 1, p. 39-47, 1994.

GODDARD, M. E.; HAYES, B. J. Genomic selection. **Journal of animal breeding and genetics**, v. 124, n. 6, p. 323–30, 2007.

GODDARD, M. Genomic selection: prediction of accuracy and maximisation of long term response. **Genetica**, v. 136, n. 2, p. 245–57, 2009.

GRATTAPAGLIA, D.; RESENDE, M. D. V. Genomic selection in forest tree breeding. **Tree Genetics & Genomes**, v. 7, n. 2, p. 241–255, 2011.

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

HABIER, D.; FERNANDO, R. L.; KIZILKAYA, K.; GARRICK, D. J. Extension of the bayesian alphabet for genomic selection. **BMC bioinformatics**, v. 12, n. 1, p. 1-12, 2011.

HALLAUER, A. R.; DARRAH, L. L. Compendium of recurrent selection methods and their application. **Critical Reviews in Plant Sciences**, v. 3, n. 1, p. 1–33, 1985.

HAYES, B. J.; BOWMAN, P. J.; CHAMBERLAIN, A. J.; GODDARD, M. E. Genomic selection in dairy cattle: progress and challenges. **Journal Dairy Science**, v. 92, n. 1, p. 433–443, 2009.

HAYES, B. J.; COGAN, N. O. I.; PEMBLETON, L. W.; GODDARD, M. E.; WANG, J.; SPANGENBERG, G. C.; FORSTER, J. W. Prospects for genomic selection in forage plant species. **Plant Breeding**, v. 132, n. 2, p. 133–143, 2013.

HASTBACKA, J.; CHAPPELLE, A.; MAHTANI, M. M.; CLINES, G.; REEVE-DALY, M. P.; DALY, M.; HAMILTON, B. A.; KUSUMI, K.; TRIVEDI, B.; WEAVER, A. The

diastrophic dysplasia gene encodes a novel sulfate transporter: positional cloning by fine-structure linkage disequilibrium mapping. **Cell**, v. 78, n. 1, p. 1073–1087, 1994.

HEFFNER, E. L.; SORRELLS, M. E.; JANNINK, J. L. Genomic selection for crop improvement. **Crop Science**, v. 49, n. 1, p. 1-12, 2009.

HEFFNER, E. L.; LORENZ, A. J.; JANNINK, J. L.; SORRELLS, M. E. Plant breeding with genomic selection: Potential gain per unit time and cost. **Crop Science**, v. 50, n. 50, p. 1681–1690, 2010.

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

HENDERSON, C. R. **Sire evaluation and genetic trends**. In: Animal Breeding and Genetics Symposium in Honor of J. Lush. Champaign: American Society of Animal Science, 1973. p. 10–41.

HESLOT, N.; YANG, H. P.; SORRELLS, M. E.; JANNINK, J. L. Genomic selection in plant breeding: a comparison of models. **Crop Science**, v. 52, n. 1, p. 146-160, 2012.

HILL, W. G.; ROBERTSON, A. Linkage disequilibrium in finite populations. **Theoretical and Applied Genetics**, v. 38, n. 1, p. 226–231, 1968.

HOLLAND, J. B. Implementation of molecular markers for quantitative traits in breeding programs - challenges and opportunities. In: FISCHER, T. **New directions for a diverse planet: ProceedinGWS for the 4th International Crop Science Congress**. Brisbane. 2004. p. 1–13.

HOSPITAL, F.; MOREAU, L.; LACOUDRE, F.; CHARCOSSET, A.; GALLAIS, A. More on the efficiency of marker-assisted selection. **Theoretical and Applied Genetics**, v. 95, n. 1, p. 1181–1189, 1997.

JANNINK, J. L.; LORENZ, A. J.; IWATA, H. Genomic selection in plant breeding: from theory to practice. **BriefinGWS in functional genomics**, v. 9, n. 2, p. 166–77, 2010.

JEHAN, T.; LAKHANPAUL, S. Single nucleotide polymorphism (SNP)– Methods and applications in plant genetics: A review. **Indian Journal of Biotechnology**, v. 5, n. 4, p. 435–459, 2006.

JENKINS, S.; GIBSON, N. High-throughput SNP genotyping. **Comparative and Functional Genomics**. v. 3, n. 1, p. 57–66, 2002.

KOHAVI, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: ProceedinGWS of the 14th international joint conference on artificial intelligence, 1995, San Francisco. **Anais...** CA, USA: Morgan Kaufmann Publishers Inc, 1995, p. 1137–1143.

LEE, M. DNA markers and plant breeding programs. **Agronomy Journal**, v. 55, n. 1, p. 265–344, 1995.

LEE, S. H.; WERF, J. H. J. V. D.; HAYES, B. H.; GODDARD, M. E.; VISSCHER, P. M. Predicting unobserved phenotypes for complex traits from whole-genome SNP data. **PLoS genetics**, v. 4, n. 10, p. 1-11, 2008.

LONG, N.; GIANOLA, D.; ROSA, G. J. M.; WEIGEL, K. A.; AVENDAÑO, S. Machine learning classification procedure for selecting SNPs in genomic selection: application to early mortality in broilers. **Journal of Animal Breeding and Genetics**, v. 124, n. 6, p. 377–389, 2007.

LORENZANA, R. E.; BERNARDO, R. Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. **Theoretical and Applied Genetics**, v. 120, n. 1, p. 151–161, 2009.

MCCOUCH, S. R.; ZHAO, K.; WRIGHT, M.; TUNG, C. W.; EBANA, K.; THOMSON, M.; REYNOLDS, A.; WANG, D.; DECLERCK, G.; ALI, L.; MCCLUNG, A.; EIZENGA, G.; BUSTAMANTE, C. Development of genome-wide SNP assays for rice. **Breeding Science**, v. 535, n. 60, p. 524–535, 2010.

MEUWISSEN, T. H. Accuracy of breeding values of ‘unrelated’ individuals predicted by dense SNP genotyping. **Genetics Selection Evolution**, v. 41, n. 1, p. 1-9, 2009.

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

MEUWISSEN, T.; HAYES, B. J.; GODDARD, M. E. Accelerating improvement of livestock with genomic selection. **Annual Review of Animal Biosciences**, v. 1, n. 1, p. 221–237, 2013.

MILLER, M. R.; DUNHAM, J. P.; AMORES, A.; CRESKO, W. A.; JOHNSON, E. A. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. **Genome Research**, v. 17, n. 1, p. 240–248, 2007.

MOOSE, S. P.; MUMM, R. H. Molecular plant breeding as the foundation for 21st century crop improvement. **Plant Physiology**, v. 147, n. 3, p. 969–77, 2008.

MORAIS JÚNIOR, O. P. **Variabilidade e progresso genético com seleção recorrente em arroz de terras altas**. 2013. 163 f. Dissertação (Mestrado em Genética e Melhoramento de Plantas) – Escola de Agronomia e Engenharia de Alimentos, Universidade Federal de Goiás, Goiânia, 2013.

MRODE, R.; COFFEY, M.; BERRY, D. P. **Understanding Genomic Evaluations from Various Evaluation Methods and GMACE**. Riga: Latvia, Interbull bulletin, n. 42, p. 52-55, 2010.

- MUIR, W. M. Comparison of genomic and traditional BLUP- estimated breeding value accuracy and selection response under alternative trait and genomic parameters. **Journal Animal and Breeding Genetics**, v. 124, n. 1, p. 342–355, 2007.
- NAKAYA, A.; ISOBE, S. N. Will genomic selection be a practical method for plant breeding? **Annals of botany**, v. 110, n. 6, p. 1303–1316, 2012.
- PARK, T.; CASELLA, G. The Bayesian LASSO. **Journal of the American Statistical Association**, v. 103, n. 482, p. 681–686, 2008.
- PIEPHO, H. P.; MÖHRING, J.; MELCHINGER, A. E.; BÜCHSE, A. BLUP for phenotypic selection in plant breeding and variety testing. **Euphytica**, v. 161, n. 1-2, p. 209–228, 2008.
- PODLICH, D. W.; WINKLER, C. R.; COOPER, M. K. Mapping as you go : an effective approach for marker-assisted selection of complex traits. **Crop Science**, v. 44, n. 1, p. 1560–1571, 2004.
- RAMALHO, M. A. P.; ABREU, Â. F. B.; SANTOS, J. B.; NUNES, J. A. R. **Aplicações da genética quantitativa no melhoramento de plantas autógamas**. Lavras: Universidade Federal de Lavras, 2012. 522 p.
- RESENDE JÚNIOR, M. F. R. **Seleção genômica ampla no melhoramento vegetal**. 2010. 78 f. Dissertação (Mestrado em Genética e Melhoramento de Plantas) – Universidade Federal de Viçosa, Viçosa, 2010.
- RESENDE, M. D. V. **Matemática e estatística na análise de experimentos e no melhoramento genético**. Colombo, PR: Embrapa Florestas, 2007. 561 p.
- RESENDE, M. D. V.; SILVA, F. F.; VIANA, J. M. S.; PETERNELLI, L. A.; RESENDE JÚNIOR, M. F. R. VALLE, P. M. **Métodos estatísticos na seleção genômica ampla**. Colombo, PR: Embrapa Florestas, 2011. 107 p.
- RESENDE, M. D. V.; RESENDE JÚNIOR, M. F. R.; AGUIAR, A. M.; ABAD, J. I. M.; MISSIAGGIA, A. A.; SANSALONI, C.; PETROLI, C.; GRATTAPAGLIA, D. **Computação da Seleção Genômica Ampla (GWS)**. Colombo, PR: Embrapa Florestas, 2010. 79 p.
- 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**, v. 1, n. 56, p. 63–77, 2008.
- 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, MG: Editora UFV, 2012, 291 p.
- RESENDE, M. D. V. **Genômica Quantitativa e Seleção no Melhoramento de Plantas Perenes e Animais**. Colombo, PR: Embrapa Florestas, 2008. 324 p.

- RUTKOSKI, J. E.; HEFFNER, E. L.; SORRELLS, M. E. Genomic selection for durable stem rust resistance in wheat. **Euphytica**, v. 179, n. 1, p. 161–173, 2011.
- SCHAEFFER, L. R. Strategy for applying genome-wide selection in dairy cattle. **Journal of animal breeding and genetics**, v. 123, n. 4, p. 218–223, 2006.
- SOLBERG, T. R.; SONESSON, A.; WOOLLIAMS, J.; MEUWISSEN, T. H. E. Genomic selection using different marker types and density. In: World Congress of Genetics Applied to Livestock Production, 8., 2006. Proceeding. **Anais...** Belo Horizonte: Ed. Da UFMG, 2006.
- SOLBERG, T. R.; SONESSON, A. K.; WOOLLIAMS, J. A.; MEUWISSEN, T. H. E. Genomic selection using different marker types and densities. **Journal of animal science**, v. 86, n. 10, p. 2447–2454, 2008.
- SOLBERG, T. R.; SONESSON, A. K.; WOOLLIAMS, J. A.; MEUWISSEN, T. H. E. Reducing dimensionality for prediction of genome-wide breeding values. **Genetics Selections Evolution**, v. 41, n. 29, p. 1-8, 2009.
- TIBSHIRANI, R. Regression shrinkage and selection via the Lasso. **Journal of the Royal Statistical Society Series B**, v. 58, n. 1, p. 267–288, 1996.
- VAN BERLOO, R.; STAM, P. Marker-assisted selection in autogamous RIL populations: a simulation study. **Theoretical and Applied Genetics**, v. 96, n. 1, p. 147–154, 1998.
- VAN BERLOO, R.; STAM, P. Simultaneous marker-assisted selection for multiple traits in autogamous crops. **Theoretical and Applied Genetics**, v. 102, n. 6-7, p. 1107–1112, 2001.
- WENZL, P.; CARLING, J.; KUDRNA, D.; JACCOUD, D.; HUTTNER, E.; KLEINHOF, A.; KILIAN, A. Diversity Arrays Technology (DArT) for whole-genome profiling of barley. **Proceedings of the National Academy of Sciences of the United States of America**, v. 101, n. 1, p. 9915–9920, 2004.
- WILLIAM, H. M.; TRETOWAN, R.; CROSBY-GALVAN, E. M. Wheat breeding assisted by markers: CIMMYT's experience. **Euphytica**, v. 157, n. 1, p. 307–319, 2007.
- WONG, C. K.; BERNARDO, R. Genomewide selection in oil palm: increasing selection gain per unit time and cost with small populations. **Theoretical and applied genetics**, v. 116, n. 6, p. 815–824, 2008.
- ZHONG, S.; DEKKERS, J. C. M.; FERNANDO, R. L.; JANNINK, J. L. Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a Barley case study. **Genetics**, v. 182, n. 1, p. 355–364, 2009.
- ZOU, H.; HASTIE, T. Regularization and variable selection via the elastic net. **Journal of the Royal Statistical Society**, v. 67, n. 2, p. 301–320, 2005.