

MELHORAMENTO DE PRECISÃO

*Aplicações e perspectivas
na genética de plantas*

**RAFAEL TASSINARI RESENDE
CLAUDIO BRONDANI**

Editores Técnicos



*Empresa Brasileira de Pesquisa Agropecuária
Embrapa Arroz e Feijão
Ministério da Agricultura e Pecuária*

Melhoramento de Precisão

Aplicações e perspectivas na genética de plantas

Rafael Tassinari Resende
Claudio Brondani

Editores Técnicos

*Embrapa
Brasília, DF
2023*

Rosana Pereira Vianello, Rafael Tassinari Resende,
Claudio Brondani

Introdução

Genômica é a parte da genética dedicada ao estudo do genoma completo de organismos, e convencionou-se aplicar esse conceito a toda e qualquer análise molecular baseada no DNA dos indivíduos. Os avanços de tecnologias genômicas ampliaram a compreensão da variação de caracteres, tanto de herança mendeliana quanto de herança complexa, e regiões genômicas associadas a tais caracteres têm sido amplamente identificadas mediante estudos de mapeamento de ligação e análises de associação (Lin et al., 2022). A combinação da tecnologia de marcadores moleculares e do melhoramento de plantas é conhecida por “melhoramento molecular” e envolve um conjunto de atividades que visam o desenvolvimento de novas cultivares comerciais, mais produtivas e com maior valor agrônômico agregado (Sharma et al., 2021).

As tecnologias de marcadores moleculares são aplicadas no melhoramento genético de plantas para diversas finalidades e compreendem tanto o seu uso em atividades de rotina, que visam aumentar a eficiência de seleção, quanto acelerar o tempo de obtenção de determinada cultivar comercial. A genotipagem também pode ser usada na caracterização de germoplasma, na determinação de identidade genética e pureza de sementes, no estudo da heterose, na introgressão de genes assistida por marcadores e no retrocruzamento assistido por marcadores. Adicionalmente, pode ser aplicada em atividades de pesquisa que busquem a identificação de genes de interesse agrônômico para caracteres pouco complexos por meio das análises de mapeamento de QTLs (*Quantitative Trait Loci*), Estudos de Associação Genômica Ampla (do inglês *genome-wide association studies* – GWAS) e na piramidação gênica assistida por marcadores previamente identificados como associados a caracteres de interesse.

Em especial, no melhoramento vegetal, a Seleção Genômica Ampla (SGA, do inglês *genome-wide selection* – GWS) é uma técnica cada vez mais popular para o avanço de gerações e o desenvolvimento de cultivares, na qual os marcadores moleculares são integrados ao esquema convencional de melhoramento vegetal e/ou utilizados para substituir a seleção fenotípica convencional (Grattapaglia, 2022).

A eficácia na aplicação de análise genômica para a identificação de regiões genômicas responsáveis pela expressão de caracteres fenotípicos em programas de melhoramento é fortemente dependente do tamanho do genoma da espécie, do seu hábito reprodutivo, da população de melhoramento utilizada (tamanho e relacionamento genético entre os indivíduos), da natureza genética do caráter fenotípico (tipo de herança e influência ambiental sobre ele) e da robustez e cobertura genômica das marcas moleculares empregadas. Além disso, o desequilíbrio de ligação (DL) genômico é um importante fator a ser considerado e, portanto, será bem enfatizado ao longo deste capítulo. O meio de desenvolvimento também afetará a estabilidade das regiões genômicas controladoras das características quantitativas (QTLs) e deve ser investigado no contexto de interações dos Genótipos com Ambientes ($G \times A$), com o tempo de desenvolvimento e maturação ($G \times T$), com as tecnologias de manejo aplicadas ($G \times M$), ou mesmo as interações conjuntas entre todos esses fatores ($G \times A \times T \times M$). É importante ressaltar que também deve ser considerado os *backgrounds* genéticos distintos sob a perspectiva da ocorrência de dominância alélica e interações gênicas, como a epistasia (Anilkumar et al., 2022).

Compreendendo a importância da análise genômica no contexto do Melhoramento de Precisão e da Agricultura de Precisão, este artigo de revisão tem como objetivo examinar os principais marcadores moleculares utilizados, as tecnologias disponíveis para identificação desses marcadores e explorar a aplicação dessas informações na caracterização molecular de genitores e linhagens, na seleção assistida por marcadores e nos estudos de associação genômica ampla. Busca-se fornecer uma visão abrangente dos avanços da análise genômica no Melhoramento de Precisão, contribuindo para o avanço da eficiência e produtividade agrícola.

Marcadores Moleculares mais Utilizados em Análise Genômica

O que se deseja de um marcador molecular é que o sistema de genotipagem utilizado seja de custo acessível, rápido, reproduzível e capaz de discriminar geneticamente os acessos. Os marcadores microssatélites (ou SSR, *simple*

sequence repeats) foram desenvolvidos na década de 1980, e foram rapidamente adotados por serem capazes de detectar vários alelos por loco, apresentarem um alto nível de polimorfismo decorrente das elevadas e heterogêneas taxas de mutação e um padrão de segregação codominante, ou seja, capaz de diferenciar indivíduo homocigoto do heterocigoto para determinado loco com elevada precisão (Kumar et al., 2022). Embora os marcadores SSRs sejam uma valiosa ferramenta de análise genética para responder importantes questões biológicas, como determinação de parentesco e diversidade genética, alguns estudos têm demonstrado que estimativas mais precisas de diferenciação e estruturação podem ser obtidas a partir da análise com marcadores SNPs (*single nucleotide polymorphisms*, ou polimorfismos de nucleotídeo único), apesar de que a escolha por determinado marcador deve ser cuidadosamente avaliada para cada espécie (Zimmerman et al., 2020). Adicionalmente, a análise por marcadores SSR demandam equipamento e pessoal especializado, devido à automação reduzida e à impossibilidade técnica de genotipar centenas de locos simultaneamente, dentre outros.

Os marcadores baseados em SNPs são o tipo mais comum de variação no DNA e a frequência de um SNP pode variar consideravelmente ao longo do genoma em plantas autógamas e alógamas, podendo abranger de algumas centenas a milhares de pares de base (Sharma et al., 2021), e apresentam natureza codominante. SNPs têm alta capacidade multiplex, possibilitando a genotipagem de centenas a milhares de SNPs ao mesmo tempo, o que reduz o custo por *data-point*. Os avanços nas tecnologias de sequenciamento, proporcionando incremento no desempenho do processo e acurácia das sequências, têm permitido a criação de grandes conjuntos de dados de SNPs, aumentando consideravelmente o número de locos facilmente amostrados, com menor custo em comparação ao desenvolvimento e genotipagem por microssatélites (Singh et al., 2022). Além disso, os avanços nas metodologias estatísticas e o aumento do poder de computação levaram à adoção rápida dessa tecnologia (Yu; Chung, 2021), e, conseqüentemente, o seu uso tem sido ampliado pelos programas de melhoramento. Para o arroz, uma série de dados públicos possibilita que sejam escolhidos marcadores informativos sem a necessidade de realizar nenhum sequenciamento adicional, já que, por exemplo, mais de 3.000 acessos foram resequenciados e tornados públicos (Li et al., 2014). Adicionalmente, chips comerciais de diferentes densidades (quantidade de SNPs) estão disponíveis para a caracterização de genótipos a preços competitivos (Thudi et al., 2021).

A análise molecular de plantas com características contrastantes ainda é considerada uma das ferramentas mais poderosas para identificar, isolar e usar os genes relacionados à expressão de caracteres de interesse e QTLs. Da mesma

■ Melhoria de precisão: aplicações e perspectivas na genética...

forma, populações naturais ou populações fechadas (derivadas de cruzamentos direcionados) podem ser exploradas para descobrir marcadores moleculares ligados a genes agronomicamente importantes por meio de mapeamento de associação pela análise de desequilíbrio de ligação (DL). O número de marcadores para alcançar cobertura representativa de todo o genoma depende da determinação do DL, que é variável tanto entre populações de uma mesma espécie, quando dentro do genoma da própria espécie (Hyten et al., 2010). O interesse no estudo de DL, ou seja, associação não aleatória de alelos, em plantas cultivadas aumentou dramaticamente nos últimos anos devido a dois fatores principais. Primeiro, as tecnologias genômicas permitem a identificação rápida de haplótipos (segmentos cromossômicos transmitidos em conjunto para a próxima geração) em muitos locos mendelianos, seja por sequenciamento de DNA ou pela detecção de marcadores SNP. Em segundo lugar, na presença de DL significativo, é possível identificar regiões genômicas que estão associadas a um caráter de interesse (por exemplo, resistência a doenças) por varredura de genoma de indivíduos de uma população existente. Diante disso, estabelecer a magnitude do DL é vantajoso para determinar o número de marcas necessárias para realizar estudos de seleção assistida, associação e seleção genômica. Por outro lado, devido ao DL, variantes causais subjacentes à variação fenotípica são difíceis de identificar, dificultando a condução da tecnologia de edição de genomas no melhoramento de plantas (Wang et al., 2020).

Tecnologias Disponíveis para Identificação de Marcadores Genômicos

Vários tipos de marcadores moleculares estão disponíveis, obtidos por diferentes métodos de detecção, como a hibridização, a reação da polimerase em cadeia (*polymerase chain reaction*, PCR) e o sequenciamento de DNA (Zhang et al., 2022). Com o advento da PCR na década de 1980, diversas tecnologias moleculares passaram a integrar rotineiramente os programas de melhoramento genético vegetal (Figura 1). Embora quase todos os tipos de marcadores moleculares clássicos ainda sejam usados em alguma escala, e a depender do propósito, as tecnologias de sequenciamento de nova geração (*Next Generation Sequencing*, NGS), aliadas à redução dos custos e de mão de obra e ao aumento da acurácia, têm permitido rápidos avanços nos processos de análise genômica.

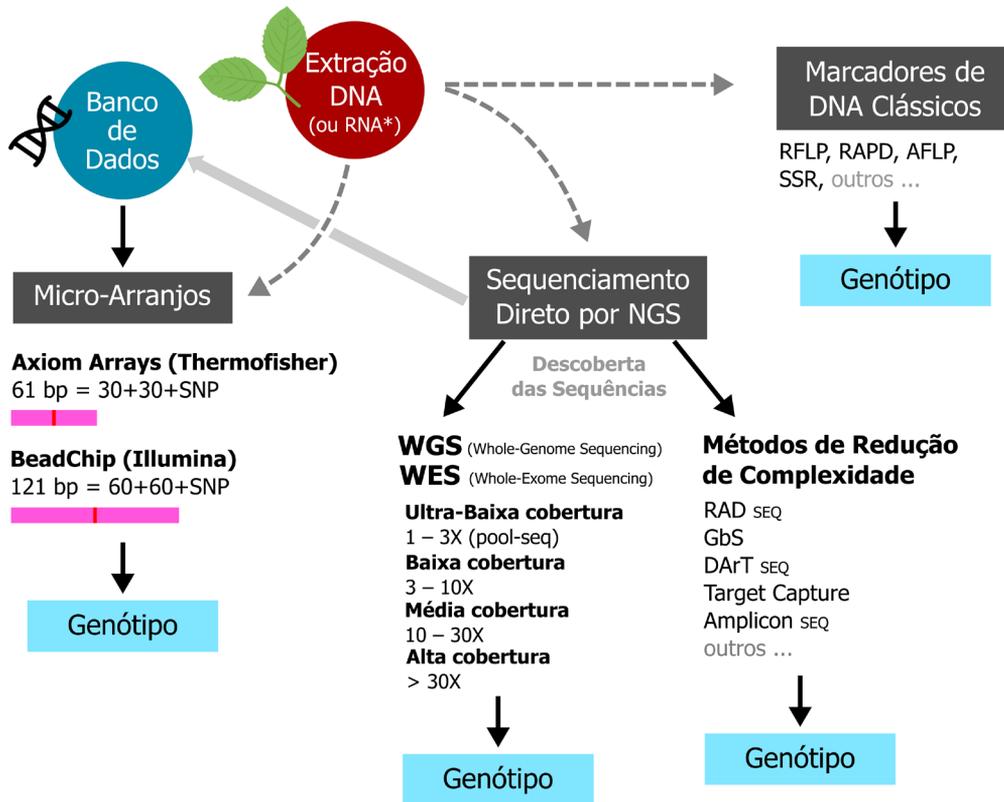


Figura 1. Organograma simplificado dos processos de genotipagem, considerando as novas tecnologias e marcadores moleculares clássicos. Em todas elas, “Genótipo” representa a disponibilização de marcadores genéticos para as amostras (Com colaboração do Dr. Orzenil Silva-Júnior).

As plataformas de *NGS*, cuja comercialização iniciou no ano de 2004 com o modelo Roche GS20 (empresa 454 Life Sciences, EUA), ainda estão em franca evolução, gerando informações com um alcance de bilhões de pares de bases em uma única corrida, e fragmentos com comprimentos de leitura que podem ser acima de 10Kb, permitindo análise ampla dos genomas, com precisão ao nível de nucleotídeos/pares de base (Hu et al., 2021). O processo de sequenciamento de genoma estrutural começa com o isolamento de DNA a partir da coleta de amostras de tecidos vegetais, como folhas, caules ou raízes. Em seguida, o DNA é tipicamente fragmentado mecanicamente ou por enzimas, ligado a adaptadores, e as bibliotecas genômicas são construídas (o “Banco de dados” da Figura 1). A biblioteca de DNA é então colocada no sequenciador NGS, que “lê” as sequências de DNA. Os dados de

■ Melhoramento de precisão: aplicações e perspectivas na genética...

sequenciamento obtidos são geralmente alinhados a um genoma de referência da espécie, possibilitando a identificação de variantes nesse genoma, como os SNPs, os *indels* (inserções/deleções de fragmentos de DNA) e variações estruturais. O acesso crescente às tecnologias NGS tem permitido aumento do volume de projetos de genômica estrutural e funcional (genoma expresso) em todo o mundo, viabilizando o sequenciamento de genomas de espécies modelos e não modelos. As descobertas científicas derivadas da aplicação de NGS resultaram em grande impacto na área de melhoramento genético/seleção genômica, proporcionando incremento na precisão das análises genético-estatísticas.

O rápido desenvolvimento do NGS propiciou o advento da tecnologia de genotipagem por sequenciamento, comumente denominada de GbS (*Genotyping by Sequencing*; Elshire et al., 2011). Esse processo de genotipagem utilizando sequenciamento pode ser realizado pelo resequenciamento do genoma completo (*Whole-Genome Resequencing* – WGR), propiciando alta resolução, ou pelo sequenciamento de uma representação reduzida (*Reduced-Representation Sequencing* - RRS), em que são sequenciadas apenas regiões específicas do genoma obtidas por redução de complexidade pelo uso de enzimas de restrição, de sondas de captura, amplificação de alvos via PCR ou, ainda, pelo sequenciamento do genoma transcrito (RNA-seq). Uma das principais aplicações de GbS é justamente no melhoramento de plantas, fornecendo uma ferramenta rápida e de baixo custo para estudos de associação genética, na identificação de QTLs e criação de mapas de alta densidade de marcadores, na seleção genômica, caracterização de germoplasma e estudos de diversidade genética (Rayaprolu et al., 2022).

Diferentes abordagens de genotipagem por sequenciamento em plantas estão disponíveis atualmente e possibilitam genotipar milhares de marcadores em grande volume de amostras a custos bastante acessíveis (Scheben et al., 2017). Elas variam principalmente no método utilizado para a redução da complexidade do genoma e se é possível ou não a seleção antecipada de quais e quantas regiões alvo no genoma serão genotipadas. Alguns exemplos são:

- **RADseq** (*Restriction-site Associated DNA sequencing*) é uma técnica de sequenciamento de DNA que utiliza enzimas de restrição para fragmentar o genoma. O objetivo é obter regiões específicas para sequenciar, geralmente aquelas com alta variabilidade genética, mas não é possível selecionar as regiões que serão sequenciadas além da escolha da enzima a ser utilizada. A principal vantagem desse método é sua capacidade de genotipar grande número de indivíduos simultaneamente e com baixo custo. Ele é utilizado para estudos

de diversidade genética, mapas genéticos e análise de associação genômica. Além disso, ele também é utilizado para a análise de genomas de espécies não modelo, em que a informação genética prévia é limitada.

- **Target Capture**, é um método que se concentra em selecionar regiões específicas do genoma para análise, usando sondas específicas que hibridizam com as regiões alvo no genoma e concentram o sequenciamento nesses alvos. Isso permite análise mais precisa e eficiente de regiões genômicas onde estão localizados genes candidatos ou marcadores moleculares pré-definidos, por exemplo.
- **Amplicon seq**, como o anterior, é um método de sequenciamento que se concentra em regiões específicas do genoma, como genes, marcadores moleculares ou regiões conservadas. O processo inclui a amplificação dessas regiões usando PCR e, em seguida, o sequenciamento das amplificações geradas. Isso permite análise detalhada de regiões específicas do genoma, como a variação genética, a evolução e a função dos genes.
- **DArTseq** (*Diversity Array Technology sequencing*), que assim como RADseq utiliza enzimas de restrição para reduzir a complexidade do genoma, seguido pela detecção via NGS, gerando grande quantidade de marcadores genéticos a custos bastante acessíveis.
- **RNA-seq** (*RNA sequencing*) se refere ao sequenciamento do transcrito utilizando a técnica de NGS para a análise dos padrões de expressão gênica, sem que seja necessário conhecimento prévio do genoma. Tal procedimento consiste no sequenciamento do cDNA (DNA complementar), que é obtido a partir do mRNA (RNA mensageiro) de uma determinada amostra biológica. Esse método permite a análise de toda a expressão gênica de uma amostra, incluindo a detecção de transcritos não codificantes, como miRNAs (micro-RNAs) e lncRNA (*long non-coding* RNA). Além disso, é possível identificar variações gênicas, como *splicing* alternativo e variações de expressão gênica. O RNA-seq é utilizado para estudar a regulação da expressão gênica em diferentes condições biológicas e para identificar diferenças de expressão entre diferentes genótipos ou entre tecidos ou estágios de desenvolvimento vegetal.

As abordagens descritas anteriormente possibilitam e geralmente são utilizadas para a descoberta de milhares de SNPs distribuídos no genoma e, conseqüentemente, permitem o desenvolvimento de plataformas de genotipagem em larga escala para SNPs específicos. A tecnologia de painéis fixos de SNPs,

■ Melhoria de precisão: aplicações e perspectivas na genética...

que podem ser desenvolvidos e analisados por meio de diversas plataformas de genotipagem, utilizando ou não sequenciamento (SNP-chips, SNPs-assays, Flex-Seq, por exemplo), permitem aplicações em atividades de rotina e pesquisa pela integração permanente dos dados moleculares obtidos em estudos distintos, apresentam elevada reprodutibilidade dentro e entre laboratórios, maior rapidez na análise dos resultados, facilidade de acesso pelos usuários e custos reduzidos (Sun et al., 2021). Os SNP chips, que são os DNA microarrays, correspondem a um suporte de vidro (chip) contendo milhares de oligonucleotídeos (sondas), que correspondem a SNPs específicos roboticamente imobilizados que podem ser escolhidos a partir de um banco de dados de SNPs ou de clones de cDNA previamente gerados. A genotipagem por *array* também pode ser utilizada para detectar variações no número de cópias de um gene (*copy number variation*, CNV) e variações estruturais do genoma, como deleções. A tecnologia de painéis fixos reunindo grande número de SNPs, que possibilita estudos diversos, tem sido desenvolvida e adotada com sucesso para arroz (Lv et al., 2021), eucalipto (Silva-Junior et al., 2015), guandu (Singh et al., 2022), dentre outros. Entretanto, a seleção assistida para algumas dezenas de locos de interesse agrônômico ainda é baseada, prioritariamente, em sistema de genotipagem individualizado de baixo rendimento, como KASP™ e sondas de hidrólise (He et al., 2014), as quais possuem custos elevados por *data-point*, o que dificulta a aplicação em número elevado de marcadores por amostra (Rasheed et al., 2017). Nos últimos anos, sistemas de genotipagem que possibilitam a multiplexagem de um número intermediário de SNPs estão se tornando mais acessíveis, ocupando uma lacuna entre microarrays, no qual milhares de SNPs são genotipados, e a PCR em tempo real, em que poucas dezenas de SNPs são genotipados. A multiplexagem intermediária de SNPs baseia-se no sequenciamento NGS por PCR, hibridação e captura (*target sequencing*) que permitem selecionar regiões específicas do genoma, tornando as análises ainda mais rápidas e menos onerosas. Algumas dessas tecnologias são comercialmente conhecidas como AgriSeq™ Targeted-GbS (Thermo Fisher), Capture-Seq and Flex-Seq (Rapid Genomics), SNP-TAGs (Diversity Arrays), SNPSelect (KeyGene), dentre outras (Semalayiappan et al., 2023).

Na Figura 2 é ilustrada a distribuição de marcadores SNPs genotipados em uma plataforma do tipo BeadChip (Illumina Inc., San Diego, CA, USA) após o controle de qualidade para uma população híbrida de *Eucalyptus grandis* × *E. urophylla*. O eucalipto é um gênero que possui inúmeras espécies nativas na Oceania. O *E. grandis* tem um genoma de aproximadamente 697 Mb, enquanto que no *E. urophylla* é de aproximadamente 626 Mb.

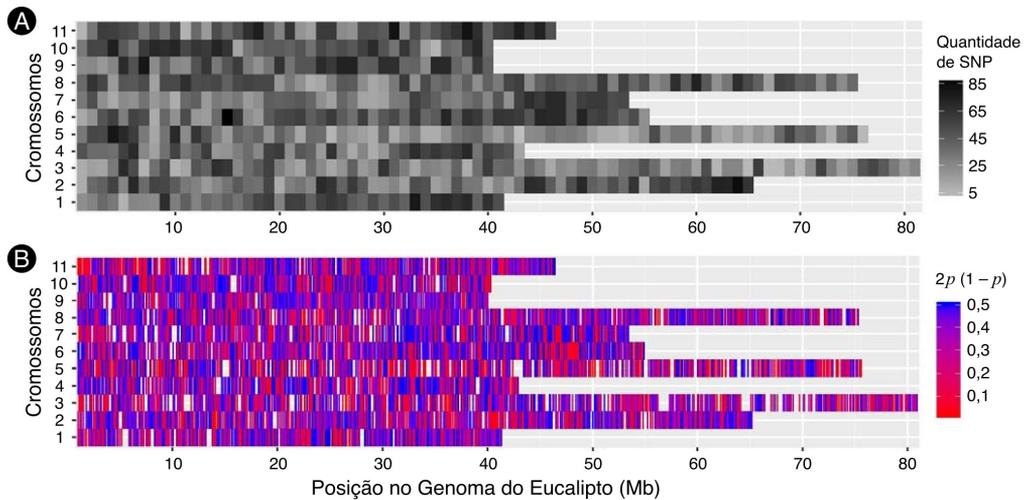


Figura 2. Distribuição de 24.806 marcados SNP polimórficos ao longo dos onze cromossomos de uma população de eucalipto. A parte “A” da figura mostra a concentração de SNPs por janela de 1 Mb. A parte “B” da figura mostra a heterozigose média dos SNPs em uma janela de 100 kb.

Fonte: Adaptado de Resende et al. (2017).

Análise Genômica no Melhoramento de Plantas

O uso de marcadores moleculares para a caracterização genética de germoplasma por meio da obtenção das estimativas dos parâmetros genéticos populacionais é considerado um caso de sucesso da genômica aplicada aos programas de melhoramento de plantas. Essas análises tornaram-se parte importante do melhoramento molecular e, de modo bastante preciso, o gerenciamento do germoplasma assistido pela genômica possibilita explorar a diversidade genética de modo direcionado na base de programas de melhoramento genético (Milner et al., 2019). Dados moleculares também são utilizados para a prospecção de locos sob assinatura de seleção que atuam moldando o genoma em função da seleção natural e artificial do germoplasma (Civan et al., 2021), o que pode aumentar a frequência de alelos favoráveis para genes específicos visando a ampliação da base genética de modo direcionado (Sharma et al., 2021).

O melhoramento genético assistido pela genômica é demonstrado de modo crescente na literatura (Varshney et al., 2021). As plataformas baseadas em genotipagem via sequenciamento (GbS) e painéis fixos de SNPs têm ampliado o uso das ferramentas moleculares no melhoramento de plantas, permitindo

■ Melhoria de precisão: aplicações e perspectivas na genética...

potencialmente a aplicação da seleção assistida por marcadores. No entanto, é importante ressaltar que a implementação e adoção de marcadores moleculares no melhoramento requerem uma abordagem integrada e devem levar em consideração os custos envolvidos, os quais incluem a identificação, validação e operacionalização de um conjunto de marcadores úteis para as análises moleculares, a infraestrutura laboratorial necessária, as capacidades analíticas e os recursos humanos disponíveis. A obtenção de um *data-point* na análise de marcadores moleculares tem um custo aproximado de 50 centavos de dólar, sem considerar os custos associados ao desenvolvimento do marcador ou à determinação da real associação entre um alelo do marcador e uma característica relevante no germoplasma de interesse. Portanto, é fundamental considerar a necessidade de validação de cada marcador e determinar a reprodutibilidade dos resultados obtidos, a fim de garantir a confiabilidade da aplicação prática. Diante disso, a partir da definição de um conjunto de análises que efetivamente tragam benefícios para a seleção assistida, é importante enfrentar o desafio de reduzir os custos da análise genética e aumentar a sua agilidade, garantindo que os métodos de marcadores moleculares sejam adequados ao contexto em que serão usados (Sharma et al., 2021).

Caracterização Molecular de Genitores e Linhagens em um Programa de Melhoramento

Bancos de dados de frequências alélicas podem ser montados a partir da caracterização de um conjunto selecionado de marcadores SNPs, o que permite administrar a variabilidade genética disponível para os programas pelo direcionamento de cruzamentos, além de introduzir genótipos que possam aumentar a possibilidade de serem obtidas novas combinações gênicas favoráveis e, com isso, via cruzamentos e seleção, darem origem a linhagens e cultivares superiores. Em um programa de melhoramento, a identificação de acessos e linhagens é muito importante. Em experimentos de melhoramento, várias linhagens são avaliadas e, portanto, elas podem ser contaminadas devido à mistura de amostras de sementes e polinização cruzada em campo. Isso pode levar à identificação incorreta das sementes das linhagens. Tais acessos são difíceis de distinguir, pois diferem em poucas características morfológicas. Contudo, marcadores moleculares podem facilmente distinguir esses genótipos intimamente relacionados. Outra aplicação importante dos marcadores moleculares em um programa de melhoramento é a sua utilização como descritores acessórios em testes de DHE (distinguilidade, homogeneidade e estabilidade), necessário para o registro e proteção de cultivares. O atual sistema de proteção de cultivares

vegetais baseia-se na descrição morfológica. As avaliações de DHE determinam se uma nova variedade é distinguível das cultivares conhecidas e se exibe uniformidade e estabilidade fenotípica suficientes durante dois ciclos de cultivo independentes. No entanto, a avaliação do DHE é dispendiosa, demorada e muitas vezes restrita a um número relativamente pequeno de características que podem ser influenciadas pelas condições ambientais. Isso exige a adoção de um sistema baseado em DNA, o que é endossado pela União Internacional para a Proteção de Novas Variedades de Plantas (UPOV). Além de seu papel complementar no teste DHE, os marcadores de DNA são um bom substituto de características morfológicas na definição de limites entre variedades independentes e essencialmente derivadas (Jamali et al., 2019). Contudo, para essa última aplicação se tornar útil e publicamente aceita, uma série de avaliações e validações devem ser realizadas. Em arroz, Liu et al. (2022) concluíram que as análises envolvendo tanto descritores fenotípicos quanto marcadores moleculares aumentarão a eficiência de testes de DHE.

Seleção Assistida por Marcadores (SAM)

A vantagem da SAM associado ao melhoramento clássico é que a seleção de plantas pode ser obtida a partir de qualquer tecido vegetal e realizadas em um estágio inicial de desenvolvimento, mesmo para caracteres que são expressos em plantas adultas, economizando tempo e recursos. Além disso, os marcadores moleculares não são afetados pelas condições de cultivo e/ou ambientais, tornando a SAM bastante confiável, quando comparada com a fenotipagem para a característica-alvo. Para a condução de SAM, são necessárias instalações, equipamentos e reagentes (o que envolve custos de implantação e manutenção), e pessoal qualificado. Adicionalmente, trabalhando com uma população de plantas em larga escala para investigação por vários marcadores, é necessário um sistema eficiente de extração de DNA e validação de marcadores (Sharma et al., 2021). O sucesso da SAM depende de vários fatores, incluindo a base genética dos caracteres, o número de indivíduos que podem ser analisados, a eficiência de seleção e a validação prévia no germoplasma que será avaliado. Em determinadas situações, o marcador desenvolvido pode não ser útil/funcional em populações com backgrounds genéticos diferentes dentro de um programa de melhoramento, demandando a identificação de novos marcadores para realizar a SAM (Gomes-Messias et al., 2022). SAM para caracteres de herança simples é muito útil nos programas de melhoramento, por propiciar agilidade ao desenvolvimento de cultivares (Kumawat et al., 2020). Para caracteres de herança mais complexa, como produtividade, existem severas

■ Melhoramento de precisão: aplicações e perspectivas na genética...

limitações na utilização eficiente de SAM, contudo, têm sido reportadas aplicações bem-sucedidas na melhoria de caracteres quantitativos, como a tolerância a estresses abióticos - seca, frio e salino (Ullah et al., 2022).

Caracteres relacionados a resistência a doenças e à qualidade de alguns produtos agrícolas são alguns exemplos importantes de aplicação rotineira de SAM. Yang et al. (2019), por exemplo, introgridiram o gene *Pi2*, que confere resistência durável à brusone do arroz, por meio da estratégia de retrocruzamento assistido por marcadores (RAM), em que tanto o gene foi selecionado na progênie (identificação da presença do gene via PCR por dois conjuntos de marcadores localizados nas extremidades do gene), quanto o background genético do genitor recorrente, pela genotipagem utilizando um chip de 6K SNPs. Singh et al. (2022) descreveram os benefícios do emprego do retrocruzamento assistido no Instituto Indiano de Pesquisa Agrícola (ICAR - Indian Agricultural Research Institute), que têm possibilitado o desenvolvimento de variedades comerciais incorporando genes de resistência/tolerância a diferentes caracteres, com redução de até 50% do tempo necessário via melhoramento clássico. Em outro exemplo, Liu et al. (2019) desenvolveram e validaram um marcador baseado em PCR para identificar variedades de arroz com baixo teor de amilose, a partir de um SNP (A/G) presente no gene *Waxy*. Para o feijão, o programa de melhoramento genético da Embrapa se baseia fortemente na SAM para a seleção de genótipos resistentes a diversas doenças e qualidade de grãos (Gomes-Messias et al., 2022). A operacionalização das ferramentas moleculares junto ao melhoramento genético de feijão também foi recentemente demonstrada pelo desenvolvimento de uma cultivar de feijão transgênica com resistência múltipla a viroses (Bean Golden Mosaic Virus - BGMV, Bean Common Mosaic Virus - BCMV e Cowpea Mild Mottle Virus – CPMMV) e reunindo caracteres agrônômicos favoráveis, que está em fase final de registro (Silva et al., 2022).

O procedimento para integrar vários genes ao mesmo tempo, ou QTLs em um único genótipo, é conhecido como piramidação. Essa estratégia é aplicada, por exemplo, na incorporação de diversos genes de resistência a doenças em uma única linhagem (Sharma et al., 2021). Embora ainda seja concebível utilizar o melhoramento tradicional, é impensável rastrear fenotipicamente cada planta para todos os caracteres que estão sendo piramidados (Hassan et al., 2021). Marcadores de DNA devem estar perto da variante causal e, dada a abundância no genoma, marcadores SNPs podem ser relacionados a praticamente todos os genes de espécies vegetais. Marcadores SNPs associados a caracteres específicos podem ser utilizados em larga escala por chips de DNA, mas também, dependendo do caso, podem ser convertidos em reações em que um gene por vez é caracterizado. Steele et al. (2018) desenvolveram e validaram, a partir de SNPs associados a caracteres de interesse, como resistência a doenças e qualidade de

grão, 39 KASP (*Kompetitive allele-specific PCR*) para uso na SAM em arroz. Linhagens de soja contendo genes de resistência para a ferrugem asiática piramidados tendem a conferir maior resistência que genótipos com um único gene (Panho et al., 2022).

Em suma, verifica-se que os marcadores utilizados na seleção assistida podem economizar tempo pela seleção em gerações de entressafra, sem avaliação agrônômica e, acima de tudo, são insubstituíveis para o manejo de recombinações, a fim de acumular alelos favoráveis o mais rápido possível em um único genótipo. Para explorar plenamente seu valor, novos esquemas de seleção recorrente ou construção de genótipos recorrentes devem ser planejados. Para que marcadores moleculares sejam relacionados a caracteres de interesse, estudos preliminares envolvendo genotipagens de populações segregantes ou coleções de germoplasma devem ser realizados. Essa etapa é dispendiosa e demorada, e ainda depende de fenotipagens feitas criteriosamente. Contudo, fenotipagens e genotipagens podem resultar no desenvolvimento de séries de marcadores que podem acelerar o programa de melhoramento. Outro ponto importante é qual a metodologia de análise deve ser utilizada. Populações segregantes fechadas, isto é, oriundas da utilização de dois ou mais parentais, formam a base de estudos de mapeamento de QTLs (Kumar et al., 2017). Outras estratégias têm sido utilizadas, como o mapeamento associativo, ou GWAS (*genome-wide association studies*) (Liu et al., 2020), *Machine Learning* e seleção genômica.

Estudos de Associação Genômica Ampla (Genome-Wide Association Studies – GWAS)

O estudo de associação genômica ampla (GWAS) é uma metodologia usada para detectar associações entre as variações genéticas e uma determinada característica de interesse. A metodologia de GWAS busca tirar o máximo de proveito de antigos eventos de recombinação, que são revelados pela análise do Desequilíbrio de Ligação (DL) ao longo do genoma para detectar a variação natural relacionada aos caracteres complexos em diversas culturas (Kumar et al., 2022). O conhecimento do DL é fundamental para determinar a densidade de marcadores necessária para os estudos de GWAS e, conseqüentemente, para identificar locos relacionados aos caracteres com uma resolução alta (Huang; Han, 2014). Diante disso, o sucesso do GWAS depende da resolução que determina a capacidade de detectar associações entre os marcadores genotipados e as variantes causais, requerendo densidade específica de marcadores consistentes com a extensão de LD no genoma. Somado a isso, tamanhos amostrais maiores têm impacto significativo no poder estatístico da análise e detecção de locos com

■ Melhoramento de precisão: aplicações e perspectivas na genética...

efeitos menores (Robinson et al., 2014). Outro ponto relevante diz respeito à adequação de modelos e metodologias estatísticas para a estimação dos efeitos dos locos. O desenvolvimento da estrutura de modelos mistos para GWAS reduziu drasticamente o número de falsos positivos em comparação com outros métodos (Wen et al., 2018) e, a partir disso, muitos métodos têm sido desenvolvidos desde então para aumentar a velocidade computacional ou melhorar o poder estatístico do GWAS (Uffelmann et al., 2021). Os dados genotípicos reunindo milhares de SNPs distribuídos em todo genoma, associados a uma população diversa, torna-se um recurso permanente, requerendo apenas que essa população seja fenotipada para outros caracteres de interesse, constituindo recurso valioso para a descoberta de genes.

Estudos de GWAS integrando novas ferramentas genômicas, fenômicas, biotecnológicas e aprimorados métodos computacionais têm ampliado o poder de detecção dos genes subjacentes a características complexas (Alseekh et al., 2021). Os estudos de GWAS foram inicialmente desenvolvidos no contexto da genética humana e levaram à identificação de inúmeras variantes genéticas associadas principalmente a doenças, possibilitando o desenvolvimento de novas abordagens terapêuticas e levando a avanços importantes, como a medicina personalizada, com a identificação de alvos-terapêuticos individualizados e desenvolvimento de novas medicações e estratégias de terapia genética, revertidos eventualmente em tratamentos clínicos (Rao et al., 2021). Em plantas, análises de GWAS investigando caracteres agrônômicos têm sido conduzidos para inúmeras espécies como a soja (*Glycine max*), arroz (*Oryza sativa*), feijão (*Phaseolus vulgaris*), milho (*Zea mays* L.), trigo (*Triticum aestivum* L.), dentre outras (Alseekh et al., 2021). Por meio da análise de GWAS, Wang et al. (2020) identificaram o gene *ZmVPP1*, que codifica uma pirofosfatase H (+) do tipo vacuolar e desenvolveram uma planta transgênica de milho tolerante à seca, que apresenta eficiência fotossintética aprimorada e desenvolvimento radicular mais robusto. Estudos de associação têm possibilitado a identificação de regiões associadas com muitas características agrônômicas e fisiológicas, e processos evolutivos, incluindo dias para a floração, arquitetura de plantas, resistência a doenças, fertilidade, qualidade do grão, tolerância a estresses abióticos, teor e rendimento de óleo, dentre outros (Valdisser et al., 2017).

Muitas vezes, o loco causal de uma característica de interesse não é amostrado na genotipagem ou não há resolução para detectá-lo, mas há locos genômicos candidatos identificados via GWAS que são detectados por estarem em equilíbrio de ligação parcial com tal loco causal. Nesse sentido, uma abordagem mais eficaz para transpor as limitações das análises individuais dos SNPs e aumentar a resolução de regiões genômicas candidatas é considerar os haplótipos para a análise de todo

o genoma (Qian et al., 2017). Os SNPs são bialélicos, com baixa informatividade e taxa mutacional e, a rigor, os alelos significativos frequentemente não representam as variantes causais, e sim os alelos raros que determinam os fenótipos extremos. Adicionalmente, quando os SNPs são analisados independentemente, grande número de testes são gerados, causando redução do poder estatístico. Contudo, na análise baseada em haplótipos o número de testes é reduzido, preservando o poder do teste, e a taxa de falso positivos pode ser controlada. Além disso, os blocos de haplótipos são geralmente “multialélicos”, mais informativos e podem, portanto, capturar melhor o DL com QTLs multialélicos; as interações epistáticas entre SNPs em um loco; as informações da história evolutiva e, por analisar uma série alélica existente em um determinado loco, a detecção do loco causal, do que um marcador individual (Bhat et al., 2021). Por exemplo, Hamblin e Jannink (2011) relataram que a abordagem do haplótipo aumentou o efeito alélico e a variação fenotípica explicada em 34% e 50%, respectivamente, em comparação com análise de SNPs individuais.

É importante destacar que uma das principais críticas à análise de GWAS é que os marcadores identificados como associados ao caráter explicam apenas uma fração da herdabilidade das características de herança complexa. Apesar das grandes variações existentes nas estimativas de herdabilidade, que não podem ser negligenciadas, mesmo que o GWAS não possa explicar tudo, a análise revela uma maneira de identificar associações relevantes. Existem muitos fatores que podem explicar essa porção da variação genética não explicada pelos marcadores, a qual varia muito em função da população analisada. Dentre esses fatores, destaca-se a presença das variantes raras que dependem de tamanhos e estratégias amostrais adequados para serem detectadas, bem como os seus efeitos estimados; a baixa resolução nas análises em função da relação entre o número de marcadores e a extensão do DL; o limitado número de variantes amostradas no genoma (principalmente as variantes raras de baixa frequência), com foco em SNPs, e geralmente omitindo outras classes como indels e variações estruturais pouco amostradas; efeito conjunto dos genes sendo negligenciado; os fatores ambientais como uma importante fonte de variação do fenótipo; os efeitos genéticos aditivos e não aditivos sobre a herdabilidade; o efeito das variações resultante de interações epistáticas entre genes; e o efeito das variações epigenética que controlam o fenótipo (Zhou et al., 2022). Devido a isso, o desenvolvimento de métodos aprimorados para a condução do GWAS tem sido contínuo nos últimos 25 anos. Além dos constantes avanços nos modelos de análise dos dados, conforme descrito anteriormente, a disponibilidade e união de sequências completas do genoma para múltiplos acessos de uma mesma espécie têm permitido a montagem de pangenomas, que estão se tornando uma nova

■ Melhoramento de precisão: aplicações e perspectivas na genética...

referência para análise de GWAS, que passa a ser denominada de PWAS (*Pangenome Wide Association Study*). As reduções de custos do NGS e os avanços nos algoritmos de montagem de sequências têm possibilitado a criação de vários genomas de referência, juntamente com um catálogo de todas as formas de variações genéticas em espécies de plantas com genomas grandes e complexos ou poliploides. Como consequência, inúmeras variantes genômicas, incluindo variações estruturais (SVs), que compreendem grandes deleções, inserções, rearranjos (intercâmbios), PAVs (variações de presença e ausência) e CNVs (variações no número de cópias) têm dado um novo impulso às análises de GWAS por possibilitar explorar variações genéticas adicionais que até então não estavam disponíveis para análise (Jayakodi et al., 2021). Recentemente, Zhou et al. (2022) demonstraram que a análise de grande número de variantes no genoma de tomate (gerados por meio da análise de 32 genomas de referência - pangenoma) proporcionou aumento da ordem de 24% na herdabilidade estimada. A inclusão de inúmeras variantes estruturais causais identificadas usando o pangenoma gráfico proporcionou tanto um aumento considerável na resolução do desequilíbrio de ligação, quanto o poder de identificar fatores genéticos subjacentes a características agronomicamente importantes, levando, por exemplo, à identificação de novas variações em dois genes que potencialmente contribuem para o teor de sólidos solúveis nos frutos. Esses avanços estão resultando em melhoria significativa na metodologia de GWAS para detecção mais ampla e robusta de associações que são de grande relevância para o melhoramento de plantas.

Seleção Genômica Ampla – SGA (*Genome-Wide Selection - GWS*)

A partir das técnicas de alto rendimento de genotipagem, a disponibilidade de grandes quantidades de SNPs por amostra nos permite explorar ao máximo a variância genética existente nas populações de melhoramento (Resende et al., 2017; Grattapaglia, 2022). Tal como os procedimentos já descritos neste capítulo, a seleção genômica também é possível graças ao Desequilíbrio de Ligação (DL) presente entre marcadores, isto é, falando de forma mais direta, a correlação que existe entre dois marcadores em uma população genotipada com marcadores SNP. Isso fará com que, mesmo não estando diretamente dentro de um QTL, muitos marcadores adjacentes ao QTL conseguirão testemunhar algo sobre a segregação dos genes relacionados com a expressão do caráter de interesse.

A partir da aplicação de técnicas de Seleção Genômica Ampla (SGA), é possível prever os “fenótipos” de testes de melhoramento experimentais, sem que esses testes tenham sequer sido montados. Isto é, apenas com base no DNA dos indivíduos que

hipoteticamente *iriam* para campo. É possível prever seu comportamento sem que eles tenham sido efetivamente plantados. Embora em um primeiro momento isso possa parecer algo irrealista, faz sentido se lembrarmos que provém dos genes uma parte importante da informação que irá culminar no fenótipo final. Manejando-se e/ou corrigindo-se adequadamente a fração ambiental do fenótipo (aliás, como se sabe, Fenótipo = Genótipo + Ambiente), os modelos de SGA irão demonstrar boas habilidades preditivas.

É importante destacar que a SGA, em termos de capacidade preditiva (ou o ranqueamento fidedigno dos melhores materiais genéticos) não necessariamente será superior à seleção fenotípica que é realizada em campo (Heffner et al., 2011). Isso acontece porque a SGA é geralmente aplicada como uma *seleção indireta ultra precoce*, em que se almeja alcançar a *seleção direta* (um “*benchmark*”) feita efetivamente no campo (veja preliminarmente a Figura 3-A). No entanto, ela pode ser sim mais vantajosa do que a seleção direta feita em campo, principalmente por cinco motivos: *i*) economia de tempo, afinal, a partir de propágulos vegetais iniciais já é possível se realizar seleção *precoce* genômica; *ii*) economia de recursos, mão de obra e insumos que seriam dispendidos em todo processo de implantação, mensuração fenotípica e colheita/transporte nos testes experimentais de melhoramento; e *iii*) possibilidade de avaliar maior número de materiais genéticos que eventualmente não iriam para o campo e, portanto, não seriam testados, por exemplo, ao invés de se levar 2.000 materiais genéticos para serem avaliados em campo, pode-se enviar 4.000 para serem genotipados e ter predito seus fenótipos genomicamente; *iv*) prever traços de difícil mensuração, como por exemplo volume de raiz em mandioca e volume da madeira em árvores altamente bifurcadas; *v*) correção de eventuais erros de pedigree na construção da matriz de parentesco (A), sendo essa informação recuperada por meio da matriz genômica (G), ou mesmo realizar concatenação das matrizes $A + G$ em uma “super” matriz chamada “ H ” (Legarra et al., 2014).

A capacidade preditiva pode ser entendida com um exemplo didático contendo quarenta materiais genéticos e uma capacidade preditiva $r_{\hat{g}g}$ de aproximadamente 50% (Figura 3). Alguns indivíduos fenotipicamente bons podem ficar de fora do crivo da seleção genômica e, mesmo assim, vislumbrar-se capacidades preditivas satisfatórias. Observa-se na Figura 3-B embaralhamento dos melhores e piores fenótipos quando preditos genomicamente, na Figura 3-C, como se relacionam os valores fenotípicos observados (ou os valores genéticos estimados a partir de médias ou por *Best Linear Unbiased Prediction* – BLUP – de um experimento). Verifica-se que o termo “seleção” da SGA pode ser confundido com “exclusão” genômica, que na verdade é o que acontece na maioria dos casos quando se usa a SGA para eliminar os piores indivíduos, e não para efetivamente selecionar os melhores.

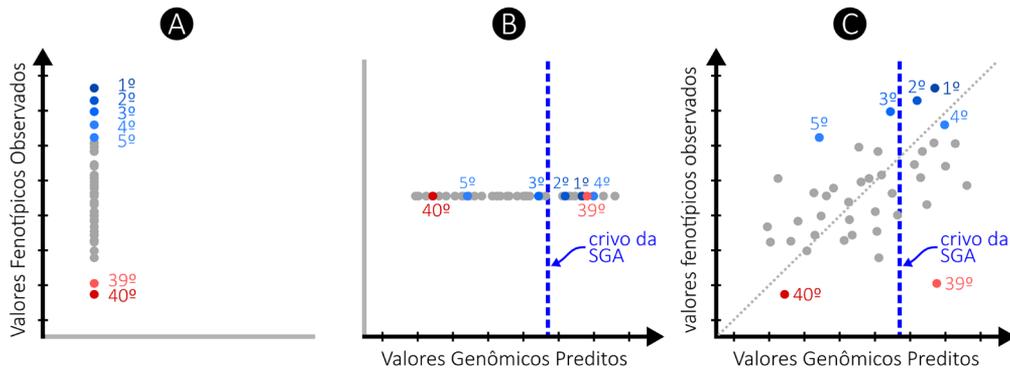


Figura 3. Relacionamento entre valores fenotípicos (médias ou BLUP de quarenta materiais genéticos) observados nos experimentos versus os valores preditos genomicamente. Na parte “A” é mostrado apenas o ranqueamento dito “verdadeiro”, a partir dos valores mensurados em campo. Na parte “B” são mostrados apenas a predição genômica, no entanto fazendo-se analogia com o ranqueamento dos materiais genéticos vistos em “A”. Na parte “C” é mostrado a relação entre as partes “A” e “B”. A linha azul pontilhada é o crivo de seleção da SGA, à direita os materiais genéticos selecionados, e à esquerda os descartados. A linha cinza pontilhada da parte “C” dá uma dimensão da capacidade preditiva do modelo SGA.

Muito se diz sobre o número adequado de marcadores a serem utilizados em um processo de seleção genômica, ou mesmo a quantidade ideal de indivíduos/materiais genéticos a serem fenotipados e genotipados (Werner et al., 2020; Merrick et al., 2022). A verdade é que não existe uma “receita de bolo” para esse processo. Isso irá depender de diversos fatores inerentes à cultura vegetal, à população que se deseja melhorar, ao objetivo do melhoramento (como, por exemplo, para desenvolvimentos de linhagens, híbridos e clones), e sobre o caractere fenotípico que se deseja melhorar (Silva et al., 2021). Entre algumas equações com a finalidade de planejar o programa melhoramento acoplado à seleção genômica, Resende et al. (2012) propuseram esta equação: $\hat{r}_{gg} = \sqrt{\left\{1 + \frac{4N_e L}{n_m} + \frac{2N_e L (4N_e L + n_m)}{n_m^2}\right\}^{-1}}$, em que: L é o tamanho do genoma da espécie; n_m é o número de marcadores do tipo SNP; h^2 é o coeficiente de herdabilidades do caractere fenotípico (podendo ser a herdabilidade no sentido amplo ou no sentido restrito); N é o tamanho real da população; N_e é o tamanho efetivo da população. O resultado irá entregar a capacidade preditiva (ou acurácia esperada) da aplicação de modelos de seleção genômica. É recomendável que antes de iniciar qualquer processo de seleção genômica, se planeje, e tenha uma dimensão dos possíveis resultados, caso contrário, ao invés de economia de recursos e tempo, pode-se ter um grande prejuízo!

Nessa altura do capítulo, é importante destacar que a validação estatístico-matemática dos modelos de Seleção Genômica Ampla (SGA) é uma tarefa necessária para avaliar a capacidade preditiva dos modelos. Dentre os métodos disponíveis para computar a acurácia

preditiva, a correlação simples entre os valores fenotípicos observados e os valores preditos pelo modelo é uma medida direta da acurácia preditiva do modelo e é considerada uma forma confiável de avaliar o desempenho do modelo de SGA. A acurácia também pode ser calculada ponderando-se com base nas herdabilidades dos caracteres fenotípicos e/ou genômicos, com a premissa de corrigir eventuais efeitos de encolhimento da predição (efeito de *shrinkage*) (Müller et al., 2015). No entanto, é preciso ter cuidado com a inclusão dessas quantidades nas equações de acurácia, pois isso pode causar superestimação da acurácia de SGA em caracteres de baixa herdabilidade ou subestimação da capacidade preditiva de caracteres com alta herdabilidade. Para contornar essa questão, uma opção razoável é a utilização da correlação de Pearson, uma conhecida medida de correlação, ou mesmo correlações de Spearman (entre rankings genotípicos).

Sobre o tamanho amostral da SGA, em geral, tem-se a experiência de que bem mais do que os 1.000 indivíduos indicados de maneira estereotipada na literatura são necessários para ajustar bons modelos de SGA. Além disso, após observar muitos esforços ajustando-se diversos tipos/abordagens de modelos preditivos (como por exemplo os Bayesianos – Bayes A, B, C, LASSO – e os via Inteligência Artificial ou *Machine Learning*), há pouco incremento de capacidade preditiva em cima dos métodos GBLUP ou RRBLUP inicialmente descritos por Meuwissen et al. (2001). Na verdade, há situações em que se pode sim atingir melhores ajustes a partir de métodos mais rebuscados frente aos clássicos GBLUP/RRBLUP aqui citados (Montesinos-López et al., 2021). Contudo, é fundamental destacar aqui que outros esforços, como o manejo das populações de melhoramento e estratégias sobre como alimentar os modelos e validá-los, são vitais no sucesso da SGA. Além disso, a adequada exploração dos componentes aditivos e não aditivos dos caracteres fenotípicos pretendidos, como também a maneira de adequar a SGA à fase específica do programa de melhoramento, são fatores que geralmente proporcionarão maiores benefícios comparados à disputa de maiores capacidades preditivas entre métodos frequentistas × bayesianos × inteligência artificial.

Na Figura 4 é apresentado o modelo linear misto básico $y = X\beta + Zg + e$, como ilustrativo das inúmeras metodologias que poderiam ser aplicadas na predição genômica. No caso, y é o vetor de dados fenotípicos, β é o vetor de efeitos fixos (como por exemplo: repetições/blocos experimentais, locais, medidas repetidas ao longo do tempo, entre outras); g é o vetor de efeitos aleatórios genéticos (os materiais genéticos, podendo ser linhagens, híbridos, entre outros); e é o vetor aleatório de resíduos; X e Z são matrizes de incidência sobre os efeitos fixos e aleatórios, respectivamente. Não é foco deste capítulo discutir quando atribuir certos efeitos como de natureza fixa ou aleatória, mas é consenso que os materiais genéticos devam entrar como natureza aleatória para viabilizar a execução dos modelos mistos com a finalidade de seleção genômica.

■ Melhoramento de precisão: aplicações e perspectivas na genética...

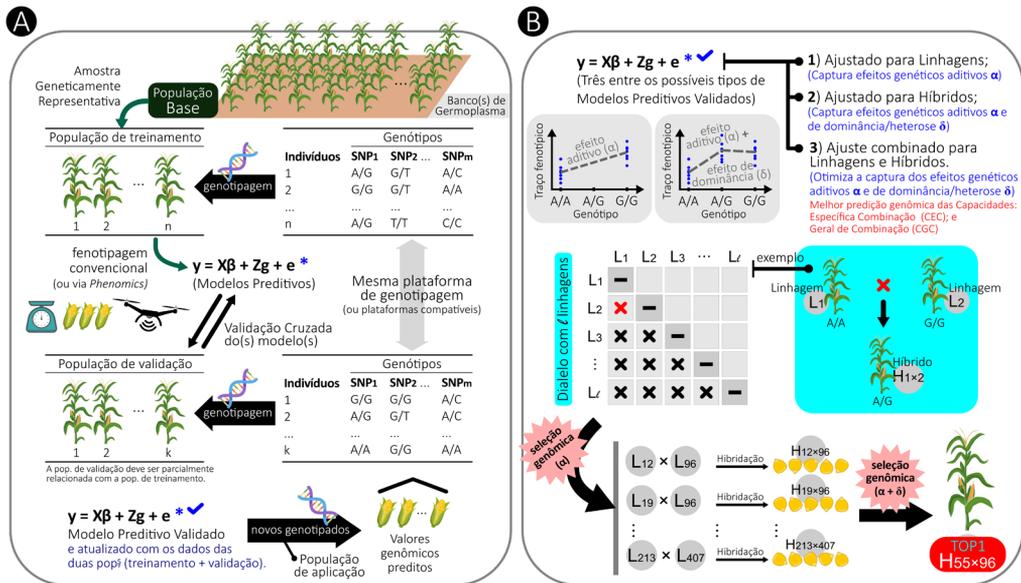


Figura 4. Esquema reduzido de um processo de Seleção Genômica Ampla (SGA), utilizando o milho como exemplo. Na parte A da figura é mostrado o processo de ajuste e validação de modelos genômicos preditivos. Na parte B da figura é mostrado possíveis esquemas de utilização, tanto para predição de linhagens endogâmicas, como na predição de híbridos baseando-se nas melhores linhagens ou em modelos validados com informações de híbridos.

No exemplo de procedimento mostrado na Figura 4, vê-se a SGA sendo utilizada com duas finalidades, uma na formação de linhagens endogâmicas melhoradas (L_1, L_2, \dots, L_r), e outra na formação de híbridos heteróticos melhorados ($H_{1 \times 2}, H_{1 \times 2'}, \dots$). Em ambos os casos, tais materiais genéticos poderão vir a se tornar, caso desejado pelo melhorista, cultivares registradas, obviamente após devidos testes em campo experimental, a partir, por exemplo, de ensaios em VCU (Valor de Cultivo e Uso). Um modelo preditivo genômico, tal como todo modelo preditivo, irá prever aquilo que “conhece”. Se você o alimentar com dados de testes de progênes, ele irá te entregar valores preditos de teste de progênes de milho, bem como se você o alimentar com valores de testes clonais de eucalipto, da mesma forma ele irá retornar valores preditivos compatível com os fenótipos crescidos em um teste clonal. Portanto, é preciso ter muita cautela com a população base que se irá utilizar para ajustar tais modelos.

A necessidade de um relacionamento genético parcial entre as populações de treinamento, validação e aplicação é de fato um ponto negativo da SGA, afinal o DL, que é a força motriz da genômica, se perde e se cria facilmente entre populações distintas, ou mesmo após o avanço de algumas gerações de uma mesma população (Werner et al.,

2020). No entanto, o ideal é manejar os dados de ajuste de modelo conforme será o objetivo do programa. O ideal é inicialmente mapear toda base genética do programa, inserir novos materiais ou, principalmente, descartar os materiais pouco desejáveis (isto irá eliminar ruídos desnecessários da análise). É oportuno mencionar que os materiais genéticos desenvolvidos com base em SGA necessariamente serão relacionados com a base genética inicial que se possui (Grattapaglia, 2022). E, na verdade, isso é razoável quando se pensa que empresas têm, ou deveriam ter, bases genéticas (bancos de germoplasma) bem definidos. Dificilmente um bom modelo de SGA irá prever bem materiais genéticos de outras bases genéticas (como de outras empresas, países e regiões).

O mesmo acontece ao se utilizar apenas um ou poucos ambientes no modelo preditivo genômico, pois ele obviamente estará apto a apenas prever o comportamento dos materiais genéticos para aqueles poucos ambientes que ele conhece. Essa falta de representatividade nos dados de entrada dos modelos SGA, somada à validação com uma partição dos próprios dados, poderá fatalmente dar falsa sensação de que o modelo é bom (com altas, porém viciadas, capacidades preditivas), primeiro porque a população de validação será completamente relacionada com a população de treinamento e, segundo, porque os dados fenotípicos cresceram em um ou poucos ambientes. Duas estratégias podem ser tomadas para contornar essa questão: *i*) criar um modelo genômico multiambiente capaz de retornar previsões de indivíduos com alta *estabilidade*, isto é, o valor esperado do material genético é bom independentemente do ambiente; *ii*) utilizar modelos que incorporem interações Genótipos × Ambientes (G×A) – há algumas classes de modelos dessa natureza, porém é razoável citar aqui aqueles que lidam com múltiplos dados ambientais no escopo da Ambientômica (Resende et al., 2021; Costa-Neto et al., 2023), que são capazes de retornar previsões de indivíduos com alta estabilidade e *adaptabilidade* a diferentes locais. Esses modelos podem, portanto, prever materiais melhorados em escala sítio-específica.

Nesse contexto, é importante escolher os caracteres certos para alimentar o modelo de forma operacional ou industrial. Alguns caracteres são mais fáceis de medir do que outros, mas isso pode levar a baixa correlação genética com o caractere real de interesse, o que é um problema sério que muitas vezes é ignorado. Por exemplo, os caracteres fenotípicos de testes genéticos de melhoramento (i.e., progênies, híbridos, clonais entre outros) podem não se correlacionar bem com o desempenho real em campo. É importante alimentar o modelo SGA com dados operacionais, mas muitas vezes há poucos dados disponíveis para genótipos comerciais. Nesse caso, integrar dados de testes e comerciais pode ajudar a computar a correlação genética entre os dois tipos de dados. Dessa forma, é possível lidar bem com o problema e obter melhores resultados na produção de grãos, na produção florestal, horticultura, fruticultura e em outros setores.

■ Melhoria de precisão: aplicações e perspectivas na genética...

Em qualquer programa de melhoramento que englobe a seleção genômica, diversos tipos de caracteres fenotípicos serão melhorados, de preferência de forma simultânea. Dados de fenotipagem em larga escala também podem compor os modelos de SGA, como os coletados por sensores embarcados em drones ou aqueles preditos via *Near Infrared Spectroscopy* (NIRS). Mais detalhes podem ser observados no Capítulo 9, sobre fenômica, desta obra. Cada caractere fenotípico tem sua particularidade, bem como natureza genética, herança e assim por diante. Os caracteres irão se relacionar diretamente com o tipo de material genético a ser trabalhado, bem com o objetivo do programa. Mas uma coisa é certa: quanto maior forem as herdabilidades do caractere (seja aditiva no sentido restrito – h_a^2 , como a total no sentido amplo – h^2) melhor o modelo de SGA irá funcionar e entregar bons resultados (Zhang et al., 2017). Num programa de seleção recorrente (SRR), nas fases iniciais, a variabilidade genética tende a ser maior e, por serem muitos materiais avaliados, provavelmente com baixa repetição, as herdabilidades tendem a ser melhores. O paradoxo é que mais para o final do programa, embora a base genética esteja afinada (após alguns ciclos de seleção), a quantidade de materiais genéticos é bem melhor e, portanto, estão mais repetidos experimentalmente falando.

Logo, os modelos de SGA devem ser manejados levando-se em consideração que as herdabilidades devem ser, na medida do possível, elevadas ao máximo para se otimizar a eficiência da seleção. Isso será alcançado, em geral, de duas maneiras: *i*) aumento da variação genética; e/ou *ii*) redução da variância ambiental por meio de melhor controle residual ou aumento no número de repetições. Nas fases iniciais do programa (seja de autógamias ou alógamas, anuais ou perenes), em geral, preza-se pelos efeitos aditivos (α), pois se está falando de algumas etapas de cruzamento/recombinação e seleção (efeitos não aditivos se perdem mais facilmente nessas etapas). No entanto, nas fases finais dos programas, os efeitos não aditivos (δ) são também desejados, uma vez que os cultivares gerados em geral são híbridos com algum grau de heterose (que é um fenômeno de dominância) e materiais geneticamente mais uniformes (Labroo et al., 2021). Nesse ponto, lançando mão da Figura 4-B, pode-se utilizar um modelo para predição genômica aditiva, em que se conseguirá predizer bem indivíduos segregantes em etapas iniciais do programa, mesmo com o objetivo sendo a obtenção de linhagens/linhas puras. Pode-se também utilizar modelos que comportem aditividade e dominância utilizando apenas os materiais em final de programa. Ou, de forma moderna, integrar os dados de todas as etapas, maximizando-se e interconectando todo processo de seleção.

Essa mistura de diferentes fontes de dados fenotípicos fornecerá inúmeras vantagens, começando pela utilização de diversos ambientes e, com isso, predizer o comportamento de estabilidade e adaptabilidade dos genótipos. Além disso, como se está falando de uma mesma base genética (isto é, fracionando-se os modelos SGA ou usando um único

com todos os dados), o tamanho efetivo da população – N_e – não deverá se alterar drasticamente, porém, o tamanho da população total (N) aumenta, aproveita-se o melhor dos mundos: alta variabilidade das populações iniciais do programa e maior quantidade de repetições experimentais nos estágios finais do programa. Desse modo, haverá efeito direto no aumento da capacidade preditiva do modelo.

A utilização de abordagens multiômicas, muitas vezes combinadas com a genômica, também é uma ferramenta importante para a predição genética de plantas. Essas abordagens podem combinar dados transcriptômicos, proteômicos, metabolômicos, os próprios genômicos e muitas outras ômicas, para conhecer preditivamente características dos genótipos em estudo. A premissa é que a agregação de informação em mais alto nível de precisão com o fenótipo final pode melhorar a capacidade preditiva dos modelos, como na utilização de marcadores exômicos, que são aqueles que efetivamente traduzem-se em proteínas (Hashmi et al., 2015). Com as demais estruturas moleculares o raciocínio é o mesmo. Por exemplo, a predição de compostos de sabor (açúcares, ácidos e voláteis) em mirtilo e tomate, com base em metabolômica, mostram resultados bem promissores (Colantonio et al., 2022), ou com genômica no paladar do café (Ferrão et al., 2023). Outro exemplo são as abordagens transcriptômicas + proteômicas + metabolômicas + genômicas funcionais para o estudo do estresse abiótico em hortaliças (Zhuang et al., 2014). A análise conjunta desses diferentes níveis de informação aumenta as acurácias preditivas dos modelos, sendo capazes até de prever características muito afetadas pelo ambiente, ou até mesmo subjetivas, como o paladar de determinados produtos agrícolas.

Inteligência Artificial e Machine Learning na Análise Genômica

A técnica de GWAS já identificou milhares de variantes genéticas relacionadas a caracteres de interesse à agricultura. No entanto, o potencial do GWAS ainda é pouco explorado devido a limitações metodológicas relacionadas com a presença de epistasia, SNPs associados a pequenos efeitos e a não distinção de variantes causais de outros SNPs associados por desequilíbrio de ligação. Variantes identificadas pelo GWAS representam apenas uma proporção modesta da herdabilidade de características complexas. Uma análise GWAS típica pode envolver 200.000–2.000.000 SNPs, e o grande volume de testes de significância para SNPs individuais pode levar a que eventos aleatórios sejam falsamente significativos, e se o efeito de SNPs causais individuais for muito pequeno, eles podem não passar dos limites de significância mais rigorosos. Essas limitações podem ser contornadas com a aplicação de métodos de *machine learning* em GWAS (Sun et al., 2021).

■ Melhoria de precisão: aplicações e perspectivas na genética...

Machine Learning (ML) é baseado no desenvolvimento de modelos algorítmicos que podem aprender com a experiência a identificar associações genótipo:fenótipo e é principalmente utilizado para interpretar grande volume de dados genômicos (Sun et al., 2021). Aplicações recentes de ML no melhoramento de plantas incluem redução de dimensionalidade de dados, inferência sobre rotas de regulação de genes, descoberta e priorização de genes, análise de fenômica de plantas e previsão genômica de fenótipos de plantas. Adicionalmente, utilizando *big data*, o ML é capaz de modelar as relações complexas de dados genotípicos, fenotípicos e ambientais (Yan; Wang, 2023). Cada algoritmo de ML pode aprender o padrão de um conjunto de dados de treinamento de uma maneira específica e prever o desempenho de uma variável de interesse (produtividade, por exemplo) no conjunto de dados de teste. *Random Forest* (RF) é um dos algoritmos de ML mais amplamente usados e pode prever o resultado de determinada variável em estudo utilizando os resultados médios de árvores de decisão idênticas obtidas das amostras de bootstrap do conjunto de dados de treinamento. *Support vector regression* (SVR), como a forma de regressão da *support vector machine*, fornece diferentes conjuntos de hiperplanos para selecionar a melhor linha de regressão com o mínimo de erros possível no modelo. Embora todos os algoritmos de ML tenham revolucionado os métodos de análise de *big data*, estudos recentes mostraram que o uso individual de algoritmos de ML pode estar sujeito a superajuste e aumento da taxa de falsos positivos (Yoosefzadeh-Najafabadi et al., 2022).

As ferramentas de predição genômica são utilizáveis em programas de melhoramento genético com base em métodos estatísticos, como o GBLUP (*genomic best linear unbiased prediction*), o qual tem sido amplamente utilizado em seleção genômica (SGA). Apesar do sucesso dos métodos lineares em SGA, eles podem enfrentar desafios devido à alta dimensionalidade dos dados do marcador versus o número de indivíduos e à presença de relações complexas difíceis de explicar. Para melhorar os modelos lineares em SGA, tem havido aumento no uso de métodos não lineares, como modelos ML e *deep learning* (DL), para prever fenótipos de plantas. No entanto, essas ferramentas não são projetadas para capturar relacionamentos não lineares em conjuntos de dados multidimensionais. Os algoritmos de ML têm o potencial de superar a precisão de previsão das ferramentas atuais usadas para previsão de genótipo para fenótipo, devido à sua capacidade de extrair recursos de dados de forma autônoma. Por exemplo, a *Random Forest* do método ML pode capturar padrões em dados de alta dimensão para fornecer previsões precisas e pode levar em conta efeitos não aditivos. Seu uso como modelo para seleção genômica também tem demonstrado desempenho superior em comparação a modelos lineares como *Bayesian Least Absolute Shrinkage and Selection Operator* (LASSO) e *Ridge Regression* BLUP (RR-BLUP), dependendo da arquitetura

genética da característica estudada. Outros modelos de ML que mostraram potencial para seleção genômica incluem redes neurais convolucionais e redes neurais profundas (Danilevicz et al., 2022). Mais detalhes e aplicações sobre a Inteligência Artificial (I.A.) na análise genômica podem ser acompanhados no Capítulo 5 deste livro.

Considerações Finais

Mediante a integração das modernas tecnologias de análise genômica e do melhoramento tradicional de plantas, novas abordagens estão sendo propostas para enfrentar os desafios agrícolas globais, como segurança alimentar, mudança climática e agricultura sustentável. Por meio dos avanços na análise genômica tem sido possível implementar, de forma eficiente e crescente, técnicas de biologia molecular, métodos eficazes de melhoramento (como a seleção genômica) e a engenharia genética, acelerando o processo de melhoramento de plantas de maneira segura, sustentável e precisa. Consequentemente, isso tem permitido o desenvolvimento de novas cultivares com características desejáveis e superiores, com maior potencial de rendimento, tolerância a estresses bióticos e abióticos e menor dependência de insumos químicos. Além disso, o melhoramento molecular de plantas tem auxiliado na preservação e utilização eficiente dos recursos genéticos vegetais, pois, pela identificação e caracterização da diversidade genética, é possível propor estratégias eficientes de preservação, conservação e manipulação dessa diversidade, garantindo maior sustentabilidade aos programas de melhoramento por meio da ampliação da base genética, além de possibilitar selecionar e introduzir alelos favoráveis para fenótipos desejáveis. Por fim, tecnologias de edição de genoma, como CRISPR (tecnologia detalhada no Capítulo 12 deste livro), têm possibilitado aos cientistas fazerem alterações precisas e controladas no genoma das plantas, sem a introdução de genes de espécies não relacionadas, acelerando o processo de melhoramento genético.

Entre as perspectivas de avanço para a área da genômica, pode-se citar: *i*) contínuo aprimoramento das ferramentas moleculares e redução de custos de genotipagem, ampliando seu uso para maior número de espécies trabalhadas por programas de melhoramento genético; *ii*) utilização de quantitativos cada vez maiores de marcadores genômicos nas análises de melhoramento, possivelmente na casa das centenas de milhares de marcas SNP; *iii*) incremento do uso da tecnologia de edição de genoma, como CRISPR, para realizar alterações precisas e controladas no genoma das plantas e, consequentemente, acelerar o processo de desenvolvimento de plantas com fenótipo melhorado; *iv*) aplicação da genômica para melhorar características com impacto econômico, como a produção de biocombustíveis, a geração de

■ Melhoria de precisão: aplicações e perspectivas na genética...

alimentos nutracêuticos, que proporcionam benefícios médicos e de saúde, e o desenvolvimento de processos mais sustentáveis e econômicos de produção, como por exemplo, cultivares com maior resistência a doenças obtidas por seleção assistida; e v) continuidade da utilização de técnicas de inteligência artificial e aprendizado de máquina na análise genômica, visando automatizar e otimizar o processo de análise de grandes conjuntos de dados, identificando padrões e relações complexas entre os genes e os fenótipos, que são informações valiosas para o desenvolvimento de estratégias de melhoramento genético mais eficazes; vi) incorporação cada vez maior de dados via fenotipagem em larga-escala (como os obtidos por drones e NIRS) nos modelos de seleção genômica e, também, dados de ambientômica. Tais perspectivas prometem impulsionar ainda mais a área da genômica vegetal, e o contínuo progresso tecnológico abre caminho para um futuro promissor das aplicações do melhoramento molecular.

Referências

ALSEEKH, S.; KOSTOVA, D.; BULUT, M.; FERNIE, A. R. Genome-wide association studies: assessing trait characteristics in model and crop plants. **Cellular and Molecular Life Sciences**, v. 78, p. 5743–5754, July 2021. DOI: <https://doi.org/10.1007/s00018-021-03868-w>.

ANILKUMAR, C.; SAH, R. P.; AZHARUDHEEN, T. P. M.; BEHERA, S.; SING, N.; PRAKASH, N. R.; SUNITHA, N. C.; DEVANNA, B. N.; MARNDI, B. C.; PATRA, B. C.; NAIR, S. K. Understanding complex genetic architecture of rice grain weight through QTL-meta analysis and candidate gene identification. **Scientific Reports**, v. 12, 13832, 2022. DOI: <https://doi.org/10.1038/s41598-022-17402-w>.

BHAT, J. A.; YU, D.; BOHRA, A.; GANIE, S. A.; VARSHNEY, R. K. Features and applications of haplotypes in crop breeding. **Communications Biology**, v. 4, 1266, Nov. 2021. DOI: <https://doi.org/10.1038/s42003-021-02782-y>.

CIVAN, P.; RINCENT, R.; DANGUY-DES-DESERTS, A.; ELSEIN, J. M.; BOUCHET, S. Population genomics along with quantitative genetics provides a more efficient valorization of crop plant genetic diversity in breeding and pre-breeding programs. In: RAJORA, O. P. (ed.). **Population genomics: crop plants, population genomics**. Cham: Springer, 2021. DOI: https://doi.org/10.1007/13836_2021_97.

COLANTONIO, V.; FERRÃO, L. F. V.; TIEMAN, D. M.; BLIZNYUK, N.; SIMS, C.; KLEE, H. J.; MUNOZ, P.; RESENDE JUNIOR, M. F. R. Metabolomic selection for enhanced fruit flavor. **Proceedings of the National Academy of Sciences**, v. 119, e2115865119, Dec. 2022. DOI: <https://doi.org/10.1073/pnas.2115865119>.

COSTA-NETO, G.; CRESPO-HERRERA, L.; FRADGLEY, N.; GARDNER, K.; BENTLEY, A. R.; DREISIGACKER, S.; FRITSCHÉ-NETO, R.; MONTESINOS-LÓPEZ, O. A.; CROSSA, J. Envirom-wide associations enhance multi-year genome-based prediction of historical wheat breeding data. **G3 Genes|Genomes|Genetics**, v. 13, jkac313, Feb. 2023. DOI: <https://doi.org/10.1093/g3journal/jkac313>.

DANILEVICZ, M. F.; GILL, M.; ANDERSON, R.; BATLEY, J.; BENNAMOUN, M.; BAYER, P. E.; EDWARDS, D. Plant genotype to phenotype prediction using machine learning. **Frontiers in Genetics**, v. 13, 822173, May 2022. DOI: <https://doi.org/10.3389/fgene.2022.822173>.

ELSHIRE, R. J.; GLAUBITZ, J. C.; SUN, Q.; POLAND, J. A.; KAWAMOTO, K.; BUCKER, E. S.; MITCHELL, S. E. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. **PLoS ONE**, v. 6, e19379, May 2011. DOI: <https://doi.org/10.1371/journal.pone.0019379>.

FERRÃO, M. A. G.; FONSECA, A. F. A.; VOLPI, P. S.; SOUZA, L. C.; COMÉRIO, M.; VERDIN FILHO, A. C. et al. Genomic-assisted breeding for climate-smart coffee. **The Plant Genome**, e20321, 2023. DOI: <https://doi.org/10.1002/tpg2.20321>.

GOMES-MESSIAS, L. M.; VIANELLO, R. P.; MONTEIRO-JÚNIOR, J. P.; RODRIGUES, L. A.; MOTA, A. P. S.; PEREIRA, H. S.; MELO, L. C.; RAATZ, B.; SOUZA, T. L. P. O. Molecular characterization of parental lines and validation of SNP markers for anthracnose and angular leaf spot in common bean. **Euphytica**, v. 218, 49, Apr. 2022. DOI: <https://doi.org/10.1007/s10681-022-03002-2>.

GRATTAPAGLIA, D. Twelve years into genomic selection in forest trees: climbing the slope of enlightenment of marker assisted tree breeding. **Forests**, v. 13, 1554, Sept. 2022. DOI: <https://doi.org/10.3390/f13101554>.

HAMBLIN, M. T.; JANNINK, J. L. Factors affecting the power of haplotype markers in association studies. **The Plant Genome**, v. 4, n. 2, July 2011. DOI: <https://doi.org/10.3835/plantgenome2011.03.00082011>.

HASHMI, U.; SHAFQAT, S.; KHAN, F.; MAJID, M.; HUSSAIN, H.; KAZI, A. G.; JOHN, R.; AHMAD, P. Plant exomics: concepts, applications and methodologies in crop improvement. **Plant Signaling & Behavior**, v. 10, e976152, Jan. 2015. DOI: <https://doi.org/10.4161/15592324.2014.976152>.

HASSAN, M. A.; YANG, M.; RASHEED, A.; TIAN, X.; REYNOLDS, M.; XIA, X.; XIAO, Y.; HE, Z. Quantifying senescence in bread wheat using multispectral imaging from an unmanned aerial vehicle and QTL mapping. **Plant Physiology**, v. 187, n. 4, p. 2623-2636, Dec. 2021. DOI: <https://doi.org/10.1093/plphys/kiab431>.

HE, C.; HOLME, J.; ANTHONY, J. SNP genotyping: the KASP assay. **Methods in Molecular Biology**, v. 1145, p. 75–86, Jan. 2014. DOI: https://doi.org/10.1007/978-1-4939-0446-4_7.

HEFFNER, E. L.; JANNINK, J. L.; IWATA, H.; SOUZA, E.; SORRELLS, M. E. Genomic selection accuracy for grain quality traits in biparental wheat populations. **Crop Science**, v. 51, n. 6, p. 2597-2606, Nov. 2011. DOI: <https://doi.org/10.2135/cropsci2011.05.0253>.

HU, T.; CHITNIS, N.; MONOS, D.; DINH, A. Next-generation sequencing technologies: an overview. **Human Immunology**, v. 82, n. 11, p. 801-811, Nov. 2021. DOI: <https://doi.org/10.1016/j.humimm.2021.02.012>.

HUANG, X. H.; HAN, B. Natural variations and genome-wide association studies in crop plants. **Annual Review of Plant Biology**, v. 65, p. 531-551, Apr. 2014. DOI: <https://doi.org/10.1146/annurev-arplant-050213-035715>.

HYTEN, D. L.; CHOI, I. Y.; SONG, Q.; SPECHT, J. E.; CARTER, T. E.; SHOEMAKER, R. C.; HWANG, E. Y.; MATUKUMALLI, L. K.; CREGAN, P. B. A high density integrated genetic linkage map of soybean and the development of a 1536 Universal Soy Linkage Panel for quantitative trait locus map-ping. **Crop Science**, v. 50, n. 3, p. 960–968, May-June, 2010. DOI: <https://doi.org/10.2135/cropsci2009.06.0360>.

JAMALI, S. H.; COCKRAM, J.; HICKEY, L.T. Insights into deployment of DNA markers in plant variety protection and registration. **Theoretical and Applied Genetics**, v. 132, p. 1911–1929, May 2019. DOI: <https://doi.org/10.1007/s00122-019-03348-7>.

JAYAKODI, M.; SCHRIEBER, M.; STEIN, N.; MASCHER, M. Building pan-genome infrastructures for crop plants and their use in association genetics. **DNA Research**, v. 28, dsaa030, Feb. 2021. DOI: <https://doi.org/10.1093/dnares/dsaa030>.

KUMAR, B.; RAKSHIT, S.; KUMAR, S.; SINGH, B. K.; LAHKAR, C.; JHA, A. K.; KUMAR, K.; KUMAR, P.; CHOUDHARY, M.; SINGH, S. B.; AMALRAJ, J. J.; PRAKASH, B.; KHULBE, R.; KAMBOJ, M. C.; CHIRRAVURI, N. N.; HOSSAIN, F. Genetic diversity, population structure and linkage disequilibrium analyses in tropical maize using genotyping by sequencing. **Plants**, v. 11, 799, Mar. 2022. DOI: <https://doi.org/10.3390/plants11060799>.

■ Melhoria de precisão: aplicações e perspectivas na genética...

KUMAR, J.; GUPTA, D. S.; GUPTA, S.; DUBEY, S.; GUPTA, P.; KUMAR, S. Quantitative trait loci from identification to exploitation for crop improvement. **Plant Cell Reports**, v. 36, p. 1187–1213, 2017. DOI: <https://doi.org/10.1007/s00299-017-2127-y>.

KUMAWAT, G.; KUMAWAT, C. K.; CHANDRA, K.; PANDEY, S.; CHAND, S.; MISHRA, U. N.; LENKA, D.; SHARMA, R. Insights into marker assisted selection and its applications in plant breeding. In: ABDURAKHMONOV, I. Y. (ed.). **Plant breeding: current and future views**. London: IntechOpen, 2020. DOI: <https://www.doi.org/10.5772/intechopen.95004>.

LABROO, M. R.; STUDER, A. J.; RUTKOSKI, J. E. Heterosis and hybrid crop breeding: a multidisciplinary review. **Frontiers in Genetics**, v. 12, p. 643761, 2021. DOI: <https://doi.org/10.3389/fgene.2021.643761>.

LEGARRA, A.; CHRISTENSEN, O. F.; AGUILAR, I.; MISZTAL, I. Single step, a general approach for genomic selection. **Livestock Science**, v. 166, p. 54-65, Aug. 2014. DOI: <https://doi.org/10.1016/j.livsci.2014.04.029>.

LI, J. Y.; WANG, J.; ZEIGLER, R. The 3,000 rice genomes project: new opportunities and challenges for future rice research. **GigaScience**, v. 3, n.1, p. 1-3, Dec. 2014. DOI: <https://doi.org/10.1186/2047-217X-3-8>.

LIN, M.; FICKE, A.; DIESETH, J. A.; LILLEMO, M. Genome-wide association mapping of septoria nodorum blotch resistance in Nordic winter and spring wheat collections. **Theoretical and Applied Genetics**, v. 135, p. 4169–4182, Sept. 2022. DOI: <https://doi.org/10.1007/s00122-022-04210-z>.

LIU, H.; RAO, D.; GUO, T.; GANGURDE, S. S.; HONG, Y.; CHEN, M.; HUANG, Z.; JIANG, Y.; XU, Z.; CHEN, Z. Whole genome sequencing and morphological trait-based evaluation of UPOV option 2 for DUS testing in rice. **Frontiers in Genetics**, v. 13, 945015, Aug. 2022. DOI: <https://doi.org/10.3389/fgene.2022.945015>.

LIU, S.; ZHONG, H.; MENG, X.; SUN, T.; LI, Y.; PINSON, S. R. M.; CHANG, S. K. C.; PENG, Z. Genome-wide association studies of ionomic and agronomic traits in USDA mini core collection of rice and comparative analyses of different mapping methods. **BMC Plant Biology**, v. 20, 441, Sept. 2020. DOI: <https://doi.org/10.1186/s12870-020-02603-0>.

LIU, Y.; ZHANG, A.; WANG, F.; WANG, J.; BI, J.; KONG, D.; ZHANG, F.; LUO, L.; LIU, G.; YU, X. Development and validation of a PCR-based functional marker system for identifying the low amylose content-associated gene *Wx^{hp}* in rice. **Breeding Science**, v. 69, n. 4, p. 702-706, Dec. 2019. DOI: <https://doi.org/10.1270/jsbbs.19043>.

LV, D.; ZHANG, C.; YV, R.; YAO, J.; WU, J.; SONG, X.; JIAN, J.; SONG, P.; ZHANG, Z.; HAN, D.; SUN, D. Utilization of a wheat50K SNP Microarray-derived high-density genetic map for QTL mapping of plant height and grain traits in wheat. **Plants**, v. 10, 1167, June 2021. DOI: <https://doi.org/10.3390/plants10061167>.

MERRICK, L. F.; HERR, A. W.; SANDHU, K. S.; LOZADA, D. N.; CARTER, A. H. Optimizing plant breeding programs for genomic selection. **Agronomy**, v. 12, 714, Mar. 2022. DOI: <https://doi.org/10.3390/agronomy12030714>.

MEUWISSEN, T. H. E.; 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, Apr. 2001. DOI: <https://doi.org/10.1093/genetics/157.4.1819>.

MILNER, S. G.; JOST, M.; TAKETA, S.; MAZÓN, E. R.; HIMMELBACH, A.; OPPERMANN, M.; WEISE, S.; KNÜPFER, H.; BASTERRECHEA, M.; KÖNIG, P.; SCHÜLLER, D.; SHARMA, R.; PASAM, R. K.; RUTTEN, T.; GUO, G.; XU, D.; ZHANG, J.; HERREN, G.; MÜLLER, T.; KRATTINGER, S. G.; KELLER, B.; JIANG, Y.; GONZÁLEZ, M. Y.; ZHAO, Y.; HABEKUB, A.; FÄRBER, S.; ORDON, F.; LANGE, M.; BÖRNER, A.; GRANER, A.; REIF, J. C.; SCHOLZ, U.; MASCHER, M.; STEIN, N. Genebank genomics highlights the diversity of a global barley collection. **Nature Genetics**, v. 51, 319–326, Feb. 2019. DOI: <https://doi.org/10.1038/s41588-018-0266-x>.

- MONTESINOS-LÓPEZ, O. A.; MONTESINOS-LÓPEZ, A.; PÉREZ-RODRÍGUEZ, P.; BARRÓN-LÓPEZ, A.; MARTINI, J. W. R.; FAJARDO-FLORES, S. B.; GAYTAN-LUGO, L.; SANTANA-MANCILLA, P. C.; CROSSA, J. A review of deep learning applications for genomic selection. **BMC Genomics**, v. 22, 19, Jan. 2021. DOI: <https://doi.org/10.1186/s12864-020-07319-x>.
- MÜLLER, D.; TECHNOW, F.; MELCHINGER, A. E. Shrinkage estimation of the genomic relationship matrix can improve genomic estimated breeding values in the training set. **Theoretical and Applied Genetics**, v. 128, n. 4, p. 693-703, Apr. 2015. DOI: <https://doi.org/10.1007/s00122-015-2464-6>.
- PANHO, M. C.; FERNANDES, R. A. T.; MENEGAZZI, C. P.; CAMPAGNOLLI, O. R.; QUADRA, F. C.; MADELLA, L. A.; MEIRA, D.; MALONE, G.; BRITO JUNIOR, S. L.; BENIN, G. Rpp-Gene pyramiding confers higher resistance level to Asian soybean rust. **Euphytica**, v. 218, 172, Nov. 2022. DOI: <https://doi.org/10.1007/s10681-022-03123-8>.
- QIAN, L.; HICKEY, L. T.; STAHL, A.; WERNER, C. R.; HAYES, B.; SNOWDON, R. J.; VOOS-FELS, K. P. Exploring and harnessing haplotype diversity to improve yield stability in crops. **Frontiers in Plant Science**, v. 8, 1534, Sept. 2017. DOI: <https://doi.org/10.3389/fpls.2017.01534>.
- RAO, S.; YAO, Y.; BAUER, D. E. Editing GWAS: experimental approaches to dissect and exploit disease-associated genetic variation. **Genome Medicine**, v. 13, 41, Mar. 2021. DOI: <https://doi.org/10.1186/s13073-021-00857-3>.
- RASHEED, A.; HAO, Y.; XIA, X.; KHAN, A.; XU, Y.; VARSHNEY, R. K.; HE, Z. Crop breeding chips and genotyping platforms: progress, challenges, and perspectives. **Molecular Plant**, v. 10, n. 8, p. 1047-1064, Aug. 2017. DOI: <https://doi.org/10.1016/j.molp.2017.06.008>.
- RAYAPROLU, L.; DESHPANDE, S. P.; GUPTA, R. Genotyping-by-sequencing (GBS) method for accelerating marker-assisted selection (MAS) program. In: WANI, S. H.; KUMAR, A. (ed.). **Genomics of cereal crops**. New York: Humana, 2022. p. 245-257. DOI: https://doi.org/10.1007/978-1-0716-2533-0_12.
- 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: Ed. UFV, 2012. 291 p.
- RESENDE, R. T.; PIEPHO, H.-P.; ROSA, G. J. M.; SILVA-JUNIOR, O. B.; SILVA, F. F.; RESENDE, M. D. V.; GRATTAPAGLIA, D. *Enviroomics* in breeding: applications and perspectives on envirotypic-assisted selection. **Theoretical and Applied Genetics**, v. 134, p. 95-112, Jan. 2021. DOI: <https://doi.org/10.1007/s00122-020-03684-z>.
- RESENDE, R. T.; RESENDE, M. D. V.; SILVA, F. F.; AZEVEDO, C. F.; TAKAHASHI, E. K.; SILVA-JUNIOR, O. B.; GRATTAPAGLIA, D. Assessing the expected response to genomic selection of individuals and families in Eucalyptus breeding with an additive-dominant model. **Heredity**, v. 119, n. 4, p. 245-255, July 2017. DOI: <https://doi.org/10.1038/hdy.2017.37>.
- ROBINSON, J. D.; COFFMAN, A. J.; HICKERSON, M. J.; GUTENKUNST, R. N. Sampling strategies for frequency spectrum-based population genomic inference. **BMC Evolutionary Biology**, v. 14, 254, Dec. 2014. DOI: <https://doi.org/10.1186/s12862-014-0254-4>.
- SCHEBEN, A.; BATLEY, J.; EDWARDS D. Genotyping-by-sequencing approaches to characterize crop genomes: choosing the right tool for the right application. **Plant Biotechnology Journal**, v. 15, n. 2, p. 149-161, Feb. 2017. DOI: <https://doi.org/10.1111/pbi.12645>.
- SEMALAIYAPPAN, J.; SELVANAYAGAM, S.; RATHORE, A.; GUPTA, S. K.; CHAKRABORTY, A.; GUJJULA, K. R.; HAKTAN, S.; VISWANATH, A.; MALIPATIL, R.; SHAH, P.; GOVINDARAJ, M.; IGNACIO, J. C.; REDDY, S.; SINGH, A. K.; THIRUNAVUKKARASU, N. Development of a new AgriSeq 4K mid-density SNP genotyping panel and its utility in pearl millet breeding. **Frontiers in Plant Science**, v. 13, 1068883, Jan. 2023. DOI: <https://doi.org/10.3389/fpls.2022.1068883>.

■ Melhoria de precisão: aplicações e perspectivas na genética...

- SHARMA, S.; SCHULTHESS, A. W.; BASSI, F. M.; BADAIEVA, E. D.; NEUMANN, K.; GRANER, A.; ÖZKAN, H.; WERNER, P.; KNÜPFER, H.; KILIAN, B. Introducing beneficial alleles from plant genetic resources into the wheat germplasm. **Biology**, v. 10, 982, Sept. 2021. DOI: <https://doi.org/10.3390/biology10100982>.
- SILVA, L. A.; PEIXOTO, M. A.; PEIXOTO, L. A.; ROMERO, J. V.; BHERING, L. L. Multi-trait genomic selection indexes applied to identification of superior genotypes. **Bragantia**, v. 80, e3621, 2021. DOI: <https://doi.org/10.1590/1678-4499.20200381>.
- SILVA, R. S.; FARIA, J. C.; KNUPP, A. M.; AGUIAR, M. S.; PEREIRA, H. S.; FERREIRA, A. L.; ZAIDEM, A. L. M.; PINHEIRO, P. V.; MELO, L. C.; SOUZA, T. L. P. O. Development and selection of transgenic advanced lines of carioca seeded common bean with multiple resistance to viruses. **Euphytica**, v. 218, 67, May 2022. DOI: <https://doi.org/10.1007/s10681-022-03017-9>.
- SILVA-JUNIOR, O. B.; FARIA, D. A.; GRATTAPAGLIA, D. A flexible multi-species genome-wide 60K SNP chip developed from pooled resequencing of 240 *Eucalyptus* tree genomes across 12 species. **New Phytologist**, v. 206, n. 4, p. 1527-1540, June 2015. DOI: <https://doi.org/10.1111/nph.13322>.
- SINGH, M.; NARA, U.; KUMAR, A.; THAPA, S.; JASWAL, C.; SINGH, H. Enhancing genetic gains through marker-assisted recurrent selection: from phenotyping to genotyping. **Cereal Research Communications**, v. 50, p. 523-538, Sept. 2022. DOI: <https://doi.org/10.1007/s42976-021-00207-4>.
- STEELE, K. A.; QUINTON-TULLOCH, M. J.; AMGAI, R. B.; DHAKAL, R.; KHATIWADA, S. P.; VYAS, D.; HEINE, M.; WITCOMBE, J. R. Accelerating public sector rice breeding with high-density KASP markers derived from whole genome sequencing of *indica* rice. **Molecular Breeding**, v. 38, 38, Mar. 2018. DOI: <https://doi.org/10.1007/s11032-018-0777-2>.
- SUN, S.; DONG, B.; ZOU, Q. Revisiting genome-wide association studies from statistical modelling to machine learning. **Briefings in Bioinformatics**, v. 22, bbaa263, Jul. 2021. DOI: <https://doi.org/10.1093/bib/bbaa263>.
- THUDI, M.; PALAKURTHI, R.; SCHNABLE, J. C.; CHITIKINENI, A.; DREISIGACKER, S.; MACE, E.; SRIVASTAVA, R. K.; SATYAVATHI, C. T.; ODENY, D.; TIWARI, V. K.; LAM, H. M.; HONG, Y. B.; SINGH, J. V. K.; LI, G.; XU, Y.; CHEN, X.; KAILA, S.; NGUYEN, H.; SIVASANKAR, S.; JACKSON, S. A.; CLOSE, T. J.; SHUBO, W.; VARSHNEY, R. K. Genomic resources in plant breeding for sustainable agriculture. **Journal of Plant Physiology**, v. 257, 153351, Feb. 2021. DOI: <https://doi.org/10.1016/j.jplph.2020.153351>.
- UFFELMANN, E.; HUANG, Q. Q.; MUNUNG, N. S.; VRIES, J.; OKADA, Y.; MARTIN, A. R.; MARTIN, H. C.; LAPPALAINEN, T.; POSTHUMA, D. Genome-wide association studies. **Nature Reviews Methods Primers**, v. 1, 59, Aug. 2021. DOI: <https://doi.org/10.1038/s43586-021-00056-9>.
- ULLAH, M. A.; ABDULLAH-ZAWAWI, M.-R.; ZAINAL-ABIDIN, R.-A.; SUKIRAN, N. L.; UDDIN, M. I.; ZAINAL, Z. A review of integrative omic approaches for understanding rice salt response mechanisms. **Plants**, v. 11, 1430, May 2022. DOI: <https://doi.org/10.3390/plants11111430>.
- VALDISSER, P. A. M. R.; PEREIRA, W. J.; ALMEIDA FILHO, J. E.; MÜLLER, B. S. F.; COELHO, G. R. C.; MENEZES, I. P. P.; VIANNA, J. P. G.; ZUCCHI, M. I.; LANNA, A. C.; COELHO, A. S. G.; OLIVEIRA, J. P.; MORAES, A. C.; BRONDANI, C.; VIANELLO, R. P. In-depth genome characterization of a Brazilian common bean core collection using DArTseq high-density SNP genotyping. **BMC Genomics**, v. 18, 423, May 2017. DOI: <https://doi.org/10.1186/s12864-017-3805-4>.
- VARSHNEY, R. K.; BOHRA, A.; YU, J.; GRANER, A.; ZHANG, Q.; SORRELLS, M. E. Designing future crops: genomics-assisted breeding comes of age. **Trends in Plant Science**, v. 26, n. 6, p. 631-649, June 2021. DOI: <https://doi.org/10.1016/j.tplants.2021.03.010>.
- WANG, B.; LIN, Z.; LI, X.; ZHAO, Y.; ZHAO, B.; WU, G.; MA, X.; WANG, H.; XIE, Y.; LI, Q.; SONG, G.; KONG, D.; ZHENG, Z.; WEI, H.; SHEN, R.; WU, H.; CHEN, C.; MENG, Z.; WANG, T.; LI, Y.; LI, X.; CHEN, Y.; LAI, J.; HUFFORD, M. B.; ROSS-IBARRA, J.; HE, H.; WANG, H. Genome-wide selection and genetic improvement during modern maize breeding. **Nature Genetics**, v. 52, p. 565-571, June 2020. DOI: <https://doi.org/10.1038/s41588-020-0616-3>.

- WEN, Y.-J.; ZHANG, H.; NI, Y.-L.; HUANG, B.; ZHANG, J.; FENG, J.-Y.; WANG, S.-B.; DUNWELL, J. M.; ZHANG, Y.-M.; WU, R. Methodological implementation of mixed linear models in multi-locus genome-wide association studies. **Briefings in Bioinformatics**, v. 19, n. 4, p. 700–712, July 2018. DOI: <https://doi.org/10.1093/bib/bbw145>.
- WERNER, C. R.; GAYNOR, R. C.; GORJANC, G.; HICKEY, J. M.; KOX, T.; ABBADI, A.; LECKBAND, G.; SNOWDON, R. J.; STAHL, A. How population structure impacts genomic selection accuracy in cross-validation: implications for practical breeding. **Frontiers in Plant Science**, v. 11, 592977, Dec. 2020. DOI: <https://doi.org/10.3389/fpls.2020.592977>.
- YAN, J.; WANG, X. Machine learning bridges omics sciences and plant breeding. **Trends in Plant Science**, v. 28, n. 2, p. 199-199-210, Feb. 2023. DOI: <https://doi.org/10.1016/j.tplants.2022.08.018>.
- YANG, D.; TANG, J.; YANG, D.; CHEN, Y.; ALI, J.; MOU, T. Improving rice blast resistance of Feng39S through molecular marker-assisted backcrossing. **Rice**, 12, 70, Sept. 2019. DOI: <https://doi.org/10.1186/s12284-019-0329-3>.
- YOOSEFZADEH-NAJAFABADI, M.; RAJCAN, I.; ESKANDARI, M. Optimizing genomic selection in soybean: an important improvement in agricultural genomics. **Heliyon**, v. 8, e11873, Nov. 2022. DOI: <https://doi.org/10.1016/j.heliyon.2022.e11873>.
- YU, J.-K.; CHUNG, Y.-S. Plant variety protection: current practices and insights. **Genes**, v. 12, 1127, July 2021. DOI: <https://doi.org/10.3390/genes12081127>.
- ZHANG, A.; WANG, H.; BEYENE, Y.; SEMAGN, K.; LIU, Y.; CAO, S.; CUI, Z.; RUAN, Y.; BURGUEÑO, J.; SAN VICENTE, F.; OLSEN, M.; PRASANNA, B. M.; CROSSA, J.; YU, H.; ZHANG, X. Effect of trait heritability, training population size and marker density on genomic prediction accuracy estimation in 22 biparental tropical maize populations. **Frontiers in Plant Science**, v. 8, 1916, Nov. 2017. DOI: <https://doi.org/10.3389/fpls.2017.01916>.
- ZHANG, H.; JIANG, H.; HU, Z.; SONG, Q.; CHARLES AN, Y.-G. Development of a versatile resource for post-genomic research through consolidating and characterizing 1500 diverse wild and cultivated soybean genomes. **BMC Genomics**, v. 23, 250, Mar. 2022. DOI: <https://doi.org/10.1186/s12864-022-08326-w>.
- ZHOU, Y.; ZHANG, Z.; BAO, Z. et al. Graph pangenome captures missing heritability and empowers tomato breeding. **Nature**, v. 606, p. 527–534, June 2022. DOI: <https://doi.org/10.1038/s41586-022-04808-9>.
- ZHUANG, J.; ZHANG, J.;-- HOU, X.-L.; WANG, F.; XIONG, A.-S. Transcriptomic, proteomic, metabolomic and functional genomic approaches for the study of abiotic stress in vegetable crops. **Critical Reviews in Plant Sciences**, v. 33, n. 2-3, p. 225-237, 2014. DOI: <https://doi.org/10.1080/07352689.2014.870420>.
- ZIMMERMAN, S. J.; ALDRIDGE, C. L.; OYLER-MCCANCE, S. J. An empirical comparison of population genetic analyses using microsatellite and SNP data for a species of conservation concern. **BMC Genomics** 21, 382, June 2020. DOI: <https://doi.org/10.1186/s12864-020-06783-9>.