

MELHORAMENTO DE PRECISÃO

*Aplicações e perspectivas
na genética de plantas*

**RAFAEL TASSINARI RESENDE
CLAUDIO BRONDANI**

Editores Técnicos



Embrapa

*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*

Embrapa
Parque Estação Biológica
Av. W3 Norte (final)
70770-901 Brasília, DF
www.embrapa.br
www.embrapa.br/fale-conosco/sac

Responsável pelo conteúdo e edição

Embrapa Arroz e Feijão
Rodovia GO-462, Km. 12, Zona Rural
Caixa Postal 179
75375-000 Santo Antônio de Goiás, GO

Comitê Local de Publicações
da Embrapa Arroz e Feijão

Presidente
Roselene de Queiroz Chaves

Membros
Tereza Cristina de Oliveira Borba
Newton Cavalcanti de Noronha Júnior
Luís Fernando Stone
Ana Lucia Delalibera de Faria

Supervisão editorial
Rafael Tassinari Resende
Claudio Brondani

Revisão de texto
Luís Fernando Stone

Normalização bibliográfica
Ana Lucia Delalibera de Faria

Capa
Fábio Fernandes Noletto

Editoração eletrônica
Fabiano Severino

1ª edição

1ª impressão (2023): 100 exemplares
Publicação digital (2023): PDF

Todos os direitos reservados.

A reprodução não autorizada desta publicação, no todo ou em parte,
constitui violação dos direitos autorais (Lei nº 9.610).

Dados Internacionais de Catalogação na Publicação (CIP)

Embrapa Arroz e Feijão

Melhoramento de precisão : aplicações e perspectivas na genética de plantas / Rafael Tassinari
Resende, Claudio Brondani, editores técnicos. – Brasília, DF: Embrapa, 2023.

PDF (326 p.). il. color

ISBN 978-65-5467-005-0

1. Agricultura de precisão. 2. Melhoramento genético vegetal. 3. Automação de processos.
4. Otimização da seleção. 5. Inteligência artificial. 6. Multiômicas. 7. Ciência de dados. 8. Inovação
genética. 9. Biotecnologia vegetal. 10. Abordagens multidisciplinares. I. Resende, Rafael Tassinari. II.
Brondani, Claudio. III. Embrapa Arroz e Feijão.

CDD 631.5233

Ana Lúcia Delalibera de Faria (CRB-1/324)

©Embrapa, 2023

AUTORES

Adriano Teodoro Bruzi

Engenheiro-agrônomo, doutor em Genética e Melhoramento de Plantas, professor da Universidade Federal de Lavras (Ufla), Lavras, MG

Alexandre Bryan Heinemann

Engenheiro-agrônomo, doutor em Irrigação e Drenagem, pesquisador da Embrapa Arroz e Feijão, Santo Antônio de Goiás, GO

Alexandre Lima Nepomuceno

Engenheiro-agrônomo, doutor em Biologia Molecular, pesquisador da Embrapa Soja, Londrina, PR

Ana Carolina Campana Nascimento

Estatística, doutora em Economia Aplicada, professora da Universidade Federal de Viçosa (UFV), Viçosa, MG

Camila Azevedo Ferreira

Matemática, doutora em Estatística Aplicada e Biométrica, professora da Universidade Federal de Viçosa (UFV), Viçosa, MG

Claudio Brondani

Engenheiro-agrônomo, doutor em Biologia Molecular, pesquisador da Embrapa Arroz e Feijão, Santo Antônio de Goiás, GO

Cosme Damião Cruz

Engenheiro-agrônomo, doutor em Agronomia, professor da Universidade Federal de Viçosa (UFV), Viçosa, MG

Fabrcio Barbosa Monteiro Arraes

Bi6logo, doutor em Biologia Celular e Molecular, pesquisador da Sempre AgTech, Bras6lia, DF

Filipe In6cio Matias

Engenheiro-agr6nomo, doutor em Gen6tica e Melhoramento de Plantas, pesquisador da Phenome-Force: Plant Phenotyping Community, Uberl6ndia, MG

Flavia Maria Avelar Goncalves

Engenheira-agr6noma, doutora em Gen6tica e Melhoramento de Plantas, professora da Universidade Federal de Lavras (Ufla), Lavras, MG

Germano Martins Ferreira Costa-Neto

Engenheiro-agr6nomo, doutor em Gen6tica e Melhoramento de Plantas, p6s-doutorando em Cornell University, Estados Unidos da Am6rica

Gustavo Eduardo Marcatti

Engenheiro florestal, doutor em Ci6ncia Florestal, professor da Universidade Federal de S6o Jo6o de Rei (UFSJ), Sete Lagoas, MG

Gustavo Evangelista Oliveira

Engenheiro-agr6nomo, doutor em Gen6tica e Melhoramento de Plantas, pesquisador da Enza Zaden Pesquisa e Desenvolvimento de Sementes, Holambra, SP

Hugo Bruno Correa Molinari

Engenheiro-agr6nomo, doutor em Agronomia, pesquisador da Sempre AgTech, Bras6lia, DF

Jo6o Gabriel Zanon Paludeto

Engenheiro florestal, mestre em Ci6ncias Florestais, pesquisador da Klabin S.A., Tel6maco Borba, PR

Jo6o Lu6s da Silva Filho

Engenheiro-agr6nomo, doutor em Gen6tica e Melhoramento de Plantas, pesquisador da Embrapa Algod6o, Campina Grande, PB

Jo6o Otavio Firigato

Engenheiro da computa6o, mestre em Geografia, pesquisador da Universidade Federal de Mato Grosso do Sul, Tr6s Lagoas, MS

José Maria Villela Pádua

Engenheiro-agrônomo, doutor em Genética e Melhoramento de Plantas, professor da Universidade Federal de Lavras (Ufla), Lavras, MG

Lázaro José Chaves

Engenheiro-agrônomo, doutor em Genética e Melhoramento de Plantas, professor da Universidade Federal de Goiás (UFG), Goiânia, GO

Leonardo de Castro Santos

Engenheiro-agrônomo, doutor em Agronomia, professor do Instituto Federal de Educação, Ciência e Tecnologia Goiano (IF Goiano), Rio Verde, GO

Leonardo Volpato

Engenheiro-agrônomo, doutor em Genética e Melhoramento de Plantas, professor da Michigan State University, East Lansing, Michigan, Estados Unidos da América

Magno Antonio Patto Ramalho

Engenheiro-agrônomo, doutor em Genética e Melhoramento de Plantas, professor da Universidade Federal de Lavras (Ufla), Lavras, MG

Marcelo Gonçalves Narciso

Engenheiro eletrônico, doutor em Computação Aplicada, pesquisador da Embrapa Arroz e Feijão, Santo Antônio de Goiás, GO

Marco Antônio Marcelino Bahia

Engenheiro florestal, mestre em Genética e Melhoramento de Plantas, doutorando da Universidade Federal de Goiás (UFG), Goiânia, GO

Marcos Santos

Engenheiro naval, doutor em Engenharia de Produção, professor do Instituto Militar de Engenharia (IME), Rio de Janeiro, RJ

Mayla Daiane Correa Molinari

Biomédica, doutora em Genética e Biologia Molecular, pesquisadora da Sempre AgTech, Brasília, DF

Moysés Nascimento

Estatístico, doutor em Estatística e Experimentação Agropecuária, professor da Universidade Federal de Viçosa (UFV), Viçosa, MG

Priscila Ferreira Batista

Bióloga, doutora em Ciências Agrárias, professora do Instituto Federal de Educação, Ciência e Tecnologia Goiano (IF Goiano), Rio Verde, GO

Rafael Tassinari Resende

Engenheiro florestal, doutor em Genética e Melhoramento de Plantas e doutor em Ciências Florestais, professor da Universidade Federal de Goiás (UFG), Goiânia, GO

Renata Fuganti-Pagliarini

Bióloga, doutora em Genética e Melhoramento de Plantas, pesquisadora da Piccola Scientific Research, Saskatoon, Saskatchewan, Canadá

Ricardo Augusto Diniz Cabral Ferreira

Engenheiro-agrônomo, doutor em Genética e Melhoramento de Plantas, pesquisador da Syngenta, Uberlândia, MG

Roberto Fritsche-Neto

Engenheiro-agrônomo, doutor em Genética e Melhoramento de Plantas, professor da Louisiana State University (LSU), Baton Rouge, Louisiana, Estados Unidos da América

Roberto Tarazi

Biólogo, doutor em Genética e Melhoramento de Plantas, professor da Universidade Federal do Rio de Janeiro, Duque de Caxias, RJ

Rosana Pereira Vianello

Bióloga, doutora em Biologia Molecular, pesquisadora da Embrapa Arroz e Feijão, Santo Antônio de Goiás, GO

Tavvs Micael Alves

Engenheiro-agrônomo, doutor em Entomologia, professor do Instituto Federal de Educação, Ciência e Tecnologia Goiano (IF Goiano), Rio Verde, GO

Esta obra é dedicada ao Professor Dr. Fabyano Fonseca e Silva (*in memoriam*), por suas enormes contribuições ao melhoramento animal e de plantas, e exemplo de humildade e dedicação ao ensino e pesquisa.

PREFÁCIO

Vivemos na era da Agricultura de Precisão, na qual temos testemunhado mudanças significativas devido ao rápido avanço tecnológico. No campo da fenotipagem, anteriormente era obrigatório realizar medições manuais de campo, experimento por experimento, parcela por parcela. Atualmente, essa tarefa pode ser facilitada através do uso de drones equipados com sensores espectrais. No campo da seleção genética, em que o objetivo é concentrar alelos favoráveis para expressar características desejadas, a busca por esses alelos é realizada por meio de cruzamentos em várias gerações sucessivas de melhoramento. Mas hoje, graças às ferramentas de edição gênica, é possível identificar e incorporar genes específicos com precisão, ao nível de bases de DNA.

De fato, entramos na era do Melhoramento de Precisão, que mais do que ciências de engenharia genética, hoje lançamos mão de múltiplas altas-tecnologias de campo e laboratório, tais como a inteligência artificial, geoprocessamento e marcadores genômicos hiper densos. Estes instrumentos modernos permitem a coleta de grandes volumes de dados, análise precisa e rápida, e a tomada de decisões mais informatizadas e direcionadas do que nunca! Além disso, a combinação desses recursos potencializa ainda mais a seleção de características desejáveis, acelerando o desenvolvimento de variedades de plantas mais produtivas, resistentes a doenças e adaptadas às mais diversas condições ambientais.

Editorada pelos professores e pesquisadores Dr. Rafael Tassinari Resende, idealizador da “Ambientômica” no melhoramento vegetal, e Dr. Cláudio Brondani, uma autoridade na genômica vegetal brasileira, esta obra é de extrema importância para todos aqueles interessados no aprimoramento genético das culturas. O livro oferece uma visão abrangente do estado da arte na organização e otimização dos programas de melhoramento, além de novos *insights* tecnológicos para evolução da área. Seus

capítulos contam com a contribuição de inúmeros autores especialistas, incluindo os próprios editores. Dessa forma, esta publicação aborda de maneira inovadora e original diversos temas relacionados ao melhoramento genético na era da Agricultura de Precisão.

Este livro se destaca pela análise de diversos aspectos-chave no campo do melhoramento genético. Entre os seus pontos fortes, merecem evidência a abordagem detalhada da gestão dos programas de melhoramento, o desenvolvimento tecnológico e a inovação, bem como a exploração dos avanços em ciência de dados, inteligência artificial e conceitos da pesquisa operacional. Além disso, o livro trata dos impactos da Internet das Coisas (IoT) e apresenta uma visão aprofundada das disciplinas fenômica e genômica. Também são abordadas técnicas emergentes, como o *speed breeding*, edição gênica, ambientômica, e muitas outras novas *hightechs* do melhoramento. Esses temas são de relevância para impulsionar o desenvolvimento de novas cultivares, permitindo progressos significativos e otimizados dos resultados alcançados nessa área vital da agricultura.

Este livro é uma leitura obrigatória para todos os apaixonados pelo melhoramento de plantas. Aproveite ao máximo as valiosas informações e contribuições que esta obra proporciona. Boa leitura!

Aluízio Borém

Professor Titular de Melhoramento de Plantas
Universidade Federal de Viçosa (UFV)

SUMÁRIO

Capítulo	1	O Melhoramento na Era de Agricultura de Precisão	13
Capítulo	2	Gestão Otimizada dos Programas	41
Capítulo	3	Empreendedorismo e Inovação.....	67
Capítulo	4	Ciência de Dados	87
Capítulo	5	Inteligência Artificial	121
Capítulo	6	Pesquisa Operacional	145
Capítulo	7	Internet das Coisas	173
Capítulo	8	Ambientômica	195
Capítulo	9	Fenômica	227
Capítulo	10	Genômica	255
Capítulo	11	Speed Breeding	287
Capítulo	12	Edição de Genoma.....	309

Rafael Tassinari Resende, Claudio Brondani,
Lazaro José Chaves

Introdução

Agricultura moderna enfrenta desafios constantes, como a necessidade de aumentar a produção de insumos para uma população em contínuo crescimento, buscando por soluções sustentáveis para garantir as demandas por alimentos, bioenergias, fibras, medicamentos, forragens, entre outras. Nesse cenário, o Melhoramento Genético de plantas desempenha relevante papel, sempre buscando desenvolver cultivares com características desejáveis, como maior produtividade, resistência a insetos-pragas e doenças e adaptabilidade a diferentes condições ambientais (Hickey et al., 2019; Ramalho et al., 2021).

No decorrer da história, de forma análoga às “Eras da Revolução Industrial”, o Melhoramento Genético de plantas também evoluiu em diferentes fases, refletindo avanços tecnológicos e científicos. Pode-se dizer que o Melhoramento Genético 1.0, associado à domesticação promovida de paisagens, marcou o início das interações humanas com as plantas, em que a seleção de plantas úteis e a eliminação das competidoras ocorriam de forma rudimentar. Essa primeira era seguiu até o surgimento da agricultura (entre 10.000 a.C. e 8.000 a.C.), havendo mudanças significativas, como a seleção intencional por características desejáveis e a adaptação de plantas ao cultivo humano (Wallace et al., 2018). Em seguida, o Melhoramento Genético 2.0 combinou a genética mendeliana e a biometria, permitindo a previsão dos resultados da seleção e o desenvolvimento de métodos mais sistemáticos de melhoramento.

Tal como com o avanço da tecnologia na Revolução Industrial 3.0, o Melhoramento Genético 3.0 também incorporou técnicas novas, no caso, moleculares e genômicas, como a identificação de marcadores moleculares e a análise de dados genômicos

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

de alto rendimento. Essas abordagens aceleraram o processo de melhoramento, possibilitando a identificação de múltiplos efeitos gênicos relacionados à expressão dos fenótipos desejáveis e o desenvolvimento de cultivares com maior rigor. Agora, na era do Melhoramento de Precisão, o Melhoramento Genético 4.0 está em grande expansão, impulsionado pela convergência de tecnologias avançadas, como a genotipagem em larga escala, Inteligência Artificial (IA), Sistemas de Informações Geográficas (SIG), drones, sensoriamento remoto e análise de dados multiômicos (Crossa et al., 2021; Resende et al., 2021; Varshney et al., 2021).

A qualquer momento estaremos entrando no Melhoramento Genético 5.0. Aliás, quem sabe já não estejamos vivendo esta era? Essa fase do aprimoramento de plantas promete uma revolução na forma como se desenvolvem cultivares. Com a automação total dos processos e o avanço da robótica, espera-se que o melhoramento genético seja conduzido por meio de sistemas inteligentes e robôs especializados (Pearson et al., 2022; Najafabadi et al., 2023). Essas tecnologias, quando implementada nas rotinas de melhoramento, permitirão uma coleta eficiente de dados fenotípicos de altíssimo rendimento, acelerando a identificação e a mensuração dos fenótipos-alvo das plantas. Além disso, a análise de dados se beneficiará de algoritmos avançados e inteligência artificial, proporcionando uma tomada de decisão mais precisa no processo de seleção genética.

Neste primeiro capítulo, vamos explorar a interseção entre o Melhoramento Genético de plantas e a Agricultura de Precisão, apresentando técnicas e conceitos fundamentais. Abordaremos a história do melhoramento genético de plantas, culminando na era atual, que será explorada nos capítulos subsequentes deste livro. Faremos uma analogia à Equação do Melhorista, relacionando-a com questões contemporâneas do Melhoramento de Precisão. Além disso, apresentaremos os resultados de uma pesquisa realizada com profissionais da área, investigando suas perspectivas em relação ao Melhoramento de Precisão. Demonstraremos como a combinação dessas abordagens está transformando a agricultura, impulsionando a sustentabilidade, produtividade e adaptabilidade das plantas cultivadas. Nosso objetivo com este livro é auxiliar os melhoristas a otimizarem o processo de seleção, reduzindo o tempo necessário e maximizando os ganhos obtidos.

Histórico do Melhoramento Genético

A interferência das populações humanas sobre as plantas e seu habitat antecedem a origem da agricultura, quando a composição de plantas de uma comunidade vegetal

era alterada pela interferência humana, favorecendo a ocorrência de plantas úteis e eliminando ou desfavorecendo as plantas competidoras. Esse processo é conhecido como domesticação promovida de paisagens (Clement, 2001) e a alteração genética decorrente nas plantas foi, presumivelmente, de baixo impacto imediato, uma vez que as plantas continuaram a depender de seu sistema natural de dispersão de sementes ou outras estruturas de propagação para a manutenção das populações naturais.

A origem da agricultura, cujos mais antigos registros fósseis datam de aproximadamente 12 mil anos, trouxe uma mudança radical na relação dos humanos com as plantas cultivadas, do ponto de vista evolutivo. Se antes as populações de plantas silvestres dependiam de seus eficientes mecanismos de dispersão de sementes para sobreviverem, as populações cultivadas passaram a ter os humanos, no início mulheres provavelmente, como únicos dispersores. Com isso, a seleção, automática e não intencional, passou a favorecer genótipos com menor poder de dispersão de sementes, sementes com germinação mais rápida e com maior vigor. Esse fato, por si só, acarretou uma série de mudanças morfofisiológicas nas populações cultivadas, como a retenção de sementes, maior uniformidade de maturação, aumento do tamanho das sementes, perda de dormência e estruturas de proteção de sementes. Esse conjunto de mudanças, que ocorreram em diferentes espécies e em diferentes centros de domesticação, ficaram conhecidas como síndrome da domesticação. Dentre essas características decorrentes da domesticação, algumas como dormência de sementes possuem controle genético mais simples e ocorreram em relativamente curto espaço de tempo, algumas centenas de anos, enquanto outras, como retenção de sementes, de controle genético mais complexo, deram-se em um período mais longo de tempo.

Com o avanço da agricultura como atividade humana, as plantas cultivadas foram levadas para regiões distantes do seu centro de origem e foram submetidas a diferentes pressões de seleção para adaptação às condições de solo, variáveis climáticas e modos de cultivo. Adicionalmente, a seleção intencional pelos agricultores passou a favorecer certos tipos de variantes que chamavam a atenção por algum atributo de interesse. Esse processo, que se expandiu por milhares de anos, pelas diferentes regiões da terra a partir dos centros de domesticação, levou a uma enorme variabilidade morfológica nas espécies cultivadas em comparação com seus ancestrais silvestres. Esse processo de expansão da variabilidade fenotípica e genotípica ganhou novo impulso com as navegações a partir do século XVI, com intenso intercâmbio de espécies domesticadas entre os continentes e adaptação a novos ambientes de cultivo. Essa evolução sob cultivo em múltiplos ambientes, com diferentes técnicas de cultivo e para atender aos interesses particulares de cada agrupamento humano, resultou em milhares de variedades locais ou crioulas, para as diferentes espécies

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

cultivadas. A importância dessa enorme diversidade de plantas para a sustentabilidade da agricultura só foi amplamente reconhecida a partir dos trabalhos de N. I. Vavilov, na primeira metade do século XX (Hummer; Hancock, 2015), o que impulsionou uma política global de conservação de recursos genéticos vegetais. A diversidade genética, hoje armazenada em bancos de germoplasma ao redor do mundo, constitui a matéria-prima para o melhoramento de plantas, mesmo na era do melhoramento de precisão. Adicionalmente, técnicas de genética molecular e genômica são cada vez mais empregadas na caracterização e utilização das coleções de germoplasma, como importante ferramenta auxiliar do melhoramento genético.

Apesar da enorme diversidade fenotípica decorrente da domesticação e expansão, a origem das populações cultivadas se deu a partir de populações locais silvestres, geneticamente uniformes, e representando uma estreita base genética original. Estudos genômicos têm identificado genes responsáveis por várias das mudanças decorrentes da domesticação, em diferentes espécies cultivadas. Isto possibilita a proposta de domesticação *de novo* ou redomesticação (Fernie; Yan, 2019), como forma de ampliar a base genética de populações cultivadas a partir de seus ancestrais selvagens, utilizando ferramentas genômicas atuais, como a edição gênica.

O procedimento para a seleção intencional de plantas que prevaleceu durante o longo período de domesticação foi, provavelmente, a escolha de indivíduos com características de interesse para coleta de sementes ou outros propágulos, o que se conhece como seleção massal. Atribui-se o marco de surgimento dos métodos sistematizados de melhoramento aos trabalhos de Louis de Vilmorin na França, em meados do século XIX. O método, empregado para selecionar beterrabas com maior teor de açúcar, consistia no plantio de progênies em fileiras individuais, sendo parte das plantas utilizadas para avaliar o teor de açúcar, e posterior coleta de sementes das outras plantas das fileiras superiores. Esse método, conhecido como Princípio do Isolamento de Vilmorin, foi posteriormente empregado para seleção de teores de óleo e proteína em milho a partir de 1896, sendo denominado método espiga-por-fileira. Aprimorado com o uso de delineamentos experimentais, o método é hoje denominado seleção entre e dentro de famílias de meios-irmãos e continua sendo muito utilizado no melhoramento de plantas alógamas.

Os diferentes métodos de melhoramento de plantas utilizados atualmente surgiram no início do século XX, após a redescoberta das leis de Mendel. O surgimento da genética como ciência trouxe melhor compreensão dos mecanismos de herança dos caracteres e a possibilidade de previsão dos resultados da seleção a partir de populações variáveis ou de cruzamentos dirigidos. A junção da genética mendeliana

com a biometria, que vinha sendo aprimorada desde meados do século XIX, deu origem à genética quantitativa ou genética biométrica, principal ciência de base para o melhoramento de plantas. A adoção do melhoramento de precisão pode trazer vantagens competitivas quando adotado, como aumento da eficiência e velocidade no desenvolvimento de novas cultivares e maior precisão na seleção de caracteres de interesse, resultando em materiais genéticos superiores em termos de eficiência e rentabilidade.

Na era da agricultura de precisão, o melhoramento genético de plantas continua tendo o papel principal na busca por cultivares mais produtivas, resistentes a doenças e com alto desempenho em diferentes condições ambientais. Com o auxílio de avanços tecnológicos, como Sistemas de Informações Geográficas, fenotipagem em larga escala e genômica, os programas de melhoramento têm a oportunidade de alcançar ganhos de seleção surpreendentes. Combinações estratégicas de inteligência artificial, computação, sensoriamento remoto e automação de processos têm acelerado o ritmo das descobertas e impulsionado a eficiência dos ciclos de seleção (Bernardo, 2020). Nesse contexto, os agricultores e pesquisadores têm à sua disposição um vasto arsenal de ferramentas que permitem a identificação de características desejáveis em plantas, conjuntamente com a análise genômica em larga escala e a tomada de decisões baseada em dados precisos. A convergência dessas tecnologias está abrindo novas perspectivas para o melhoramento genético, permitindo o desenvolvimento de cultivares cada vez mais adaptadas, sustentáveis e produtivas, capazes de enfrentar os desafios crescentes da agricultura moderna.

Embora o melhoramento de precisão seja uma abordagem promissora e complementar ao melhoramento clássico, é improvável que substitua completamente essa forma tradicional de melhoramento de plantas. Ambas as abordagens têm seus pontos fortes e são úteis em diferentes contextos. O melhoramento clássico, também conhecido como melhoramento convencional, tem sido usado há muito tempo para desenvolver novas cultivares de plantas, e envolve a seleção e o cruzamento de genótipos com características desejáveis para melhorar suas qualidades agronômicas, com um histórico comprovado de sucesso. O melhoramento de precisão pode acelerar o processo de melhoramento e facilitar a seleção de plantas com características desejáveis, enquanto o melhoramento clássico continua sendo uma valiosa ferramenta para desenvolver novas cultivares e combinar características complexas e, portanto, por serem complementares, podem ser usados em conjunto.

A Equação do Melhorista

Neste livro, não se pretende desvalorizar as técnicas e práticas de melhoramento genético clássico, rotulando-as como arcaicas ou obsoletas. Muito pelo contrário, o propósito é reconhecer e apreciar sua importância, ao mesmo tempo em que se incorpora novas possibilidades de coleta e análise de dados. A contraposição do melhoramento clássico ou convencional com o melhoramento moderno constitui uma falsa controvérsia. A criação de novas cultivares se dá pelos métodos clássicos de melhoramento. A efetividade desses métodos tem sido continuamente incrementada com novas técnicas e procedimentos, desde ajustes experimentais até as técnicas do melhoramento de precisão.

O objetivo final do melhoramento de plantas é obter populações de alto desempenho para serem utilizadas diretamente como cultivares ou como populações base para extração de linhagens ou clones. A média de um caráter na população melhorada será, por definição, a média da população original acrescida do ganho por seleção. Uma população base ideal será aquela com média favorável e com potencial para gerar maiores ganhos genéticos adicionais. Para fins de seleção recorrente, geralmente são utilizados compostos ou sintéticos como população base. A escolha de parentais utilizando métodos de predição é uma forma eficiente de assegurar a obtenção de populações com propriedades adequadas. No caso de seleção de linhas puras ou clones a partir de cruzamentos dirigidos, a qualidade da população de seleção também vai depender da escolha apropriada de parentais.

A equação para predição de ganho por seleção, denominada equação do melhorista (*breeder's equation*), é considerada a principal aplicação da genética quantitativa no melhoramento genético de plantas e animais (Cobb et al., 2019). A equação tem origem no conceito de regressão para a média de Francis Galton, em seus estudos relacionando caracteres de plantas, animais e humanos, entre gerações, na segunda metade do século XIX. O conceito de Galton recebeu tratamento matemático apropriado por Karl Pearson, gerando os coeficientes de regressão e de correlação (Stanton, 2001). Na sua forma mais simples, a equação do ganho por ciclo de seleção pode ser descrita por $\Delta G = ds \times h^2$, em que ds é o diferencial de seleção, ou seja, a diferença entre a média dos indivíduos ou progênies selecionados e a média geral da população, e h^2 é o coeficiente de herdabilidade específico para cada população, caráter e método de melhoramento. O coeficiente de herdabilidade é, por natureza, um coeficiente de regressão dos valores do caráter medidos nas unidades da população melhorada em relação aos valores do mesmo caráter nas unidades aparentadas na geração de seleção. Se denotarmos por X_i os valores na geração de seleção

e por X'_i os valores correspondentes na população melhorada, o coeficiente de herdabilidade corresponde a $h^2 = \frac{Cov_{xx'}}{\sigma_{Fx}^2}$. A covariância de X com X' é uma medida da semelhança entre parentes na geração de seleção e na população melhorada, sendo uma função de componentes da variância genética da população de referência, para o caráter em questão (Carena et al., 2010). Para os métodos mais usuais de seleção intrapopulacional, a covariância corresponde à variância genética aditiva disponível entre as unidades de seleção, indivíduos ou progênies, com seleção em ambos os sexos. O denominador da equação refere-se à variância fenotípica entre as unidades de seleção. O diferencial de seleção pode ser expresso em unidades de desvio padrão fenotípico fazendo $i = \frac{ds}{\sigma_{F_x}}$. Supondo seleção truncada e distribuição normal dos dados, o valor de i pode ser obtido a partir da proporção de unidades selecionadas. Para a comparação entre métodos e procedimentos de seleção, a equação pode ser adaptada para acomodar o tempo gasto em cada ciclo e possível endogamia na população melhorada. Com isso, a equação do melhorista pode ser reescrita por $\Delta G = (i \times r \times \sigma_g)/L - dF$. Essa equação encapsula os principais elementos que influenciam o ganho genético no processo de melhoramento de plantas. ΔG representa o objetivo final de todo esforço, o ganho genético alcançado ao longo do tempo. A intensidade de seleção (i) reflete a proporção de indivíduos selecionados, enquanto a acurácia da seleção (r) destaca a confiabilidade das predições. O desvio padrão genético (σ_g) mede a diversidade genética disponível para seleção de acordo com o método empregado. O tempo gasto por ciclo de seleção (L) pondera a rapidez do processo de melhoramento e a depressão por endogamia (dF) considera os efeitos negativos da reprodução entre parentais geneticamente relacionados. Compreender e otimizar esses termos é essencial para maximizar os ganhos genéticos, acelerar o processo de melhoramento e desenvolver variedades superiores e adaptadas às demandas atuais.

Pesquisa Realizada com Melhoristas Brasileiros sobre o Melhoramento de Precisão

Sinopse dos dados coletados

A fim de ilustrar o cenário do Melhoramento de Precisão no Brasil, foi realizada uma pesquisa via formulário Google, amplamente divulgada em universidades, instituições de pesquisa e algumas empresas dedicadas ao melhoramento genético de plantas. Neste formulário, obteve-se 238 respostas entre os dias 14 de fevereiro a 14 de maio

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

de 2023, e os resultados foram reunidos, sumarizados no decorrer do presente texto e ilustrados a partir do pacote ggplot2 do R (Wickham, 2016). Ao todo foram quinze perguntas, iniciando com o cargo/profissão/atividade do profissional. As possibilidades de resposta incluíam: estudante de pós-graduação em melhoramento de plantas ou área correlata, técnico ou especialista, docente/professor, pesquisador, consultor, cientista/analista de dados, empresário, líder/chefe de pesquisa ou setor, e em busca de recolocação. Solicitou-se aos participantes que selecionassem apenas sua ocupação principal, e, no caso de aposentados, que marcassem a ocupação antes da aposentadoria. Como se obteve apenas duas respostas de empresários, as análises feitas para esse público foram desconsideradas para evitar viés no resultado. Também se perguntou as idades, em intervalos etários de 5 em 5 anos ({21-25, 26-30, ..., 71-75}). Porém, como foram poucos os respondentes com mais de 60 anos (isto é, nas classes 61-65, 66-70 e 71-75), eles foram agrupados como “>60” anos (Figura 1A).

Adicionalmente foi perguntado sobre a formação dos entrevistados (Figura 1B). Outra pergunta foi sobre as instituições e segmentos que os profissionais atuavam, ou seja, se nos setores públicos ou privados, e se eram de instituições ou áreas de pesquisa ou ensino, ou se o profissional atuava como autônomo(a). Na primeira pergunta, sobre o tipo de instituição, as respostas foram: “instituição pública”, com 62,2% dos participantes selecionando essa opção; “instituição privada” foi selecionada por 31,5% dos respondentes; “Não se aplica” foi a opção de 4,6% dos participantes; e “Autônomo”, escolhido por apenas 1,7% dos respondentes. Na segunda pergunta, as possibilidades de resposta ficaram assim distribuídas: setor (ou instituição) de Pesquisa foi selecionada por 54,6% dos participantes; Ensino/pesquisa/extensão (ou Universidades) foi escolhida por 31,5% dos respondentes; Setor Produtivo (como empresas, companhias ou indústrias) foi a opção para 8% dos participantes; e “Outros”, escolhido por 5,9% dos respondentes.

Profissionais que responderam à pesquisa lidam com diversas culturas agrícolas, e na Figura 2 é mostrada a distribuição desses profissionais. Note-se que um profissional poderá trabalhar com mais de uma cultura e, por esse motivo, o somatório das porcentagens do gráfico é superior a 100. São exemplos dessas espécies, as autóginas anuais: soja, arroz, trigo, feijão, centeio, cevada; alógamas anuais: milho, girassol e algumas hortaliças; alógamas perenes: coco, eucalipto, dendê, teca e mogno; espécies de propagação vegetativa: batata, morango, mandioca, banana, cana-de-açúcar; espécies de propagação mista: cacau, abacate, limão, manga; e autóginas perenes: pêssego e café arábica.

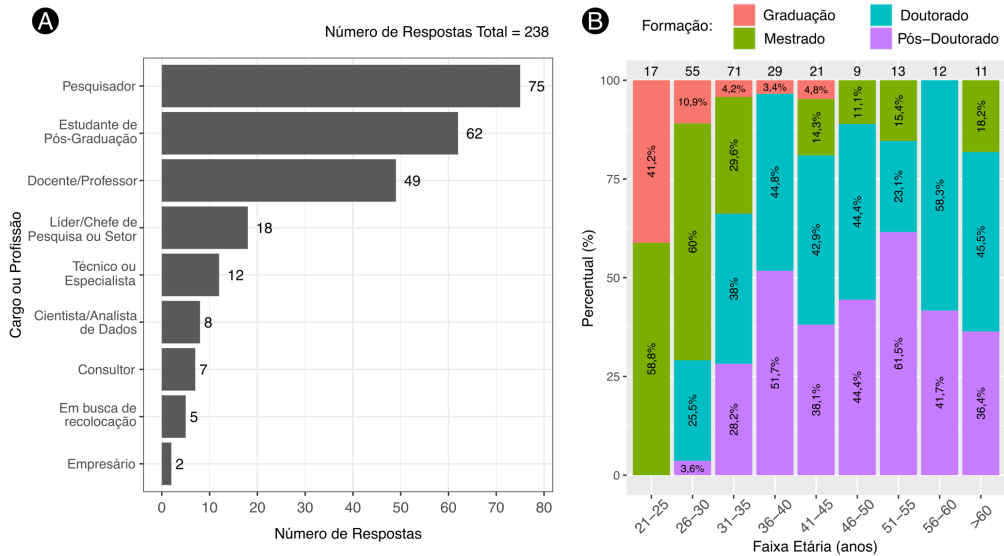


Figura 1. Distribuição das 238 respostas. Na parte A estão as diferentes Profissões (Cargos ou Posições) dos que responderam à pesquisa. Na parte B estão as distribuições das faixas etárias (em anos) e a formação do profissional (última titulação obtida: graduação, mestrado, doutorado e pós-doutorado). Observação: mesmo que “pós-doutorado” não seja considerado como um grau de titulação, incluiu-se na pesquisa a título de informação, bastando somar as quantidades de Doutorado + Pós-Doutorado para um resultado mais condizente com as titulações.

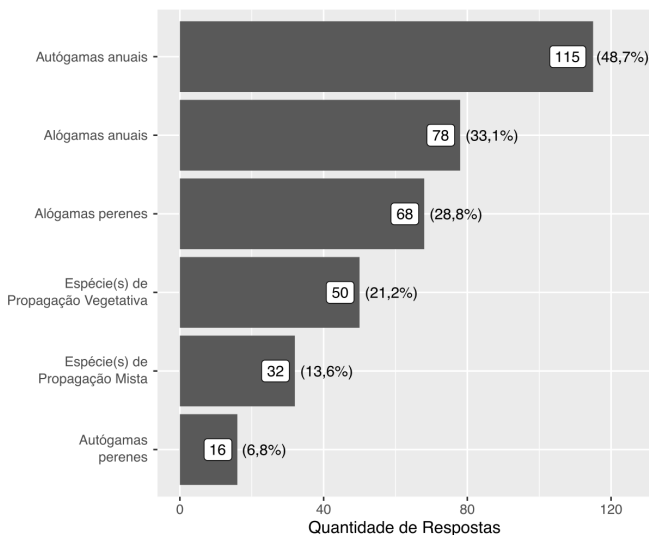


Figura 2. Distribuição das respostas por cultivo agrícola de atuação do profissional. Observação: um determinado profissional pode trabalhar com mais de uma espécie ao mesmo tempo.

Inovação e Empreendedorismo no Melhoramento de Precisão

O desenvolvimento tecnológico e a inovação têm um papel fundamental na otimização da equação do melhorista. Novas tecnologias podem ser aplicadas para aprimorar a coleta de dados, melhorar a precisão das medições fenotípicas e genotípicas, e acelerar o processo de melhoramento. Além disso, a inovação em métodos de análise e interpretação de dados genômicos e fenotípicos pode fornecer *insights* para identificar marcadores genéticos relevantes e entender as interações entre genes e características de interesse. Na Tabela 1 é revelado um panorama sobre o conhecimento dos entrevistados em relação às startups relacionadas ao melhoramento de plantas. Alguns entrevistados afirmam fazer parte ou serem proprietários de startups nesse campo, o que sugere participação ativa no setor. Além disso, uma parcela dos entrevistados demonstrou ter conhecimento específico sobre startups voltadas para o melhoramento de plantas, indicando familiaridade com o tema. No entanto, a grande maioria dos entrevistados (quase 58%) afirmou não conhecer nenhuma startup nesse âmbito. Esse resultado ressalta a necessidade de aumentar a divulgação e conscientização sobre as startups que atuam no melhoramento genético de plantas, a fim de promover maior compreensão e envolvimento dos profissionais nessa área de inovação.

Tabela 1. Resposta dos participantes à pergunta: “Você conhece Startups (ie., uma empresa emergente e inovadora que busca desenvolver um modelo de negócios aplicado) voltadas para o Melhoramento Genético de Plantas?”.

Respostas	Frequência
Sim, inclusive faço parte de uma (ou sou proprietário)	5 (2,10%)
Sim, conheço uma ou mais Startups específicas para o Melhoramento	46 (19,33%)
Sim, conheço uma ou mais Startups com aptidão para o Melhoramento	49 (20,59%)
Não. Sei que existem Startups, mas não conheço nenhuma voltada para o Melhoramento de Plantas	138 (57,98%)
	238 (100%)

Os profissionais que responderam conhecer ou pertencer a startups mencionaram que essas estão envolvidas em uma ampla variedade de áreas relacionadas ao melhoramento genético de plantas, e atuam desde a avaliação de ensaios de valor de cultivo e uso, prestação de serviços de sequenciamento de DNA e mapeamento

genético, análises estatísticas e desenvolvimento de variedades, transgenia, soluções de fenotipagem, análise de dados, edição genômica e produção de sementes. Além disso, também foram mencionadas startups que atuam em áreas específicas, como reflorestamento, polinização via drones, cultivo em ambientes controlados, e várias outras atividades diretamente ou indiretamente relacionadas ao melhoramento de plantas.

Habilidades Tecnológicas Pertencentes ao Pipeline do Melhoramento de Precisão

A rotina de trabalho do melhorista vegetal moderno é marcada pela utilização de diversas tecnologias de coleta e obtenção de dados, tais como fenotipagem automatizada, bioinformática, genotipagem via milhares de marcadores do tipo SNP (*Single Nucleotide Polymorphism*), caracterização ambiental, simulação de dados e integração com a “Internet das Coisas” (IoT – Internet of Things). Além disso, é importante dominar linguagens de programação como R, Python, C, C++ ou C#, bem como utilizar plataformas de hospedagem de código e gerenciamento de bancos de dados. Os procedimentos de análise de dados incluem estatísticas descritivas, testes de médias, estatística multivariada, modelos lineares e não lineares, métodos bayesianos e técnicas de inteligência artificial. Essas abordagens são fundamentais para o avanço do melhoramento genético de plantas e permitem uma análise mais precisa e eficiente dos dados coletados (Figura 3).

Foi perguntado quais tecnologias de coleta/obtenção de dados fazem parte da rotina de trabalho dos profissionais do melhoramento, e as possibilidades de resposta foram: fenotipagem em larga escala/automatizada com sensores embarcados em drones, robôs terrestres ou plataformas fixas. Além disso, a genotipagem via SNP usando *Next Generation Sequencing* (NGS) com Chips ou *Arrays* e a bioinformática com atividades como alinhamento de sequências, identificação de SNPs e anotação de genomas também foram mencionadas. Outras opções foram a caracterização ambiental com estações meteorológicas e dados de satélites orbitais; a simulação de dados usando técnicas como *Monte Carlo*, *Bootstrap* e Redes Bayesianas; bem como sistemas integrados com IoT. Para essa pergunta, foi possível selecionar mais de uma resposta (Figura 3-A).

Foi perguntado sobre o domínio de linguagens de programação e o uso de plataformas de hospedagem de código e de gerenciamento de banco de dados. As possíveis respostas incluíram o conhecimento avançado ou básico nas seguintes linguagens: R, Python, C, C++ ou C#. Também foi possível selecionar se utilizavam

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

plataformas de hospedagem e compartilhamento de códigos (por exemplo, *GitHub*), plataformas de gerenciamento de dados (por exemplo, SQL), ou se o profissional não utilizava ou desconhecia as opções mencionadas. Era permitido marcar mais de uma alternativa. Os participantes da pesquisa puderam incluir outras linguagens de programação que porventura utilizassem em suas rotinas de trabalho (Figura 3-B).

Na sequência, foi perguntado quais procedimentos de análise de dados são utilizados. As possibilidades de respostas incluíram desde análises clássicas e essenciais, como estatísticas descritivas e análises gráficas, ANOVA juntamente com testes de médias (como os testes Tukey e Scott-Knott), estatística multivariada (como Análise de Componentes Principais – PCA, Análise Fatorial – FA e análises de agrupamento), modelos lineares (como análise de regressão simples, quadrática e cúbica), modelos não lineares (como Sigmoidal Logístico, Gompertz e Weibull), modelos mistos (lineares e não lineares), até métodos bayesianos e técnicas de inteligência artificial (tais como aprendizado de máquina, redes neurais, lógica *Fuzzy* e *Random Forest*). Foi possível marcar mais de uma alternativa (Figura 3-C).

Com base nos dados coletados com os participantes da pesquisa, é possível discutir algumas tendências relacionadas às tecnologias de coleta de dados (Figura 3-A). É interessante notar que a genotipagem em larga escala (via SNP) é amplamente utilizada, apesar de seu alto custo quando avaliados centenas a milhares de marcadores. Isso pode ser atribuído ao seu potencial para fornecer informações importantes sobre a composição genética de indivíduos e populações. Por outro lado, a subutilização da ambientipagem via sensores orbitais, que muitas vezes é gratuita, pode ser devido à falta de conhecimento em SIG ou infraestrutura adequada para acessar, processar e interpretar esses dados (Resende et al., 2021). A bioinformática, embora tenha a necessidade de alta especialização, é uma área bem explorada, destacando sua importância como ferramenta para análise de dados genéticos e genômicos. Quanto à fenotipagem, a preferência pelo uso de drones em comparação com plataformas terrestres pode ser atribuída à sua capacidade de fornecer uma visão abrangente e detalhada das características fenotípicas, permitindo cobertura eficiente de grandes áreas, além de ainda serem mais acessíveis em termos de custo (Reynolds et al., 2019). No entanto, é possível especular que a adoção de plataformas terrestres pode aumentar à medida que novas tecnologias e abordagens sejam desenvolvidas e se popularizem, possibilitando maior diversidade e precisão na coleta de dados fenotípicos.

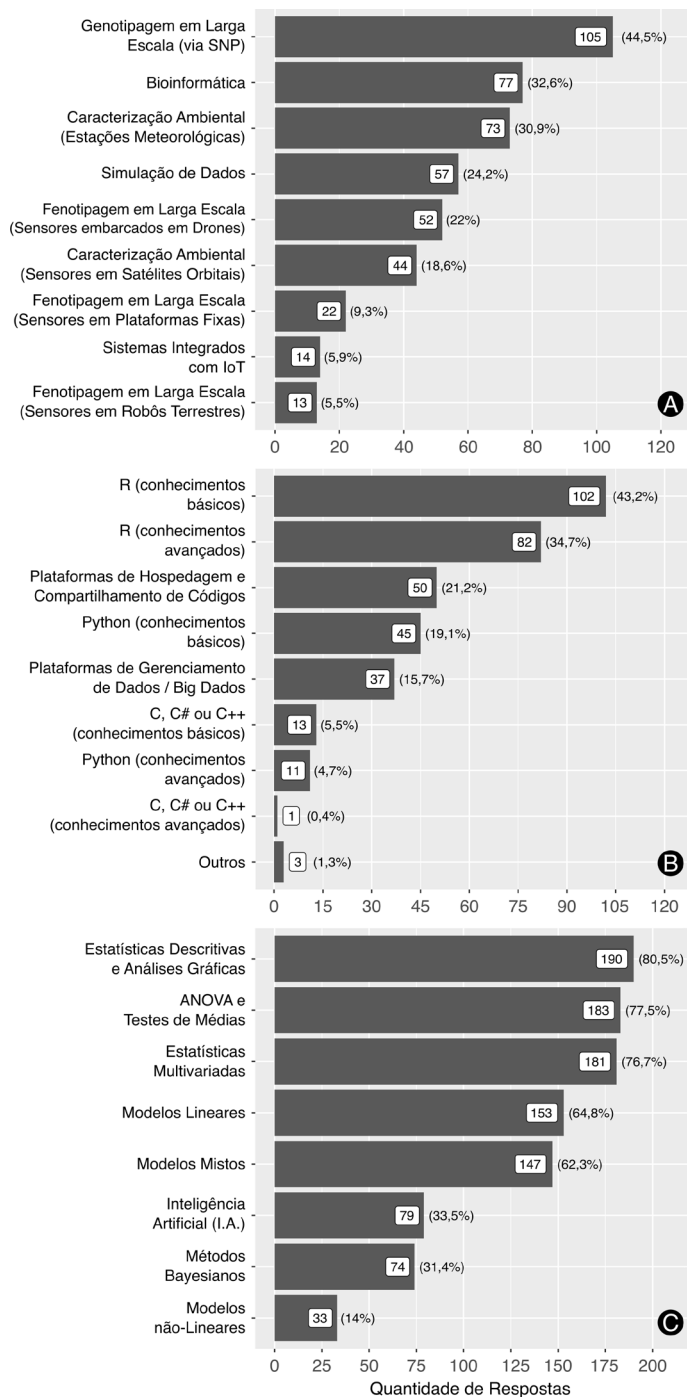


Figura 3. Resposta sobre quais ferramentas tecnológicas são utilizadas por profissionais do melhoramento genético de plantas. **Parte A**, respostas sobre conhecimento e utilização de ferramentas de aquisição de dados. **Parte B**, respostas sobre conhecimento e utilização de ferramentas computacionais de programação, manejo e gerenciamento de dados. **Parte C**, respostas sobre conhecimento e utilização de ferramentas de manipulação e análise dos dados.

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

Com base nas respostas obtidas sobre domínios de linguagens computacionais (Figura 3-B), é evidente o predomínio do uso da linguagem R. No entanto, há pouca adesão a outras facilidades, como plataformas de hospedagem e gerenciamento de códigos, bem como plataformas de gerenciamento de dados e *big data*. Além disso, a baixa utilização de linguagens de nível básico, como C, C# ou C++, pode ser atribuída à sua complexidade e necessidade de habilidades especializadas. Surpreendentemente, embora o Python seja atualmente uma unanimidade no mercado de trabalho mundial (Afzal et al., 2020), profissionais brasileiros do melhoramento de plantas exploram pouco essa linguagem, talvez devido à falta de conscientização sobre suas capacidades ou à preferência por ferramentas mais difundidas, o R. No entanto, com a crescente demanda global por habilidades em Python, é possível que sua adoção aumente no contexto do melhoramento de plantas no futuro, à medida que suas vantagens e aplicabilidades forem mais reconhecidas e exploradas.

Com base nas respostas observadas sobre o uso de ferramentas de análise de dados na pesquisa realizada (Figura 3-C), é interessante notar que a maioria dos profissionais emprega análises básicas, como estatísticas descritivas, análises gráficas e a ANOVA. No entanto, alguns profissionais afirmaram usar ferramentas robustas, como estatística Bayesiana e Inteligência Artificial, mas curiosamente não fazem uso das ferramentas básicas, que deveriam ser fundamentais para qualquer estudo. Esse fenômeno pode indicar a falta de compreensão da importância das análises básicas para esse grupo. Além disso, é surpreendente constatar que, embora se esperasse que os Métodos Bayesianos estivessem recebendo atenção equivalente à adoção dos Modelos Mistos, quase o dobro de profissionais ainda preferem os clássicos Modelos Mistos em comparação aos métodos baseados no teorema de Bayes, resultado esse também observado anteriormente por Van Eeuwijk et al. (2016). Isso pode refletir a preferência por abordagens estatísticas tradicionais e uma possível resistência à adoção de métodos estatísticos mais especializados/avançados. No entanto, é importante que os programas de pós-graduação em genética e melhoramento de plantas promovam a utilização e a compreensão das ferramentas básicas e robustas de análise de dados, incentivando a aplicação adequada dessas técnicas para obter resultados mais precisos e confiáveis.

A mineração, o manejo de dados e o uso de *big data* também desempenham importante papel na otimização da equação do melhorista. Com o aumento da disponibilidade de dados genéticos e fenotípicos em larga escala, é possível utilizar técnicas avançadas de análise de dados para identificar padrões, correlações e relações complexas entre caracteres. Isso permite a seleção mais precisa e eficiente de plantas superiores, levando a um ganho genético mais significativo.

Os conceitos da Pesquisa Operacional (PO) e a otimização matemática de processos podem contribuir para otimizar a equação do melhorista ao aplicar técnicas matemáticas e algoritmos avançados para encontrar a melhor combinação de variáveis e parâmetros no processo de seleção de plantas. Por meio da modelagem matemática e da simulação é possível explorar diferentes cenários e identificar estratégias ótimas que maximizem o ganho genético. Pela Figura 4, observa-se que Técnicos e Especialistas e, em especial, Cientistas/Analistas de Dados, são os profissionais que mais têm contato e utilizam a Pesquisa Operacional. No entanto, pouco ainda é explorado nessa área de conhecimento, que é muito utilizada por profissionais da área de engenharia de processos. O capítulo 6 deste livro é dedicado a PO e possui exemplos que podem ajudar no *pipeline* de atividades do melhorista de plantas.

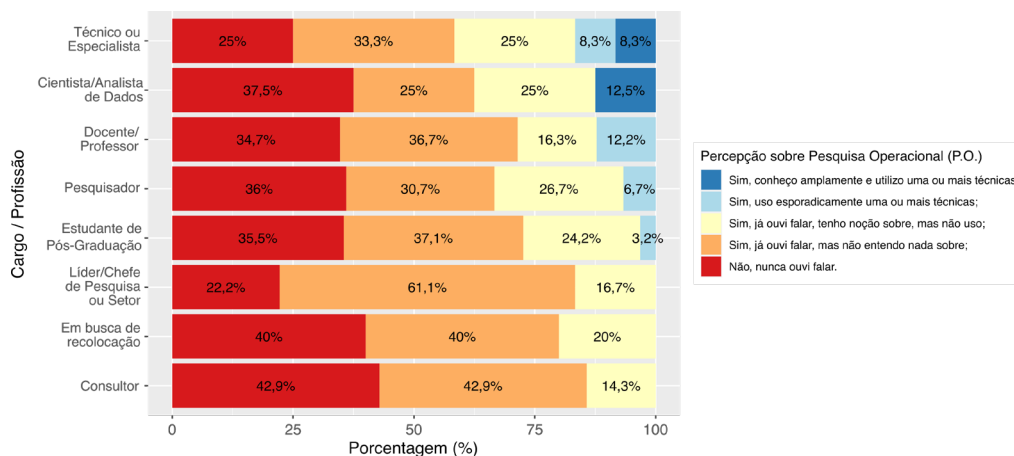


Figura 4. Percepção ou entendimento dos profissionais sobre a Pesquisa Operacional (PO). As respostas foram subdivididas entre possíveis cargos ou profissões dos entrevistados.

A Internet das Coisas (IoT), apresentada no capítulo 7, desempenha um papel fundamental no aprimoramento da rotina do melhorista de plantas, proporcionando vantagens significativas na coleta e monitoramento de dados. Ao possibilitar a obtenção de informações em tempo real sobre as plantas, em diferentes ambientes e condições, a IoT viabiliza o monitoramento contínuo e em larga escala das características fenotípicas de interesse. Esse acesso em tempo real a dados precisos e atualizados é de extrema importância para a análise fenômica, permitindo uma compreensão mais abrangente e detalhada do desempenho das plantas em resposta a diversos estímulos. Além disso, essas informações em tempo real fornecem *insights* valiosos para o ajuste de estratégias de seleção genética, possibilitando uma abordagem

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

mais precisa e adaptável ao processo de melhoramento genético. Dessa forma, a IoT também potencializa a equação do melhorista, elevando a precisão das decisões tomadas e impulsionando o desenvolvimento de variedades de plantas superiores em termos de produtividade, resistência a insetos-pragas e doenças, e adaptação a diferentes condições climáticas.

Procedimentos de Inteligência Artificial e *Machine Learning* (aprendizado de máquina), temas apresentados no capítulo 5, podem ser empregados para melhorar a acurácia da seleção, a previsão de características desejáveis e a tomada de decisões no processo de melhoramento. Algoritmos de aprendizado de máquina podem aprender a partir dos dados históricos e identificar padrões complexos, permitindo a identificação de genótipos promissores e a realização de seleção assistida por computador de forma mais precisa e eficiente. Essas tecnologias permitem analisar grandes volumes de dados genéticos e fenotípicos, identificar padrões complexos e fazer previsões precisas sobre o desempenho das plantas. Por meio do treinamento de algoritmos, a IA e o *Machine Learning* podem auxiliar na classificação de plantas com base em características desejáveis, na detecção de doenças ou estresses ambientais e na identificação de cruzamentos promissores. Isso agiliza o processo de seleção, tornando-o mais eficiente e permitindo a identificação de combinações genéticas com alto potencial de sucesso. Além disso, a IA e o *Machine Learning* também possibilitam a automação de tarefas rotineiras, como a análise de dados e a geração de relatórios, liberando tempo e recursos para os melhoristas se concentrarem em atividades estratégicas e de tomada de decisão.

As métricas fenotípicas e a análise fenômica, apresentadas no capítulo 9, desempenham um papel cada vez mais relevante na otimização da equação do melhorista, pois fornecem informações detalhadas e abrangentes sobre as características expressas pelas plantas. Para obter informações ainda mais detalhadas sobre as plantas, os melhoristas modernos recorrem ao uso de drones e sensoriamento remoto. O sensoriamento remoto por meio de drones permite a captura de dados precisos e em alta resolução, tanto em nível macro (em perspectiva geográfica) quanto micro (em perspectivas de parcelas experimentais ou de organismo), ao monitorar de perto as plantas e seu ambiente. Essa abordagem proporciona uma visão abrangente das características fenotípicas, incluindo aspectos como altura, densidade de folhas, vigor e tolerância ao déficit hídrico. Também é possível identificar não apenas características visíveis, mas também variações relacionadas ao metabolismo, fisiologia e respostas relacionadas com a interação ambiental. Essas informações detalhadas são relevantes para compreender as relações entre o fenótipo e o genótipo, permitindo uma seleção mais precisa e embasada nas características desejadas. Além disso,

a análise fenômica possibilita a identificação de marcadores fenotípicos associados a características de interesse, acelerando o processo de seleção e melhorando a eficiência do melhoramento genético de plantas.

O Speed Breeding (SB) ou melhoramento acelerado, apresentado no capítulo 11, é uma abordagem inovadora no processo de melhoramento, combinando condições controladas de crescimento, indutores químicos e climatização ambiental para acelerar o florescimento e desenvolvimento das plantas. Isso permite obter múltiplas gerações em curto prazo, facilitando a seleção precoce de características desejáveis e reduzindo o tempo necessário para seleções sucessivas. O SB aumenta a eficiência do processo de melhoramento, impulsionando o desenvolvimento de variedades agrícolas superiores adaptadas a diferentes condições e necessidades. Na Figura 5 é mostrado o resultado da resposta dos participantes da enquete ao serem questionados sobre SB, estratificando-se por diferentes posições ocupacionais. Note-se que Líderes de Pesquisa, em especial de empresas privadas, têm maior conhecimento e experiência na aplicação dessas técnicas.

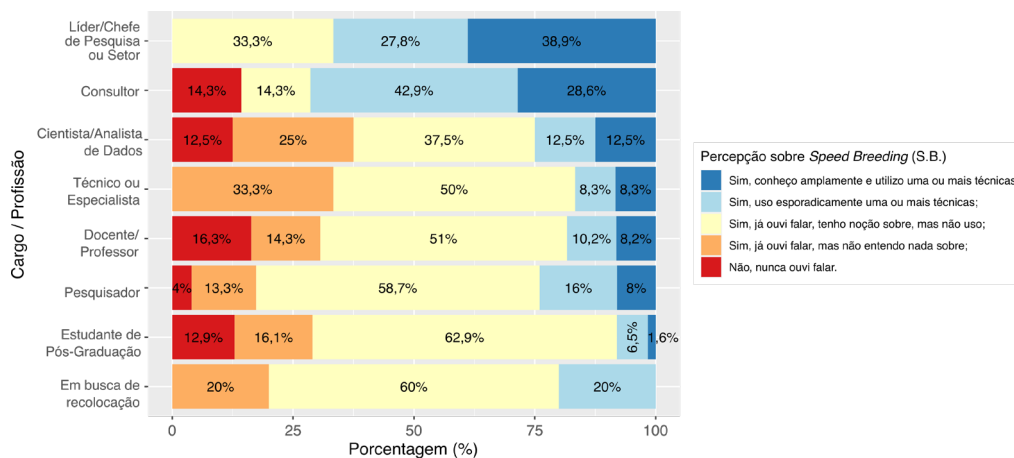


Figura 5. Percepção ou entendimento dos profissionais sobre Speed Breeding (SB). As respostas foram subdivididas entre possíveis cargos ou profissões dos entrevistados.

A bioinformática desempenha um papel importante na análise e interpretação dos dados genômicos, permitindo a identificação de locos candidatos e a compreensão de suas funções biológicas, ademais, auxilia diretamente no desenvolvimento de tecnologias de genotipagem, que hoje já fazem parte da rotina de melhoramento das instituições de sucesso no melhoramento vegetal. Essas informações genéticas e

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

genômicas, tema apresentado no capítulo 10, contribuem para aprimorar a seleção de plantas superiores e otimizar a equação do melhorista. Este livro abordará tanto o desenvolvimento de marcadores modernos via genotipagem de larga-escala, como a aplicação desses marcadores em complexos e eficientes modelos de seleção genômica ampla.

A edição gênica (capítulo 12), como parte da engenharia genética de precisão, juntamente com as novas tecnologias de melhoramento genético, apresenta um potencial revolucionário para otimizar a equação do melhorista. Por meio de técnicas avançadas, como CRISPR-Cas9, é possível realizar modificações precisas no genoma das plantas, permitindo alterar características específicas de maneira mais rápida e direta. Essas tecnologias têm o poder de acelerar significativamente o processo de melhoramento, possibilitando a geração mais eficiente de plantas com características desejáveis. No entanto, poucos entrevistados disseram fazer uso dessas ferramentas em sua rotina de trabalho (Figura 6), provavelmente devido a uma combinação de fatores, como restrições regulatórias, e limitações de acesso a recursos e tecnologias avançadas, e, eventualmente, restrições pessoais ao uso de organismos geneticamente editados. É importante promover discussões e buscar soluções para que as ferramentas de edição gênica possam ser mais amplamente adotadas no melhoramento genético de plantas, impulsionando assim a inovação e o avanço na agricultura.

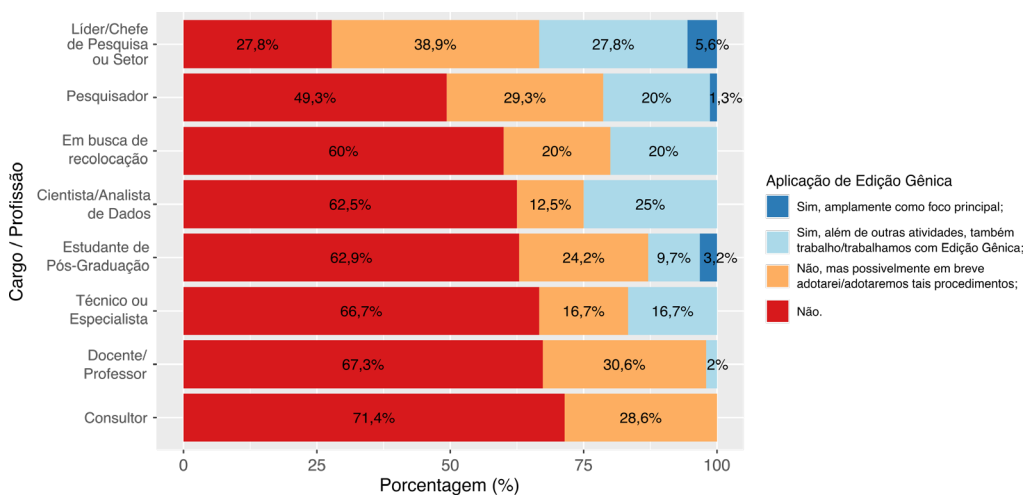


Figura 6. Aplicação de técnicas de Edição Gênica (engenharia genética de precisão) pelos profissionais de melhoramento genético de plantas. As respostas foram subdivididas entre possíveis cargos ou profissões dos entrevistados.

A integração de Sistemas de Informações Geográficas e técnicas de geoprocessamento também apresentam grande potencial na otimização da equação do melhorista, permitindo uma abordagem avançada conhecida como geoinformática. Por meio da análise de dados ambientais e da aplicação de modelos espaciais, é possível compreender a interação de genótipos com ambientes ($G \times A$) e identificar regiões ou ambientes específicos onde certos genótipos são mais adaptados. A análise ambientômica, apresentada no capítulo 8, combinada com a identificação de marcadores genéticos associados à adaptabilidade, oferece uma base sólida para a seleção de plantas com desempenho otimizado em diferentes condições ambientais (Resende et al., 2021). Dessa forma, o uso de técnicas de SIG potencializa a capacidade dos melhoristas em direcionar seus esforços de seleção, maximizando assim o ganho genético em ambientes variados e contribuindo para o desenvolvimento de variedades vegetais mais resilientes e adaptadas às demandas regionais.

Concatenando todas as áreas de conhecimento anteriores, tem-se uma gestão otimizada do programa de melhoramento genético, sendo fundamental para maximizar o potencial da equação do melhorista. Isso envolve a definição de objetivos claros, o planejamento estratégico das atividades de seleção e cruzamento, a alocação eficiente de recursos financeiros e a utilização adequada das ferramentas tecnológicas disponíveis. Uma gestão eficiente e otimizada garante que todas as etapas do processo de melhoramento sejam integradas de forma sinérgica, potencializando o ganho genético obtido.

Habilidades Tecnológicas analisadas conjuntamente

A partir das respostas dos profissionais a respeito da utilização das tecnologias incluídas no questionário, foi elaborado o índice de habilidades tecnológicas, obtido a partir da soma, por indivíduo, das tecnologias utilizadas considerando ferramentas tecnológicas (drones, técnicas de ambientipagem, *Next-Generation Sequencing* - NGS, Inteligência Artificial, CRISPR, Pesquisa Operacional, R, Python, Estatística Bayesiana, etc.). Os índices foram agrupados por Faixa Etária e Cargo ou Profissão (Figura 7) e por Tipo e segmento institucional (Figura 8).

O índice mostrou que profissionais com idades entre 31 e 45 anos apresentaram maior conhecimento e habilidades com as novas tecnologias apresentadas, e é um reflexo do contexto atual em que essas tecnologias estão se tornando cada vez mais integradas à rotina de trabalho no campo do melhoramento genético de plantas (Figura 7-A). Essa faixa etária pode ter sido exposta a essas tecnologias durante sua formação acadêmica ou posteriormente, permitindo-lhes adquirir habilidades e conhecimentos atualizados. É

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

importante ressaltar que isso não significa que profissionais mais jovens (21-30 anos) ou mais experientes (acima de 46 anos) não possuam habilidades relevantes ou estejam excluídos do uso dessas tecnologias. No entanto, é possível que profissionais mais jovens estejam em um estágio inicial de suas carreiras, ainda adquirindo experiência e aprofundando seus conhecimentos nessas áreas específicas, ou mesmo devido a currículos defasados de algumas universidades. Por outro lado, profissionais mais antigos podem ter tido menos exposição às tecnologias mais recentes por estarem em um estágio na carreira em que tal conhecimento não é exigido ou necessário. É possível ainda, verificar uma variação muito grande dentro de cada classe, mostrando que a faixa etária é apenas um dos fatores que influenciam nas habilidades tecnológicas.

Diante disso, é importante que seja oportunizado aos profissionais a educação continuada e o compartilhamento de conhecimentos entre profissionais de diferentes faixas etárias. A colaboração entre gerações pode enriquecer a aplicação dessas tecnologias no melhoramento genético de plantas, combinando a experiência e conhecimento dos profissionais mais experientes com a familiaridade e afinidade dos profissionais mais jovens com as tecnologias emergentes. Essa troca de conhecimentos e experiências contribui para a evolução e inovação contínuas no agronegócio, capacitando as equipes de melhoramento genético a aproveitarem todo o potencial dessas ferramentas para impulsionar os avanços na agricultura e atender às demandas crescentes por alimentos e demais insumos vegetais.

Os resultados da pesquisa também indicam que profissionais das ciências de dados (tema apresentado no capítulo 4) e líderes de pesquisa apresentaram maiores afinidades com as novas tecnologias apresentadas no questionário, o que é coerente com a natureza de suas atribuições profissionais e o contexto em que atuam (Figura 7-B). Profissionais das ciências de dados, por estarem diretamente envolvidos na análise e interpretação de informações complexas, estão mais familiarizados com a utilização de tecnologias avançadas, incluindo aquelas relacionadas ao melhoramento genético de plantas. Líderes de grupos de pesquisa também são frequentemente responsáveis por supervisionar projetos de pesquisa e desenvolvimento, o que exige conhecimento e compreensão das tecnologias mais recentes disponíveis. Além disso, eles têm um papel importante na promoção da inovação e adoção de novas abordagens em suas equipes de trabalho.

Por outro lado, docentes/professores, estudantes de pós-graduação, consultores e profissionais em busca de recolocação podem não ter a mesma exposição diária a essas tecnologias ou a mesma demanda em seus respectivos papéis. Docentes/professores podem estar mais focados na transmissão de conhecimentos gerais e fundamentos teóricos, enquanto estudantes de pós-graduação estão em um estágio

de aprendizado e aprofundamento de conceitos básicos, antes de se especializarem em áreas mais específicas. Consultores e profissionais em busca de colocação profissional podem não estar diretamente envolvidos em projetos de pesquisa ativos ou não terem acesso às mesmas oportunidades de atualização de habilidades. Ter conhecimento em melhoramento de precisão pode aumentar a chance de conseguir posições mais importantes e de maior responsabilidade, pois o gerenciamento nessa área requer um entendimento sólido dos princípios genéticos e moleculares subjacentes. Gerenciar um programa de melhoramento de precisão envolve liderar equipes de profissionais com especialidades distintas e complementares, e, portanto, é fundamental ter conhecimento para coordenar os esforços, estabelecer metas claras e garantir a execução eficiente das atividades do programa. Por fim, gerenciar um programa de melhoramento de precisão envolve ter uma visão estratégica para identificar oportunidades de melhoria de processos e inovação, bem como antecipar tendências e desafios futuros na área. A capacidade de desenvolver e implementar estratégias inovadoras para o melhoramento de precisão pode ser um grande diferencial de mercado e, com isso, abrir portas para empregos mais importantes e desafiadores.

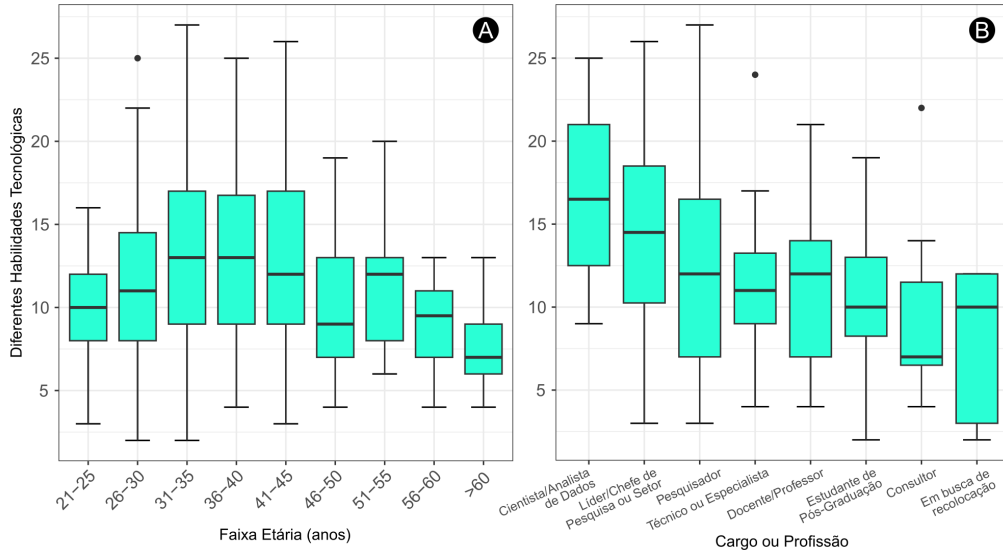


Figura 7. Boxplots mostrando a distribuição dos índices de Habilidades Tecnológicas (eixo y) entre diferentes atributos da pesquisa. Na parte **A** é mostrada a aptidão para diferentes habilidades tecnológicas de acordo com as diferentes idades dos respondentes da pesquisa. Na parte **B** é mostrada a aptidão para diferentes habilidades tecnológicas de acordo com as profissões (ou cargo ou posição) dos respondentes da pesquisa.

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

Integrações e parcerias entre universidades e empresas privadas foram abordadas, solicitando-se a percepção ou opinião dos participantes (Tabela 2). As respostas permitiam marcar até duas opções, incluindo: 1) parcerias para pesquisas de base sem ganhos operacionais diretos para as empresas; 2) parcerias com ganhos significativos para a(s) empresa(s); 3) parcerias sem contrapartida financeira ou com contrapartida mínima; 4) incentivo a muitas colaborações dessa natureza; 5) desaprovação devido ao alto custo e resultados insatisfatórios; 6) preferência pela realização exclusiva de Pesquisa e Desenvolvimento (P&D) pela empresa; e 7) avaliação cautelosa e imposição de restrições quando necessário.

Tabela 2. Percepção dos entrevistados sobre integrações/parcerias entre universidades com empresas privadas.

Percepção	Respostas	
Deveriam ser muito incentivadas. Há espaço/campo para muitas parcerias dessa natureza	197	82,77%
Deveriam ser feitas. Mesmo que apenas para pesquisas de base (sem ganhos operacionais diretos para as empresas)	91	38,24%
Estas parcerias devem ser avaliadas com muito cuidado, impondo-se restrições sempre que necessário	20	8,40%
Deveriam ser feitas. Porém, apenas se houver ganhos significativos para a(s) Empresa(s)	18	7,56%
Deveriam ser feitas. Porém, sem contrapartida financeira por parte da empresa, ou com contrapartida mínima	9	3,78%
Não deveriam ser feitas. Em geral, gasta-se muito dinheiro e raramente há resultados relevantes para ambas as partes	0	0,00%
Não deveriam ser feitas. Os profissionais/pesquisadores do quadro de funcionários da empresa devem atender toda demanda de P&D	0	0,00%

Com base nos resultados da pesquisa, observou-se que profissionais de instituições privadas demonstraram maior afinidade e proximidade com as novas tecnologias discutidas aqui, em comparação com profissionais de instituições públicas e profissionais autônomos (Figura 8-A). Essa diferença pode ser atribuída a diversos fatores, como o acesso a recursos e investimentos em tecnologia, a cultura organizacional voltada para a inovação e a ênfase na competitividade do mercado. Além disso, um paralelo interessante pode ser feito em relação aos setores de pesquisa, especialmente dentro dessas instituições privadas (Figura 8-B). É provável que esses setores estejam mais

envolvidos e tenham rotinas mais intensas com atividades tecnológicas devido à natureza de sua função, que requer constantes avanços científicos e tecnológicos para se manterem competitivos e atenderem às demandas do mercado.

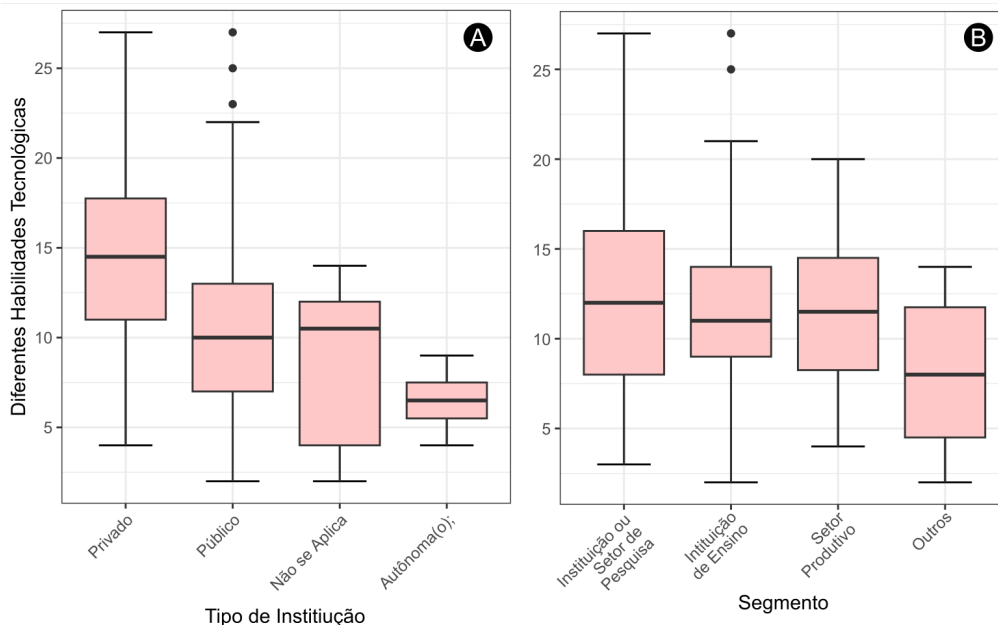


Figura 8. Boxplots mostrando a distribuição dos índices de Habilidades Tecnológicas (eixo y) entre diferentes atributos da pesquisa. Na parte A é mostrada a aptidão para diferentes habilidades tecnológicas de acordo com os diferentes tipos de instituição dos respondentes da pesquisa. Na parte B é mostrada a aptidão para diferentes habilidades tecnológicas de acordo com os segmentos institucionais dos respondentes da pesquisa.

No entanto, é importante ressaltar que essas observações são baseadas nos resultados específicos dessa pesquisa, e outras variáveis podem influenciar a relação entre o setor de atuação e a proximidade com as novas tecnologias. Mais estudos são necessários para aprofundar essa compreensão e considerar outros fatores que podem estar em jogo, como o nível de investimento em pesquisa e desenvolvimento, as políticas governamentais e o acesso a recursos tecnológicos.

Com base nos resultados, a maioria dos participantes (quase 83%) apoia fortemente as parcerias entre universidades e empresas privadas, reconhecendo seu potencial e importância. Cerca de 38,3% concordam que as parcerias podem ser feitas para pesquisas de base, mesmo sem benefícios operacionais diretos. Uma minoria (7,6%)

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

destaca a necessidade de ganhos significativos para as empresas, enquanto outros (8,4%) enfatizam a importância da avaliação cuidadosa e da imposição de restrições, quando apropriado. Não houve participantes que se opuseram às parcerias ou acreditassem que a empresa deveria assumir toda a responsabilidade por P&D, o que sugere um amplo apoio à colaboração entre universidades e empresas privadas para impulsionar a pesquisa e o desenvolvimento.

As universidades brasileiras têm o potencial para formar profissionais aptos para o melhoramento de precisão, por fornecer conhecimentos teóricos e práticos em áreas relevantes, como genética, biologia molecular, estatística, dentre outras. Contudo, é importante ressaltar que as habilidades específicas em melhoramento de precisão, como o uso de técnicas moleculares avançadas e ferramentas de genômica, geralmente requerem um treinamento adicional e prática diária. Nesse sentido, instituições como a Empresa Brasileira de Pesquisa Agropecuária (Embrapa) e outros centros de pesquisa e desenvolvimento no Brasil desempenham um papel importante na formação de profissionais em diversas áreas, incluindo melhoramento de precisão. Por meio de programas de treinamento, workshops, cursos e colaborações, a Embrapa e outras instituições de pesquisa podem fornecer aos profissionais a oportunidade de adquirir habilidades práticas e conhecimentos especializados em melhoramento de precisão. Por fim, é importante destacar que o aprendizado contínuo e o acompanhamento dos avanços tecnológicos e científicos são essenciais para se manter atualizado em melhoramento de precisão, uma vez que técnicas e ferramentas estão em constante evolução.

O Brasil tem presença significativa e reconhecida na área de melhoramento de plantas, e possui instituições de pesquisa e universidades renomadas em ciências agrícolas, as quais contribuem para os avanços científicos e tecnológicos nessa área. No Brasil, existem professores e pesquisadores qualificados e experientes em melhoramento genético, alguns já com trabalhos em melhoramento de precisão, e, desse modo, o Brasil já possui expertise para ampliar e disseminar o conhecimento na área. No entanto, é importante observar que, dependendo do nível de especialização e das necessidades específicas, alguns pesquisadores brasileiros podem buscar oportunidades de treinamento e colaborações internacionais para aprimorar suas habilidades em melhoramento de precisão.

Considerações Finais

Os avanços tecnológicos e as novas ferramentas estão revolucionando o campo do melhoramento genético de plantas. Como foi observado ao longo da discussão

baseada nos resultados do questionário, ficou evidente o potencial transformador dessas tecnologias na otimização da equação do melhorista. A partir da integração de técnicas, como a edição gênica, sistemas de informações geográficas, bioinformática e o uso de *big data*, os profissionais em melhoramento genético vegetal têm acesso a uma gama cada vez maior de informações genéticas e fenotípicas. Isso possibilita a identificação precisa de características-chave, a compreensão de suas relações com o genótipo e a seleção mais eficiente de plantas superiores. Além disso, a incorporação de novas abordagens, como o Speed Breeding, o uso de drones e o sensoriamento remoto, proporciona a coleta de dados em tempo real sobre as plantas em diferentes ambientes e condições. Isso amplia o conhecimento sobre as características fenotípicas relevantes e permite o ajuste de estratégias de seleção, resultando em um ganho genético mais significativo. No entanto, é importante ressaltar que a adoção dessas tecnologias não se limita apenas à atualização do conhecimento. A capacitação e a formação contínua dos profissionais são fundamentais para aproveitar ao máximo essas ferramentas e promover um melhoramento genético de sucesso.

Apesar dos progressos, é necessário um cuidado na interpretação dos resultados da pesquisa, pois existem variáveis adicionais que podem influenciá-los, como a demanda do mercado de trabalho e a competitividade geral do setor. Portanto, mais estudos são necessários para uma compreensão abrangente e precisa das relações entre as novas tecnologias e as diferentes categorias de profissionais. Em conclusão, as inovações tecnológicas discutidas neste capítulo têm o potencial de impulsionar o campo do melhoramento genético de plantas, permitindo a seleção mais precisa, eficiente e adaptada às necessidades específicas. A atualização constante e a incorporação dessas ferramentas prometem avanços significativos no desenvolvimento de cultivares superiores, contribuindo para a segurança alimentar e a sustentabilidade agrícola, diante de um cenário de aumento de risco devido às mudanças climáticas atuais e futuras.

Este livro apresenta doze capítulos que, no nosso entendimento, fazem parte do que consideramos atualmente “melhoramento de precisão”. Esta obra foi pensada para alcançar interessados com nível de formação variado, desde estudantes de graduação até profissionais atuantes em melhoramento. Para facilitar a compreensão e entendimento, esta obra foi elaborada em língua portuguesa, fornecendo um panorama geral das áreas abordadas, e informações que permitem que quem quiser se aprofundar no tema de interesse, terá condições de fazê-lo. Esperamos com isso contribuir para que os profissionais possam conhecer e trazer esse conhecimento para o seu dia a dia, e melhorar o resultado de seu trabalho. Para atuar em melhoramento de precisão, o profissional deve adquirir conhecimentos e habilidades adicionais relacionados a

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

técnicas moleculares, análise de dados obtidos pelas ômicas, e dominar ferramentas avançadas de melhoria genética. Em adição, estudantes de graduação, ao tomarem conhecimento de áreas de estudo até então desconhecidas para eles, poderão eventualmente vir a se especializar naquelas que despertaram curiosidade e afinidade, dando início a uma nova geração de profissionais que irão promover o desenvolvimento do nosso país. Tenham todos uma ótima leitura e resultados!

Referências

- AFZAL, H.; KAMRAN, A.; NOREEN, A. Survival analysis of Python and R within the job market trend. **Journal of Information Technology and Computing**, v. 1, n. 1, p. 31-40, 2020. DOI: <https://doi.org/10.48185/jitc.v1i1.94>.
- BERNARDO, R. Reinventing quantitative genetics for plant breeding: something old, something new, something borrowed, something BLUE. **Heredity**, v. 125, n. 6, p. 375-385, Dec. 2020. DOI: <https://doi.org/10.1038/s41437-020-0312-1>.
- CARENA, M. J.; HALLAUER, A. R.; MIRANDA-FILHO, J. B. **Quantitative genetics in maize breeding**. New York: Springer, 2010. 664 p.
- CLEMENT, C. R. Melhoria de espécies nativas. In: NASS, L. L.; VALOIS, A. C. C.; MELO, I. S.; VALADARES-INGLIS, M. C. (ed.). **Recursos genéticos e melhoria- plantas**. Rondonópolis: Fundação MT, 2001. p. 423-441.
- COBB, J. N.; JUMA, R. U.; BISWAS, P. S.; ARBELAEZ, J. D.; RUTKOSKI, J.; ATLIN, G.; HAGEN, T.; QUINN, M.; NG, E. H. Enhancing the rate of genetic gain in public-sector plant breeding programs: lessons from the breeder's equation. **Theoretical and Applied Genetics**, v. 132, p. 627-645, Mar. 2019. DOI: <https://doi.org/10.1007/s00122-019-03317-0>.
- CROSSA, J.; FRITSCHÉ-NETO, R.; MONTESINOS-LOPEZ, O. A.; COSTA-NETO, G.; DREISIGACKER, S.; MONTESINOS-LOPEZ, A.; BENTLEY, A. R. The modern plant breeding triangle: optimizing the use of genomics, phenomics, and enviromics data. **Frontiers in Plant Science**, v. 12, 651480, Apr. 2021. DOI: <https://doi.org/10.3389/fpls.2021.651480>.
- FERNIE, A. R.; YAN, J. *De novo* domestication: an alternative route toward new crops for the future. **Molecular Plant**, v. 12, n. 5/6, p. 615-631, May 2019. DOI: <https://doi.org/10.1016/j.molp.2019.03.016>.
- HICKEY, L. T.; HAFEEZ, A. N.; ROBINSON, H.; JACKSON, S. A.; LEAL-BERTIOLI, S. C. M.; TESTER, M.; GAO, C.; GODWIN, I. D.; HAYES, B. J.; WULFF, B. B. H. Breeding crops to feed 10 billion. **Nature Biotechnology**, v. 37, p. 744-754, July 2019. DOI: <https://doi.org/10.1038/s41587-019-0152-9>.
- HUMMER, K. E.; HANCOCK, J. F. Vavilovian centers of plant diversity: implications and impacts. **HortScience**, v. 50, n. 6, p. 780-783, Jun. 2015. DOI: <https://doi.org/10.21273/HORTSCI.50.6.780>.
- NAJAFABADI, M. Y.; HESAMI, M.; ESKANDARI, M. Machine learning-assisted approaches in modernized plant breeding programs. **Genes**, v. 14, 777, Mar. 2023. DOI: <https://doi.org/10.3390/genes14040777>.
- PEARSON, S.; CAMACHO-VILLA, T. C.; VALLURU, R.; GAJU, O.; RAI, M. C.; GOULD, I.; BREWER, S.; SKLAR, E. Robotics and autonomous systems for net zero agriculture. **Current Robotics Reports**, v. 3, n. 2, p. 57-64, June 2022. DOI: <https://doi.org/10.1007/s43154-022-00077-6>.
- RAMALHO, M. A. P.; MARQUES, T. L.; LEMOS, R. C. Plant breeding in Brazil: retrospective of the past 50 years. **Crop Breeding and Applied Biotechnology**, v. 21, e383021S3, May/June 2021. DOI: <https://doi.org/10.1590/1984-70332021v21Sa16>.

RESENDE, R. T.; PIEPHO, H.-P.; ROSA, G. J. M.; SILVA-JUNIOR, O. B.; SILVA, F. F.; RESENDE, M. D. V.; GRATTAPAGLIA, D. Enviromics 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>.

REYNOLDS, D.; BARET, F.; WELCKER, C.; BOSTROM, A.; BALL, J.; CELLINI, F.; LORENCE, A.; CHAWADE, A.; KHAFIF, M.; NOSHITA, K.; MUELLER-LINOW, M.; ZHOU, J.; TARDIEU, F. What is cost-efficient phenotyping? Optimizing costs for different scenarios. **Plant Science**, v. 282, p. 14-22, May 2019. DOI: <https://doi.org/10.1016/j.plantsci.2018.06.015>.

STANTON J. M. Galton, Pearson, and the Peas: a brief history of linear regression for statistics instructors. **Journal of Statistics Education**, v. 9, n. 3, p. 1-13, Jan. 2001. DOI: <https://doi.org/10.1080/10691898.2001.11910537>.

VAN EEUWIJK, F. A.; BUSTOS-KORTS, D. V.; MALOSETTI, M. What should students in plant breeding know about the statistical aspects of genotype× environment interactions? **Crop Science**, v. 56, n. 5, p. 2119-2140, Sept./Oct. 2016. DOI: <https://doi.org/10.2135/cropsci2015.06.0375>.

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>.

WALLACE, J. G.; RODGERS-MELNICK, E.; BUCKLER, E. S. On the road to breeding 4.0: unraveling the good, the bad, and the boring of crop quantitative genomics. **Annual Review of Genetics**, v. 52, p. 421-444, Nov. 2018. DOI: <https://doi.org/10.1146/annurev-genet-120116-024846>.

WICKHAM, H. **Ggplot2**: elegant graphics for data analysis. 2nd ed. New York: Springer, 2016. p. 241-253.

José Maria Villela Pádua, Adriano Teodoro Bruzi, Ricardo Augusto Diniz Cabral Ferreira, Gustavo Evangelista Oliveira

Introdução

Os programas de melhoramento de plantas apresentaram evolução bastante acentuada nos últimos anos. Esse incremento não foi somente em termos de conhecimento ou tecnologias, mas também de dimensionamento ou tamanho dos mesmos. Hoje a atuação de muitos programas é da ordem global, principalmente devido a muitas fusões entre empresas que aconteceram nos últimos anos. Um dos lemas colocados por muitos programas é o “jogo de números”, o que faz com que o tamanho dos programas muitas vezes seja de grande magnitude, isto é, o número de dados obtidos seja muito grande (Vieira; Chen, 2021). Esses dados são oriundos de diferentes esferas: de fenômica, genômica e ambientômica, e ao conjunto foi dado o nome “melhoramento de plantas moderno” (Xu, 2016; Crossa et al., 2021; Resende et al., 2021).

Esse melhoramento de plantas moderno, também chamado de “melhoramento 4.0”, é caracterizado pelo desenvolvimento e uso de diferentes metodologias, incluindo técnicas como o Speed Breeding e a edição de genomas (Ramstein et al., 2019). A aplicação de tais tecnologias irá exigir ainda mais conhecimento em diferentes áreas dos melhoristas. Como essas abordagens dependerão de novos tipos de dados, como sequências de DNA, fenotipagem de alto rendimento, predições, entre outros, o “melhoramento 4.0” demandará análises complementares aos estudos genéticos quantitativos tradicionais, sendo baseado em técnicas como as de Inteligência Artificial, IoT (Internet of Things), de aprendizado de máquina, entre outras que fazem uso eficiente de dados.

A gestão otimizada de um programa de melhoramento genético envolve uma série de atividades que visam maximizar a eficiência e eficácia do processo de seleção

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

de características desejáveis em uma determinada espécie. Algumas das atividades-chave para a gestão otimizada de um programa de melhoramento genético incluem: definição dos objetivos do programa de melhoramento; organização do germoplasma; geração de variabilidade – planejamento de bloco de cruzamentos; condução do processo ou desenvolvimento do germoplasma, avaliações e seleções, entre outros.

Por tudo isso destacado, a gestão otimizada de um programa de melhoramento genético exige cada vez mais a combinação e integração de inúmeros conhecimentos técnicos, experiência prática e recursos financeiros para ser bem-sucedida. É importante que os responsáveis pela gestão do programa estejam sempre atualizados com as últimas tecnologias e métodos disponíveis para garantir que o programa seja eficiente e eficaz. A maioria dessas metodologias serão comentadas mais detalhadamente nos próximos capítulos. Neste capítulo, será abordado o funcionamento de um programa de melhoramento levando em conta a gestão otimizada das informações.

Organização Geral dos Programas de Melhoramento

Dentro de um programa de melhoramento de plantas existem várias etapas envolvidas (Bernardo, 2020), e dentre as mais importantes, estão o planejamento e a organização do programa de melhoramento. Essa organização envolve etapas ou fases chamadas, por alguns, de fases iniciais, intermediárias e fases finais (Figura 1). Antes de discorrer um pouco sobre essas etapas, é preciso salientar que qualquer programa de melhoramento leva em torno de 7 a 10 anos para terminar, ou seja chegar a cultivares, portanto, os objetivos precisam ser bastante claros e bem definidos para se ter sucesso ao longo do processo (Ramalho et al., 2012).

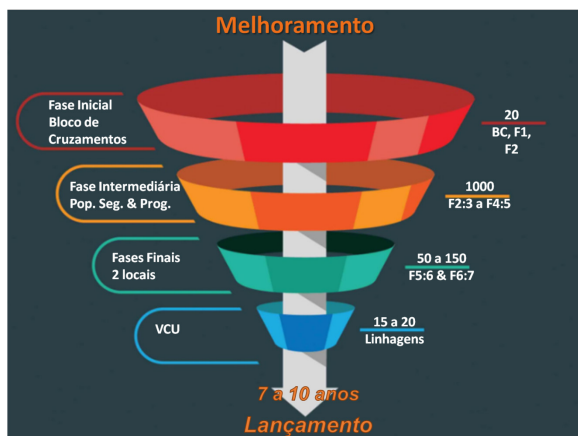


Figura 1. Exemplo da organização de um programa de melhoramento de plantas autógamas distribuído em quatro etapas sequenciais. VCU = Valor de Cultivo e Uso.

A fase inicial contempla a decisão de qual estratégia será utilizada, seja ela introdução de cultivares, seleção de linhas puras ou a hibridação. Diante dessa decisão surgem vários outros questionamentos, como, por exemplo, quais genitores utilizar, quais métodos de recombinação, dentre várias outras questões. E, como não existe uma resposta totalmente certa – é a velha frase do agrônomo “depende” – isso nos leva a crer que quanto maior o conhecimento aplicado maior a chance de sucesso.

O progresso no melhoramento de plantas para resistência a doenças pode ser relativamente rápido se elas forem controladas por poucos genes. No entanto, a resistência de plantas a doenças é um dos vários objetivos do melhoramento, que de outro lado necessita desenvolver genótipos com níveis aceitáveis de produtividade, adaptação, qualidade e outros caracteres agronômicos favoráveis, ou seja, o trabalho com várias características se torna um desafio, principalmente quando as correlações entre as características são desfavoráveis ao ganho simultâneo.

Melhoramento de espécies autógamas

As espécies vegetais autógamas são caracterizadas por apresentarem um processo de reprodução que ocorre predominantemente por autofecundação, ou seja, o cruzamento entre flores da mesma planta. Nas espécies vegetais autógamas, a autopolinização é o principal mecanismo de reprodução, o que contribui para a manutenção da pureza genética e estabilidade dos genótipos ao longo das gerações. Aqui se utilizará como exemplo um programa de melhoramento de trigo para resistência às doenças, produtividade e qualidade de grãos. Antes de iniciar um programa para o desenvolvimento de cultivares é necessário fazer levantamento prévio, no caso do trigo, das doenças que ocorrem na cultura e priorizá-las segundo os danos econômicos causados, para a região de atuação, bem como das cultivares mais produtivas (“adaptadas” e recomendadas) e também da qualidade desejada pelos compradores (moinhos da região). Nesse sentido, para se obter sucesso, é fundamental a organização e conhecimento do germoplasma, sendo uma das principais e mais importantes tarefas nessa fase (Carena, 2021). Entre as informações úteis para conhecer e selecionar genitores podem ser utilizados dados agronômicos, morfológicos, dados genômicos, entre outros.

A cultivar de trigo mais cultivada e aceita pelos moinhos na mesorregião Sul/Sudoeste de Minas e Campo das Vertentes é a BRS 264 (Pádua et al., 2022). Essa cultivar apresenta algumas características agronômicas e de qualidade que atendem a maioria das necessidades dos produtores, com exceção de uma: é altamente suscetível à brusone da espiga (*Pyricularia oryzae* patótipo *Triticum* – PoTI). Nesse

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

caso, já se sabe que o controle genético desse patógeno é poligênico, logo se trabalha com acúmulo de alelos favoráveis. Nesse caso, o primeiro ponto é decidir quem serão os genitores envolvidos na formação da população multiparental recombinada. Considerando informações já existentes do comportamento de cultivares (Maciel et al., 2022) e a interação com outros setores da Ufla e Epamig foram identificadas as linhagens genitoras da população base. Em princípio, foram envolvidas 20 linhagens cruzadas no esquema de Top Cross com a cultivar BRS 264. A intenção foi obter 20 populações de F_1 e avançá-las para a F_2 .

Nas fases iniciais e intermediárias, no caso do trigo, são selecionadas principalmente características de alta a média herdabilidade e, nesse momento, são utilizadas ferramentas úteis como é o caso dos marcadores moleculares. O uso de marcadores moleculares, ligados a genes, não é nenhuma novidade e já é realidade na seleção assistida ao melhoramento de trigo. Marcadores moleculares são sequências de DNA que permitem diferenciar, no mínimo, dois ou mais indivíduos, e em trigo estão disponíveis alguns marcadores que auxiliam na seleção de populações e progênies que apresentem as propriedades desejadas antes que elas se manifestem, entre eles os genes de resistência a brusone (2NS, Rmg e alguns QTL – Cruz; Valent, 2017). A seleção genômica também tem sido implementada com sucesso em alguns programas públicos (Wartha; Lorenz, 2021). Além dos marcadores, uma ferramenta possível de ser utilizada nessa fase inicial é o Speed Breeding (Watson et al., 2018). Essas ferramentas de melhoramento acelerado têm permitido a redução do tempo necessário para condução de um ciclo de melhoramento em diversas culturas, como: trigo, aveia, cevada, dentre algumas outras culturas.

A fase intermediária, envolve a avaliação das progênies da geração $F_{2,3}$ até a geração $F_{4,5}$, e, no caso do trigo, envolve as seleções e também o aumento da quantidade de sementes. Nessa fase intermediária, são avaliadas progênies com maior número de plantas, possibilitando fazer avaliações mais precisas, bem como trabalhar com experimentos maiores. As decisões nesse momento envolvem definir qual método de melhoramento utilizar, tamanho de parcela ideal, número de repetições, delineamento estatístico, entre outras. É necessário ressaltar que os vários métodos de melhoramento disponíveis para o desenvolvimento de cultivares apresentam benefícios e limitações, e a escolha do método mais adequado irá depender, entre outros fatores, do tipo de planta, das características a serem alvo de seleção, assim como dos recursos disponíveis (Ramalho et al., 2012).

Nesse momento, além dos dados fenotípicos e genotípicos, é mais comum iniciar a utilização de algumas ferramentas como o uso de imagens (que também podem ser utilizadas nas fases iniciais). Ferramentas como as de imagens de alta qualidade

permitted today that different phenotypes are screened on a large scale (HTP – High-throughput phenotyping). Considering commercial programs that test a large number of treatments per year and considering evaluations of characteristics that are complex and costly, it has been possible to obtain HTP of good quality to help in the decision of the improvement programs. An example of this has been the use in the wheat program of Cimmyt (Centro Internacional de Mejoramiento de Milho e Trigo, México) for diseases, plant height, among others (Volpato et al., 2021). In this context, it is necessary to evaluate which characteristics have a good explanation by the image, the cost/benefit relationship of this tool, as well as the people and equipment necessary to process, store, interpret and use this information.

In the final phase of an improvement program a smaller number of genotypes, but in a larger number of repetitions and locations, is one of the main challenges of the breeders, such as the greater demand for resources and the interaction Genotypes × Environments. To comment on this issue, we will continue with the example of tropical wheat in Brazil. There is a large number of mentions about the potential for expansion of the culture in the tropical region of Brazil, but the challenges under tropical conditions are great and the number of works to better characterize this environment is very scarce (Coelho et al., 2010; Condé et al., 2010). In these conditions, only a little more efficiency can be achieved if the institutions involved with the research associate to optimize the available resources. In a similar situation, the strategy that the north-Americans found was to implement a research network in a way that many pieces of information are generated and thus making the improvement process more efficient, based on predictions (Lawrence-Dill et al., 2019).

In the specific case of wheat under tropical conditions, there is a great limitation in terms of the number of institutions involved and resources (human and financial), and this makes the works of the different improvement or research programs even more challenging. Therefore, there is a great benefit in adopting a system similar to the one reported above for the wheat chain under tropical conditions. This would be the first step to advance the understanding of this environment, bringing together the union of data from environmental, phenomic and genomic and unfolding the positive and negative points, visualizing new frontiers and subsidizing the information for improvement in this region. The final phases still include experiments of VCU (Valor de Cultivo e Uso) for the registration of new cultivars.

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

Além dessas etapas, tem-se também fase de pós-melhoramento, que engloba duas etapas. A primeira é que após a geração $F_{6,7}$, as progênies mais promissoras deverão ter suas sementes multiplicadas para ser o pipeline de produção e abastecimento de sementes das novas cultivares que irão para o mercado. A segunda fase contempla os esforços para a adoção das novas cultivares, na qual as equipes de desenvolvimento de mercado atuam em algumas linhas, como por exemplo: 1) Caracterização e definição do melhor pacote de manejo para a nova cultivar; 2) Dias de campo; 3) Vitruvina tecnológica; 4) Visitas técnicas; 5) Mídia eletrônica; entre outros. A introdução e adoção das novas cultivares por parte dos agricultores não é um processo simples, e pelas razões discutidas acima, não parece haver uma estratégia baseada em ferramenta única que determinará o sucesso do processo. Com o acirramento da competitividade do mercado de cultivares, uma estratégia de múltiplas ferramentas que procure contemplar os diferentes padrões de agricultores aliado às demandas dos compradores/consumidores parece ter maiores chances de sucesso no contexto agrícola atual.

Nesse contexto, na tomada de decisão há aspectos além dos técnicos, quais sejam a definição do plano de marketing, como definição do tamanho de mercado e precificação, além dos aspectos de produção de sementes. Paralelamente, as linhagens que estão sendo desenvolvidas e testadas pelos melhoristas no funil estão sendo purificadas, multiplicadas e estudadas no aspecto de produção de sementes. Assim, quando uma decisão de lançar uma cultivar é tomada, é necessário que já se tenham disponíveis sementes básicas e uma receita de produção de sementes eficiente para suportar a demanda dessa cultivar pelos agricultores.

Melhoramento de espécies alógamas

Diferentemente do melhoramento de plantas autógamas, existem espécies de plantas alógamas em que o objetivo final de se obter uma cultivar seria, majoritariamente, de variedades de polinização aberta ou híbridos. Nesse sentido, para a obtenção de híbridos após a obtenção de progênies endogâmicas ou linhagens, são realizados cruzamentos para se obter os híbridos. A planta mais elucidada no melhoramento de alógamas é o milho, logo se irá utilizá-la como exemplo. O mercado brasileiro de sementes de milho é de aproximadamente 96% de híbridos, sendo assim, o primeiro passo para obtenção de híbridos, como mencionado anteriormente, é a obtenção de linhagens. Nesse sentido, tudo que já foi mencionado é válido para a obtenção de linhagens. A diferença mais evidente é que, no caso do milho, já estão sendo utilizadas rotineiramente metodologias para se obter linhagens mais rápidas, como é o caso dos Duplos Haplóides (DH) (Figura 2). No ano de 1924, surgiu o

primeiro relato da ocorrência de plantas haplóides em tabaco (Clausen; Mann, 1924), posteriormente foram encontradas também em trigo e outras espécies (Wedzony et al., 2009). Somente recentemente essa tecnologia vem sendo difundida e utilizada nos programas de melhoramento no Brasil. Sua contribuição se baseia no avanço do processo de formação de plantas completamente homocigotas numa única etapa, contrapondo-se ao demorado processo tradicional de autofecundação. Ela permite que o processo de formação de uma nova variedade seja abreviado, diminuindo em até quatro anos as etapas de obtenção de uma nova cultivar, agilizando o sistema de seleção. A tecnologia de DH pode ser descrita por plantas geralmente originadas *in vitro* de gametas masculinos ou femininos, não fecundados, resultando na formação de plantas com apenas um lote cromossômico (vindo de um único gameta). Por serem haplóides e apresentarem apenas a metade do número de cromossomos, na sua grande maioria, essas plantas são, naturalmente, estéreis. Após duplicação dos cromossomos, elas recuperam a fertilidade, se tornando completamente homocigotas. No caso do trigo, por exemplo, existem duas maneiras mais comuns de induzir os haplóides: por cultura de anteras ou por cruzamento com indutor de haplóides e posterior eliminação cromossômica. Vale ressaltar que existem diversos exemplos de uso e aplicação dos DH nos diferentes programas de melhoramento (Chaikam et al., 2019). Além da diminuição no tempo para alcance da homocigose, o método também permite maior eficiência no processo de seleção de genótipos com características superiores, sendo possível identificar, rapidamente, genes deletérios (que poderiam ser mascarados pela heterocigose), diminuindo o número de indivíduos necessários para identificar um possível “genótipo-alvo” (com características desejáveis).

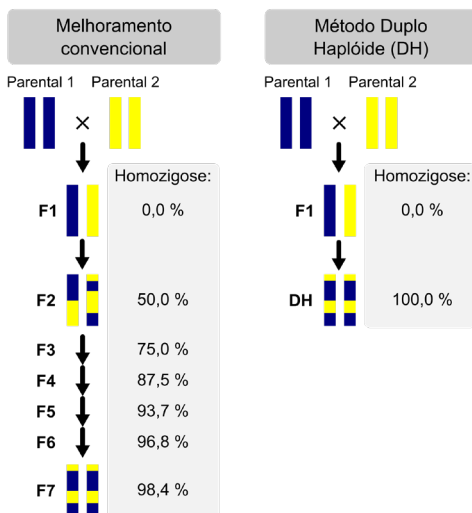


Figura 2. Comparação do uso da metodologia de Duplos Haplóides (DH) com a metodologia convencional. São mostrados os níveis de homocigose das populações de melhoramento ao longo das gerações.

Fonte: Adaptado de Teingham e La Borde (2017).

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

Além do uso de DHs em milho, a seleção genômica ampla (SGA ou simplesmente SG) vem sendo largamente comentada e utilizada em alguns programas de melhoramento nacionais e mundiais de milho (Marinho et al., 2022). Essa técnica preconiza o uso de um número “x” de marcas espalhadas no genoma da espécie em estudo para se correlacionar ao fenótipo em fases de treinamento, para posterior seleção da população (Crossa et al., 2017). A seleção genômica tem oferecido oportunidades de maximizar os ganhos genéticos e acelerar o ciclo de melhoramento. A seleção genômica é um tipo de metodologia de melhoramento em que os melhores materiais são selecionados a partir de seus valores genéticos preditos a partir de dados obtidos de marcadores moleculares, sendo os mais comuns os Single Nucleotide Polymorphisms – SNPs. Para isso é necessária que se tenha uma chamada população de treinamento e uma população de validação para calibrar os valores genotípicos vs os fenotípicos. Valores genéticos aditivos preditos em gerações precoces (DH ou $F_{2,3}$) são utilizados para seleção precoce e redução do intervalo dos ciclos de seleção. Nesse caso, os valores genéticos aditivos são utilizados como Breeding Values como critério de seleção dos indivíduos / linhagens de uma população. Além da redução do intervalo dos ciclos proporcionada pela seleção em gerações precoces, também existe a possibilidade da maximização dos ganhos por meio da condução de populações muito maiores, já que a análise por marcadores para predição dos valores genéticos pode ser efetuada ainda em fases iniciais do programa de melhoramento, possibilitando a seleção precoce e descarte de não candidatos.

Para elucidar melhor os comentários anteriores, tome-se como exemplo um programa que implementa ambos (DH e SGA) em sua rotina. Após a obtenção dos DHs (vejam que se utiliza a SGA já para a seleção dos DHs), são realizados os chamados Top crosses (TC) com testadores do grupo heterótico complementar a cada um dos grupos (Lado esquerdo – em azul e direito – em verde – da Figura 3). Esses testadores são normalmente linhagens elite de cada um dos grupos, que apresentam alta capacidade geral de combinação (CGC). Após serem sintetizados, os híbridos irão para testes em campo e serão selecionados os melhores, os quais, após as reuniões de avanço, irão avançar para ensaios finais e VCU e, posteriormente, para registro e lançamento do(s) melhor(es). Vejam que o processo de seleção e registro é semelhante ao relatado para autógamias, a diferença é que as linhagens convencionais ou DH não são testadas em vários locais, mas sim os híbridos oriundos dos TC_1 e TC_2 .

Melhoramento de espécies de propagação vegetativa

As espécies de propagação vegetativa têm como objetivo o uso de cultivares comerciais sendo clones. Isso faz com que os indivíduos sejam geneticamente

idênticos à planta mãe, ou efetivamente clones da planta mãe. Uma parte das espécies de propagação vegetativa (ou de propagação assexuada) são alógamas, como a uva (*Vitis* spp.), banana (*Musa* spp.), mandioca (*Manihot esculenta*), cana-de-açúcar (*Saccharum officinarum*) e algumas espécies florestais como as espécies do gênero *Eucalyptus* spp., o qual será exemplificado aqui. É importante destacar que o eucalipto, embora tenha predomínio de propagação vegetativa significativo, recentemente, muitos pesquisadores brasileiros têm dedicado esforços no melhoramento de linhagens e composição de híbridos, algo semelhante ao que ocorre no programa de melhoramento de milho mostrado na Figura 3.

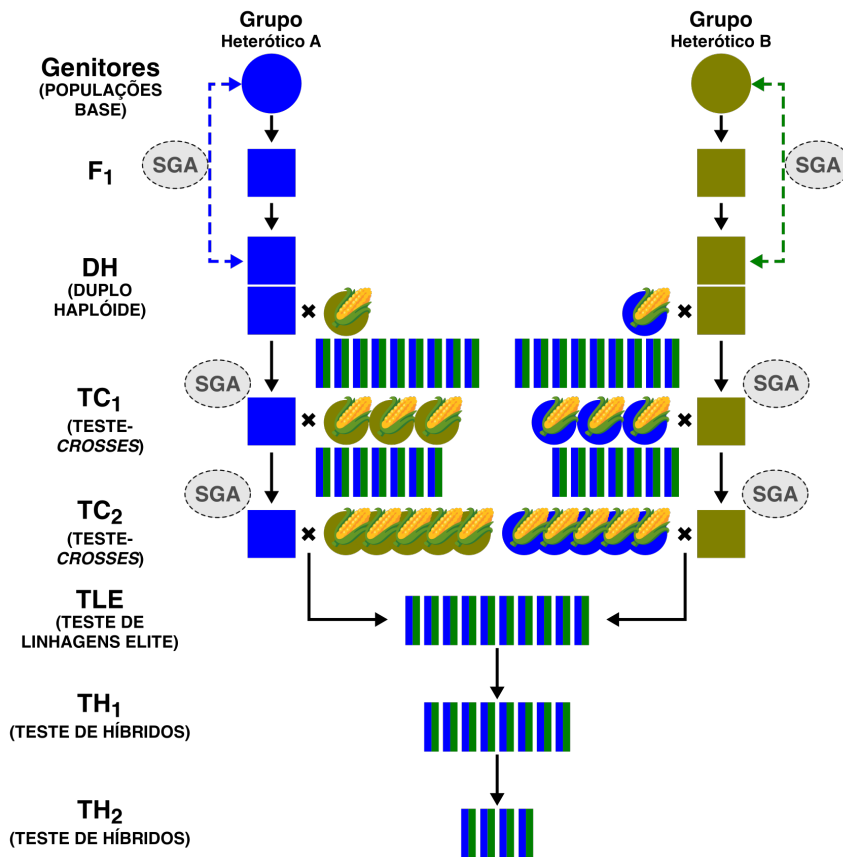


Figura 3. Estrutura de um programa de melhoramento de milho, partindo de dois grupos heteróticos distintos (parte azul e verde) até chegar em híbridos melhorados com elevado grau de heterose. A sigla SGA significa Seleção Genômica Ampla.

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

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

Um processo simplificado de seleção utilizado no melhoramento florestal é apresentado na Figura 4. Inicialmente, são identificados os melhores indivíduos, que podem ser provenientes de germoplasma da Oceania, como a Austrália, ou de bancos dentro das grandes companhias florestais brasileiras. Esses indivíduos passam por múltiplos testes de progênes, seguindo um ciclo de Seleção Recorrente, que pode ser Recíproca/Interpopulacional (SRR), Intrapopulacional (SRI) ou Intrapopulacional em População Sintética (SRIPS). Em seguida, os melhores materiais genéticos são selecionados e têm duas destinações possíveis: 1) o melhoramento para a obtenção de híbridos clonais (maximizando-se a heterose entre o cruzamento de diferentes espécies de eucalipto); ou 2) a manutenção de pomares de multiplicação ou a própria recombinação das melhores plantas dentro das melhores populações. É importante observar que, após os cruzamentos, as plantas S_0 são testadas e as melhores são selecionadas para serem clonadas e continuar os ensaios finais e VCU. A genômica florestal brasileira também merece destaque aqui, devido a suas pronunciadas tecnologias de genotipagem via SNP e de SGA (Grattapaglia, 2022).

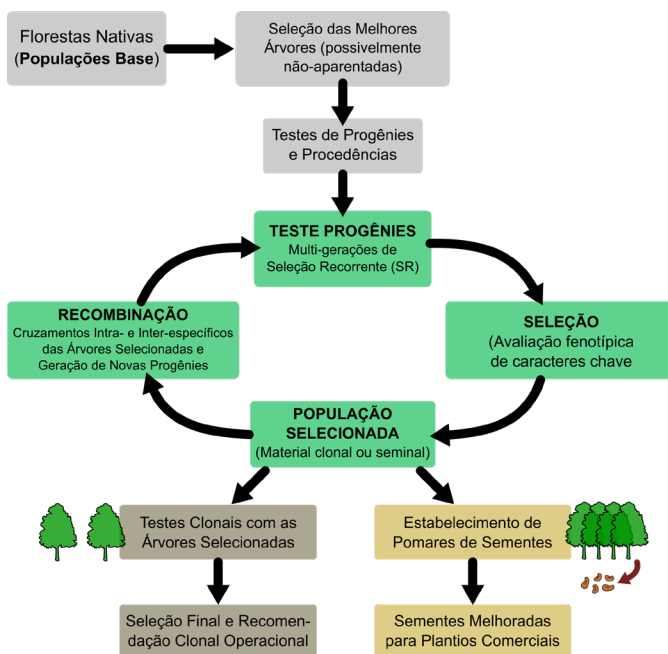


Figura 4. Estrutura simplificada de um programa de melhoramento de Eucalipto do tipo Seleção Recorrente, em que duas vias são possíveis, uma para desenvolvimento de clones elite e a outra para composição de pomares que fornecerão sementes melhoradas.

Fonte: Adaptado de Grattapaglia e Kirst (2008).

Para maiores detalhes sobre o melhoramento de autógamias, alógamas e de espécies de propagação vegetativa, sugere-se a obra de Recursos Genéticos e Melhoramento de Plantas (Nass et al., 2001) e Melhoramento genético de plantas de propagação assexuada (Resende; Barbosa, 2005).

Ferramentas Utilizadas para Gerenciamento do Programa

Um programa de melhoramento é de natureza complexa, principalmente em uma organização privada. Nesse sentido, estão envolvidos uma estrutura física, recursos humanos e financeiros que envolvem uma gestão cada vez mais tecnificada. Nesse contexto, são inúmeras ferramentas disponíveis para o gerenciamento de informações geradas por esses programas, e abrangem desde as tradicionais planilhas feitas por computador e preenchidas à mão, até os mais sofisticados softwares. Entre eles existem diferenças marcantes e com isso é preciso avaliar qual a real necessidade de cada programa de melhoramento. Entre as principais atividades de um software de melhoramento estão a organização, descrição e controle do germoplasma; planejamento e montagem de experimentos; análise de dados fenotípicos e genotípicos; controle de estoques de sementes, os quais serão detalhados um pouco dentro de alguns dos principais softwares.

Quando mencionamos a estruturação de um time de ciência de dados, as ferramentas de coleta, estruturação e organização dos dados, além das ferramentas de análises, devem ser escolhidas com base em requisitos técnicos respeitando as características de cada empresa/organização. Dessa maneira, é de fundamental importância ter um time de ciência de dados estruturado suportando o melhorista, como também toda a cadeia ligada ao melhoramento. Com uma estrutura semelhante à da Figura 5, o melhorista terá suporte na parte de organização de dados, análises e sumarização, podendo focar seus esforços em planejamento estratégico e interpretação de informações e as tomadas de decisões sobre ferramentas e sistemas de informação a serem implementados de modo assertivo e sustentável ao longo dos anos. Nesse sentido, as técnicas de análise de dados e a importância de integrar informações de fontes diferentes para obter uma visão completa do programa são tratadas pelo time de ciência dos dados. O armazenamento e gestão dos resultados e avaliações na forma de meta-análises são questões que vêm sendo aplicadas no melhoramento, bem como tecnologias, como inteligência artificial e o aprendizado de máquinas, podem ajudar em tomadas de decisão estratégicas. No âmbito da era digital e da *Big data*, novas tecnologias têm emergido para lidar com isso. Nesse contexto, merece destaque a inteligência artificial, que é a ciência capaz de mimetizar as habilidades humanas. Muitas empresas como o Google, Facebook e

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

Netflix têm utilizado a inteligência artificial no sentido de mimetizar perfis de consumidor. No melhoramento de plantas, no âmbito da inteligência artificial, vários métodos como o *Machine Learning*, *Deep Learning*, entre outros têm surgido. As aplicações são no sentido de juntar dados de genômica, fenotipagem e ambientipagem, considerando o treinamento de modelos para aprender com os dados, identificar padrões e tomar decisões mais assertivas. Mais detalhes sobre as técnicas de Ciência de Dados e Inteligência Artificial, destacadas na Figura 5, serão vistos com maior riqueza de detalhes nos demais capítulos deste livro.

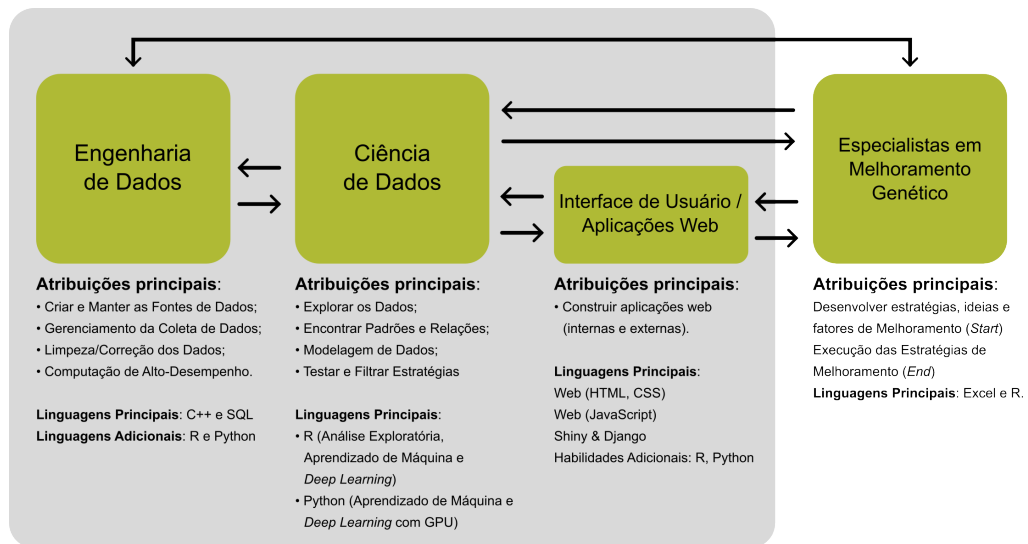


Figura 5. Exemplo de estruturação de uma equipe de análise de dados e sua interface com profissionais do Melhoramento Genético de Plantas. Mais detalhes sobre essas atribuições dos profissionais da Ciência de Dados serão vistos no decorrer dos capítulos do presente livro.

Fonte: Adaptado de: <https://www.business-science.io/business/2018/09/18/data-science-team.html>.

Softwares de gerenciamento

Como mencionado nos tópicos anteriores, os programas de melhoramento tomaram uma magnitude muito grande quando se trata de informações e dados gerados. Nesse sentido, o tempo do melhorista para lidar com todas essas informações tende a ser muito grande, e sabemos que nas diferentes organizações ele não consegue dedicar 100% do seu tempo para lidar com tudo isso, o que faz com que as ferramentas e softwares de gestão sejam estritamente necessárias. Muitas são as opções de ferramentas e softwares disponíveis para gerenciamento de informações dentro de um

programa de melhoramento de plantas. Porém, é preciso ter conhecimento do programa de melhoramento em que se deseja utilizá-los, pois às vezes o custo-benefício ou a eficiência operacional fazem com que eles não sejam passíveis de utilização. Outro ponto importante é a integração de sistemas, pois na maioria dos casos as ferramentas não conversam entre si, o que faz com que o uso de duas ou mais plataformas seja difícil. Em nossa visão, isso não será modificado facilmente, pois envolve interesses dos desenvolvedores. Essas ferramentas têm um apreço grande na otimização de processos dentro do programa, como por exemplo a automação de tarefas repetitivas e a eliminação de redundâncias, para uma gestão mais eficiente e eficaz.

O processo de melhoramento genético pode se beneficiar de uma série de softwares especializados (Tabela 1). O BMS Pro, disponível na plataforma *Integrated Breeding Platform* (IBP), oferece recursos avançados para a gestão de dados genéticos, permitindo aos pesquisadores armazenar, acessar e analisar informações para a seleção e melhoramento de plantas. O software Phenome oferece uma solução abrangente para o gerenciamento de dados fenotípicos, possibilitando a captura precisa de informações relacionadas a características observáveis de plantas e animais. O Agronomix, desenvolvido pela Genovix, é uma ferramenta poderosa para análise estatística e tomada de decisões baseadas em dados genéticos, oferecendo recursos avançados para a identificação de linhagens e cruzamentos promissores. O E-brida é um software projetado especificamente para o melhoramento genético de plantas, fornecendo ferramentas avançadas para a análise de dados genéticos e fenotípicos. O Doriane é um software abrangente que oferece suporte a várias etapas do processo de melhoramento genético, desde a coleta e gerenciamento de dados até a análise e tomada de decisões.

Para a visualização e análise de dados, o POWER BI, uma ferramenta da Microsoft, pode ser utilizado para criar painéis interativos e relatórios personalizados, permitindo compreensão mais profunda dos dados e facilitando a comunicação de resultados entre os envolvidos no processo de melhoramento genético. A título de exemplo, o [Google imagens](#) possui visualizações gráficas do POWER BI. Esse é de fato um software aplicado para mostrar o panorama de dados, experimentos e o progresso do melhoramento genético da empresa ou instituição. Por meio desse software é fácil visualizar a quantidade de dados em diferentes localidades, em diferentes distribuições geográficas, os diferentes caracteres fenotípicos avaliados, ou mesmo os diferentes estágios do melhoramento da empresa, como: quantidade de testes de progênies, testes de hibridação, testes clonais etc. Além disso, é uma excelente forma de visualizar os padrões de balanceamento e desbalanceamento de materiais genéticos entre diferentes experimentos.

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

Tabela 1. Exemplos de softwares utilizados no gerenciamento de informações nos programas de melhoramento.

Software	Principais Funcionalidades	Fonte
BMS Pro (Integrated Breeding Platform – IBP)	Criação de banco de dados e consultas Atividades de Reprodução Análise estatística Dados de genotipagem Gerenciamento de programas e usuários Captura eletrônica de dados	https://integratedbreeding.net/
Phenome	Sistema Principal: PhenomeOne Aplicativo Móvel: PhenoTop Inventário Análise de dados Genômica PhenomeBI Phenogene	https://phenome-networks.com/
Agronomix (Genovix)	Teste de variedades Melhoramento de plantas Inventário de sementes Bioinformática SaaS Baseado em nuvem Soluções móveis	https://genovix.io/ ; https://www.agronomix.com/
E-brida	Registo de dados online e offline Compartilhe seu conhecimento de reprodução internamente Gerencie seus custos de reprodução Registre a observação com laptop, tablet ou telefone no campo Minimizar atos administrativos Crie um método de trabalho uniforme Ancestrais e descendentes em uma visão geral	https://www.e-brida.com/
Doriane	Acompanha várias operações: cruzamentos e hibridações, produção de linhagem pura, cruzamentos de teste e ensaios, relatórios Novas tecnologias Sistema ETL ¹ Sistema Colaborativo Integrado	https://www.doriane.com/
POWER BI	Visualização de dados fenotípicos e genéticos Monitoramento do número de experimentos implantados Visualização gráfica dos locais com experimentos Contabilização dos traços fenotípicos mensurados Criação de painéis interativos de dados Compartilhamento de informações e colaboração entre diferentes profissionais	https://powerbi.microsoft.com/pt-br/

¹ ETL = Extraction (Extração)/ Transformation (Transformação)/ Loading (Carregamento)

Além desses softwares de gerenciamento, algumas empresas optam pelo desenvolvimento de programas dentro da empresa (*in company*) pensando na otimização dos processos, ou seja, na personalização do produto e, além disso, na segurança da informação. Existem softwares como o Microsoft Excel e Programa R, por exemplo, contendo diferentes pacotes que são utilizados para gerenciamento de informações em um programa de melhoramento. Como exemplo, planilhas podem ser geradas no Excel, compartilhadas no Google Drive, e podem ser editadas em celulares ou tablets. Também alguns pacotes do R são bastante úteis em fazer sorteios e, principalmente, analisar dados obtidos.

Porém, é de fundamental importância entender que softwares como Excel e outros com características semelhantes não podem ser tratados como uma forma segura de banco de dados, pois não foram desenvolvidos com essa finalidade. Hoje em dia há plataformas de nuvem, a um preço acessível, como Azure, AWS, entre outras, que permitem criar banco de dados de maneira estruturada, utilizando na maioria das vezes linguagem de programação em SQL, que permite segurança da informação, integração com outras ferramentas e sustentabilidade em números de data points a serem armazenados. Maior aprofundamento nesse tema será abordado nos demais capítulos deste livro.

Organização dos Dados de Melhoramento

Organização de dados fenotípicos

Nas diferentes etapas do programa de melhoramento são coletados diversos tipos de dados de avaliações fenotípicas. Nesse caso, é preciso estar atento ao que se tem de objetivo e prioridade em cada etapa. Para exemplificar a organização dos dados, será tomado como exemplo um experimento de tomate. São inúmeras avaliações necessárias para se fazer o registro de uma nova cultivar de tomate no Brasil. Nesse sentido, as principais avaliações são de dados agrônômicos e de qualidade. Ao se utilizar algum software, tal como algum dos listados anteriormente (Tabela 1), após a definição dos locais, tratamentos, repetições, delineamento e etc., serão criadas colunas com cada variável a ser avaliada (Figura 6). Outro ponto importante, são as coletas de dados em plataformas remotas, como os celulares, tablets e até mesmo nos próprios equipamentos agrícolas, como as colhedoras de parcelas, em que as mesmas são processadas, analisadas e avaliadas, gerando as informações necessárias em tempo real (Figura 7).

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

#	I	GID	Name	Generation	Method code	Female code	Male code	Female plot	Male plot	TY_1	TY_3	I2	Ver	TMV	MI																																												
1		4620826	CR18-F1-2-x	F1	CR	2	97	541	540	R	S	R	R	R	R																																												
<table border="1"> <thead> <tr> <th>Role</th> <th>GID</th> <th>Entry code</th> <th>Name</th> <th>Group</th> <th>Source</th> <th>Generation</th> <th>Firmness</th> <th>Cracking</th> <th>Fruit size</th> <th>No.</th> </tr> </thead> <tbody> <tr> <td>c</td> <td>462082</td> <td></td> <td>CR18-F1-2-x</td> <td></td> <td></td> <td>F1</td> <td>5</td> <td>5</td> <td>4</td> <td></td> </tr> <tr> <td>f</td> <td>461501</td> <td>2</td> <td>Sel18-F6-501-B1</td> <td></td> <td>Dani1272</td> <td>F6</td> <td>5</td> <td>4</td> <td>4</td> <td></td> </tr> <tr> <td>m</td> <td>461511</td> <td>97</td> <td>Sel18-F6-634-SB2</td> <td></td> <td>Star56456</td> <td>F6</td> <td>5</td> <td>6</td> <td>4</td> <td></td> </tr> </tbody> </table>																Role	GID	Entry code	Name	Group	Source	Generation	Firmness	Cracking	Fruit size	No.	c	462082		CR18-F1-2-x			F1	5	5	4		f	461501	2	Sel18-F6-501-B1		Dani1272	F6	5	4	4		m	461511	97	Sel18-F6-634-SB2		Star56456	F6	5	6	4	
Role	GID	Entry code	Name	Group	Source	Generation	Firmness	Cracking	Fruit size	No.																																																	
c	462082		CR18-F1-2-x			F1	5	5	4																																																		
f	461501	2	Sel18-F6-501-B1		Dani1272	F6	5	4	4																																																		
m	461511	97	Sel18-F6-634-SB2		Star56456	F6	5	6	4																																																		
2		4620833	CR18-F1-2-x(1)	F1	CR	2	78	502	500	H	S	R	H	R	H																																												
3		4620834	CR18-F1-2-x(2)	F1	CR	2	79	510	508	H	H	R	R	R	H																																												
4		4620835	CR18-F1-2-x(3)	F1	CR	2	80	518	516	R	S	H	H	R	H																																												
5		4620836	CR18-F1-2-x(4)	F1	CR	2	81	526	524	R	H	H	R	R	H																																												

Experimental designs

Design: Factorial design

Repl: Alpha design type (0,1)

Rand: Randomized Balanced Incomplete Block I

Rand: Completely Randomized Design

Facto: Cyclic designs

Facto: Augmented block design

Seed: Graeco - latin square design

Seed: Lattice designs

Subplots: Yes No

Completely Randomized Design

Kinds: default

Replicates:

Replicates:

Figura 6. Exemplo de planilhas para coleta de dados fenotípicos. Na parte A é mostrada a organização de um banco de dados de cruzamentos. Na parte B, é mostrada como executar algumas análises preliminares para diferentes delineamentos experimentais.

Fonte: <https://phenome-networks.com/>

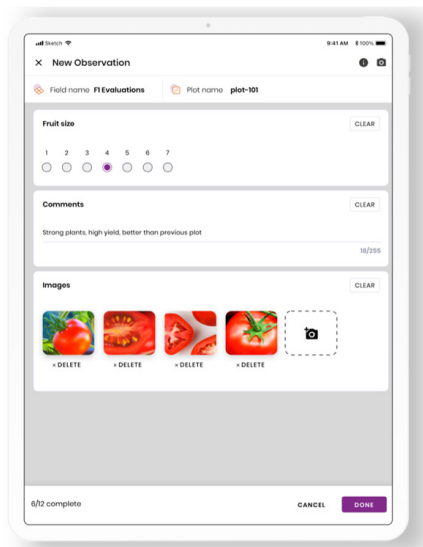


Figura 7. Exemplo de aplicativo para coleta de dados fenotípicos em dispositivo móvel como um simples celular.

Fonte: <https://phenome-networks.com/>

Organização de dados genéticos (moleculares)

Em programas consolidados de melhoramento genético, o uso dos marcadores moleculares é uma rotina. Não somente em se tratando de marcadores genéticos simples, ou seja, voltados para características de herança mono ou oligogênicas, mas também para as poligênicas. Os dados de caracterização genética por marcadores moleculares podem ser utilizados para a Seleção Assistida por Marcadores; Seleção Genômica ou *Genome-wide association studies*), por exemplo. Em consonância com o relatado anteriormente, em alguns softwares de gerenciamento, uma vez definidos os detalhes do experimento e quais as análises genéticas serão realizadas, também haverá o processamento das informações. No capítulo sobre genômica do presente livro será abordado diversas análises modernas sobre genômica no melhoramento de plantas.

Organização de dados ambientais

Além das já bem consolidadas Fenômica e Genômica, a “Ambientômica” apareceu como uma nova ômica modeladora da interação entre múltiplos dados ambientais × Genótipos (Resende et al., 2021, 2022). E não diferente do que foi listado acima, os dados ambientais têm ganhado cada vez mais importância no melhoramento de plantas (Costa-Neto; Fristche-Neto, 2021; Resende et al., 2021).

Para uma boa organização de dados ambientais nesse nosso contexto de melhoramento, seguem algumas ideais a serem desenvolvidas: i) Definição do objetivo: É importante ter claro o objetivo do estudo e o que se deseja medir e avaliar para definir quais dados são relevantes; ii) Coleta de dados: Realizar a coleta de dados ambientais relevantes, como temperatura, umidade, luminosidade, solo, entre outros. É importante registrar a data e hora de coleta, localização e identificação da planta. Porém, se for com finalidade de Ambientômica é necessário coletar e gerir toda gama de variáveis ambientais disponíveis. É importante que a empresa tenha bom relacionamento com a área de SIG (Sistemas de Informações Geográficas); iii) Armazenamento de dados: Armazenar os dados em um formato estruturado, como tabelas ou planilhas, para permitir a análise e a visualização dos dados; iv) Análise de dados: Realizar a análise dos dados coletados para identificar padrões e relações. É importante avaliar o impacto dos fatores ambientais na planta e como eles podem afetar o desenvolvimento e a produção; v) Integração de dados: Integrar os dados ambientais com outros dados relevantes, como informações genéticas, para obter compreensão mais abrangente da planta, vi) Documentação: Documentar todo o

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

processo de coleta, armazenamento e análise de dados para permitir a revisão e a reprodução dos resultados no futuro.

Um ponto importante de se mencionar é que falta maior diálogo do setor de geociências/geoprocessamento e o setor de melhoramento. Nesse sentido, alguns pontos que merecem destaque são: i) Comunicação: Estabelecer canais de comunicação abertos e efetivos entre os dois setores para permitir a troca de informações e ideias; ii) Identificação de necessidades: Identifique as necessidades e objetivos de cada setor e como eles podem complementar um ao outro; iii) Projetos colaborativos: Desenvolver projetos colaborativos que permitam a integração de dados geocientíficos com informações sobre o melhoramento vegetal; iv) Treinamento: Oferecer treinamento aos funcionários para que eles possam compreender as metodologias e ferramentas utilizadas por ambos os setores; v) Conferências e workshops: Promover conferências e workshops para que os profissionais de ambos os setores possam se conhecer e compartilhar suas experiências; vi) Tecnologia: Investigar e adotar tecnologias que possam facilitar a integração de dados e informações entre os dois setores; vii) Envolvimento de líderes: Envolver líderes de ambos os setores para que eles possam apoiar e garantir o sucesso da colaboração.

Exemplo de Aplicação

Apesar de tratar os dados fenotípicos, genéticos e ambientais em pontos separados acima, no melhoramento o acompanhamento e a ligação entre as diferentes etapas são fundamentais. Com isso, será elucidado aqui um processo de melhoramento que visa a confiabilidade, rastreabilidade e agilidade na busca de informações e tomada de decisão. Esse processo será dividido em quatro etapas interligadas e cíclicas: Gerenciamento de Germoplasma → Campos de Avaliação/Produção → Dados → Inventário → Gerenciamento de Germoplasma.

Gerenciamento de Germoplasma

Esta etapa representa a interface entre o melhorista e o banco de germoplasma, onde todas as informações relevantes, como fenótipos, genótipos e pedigrees, estão disponíveis e associadas a cada acesso genético (populações, progênies, linhagens, híbridos, clones, entre outros). Durante esse processo, várias ações são realizadas para garantir a eficiência e a utilização adequada do germoplasma. Em primeiro lugar, a gestão e classificação do germoplasma são realizadas, levando em

consideração as características específicas de cada cultura. Em culturas alógamas, como o milho, a classificação é baseada em grupos heteróticos, que consideram a compatibilidade genética entre os acessos para a formação de híbridos produtivos. Já em outras culturas, principalmente as autógamias, a classificação pode ser feita com base em critérios como grupo de maturação, tipologia de planta ou características específicas.

O controle populacional/parental e o pedigree são essenciais para rastrear a origem e a genealogia de cada acesso genético. Isso envolve o registro e o acompanhamento das gerações e dos parentais que deram origem a um determinado acesso. Essas informações são fundamentais para entender a história e as características genéticas de um material, bem como para direcionar cruzamentos e seleções futuras.

Outra ação importante é a rastreabilidade e o armazenamento de dados genotípicos e fenotípicos associados aos acessos genéticos. Durante o gerenciamento do germoplasma, é fundamental garantir a correta documentação e armazenamento dos dados relacionados aos acessos. Isso inclui informações como produção/ produtividade, desempenho dos materiais em diferentes ambientes, detalhes sobre resistência a estresses bióticos e abióticos, características morfológicas relevantes, dados de marcadores moleculares e *fingerprint* genético. Esses dados são valiosos para a avaliação e a seleção dos materiais, auxiliando na tomada de decisões informadas ao longo do programa de melhoramento.

Campos de Avaliação/Produção

Aqui o melhorista planeja e executa diferentes campos (áreas de produção ou avanço genético e/ou ensaios de avaliação). Cada campo é composto por um grupo específico de genótipos e tem um objetivo bem definido. Nesse sentido, o acompanhamento e a rastreabilidade das atividades relacionadas a cada campo são essenciais para o sucesso do programa. Essas atividades incluem a avaliação e seleção de populações, progênies, híbridos e linhagens, onde são realizadas avaliações minuciosas das diferentes populações, progênies, híbridos e linhagens presentes no campo, aplicando critérios específicos para avaliar características desejáveis. Com base nessas avaliações, são feitas seleções criteriosas visando identificar os genótipos mais promissores.

Após a seleção dos genótipos promissores, é necessário produzir sementes para avançar no programa de melhoramento. Isso pode ser feito por meio de autofecundação, em que as plantas selecionadas são cruzadas com elas mesmas para garantir a

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

fixação das características desejáveis, ou por meio de cruzamentos controlados entre diferentes genótipos, visando combinar características complementares.

Um processo fundamental no campo de melhoramento genético de plantas é o *grow-out*, que consiste no cultivo e monitoramento das plantas selecionadas até que atinjam um estágio específico de desenvolvimento. Durante esse período, as plantas são cuidadosamente observadas, avaliadas e selecionadas com base em critérios estabelecidos. O *grow-out* permite verificar o desempenho das plantas em condições de campo, considerando fatores como interações genótipo-ambiente e estabilidade das características desejáveis. Além disso, também é utilizado para purificar linhagens, ou seja, realizar a seleção e fixação de características desejáveis em populações homogêneas.

A purificação de linhagens é um processo de seleção contínua realizado no campo para eliminar variações indesejáveis e garantir a estabilidade e uniformidade das linhagens desenvolvidas. Envolve a seleção e autopolinização das plantas escolhidas, repetidas vezes, visando a fixação das características desejáveis. Esse processo é repetido por várias gerações até que se obtenha uma linhagem pura, ou seja, uma população que apresenta alta uniformidade genética e estabilidade nas características de interesse. Posteriormente o desenvolvimento das linhagens, estas podem ser hibridizadas, compondo genótipos híbridos com elevado vigor de desempenho em comparação com seus genitores.

Dados

No escopo da organização e gestão de um programa de melhoramento de plantas, as ações relacionadas aos dados podem ser divididas em cinco etapas principais. A primeira etapa é o planejamento e definição dos ensaios, onde são estabelecidas as metodologias a serem utilizadas, como o delineamento experimental, a escolha das unidades experimentais, a definição das variáveis a serem avaliadas e os critérios de medição. Essa fase é crucial para garantir a qualidade e a comparabilidade dos dados coletados.

Em seguida, temos a avaliação/aquisição dos dados, na qual são realizadas as avaliações das plantas ou populações em estudo, utilizando critérios específicos previamente definidos. São coletados dados sobre características relevantes, como produtividade, resistência a doenças, tolerância a estresses, entre outras. A coleta de dados pode ser feita por meio de observações diretas, medições ou análises laboratoriais, dependendo das características em avaliação. Após a

coleta dos dados, é realizada a etapa de análise dos dados. Nessa fase, uma análise estatística é conduzida com o objetivo de extrair informações e inferências a partir dos dados coletados. Essa análise estatística é adaptada aos objetivos de cada experimento e pode incluir técnicas como análise de variância, regressão, correlação, entre outras. Essa etapa permite uma interpretação objetiva dos resultados obtidos.

A interpretação dos resultados e a tomada de decisão são realizadas a partir da combinação dos resultados das análises estatísticas com outras informações relevantes sobre os experimentos, como informações sobre os tratamentos aplicados, condições de cultivo e características dos genótipos avaliados. Essa interpretação permite obter insights sobre o desempenho das plantas, identificar genótipos promissores e tomar decisões sobre os próximos passos do programa de melhoramento. Geralmente, essa etapa envolve reuniões de seleção e avanço realizadas por equipes multidisciplinares das empresas, que combinam conhecimentos técnicos e científicos para tomar decisões embasadas nos resultados. A gestão está intimamente ligada ao gerenciamento de germoplasma. Isso envolve o manejo, conservação e disponibilização de recursos genéticos vegetais. São estabelecidos bancos de germoplasma, realizada a coleta e conservação de amostras representativas da diversidade genética das plantas, documentação e registro dos acessos, além da disponibilização desses recursos para uso no programa de melhoramento.

Inventário

O inventário desempenha um papel fundamental na rastreabilidade e na proteção da propriedade intelectual. Nos últimos anos, essa etapa tem adquirido uma importância significativa nas empresas de melhoramento, visando garantir a segurança e dificultar a perda ou o roubo de germoplasma. O controle adequado do inventário permite rastrear todo o processo, fornecendo informações detalhadas sobre a origem, a produção, a colheita, o processamento, o armazenamento e a manipulação de lotes específicos de sementes, vinculados a cada acesso de germoplasma.

Conforme mencionado anteriormente, todas as etapas do programa de melhoramento estão interligadas, e os dados gerados em cada uma delas são interdependentes. A qualidade dos softwares de gestão de programas de melhoramento está diretamente relacionada à capacidade de manter os dados de inventário, populações, gerações, genótipos, campos, locais e anos, de forma

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

simples e acessível. É crucial que esses sistemas de gestão sejam capazes de armazenar e recuperar informações de maneira eficiente, garantindo a integridade e a disponibilidade dos dados em todo o processo de melhoramento.

É importante ressaltar que, dentro dessas etapas, todas as informações são mantidas, permitindo o acompanhamento e o controle do pedigree dos materiais genéticos. Essas informações podem ser utilizadas para análises combinadas entre as gerações, possibilitando uma avaliação mais abrangente do desempenho dos genótipos. Além disso, os diferentes tipos de informações obtidos ao longo do programa de melhoramento, como dados fenotípicos e genotípicos, podem ser empregados na tomada de decisões estratégicas.

Considerações Finais

A genética e o melhoramento de plantas são áreas da ciência em que o conhecimento evolui muito rápido e isso traz oportunidades em diversas áreas. Esse fato se caracteriza pela multidisciplinaridade e pela rapidez de informações geradas. Nessa área, os resultados se refletem, positivamente, nos diversos setores agrícolas, na indústria alimentícia e no meio ambiente. Inovações tecnológicas e novas ferramentas que permitam o desenvolvimento de novas cultivares de maneira mais organizada, rápida e eficaz certamente contribuirão para o agronegócio brasileiro. Cabe realçar que é necessário trabalhar com a integração de todas essas informações geradas (Venske et al., 2021). Sabe-se, também, que a quantidade de informações obtidas é muito grande e fica evidente que os melhoristas do futuro terão de ser altamente qualificados, e rodeados por uma equipe multidisciplinar bem treinada.

Com isso, a gestão otimizada das informações dos programas de melhoramento tende a ganhar cada vez mais espaço e importância nos programas comerciais de melhoramento de plantas. Entende-se que, em termos de otimização da inteligência artificial e aprendizado de máquina, o monitoramento de condições ambientais mais localizadas e em tempo real, a integração de sistemas, o uso de tecnologias verdes e a maior integração entre setores serão fundamentais para os melhores resultados. Nesse sentido, espera-se que num futuro próximo as predições de comportamento de genótipos possam ser maximizadas (utilizando dados das diferentes origens), com elevadas taxas de acurácia dentro do melhoramento de precisão, e contribuam para que os programas de melhoramento incrementem os ganhos genéticos atrelados ao incremento de eficiência operacional (Figura 8).

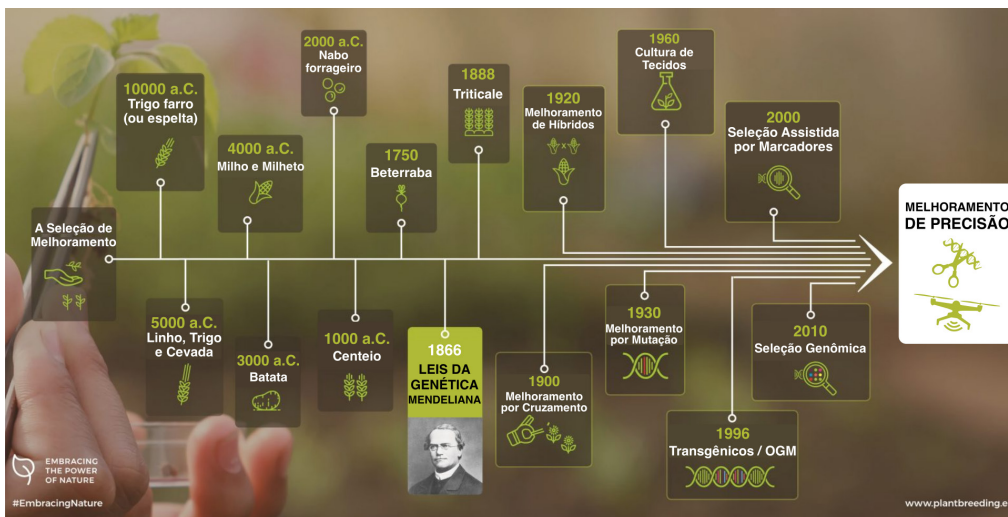


Figura 8. Evolução do melhoramento de plantas desde 10.000 a.C. até os tempos atuais, culminando da era do Melhoramento de Precisão.

Fonte: Adaptado de <https://www.euroseeds.eu/subjects/plant-breeding-innovation/>

Referências

- BERNARDO, R. **Breeding for quantitative traits in plants**. 3rd ed. Woodbury: Stemma Press, 2020. 422 p.
- CARENA, M. J. Germplasm enhancement and cultivar development: the need for sustainable breeding. **Crop Breeding and Applied Biotechnology**, v. 21, e385621S4, July 2021. DOI: <https://doi.org/10.1590/1984-70332021v21Sa17>.
- CHAIKAM, V.; MOLENAAR, W.; MELCHINGER, A. E.; BODDUPALLI, P. M. Doubled haploid technology for line development in maize: technical advances and prospects. **Theoretical and Applied Genetics**, v. 132, p. 3227–3243, Sept. 2019. DOI: <https://doi.org/10.1007/s00122-019-03433-x>.
- CLAUSEN, R. E.; MANN, M. C. Inheritance in *Nicotiana Tabacum*: V. The occurrence of haploid plants in interspecific progenies. **Proceedings of the National Academy of Sciences of the United States of America**, v. 10, n. 4, p. 121-124, Apr. 1924. DOI: <https://doi.org/10.1073/pnas.10.4.121>.
- COELHO, M. A. O.; CONDÉ, A. B. T.; YAMANAKA, C. H.; CORTE, H. R. Avaliação da produtividade de trigo (*Triticum aestivum* L.) de sequeiro em Minas Gerais. **Bioscience Journal**, v. 26, n. 5, p. 717-723, Sept./Oct. 2010.
- CONDÉ, A. B. T.; COELHO, M. A. O.; YAMANAKA, C. H.; CORTE, H. R. Adaptabilidade e estabilidade de genótipos de trigo sob cultivo de sequeiro em Minas Gerais. **Pesquisa Agropecuária Tropical**, v. 40, n. 1, p. 45-52, jan./mar. 2010.

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

COSTA-NETO, G.; FRITSCHÉ-NETO, R. Enviromics: bridging different sources of data, building one framework. **Crop Breeding and Applied Biotechnology**, v. 21, e393521S12, Aug. 2021. DOI: <https://doi.org/10.1590/1984-70332021v21Sa25>.

CROSSA, J.; FRITSCHÉ-NETO, R.; MONTESINOS-LOPEZ, O. A.; COSTA-NETO, G., DREISIGACKER, S.; MONTESINOS-LOPEZ, A.; BENTLEY, A. R. The modern plant breeding triangle: optimizing the use of genomics, phenomics, and enviromics data. **Frontiers in Plant Science**, v. 12, 651480, Apr. 2021. DOI: <https://doi.org/10.3389/fpls.2021.651480>.

CROSSA, J.; PÉREZ-RODRÍGUEZ, P.; CUEVAS, J.; MONTESINOS-LÓPEZ, O.; JARQUÍN, D.; CAMPOS, G.; BURGUEÑO, J.; GONZÁLEZ-CAMACHO, J. M.; PÉREZ-ELIZALDE, S.; BEYENE, Y.; DREISIGACKER, S.; SINGH, R.; ZHANG, X.; GOWDA, M.; ROORKIWAL, M.; RUTKOSKI, J.; VARSHNEY, R. K. Genomic selection in plant breeding: methods, models, and perspectives. **Trends in Plant Science**, v. 22, n. 11, p. 961-975, Nov. 2017. DOI: <https://doi.org/10.1016/j.tplants.2017.08.011>.

CRUZ, C. D.; VALENT, B. Wheat blast disease: danger on the move. **Tropical Plant Pathology**, v. 42, p. 210-222, May 2017. DOI: <https://doi.org/10.1007/s40858-017-0159-z>.

GAYNOR, R. C.; GORJANC, G.; BENTLEY, A. R.; OBER, E. S.; HOWELL, P.; JACKSON, R.; MACKAY, I. J.; HICKEY, J. M. A two-part strategy for using genomic selection to develop inbred lines. **Crop Science**, v. 57, n. 5, p. 2372-2386, Sept./Oct. 2017. DOI: <https://doi.org/10.2135/cropsci2016.09.0742>.

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>.

GRATTAPAGLIA, D.; KIRST, M. Eucalyptus applied genomics: from gene sequences to breeding tools. **New Phytologist**, v. 179, n. 4, p. 911-929, Sept. 2008. DOI: <https://doi.org/10.1111/j.1469-8137.2008.02503.x>.

LAWRENCE-DILL, C. J.; SCHNABLE, P. S.; SPRINGER, N. M. Idea factory: the maize genomes to fields initiative. **Crop Science**, v. 59, n. 4, p. 1406-1410, July/Aug. 2019. DOI: <https://doi.org/10.2135/cropsci2019.02.0071>.

MACIEL, J. L. N.; SANTOS, G. B.; PIZOLOTTO, C. A.; KOVALESKI, M.; SILVA, A. N.; DEUNER, C. C.; COSTA, I. V. D. Resistance of Brazilian wheat cultivars to blast under controlled condition. **Ciência Rural**, v. 52, e20210417, 2022. DOI: <https://doi.org/10.1590/0103-8478cr20210417>.

MARINHO, C. D.; COELHO, I. F.; PEIXOTO, M. A.; CARVALHO JUNIOR, G. A.; RESENDE JUNIOR, M. F. R. Genomic selection as a tool for maize cultivars development. **Revista Brasileira de Milho e Sorgo**, v. 21, e1285, mar. 2022. DOI: <https://doi.org/10.18512/rbms2022v21e1285>.

NASS, L. L. Utilização de recursos genéticos vegetais no melhoramento. In: NASS, L. L.; VALOIS, A. C. C.; MELO, I. S.; VALADARES-INGLIS, M. C. (ed.). **Recursos genéticos e melhoramento-plantas**. Rondonópolis: Fundação MT, 2001. p. 29-55.

PÁDUA, J. M. V.; CONDÉ, A. B. T.; MARTINS, F. A. D.; NUNES, J. A. R.; AGOSTINETTO, A. **Cultivo do trigo tropical – safra 2022: aprendizados e recomendações para a região Sul de Minas Gerais e Campo das Vertentes**. Belo Horizonte: Epamig, 2022. 6 p. (Epamig. Circular técnica, 365).

RAMALHO, M. A. P.; ABREU, A. F. B.; SANTOS, J. B.; NUNES, J. A. R. **Aplicações da genética quantitativa no melhoramento de plantas autógamas**. Lavras: Ufla, 2012. 522 p.

RAMSTEIN, G. P.; JENSEN, S. E.; BUCKLER, E. S. Breaking the curse of dimensionality to identify causal variants in Breeding 4. **Theoretical and Applied Genetics**, v. 132, p. 559-567, 2019. DOI: <https://doi.org/10.1007/s00122-018-3267-3>.

RESENDE, M. D. V.; BARBOSA, M. H. P. **Melhoramento genético de plantas de propagação assexuada**. Colombo: Embrapa Florestas, 2005. 130 p. <http://www.alice.cnptia.embrapa.br/alice/handle/doc/286194>.

RESENDE, R. T.; CHENU, K.; RASMUSSEN, S. K.; HEINEMANN, A. B.; FRITSCHÉ-NETO, R. Editorial: Enviromics in plant breeding. **Frontiers in Plant Science**, v. 13, 935380, June 2022. DOI: <https://doi.org/10.3389/fpls.2022.935380>.

RESENDE, R. T.; PIEPHO, H.-P.; ROSA, G. J. M.; SILVA-JUNIOR, O. B.; SILVA, F. F.; RESENDE, M. D. V.; GRATTAPAGLIA, D. Enviromics 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>.

TEINGTHAM, K.; LA BORDE, N. D. Is doubled haploid production in sorghum impossible? **Applied Science and Engineering Progress**, v. 10, n. 4, p. 247–256, 2017. DOI: [10.14416/ij.ijast.2017.11.001](https://doi.org/10.14416/ij.ijast.2017.11.001).

VENSKE, E.; VERDI, C. A.; PETROLI, C. D.; PEGORARO, C.; MAIA, L. C.; OLIVEIRA, A. C. The 4. (n) wheat breeding era: genomic-based predictions becoming part of the daily routine and beyond. **Biotechnology Research and Innovation**, v. 5, e2021007, 2021. DOI: <http://dx.doi.org/10.4322/biori.21050202>.

VIEIRA, C. C.; CHEN, P. The numbers game of soybean breeding in the United States. **Crop Breeding and Applied Biotechnology**, v. 21, e387521S10, July 2021. DOI: <https://doi.org/10.1590/1984-70332021v21Sa23>.

VOLPATO, L.; PINTO, F.; GONZÁLEZ-PÉREZ, L.; THOMPSON, I. G.; BORÉM, A.; REYNOLDS, M.; GÉRARD, B.; MOLERO, G.; RODRIGUES JUNIOR, F. A. High throughput field phenotyping for plant height using UAV-based RGB imagery in wheat breeding lines: feasibility and validation. **Frontiers in Plant Science**, v. 12, 591587, Feb. 2021. DOI: <https://doi.org/10.3389/fpls.2021.591587>.

WARTHA, C. A.; LORENZ, A. J. Implementation of genomic selection in public-sector plant breeding programs: current status and opportunities. **Crop Breeding and Applied Biotechnology**, v. 21, e394621S15, Sept. 2021. DOI: <https://doi.org/10.1590/1984-70332021v21Sa28>.

WATSON, A.; GHOSH, S.; WILLIAMS, M. J.; CUDDY, W. S.; SIMMONDS, J.; REY, M. D.; HATTA, M. A. M.; HINCHLIFFE A.; STEED, A.; REYNOLDS, D.; ADAMSKI, N. M.; BREAKSPEAR, A.; DOROLEV, A.; RAYNER, T.; DIXON, L. E.; RIAZ, A.; MARTIN, W.; RYAN, M.; EDWARDS, D.; BATLEY, J.; RAMAN, H.; CARTER, J.; ROGERS, C.; DOMONEY, C.; HICKEY, L. T. Speed breeding is a powerful tool to accelerate crop research and breeding. **Nature Plants**, v. 4, p. 23–29, Jan. 2018. DOI: <https://doi.org/10.1038/s41477-017-0083-8>.

WEDZONY, M.; FORSTER, B. P.; ZUR, I.; GOLEMIEC, E.; SZCHYNSKA-HEBDA, M.; DUBAS, E.; GOTEBIOWSKA, G. Progress in doubled haploid technology in higher plants. In: TOURAEV, A.; FORSTER, B. P.; JAIN, S. M. (ed.). **Advances in haploid production in higher plants**. Dordrecht: Springer, 2009. Cap. 1, p. 1-33.

XU, Y. Envirotyping for deciphering environmental impacts on crop plants. **Theoretical and Applied Genetics**, v. 129, p. 653-673, Apr. 2016. DOI: <https://doi.org/10.1007/s00122-016-2691-5>.

Tavvs Micael Alves, Leonardo de Castro Santos,
Priscila Ferreira Batista

Introdução

A economia mundial enfrenta vários desafios para melhorar o nível de desenvolvimento econômico e qualidade de vida de uma população de mais 9,7 bilhões de pessoas até 2050 (United Nations, 2023). O crescimento global do Produto Interno Bruto (PIB) vem desacelerando com as tensões comerciais entre as principais economias globais e níveis crescentes de dívida pública (International Monetary Fund, 2022). A pandemia da covid-19 também pressionou o desenvolvimento tecnológico de empresas de todos os setores, incluindo aquelas envolvidas em programas de melhoramento de plantas (Popkova et al., 2022). Além desses desafios, a iniciativa privada enfrenta riscos e oportunidades específicas da inovação tecnológica na agricultura (Leal; Figueiredo, 2021). Por exemplo, a regulamentação sobre organismos geneticamente modificados pode ser complexa e onerosa, exigindo altos custos relativos a análises de biossegurança e o tempo de aprovação. Além disso, muitos investidores podem ver o desenvolvimento de cultivares como um investimento de alto risco, dificultando o acesso ao capital. As empresas que desenvolvem cultivares comerciais de espécies de interesse econômico também enfrentam a rápida mudança nas preferências dos consumidores por produtos orgânicos, o impacto das mudanças climáticas na produtividade agrícola, a escassez de mão de obra devido à automação do trabalho agrícola por máquinas e drones, e a necessidade de investimentos significativos em infraestrutura e treinamento de pessoal (Devaux et al., 2021; Oliveira et al., 2023). Por fim, há falta de coordenação entre instituições de pesquisa do setor público e empresas privadas em relação aos esforços de desenvolvimento e comercialização. Essa falta de colaboração pode resultar em duplicação de esforços, desperdício de recursos e atrasos no processo de desenvolvimento.

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

Os desafios de inovação e desenvolvimento tecnológico também pressionam as empresas do setor agrícola a acompanharem o rápido avanço de tecnologias disruptivas envolvendo inteligência artificial (IA), internet das coisas (IoT), *blockchain*, *big data* e robótica. Esse capítulo abordará os principais desafios da inovação, formação profissional, propriedade intelectual (PI) e empreendedorismo para utilização de novas tecnologias em programas de melhoria de precisão. Investir em tecnologias digitais, criação de uma cultura inovadora e delineamento de rotas tecnológicas podem garantir o sucesso econômico e superação dos desafios da indústria e academia.

Invenção, Inovação e Propriedade Intelectual

Os termos inovação e invenção são conceitos relacionados, mas possuem diferenças (Tirmizi et al., 2020). Invenção refere-se à criação de algo novo, como um dispositivo, método ou processo que não existia antes. A inovação se apoia em avanços da pesquisa e desenvolvimento (P&D), mas é realizada por meio de decisões de gestão no sentido de dedicar recursos à mudança de posição concorrencial da empresa em uma ou várias dimensões do sistema comercial e industrial. A inovação envolve investimento significativo de tempo, dinheiro e recursos humanos para desenvolver, testar e refinar uma invenção, tornando-a viável, lucrativa e útil para o mercado (Roberts, 2007; Solheim, 2022). Portanto, inovação não é uma ciência, tecnologia ou invenção. A inovação empresarial é sistêmica e exige a consideração minuciosa de todos os aspectos da atividade econômica.

Propriedade intelectual (PI) é um sistema estabelecido para garantir a propriedade ou direitos exclusivos decorrentes de esforços intelectuais na indústria, ciência, literatura e arte (Bagnato et al., 2016). Fonseca (1998) define a patente como um monopólio legal de uso exclusivo de uma invenção e com duração de 20 anos no Brasil. É uma propriedade e, como tal, pode ser comprada, vendida, transferida ou licenciada. A proteção impede a cópia da invenção e qualquer tipo de exploração sem o consentimento do autor. Com a obtenção de uma patente, marca, desenho ou modelo, segredo comercial ou industrial, e direito autoral, as empresas transformam as ideias em propriedade intelectual (Quintino et al., 2021).

Inovação Tecnológica e a Competitividade das Empresas

O fundamento econômico da proteção moderna da PI está relacionado com benefícios econômicos. Porém, as características de “bem público” devem ser neutralizadas para que seja possível obter benefícios financeiros. A propriedade intelectual permite essa

neutralização ao estabelecer propriedade sobre conhecimentos e informações, e isso transforma conhecimentos e informações em bens privados durante períodos limitados (Bagnato et al., 2016). Portanto, a proteção da PI resulta em privilégios aos criadores pela exclusão da utilização do direito por outras pessoas que não possuem autorização.

Propriedade intelectual é a base do direito à remuneração (Brasil, 1996). A proteção da PI fornece incentivos que asseguram a geração de novas invenções e criações, estimulando a inovação e promovendo o crescimento econômico sustentável. Se não há meios para proteger os conhecimentos, os inventores e/ou inovadores podem sofrer uma desvantagem, pois não recuperariam o investimento da inovação. O investimento no desenvolvimento de uma nova cultivar, por exemplo, pode requerer vários aportes de recursos financeiros durante 20 anos (Sampaio, 1998). Uma vez que ideias valiosas e mentes inventivas não estão amplamente disponíveis no mercado, o registro de ideias e tecnologias assegura sua propriedade e coloca valores tangíveis aos esforços intelectuais (Barrett, 2005). O desafio de levar soluções ao mercado é ainda maior no âmbito público do que no setor privado. Muitos gestores públicos ainda precisam entender que ideias novas e criativas oriundas do setor público podem ser integradas a negócios de sucesso e benefícios para a sociedade. A falta de cooperação e diálogo entre os setores público e privado é uma barreira para a transferência de tecnologia e inovação entre os setores.

As empresas privadas muitas vezes possuem soluções criativas e inovadoras que poderiam ser adaptadas e incorporadas ao setor público (e vice-versa), mas a falta de interação e de entendimento mútuo dificulta a transferência de conhecimento (Diniz et al., 2018). Para superar os desafios da implementação de ideias inovadoras no setor público, é preciso investir em capacitação e incentivos para gestores públicos e promover uma cultura de diálogo e cooperação entre o setor público e privado. Ainda que em ritmo muito aquém das inovações no setor privado, o movimento em favor da inovação no setor público vem ocorrendo há alguns anos em vários países, incluindo o Brasil. Pode-se citar a existência de alguns prêmios que estimulam a criação de ideias inovadoras, como o Concurso de Inovação no Setor Público, promovido anualmente desde 1996 pela Escola Nacional da Administração Pública – ENAP (Santos, 2022). Fernandes (2019) verificou que a maioria dos casos de inovação está voltada para a melhoria operacional de processos.

Para as empresas, a inovação é tipicamente realizada em função das decisões do mercado sobre recursos financeiros e humanos (investimento). O desenvolvimento de novos produtos e processos requer prazos cada vez mais curtos para o lançamento no mercado. No entanto, o momento de entrada no mercado (*time to market*) precisa ser

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

balanceado com a qualidade da primeira versão do produto colocada para lançamento (McNally et al., 2011). O ciclo de vida dos produtos também tem diminuído. Há um processo seletivo rigoroso em que apenas algumas ideias e invenções resultarão em inovação. Portanto, a inovação deve ser uma atividade contínua. Uma empresa líder em determinado mercado perderá rapidamente sua posição a partir do momento em que deixar de inovar. Investir em tecnologias digitais e formação profissional são essenciais para criar novas rotas de desenvolvimento tecnológico e ambientes de negócios capazes de superar os atuais e futuros desafios econômicos (Santos et al., 2016).

Não há fronteiras para a difusão do conhecimento em um mundo cada vez mais globalizado. Por um lado, a comunicação facilitada impõe um risco adicional de essas inovações serem copiadas, utilizadas ou cair na posse de concorrentes a custos relativamente baixos. Por outro lado, a comunicação instantânea favorece a participação integrada de empresários, empreendedores, inventores, cientistas e artistas. Esses dispõem de meios e materiais para inventar, proteger, divulgar e comercializar suas criações e inovações a nível global. As ideias precisam ser desenvolvidas, testadas, validadas, transformadas em produtos, processos ou serviços inovadores. Também precisam ser valoradas, divulgadas e comercializadas com sucesso. Portanto, antes da decisão sobre o investimento em tempo, talento, recurso econômico e financeiro no desenvolvimento de um novo empreendimento/produto/processo, é primordial e essencial se avaliar o potencial mercadológico da ideia.

Além de ter acesso ao conhecimento técnico, é importante que as organizações entendam que tipo de ambiente existe ao seu redor. Isso inclui a análise de fatores internos, como a cultura da empresa e estrutura organizacional. Também há necessidade de avaliar variáveis externas, como tendências de mercado, cenário competitivo, preferências do cliente, etc. Avaliar o nível de maturidade tecnológica (TRL - *technology readiness level*) da solução ajuda a determinar os tipos de estratégias mais adequados para transformar a atual ideia ou tecnologia em uma inovação disruptiva (Embrapa, 2023). Uma vez escolhida a estratégia apropriada, traduzir a rota tecnológica em decisões de negócios é fundamental para iniciar a transformação digital de organizações públicas e privadas, como exemplificado para o melhoramento de precisão (Figura 1). A inovação envolve motivar as pessoas, definir objetivos claros, delinear o que se deseja alcançar, usar ferramentas e técnicas modernas, planejar as fases de execução, definir etapas críticas e metas mensuráveis, e realizar o acompanhamento rigoroso dos processos, orçamento e cronograma. Essas estratégias podem ser viabilizadas por diferentes modelos de negócios como a cooperação tecnológica, fornecimento de tecnologia, licenciamento de direitos sobre patentes e cultivares, dentre outros.



GMO = *Genetically Modified Organism* (Organismo Geneticamente Modificado);
 IoT = Internet of Things (Internet das Coisas); IA = Inteligência Artificial;
 TT = Transferência de Tecnologia; PI = Propriedade Intelectual.

Figura 1. Panorama da inovação em melhoramento de precisão.

Parcerias com Universidades e o Fomento à Inovação

A indústria é um dos principais interessados no ecossistema de inovação e na qualidade da pesquisa e desenvolvimento. Vários novos produtos e serviços colocados no mercado dependem de resultados de P&D e, por isso, parcerias entre a indústria e academia podem fornecer suporte valioso a pesquisadores e empresas. Esse suporte pode ajudar a mitigar os riscos associados à inovação e garantir que novas ideias sejam comercializadas com sucesso. A colaboração entre empresas, universidades e startups é essencial para a inovação (Alänge et al., 2022).

Startups são empresas jovens que buscam um modelo de negócio escalável (Blank; Dorf, 2012). As startups se empenham em inovar em um mercado pouco explorado ou com alta concorrência, apresentando um perfil de risco mais elevado. A agilidade e flexibilidade são características importantes dessas empresas, que buscam solucionar problemas e atender às necessidades dos clientes de forma rápida e eficiente, muitas vezes utilizando métodos ágeis de gestão e desenvolvimento de produtos (Ries, 2011; Cohn, 2017). As startups também são importantes no processo de inovação, pois muitas vezes trazem novos produtos e serviços para o mercado baseados em ideias e tecnologias inovadoras.

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

As startups podem se beneficiar de parcerias com universidades para mitigar os riscos da criação de um novo negócio. As startups afiliadas a universidades têm acesso a financiamento, orientação e outros recursos que podem ajudá-las a levar suas ideias ao mercado (Barrett, 2005). Além disso, muitas universidades possuem incubadoras e aceleradoras que fornecem suporte e recursos especificamente projetados para ajudar no sucesso das startups. Portanto, as universidades fornecem um terreno fértil para inovação, empreendedorismo e progresso tecnológico por meio de suas atividades de pesquisa e ensino.

As parcerias com o setor público permitem que as empresas tenham acesso a novas tecnologias e conhecimentos para aprimorar os serviços oferecidos aos seus clientes. Um exemplo de parceria envolve uma startup de fenotipagem baseada em plataformas de monitoramento agrícola com o uso de drones e sensores, e a Universidade de São Paulo (Terramagna, 2023). Essa universidade também apoiou o desenvolvimento de outras soluções de mapeamento aéreo com drones (Pix4d Brasil, 2023). Outros exemplos são startups envolvidas no monitoramento de lavouras em parceria com a Universidade Federal de Lavras e Universidade Federal de Viçosa (Agrointeli, 2023; Smart Agri, 2023). A Rapid Genomics, uma startup sediada na Flórida, Estados Unidos, conta com sócios brasileiros especializados em biotecnologia de genotipagem. A empresa se destaca por suas tecnologias avançadas de sequenciamento de DNA de alta velocidade e precisão, que são empregadas na análise genômica de plantas, animais e microrganismos (Rapid Genomics, 2023). Outras startups também atuam em análises de dados multiômicos apoiadas pelo *know how* em bioinformática e aprendizado de máquina (BioTuring, 2023). Outra importante startup com aplicação no agronegócio atua na análise estatística de dados fenotípicos e genômicos em parceria com várias universidades e instituições de pesquisa no Brasil e no exterior (Agrotools, 2023). Algumas oportunidades para novas startups envolvem o desenvolvimento de redes de dispositivos físicos interconectados por internet das coisas (IoT) para monitorar e controlar variáveis ambientais em tempo real; estruturas de dados compostas por blocos interligados e imutáveis criados pela tecnologia *blockchain* para rastrear informações genéticas e proteger a propriedade intelectual; bem como sistemas computacionais que possuem a capacidade de utilizar inteligência artificial (IA) para automatizar a análise de dados genômicos e auxiliar na seleção de cruzamentos promissores.

Outra forma que as universidades contribuem para o desenvolvimento tecnológico é por meio da criação de empresas *spin-off*, que são criadas por professores universitários, funcionários ou alunos, que usam seu conhecimento e experiência para comercializar suas pesquisas. Essas empresas geralmente trazem novos produtos e

serviços para o mercado graças ao apoio e aos recursos das universidades. As *spin-offs* também ajudam a criar empregos e impulsionar o crescimento econômico em suas comunidades locais. As empresas de inovação aberta que estabelecem parcerias com universidades podem ser consideradas como um tipo de *spin-off*, pois a criação dessa nova empresa é impulsionada pela transferência de tecnologias, patentes e conhecimentos gerados durante a colaboração (Miranda et al., 2018).

Um aspecto importante da inovação é o compartilhamento de riscos financeiros mediante parcerias entre universidades e indústrias. A inovação aberta é um modelo de inovação em que as empresas se abrem para a colaboração externa, visando aumentar sua eficiência e capacidade de inovar (Chesbrough et al., 2006). Esse modelo de inovação está relacionado à ideia de cocriação e colaboração entre empresas e instituições de pesquisa, incluindo universidades (Van de Vrande et al., 2011). As startups enfrentam riscos significativos nos estágios iniciais de seu desenvolvimento, pois muitas vezes carecem dos recursos e conhecimentos necessários para levar suas ideias ao mercado. No Brasil, o apoio financeiro concedido pelo governo pode ser realizado indiretamente por meio de benefícios fiscais, como a Lei do Bem, Rota 2030 e Lei de Informática (Piazentini et al., 2021). Há também o apoio financeiro direto por meio de diversos mecanismos, como bolsas de recursos humanos, recursos não reembolsáveis da Empresa Brasileira de Pesquisa e Inovação Industrial (Embrapii), investimento obrigatório e subsídios econômicos. As empresas também podem acessar recursos reembolsáveis na forma de empréstimos ou buscar investimento de capital de risco, *crowdfunding* e investidores anjos. Cada tipo de financiamento tem suas próprias regras e exigências, mas geralmente as empresas precisam realizar aportes com recursos não reembolsáveis em todos os projetos.

Financiamento de capital e capital de risco são duas formas importantes de compartilhamento de riscos em inovação, pois fornecem às startups o capital necessário para crescer. O investimento de capital em startups geralmente é realizado em troca de participação na empresa (*equity*). Esse financiamento por ações e capital de risco desempenham papel crucial no desenvolvimento da tecnologia e podem retornar financeiramente aos investidores se a startup for bem-sucedida (Baeta; Leite, 1990). O financiamento público também ajuda a mitigar os riscos associados à inovação, fornecendo a pesquisadores e empresas os recursos de que precisam para o desenvolvimento de novas ideias. O financiamento público ajuda a garantir que a inovação seja focada no bem público, contratando pesquisadores e fornecendo os recursos necessários para conduzir pesquisas de ponta em áreas prioritárias para o desenvolvimento econômico e social (Brasil, 2016, 2021). Além disso, o financiamento

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

público pode ajudar a apoiar a comercialização da pesquisa universitária, fornecendo financiamento para empresas *spin-off* e startups.

Tecnologias Digitais e a Formação Profissional

Devido à competitividade e complexidade do setor agrícola, investir em tecnologias digitais e criação de uma cultura inovadora se tornam cada vez mais importantes para modernizar os programas de melhoramento de plantas. As empresas precisam de acesso a recursos suficientes para que possam alavancar os investimentos existentes enquanto fazem investimentos estratégicos em tecnologias digitais emergentes, como internet das coisas (IoT), *blockchain*, seleção genômica, automação, análise de *big data*, inteligência artificial, robótica e plataformas digitais. A combinação de várias abordagens fornece às empresas maneiras de utilizar soluções inovadoras em suas operações, permitindo-lhes flexibilidade na hora de tomar decisões sobre qual rota melhor se adapta às suas necessidades (Figura 1).

Novas tecnologias digitais podem ter aplicações potenciais na indústria, mas exigem investimento financeiro antes que qualquer retorno possa ocorrer. As tecnologias digitais emergentes podem permitir que pesquisadores e produtores não apenas coletem dados mais precisos sobre as plantas, mas também tomem decisões mais bem informadas com base em algoritmos preditivos. Essas tecnologias fornecem maior percepção das condições do solo, estratégias de controle de pragas e métodos de irrigação e, ao mesmo tempo, reduzem os custos de mão de obra associados às atividades manuais das operações. Além disso, permitem acompanhar o desempenho das plantas ao longo do tempo por meio de painéis em tempo real que permitem ajustar as atividades durante o cultivo. Esse tipo de automatização permite melhorar a eficiência do uso de recursos humanos e financeiros (Figura 1).

Para que as organizações envolvidas em programas de melhoramento de plantas capitalizem as oportunidades envolvendo tecnologias digitais, um dos primeiros passos é adotar uma agenda de aprendizado focada no desenvolvimento profissional contínuo de seus colaboradores. As *soft skills* estão se tornando cada vez mais importantes no mercado de trabalho, pois os empregadores buscam indivíduos que possuam não apenas conhecimentos técnicos, mas também a capacidade de trabalhar bem em equipe, comunicar-se de forma eficaz e adaptar-se efetivamente às constantes mudanças (Smaldone et al., 2022; Espina-Romero et al., 2023). Ao favorecer o desenvolvimento de habilidades de liderança na equipe, por exemplo, os indivíduos mais confiantes e eficazes podem assumir funções e responsabilidades mais desafiadoras. Portanto, incentivar os colaboradores a pensar criativamente em

todos os níveis da organização e no seu desenvolvimento profissional podem levar a uma maior satisfação dos indivíduos no trabalho, melhorar seus relacionamentos profissionais, e aumentar os resultados e faturamento da empresa.

A pós-graduação é outro caminho para o desenvolvimento profissional, fornecendo aos alunos conhecimentos e habilidades avançados em um campo específico. A obtenção de um diploma de pós-graduação pode levar a melhores oportunidades de emprego, salários mais altos e maior satisfação no trabalho. Além disso, a pós-graduação oferece aos indivíduos a oportunidade de interagir com outros profissionais em suas áreas e participar de projetos de pesquisa de ponta. No Brasil, o retorno financeiro do investimento em pós-graduação é maior do que o retorno do investimento em graduação em diversas áreas de formação (Barbosa Filho; Pessôa, 2008; Costa et al., 2014; Barth et al., 2016). Além disso, é importante lembrar que o impacto da pós-graduação na carreira de um profissional pode variar de acordo com a área de atuação e com a demanda do mercado de trabalho em cada setor. Assim como em outros países, é importante ressaltar que a obtenção de um diploma de pós-graduação não é uma garantia de sucesso profissional, e outros fatores, como experiência e habilidades, também desempenham papel importante na carreira do profissional.

O empreendedorismo é outro aspecto do desenvolvimento profissional que está se tornando cada vez mais importante. Empreendedores são indivíduos que iniciam e administram seus próprios negócios, assumindo os riscos e responsabilidades inerentes à sua atividade profissional. O empreendedorismo pode proporcionar aos indivíduos maior controle sobre suas carreiras, a capacidade de trabalhar em projetos pelos quais possuem grande afinidade e o potencial de retorno financeiro. Em áreas específicas, como o melhoramento de precisão, essas habilidades podem ser aplicadas para identificar oportunidades de negócio, desenvolver soluções inovadoras, liderar equipes e gerenciar projetos de forma eficiente. Além disso, o intraempreendedorismo tem se destacado como uma abordagem promissora no desenvolvimento profissional. Intraempreendedores são colaboradores que demonstram características empreendedoras dentro de uma organização, trazendo inovação, criatividade e iniciativa para impulsionar o crescimento da empresa.

Registro e Proteção de Propriedade Intelectual

A imaginação e a criatividade têm impulsionado o avanço tecnológico desde as primeiras formas de escrita até as inovações mais recentes, como o transistor, a nanotecnologia e a edição gênica de plantas. A inventividade humana permite enfrentar desafios, melhorar a qualidade de vida e explorar novas fronteiras em todas

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

as áreas do conhecimento (WIPO, 2022). No passado, as tecnologias eram protegidas exclusivamente por segredos comerciais, acordos de não divulgação e contratos com funcionários para impedir que concorrentes acessassem ou copiassem a PI. Hoje, o meio legal mais comum de proteger um invento é obter uma patente (WIPO, 2023). Conforme orientação do Instituto Nacional de Propriedade Industrial (INPI), o registro de um invento envolve atender condições gerais, tais como: demonstrar as novas características que não sejam conhecidas no corpo dos conhecimentos existentes, não ser evidente ou envolver algo que possa ser deduzido por uma pessoa com conhecimento médio naquele campo técnico (quesito de atividade inventiva), ter um uso prático ou capacidade de aplicação industrial e, finalmente, deve constituir o que se denomina “matéria patenteável” nos termos da lei aplicável, Lei nº 9.279/1996 (Brasil, 1996). As patentes são conferidas pelo escritório de patentes do país em que se deseja proteger o invento. No Brasil, os pedidos de patentes são depositados no INPI. Direitos patentários são conferidos em troca da revelação completa da tecnologia pelo inventor ao público, por meio do pedido de patente.

As patentes, sejam de invenções ou de processos, não podem ser consideradas o único instrumento da atividade inovadora e geração/registo de PI. A proteção dos segredos industriais e comerciais, bem como proteção do *know-how*, proteção de desenhos e modelos continuam sendo importantes para proteger um invento pelas indústrias e empresas (Hull, 2019). A proteção dos segredos comerciais permite a preservação da natureza confidencial da informação contra a revelação indevida e contra o uso por pessoas não autorizadas (Hull, 2019).

Os Estados membros da Organização Mundial da Propriedade Intelectual (OMPI ou WIPO em inglês) concordam sobre a importância de incentivar continuamente a inovação e assegurar a proteção como propriedade intelectual dos conhecimentos resultantes. Os Estados membros da OMPI empregaram recursos consideráveis para construir a infraestrutura global de atribuição de direitos de propriedade aos inventores (patentes, segredos comerciais e industriais), aos criadores (direitos autorais, desenhos ou modelos) e fornecedores de produtos e serviços (Castro, 2018).

Os Direitos de Obtentores aplicam-se à proteção de novas variedades vegetais e são regidos no Brasil pela Lei de Proteção de Cultivares, Lei nº 9.456/1997 (Brasil, 1997). Cultivar é descrita como a variedade de qualquer gênero ou espécie vegetal superior que seja claramente distinguível de outras cultivares conhecidas por margem mínima de descritores, por sua denominação própria, que seja homogênea e estável quanto aos descritores através de gerações sucessivas e seja de espécie passível de uso pelo complexo agroflorestral. A nova cultivar precisa ser descrita em publicação

especializada disponível e acessível ao público, bem como a linhagem componente de híbridos, resultantes do melhoramento genético realizado pelo melhorista. As cultivares protegidas não podem ser reproduzidas comercialmente sem a autorização formal do(s) inventor(es). Um caminho para que o interessado possa multiplicar as sementes e mudas da nova variedade envolve a negociação de *royalties* (Silva et al., 2021). A Lei de Proteção de Cultivares estabelece que qualquer gênero ou espécie vegetal da nova cultivar ou a cultivar essencialmente derivada poderá ser protegida (Artigo 4º).

Um sistema efetivo de proteção de obtenções vegetais visa encorajar o desenvolvimento de novas variedades de plantas para o benefício da sociedade em áreas como agricultura, horticultura e reflorestamento. A criação de novas variedades vegetais requer substancial investimento (conhecimento, trabalho, dinheiro, tempo, etc.). A União para a Proteção de Obtenções Vegetais (UPOV) estabelece a proteção de variedades vegetais mediante um “direito do obtentor”, que é uma forma *sui generis* de direito de propriedade intelectual (um sistema especial para obtenções vegetais). Em virtude do acordo TRIPS (Acordo sobre Aspectos dos Direitos de Propriedade Intelectual Relacionados ao Comércio), os membros da Organização Mundial do Comércio (OMC) são obrigados a prover proteção a obtenções vegetais, seja mediante patentes, seja por meio de um efetivo sistema *sui generis* ou por uma combinação entre ambos (Boff, 2019). A proteção é conferida em um nível nacional. Entretanto, a UPOV permite proteção em um nível supranacional, que pode reduzir consideravelmente os custos e esforços necessários para obter proteção em diversos países.

Diferentes formas de proteção à PI estão disponíveis para variedades de plantas e invenções biotecnológicas, incluindo os direitos de melhoristas de plantas ou proteção de variedades de plantas; patentes de novos genes, características e processos; patentes de utilidade em variedades vegetais; uso do direito contratual; mecanismos biológicos (culturas híbridas); e segredos comerciais. Os direitos de melhoristas de plantas constituem-se no método padrão de proteção de cultivares na maioria dos países (Silva et al., 2021). Porém, a Lei nº 9.279/1996 impede que plantas e animais sejam patenteados, incluindo os clones de espécies arbóreas e seus híbridos. Como exceção, os microrganismos transgênicos podem ser patenteados.

Não é vedada a patenteabilidade de processos para a obtenção de seres vivos, desde que atendam aos requisitos básicos de novidade, inventividade e utilidade prática industrial (Wolff, 1998). As novas espécies agrícolas, híbridos ou cultivares precisam ser registradas como cultivares comerciais pelo Registro Nacional de Cultivares (RNC). A proteção da propriedade intelectual no setor de melhoramento vegetal foi consolidada pelo RNC, com resultados imediatos para a pesquisa agropecuária e a produção de

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

sementes. Além disso, a Lei de Proteção de Cultivares criou o Serviço Nacional de Proteção de Cultivares (SNPC), responsável por gerenciar aspectos administrativos e técnicos dos materiais vegetais desenvolvidos. O SNPC mantém uma base de dados, preserva amostras vivas para fiscalização e monitora as características originais de cultivares protegidas no território nacional (Croplife, 2019).

Uma das principais atividades da UPOV é estabelecer os princípios gerais para o exame das variedades de plantas quanto à novidade (que não tenha sido comercialmente explorada durante um certo período anterior ao pedido), distinguibilidade (distinguível de qualquer outra variedade cuja existência é questão de conhecimento comum), uniformidade (plantas de uma obtenção devem ser uniformes nas suas características relevantes, salvo as variações previsíveis havidas tendo em conta as particularidades de sua multiplicação ou reprodução), estabilidade (a variedade deve permanecer sem modificações nas suas características relevantes após sucessivas reproduções ou multiplicações) e diretrizes e denominação adequadas para as espécies. Esses documentos normativos são periodicamente atualizados e estendidos a um número crescente de gêneros e espécies. No campo das patentes, o cenário internacional apresenta marcos legais como as Convenções de Paris para a Proteção da PI, de 1883 e revista em 1967, firmada por 109 países (Castro, 2018). O Brasil é signatário das duas edições do Acordo.

De acordo com o Ato de 1991 da UPOV, os materiais de multiplicação ou reprodução (a semente ou planta – ou ainda parte de planta utilizada para a reprodução ou multiplicação da variedade) requerem autorização prévia do obtentor para produção ou reprodução; preparação para reprodução ou multiplicação; oferta à venda; venda ou qualquer forma de comercialização; exportação; importação; e armazenamento. Se um obtentor não tem possibilidade de exercer seus direitos sobre espécies anuais cultivadas, o material de multiplicação ou reprodução pode ser multiplicado ou propagado sem a autorização (WIPO, 2004). Nesse caso, o obtentor pode exercer seus direitos sobre a colheita.

Desafios da Propriedade Intelectual no Brasil

A propriedade intelectual no Brasil ainda apresenta pouca relevância e seu entendimento ainda é restrito (WIPO, 2019). A pouca utilização dos documentos e informações tecnológicas das patentes e outros ativos reforça a falta de conhecimento sobre a função da PI na inovação e desenvolvimento de novas tecnologias. No Brasil, esse desafio inclui a contraditória defasagem entre a alta produção científica nacional e o baixo número de patentes. Isso se deve a um reflexo da desconexão das

empresas, universidades e centros de pesquisa, e o desconhecimento do potencial da PI para incrementar o desenvolvimento econômico. Em consonância, existem gastos desnecessários com a compra de tecnologias externas que estão em domínio público, demonstrando a falta de conhecimento na busca de ativos de propriedade intelectual que possuem registro ou dos quais já se tem o pleno conhecimento industrial. Desperdiça-se parte do potencial inventivo e, por consequência, do potencial industrial inovador das entidades nacionais que lidam com pesquisa, tecnologia e inovação. Ainda existem vários pontos que impedem o país de atingir um ambiente favorável para a proteção de invenções, particularmente na área de biotecnologia (Biominas Brasil, 2018).

Um outro desafio da PI em programas de melhoramento é o tempo necessário para que uma patente seja analisada e concedida no país. Esse fato é basicamente atribuído ao aumento do número e complexidade das patentes depositadas e à limitada quantidade de examinadores. A falta de uma cultura de proteção das inovações reduz o número de patentes depositadas no exterior e dentro do país pelas instituições brasileiras. A ineficiência na produção de patentes também está associada a pesquisas científicas com poucas aplicações tecnológicas e originalidade e, como resultado, o Brasil aparece apenas na 66ª posição em relação à eficiência de produção de novas tecnologias mundialmente (Quintino et al., 2021). Os projetos de P&D financiados pelo setor público geram apenas 30% de publicações técnico-científicas em colaboração com grupos internacionais e 1% em colaboração com a indústria (Cross et al., 2017). A baixa performance das instituições brasileiras no ranking de patentes é agravada pela ausência de pesquisadores no corpo técnico das empresas. Em países desenvolvidos, as publicações em colaboração com a iniciativa privada podem corresponder até 80% do total. Além de haver poucos doutores, os investimentos em P&D das empresas são muito baixos (Biominas Brasil, 2018). Por fim, as instituições adotam a estratégia de explorar apenas o mercado brasileiro, ou não apresentam conhecimento e recurso para depositar a patente no exterior.

As opiniões sobre a interação entre instituições públicas e privadas têm sido contraditórias (Paiva; Melo, 2008). Um desafio importante para geração de PI envolve criar regras claras sobre a interação dos professores e laboratórios de instituições públicas com a iniciativa privada (Thamer; Lazzarini, 2015). Também há vários desafios a serem superados sobre as regras para professores que decidem criar novas tecnologias e abrir sua própria empresa em áreas próximas de sua formação acadêmica. Entretanto, não é objetivo deste capítulo resolver as contradições sobre o funcionamento das instituições públicas. Uma discussão mais aprofundada envolvendo representantes da sociedade civil pode ajudar a esclarecer essas questões e identificar

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

oportunidades para aprimorar a interação entre instituições públicas e privadas, incluindo como poderá ser viabilizado o aporte financeiro da iniciativa privada nas universidades, a liberação institucional de professores para projetos tecnológicos, o papel do financiamento público ou privado no fomento a essas iniciativas, o retorno esperado para a sociedade e o potencial dos alunos em se tornarem fundadores de startups a partir de projetos acadêmicos.

Os professores em regime de dedicação exclusiva trabalhando em projetos institucionais de pesquisa e inovação geralmente podem receber recursos financeiros de empresas privadas, desde que geridos por fundação de apoio ou receber diretamente de empresa privada desde que autorizado pela sua instituição. O recurso financeiro fornecido ao professor incentiva a produção e transferência de propriedade intelectual. Mas o país ainda precisa considerar a criação de novos planos de carreira para indivíduos contratados para atuar no ensino, pesquisa ou extensão conforme a necessidade da sociedade na região de sua unidade acadêmica. Ao contrário de países desenvolvidos, a educação superior brasileira não possui professores com dedicação exclusiva para pesquisa ou extensão em universidades federais e institutos federais de educação, ciência e tecnologia.

Gestão da Inovação e Transferência de Tecnologias

A propriedade intelectual deve ser vista como um instrumento de política econômica. As patentes, por exemplo, destinam-se a recompensar o inventor, gerando um incentivo para a inovação e contribuindo para o investimento em mais projetos de P&D. A proteção das marcas baseia-se na ideia de distinguir os produtos e os serviços de uma empresa dos produtos e serviços de outras empresas. As marcas podem proteger o público de fraude e confusão, além do que o valor de uma marca está intimamente ligado à qualidade de seus produtos e afeta as decisões de compra pelos clientes (Qiao et al., 2022). A gestão da inovação encontra-se bastante relacionada à gestão da propriedade intelectual. A PI envolve processos tais como o depósito de pedidos e o acionamento, manutenção e aplicação dos direitos. O objetivo da gestão da inovação é a proteção dos resultados de projetos de P&D inovadores. Importante notar que nem todas as ideias comercialmente viáveis podem ser ou serão patenteadas. Portanto, algumas ideias são mantidas em segredo, especialmente na fase inicial, a fim de prevenir o enfraquecimento de uma vantagem competitiva.

A gestão da inovação é importante para identificar os aspectos técnicos e econômicos da PI e determinar as medidas necessárias para assegurar a proteção. As empresas podem utilizar a PI como um ativo essencial em modelos de negócios

baseados em conhecimentos e, para tirar benefícios da PI, as empresas devem associar esse desenvolvimento a regimes de apropriação e utilização da PI, a qual é uma fonte de diferenciação sustentável e de vantagem competitiva para as empresas e indivíduos que inovam.

A gestão da inovação concilia a oportunidade da inovação com o contexto comercial e aloca recursos para a obtenção de retorno lucrativo e sobrevivência das organizações. A gestão da inovação atua no campo da economia, marketing, operações, finanças, contabilidade, tecnologia, forma organizacional, distribuição e no de todas as suas combinações possíveis (Tidd; Bessant, 2015). Com isso, a gestão da inovação não se apoia somente na apropriação e proteção dos conhecimentos criados pela empresa (Candelin-Palmqvist et al., 2012). Ela combina essa apropriação e proteção com as melhores oportunidades do mercado consumidor. Os principais fatores a serem considerados envolvem o tamanho do mercado, disponibilidade de capital e outros conhecimentos tácitos ou codificados, tecnologias específicas, regimes de apropriabilidade, ciclo de vida da inovação e o contexto jurídico do país.

A comercialização requer o licenciamento da inovação, e a proteção de PI é apoiada pela criação de sistemas de proteção múltipla (Figura 1). A transferência de tecnologia (TT) pode ocorrer por vias formais, mediante a compra de patentes ou licença, operações de fusão e aquisição de empresas ou outras formas de cooperação entre empresas. Independentemente da forma de transferência, a empresa precisa realizar investimentos tangíveis e intangíveis, que permitam adquirir novas máquinas e equipamentos e aperfeiçoar métodos de gestão (Etzkowitz; Zhou, 2021). Exemplos de transferência de tecnologia incluem a utilização de marcas para aumentar as vantagens comerciais da proteção de patentes e a transferência da notoriedade existente da marca de um produto para novos produtos associados (WIPO, 2023).

Considerações Finais

Os programas de melhoramento de plantas devem estar preparados para enfrentar os desafios relacionados à inovação, propriedade intelectual e empreendedorismo. A modernização e o avanço desses programas dependem da utilização de tecnologias digitais e da formação profissional adequada. Com a crescente competitividade e complexidade do setor, é crucial que empresas e organizações invistam em recursos e estratégias para potencializar seus investimentos e adotar tecnologias emergentes, tais como internet das coisas, *blockchain*, seleção genômica, automação, análise de *big data*, inteligência artificial, robótica e plataformas digitais. Essas abordagens oferecem

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

soluções inovadoras que permitem maior eficiência operacional, tomada de decisão baseada em análise de dados e redução de custos com mão de obra. Além disso, o empreendedorismo tem importância significativa no desenvolvimento profissional, uma vez que proporciona aos indivíduos maior controle sobre suas carreiras e a oportunidade de se envolver em projetos de seu interesse. A proteção da propriedade intelectual, por meio de patentes, segredos comerciais e proteção de cultivares, é essencial para incentivar a inovação e garantir os direitos dos inventores e melhoristas de plantas. Essas medidas contribuem para a modernização dos programas de melhoria de plantas, impulsionando a inovação, a eficiência operacional e a competitividade das empresas e organizações nesse campo.

Referências

- AGROINTELI. Disponível em: <https://www.agrointeli.com.br/>. Acesso em: 17 abr. 2023.
- AGROTOOLS. Disponível em: <https://agrottools.com.br/agtech-para-agronegocio/>. Acesso em: 20 jul. 2023.
- ALÂNGE, S.; STEIBER, A.; CORVELLO, V. Evaluating corporate-startup collaboration: a government Perspective. **Evaluation and Program Planning**, v. 95, 102176, Dec. 2022. DOI: <https://doi.org/10.1016/j.evalprogplan.2022.102176>.
- BAETA, A. M. C.; LEITE, J. B. D. Capital de risco e desenvolvimento tecnológico. **Revista de Administração de Empresas**, v. 30, n. 2, p. 73-78, abr./jun. 1990. DOI: <https://doi.org/10.1590/S0034-75901990000200007>.
- BAGNATO, V. S.; ORTEGA, L. M.; SOUZA, M. A.; MURAKAWA, L. S. G. **Guia Prático I: introdução à propriedade intelectual**. 2016. Disponível em: http://www.inovacao.usp.br/wp-content/uploads/sites/300/2017/10/CARTILHA_PI_bom.pdf. Acesso em: 29 abr. 2023.
- BARBOSA FILHO, F. H.; PESSÔA, S. **Retorno da educação no Brasil**. 2008. Disponível em: https://repositorio.ipea.gov.br/bitstream/11058/3357/4/PPE_v38_n01_Retorno.pdf. Acesso em: 21 jul. 2023.
- BARRETT, W. A. Building a strategy for maximizing intellectual property value. **Bioentrepreneur**, Jan. 2005. DOI: <https://doi.org/10.1038/bioent842>.
- BARTH, T. G.; ENSSLIN, S. R.; BORGERT, A. Os benefícios pessoais da pós-graduação stricto sensu: uma análise na percepção de mestres em contabilidade. **Revista de Educação e Pesquisa em Contabilidade (REPeC)**, v. 10, n. 1, p. 106-128, jan./mar. 2016. DOI: <https://doi.org/10.17524/repec.v10i1.1326>.
- BIOMINAS BRASIL. **Desafios da propriedade industrial no Brasil**. 2018. Disponível em: <https://biominas.org.br/blog/desafios-da-propriedade-intelectual-no-brasil-2/>. Acesso em: 03 abr. 2023.
- BIOTURING. Disponível em: <https://bioturing.com/>. Acesso em: 21 jul. 2023.
- BLANK, S.; DORF, B. **The startup owner's manual: the step-by-step guide for building a great company**. California: K&S Ranch, 2012. 571 p.
- BOFF, S. O. Proteção jurídica das cultivares na UPOV, no TRIPs e os limites na legislação brasileira: o embate entre interesse público e privado. **Anuario Mexicano de Derecho Internacional**, v. 19, p. 433-456, enero/dic. 2019. DOI: <https://doi.org/10.22201/ij.24487872e.2019.19.13330>.

BRASIL. Lei nº 9.279, de 14 de maio de 1996. Regula direitos e obrigações relativos à propriedade industrial. **Diário Oficial [da] República Federativa do Brasil**, Brasília, DF, 15 maio 1996. Seção 1, p. 8353-8366. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/19279.htm. Acesso em: 8 fev. 2023.

BRASIL. Lei nº 9.456, de 25 de abril de 1997. Institui a Lei de Proteção de Cultivares e dá outras providências. **Diário Oficial [da] República Federativa do Brasil**, Brasília, DF, 28 abr. 1997. Seção 1, p. 8241-8246. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/19456.htm. Acesso em: 8 fev. 2023.

BRASIL. Ministério da Ciência, Tecnologia e Inovações. **Parques tecnológicos e de inovação no Brasil**. 2021. Disponível em: <https://anprotec.org.br/site/wp-content/uploads/2022/01/ParquesTecnologicosBrasil-2021-Final-vr.pdf>. Acesso em: 29 abr. 2023.

BRASIL. Ministério da Ciência, Tecnologia, Inovações e Comunicações. **Estratégia nacional de ciência, tecnologia e inovação 2016-2022**. 2016. Disponível em: http://www.finep.gov.br/images/a-finep/Politica/16_03_2018_Estrategia_Nacional_de_Ciencia_Tecnologia_e_Inovacao_2016_2022.pdf. Acesso em: 29 abr. 2023.

CANDELIN-PALMQVIST, H.; SANDBERG, B.; MYLLY, U.-M. Intellectual property rights in innovation management research: a review. **Technovation**, v. 32, n. 9-10, p. 502-512, Sept./Oct. 2012. DOI: <https://doi.org/10.1016/j.technovation.2012.01.005>.

CASTRO, E. M. M. **O acordo TRIPS e a saúde pública**: implicações e perspectivas. Brasília, DF: Fundação Alexandre de Gusmão, 2018. 533 p.

CHESBROUGH, H. W.; VANHAVERBEKE, W.; WEST, J. **Open innovation**: researching a new paradigm. Oxford: Oxford University Press, 2006. 373 p.

COHN, M. **Succeeding with agile**: software development using Scrum. Boston: Addison-Wesley, 2017. 512 p.

COSTA, C. M. M.; CHAGAS, H. M. A.; MATSUKURA, T. S.; VIEIRA, G. I.; MARQUEZE, E. C.; LÓPEZ, C. G.; GHELARDI, I. R.; LEFÈVRE, A.; LEFÈVRE, F. Contribuições da pós-graduação na área da saúde para a formação profissional: relato de experiência. **Saúde e Sociedade**, v. 23, n. 4, p. 1471-1481, out./dez. 2014. DOI: <https://doi.org/10.1590/S0104-12902014000400028>.

CROPLIFE. **Proteção de plantas e mudas (regulamentação de germoplasma)**. 2019. Disponível em: <https://croplifebrasil.org/sementes-e-mudas/protacao-de-plantas-e-mudas/>. Acesso em: 8 fev. 2023.

CROSS, D.; THOMSON, S.; SINCLAIR, A. **Research in Brazil**: a report for CAPES by Clarivate Analytics. 2017. Disponível em: <https://www.gov.br/capes/pt-br/centrais-de-conteudo/17012018-capes-incitesreport-final-pdf>. Acesso em: 8 fev. 2023.

DEVAUX, A.; GOFFART, J.-P.; KROMANN, P.; ANDRADE-PIEDRA, J.; POLAR, V.; HAREAU, G. The potato of the future: opportunities and challenges in sustainable agri-food systems. **Potato Research**, v. 64, p. 681-720, Dec. 2021. DOI: <https://doi.org/10.1007/s11540-021-09501-4>.

DINIZ, D. M.; CRUZ, M. A.; CORREA, V. S. Fatores críticos da transferência de conhecimento entre universidade e empresa (U-E). **Revista Eletrônica de Administração**, v. 24, n. 2, p. 230-252, maio/ago. 2018. DOI: <http://dx.doi.org/10.1590/1413-2311.210.83919>.

EMBRAPA. Secretaria de Pesquisa e Desenvolvimento. Sistema Embrapa de Gestão. **Manual sobre o uso da escala TRL/MRL**. Disponível em: <https://cloud.cnpgc.embrapa.br/nap/files/2018/08/EscalaTRL-MRL-17Abr2018.pdf>. Acesso em: 3 abr. 2023.

ESPINA-ROMERO, L.; FRANCO, S. L. A.; CONDE, H. O. D.; GUERRERO-ALCEDO, J. M.; PARRA, D. E. R.; RAMÍREZ, J. C. R. Soft skills in personnel training: report of publications in Scopus, topics explored and future research agenda. **Heliyon**, v. 9, e15468, Apr. 2023. DOI: <https://doi.org/10.1016/j.heliyon.2023.e15468>.

ETZKOWITZ, H.; ZHOU, C. Licensing life: the evolution of Stanford university's technology transfer practice. **Technological Forecasting and Social Change**, v. 168, 120764, July 2021. DOI: <https://doi.org/10.1016/j.techfore.2021.120764>.

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

FERNANDES, S. A. S. **Inovação no Setor Público**: uma análise da inovação no âmbito da Administração Pública Federal sob o enfoque da co-criação. 2019. 132 p. Dissertação (Mestrado em Administração Pública). Universidade Federal de Lavras. Lavras.

FONSECA, A. Concorrência e propriedade intelectual. **Revista da ABPI**, n. 36, p. 3-24, 1998.

HULL, J. **Proteção de segredos comerciais**: como as empresas podem responder ao desafio de tomar “medidas razoáveis”. 2019. Disponível em: https://www.wipo.int/wipo_magazine/pt/2019/05/article_0006.html. Acesso em: 8 fev. 2023.

INTERNATIONAL MONETARY FUND. **World economic outlook**: countering the cost-of-living crisis. 2022. Disponível em: <https://www.imf.org/en/Publications/WEO/Issues/2022/10/11/world-economic-outlook-october-2022>. Acesso em: 7 fev. 2023.

LEAL, C. I. S.; FIGUEIREDO, P. N. Inovação tecnológica no Brasil: desafios e insumos para política públicas. **Revista de Administração Pública**, v. 55, n. 3, p. 512-537, maio/jun. 2021. DOI: <http://dx.doi.org/10.1590/0034-761220200583>.

MCNALLY, R. C.; AKDENIZ, M. B.; CALANTONE, R. J. New product development processes and new product profitability: exploring the mediating role of speed to market and product quality. **Journal of Product Innovation Management**, v. 28, n. S1, p. 63-77, Nov. 2011. DOI: <https://doi.org/10.1111/j.1540-5885.2011.00861.x>.

MIRANDA, F. J.; CHAMORRO, A.; RUBIO, S. Re-thinking university spin-off: a critical literature review and a research agenda. **The Journal of Technology Transfer**, v. 43 n. 4, p. 1007-1038, Aug. 2018. DOI: <https://doi.org/10.1007/s10961-017-9647-z>.

OLIVEIRA, S. F.; PRADO, R. B.; MONTEIRO, J. M. G. Impactos das mudanças climáticas na produção agrícola e medidas de adaptação sob a percepção de atores e produtores rurais de Nova Friburgo, RJ. **Interações**, v. 23, n. 4, p. 1179-1201, out./dez. 2023. DOI: <https://doi.org/10.20435/inter.v23i4.3548>.

PAIVA, K. C. M.; MELO, M. C. O. L. Competências, gestão de competências e profissões: perspectivas de pesquisas. **Revista de Administração Contemporânea**, v. 12, n. 2, p. 339-368, jun. 2008. DOI: <https://doi.org/10.1590/S1415-65552008000200004>.

PIAZENTINI, M. S.; PETERS, M. P. L.; RAMOS, F. M. Incentivos fiscais para inovação no agronegócio. In: MENDONÇA, M. de S. (org.). **Agronegócio & sustentabilidade**: métodos, técnicas, inovação e gestão. Guarujá: Científica Digital, 2021. p. 186-191. DOI: <https://doi.org/10.37885/211106690>.

PIX4D BRASIL. Disponível em: <https://www.pix4d.com/> Acesso em: 17 abr. 2023.

POPKOVA, E.; BOGOVIZ, A. V.; LOBOVA, S. V.; CHILILOV, A. M.; SOZINOVA, A. A.; SERGI, B. S. Changing entrepreneurial attitudes for mitigating the global pandemic's social drama. **Humanities and Social Sciences Communications**, v. 9, 141, Apr. 2022. DOI: <https://doi.org/10.1057/s41599-022-01151-2>.

QIAO, Y.; YIN, X.; XING, G. Impact of perceived product value on customer-based brand equity: Marx's Theory value-based perspective. **Frontiers in Psychology**, v. 13, 931064, June 2022. DOI: <https://doi.org/10.3389/fpsyg.2022.931064>.

QUINTINO, H. M. S.; HOLANDA, F. S. R.; MOURA, F. R.; SANTANA, J. R.; VIDAL, L. D. World efficiency in the potential production of new technologies under intellectual property assets. **Technology in Society**, v. 65, 101601, May 2021. DOI: <https://doi.org/10.1016/j.techsoc.2021.101601>.

RAPID GENOMICS. Disponível em: <https://rapid-genomics.com/about-rapid-genomics/>. Acesso em: 17 abr. 2023.

RIES, E. **The lean startup**: how today's entrepreneurs use continuous innovation to create radically successful businesses. New York: Crown Business, 2011. 295 p.

ROBERTS, E. B. Managing invention and innovation. **Research-Technology Management**, v. 50, n. 1, p. 35-54, 2007. DOI: <https://doi.org/10.1080/08956308.2007.11657418>.

- SAMPAIO, M. J. A. Propriedade intelectual de plantas: a nova lei de proteção de cultivares e suas decorrências imediatas. In: BORÉM, A; DEL GIÚDICE M. P; SAKIYAMA, N. S; SEDIYAMA, T.; MOREIRA, M. A; PORTUGAL, R. S. (ed.). **Biossegurança, proteção de cultivares, acesso a recursos genéticos e propriedade industrial na agropecuária**. Viçosa, MG: Universidade Federal de Viçosa, 1998. p.145- 158.
- SANTOS, D. P. Inovação no setor público: tendências e casos bem-sucedidos de inovações no Brasil e no mundo. **Revista Científica Multidisciplinar Núcleo do Conhecimento**, v. 7, n. 1, p. 92-103, set. 2022.
- SANTOS, E. S.; SOUZA, F. T.; FERNANDES, A. C. A. Tecnologia da informação como vantagem competitiva para as empresas. **Revista Científica Multidisciplinar Núcleo do Conhecimento**, v. 1, n. 2, p. 93-112, 2016.
- SILVA, E. S. S.; SILVA-MANN, R.; CALAZANS, C. C. Royalties para cultivares, legislação e regulação: uma meta-análise. **Research, Society and Development**, v. 10, e31710414231, 2021. DOI: <https://doi.org/10.33448/rsd-v10i4.14231>.
- SMALDONE, F.; IPPOLITO, A.; LAGGER, J.; PELLICANO, M. Employability skills: profiling data scientists in the digital labour market. **European Management Journal**, v. 40, n. 5, p. 671-684, Oct. 2022. DOI: <https://doi.org/10.1016/j.emj.2022.05.005>.
- SMART AGRI. Disponível em: <https://smartagri.com.br/>. Acesso em: 17 abr. 2023.
- SOLHEIM, M. C. W. Diversity and inclusion is a must to make innovation work for all. **Nature**, v. 612, n. 7939, p. S11, Dec. 2022. DOI: <https://doi.org/10.1038/d41586-022-04205-2>.
- TERRAMAGNA. Disponível em: <https://terramagna.com.br/>. Acesso em: 17 abr. 2023.
- THAMER, R.; LAZZARINI, S. G. Projetos de parceria público-privada: fatores que influenciam o avanço dessas iniciativas. **Revista de Administração Pública**, v. 49, n. 4, p. 819-846, jul./ago. 2015. DOI: <https://doi.org/10.1590/0034-7612119746>.
- TIDD, J.; BESSANT, J. **Gestão da inovação**. 5. ed. Porto Alegre: Bookman, 2015. 648 p.
- TIRMIZI, S. M. A.; MALIK, Q. A.; HUSSAIN, S. S. Invention and open innovation processes, and linkages: a conceptual framework. **Journal of Open Innovation: Technology, Market, and Complexity**, v. 6, n. 4, p. 159, Dec. 2020. DOI: <https://doi.org/10.3390/joitmc6040159>.
- UNITED NATIONS. **UN population division data portal**: interactive access to global demographic indicators. Disponível em: <https://population.un.org/dataportal/home>. Acesso em: 7 fev. 2023.
- VAN DE VRANDE, V.; VANHAVERBEKE, W.; DUYSTERS, G. Additivity and complementarity in external technology sourcing: the added value of corporate venture capital investments. **IEEE Transactions on Engineering Management**, v. 58, n. 3, p. 483-496, Aug. 2011. DOI: <https://doi.org/10.1109/tem.2010.2091134>.
- WIPO. World Intellectual Property Organization. **Guide to the international patent classification**. 2019. Disponível em: https://www.wipo.int/edocs/pubdocs/en/wipo_guide_ipc_2019.pdf. Acesso em: 8 fev. 2023.
- WIPO. World Intellectual Property Organization. **Seed ordinance n° 15/2004/PL-UBTVQH11**. 2004. Disponível em: <https://www.wipo.int/wipolex/ru/text/450626>. Acesso em: 8 fev. 2023.
- WIPO. World Intellectual Property Organization. **Transferring technology from lab to market**. 2022. Disponível em: <https://www.wipo.int/technology-transfer/en/access-market.html>. Acesso em: 9 fev. 2023.
- WIPO. World Intellectual Property Organization. **What is a patent?** Disponível em: <https://www.wipo.int/patents/en/>. Acesso em: 8 fev. 2023.
- WOLFF, M. T. Diretiva da comissão europeia sobre a proteção legal de invenções biotecnológicas. **Revista da ABPI**, n. 37, p. 13-17, 1998.

Marco Antônio Marcelino Bahia, João Gabriel Zanon Paludeto,
Gustavo Eduardo Marcatti, Rafael Tassinari Resende

Introdução

Para aperfeiçoar o processo de melhoramento genético de plantas, além de primar por estratégias bem encaixadas à natureza genética da cultura, focando nos caracteres fenotípicos a serem avaliados, é cada vez mais necessário sofisticar os programas de melhoramento, lançando mão de ferramentas de manejo de dados. Tais ferramentas contemplam a computação, programação, mineração de dados e a aptidão para lidar com grandes volumes de dados (os chamados *Big Data*). Quando utilizadas em conjunto, essas ferramentas possibilitam a integração e análise de informações em larga escala, o que pode derivar em resultados mais precisos e eficientes para o melhorista de plantas (Bernardo, 2020). Aqui um dos termos chave é a “automatização de processos”, em que por meio da programação computacional, usando linguagens como o R e o Python, é possível implementar soluções para os desafios da rotina do melhorista, como a criação ou utilização de funções ou programas específicos para lidar com planejamentos experimentais e com a análise dos múltiplos dados genéticos (Crossa et al., 2021).

Uma rotina básica de aplicação da computação no melhoramento genético de plantas é a modelagem fenotípica, em que se pode usar métodos estatísticos e/ou algoritmos de inteligência artificial programados para prever as características de plantas resultantes de determinado cruzamento alógamo ou autógamo (Niazian; Niedbała, 2020). Outra ferramenta de grande potencial é a seleção genômica, a qual permite a seleção com base no DNA das plantas para as características desejadas, antes mesmo de serem colocadas em experimentos de campo (Merrick et al., 2022). Ou mesmo a ambientômica, que permite a predição genotípica com base em múltiplas informações ambientais (Resende et al., 2022). Para aplicar

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

todos esses métodos, é preciso lidar com grandes quantidades de dados, o que torna a mineração de dados um processo necessário para a descoberta de informações úteis e padrões ocultos. A mineração de dados consiste no processo de descoberta de informações úteis e padrões em grandes conjuntos de dados, enquanto o manejo de dados refere-se à gestão, processamento e armazenamento de dados para torná-los acessíveis e utilizáveis em diferentes contextos.

O armazenamento e gerenciamento de dados é uma área importante de aplicação no melhoramento genético de plantas. As ferramentas de *Big Data* permitem o armazenamento de enormes volumes de informações, sendo utilizadas em diversas áreas, como as ciências agrárias (Kamilaris et al., 2017). Isso possibilita análise a longo prazo dos resultados, permitindo a identificação de tendências e aprimoramento contínuo do processo de melhoramento genético. Entre outras aplicações, é possível citar a predição de características fenotípicas em híbridos, realização de sequenciamentos genéticos via bioinformática, criação de plataformas de genotipagem com SNPs (*Single-Nucleotide Polymorphisms*), armazenamento amplo de informações climáticas e identificação de efeitos de marcadores moleculares para características agrônômicas importantes. A compilação de todas essas tecnologias ao melhoramento genético vegetal permite melhor compreensão das características genéticas e ambientais que afetam o desenvolvimento e a produtividade das plantas. Isso ajuda a aumentar a eficiência do processo de seleção e a obtenção de cultivares mais produtivas e adaptadas. Além disso, o domínio de técnicas de manejo de dados permite a análise de grandes quantidades de dados genéticos e ambientais, tornando possível identificar padrões e tendências que poderiam ser facilmente perdidos em análises manuais.

As técnicas e tecnologias que permeiam a boa gestão de dados e *Big Data* são capazes de auxiliar em todas as etapas do melhoramento genético de plantas, tanto em alógamas quanto autógamas. Essas tecnologias permitem a integração e análise de informações em larga escala, auxiliando na tomada de decisão do melhorista e, conseqüentemente, aumentando o sucesso na seleção e produção de cultivares de elevada performance. Neste capítulo serão abordadas ferramentas e técnicas para gerenciar e utilizar dados com segurança em projetos de melhoramento. Serão apresentadas as boas práticas no manejo de dados, algumas linguagens de programação, softwares, ambientes de desenvolvimento integrado (IDE) e plataformas de gerenciamento de código. Também será discutida a segurança e os direitos de uso dos dados, incluindo repositórios de dados de melhoramento.

Boas Práticas no Uso e Manejo de Dados (*Data Wrangling*)

Antes de aprofundarmos nas inúmeras possibilidades de análises de grandes conjuntos de dados no melhoramento genético, é preciso apresentar as boas práticas nesse contexto. A qualidade dos dados é fundamental para a análise e interpretação dos dados em todos os campos. Porém, a falta de padrões de qualidade de dados pode levar a resultados imprecisos, incorretos, distorções e interpretações equivocadas. É importante garantir a precisão, consistência e integridade dos dados, além de documentar claramente as fontes, métodos e limitações de cada dado. Por exemplo, no melhoramento genético de plantas, é importante seguir padrões de qualidade na obtenção e tratamento de dados para garantir a precisão das avaliações fenotípicas e genotípicas dos materiais genéticos. Isso inclui a padronização de métodos de avaliação, a documentação clara das condições de teste e a utilização de controles internos para minimizar a variabilidade não intencional.

Ademais, é importante seguir boas práticas na análise de dados, como a validação de hipóteses, a utilização de métodos estatísticos apropriados e a verificação da robustez dos resultados. A utilização de ambientes computacionais especializados, como o R ou o Python, e seus pacotes apropriados, são meios para garantir a qualidade dos dados e a realização de análises precisas. Seguir padrões de qualidade e boas práticas na análise de dados é vital para garantir a precisão e validade dos resultados no melhoramento genético de plantas. Isso permite tomadas de decisões baseada em evidências sólidas e a melhoria contínua da eficiência e eficácia dos programas de melhoramento genético. Nesse contexto, *Data wrangling* refere-se ao processo de coleta, limpeza, transformação e organização de dados brutos em um formato adequado para análise. Também é conhecido como pré-processamento de dados.

A utilização de ferramentas analíticas avançadas, como Inteligência Artificial, aprendizado de máquina (*Machine Learning*), *Deep Learning* e Redes Bayesianas, entre outras, é insuficiente quando aplicadas a conjuntos de dados de má qualidade ou com estruturação inadequada (isso reflete no *meme* mostrado na Figura 1). Essas abordagens tecnológicas dependem de conjuntos de dados confiáveis, precisos e bem estruturados para gerar resultados significativos. Conseqüentemente, é importante investir em procedimentos sólidos de coleta, limpeza e organização dos dados antes de aplicar técnicas mais avançadas (isto é, realizar *data wrangling*), a fim de obter resultados relevantes aproveitando ao máximo as ferramentas de análise existentes.

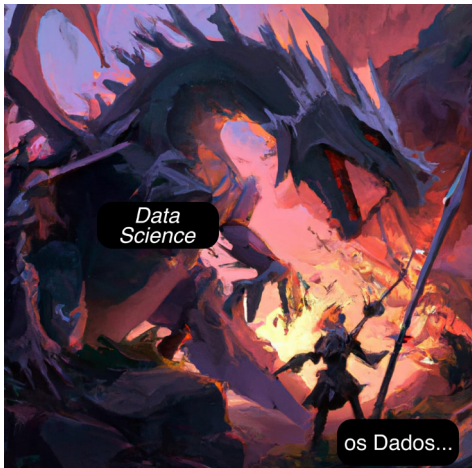


Figura 1. A incongruência da Ciência de Dados (*Data Science*): Análises rebuscadas nem sempre são suficientes quando confrontadas com dados de baixa qualidade. A imagem brinca com a contradição entre o alto poder das ferramentas de Ciência de Dados e a importância crítica de dados confiáveis.

Fonte: Adaptado de LinkedIn (2023).

A organização de dados é fundamental em qualquer análise, no melhoramento genético de plantas não poderia ser diferente. Conjuntos de dados bem-organizados permitem uma fácil e eficiente manipulação, além de minimizar erros e facilitar a identificação de padrões e tendências. Ao realizar *data wrangling*, é importante seguir algumas dicas para garantir a qualidade e organização dos dados: utilize nomes claros e representativos para as colunas e linhas. Evite nomear as colunas dos dados com nomes extensos, se possível, uma única palavra é o suficiente. Para demais informações relevantes utilize uma planilha ou um arquivo extra de *metadados*. Inclua nesse arquivo informações detalhadas sobre as unidades de medida. Tenha linhas ou colunas para descrição das variáveis observadas e evite dados redundantes ou vazios, utilizando uma codificação clara para indicar valores ausentes, como “NA” (*Not Available*) ou “NaN” (*Not a Number*) no R ou “None” no Python. Mantenha um padrão de formatação consistente, utilize ferramentas de controle de versão e verifique a consistência dos dados. Realize transformações e agregações quando necessário. Documente todas as etapas realizadas e automatize tarefas quando possível. Seguir essas dicas auxilia na estruturação e documentação adequada dos dados, tornando-os prontos para análise.

A adoção de boas práticas de organização de dados torna as análises mais reproduzíveis, confiáveis e precisas, sendo particularmente importante no melhoramento genético de plantas, em que as decisões tomadas devem ser imprescindivelmente amparadas em dados íntegros e plenamente confiáveis, que serão base para a produção de cultivares de alta qualidade e com um alto valor agregado. Também se deve ressaltar a importância da realização de análises prévias dos dados para verificar se estão organizados de forma adequada e se

possuem informações faltantes ou *outliers* (isto é, valores extremos em um conjunto de dados que se afastam significativamente da maioria dos outros valores e podem distorcer a análise estatística). Uma vez que os dados tenham sido pré-analisados, é possível detectar erros ou observações estranhas utilizando algoritmos estatísticos. Pode-se destacar alguns exemplos, como a verificação de *outliers*, a detecção de observações fora dos limites, a verificação de redundâncias, entre outros. A utilização de ferramentas de mineração de dados, como aprendizado de máquina e análise de cluster, também podem ser úteis na detecção de incoerências no conjunto a ser analisado. O treinamento de algoritmos que detectam padrões ou *outliers* em conjuntos de *Big Data*, permitindo uma análise precisa e mais coerente, já são rotina em programas de melhoramento.

Algumas técnicas de pré-análise, como podem ser vistas na Figura 2, incluem a análise visual gráfica, pela verificação de distribuições, diagramas de caixas (*boxplot*), histogramas e gráficos de dispersão. Nessa figura, simulou-se dois conjuntos de dados (X e Y), cada um deles contendo 10 mil observações. Supondo que os dados X e Y sejam caracteres mensurados ao longo do tempo (t_1 e t_2 , respectivamente, após um tempo zero), pode-se supor que o valor fenotípico “0” sejam plantas mortas, também é possível observar um padrão de incremento entre ambos (Figura 2A). Já os *boxplots* são representações visuais de conjuntos de dados que mostram as distribuições dos valores em relação à mediana, quartis e *outliers* (observados na Figura 2B). A caixa representa o intervalo entre o primeiro e o terceiro quartil, a linha dentro da caixa representa a mediana e os “bigodes” mostram a extensão dos valores mínimo e máximo que não são *outliers*. Algumas observações foram identificadas como *outliers*, no entanto, análises mais profundas devem ser feitas no intuito de eliminá-las ou não.

Pela Figura 2C observam-se anomalias nos dados, como por exemplo, se esses valores são de crescimento das plantas e uma planta apresentou o valor zero no t_1 , é razoável desconfiar que ela não poderia ter ressuscitado em t_2 , a menos que tenha habilidade de rebrota ou qualquer outra habilidade regenerativa. Ou ainda, deve-se identificar plantas que eventualmente encolheram ao longo do tempo, e avaliar se isso faz sentido fenotipicamente. Essas, entre outras práticas, muitas vezes são mais eficazes no sentido de melhoria de qualidade nos dados em comparação à uma direta exclusão de *outliers*. Outro ponto destacado na Figura 2C é uma observação isolada que corresponde a 54,3 unidades da variável X2 e apenas 3,80 unidades da variável X1, note que essa combinação Y em função de X é improvável, dado o padrão de crescimento mostrado pela reta de regressão.

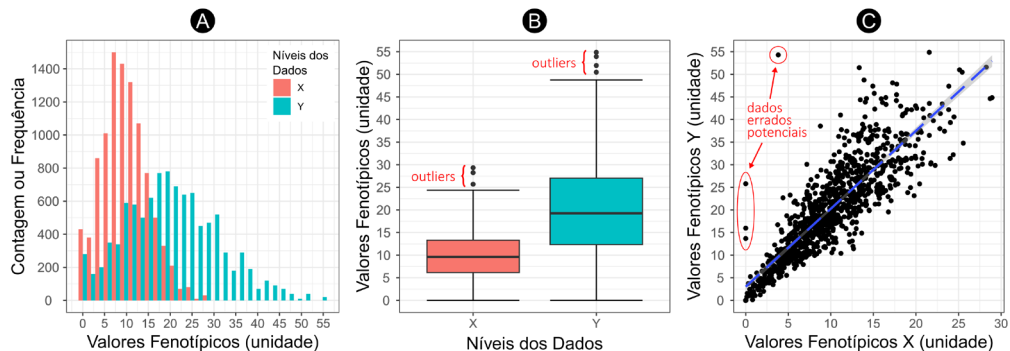


Figura 2. Análise gráfica preliminar dos dados originais mostrando dois caracteres fenotípicos fictícios X e Y. (A) Gráficos de distribuições (histogramas). (B) gráficos *boxplot*, mostrando *outliers*. (C) gráficos de dispersão, mostrando como identificar anomalias nos dados (setas vermelhas). A reta tracejada em azul é o ajuste de regressão linear com Y em função de X.

Dentre as boas práticas no uso e na mineração de dados, há uma metodologia que se destaca há mais de duas décadas, o CRISP-DM (*Cross-Industry Standard Process for Data Mining*). O CRISP-DM foi introduzido em 1996 com o objetivo de propor um método sistemático e generalizado para aplicar conceitos de mineração de dados e análise de operações industriais (Shearer, 2000). Entretanto, é uma metodologia altamente adaptável e de fácil aplicação em outros domínios e setores que lidam com grandes volumes de dados e, por isso, é o método analítico mais utilizado por especialistas em mineração de dados atualmente (Saltz, 2021).

O CRISP-DM é um processo cíclico que se divide em seis fases interativas: entendimento do negócio, entendimento dos dados, preparação dos dados, modelagem, avaliação e implantação (Figura 3).

- 1) **Entendimento do negócio:** A primeira fase consiste em compreender os recursos disponíveis e necessários, bem como as necessidades e objetivos das partes interessadas a fim de obter uma compreensão abrangente do contexto do negócio e definir expectativas realistas.
- 2) **Entendimento dos dados:** Nessa fase, o foco é obter uma visão clara dos dados disponíveis. Isso envolve explorar as fontes de dados, compreender a estrutura, checar a qualidade dos problemas e lacunas, assim como a relevância para o projeto e o próprio entendimento do negócio.

- 3) **Preparação dos dados:** Após o entendimento dos dados, é necessário prepará-los para a análise. Essa fase envolve a limpeza, tratamento de dados perdidos, padronização e transformações necessárias.
- 4) **Modelagem:** Nessa fase, os algoritmos e modelos estatísticos adequados são aplicados aos dados. Isso pode envolver o uso de técnicas de aprendizado de máquina, regressão, árvores de decisão, entre outras.
- 5) **Avaliação:** Uma vez que os modelos tenham sido criados, é importante avaliá-los de forma criteriosa, ou seja, testar os modelos com dados de validação a fim de confirmar sua acurácia e desempenho e refinar e reajustar modelos, se necessário.
- 6) **Implantação:** Na última etapa, os modelos selecionados e validados são implantados para uso prático na organização. Isso pode envolver a integração desses modelos em sistemas já existentes ou o desenvolvimento de interfaces de usuário e criação de relatórios automatizados.

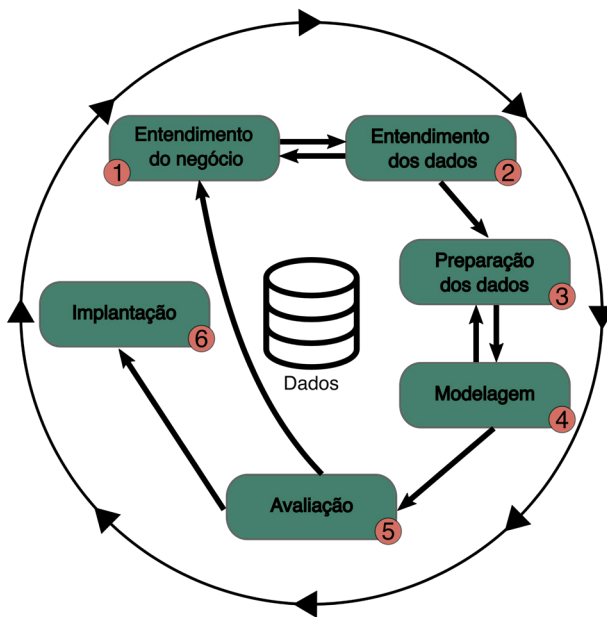


Figura 3. Metodologia CRISP-DM, processo cíclico de mineração e análise de dados.

A metodologia CRISP-DM não termina com a implantação, é um processo cíclico e interativo em que uma etapa pode influenciar não só nas etapas subsequentes, como também em algumas etapas anteriores. Além disso, é fundamental monitorar e avaliar continuamente os resultados e realizar ajustes, se necessário.

Linguagens de Programação Usadas no Melhoramento

Atualmente, a distinção entre “linguagem de programação” e “software” é desafiadora, especialmente quando observamos as linguagens de alto nível utilizadas no campo do melhoramento vegetal. Ao contrário das linguagens de baixo nível, que estão mais próximas da linguagem de máquina, como *Assembly*, *C*, *C++* e *Fortran*, as linguagens de alto nível oferecem uma variedade de funções e pacotes, que permitem a utilização de aplicativos prontos, dispensando a necessidade de desenvolvê-los do zero. No âmbito do melhoramento genético de plantas, as linguagens de programação mais comumente empregadas incluem R, Python, SAS, MATLAB e Julia. A distinção aqui residirá no fato das linguagens de programação serem associadas a ações operacionais que envolvem “linhas de comando”, enquanto que as ações com interface gráfica amigável serão consideradas softwares.

R

O R (The R Foundation, 2022) é uma linguagem de programação aberta muito utilizada na análise estatística e visualização de dados. É bastante utilizada para análise de dados genéticos, como a estimativa de parâmetros de modelos de seleção e a identificação de efeitos gênicos significativos.

A linguagem R é frequentemente utilizada em diversos tipos de análises úteis para o melhoramento genético de plantas, como análise multivariada, modelos de regressão lineares e não lineares diversos, análise genômica, associação genética, além de ser uma ferramenta gráfica poderosa para visualização dos dados e resultados, tal como mostrado na Figura 2, utilizando o pacote ggplot2 (Wickham, 2016). Dentre as vantagens do R, podemos destacar o fato de ser uma linguagem com sintaxe intuitiva e usabilidade facilitada, sendo *open-source*, o que significa gratuidade e uma ampla comunidade de desenvolvedores e usuários, se traduzindo em uma gama de pacotes disponíveis para analisar dados de melhoramento genético. No entanto, ele pode ser mais lento do que outros softwares para realizar alguns tipos de análises, porém o pacote Rcpp, desenvolvido em C++ tem solucionado efetivamente os problemas enfrentados ao executar diversas tarefas no R (Eddelbuettel; Balamuta, 2017). Existem vários pacotes no R que são úteis para o melhoramento genético de plantas. Piaskowski (2022) faz uma extensa revisão sobre vários pacotes atuais para a pesquisa agrícola. Existe ampla gama de pacotes com aplicações no melhoramento de plantas e, a título de exemplo, mencionaremos alguns:

- **lme4**: Pacote que fornece suporte para a realização de ajustes e análises de modelos mistos lineares, modelos lineares generalizados mistos e modelos não lineares. É uma evolução do pacote **nlme**, apresentando melhor desempenho operacional e novas usabilidades, como a presença da função **glmer**, relativa aos modelos lineares generalizados mistos (Bates et al., 2015). O pacote não permite a inclusão de matrizes de variância e covariância, como as matrizes de relacionamento construídas por meio de pedigree ou marcadores moleculares. Link para o CRAN: <https://CRAN.R-project.org/package=lme4>
- **synbreed**: É um compêndio de funções com foco em predição genômica que podem ser implementadas no melhoramento contínuo de plantas e animais. O pacote atende as áreas de processamento, visualização e análise de dados, sendo notável por apresentar uma implementação flexível (Wimmer et al., 2012). Link para o CRAN: <https://synbreed.r-forge.r-project.org/>
- **sommer**: Pacote que possibilita o ajuste de modelos mistos univariados, multivariados, bem como a inclusão de matrizes de relacionamento, a especificação das estruturas de variância e covariância assumidas para cada efeito aleatório e o ajuste de modelos espaciais (Covarrubias-Pazarán, 2016). Link para o CRAN: <https://cran.r-project.org/web/packages/sommer/index.html>
- **breedR**: Pacote que permite o uso de abordagens estatísticas frequentistas e bayesianas na construção de modelos mistos de predição para melhoristas e geneticistas quantitativos. Possibilita a inclusão de matrizes de relacionamento, ajuste de modelos espaciais e modelos que consideram a competição entre genótipos. Utiliza a família de programas BLUPF90 como base, o que permite um significativo ganho em tempo computacional (Muñoz, 2017). Link para o arquivo: <https://github.com/famuvie/breedR/wiki/Overview.pdf>
- **AGHmatrix**: Muito utilizado para a genética quantitativa, esse pacote cria matrizes de relacionamento a partir do pedigree das amostras e com base no nível de ploidia, retornando para o usuário diferentes matrizes de dominância de acordo com os parâmetros apresentados por Amadeu et al. (2016). Link para o CRAN: <https://cran.r-project.org/web/packages/AGHmatrix/index.html>
- **regress**: Pacote que oferece um conjunto de funções de ajuste de modelos lineares Gaussianos por meio da máxima verossimilhança do resíduo. É comumente usado para modelos de efeito aleatório e modelos multivariados. Conjunto útil de ferramentas para os melhoristas atuais. Link para o CRAN: <https://CRAN.R-project.org/package=regress>

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

- **METAN**: Pacote que possibilita testes em ambientes variados, analisando a estabilidade da amostra. Aplicado para análises biométricas, menores caminhos, correlações, clusters e diversas opções para testes de multiambiente (Olivoto; Lúcio, 2020). Link para o CRAN: <https://CRAN.R-project.org/package=metan>
- **rrBLUP**: Pacote que fornece implementações de métodos de ajuste de modelos de REML (restricted maximum likelihood) para dados de genética de populações e seleção genômica. Link para o CRAN: <https://CRAN.R-project.org/package=rrBLUP/>
- **BLR**: Pacote que fornece ferramentas para a realização de análises de modelos lineares mistos por meio de estatística Bayesiana. Link para o CRAN: <https://CRAN.R-project.org/package=BLR>
- **BGLR**: Pacote que fornece ampla coleção de modelos de regressão Bayesiana, incluindo seleção de variáveis paramétricas e semiparamétricas. É focado em análise e predição genômica, entretanto também permite o ajuste de modelos mais simples (Pérez; De Los Campos, 2014).
- **QTL**: Pacote que fornece ferramentas para a análise de dados QTL (*quantitative trait loci*), incluindo métodos para a identificação de locos quantitativos, mapeamento de QTL, dentre outras funções e aplicabilidades (Broman et al., 2003). Link para o CRAN: <https://cran.r-project.org/web/packages/qtl/index.html>

Na manipulação e transformação de dados no R, os pacotes `dplyr` e `data.table` são recomendados. O `dplyr` é um pacote criado com o objetivo de fornecer um conjunto de ferramentas intuitivas para a manipulação de dados, como funções para filtragem, seleção, ordenação e agregação, que faz parte de uma coleção de pacotes destinados a ciência de dados chamado de *tidyverse* (Wickham et al., 2019). Ele tem uma sintaxe clara e concisa que facilita o trabalho com dados. Por outro lado, o `data.table` é um pacote mais antigo, que oferece uma sintaxe mais complexa, mas é conhecido por sua velocidade e eficiência em lidar com grandes conjuntos de dados. Ele também possui ampla variedade de funções, incluindo funções para a junção de tabelas, agrupamento e agregação. Enquanto o `dplyr` é mais voltado para a facilidade de uso, o `data.table` é mais adequado para operações mais complexas e exigentes em termos de desempenho. A escolha entre os dois pacotes geralmente depende do tipo de operações que serão realizadas e do tamanho dos dados a serem manipulados.

Python

O Python é uma linguagem de programação bastante utilizada em diversos setores, podendo também ser aplicado na biotecnologia e na genética e melhoramento de plantas. Dentro da genética, seu uso é difundido para a automação de tarefas, tais como a análise de dados genotípicos, fenotípicos e ambientais, e a construção de modelos de seleção genética. O Python também é utilizado para a integração de diferentes fontes de dados, como a análise de sequências de DNA e a visualização de dados. Tal como o R, ele tem a usabilidade facilitada e uma sintaxe clara, contando ainda com diversas bibliotecas específicas disponíveis para análise de dados genéticos. A desvantagem é que o Python tende a ser menos intuitivo para usuários menos familiarizados com programação. Aqui são citadas algumas das principais bibliotecas do Python, que podem ter aplicações diretas no melhoramento genético vegetal:

- **NumPy**: Essa biblioteca fornece ampla gama de funções matemáticas para o Python, incluindo operações com matrizes, manipulação de *arrays*, operações de álgebra linear, entre outras. Entre seus diferenciais, pode-se citar a performance otimizada, usabilidade simplificada e alta compatibilidade com outras plataformas e bibliotecas (Harris et al., 2020). Link: <https://numpy.org/>
- **Pandas**: Essa biblioteca provê funcionalidades de manipulação, estruturação e análise de dados para o Python. É utilizada na preparação e limpeza de dados, sendo um dos seus diferenciais a performance e a multifuncionalidade de seus recursos (Mckinney, 2010). Link: <https://pandas.pydata.org/>
- **Scikit-learn**: Essa biblioteca fornece ampla gama de ferramentas para aprendizado de máquina, supervisionado ou não. O scikit-learn ainda provê ferramentas para ajuste de modelos, pré-processamento de dados, métodos de agrupamento, classificação, regressão, e outras funcionalidades (Pedregosa et al., 2011). Link: <https://scikit-learn.org/stable/>
- **PyGMO**: Essa biblioteca oferece ampla gama de algoritmos de otimização para ambientes massivamente paralelos. Seu uso é comum nas áreas de genética evolutiva e melhoramento, sendo extremamente útil para projetar cenários levando em consideração todo o ambiente (Pedregosa et al., 2011). Link: <https://esa.github.io/pygmo2/>

SAS

O SAS (*Statistical Analysis System*) é um software estatístico privado muito utilizado em diversos setores. O SAS oferece recursos avançados de análise estatística, incluindo

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

modelagem, otimização, mineração de dados, visualização de dados e muito mais. Além disso, é uma plataforma baseada em *SAS Language*, o que significa que é possível escrever programas personalizados para atender às necessidades específicas do seu projeto, tendo, no entanto, um custo elevado e sendo menos intuitivo para usuários menos familiarizados com programação. No contexto do melhoramento genético de plantas, o SAS é utilizado para realizar análises estatísticas avançadas, incluindo a modelagem de dados genômicos e fenotípicos, a identificação de marcadores associados a características agrônomicas desejadas, a avaliação da herdabilidade e a seleção de progênies comerciais (Vargas et al., 2012). Adicionalmente, o SAS pode ser usado para visualizar e interpretar resultados de análises, bem como para gerar gráficos e relatórios.

MATLAB

O MATLAB pode ser utilizado para análises de dados genéticos, tais como a estimativa de parâmetros de modelos de seleção e a identificação de locos quantitativos importantes, sendo útil também para a visualização de dados genéticos, como a representação de matrizes de dados de associação. É amplamente utilizado na indústria e na academia, tem uma interface amigável e de usabilidade facilitada, possuindo diversas ferramentas integradas para análise de dados genéticos, como o pacote *Statistics and Machine Learning Toolbox*. No entanto, tal como o SAS, o MATLAB também não é gratuito, e pode ser menos flexível computacionalmente do que outros softwares. Algumas das principais funções do MATLAB para o melhoramento genético de plantas são:

- **Ferramentas de análise de dados:** O MATLAB oferece diversas ferramentas para a análise de dados, incluindo funções para limpeza, manipulação e visualização de dados. Link: <https://www.mathworks.com/products/matlab/data-analysis.html>
- **Ferramentas de otimização:** Fonte de grande conjunto de algoritmos de otimização, incluindo métodos de programação linear, não linear, estimação de parâmetros, seleção de componentes, gerenciamento de energia, entre outros. Link: <https://www.mathworks.com/products/optimization.html>
- **Ferramentas de aprendizado de máquina:** Os aplicativos de aprendizado de máquina oferecidos pelo MATLAB oferecem algoritmos de aprendizado supervisionados ou não, que incluem métodos de agrupamento, classificação e regressão, funcionando também como gerenciador de redes neurais. Link: <https://www.mathworks.com/solutions/machine-learning.html>
- **Ferramentas de visualização:** MATLAB oferece ampla gama de ferramentas de visualização de dados, como gráficos 2D/3D, mapas e animações. Link: <https://www.mathworks.com/help/matlab/graphics.html>

Julia

A linguagem de programação Julia também está ganhando popularidade no melhoramento genético de plantas, lançada em 2012 e desenvolvida especificamente para a computação científica e financeira, Julia oferece uma sintaxe fácil de usar e uma gama de bibliotecas para a análise de dados, como a mineração de dados, a modelagem estatística e otimização matemática (Gao et al., 2020). Julia pode especialmente ser útil no melhoramento genético de plantas, pois oferece velocidade e eficiência para a manipulação de grandes volumes de dados. Isso é importante nesse contexto, dado que a quantidade de dados genéticos disponíveis para o melhoramento vegetal está aumentando rapidamente devido ao avanço da tecnologia de sequenciamento de DNA (Castellucci, 2017; Hunold; Steiner, 2020), ou mesmo tecnologias de fenotipagem ou ambientipagem. Esses três casos (tecnologias de genotipagem, fenotipagem no contexto da fenômica e ambientipagem) estão apresentados com maior detalhamento nos respectivos capítulos deste livro.

Tal como o R e o Python, a linguagem Julia é de código aberto, tem uma sintaxe intuitiva e é amistosa com profissionais que tenham pouca experiência com programação. Ela também tem uma grande comunidade de usuários e desenvolvedores, e há muitos pacotes disponíveis para análise de dados genéticos. Além disso, a linguagem Julia é altamente interoperável, sendo possível se usar bibliotecas e pacotes para realizar a integração com outras ferramentas (Pereira; Siqueira, 2017). A desvantagem é que, até o presente momento, essa linguagem ainda está em fase inicial de desenvolvimento, e pode não ter a mesma capacidade de análise de dados como outros softwares já estabelecidos. Quando comparada com outras linguagens de código aberto, a documentação de seus pacotes também deixa a desejar. Alguns pacotes que podem ser úteis para o melhoramento são:

- **ProgenyTestingTools.jl**: Pacote que oferece recursos para simulações de testes de progênie. Oferece também opções para operações com marcadores genômicos a partir de pedigrees da população base. Link: <https://juliapackages.com/p/progenytestingtools>
- **GenomicFeatures.jl**: Conjunto de funções para trabalhar com dados genômicos, sendo possível por meio das ferramentas desse pacote ler, analisar, manipular e comparar diversos tipos de dados. Possui uma interessante ferramenta que procura locos específicos entre os dados fornecidos. Link: <https://juliapackages.com/p/GenomicFeatures>
- **JWAS.jl**: Ferramenta que permite, por meio de métodos de regressões múltiplas, realizar estudos de associação genômica e predição genômica. Link: <https://juliapackages.com/p/jwas>

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

- **VariantVisualization.jl**: Útil para criação, visualização, manipulação e interação de mapas genéticos em alta qualidade. Link: <https://juliapackages.com/p/variantvisualization>
- **MixedModels.jl**: Pacote que reúne ferramentas que possibilitam o ajuste e análise de modelos lineares e lineares de efeito misto. Tem funções similares e foi desenvolvido pela mesma equipe que o pacote **lme4** do R.

Esses pacotes podem ser encontrados no repositório oficial de pacotes do Julia <https://pkg.julialang.org/>.

Metodologia CRISP-DM no Melhoramento

Após as definições das linguagens utilizadas no melhoramento, especialmente o R e o Python, a Figura 3 pode ser atualizada (Figura 4). É importante ressaltar que o esquema, assim como os tópicos abordados neste capítulo, não formam uma nova proposta de metodologia, afinal, foram apenas usados alguns pacotes R e bibliotecas Python para ilustrar a aplicação da metodologia. Mesmo que os pacotes e bibliotecas abordados sejam extremamente úteis e de excelente qualidade, sabe-se que existe ampla gama de alternativas em ambas as linguagens de programação, capazes de realizar as mesmas tarefas.

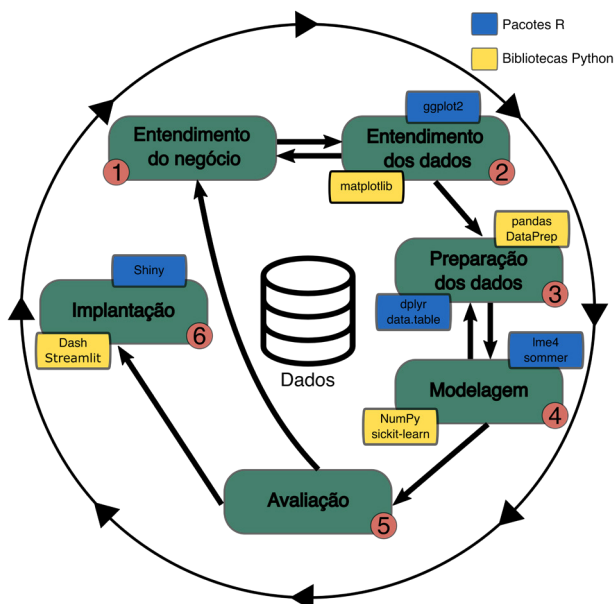


Figura 4. Metodologia CRISP-DM, atualizada com alguns exemplos de pacotes R e bibliotecas Python utilizadas no melhoramento.

As seis etapas demonstradas no tópico anterior, quando aplicadas em um contexto de melhoramento genético, podem ser atualizadas da seguinte forma:

- 1) **Entendimento do negócio:** O termo “negócio” é demasiadamente amplo, mas no contexto do melhoramento genético de plantas pode ser desde o ajuste de um bom modelo de seleção genômica até a criação de um programa de melhoramento do zero.
- 2) **Entendimento dos dados:** Ao se trabalhar com *big data*, a obtenção de uma visão clara dos dados disponíveis se torna uma tarefa difícil. Para isso, tem-se pacotes R como o ggplot2 e bibliotecas Python como a matplotlib (Hunter, 2007), que são capazes de fornecer maneiras elegantes de plotar e visualizar dados.
- 3) **Preparação dos dados:** Para esta etapa são necessários pacotes que sejam capazes de facilitar o tratamento dos dados. Nessa tarefa, no R, como já exemplificado anteriormente, destacam-se os pacotes dplyr e data.table. Já no Python, a biblioteca Pandas é bastante utilizada.
- 4) **Modelagem:** No melhoramento, existem diversos tipos de modelos e metodologias de análise utilizadas, os quais buscam se adequar às diferentes estratégias e delineamentos experimentais empregados. No R, destacam-se os pacotes lme4 e sommer. No Python, destaque para as bibliotecas NumPy e scikit-learn.
- 5) **Avaliação:** A avaliação pode ser realizada de diversas formas, desde metodologias de validação cruzada, até estatísticas como R^2 (coeficiente de determinação), erro padrão da estimativa, análises gráficas e análises a posteriori, no caso de modelos bayesianos.
- 6) **Implantação:** A última etapa, que consiste no uso efetivo do que foi construído durante a metodologia, pode compor a integração do modelo em sistemas já utilizados ou a criação de novas interfaces e aplicativos. O Shiny é um exemplo de sistema de desenvolvimento de aplicações e *dashboards* interativos em R e Python, que pode ser utilizado no desenvolvimento de uma plataforma de implantação (Chang et al., 2023). Já no Python, além do Shiny, existe maior variedade de alternativas com o mesmo propósito, como o Dash e o Streamlit. Além dos pacotes citados, vem se popularizando o uso do Power BI como ferramenta para criação de *dashboards* interativos. O Power BI também possui interface como o R e o Python, sendo capaz de executar scripts, facilitando a visualização e conectividade do que for desenvolvido.

Softwares nas Ciência de Dados

No melhoramento vegetal, o uso de softwares tem um importante papel na análise e interpretação de dados. Essas ferramentas computacionais são projetadas para extrair informações de conjuntos de dados massivos, permitindo que os melhoristas identifiquem padrões e estatísticas relevantes para aprimorar características específicas de plantas. Neste capítulo, conforme indicado anteriormente, será considerado como “software” as aplicações computacionais apresentadas com interface gráfica amigável de aponte e clique.

A popularidade das ferramentas de análise de dados na ciência vem mudando com o tempo. Na Figura 5 é apresentada a variação percentual na taxa de citação obtida pelo Google Scholar entre o período de 2019 e 2022 (Muenchen, 2023). Observa-se a representação visual da mudança percentual ao longo desses anos, em que os pacotes de software em ascensão, denominados “quentes”, são destacados em vermelho (lado direito), enquanto os pacotes em declínio, denominados “frios”, são representados em azul (lado esquerdo). Os com maior tendência de “impopularização” são as ferramentas pagas, tais como SPSS, GraphPad Prism, SAS, BMDP, Stata, Statistica e Systat. Entretanto, observa-se também um declínio no uso do R, possivelmente em decorrência do crescimento da popularidade do Python. Além disso, como esse gráfico mede o incremento entre os anos, e não a estabilidade, os softwares novos (ou não antes utilizados) possivelmente foram supervalorizados, e tiveram seus coeficientes percentuais mais elevados.

Será dado aqui um rápido destaque aos softwares gratuitos BlueSky, Jamovi e JASP. O BlueSky Statistics é um software de estatística integrado à linguagem R e de código aberto, projetado para ser fácil de usar e fornecer uma interface gráfica intuitiva, com recursos de arrastar e soltar, destinados a usuários que desejam projetar novos módulos ou funções estatísticas. O Jamovi também é um software de código aberto com interface gráfica, que oferece uma alternativa aos softwares pagos, como SPSS e SAS, sendo considerado uma “3ª geração” de planilhas estatísticas projetadas desde o início para serem fáceis de usar. Já o JASP é um software estatístico multiplataforma de código aberto, que combina o poder de computação do R com a simplicidade dos menus suspensos do SPSS, tornando-o ideal para alunos e professores.

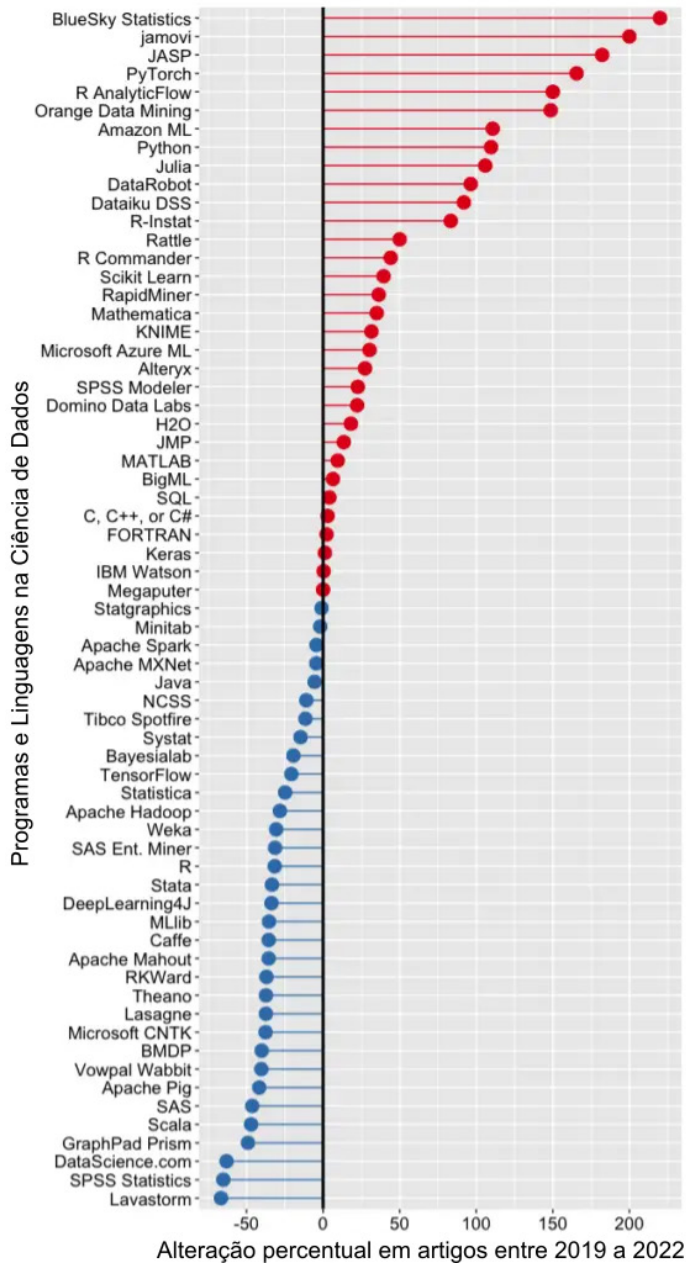


Figura 5. Variação percentual na taxa de citação obtida pelo Google Scholar entre o período de 2019 e 2022. Para melhorar a legibilidade do gráfico, os valores de crescimento do BlueSky (2960%) e do Jamovi (452%) foram encurtados.

Fonte: Adaptado de Data Science Software Reviews (2023).

Softwares livres nacionais no melhoramento

O GENES é um software desenvolvido pelo professor Dr. Cosme Damião Cruz, que tem como objetivo auxiliar os pesquisadores e profissionais do melhoramento genético de plantas. Ele permite a avaliação e a seleção de materiais genéticos de plantas baseadas em dados de rendimento e características de interesse. O software inclui vários métodos estatísticos, tais como diferentes configurações da análise de variância (ANOVA), análise de covariância, análise de correlação e análise de variância genética (Cruz, 2016). Falando sobre aplicabilidades, o software GENES contém diferentes módulos operacionais, destacando-se os módulos de análises biométricas, análises multivariadas, simulações (experimentos, amostras, número ideal de famílias e plantas, tamanho ideal de amostras, etc.), análises de diversidade genética, estatística descritiva e experimental, matrizes, módulo de inteligência computacional e aprendizado de máquina e módulo de integração com outros softwares (Cruz, 2016). Sobre integrações com outros softwares, o GENES é altamente compatível com os softwares R, Matlab e Selegen. Todas as integrações permitem novos módulos de análises pré-programados pelo GENES, o que é facilitado por tutoriais e documentação disponível gratuitamente online (Cruz, 2016).

O software SELEGEN-REML/BLUP foi desenvolvido pelo pesquisador e professor Dr. Marcos Deon Vilela de Resende (Resende, 2016). Ele foi projetado para pesquisadores e melhoristas utilizarem a análise da *Restricted* (ou *Residual*) *Maximum Likelihood* (REML) e o *Best Linear Unbiased Prediction* (BLUP) para avaliar o desempenho de materiais genéticos de plantas (Carvalho et al., 2022; Coser et al., 2022), inclusive sendo útil para o melhoramento genético animal. O software permite que o usuário defina seus objetivos de seleção e aloque parâmetros para analisar e ranquear cada característica avaliada, gerando listas de genótipos com as melhores predições de características (Pinto Júnior et al., 2005).

É importante também destacar o software brasileiro desenvolvido pelo professor Dr. Leonardo Bhering, chamado Rbio (Bhering, 2017), que é amplamente utilizado no melhoramento de plantas e possui abordagem semelhante aos softwares BlueSky, Jamovi e JASP. O Rbio é uma ferramenta útil e totalmente integrada ao R, oferecendo ampla gama de análises estatísticas acessíveis para usuários não familiarizados com a linguagem computacional do R.

Softwares internacionais no melhoramento

O ASReml (em suas versões ASReml-R e ASReml *standalone*) é um software estatístico desenvolvido pelo Centro de Genômica da Universidade de New England,

na Austrália, e é muito utilizado no melhoramento genético de plantas em todo mundo. Ele é utilizado para realizar análises estatísticas em dados genéticos, incluindo a modelagem de efeitos aleatórios e fixos, ajuste de covariáveis, análise de hibridação, seleção genômica, estimativas de parâmetros genéticos, etc. Embora ele tenha sua versão distribuída como pacote do R, o ASReml-R é um software pago, podendo ser oferecido gratuitamente para estudantes e instituições de ensino. O Genstat é um software estatístico desenvolvido pela VSN International, sendo bem direcionado para a genética e melhoramento de plantas e animais. Ele é uma ferramenta poderosa para análise de dados genéticos, como modelagem de efeitos aleatórios e fixos, análise de hibridação, estimação de parâmetros genéticos, etc. Ambos os softwares são utilizados na agroindústria e são considerados ferramentas para os cientistas e profissionais que trabalham com melhoramento genético de plantas.

O TASSEL, software criado e mantido pela Universidade do estado de Washington (EUA), é descrito como uma *pipeline* utilizada em larga escala para a análise de dados provenientes de genotipagem por sequenciamento de nova geração (Glaubitz et al., 2014). O software TASSEL é de código aberto e oferece uma variedade de ferramentas para construções de mapas genéticos, realização de análises de associação genômica, seleção genômica e análises de diversidade e identificação de SNPs (Wang et al., 2020).

O Haploview é um software projetado com o objetivo de auxiliar pesquisadores na visualização de padrões de desequilíbrio de ligação em dados oriundos da genotipagem de uma amostra, e realizar análises de associação entre SNPs e características fenotípicas (Barrett et al., 2005). O Haploview é um software de código aberto e foi projetado para trabalhar com *big data*, possuindo a capacidade de trabalhar com grandes conjuntos de dados de genotipagem e realizar análises para a imputação de dados faltantes (Barrett, 2009). Apesar de seu uso ser, na maioria dos casos, relacionados ao melhoramento animal, o Haploview é também usado no melhoramento genético de plantas (Wang et al., 2021).

Para realizar a imputação de dados genotípicos e a projeção de genótipos faltantes tem-se o software BEAGLE, que foi desenvolvido na Universidade de Auckland (Nova Zelândia) e é utilizado tanto em estudos de associação genômica como em análises de dados de sequenciamento (Ayres et al., 2012). Assim como os softwares mencionados anteriormente, o BEAGLE é de código aberto e foi projetado com foco na manipulação de grandes volumes de dados, incluindo dados de sequenciamento de nova geração. Embora seja usado principalmente por melhoristas animais, assim como o Haploview, também é funcional para o melhoramento de plantas (Niehoff et al., 2022).

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

O PLINK é um software utilizado para análise genômica em estudos de associação genética (Purcell et al., 2007). Ele fornece uma variedade de ferramentas para manipulação, análise, limpeza e visualização de dados genéticos. O PLINK é conhecido por sua eficiência e flexibilidade, permitindo a execução de ampla gama de análises genéticas, incluindo testes de associação, análises de estrutura populacional, controle de qualidade dos dados genotípicos e análises de ligação genética. Além disso, o PLINK é frequentemente utilizado em conjunto com o BEAGLE e é compatível com vários formatos de arquivos genéticos, oferecendo opções avançadas para personalização das análises.

Para genética de populações e conservação genética, tem-se a opção do software STRUCTURE (Porras-Hurtado et al., 2013). Por meio desse software, tem-se a possibilidade de utilizar inferência bayesiana para analisar dados de genotipagem e projetar a estrutura populacional de uma amostra, diferenciar populações geneticamente distintas e determinar a proporção de ascendência de um indivíduo dentro da população base. O STRUCTURE possui amplo uso na conservação genética, biologia evolutiva e estudos de impacto ante a extinção de espécies (Lancaster et al., 2006).

Ambientes de Desenvolvimento Integrado (IDE)

Um ambiente de desenvolvimento integrado (do inglês: *Integrated Development Environment* – IDE) é uma ferramenta de software que oferece um conjunto de recursos em uma interface amigável para escrever, testar e depurar códigos de programação. IDE é usado para ajudar os desenvolvedores a criar e gerenciar códigos em várias linguagens de programação, como o R, Python, Java e Julia (Han et al., 2018). Alguns exemplos de IDE popularmente usados inclui o *RStudio* para R, o *Spyder* para Python e o *Juno* para Julia (Sandhu et al., 2021; Van Dijk et al., 2021). No melhoramento genético de plantas, como em qualquer outra área de ciência, o IDE é usado para facilitar o processo de escrita e depuração de códigos. Ele fornece ferramentas para gerenciar e visualizar dados de forma eficiente, como tabelas e gráficos. Além disso, o IDE também oferece recurso de autocompletar, que sugere códigos ao desenvolvedor, e é útil para aumentar a eficiência do processo de escrita.

O IDE também ajuda na análise e interpretação dos resultados gerados pelos códigos. Ele permite aos desenvolvedores visualizar e comparar resultados de maneira interativa, ajudando na identificação de padrões e tendências. Isso pode

ser especialmente útil no melhoramento genético de plantas, em que é necessário analisar grandes quantidades de dados para identificar características genéticas importantes. Além disso, o IDE também permite a colaboração entre equipes de desenvolvedores, compartilhando e revisando códigos de forma fácil e eficiente. Isso pode ser especialmente útil no melhoramento genético de plantas, em que é comum trabalhar em equipe e em projetos de longa duração.

Plataformas de Gerenciamento de Compartilhamento de Código

As plataformas de compartilhamento, colaboração e versionamento de código são ferramentas úteis para equipes de desenvolvimento de software que trabalham juntas em projetos de código compartilhado. Elas mantêm uma cópia centralizada do código acessível a todos os desenvolvedores e registram todas as alterações feitas no código, permitindo o acesso a versões anteriores e o gerenciamento de versões (Oliveira; Dutra, 2014). Além disso, tais plataformas permitem a colaboração em tempo real, com funcionalidades que podem incluir comentários, tarefas e revisões de código, além de incorporarem ferramentas de desenvolvimento, como IDE e sistemas de construção (*build*), para uma experiência de trabalho fluida e integrada. Entre as plataformas, é possível exemplificar o *GitHub*, *GitLab* e *Bitbucket* (Balancieri et al., 2005).

O *GitHub*, talvez a mais famosa entre elas, é uma plataforma online que permite o compartilhamento, colaboração e versionamento de código. Ele é usado no melhoramento genético de plantas para ajudar a gerir, compartilhar e colaborar com outros pesquisadores e equipes diversas (Morales et al., 2020a; Covarrubias-Pazaran et al., 2022). Diferentes pesquisadores podem publicar seus códigos de análise de dados, modelos genéticos, ferramentas de mineração de dados ou bibliografia no *GitHub*, permitindo que outros membros da comunidade contribuam e usem essas ferramentas. O *GitHub* também oferece ferramentas de controle de versão, inserção de comentários e rastreamento de bugs, o que pode ajudar a manter a integridade e transparência dos projetos de melhoramento genético (Balancieri et al., 2005; Rodríguez-Álvarez et al., 2018; Morales et al., 2022).

Aqui são citadas algumas plataformas que oferecem funcionalidades similares ao *GitHub*:

- *GitLab*: é uma plataforma de código aberto e gerenciamento de projetos baseados em Git, que fornece recursos semelhantes aos do *GitHub*;

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

- *Bitbucket*: é uma plataforma de gerenciamento de código, que oferece suporte a várias linguagens de programação e fornece recursos de colaboração e controle de versão;
- *SourceForge*: é uma plataforma de hospedagem de código e gerenciamento de projetos, que oferece suporte a várias linguagens de programação e recursos de colaboração.
- *Apache Allura*: é uma plataforma de gerenciamento de projetos de software de código aberto, incluindo controle de versão, rastreamento de bugs e colaboração em projetos.

Repositórios de Dados de Melhoria

Os repositórios de dados fenotípicos, genéticos e genômicos são importantes para o melhoramento genético de plantas, porque fornecem informações para apoiar a investigação e a tomada de decisão na seleção de genótipos para programas de melhoramento. Alguns exemplos de dados fenotípicos disponíveis em repositórios on-line abrangem tanto informações sobre as características externas e internas das plantas, como altura, peso, cor, forma, resistência a doenças e tolerância a condições adversas. Esses dados são importantes para identificar as características desejáveis e selecionar as plantas mais aptas para a produção (Alonso et al., 2019).

Existem vários locais onde é possível obter dados fenotípicos de materiais genéticos em múltiplos ambientes, como:

- 1) The International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA): Esse tratado internacional oferece acesso a ampla gama de recursos genéticos de plantas, incluindo dados fenotípicos em múltiplos ambientes. O site oferece informações sobre a disponibilidade de germoplasma e outros recursos, bem como acesso a coleções de dados - <https://www.fao.org/plant-treaty/en/>
- 2) The Germplasm Resources Information Network (GRIN): Esse é um banco de dados online mantido pelo USDA Agricultural Research Service. Ele inclui informações sobre germoplasma (como dados fenotípicos, dados taxonômicos, etc.) para uma ampla variedade de plantas cultivadas e silvestres - <https://npgsweb.ars-grin.gov/gringlobal/>
- 3) The Plant Breeding and Acclimatization Institute (IHAR): Esse é um instituto de pesquisa na Polônia que mantém uma ampla gama de recursos genéticos de plantas. O site oferece informações sobre a disponibilidade de germoplasma e

outros recursos, bem como acesso a coleções de dados - <http://www.ihar.edu.pl/>

- 4) The International Maize and Wheat Improvement Center (Cimmyt): Esse centro internacional de pesquisa é dedicado ao melhoramento de trigo e milho. Ele mantém ampla gama de recursos genéticos, como dados fenotípicos em múltiplos ambientes. O site oferece informações sobre a disponibilidade de germoplasma e outros recursos, bem como acesso a coleções de dados - <https://cimmyt.org/>
- 5) The Plant Phenotyping Network: Esse é outro repositório global de dados fenotípicos de plantas, que inclui informações sobre a anatomia, fisiologia, biologia molecular, dentre outras - <http://www.plantphenotyping.org>
- 6) Germplasm Resources Information Network (GRIN): GRIN é uma rede de informações de germoplasma mantida pelo Departamento de Agricultura dos Estados Unidos (USDA), que fornece informações genotípicas e fenotípicas de muitas espécies de plantas - <https://npgsweb.ars-grin.gov/gringlobal>
- 7) The International Rice Research Institute (IRRI): O IRRI é uma organização global de pesquisa em arroz que compartilha dados genéticos, fenotípicos e genômicos de arroz para fins de pesquisa e melhoramento - <https://www.irri.org/resources/data-and-tools>.

Os dados genéticos e genômicos, como informações sobre a estrutura genética e a organização do DNA das plantas, incluindo a presença ou ausência de determinados alelos ou variantes genéticas, são importantes para entender a herança de características específicas e para identificar as fontes de variabilidade genética que podem ser utilizadas para melhorar as características de culturas de interesse agrícola. Existem vários repositórios e sites onde é possível baixar dados fenotípicos, genotípicos e genômicos relacionados ao melhoramento genético de plantas. Entre alguns dos mais conhecidos, pode-se citar:

- 1) Dryad Digital Repository - <https://datadryad.org/>: Esse é um repositório digital que fornece acesso a dados publicados associados a artigos científicos. Muitos artigos de melhoramento genético de plantas publicados em revistas científicas disponibilizaram seus dados no Dryad.
- 2) AgData Commons - <https://data.nal.usda.gov/>: Esse é um site do USDA que fornece acesso a dados de pesquisa agrícola, incluindo alguns dados de melhoramento genético de plantas.

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

- 3) GenBank - <https://www.ncbi.nlm.nih.gov/genbank/>: GenBank é uma coleção anotada de todas as sequências de DNA publicamente disponíveis, e os dados são atualizados diariamente.
- 4) Gramene - <http://www.gramene.org/>: É um banco de dados de código aberto e com curadoria para genômica comparativa de plantas e análise de rotas metabólicas, projetado para apoiar pesquisadores que trabalham em melhoramento genético e genômica vegetal.
- 5) Phytozome - <https://phytozome-next.jgi.doe.gov/>: Esse portal fornece à comunidade científica de plantas um hub para acessar, visualizar e analisar genomas de plantas já sequenciados, possibilitando, de modo fácil e intuitivo, realizar anotações e estudos de genômica.

Simulação de Dados

Diante da impossibilidade da implementação de um experimento para embasar uma pesquisa, ou caso não seja possível encontrar dados a partir um repositório, existe a alternativa de simular os próprios dados, respeitando as hipóteses estipuladas. A simulação de dados é uma ferramenta que faz sentido no contexto do melhoramento genético de plantas, permitindo que os pesquisadores estimem o comportamento de modelos teóricos de melhoramento de plantas com base em dados observados ou previstos (Li et al., 2012). Além disso, a simulação permite avaliar o impacto de hipóteses alternativas e estratégias de seleção na evolução futura da população (Daetwyler et al., 2013; Yamamoto et al., 2016). Algumas técnicas e métodos comumente usados na simulação de dados no melhoramento genético de plantas são:

- 1) **Modelos de simulação baseados em fenotipagem:** São modelos matemáticos que simulam a resposta fenotípica das plantas a diferentes condições ambientais e genéticas. Esses modelos são baseados em dados observados e relacionam variáveis genotípicas e ambientais com características fenotípicas desejadas;
- 2) **Modelos de simulação por computação de sistemas biológicos:** Essa técnica usa modelos computacionais para simular o comportamento biológico de plantas em diferentes condições ambientais e genéticas. Os modelos incluem características como a interação entre diferentes tecidos, as respostas hormonais e o transporte de nutrientes;

- 3) **Modelos de simulação baseados em genética quantitativa:** Esses modelos matemáticos simulam o comportamento genético de uma população de plantas ao longo do tempo. Eles consideram a herança quantitativa de características fenotípicas e a seleção natural;
- 4) **Modelos de simulação baseados em genômica:** Simulam a evolução do genoma de uma espécie ao longo do tempo, levando em conta fatores como mutações, recombinação e seleção natural. A simulação genômica pode ser usada para avaliar a diversidade genética e para entender como diferentes forças evolutivas afetam a evolução de um genoma.

Sistemas Gerenciadores de Banco de Dados

Sistemas gerenciadores de banco de dados (em inglês: *Database Management Systems* ou DBMS) são softwares que fornecem uma interface para criar, manter e consultar bancos de dados. Eles permitem que os usuários armazenem, acessem e atualizem informações de forma organizada e estruturada, garantindo a segurança dos dados, controle de acesso, recuperação de falhas e outros recursos importantes. Existem diversos tipos de DBMS, incluindo sistemas relacionais, sistemas de objetos, sistemas de dados distribuídos e sistemas NoSQL (Morales et al., 2020b).

O SQL é uma linguagem de programação utilizada para gerenciar e manipular dados em bancos de dados relacionais. No contexto do melhoramento genético de plantas, o SQL pode ser usado para armazenar e recuperar informações genéticas e fenotípicas, permitindo análises e manipulações mais eficientes desses dados. É possível, por exemplo, usar o SQL para combinar dados de diferentes fontes, filtrar informações irrelevantes e calcular estatísticas complexas com base nos dados armazenados. Adicionalmente, o SQL é compatível com vários sistemas gerenciadores, tornando mais fácil a integração com outras ferramentas e tecnologias (Zhao et al., 2021; Dipta et al., 2023).

Os DBMS podem ser usados para armazenar grandes quantidades de dados genotípicos e fenotípicos, e, adicionalmente, resultados de ensaios de campo podem ser registrados e gerenciados em um DBMS para facilitar sua análise e interpretação (Dorok et al., 2015). Os DBMS também podem ser usados para integrar dados de diferentes fontes, como bancos de dados públicos e registros comerciais.

Alguns dos sistemas DBMS mais conhecidos são: MySQL <https://www.mysql.com/>; PostgreSQL <https://www.postgresql.org/>; Microsoft SQL Server <https://www.microsoft.com/>.

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

[com/en-us/sql-server/sql-server-downloads](https://www.oracle.com/database/); Oracle Database <https://www.oracle.com/database/>; SQLite <https://www.sqlite.org/>; MongoDB <https://www.mongodb.com/>; e o MariaDB <https://mariadb.org/>.

Segurança e Direitos de Uso dos Dados

O uso de dados sensíveis no contexto do melhoramento genético de plantas, como informações genéticas e fenotípicas, requer atenção especial à segurança de dados e privacidade. O risco de vazamento ou uso inadequado desses dados pode ter consequências negativas para os proprietários de variedades, indústrias agrícolas e consumidores. Além disso, a legislação internacional e nacional sobre privacidade de dados deve ser seguida para garantir que as informações confidenciais sejam protegidas (Brasil, 2018; Garcia et al., 2020).

Existem várias estratégias para garantir a segurança de dados, como a criptografia de dados, acesso restrito a informações sensíveis e políticas de privacidade claras. É vital que as empresas e organizações que trabalham com dados sensíveis invistam em tecnologias avançadas de segurança da informação, como *firewalls*, autenticação de usuários e *backups* regulares. No contexto do melhoramento genético de plantas, a privacidade de dados é um assunto de suma importância, pois os dados genéticos e fenotípicos são confidenciais e podem ser usados para fins comerciais. As empresas de melhoramento genético precisam seguir as regulamentações internas e ter medidas de segurança sólidas para garantir que os dados sensíveis sejam protegidos.

É importante salientar que as tecnologias de *big data* e inteligência artificial, que são utilizadas no melhoramento genético de plantas, também precisam ser integradas com as práticas de segurança de dados (Bernardo, 2020). Por exemplo, ao coletar dados genéticos e fenotípicos usando drones ou sensores, é importante garantir que as informações sejam armazenadas de forma segura e que as políticas de privacidade sejam claramente definidas.

Mais Alguns Exemplos de Aplicação

A combinação das ferramentas apresentadas até aqui permite a integração e análise de informações em larga escala, o que é importante para o sucesso no melhoramento genético de plantas. Entre algumas aplicações da integração de ferramentas e conceitos computacionais no melhoramento genético, pode-se citar:

- 1) **Predição de características de híbridos:** utilização de algoritmos de aprendizado de máquina para prever a expressão de características em híbridos resultantes do cruzamento alógamo ou autógamo;
- 2) **Mapa genético:** criação de mapas genéticos que possibilitem a identificação de locos importantes relacionados a características produtivas e de adaptação;
- 3) **Armazenamento de informações climáticas:** armazenamento de informações climáticas relevantes ao cultivo das plantas, permitindo análise mais precisa das condições adversas que afetam o desenvolvimento e a produção;
- 4) **Estimativa de rendimento:** utilização de algoritmos de inteligência artificial para estimar o rendimento de híbridos específicos com base em informações climáticas, genéticas e produtivas;
- 5) **Seleção de germoplasma:** uso de programas de computação para selecionar plantas com características desejadas em uma população sob avaliação, sem a necessidade de colher as parcelas experimentais.

A computação, programação, mineração e manejo de dados são ferramentas que permitem o processamento e análise de grandes quantidades de dados relacionados ao melhoramento genético de plantas alógamas e autógamas. Já as ferramentas de *big data* permitem armazenar enormes quantidades de informações para caracterização do genótipo, fenótipo e ambiótipo, das plantas e ambientes de interesse. Alguns exemplos de como essas ferramentas podem ser aplicadas são:

- 1) **Modelagem de características:** programação de algoritmos de inteligência artificial para prever as características de plantas resultantes do cruzamento alógamo ou autógamo.
- 2) **Análise de genoma:** mineração de dados para identificar locos importantes relacionados a características desejadas, como rendimento e resistência a doenças.
- 3) **Seleção assistida:** uso de programas de computação para auxiliar na seleção de plantas com as características desejadas durante o melhoramento genético.
- 4) **Armazenamento de dados:** armazenamento de informações relacionadas ao cruzamento, seleção e teste de plantas, permitindo a análise a longo prazo dos resultados.

Esses e outros temas listados anteriormente serão abordados com maior detalhamento nos respectivos capítulos deste livro.

Considerações Finais

Se atentarmos ao fato de que a quantidade de dados genômicos disponíveis para plantas está crescendo rapidamente, e a integração desses dados com dados fenotípicos e de melhoramento será crucial para aprimorar as estratégias de melhoramento genético, pode-se inferir que a integração de dados genômicos será uma frente que crescerá com velocidade e força nas próximas décadas.

Aliado à integração de dados genômicos, o uso de técnicas de aprendizado de máquina para analisar grandes quantidades de dados fenotípicos e genômicos está se tornando cada vez mais comum e será uma ferramenta valiosa para identificar marcadores genéticos e prever o desempenho de linhagens sob seleção em relação às características de interesse.

Junto com diversas novas técnicas de aprendizado de máquina, o uso de inteligência artificial está cada vez mais sendo aplicado para a identificação de padrões em dados complexos, o que pode ser útil no melhoramento genético de plantas na identificação de combinações de genes que potencialmente influenciam características de interesse de melhoristas.

A integração de dados ambientais, como dados de solos, climáticos e de manejo, juntamente com dados fenotípicos e genômicos, pode fornecer uma visão mais completa sobre como as plantas respondem a diferentes condições ambientais, assim como no desenvolvimento de variedades adaptadas a mudanças climáticas. A ambientômica é também um caminho para o aumento da produção bruta de locais já sob cultivo, diminuindo a necessidade de abertura de novas áreas, resultando na preservação e conservação de ambientes naturais. Pode-se também destacar que a colaboração internacional é crescente na área de melhoramento genético de plantas, o que está levando ao compartilhamento de dados e recursos computacionais, fato que é benéfico para a velocidade do avanço e assertividade das pesquisas na área.

Referências

ALONSO, C.; RAMOS-CRUZ, D.; BECKER, C. The role of plant epigenetics in biotic interactions. **New Phytologist**, v. 221, n. 2, p. 731–737, Jan. 2019. DOI: <https://doi.org/10.1111/nph.15408>.

AMADEU, R. R.; CELLON, C.; OLMSTEAD, J. W.; GARCIA, A. A. F.; RESENDE, M. F. R.; MUÑOZ, P. R. AGHmatrix: R package to construct relationship matrices for autotetraploid and diploid species: a blueberry example. **The Plant Genome**, v. 9, n. 3, p. 1–10, Nov. 2016. DOI: <https://doi.org/10.3835/plantgenome2016.01.0009>.

AYRES, D. L.; DARLING, A.; ZWICKL, D. J.; BEERLI, P.; HOLDER, M. T.; LEWIS, P. O.; HUELSENBECK, J. P.; RONQUIST, F.; SWOFFORD, D. L.; CUMMINGS, M. P.; RAMBAUT, A.; SUCHARD, M. A. BEAGLE: an application programming interface and high-performance computing library for statistical phylogenetics. **Systematic Biology**, v. 61, n. 1, p. 170–173, Jan. 2012. DOI: <https://doi.org/10.1093/sysbio/syr100>.

BALANCIERI, R.; BOVO, A. B.; KERN, V. M.; PACHECO, R. C. S.; BARCIA, R. M. A análise de redes de colaboração científica sob as novas tecnologias de informação e comunicação: um estudo na Plataforma Lattes. **Ciência da Informação**, v. 34, n. 1, p. 64–77, jan./abr. 2005. DOI: <https://doi.org/10.1590/S0100-19652005000100008>.

BARRETT, J. C. Haploview: visualization and analysis of SNP genotype data. **Cold Spring Harbor Protocols**, v. 4, n. 10, p. 1-5, Oct. 2009. DOI: <https://doi.org/10.1101/pdb.ip71>.

BARRETT, J. C.; FRY, B.; MALLER, J.; DALY, M. J. Haploview: analysis and visualization of LD and haplotype maps. **Bioinformatics**, v. 21, n. 2, p. 263–265, Jan. 2005. DOI: <https://doi.org/10.1093/bioinformatics/bth457>.

BATES, D.; MACHLER, M.; BOLKER, B.; WALKER, S. Fitting linear mixed-effects models using lme4. **Journal of Statistical Software**, v. 67, n. 1, 2015. DOI: <https://doi.org/10.18637/jss.v067.i01>.

BERNARDO, R. Reinventing quantitative genetics for plant breeding: something old, something new, something borrowed, something BLUE. **Heredity**, v. 125, p. 375-385, Apr. 2020. DOI: <https://doi.org/10.1038/s41437-020-0312-1>.

BHERING, L. L. Rbio: a tool for biometric and statistical analysis using the R platform. **Crop Breeding and Applied Biotechnology**, v. 17, n. 2, p. 187–190, June 2017. DOI: <https://doi.org/10.1590/1984-70332017v17n2s29>.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Dispõe sobre a proteção de dados pessoais e altera a Lei nº 12.9675, de 23 de abril de 2014. **Diário Oficial da União**, n. 157-A, p. 1, ago. 2018. Seção 1. Disponível em: <https://www.in.gov.br/web/dou/-/lei-n-13-709-de-14-de-agosto-de-2018-36889940>. Acesso em: 18 fev. 2023.

BROMAN, K. W.; WU, H.; SEM, S.; CHURCHILL, G. A. R/qtl: QTL mapping in experimental crosses. **Bioinformatics**, v. 19, n. 7, p. 889–890, May 2003. DOI: <https://doi.org/10.1093/bioinformatics/btg112>.

CARVALHO, J. L. R.; VIEIRA, M. F. B.; ALBUQUERQUE, T. M.; COSTA, G. A. L.; GOMES, R. L. F.; LOPES, A. C. A. Lima bean populations assessments via Reml/Blup methodology. **Revista Caatinga**, v. 35, n. 4, p. 783–790, out./dez. 2022. DOI: <https://doi.org/10.1590/1983-21252022v35n405rc>.

CASTELLUCCI, P. B. JULIA e JuMP: novas ferramentas para programação matemática. **Pesquisa Operacional para o Desenvolvimento**, v. 9, n. 2, p. 48–61, 2017. DOI: <https://doi.org/10.4322/PODes.2017.005>.

CHANG, W.; CHENG, J.; ALLAIRE, J. J.; SIEVERT, C.; SCHLOERKE, B.; XIE, Y. et al. **shiny: Web application framework for R**. Disponível em: <https://rdr.io/cran/shiny/>. Acesso em: 19 jun. 2023.

COSER, S. M.; MOTOIKE, S. Y.; CORRÊA, T. R.; QUEIROZ, V.; COSTA, A. M.; NASCIMENTO, H. R.; MADEIRA, D. D. C.; RESENDE, M. D. V. Estimates of genetic parameters and selection gains for phenological and productive characteristics in Macaw Palm: an important plant for bioenergy. **Functional Plant Breeding Journal**, v. 4, n. 2, p. 23–33, July/Dec. 2022.

COVARRUBIAS-PAZARAN, G. Genome-Assisted prediction of quantitative traits using the r package sommer. **PLoS ONE**, v. 11, e0156744, June 2016. DOI: <https://doi.org/10.1371/journal.pone.0156744>.

COVARRUBIAS-PAZARAN, G.; GEBEYEHU, Z.; GEMENET, D.; WERNER, C.; LABROO, M.; SIRAK, S.; COALDRAKE, P.; RABBI, I.; KAYONDO, S. I.; PARKES, E.; KANJU, E.; MBANJO, E. G. N.; AGBONA, A.; KULAKOW, P.; QUINN, M.; DEBAENE, J. Breeding schemes: what are they, how to formalize them, and how to improve them? **Frontiers in Plant Science**, v. 12, 791859, Jan. 2022. DOI: <https://doi.org/10.3389/fpls.2021.791859>.

CROSSA, J.; FRITSCHÉ-NETO, R.; MONTESINOS-LOPEZ, O. A.; COSTA-NETO, G., DREISIGACKER, S.; MONTESINOS-LOPEZ, A.; BENTLEY, A. R. The modern plant breeding triangle: optimizing the use of genomics, phenomics, and enviromics data. **Frontiers in Plant Science**, v. 12, 651480, Apr. 2021. DOI: <https://doi.org/10.3389/fpls.2021.651480>.

CRUZ, C. D. Programa Genes – ampliado e integrado aos aplicativos R, Matlab e Selegen. **Acta Scientiarum. Agronomy**, v. 38, n. 4, p. 547–552, out./dez. 2016. DOI: <https://doi.org/10.4025/actasciagron.v38i4.32629>.

DAETWYLER, H. D.; CALUS, M. P. L.; PONGO-WONG, R.; DE LOS CAMPOS, G.; HICKEY, J. M. Genomic prediction in animals and plants: Simulation of data, validation, reporting, and benchmarking. **Genetics**, v. 193, n. 2, p. 347–365, Feb. 2013. DOI: <https://doi.org/10.1534/genetics.112.147983>.

DATA SCIENCE SOFTWARE REVIEWS. **r4stats.com**. Disponível em: <https://r4stats.com/2023/06/07/update-to-data-science-software-popularity/>. Acesso em: 19 jun. 2023.

DIPTA, B.; SOOD, S.; DEVI, R.; BHARDWAJ, V.; MANGAL, V.; THAKUR, A. K.; KUMAR, V.; PANDEY, N. K.; RATHORE, A.; SINGH, A. K. Digitalization of potato breeding program: improving data collection and management. **Heliyon**, v. 9, e12974, Jan. 2023. DOI: <https://doi.org/10.1016/j.heliyon.2023.e12974>.

DOROK, S.; BREB, S.; SAAKE, G.; TEUBNER, J. Flexible analysis of plant genomes in a database management system. In: INTERNATIONAL CONFERENCE ON EXTENDING DATABASE TECHNOLOGY, 18., Brussels, 2015. **Proceedings**. Brussels: EDTB, 2015. p. 509–512, 2015. DOI: <https://doi.org/10.5441/002/edbt.2015.45>.

EDDELBUETTEL, D.; BALAMUTA, J. J. Extending R with C++: a brief introduction to Rcpp. **American Statistician**, v. 72, n. 1, p. 28–36, Sept. 2017. DOI: <https://doi.org/10.1080/00031305.2017.1375990>.

GAO, K.; MEI, G.; PICCIALLI, F.; CUOMO, S.; TU, J.; HUO, Z. Julia language in machine learning: algorithms, applications, and open issues. **Computer Science Review**, v. 37, 100254, Aug. 2020. DOI: <https://doi.org/10.1016/j.cosrev.2020.100254>.

GARCIA, L. R.; AGUILERA-FERNANDES, E.; GONÇALVES, R. A. M.; PEREIRA-BARRETTO, M. R. **Lei geral de proteção de dados (LGPD): guia de implantação**. São Paulo: Blucher, 2020. 128 p.

GLAUBITZ, J. C.; CASSTEVEN, T. M.; LU, F.; HARRIMAN, J.; ELSHIRE, R. J.; SUN, Q.; BUCKLER, E. S. TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. **PLoS ONE**, v. 9, e90346, Feb. 2014. DOI: <https://doi.org/10.1371/journal.pone.0090346>.

HAN, Y.-Y.; WANG, K.-Y.; LIU, Z.-Q.; PAN, S.-H.; ZHAO, X.-Y.; WANG, S.-F. Golden seed breeding cloud platform for the management of crop breeding material and genealogical tracking. **Computers and Electronics in Agriculture**, v. 152, p. 206–214, Sept. 2018. DOI: <https://doi.org/10.1016/j.compag.2018.07.015>.

HARRIS, C. R.; MILLMAN, K. J.; VAN DER WALT, S. J.; GOMMERS, R.; VIRTANEN, P.; COURNAPEAU, D.; WIESER, E.; TAYLOR, J.; BERG, S.; SMITH, N. J.; KERN, R.; PICUS, M.; HOYER, S.; VAN KERKWIJK, M. H.; BRETT, M.; HALDANE, A.; DEL RÍO, J. F.; WIEBE, M.; PETERSON, P.; GÉRARD-MARCHANT, P.; SHEPPARD, K.; REDDY, T.; WECKESSER, W.; ABBASI, H.; GOHLKE, C.; OLIPHANT, T. E. Array programming with NumPy. **Nature**, v. 585, n. 7825, p. 357–362, Sept. 2020. DOI: <https://doi.org/10.1038/s41586-020-2649-2>.

HUNOLD, S.; STEINER, S. Benchmarking Julia’s communication performance: Is Julia HPC ready or full HPC? In: PMBS 2020: INTERNATIONAL WORKSHOP ON PERFORMANCE MODELING, BENCHMARKING AND SIMULATION OF HIGH PERFORMANCE COMPUTER SYSTEMS, 2020. **Proceedings**. 2020. DOI: <https://doi.org/10.1109/PMBS51919.2020.00008>.

HUNTER, J. D. Matplotlib: a 2D graphics environment. **Computing in Science & Engineering**, v. 9, n. 3, p. 90–95, May 2007. DOI: <https://doi.org/10.1109/MCSE.2007.55>.

KAMILARIS, A.; KARTAKOULLIS, A.; PRENAFETA-BOLDÚ, F. X. A review on the practice of big data analysis in agriculture. **Computers and Electronics in Agriculture**, v. 143, n. 2, p. 23-37, Dec. 2017. DOI: <https://doi.org/10.1016/j.compag.2017.09.037>.

LANCASTER, M. L.; GEMMELL, N. J.; NEGRO, S.; GOLDSWORTHY, S.; SUNNUCKS, P. Ménage à trois on Macquarie Island: hybridization among three species of fur seal (*Arctocephalus* spp.) following historical population extinction. **Molecular Ecology**, v. 15, n. 12, p. 3681–3692, Oct. 2006. DOI: <https://doi.org/10.1111/j.1365-294X.2006.03041.x>.

LI, X.; ZHU, C.; WANG, J.; YU, J. Computer simulation in plant breeding. **Advances in Agronomy**, v. 116, p. 219-264, 2012. DOI: <https://doi.org/10.1016/B978-0-12-394277-7.00006-3>.

LINKEDIN. **Data science academy**. Disponível em: <https://www.linkedin.com/company/data-science-academy/>. Acesso em 19 jun. 2023.

MCKINNEY, W. Data structures for statistical computing in Python. In: PYTHON FOR SCIENTIFIC COMPUTING CONFERENCE, 9., 2010, Austin. **Proceedings**. DOI: <https://doi.org/10.25080/Majora-92bf1922-00a>.

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>.

MORALES, N.; KACZMAR, N. S.; SANTANTONIO, N.; GORE, M. A.; MUELLER, L. A.; ROBBINS, K. R. ImageBreed: open-access plant breeding web–database for image-based phenotyping. **Plant Phenome Journal**, v. 3, e20004, 2020a. DOI: <https://doi.org/10.1002/ppj2.20004>.

MORALES, N.; BAUCHET, G. J.; TANTIANKJANA, T.; POWELL, A. F.; ELLERBROCK, B. J.; TECLE, I. Y.; MUELLER, L. A. High density genotype storage for plant breeding in the Chado schema of Breedbase. **PLoS ONE**, v. 15, e0140059, Nov. 2020b. DOI: <https://doi.org/10.1371/journal.pone.0240059>.

MORALES, N.; OGBONNA, A.; ELLERBROCK, B. J.; BAUCHET, G. J.; TANTIANKJANA, T.; TECLE, I. Y.; POWELL, A. F.; LYON, D.; MENDA, N.; SIMOES, C. C.; SAHA, S.; HOSMANI, P.; FLORES, M.; PANITZ, N.; PREBLE, R. S.; AGBONA, A.; RABBI, I.; KULAKOW, P.; PETETI, P.; KAWUKI, R.; ESUMA, W.; KANAABI, M.; CHELANGAT, D. M.; UBA, E.; OLOJEDE, A.; ONYEKA, J.; SHAH, T.; KARANJA, M.; EGESI, C.; TUFAN, H.; PATERNE, A.; ASFAW, A.; JANNINK, J. L.; WOLF, M.; BIRKETT, C. L.; WARING, D. J.; HERSHBERGER, J. M.; GORE, M. A.; ROBBINS, K. R.; RIFE, T.; COURTNEY, C.; POLAND, J.; ARNAUD, E.; LAPORTE, M.-A.; KULLEMBEKA, H.; SALUM, K.; MREMA, E.; BROWN, A.; BAYO, S.; UWIMANA, B.; AKECH, V.; YENCHO, C.; BOECK, B.; CAMPOS, H.; SWENNEN, R.; EDWARDS, J. D.; MUELLER, L. A. Breedbase: a digital ecosystem for modern plant breeding. **G3 Genes|Genomes|Genetics**, v. 12, jkac078, July 2022. DOI: <https://www.doi.org/10.1093/g3journal/jkac078>.

MUENCHEN, R. **Update to data science software popularity**. Disponível em: <https://r4stats.com/2023/06/07/update-to-data-science-software-popularity/>. Acesso em: 19 jun. 2023.

MUÑOZ, F. **BreedR overview**. 2017. Disponível em: <https://github.com/famuvie/breedR/wiki/Overview.pdf>. Acesso em: 19 jun. 2023.

NAZIAN, M.; NIEDBAŁA, G. Machine learning for plant breeding and biotechnology. **Agriculture**, v.10, 436, Sept. 2020. DOI: <https://doi.org/10.3390/agriculture10100436>.

NIEHOFF, T.; POOK, T.; GHOLAMI, M.; BEISSINGER, T. Imputation of low-density marker chip data in plant breeding: evaluation of methods based on sugar beet. **Plant Genome**, v. 15, e20257, Dec. 2022. DOI: <https://doi.org/10.1002/tpg2.20257>.

OLIVEIRA, E. B.; DUTRA, M. L. Um levantamento sobre do uso de ferramentas da Web 2.0 entre os estudantes da Ciência da Informação da Universidade Federal de Santa Catarina. **Encontros Bibli**:

Revista Eletrônica de Biblioteconomia e Ciência da Informação, v. 19, n. 39, p. 153–182, 2014. DOI: <https://doi.org/10.5007/1518-2924.2014v19n39p153>.

OLIVOTO, T.; LÚCIO, A. D. metan: An R package for multi-environment trial analysis. **Methods in Ecology and Evolution**, v. 11, n. 6, p. 783–789, June 2020. DOI: <https://publons.com/publon/10.1111/2041-210X.13384>.

PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O. et al. Scikit-learn: machine learning in Python. **Journal of Machine Learning Research**, v. 12, p. 2825–2830, Oct. 2011. DOI: <https://doi.org/10.48550/arXiv.1201.0490>.

PEREIRA, J. M.; SIQUEIRA, M. B. B. Linguagem de programação JULIA: uma alternativa open source e de alto desempenho ao MATLAB. **Revista Principia**, v. 1, n. 34, p. 132-140, maio 2017. DOI: <http://dx.doi.org/10.18265/1517-03062015v1n34p132-140>.

PÉREZ, P.; DE LOS CAMPOS, G. Genome-wide regression and prediction with the BGLR statistical package. **Genetics**, v. 198, n. 2, p. 483–495, Oct. 2014. DOI: <https://doi.org/10.1534/genetics.114.164442>.

PIASKOWSKI, J. **R packages for agricultural research**. 2022. Disponível em: <https://agstats.io/post/keeping-up-with-r/>. Acesso em: 19 jun. 2023.

PINTO JÚNIOR, J. E.; STURION, J. A.; RESENDE, M. D. V.; RONZELLI JÚNIOR, P. Efeitos da correção de dados na redução da heterogeneidade das variâncias genética, ambiental e fenotípica em testes de progênies de *Eucalyptus grandis* W. Hill ex Maiden. **Pesquisa Florestal Brasileira**, v. 51, p. 75–107, jul./dez. 2005.

PORRAS-HURTADO, L.; RUIZ, Y.; SANTOS, C.; PHILLIPS, C.; CARRACEDO, A.; LAREU, M. V. An overview of STRUCTURE: applications, parameter settings, and supporting software. **Frontiers in Genetics**, v. 4, 98, May 2013. DOI: <https://doi.org/10.3389/fgene.2013.00098>.

PURCELL, S.; NEALE, B.; TODD-BROWN, K.; THOMAS, L.; FERREIRA, M. A. R.; BENDER, D.; MALLER, J.; SKLAR, P.; BAKKER, P. I. W.; DAY, M. J.; SHAM, P. C. PLINK: A tool set for whole-genome association and population-based linkage analyses. **American Journal of Human Genetics**, v. 81, n. 3, p. 559–575, Sept. 2007. DOI: <https://doi.org/10.1086/519795>.

RESENDE, M. D. V. Software Selegen-REML/BLUP: a useful tool for plant breeding. **Crop Breeding and Applied Biotechnology**, v. 16, n. 4, p. 330–339, Sept. 2016. DOI: <https://doi.org/10.1590/1984-70332016v16n4a49>.

RESENDE, R. T.; CHENU, K.; RASMUSSEN, S. K.; HEINEMANN, A. B.; FRITSCHÉ-NETO, R. Editorial: Enviromics in plant breeding. **Frontiers in Plant Science**, v. 13, 935380, June 2022. DOI: <https://doi.org/10.3389/fpls.2022.935380>.

RODRÍGUEZ-ÁLVAREZ, M. X.; BOER, M. P.; Van EEUWIJK, F. A.; EILERS, P. H. C. Correcting for spatial heterogeneity in plant breeding experiments with P-splines. **Spatial Statistics**, v. 23, p. 52–71, Mar. 2018. DOI: <https://doi.org/10.1016/j.spasta.2017.10.003>.

SALTZ, J. S. CRISP-DM for data science: strengths, weaknesses and potential next steps. In: INTERNATIONAL CONFERENCE ON BIG DATA, 2021. **Proceedings**. Disponível em: <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9671634>. Acesso em: 18 fev. 2023.

SANDHU, K. S.; LOZADA, D. N.; ZHANG, Z.; PUMPHREY, M. O.; CARTER, A. H. Deep learning for predicting complex traits in spring Wheat Breeding Program. **Frontiers in Plant Science**, v. 11, 613325, Jan. 2021. DOI: <https://doi.org/10.3389/fpls.2020.613325>.

SHEARER, C. The CRISP-DM model: the new blueprint for data mining. **Journal of Data Warehousing**, v. 5, n. 4, p. 13–22, 2000.

THE R FOUNDATION. **The R Project for statistical computing**. 2022. Disponível em: <https://www.r-project.org/>. Acesso em: 19 jun. 2023.

VAN DIJK, A. D. J.; KOOTSTRA, G.; DRIJER, W.; RIDDER, D. Machine learning in plant science and plant breeding. **iScience**, v. 24, 101890, Jan. 2021. DOI: <https://doi.org/10.1016/j.isci.2020.101890>.

VARGAS, M.; COMBS, E.; ALVARADO, G.; ATLIN, G.; MATHEWS, K.; CROSSA, J. META: a suite of SAS Programs to analyze multienvironment breeding trials. **Agronomy Journal**, v. 105, n. 1, p. 11–19, Jan. 2012. DOI: <https://doi.org/10.2134/agronj2012.0016>.

WANG, B.; GUO, X.; ZHAO, P.; LIAO, W.; ZENG, C.; LI, K.; ZHOU, Y.; XIAO, J.; RUAN, M.; PENG, M.; BAI, Y.; CHEN, Y. MeMYB26, a drought-responsive transcription factor in cassava (*Manihot esculenta* Crantz). **Crop Breeding and Applied Biotechnology**, v. 21, 334432114, 2021. DOI: <https://doi.org/10.1590/1984-70332021v21n1a4>.

WANG, N.; YUAN, Y.; WANG, H.; YU, D.; LIU, Y.; ZHANG, A. et al. Applications of genotyping-by-sequencing (GBS) in maize genetics and breeding. **Scientific Reports**, v. 10, 16308, Oct. 2020. DOI: <https://doi.org/10.1038/s41598-020-73321-8>.

WICKHAM, H. **ggplot2: Elegant graphics for data analysis**. 2nd. ed. New York: Springer-Verlag, 2016. 260 p.

WICKHAM, H.; AVERICK, M.; BRYAN, J.; CHANG, W.; MCGOWAN, L.; FRANÇOIS, R.; GROLEMUND, G.; HAYES, A.; HENRY, L.; HESTER, J.; KUHN, M.; PEDERSEN, T. L.; MILLER, E.; BACHE, S. M.; MÜLLER, K.; OOMS, J.; ROBINSON, D.; SEIDEL, D. P.; SPINU, V.; TAKAHASHI, K.; VAUGHAN, D.; WILKE, C.; WOO, K.; YUTANI, H. Welcome to the Tidyverse. **Journal of Open Source Software**, v. 4, n. 43, 1686, Nov. 2019. DOI: <https://doi.org/10.21105/joss.01686>.

WIMMER, V.; ALBRECHT, T.; AUINGER, H.-J.; SCHÖN, C.-C. Synbreed: A framework for the analysis of genomic prediction data using R. **Bioinformatics**, v. 28, n. 15, p. 2086–2087, Aug. 2012. DOI: <https://doi.org/10.1093/bioinformatics/bts335>.

YAMAMOTO, E.; MATSUNAGA, H.; ONOGI, A.; KANEGAE, H. K.; MINAMIKAWA, M.; SUZUKI, A.; SHIRASAWA, K.; HIRAKAWA, H.; NUNOME, T.; YAMAGUCHI, H.; MIYATAKE, K.; OHYAMA, A.; IWATA, H.; FUKUOKA, H. A simulation-based breeding design that uses whole-genome prediction in tomato. **Scientific Reports**, v. 6, '9454, Jan. 2016. DOI: <https://doi.org/10.1038/srep19454>.

ZHAO, Y.; JIANG, B.; HUO, Y.; YI, H.; TIAN, H.; WU, H.; WANG, R.; ZHAO, J.; WANG, F. A high-performance database management system for managing and analyzing large-scale SNP data in plant genotyping and breeding applications. **Agriculture**, v. 11, 1027, Nov. 2021. DOI: <https://doi.org/10.3390/agriculture11111027>.

Moyisés Nascimento, Camila Azevedo Ferreira,
Cosme Damião Cruz, Ana Carolina Campana Nascimento

Introdução

Inteligência Artificial (IA) é a área da ciência que estuda a teoria, o desenvolvimento e a aplicação de técnicas inspiradas na inteligência humana. Tais metodologias estão fundamentadas em um sistema que incluem elementos que mimetizam o cérebro humano, ou seja, um mecanismo em que as decisões são tomadas com base nas entradas recebidas de nossos cinco sentidos (visão, audição, tato, olfato e paladar). Também tentam simular o comportamento do cérebro humano, ou seja, aprender, errar e realizar descobertas (Cruz; Nascimento, 2018). Em termos de análises, a IA relaciona-se ao processo de ajustar e prever comportamentos provenientes de qualquer tipo de dados e gerados por qualquer função, podendo ser aplicada para solucionar problemas de previsão, classificação e reconhecimento de padrões. Dentre as técnicas de IA, destacam-se as redes neurais artificiais (RNA) e a lógica nebulosa (*Fuzzy Logic*).

Por outro lado, o Aprendizado de Máquinas (AM) permite que computadores aprendam a partir de algoritmos, que por meio de alguma medida de erro, constroem regras para realizar previsões ou classificações. Dessa forma, é possível verificar que o AM faz uso de um raciocínio indutivo, o qual se aprende com base na experimentação e exemplos. Diferentemente do AM, a IA utiliza tanto o raciocínio indutivo quanto o dedutivo, que consiste em utilizar a lógica para construir um novo conhecimento a partir do conhecimento já existente acerca do fenômeno em estudo. Dentre as técnicas de AM, podemos citar as Árvores de Regressão e Classificação e seus refinamentos, como por exemplo *Bagging* e *Random Forest*.

Na área da Biometria, a previsão, a classificação e o reconhecimento de padrões são indispensáveis para os programas de melhoramento, pois auxiliam, por exemplo:

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

i) na estimação de parâmetros genéticos e predição genômica na presença de efeitos aditivos e de dominância (Sousa et al., 2022); (ii) na predição genômica na presença de efeitos de epistasia (Costa et al., 2022a); (iii) em estudos de adaptabilidade e estabilidade (Carneiro et al., 2018; Oda et al., 2021); (iv) em estudos de diversidade genética (Costa et al., 2022b); (v) no cálculo da importância de variáveis para a predição de uma variável alvo (Silva Júnior et al., 2021); (vi) na predição de variáveis categóricas (Montesinos-López et al., 2020); (vii) na detecção de doenças (Gajjar et al., 2022); (viii) na análise de imagens (Jiang; Li, 2020).

Neste capítulo serão apresentadas, de maneira sucinta, algumas metodologias que são embasadas em IA e AM. Especificamente, serão apresentadas as RNA, Lógica *Fuzzy*, as Árvores de Decisão, *Bagging* e *Random Forest*. Serão também apresentados três exemplos de aplicação, um que utiliza tais metodologias para a predição genômica em características que apresentam efeitos epistáticos e outro que faz uso de RNA para estimar os parâmetros genéticos e realizar a predição genômica na presença de efeitos aditivos e de dominância. Além disso, um exemplo no qual os mapas auto-organizáveis de Kohonen são utilizados na organização da diversidade genética também é apresentado. Finalmente, algumas perspectivas e considerações finais sob a ótica dos autores também serão apresentadas.

Metodologias

Inteligência Artificial

Redes Neurais Artificiais

A primeira arquitetura das Redes Neurais Artificiais foi proposta por McCulloch e Pitts (1943), mas essa rede se restringia a solucionar problemas de classificação do tipo binário. Posteriormente, Rosenblatt (1957) apresentou a topologia de uma RNA em que os neurônios são dispostos em uma única camada, conforme apresentado na Figura 1A e para essa rede se dá o nome de Perceptron.

O esquema da rede Perceptron se inicia com um conjunto de variáveis de entrada $(\chi_1, \chi_2, \dots, \chi_n)$ e, no interior do processo, essas variáveis são multiplicadas por pesos sinápticos $(w_0, w_1, w_2, \dots, w_n)$, conduzindo a uma quantidade denominada de porta do limiar, $w_0 + \sum_{i=1}^n w_i \chi_i$. Então, ao se aplicar uma função de ativação $g(\cdot)$ na porta do limiar, obtêm-se a variável de saída definida como $y_r = g(\sum_{i=0}^n w_i \chi_i)$, em que $\chi_0=1$. A função de ativação é escolhida de acordo com o problema a ser estudado. Alguns

exemplos são a função binária, que pode ser aplicada a problemas simples de classificação binária; a função linear, que pode ser aplicada a problemas simples em que as variáveis possuem associação linear; e as funções sigmóides e softmax, que podem ser aplicadas a problemas de associação não linear. Mas, atualmente, a função ReLU (*Rectified Linear Unit*) é uma função de ativação geral e que pode ser usada na maioria dos fenômenos. Essa função retorna 0 para todos os valores negativos, e o próprio valor para valores positivos. Na rede Perceptron, os pesos são ajustados sucessivamente para que o erro (ε), ou seja, a diferença entre a resposta desejada (y) e a saída de rede ($\varepsilon = y - y_r$), seja o menor possível.

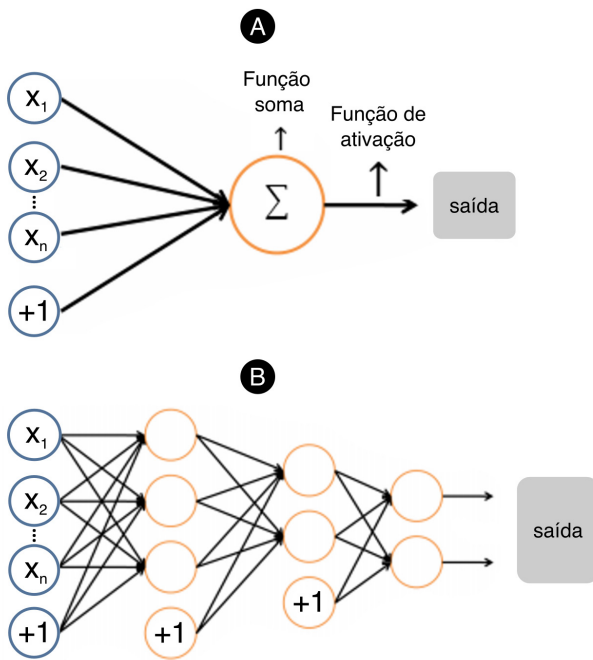


Figura 1. Arquiteturas de Redes Neurais Artificiais (RNA). Na parte A é mostrada a Representação de Perceptron com uma camada. Na parte B, o esquema de uma rede Perceptron Multicamadas.

Rosenblatt (1957) agregou à rede Perceptron uma regra de aprendizagem. No entanto, a mesma não poderia ser aplicada a problemas complexos, e essa topologia solucionaria apenas problemas linearmente separáveis. Posteriormente, com o intuito de se aplicar as redes neurais em problemas que não são linearmente separáveis, é que foi desenvolvida a rede Perceptron Multicamadas (PMN) apresentada na Figura 1B. Essa rede inclui camadas ocultas que tendem a capturar de modo mais eficiente a estrutura dos dados, seja ela linear ou não, e transmitir toda informação da camada de entrada para a camada de saída. O número de camadas ocultas e a quantidade de neurônios (unidades básicas das redes neurais artificiais) em cada camada podem ser definidos por meio de experimentação e do monitoramento das redes, uma vez que um grande número

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

de camadas e de neurônios pode ocasionar o *overfitting* (ou seja, a rede pode memorizar os dados de treinamento), e um insuficiente número de camadas e de neurônios pode não ser capaz de capturar todos os padrões presentes no conjunto de dados.

No desenvolvimento da rede Perceptron Multicamadas faz-se uso do algoritmo de treinamento *backpropagation*, proposto por Rumelhart et al. (1986). Esse algoritmo consiste em definir uma função de custo e um procedimento de otimização que permite a atualização dos pesos após cada iteração. O processo é dividido em duas fases, denominadas *forward* e *backward*. A fase *forward* consiste em receber as variáveis de entrada, definir os valores iniciais para os pesos e propagar a informação por meio das camadas ocultas até a saída de rede. Já a fase *backward* atualiza os pesos a partir da camada saída da rede em direção às variáveis de entrada, utilizando a técnica do gradiente descendente. Em termos matemáticos, para desenvolvimento desse algoritmo, as funções de ativação devem ser parcialmente ou totalmente diferenciáveis.

Lógica Fuzzy

A lógica *Fuzzy* constitui uma ferramenta para desenvolver cadeias de pensamento que, quando aplicadas ao contexto linguístico, permite reconhecer contradições e eliminar probabilidades de erro. No campo da matemática e da ciência da computação, o nome do matemático George Boole é referência quando se tem em mente o desenvolvimento de uma linguagem formal para o raciocínio. Baseado no sistema binário, que é extremamente simples, Boole demonstrou ser capaz de explorar todo o poder da lógica. No mundo moderno, essa forma de aplicar o raciocínio é tão importante que está presente no cotidiano de toda a humanidade, com diversas utilizações. A lógica booleana foi fundamental para o desenvolvimento da computação moderna. Atualmente, a maioria das máquinas modernas são alimentadas por algoritmos que descrevem raciocínios precisos e binários, como o proposto por Boole. Porém, elas são incapazes de retratar muitas situações reais para fins de tomada de decisão e, principalmente, incorporar no computador a forma de raciocínio humano que é, na maioria das vezes, impreciso.

A Lógica *Fuzzy* proposta por Zadeh (1965), também conhecida como lógica nebulosa, é a área da inteligência artificial inspirada na forma de raciocínio humano, que é baseado em aproximações e incertezas. Essa técnica provê um método de traduzir expressões verbais vagas, imprecisas e qualitativas, comuns na comunicação humana, em valores numéricos. Uma distinção básica da Lógica *Fuzzy* está relacionada a abordagens relativas à teoria de conjuntos, de forma que comparações entre a teoria de conjuntos clássicos e a teoria de conjuntos *Fuzzy* tornam-se fundamentais no entendimento dessa abordagem

para solução de problemas de predição, classificação e reconhecimento de padrões. Na teoria clássica de conjuntos é proposto valores de 0 ou 1 para determinar a pertinência de um elemento em um determinado conjunto, ao contrário da Lógica *Fuzzy*, que atribui graus de pertinência que variam entre 0 ou 1 à pertinência de um elemento em um conjunto qualquer. Assim, por exemplo, podemos classificar um grupo de linhagens ou híbridos em dois conjuntos relativos à altura cujos valores podem variar de 1,0 m a 2,0 m: conjunto A, de plantas altas, ou B de plantas baixas. Um determinado genótipo, na lógica clássica, será pertinente apenas a um determinado conjunto, por exemplo o A, com pertinência 1 e, conseqüentemente, com pertinência 0 ao conjunto B. Na Lógica *Fuzzy* pode-se admitir que esse mesmo genótipo, por exemplo com altura 1,7 m, possa ser alocado em ambos os conjuntos, com pertinência 0,7 em um conjunto e 0,3 em outro. Apesar de se ter uma informação vaga, pode-se tomar importantes decisões geradas por um raciocínio *Fuzzy*.

A área relativa à Lógica *Fuzzy* apresenta conceitos e regras que permitem a formalização do raciocínio nebuloso em sistemas computacionais, gerando o que se denomina de controladores, que funcionam como sistemas inteligentes utilizados na pesquisa e na ciência para tomadas de decisão. O controlador *Fuzzy* é um sistema no qual é possível combinar duas ou mais variáveis (variáveis de entrada) em uma única saída (variável resposta, geralmente discreta para problemas classificatórios, ou contínuas para problemas de predição), com a finalidade de facilitar o controle de uma situação que é realizada pelo ser humano. O controlador *Fuzzy* baseado em regras apresenta uma configuração básica, a qual é dividida em quatro etapas relativas ao estabelecimento da base de conhecimento, à interface de “fuzzificação”, aos sistemas de inferência fundamentado em regras e à interface de “defuzzificação”.

Os sistemas de inferência *Fuzzy* descritos na literatura são classificados em função das características gerais do seu modelo de tomada de decisão. Dentre esses, o sistema proposto por Mamdani (Mamdani; Assilian, 1975) está descrito na Figura 2.

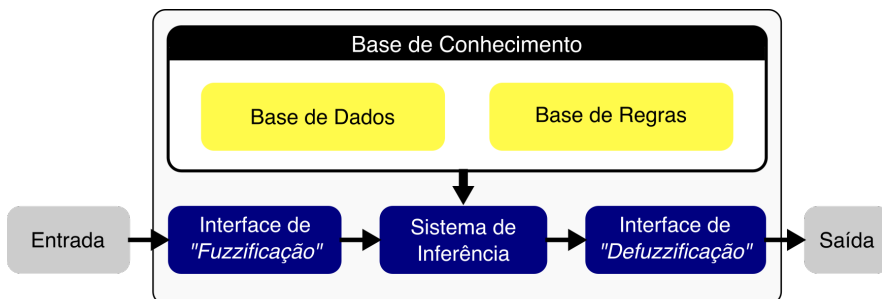


Figura 2. Esquema de um sistema de inferência *fuzzy*.

Fonte: Leite (2009).

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

A base de conhecimento é a primeira fase no projeto de um controlador *Fuzzy* e representa a configuração adotada que permitirá a tomada de decisão a respeito de determinada situação. Uma vez selecionadas as variáveis de entrada, que realmente influenciam na tomada de decisão, elas devem ser levadas em consideração sob o ponto de vista da Lógica *Fuzzy*. Assim, por exemplo, pode-se considerar a predição da característica rendimento de grãos a partir das variáveis de entrada, como altura e floração. Quando observadas sob o aspecto da Lógica *Fuzzy*, as variáveis são denominadas linguísticas, uma vez que os valores *Fuzzy* assumidos por essas variáveis são expressões verbais, representadas por conjuntos *Fuzzy* por meio de funções de pertinências. Assim, o controlador poderá, no exemplo, considerar para altura os conjuntos alto e baixo; para a variável de entrada floração, os conjuntos precoce, intermediário e tardio; e, para a variável resposta rendimento, os conjuntos selecionável ou descartável.

A segunda etapa é a “fuzzificação”. De forma geral, as variáveis de entrada e de saída de um controlador *Fuzzy* adotam valores numéricos, discretos ou contínuos. Assim, é necessária a sua conversão em valores *Fuzzy*, isto é, os valores numéricos das observações são convertidos para o intervalo $[0,1]$ e eles estão associados às pertinências em relação aos valores linguísticos de uma variável (Leite, 2009).

A terceira etapa é o estabelecimento de sistemas de inferência fundamentados em regras. A base de regras caracteriza os objetivos e a estratégia adotada para inferir sobre uma situação e permitir uma tomada de decisão adequada. As regras podem ser obtidas por meio de entrevistas com operadores humanos experientes, que detêm as informações necessárias para obter o melhor desempenho quanto a determinado processo. Assim, pode-se, por experiência, estabelecer as regras: se o genótipo é precoce OU baixo ele será descartável. Outra regra, por exemplo seria, se o genótipo é alto E tardio, ele será selecionável. Operadores de conjuntos tais como o OU e E são usados no estabelecimento de regras. Deve-se entender que as regras *Fuzzy* são responsáveis por traduzir a experiência de um especialista em determinado assunto, de forma a permitir que o computador, por meio das regras, consiga tomar decisões que são realizadas pelo ser humano.

Para que um controlador *Fuzzy* seja eficaz na tomada de decisão, é necessário que ele contenha um número de regras suficiente para contemplar todas as situações possíveis de ocorrência nas entradas. Dessa forma, pode ser necessário um grande número de regras, porém, em termos de processamento, a Lógica *Fuzzy* é muito eficiente devido ao fato de as regras serem ativadas em paralelo, o que contribui para maior velocidade de processamento dos dados.

Por fim, o controlador *Fuzzy* deve apresentar uma interface de “defuzzificação”, que consiste em um processo de transformação da área agregada de saída em um único valor que representa um compromisso entre os diferentes valores *Fuzzy* contidos na saída do controlador (Jang et al., 1997). A “defuzzificação” tem um impacto significativo no desempenho do controlador *Fuzzy*. Existem diversos métodos para a “defuzzificação”, mas o importante é escolher o método que melhor se adequa ao problema.

Os controladores *Fuzzy* são ferramentas que podem vir a contribuir tanto na seleção como na recomendação de cultivares, já que apresentam caráter multivariado comparativo. Esse tipo de ferramenta permite agregar a informação de diversas variáveis ou parâmetros em conjunto a fim de se ter um resultado único. Carneiro et al. (2018) apresentam a aplicação de controladores *Fuzzy* para auxiliar na interpretação das metodologias de adaptabilidade e estabilidade, bastante úteis na recomendação de cultivares obtidos de programas de melhoramento genético.

Aprendizado de Máquinas

Árvores de Decisão

De maneira sucinta, a técnica de aprendizado de máquinas, denotada por Árvores de Decisão, tem por objetivo construir um diagrama (árvore) que permita realizar classificações de uma variável dependente, Y (*output*), em função de uma ou mais variáveis independentes, X (*inputs*). Tal metodologia utiliza o raciocínio indutivo para elaborar, por meio de um algoritmo, a construção de regiões mutuamente exclusivas, (R_1, R_2, \dots, R_r) , que minimizam, caso a variável seja qualitativa (Árvore de Classificação), uma das três funções de custos, $C_r(A)$, apresentadas a seguir (James et al., 2013):

- Taxa de Erro Aparente: $TEA = 1 - \max_k (\hat{p}_{rk})$.
- Índice de Gini: $G = \sum_{k=1}^K \hat{p}_{rk}(1 - \hat{p}_{rk})$.
- *Deviance* (*Cross-Entropy*): $D = - \sum_{k=1}^K \hat{p}_{rk} \log \hat{p}_{rk}$,

em que \hat{p}_{rk} é a proporção estimada de observações na r -ésima região pertencentes a k -ésima classe. Em termos práticos, o uso de G ou D permite melhor avaliação da pureza do nó, ou seja, fornece uma ideia sobre a variabilidade das observações dentro dos nós. Especificamente, quanto mais próximo de zero mais “pura” (homogênea) será a região em questão. Por outro lado, a TEA é interessante quando o objetivo é realizar classificações com maior acurácia.

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

A construção de tais regiões mutuamente exclusivas, considerando todo o espaço das variáveis preditoras, é uma tarefa que exige elevado custo computacional. Dessa forma, torna-se inviável considerar cada partição possível do espaço em R regiões para se obter o menor valor de um dos três critérios descritos acima. Para contornar o custo computacional, a literatura propõe realizar um procedimento baseado em divisões binárias recursivas, na qual objetiva-se obter a variável X_p e o ponto t (dentro do conjunto de observações avaliadas) que particione o espaço em duas regiões:

$$R_1(p,t) = \{X|Xp \leq t\} \text{ e } R_2(p,s) = \{X|Xp < t\}$$

tal que o ponto t divida a p -ésima variável em duas regiões que obtenha o menor valor de um dos três critérios apresentados. Por fim, utiliza-se a variável que obteve o menor valor para o critério selecionado para a primeira divisão, em seguida, o processo é repetido para cada região gerada.

Uma questão interessante e que deve ser levada em conta é o tamanho da árvore construída. Enquanto uma árvore muito grande pode se superajustar (*overfitting*) aos dados, uma pequena pode não capturar a explicação do fenômeno de maneira adequada. Para obter uma solução para esse problema, pode-se construir uma árvore até que nenhuma região contenha mais que 5 indivíduos e em seguida podá-la (*pruning*) usando o que se denomina “custo complexidade da poda” (Hastie et al., 2009). Especificamente, tal problema é apresentado como a minimização da seguinte função custo:

$$P_\alpha(A) = \sum_{r=1}^{|A|} C_r(A) + \alpha |A|,$$

em que $C_r(A)$ é uma das três funções custo definidas anteriormente, $|A|$ é o número de nós terminais da árvore A e α é o parâmetro de penalização (*tunning parameter*), o qual pode ser escolhido por meio de um procedimento de validação cruzada.

Toda a apresentação acima pode ser expandida para a construção de Árvores de Regressão. Nesse caso, a variável dependente, Y (*output*), é quantitativa e a função custo a ser minimizada é dada pela Soma de Quadrado dos Resíduos (SQR), ou seja:

$$SQR = \sum_{j=1}^r \sum_{i \in R_j} (y_i - \hat{y}_{R_j})^2,$$

em que y_i é i -ésimo valor observado da variável quantitativa dependente Y e \hat{Y}_{R_j} é a média das observações pertencentes a região R_j .

Bootstrap Aggregating (Bagging)

O *Bootstrap* é um método de Estatística Computacional que visa acessar a incerteza sobre um estimador por meio da amostragem, com reposição, do conjunto de dados. Em termos de Árvores de Decisão, para cada amostra *bootstrap* constrói-se uma árvore e, posteriormente, considera-se como valor predito a média obtida por todas as árvores construídas, ou seja, aquela obtida por meio das N amostras *bootstrap*:

$$\hat{f}_{\text{médio}}(x) = \frac{1}{N} \sum_{n=1}^N \hat{f}^n(x)$$

em que $\hat{f}^n(x)$ é o valor predito pela n -ésima árvore construída.

Tal abordagem, sob o ponto de vista estatístico, aumenta a acurácia do valor predito, visto que a variância da média é dada pela variância original das observações dividida pela raiz quadrada do número de elementos (amostras). Uma característica interessante do método *bootstrap* é que, em média, 1/3 das observações não são utilizadas em cada árvore construída. Dessa forma, é possível avaliar a capacidade preditiva do modelo considerando esses indivíduos. Tal conjunto é denotado na literatura como *out-of-bag* (OOB) (Hastie et al., 2009).

Random Forest

No processo de construção das árvores por meio do *Bagging* todas as variáveis explicativas (*inputs*) são consideradas. Dessa forma, devido ao algoritmo utilizado ser baseado na minimização de uma função de custo, espera-se que tanto as variáveis que compõem o diagrama em árvore quanto a sua ordem de apresentação no mesmo sejam semelhantes. Tal semelhança conduz a resultados similares, ou seja, as árvores construídas apresentam dependência. Visando reduzir/eliminar tal dependência e, conseqüentemente, aumentar a acurácia do processo preditivo, Ho (1995) propôs o *Random Forest* (RF). Especificamente, essa abordagem é uma pequena modificação do *Bagging*, na qual um número inferior de variáveis explicativas (*inputs*, $m < p$) são utilizadas na construção das árvores obtendo valores preditos mais independentes e ocasionando redução da variabilidade encontrada nas Árvores de Decisão. Hastie et al. (2009) sugerem que o número de variáveis predictoras utilizadas em cada partição seja dada da seguinte forma, para árvores de classificação $m = \sqrt{p}$ e, para árvores de regressão, $m = p/3$. Além das vantagens citadas anteriormente, em termos do aumento da acurácia, a construção de árvores de decisão por meio do *Bagging* e *Random Forest* permite ao usuário verificar a importância de variáveis explicativas por meio do decréscimo nas funções de custo, quando se retira a variável do ajuste.

Exemplos de Aplicação

Em termos práticos, os programas de melhoramento genético, seja de plantas ou animais, demandam experimentação intensiva para que valores genéticos acurados possam ser obtidos e utilizados como critérios de seleção. Esse tem sido o maior desafio para o melhorista, de reconhecer entre os genótipos avaliados aqueles que terão melhor desempenho per se ou que proporcionarão melhores descendentes em gerações segregantes (Nascimento et al., 2022). Visando elevar a acurácia do processo de seleção e reduzir ciclos de gerações, a seleção genômica (SG) tem sido utilizada (Liu et al., 2019; Singh et al., 2019). Tal metodologia, proposta por Meuwissen et al. (2001), permite incorporar informações moleculares diretamente na predição do mérito genético dos indivíduos. As metodologias tradicionais de SG (G-BLUP e Métodos Bayesianos) baseiam-se na suposição de normalidade dos valores fenotípicos ou dos efeitos de marcadores moleculares. Para superar essa limitação, Pérez e De Los Campos (2014) propuseram o uso da Regressão Linear Generalizada Bayesiana (BGLR), permitindo o uso de SG para modelos contínuos e discretos. Embora útil, a presença de fatores complicadores, como epistasia e dominância, dificultam a utilização dos modelos usuais de SG, uma vez que seus efeitos devem ser estabelecidos *a priori* no modelo (Sousa et al., 2021). Modelos baseados em IA (PMC) ou AM (Árvores de Decisão e seus possíveis refinamentos – *Bagging* e *Random Forest*) se apresentam como alternativas. Tais algoritmos não fazem suposições sobre o modelo e permitem, a partir dos dados, a captura de fatores complexos, como epistasia e dominância em modelos de previsão. Além disso, eles não requerem suposições sobre a distribuição dos valores fenotípicos.

Predição genômica por meio de métodos de IA e AM considerando características com diferentes níveis de complexidade

Visando apresentar um exemplo no qual o objetivo foi avaliar e comparar a performance preditiva de métodos baseados em IA, AM e estatísticos (RR-BLUP e Bayes B) na predição genômica, serão apresentados resultados parciais do estudo desenvolvido por Barbosa et al. (2021). Para tal avaliação, foram simuladas populações com diferentes níveis de herdabilidade e números de QTL (*Quantitative Trait Loci*) na presença de dominância e efeitos epistáticos. Foi simulado um genoma de uma população F2 formada por 1.000 indivíduos, genotipados para 2.010 marcadores de polimorfismo de nucleotídeo único (SNP). Vinte e seis características foram simuladas considerando números de QTL variando de 2 a 88, distribuídos igualmente em oito grupos de ligação, e com herdabilidade 0,3 e 0,6 (Tabela 1).

Tabela 1. Número de QTL (*Quantitative Trait Loci*) distribuídos em oito grupos de ligação com dois níveis de herdabilidade para as 26 características simuladas (T1 até T26).

h ²	QTL												
	2	4	8	16	24	32	40	48	56	64	72	80	88
0,3	T1	T3	T5	T7	T9	T11	T13	T15	T17	T19	T21	T23	T25
0,6	T2	T4	T6	T8	T10	T12	T14	T16	T18	T20	T22	T24	T26

As características foram simuladas considerando grau médio de dominância igual a 0,5 e um modelo de efeito epistático como apresentado a seguir:

$$Y_i = \mu + \sum_{j=1}^n \alpha_j + \sum_{j=1}^{n-1} \alpha_j \alpha_{j+1} + e_i$$

em que Y_i é o valor fenotípico para a observação i ; μ é a média geral da característica; α_j assume os valores $\mu + \alpha_j$, $\mu + d_j$ e $\mu - \alpha_j$ para os genótipos associados às classes AA, Aa e aa, respectivamente. As classes foram codificadas como 1, 0 ou -1, respectivamente. Na equação acima, a primeira soma da expressão refere-se à contribuição individual do marcador por meio de seus efeitos aditivos e dominantes, e a segunda soma representa os efeitos multiplicativos correspondentes às interações epistáticas entre pares de locos, em que $\alpha_j \alpha_{j+1}$ é o efeito multiplicativo do alelo favorável no locus j e $j+1$ e n varia de 2 a 88. A ação gênica teve distribuição uniforme, de modo que cada loco j teve a mesma contribuição para a característica em consideração. O e_i é o efeito ambiental, gerado de acordo com a estrutura de variância dos resíduos dada por $e_i \sim N(0, V_e)$ com $V_e = (1 - h^2)V_g/h^2$.

Como resultado, foi possível observar que o desempenho das metodologias avaliadas é influenciado pela complexidade das características analisadas, e, dessa forma, diferentes resultados podem ser encontrados devido a essa complexidade (Figura 3).

Os métodos de AM mostraram ser ferramentas alternativas interessantes para prever os valores genéticos genômicos com controle gênico epistático em características com diferentes graus de herdabilidade e diferentes números de genes controladores. Especificamente, as árvores de decisão e o *Random Forest* apresentaram valores de acurácia preditiva superiores as demais metodologias para características com efeitos não aditivos nos cenários de 2 a 16 QTL (Figuras 3A e 3B). Já as PMC e os métodos de AM apresentaram resultados semelhantes

aos obtidos por meio de métodos tradicionais (RR-BLUP e Bayes B) na predição de valores genéticos genômicos para os cenários de 16 até 88 QTL (Figuras 3A e 3B).

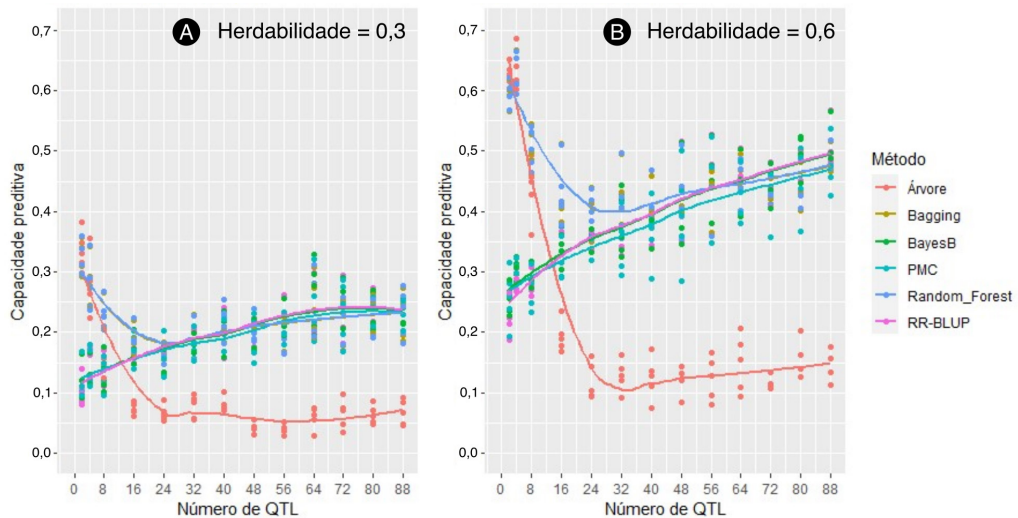


Figura 3. Capacidade preditiva estimada pelas Árvore de Decisão (Árvore), *Bagging*, Bayes B, Redes Perceptron Multicamadas (PMC) e RR-BLUP em cenários com números de QTL variando de 2 a 88 com efeitos aditivos, dominantes e epistáticos e herdabilidade 0,3 (a) e 0,6 (b).

De maneira geral, esses resultados indicam que para características com menor número de QTL, os efeitos multiplicativos dos genes controladores (epistasia) podem ser mais importantes, visto que o efeito individual de cada gene é maior do que nas características controladas por um maior número de QTL. Assim, as metodologias de AM são capazes de capturar esses efeitos epistáticos em características controladas por poucos QTL e apresentaram melhores resultados quando comparados com os métodos tradicionais de SG (Figura 3). Adicionalmente, tais metodologias não requerem especificação da arquitetura da característica e, portanto, uma primeira hipótese seria que esses métodos poderiam naturalmente capturar os efeitos não aditivos conhecidos no genoma simulado que não foram especificados nos métodos RR-BLUP e Bayes B. Por outro lado, o aumento no número de QTL faz com que o efeito de cada um deles seja reduzido, eliminando possíveis vantagens de métodos que capturam os efeitos não aditivos na predição do valor genômico.

Efeitos de marcadores e estimativas de herdabilidade via redes neurais artificiais

Para exemplificar e apresentar resultados sobre a obtenção dos efeitos dos marcadores e estimativas de herdabilidade por meio de redes neurais artificiais, serão apresentados resultados parciais do estudo desenvolvido por Sousa et al. (2022). Nesse estudo, o objetivo principal foi explorar os efeitos dos marcadores do tipo SNP e estimar a herdabilidade proveniente do modelo genômico aditivo-dominante por meio de PMC. Além disso, os resultados foram confrontados com aqueles provenientes do método estatístico mais usual em SG, o G-BLUP.

Antes de apresentar os resultados, vale a pena ressaltar a grande flexibilidade para o ajuste de efeitos não aditivos e a não necessidade de pressupor qualquer relação funcional entre o *output* e *inputs a priori* nas técnicas de IA (ex.: PMC) e AM (ex.: AD). Tais características são interessantes para estudos de SG, principalmente na falta de normalidade dos valores fenotípicos e na presença de efeitos não aditivos no modelo. Apesar dessa vantagem, informações sobre como acessar os efeitos de marcadores e estimar parâmetros genéticos, como por exemplo herdabilidade, são limitadas. Glória et al. (2016), utilizando dados simulados, obtiveram valores de herdabilidade, os quais foram estimados considerando a importância relativa dos marcadores ajustados por uma RNA e conceitos de genética quantitativa (Olden et al., 2004). Apesar de interessante, a melhor RNA desse estudo apresentou apenas uma única camada oculta com um único neurônio e a função de ativação identidade, ou seja, um modelo semelhante ao utilizado nos métodos tradicionais de SG. Diante desses resultados, algumas questões podem ser apresentadas: i) é possível obter a importância de marcadores na predição de características alvo usando RNA com arquiteturas mais complexas, como por exemplo, RNA com 3 camadas ocultas, diferentes números de neurônios e funções de ativação; e ii) é possível, a partir de tais informações, estimar valores de herdabilidade da parte aditiva e dominante de uma característica? Visando responder tais questões, Sousa et al. (2022), utilizando dados provenientes de um programa de melhoramento de café canéfora, ajustaram modelos de redes neurais do tipo PMC, obtiveram os efeitos dos marcadores do tipo SNP, estimaram valores de herdabilidade provenientes do modelo genômico aditivo dominante e, finalmente, compararam os resultados obtidos com aqueles provenientes do modelo G-BLUP. Especificamente, foram avaliadas 165 plantas genotipadas para 14.387 marcadores do tipo SNP. Nesse estudo, foram consideradas as características produtividade e resistência à ferrugem.

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

Devido ao grande número de marcadores e a grande demanda computacional requerida por uma RNA, foi realizada uma seleção de *inputs* (marcadores) para cada uma das características por meio do *Bagging*. Assim, o treinamento da rede foi realizado considerando 1.086 e 1.302 marcadores para, respectivamente, produtividade e resistência à ferrugem. Para produtividade, a melhor RNA (capacidade preditiva = 0,72) apresentava 4, 15 e 33 neurônios para a primeira, segunda e terceira camadas ocultas, respectivamente. Para resistência à ferrugem, a melhor RNA (capacidade preditiva = 0,88) possuía arquitetura com 13, 20 e 24 neurônios para a primeira, segunda e terceira camadas ocultas, respectivamente. Para todos os neurônios foi utilizada a função de ativação logística.

Para ambas as características, produtividade e resistência à ferrugem, a herdabilidade aditiva e a de dominância capturadas pela RNA foram semelhantes às obtidas pelo G-BLUP (Tabela 2). Além disso, as estimativas foram consistentes com as relatadas na literatura. Para a característica produtividade, a literatura apresenta estimativas de herdabilidade variando de 0,15 até 0,79. Já para a resistência à ferrugem, a estimativa foi próxima à relatada por Alkimim et al. (2020).

Tabela 2. Estimativas de herdabilidade aditiva e de dominância.

	Produtividade		Resistência à Ferrugem	
	RNA	G-BLUP	RNA	G-BLUP
h_a^2	0,25	0,26	0,67	0,55
h_d^2	0,06	0,05	0,30	0,22

G-BLUP, Genomic Best Linear Unbiased Predictor; h_a^2 , herdabilidade aditiva; h_d^2 , herdabilidade devido à dominância.

Organização da diversidade genética por meio de inteligência computacional

Métodos de inteligência artificial, como por exemplo, os mapas auto-organizáveis de Kohonen (Kohonen, 1990), também podem ser utilizados em estudos para organização da diversidade genética. Para exemplificar e apresentar um estudo com esse objetivo, o qual fez uso de dados fenotípicos, serão apresentados parte dos resultados de um trabalho desenvolvido por Costa et al. (2022b). Nesse estudo, o objetivo principal foi organizar a diversidade genética de cultivares de soja [*Glycine max* (L.) Merr.] lançadas no Brasil de 1998 a 2017 e, além disso, comparar a similaridade entre os mantenedores dessas cultivares com base nas informações fenotípicas.

A partir do mapa de Kohonen, em que são apresentadas as distâncias entre os neurônios vizinhos (Figura 4), observou-se que, em geral, os neurônios apresentam dissimilaridade entre eles. Já os neurônios 10, 14 e 15, apesar de apresentarem menores distâncias entre eles, apresentaram grande diversidade entre as características avaliadas. Dessa forma, cada neurônio foi considerado como um grupo.

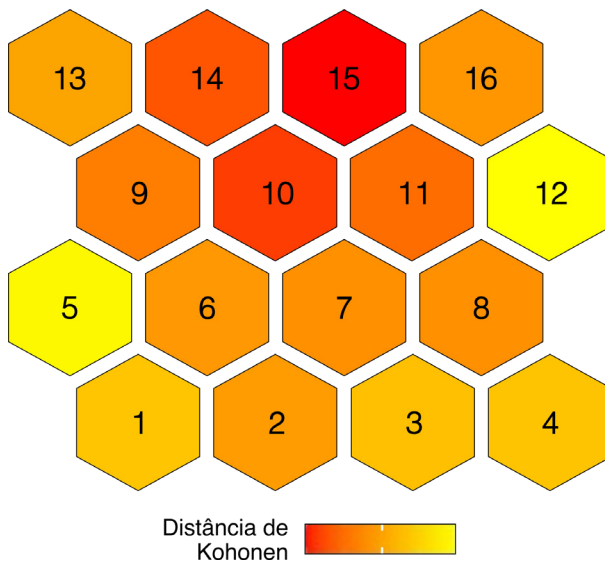


Figura 4. Distância entre neurônios vizinhos da rede de mapas auto-organizados de Kohonen. Os números centrais referem-se ao número do neurônio. O gradiente de cores refere-se à distância entre os neurônios, sendo os vermelhos os mais próximos e, os mais distantes, os amarelos.

Fonte: Costa et al. (2022b).

A organização entre mantenedores de cultivares de soja lançadas no Brasil, entre 1998 e 2017 mostrou que a diversidade genética de cultivares de soja é alta. Em termos práticos, de acordo com a Figura 5, observa-se que mantenedores de cultivares de soja em neurônios mais distantes apresentam mais dissimilaridade genética entre os cultivares lançadas pelos respectivos mantenedores. Assim, é possível inferir o perfil de cada mantenedor e sua base populacional genética para o desenvolvimento de cultivares. Especificamente, foi possível verificar que entre os mantenedores, o Soymax apresentou a base genética mais restrita. Já a Embrapa, que é uma instituição pública, apresentou grande diversidade genética em sua base populacional e foi a que mais se aproximou das empresas privadas em similaridade genética. Já as Universidades Federais de Viçosa e Uberlândia desenvolveram cultivares com alta similaridade.

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

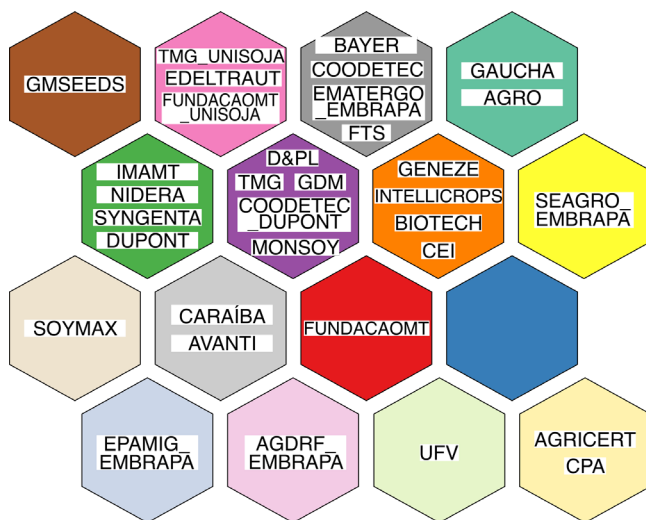


Figura 5. Organização de mantenedores de cultivares de soja lançadas no Brasil de 1998 a 2017. O padrão de similaridade e organização de variação estabelecido pelo mapa auto-organizado de Kohonen em configuração hexagonal com 16 neurônios para 36 mantenedores.

Fonte: Costa et al. (2022b).

Cenário atual da Inteligência Artificial

No contexto do rápido avanço tecnológico dos últimos anos, as ferramentas de Inteligência Artificial (IA) têm ganhado destaque como impulsionadoras de uma revolução digital. Com o progresso dos algoritmos de Aprendizado de Máquina, como as redes neurais profundas, alcançando patamares surpreendentes, a IA está se estabelecendo em diversos campos. Especificamente no processamento de linguagem natural, está se presenciando uma verdadeira transformação. Com a ascensão dos modelos de linguagem pré-treinados, como, por exemplo, o renomado GPT-3 (Dale, 2021), a IA é capaz de gerar um texto coerente e responder a perguntas com base em conhecimento prévio. Isso representa um avanço notável na compreensão e comunicação entre seres humanos e máquinas.

A sigla GPT significa *Generative Pre-trained Transformer* (transformador pré-treinado generativo, em tradução livre). O termo *transformer* refere-se a uma arquitetura de rede neural reconhecida por sua eficiência no processamento de sequências de dados, como palavras em um texto. O GPT é pré-treinado, o que significa que é treinado em grandes conjuntos de dados para aprender a entender a estrutura e as características da linguagem.

O ChatGPT, desenvolvido pela OpenAI (<https://chat.openai.com/>), atualmente utilizando a tecnologia GPT-3, tornou-se uma referência em inteligência artificial

(IA) ao reproduzir respostas semelhantes à escrita humana. Esse avanço despertou preocupações entre especialistas sobre os riscos futuros da tecnologia. Como resposta, outras empresas de tecnologia começaram a desenvolver softwares baseados no ChatGPT, com características distintas em relação ao chatbot da OpenAI. Um exemplo é o SenseChat, da China, que promete oferecer soluções de IA generativa no ambiente corporativo, intensificando a competição entre organizações na busca por eficiência e oportunidades proporcionadas por essa nova tecnologia. Essa competição contribui para a disputa entre empresas americanas e chinesas pelo domínio das tecnologias digitais de ponta em escala, impulsionando o mercado de IA e promovendo o aprimoramento contínuo dos modelos e tecnologias, especialmente em regiões de forte desenvolvimento tecnológico, como a Ásia.

Na visão computacional, outro escopo da IA hipermoderna, ferramentas extraordinárias vêm sendo lançadas. O reconhecimento de objetos, detecção facial, segmentação de imagens e a condução autônoma de veículos são exemplos de desafios enfrentados e superados com maestria (Fujiyoshi et al., 2019; Kumar et al., 2019; Qi et al., 2021). As máquinas agora enxergam o mundo com uma precisão impressionante. A aplicação da IA se estende a uma infinidade de setores. Na saúde, ela assume papel fundamental na análise de dados em grande escala, permitindo diagnósticos mais precisos e descobertas médicas inovadoras (Dias; TorkamanI, 2019). No setor financeiro, a IA é certeira na detecção de fraudes e no gerenciamento inteligente de investimentos (Bahrammirzaee, 2010). A indústria de manufatura também se beneficia dos avanços da IA, com processos automatizados e otimizados. No melhoramento, a IA já é usada na predição de caracteres de difícil mensuração, seja usando *Near Infra-Red Spectroscopy* (NIRS) (Kim, 2023), seja usando caracteres mais acessíveis para mensurar caracteres complexos ou caros (Castro et al., 2017).

A criação de figuras, imagens e até fotografias hiperrealistas por meio da inteligência artificial (IA) é um campo de pesquisa e desenvolvimento que utiliza algoritmos de aprendizado profundo, como redes neurais convolucionais (CNNs) e *generative adversarial networks* (GANs). Essas técnicas permitem que os sistemas de IA aprendam a partir de grandes conjuntos de dados visuais e reproduzam características complexas, como texturas, cores e detalhes de forma altamente realista. A Figura 6 foi criada a partir de IA, usando a plataforma *playgroundai* (<https://playgroundai.com/create>), alimentada com o texto "*plant breeding in the future*". À primeira vista, parece uma imagem perfeita, como uma fotografia obtida em um local real. No entanto, olhando com mais atenção, nota-se algumas distorções, especialmente nos textos e na tela do equipamento. Ao utilizar camadas convolucionais para capturar padrões

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

visuais e GANs para gerar amostras que se aproximem da distribuição dos dados de treinamento, a IA é capaz de produzir figuras e imagens que “enganam” o observador, imitando a aparência de fotografias reais. Esses avanços têm aplicações significativas em áreas como design gráfico, indústria cinematográfica, simulações virtuais e, até mesmo, em aplicações de arte digital, impulsionando uma nova fronteira criativa e desafiando os limites da percepção visual humana.



Figura 6. Uma imagem 100% criada por IA a partir do site <https://playgroundai.com/create>. Foi utilizado o texto “Plant breeding in the future” para geração dessa imagem.

Os avanços da IA, embora impressionantes, também apresentam desafios importantes a serem considerados. A transparência e a ética no uso da IA são questões fundamentais. À medida que os sistemas se tornam mais complexos, é

essencial garantir que as decisões tomadas pelas máquinas sejam compreensíveis e justificáveis. Além disso, preocupações relacionadas à privacidade e à segurança precisam ser abordadas com medidas sólidas de proteção. A compreensão dos modelos de IA e sua capacidade de raciocínio também estão em constante evolução. Os pesquisadores estão buscando um conhecimento mais profundo sobre como essas máquinas chegam a suas respostas e como podem explicar suas decisões de maneira acessível aos seres humanos.

Neste contexto fascinante, as ferramentas de IA estão em um estágio considerado avançado, mas é importante ressaltar que caso o leitor esteja lendo este capítulo em um ano posterior a 2023, é provável que as tecnologias atuais já tenham avançado significativamente. A expectativa é de que a IA continue evoluindo de forma exponencial, trazendo ainda mais transformações em nossa interação com o mundo. No entanto, é essencial permanecer vigilante para garantir que esses avanços sejam conduzidos de maneira responsável e em benefício da humanidade.

Considerações Finais

Programas de melhoramento genético demandam experimentação intensiva para que valores genéticos acurados possam ser obtidos e utilizados como critérios de seleção. Esse tem sido o maior desafio para o melhorista, o de reconhecer entre os genótipos avaliados aqueles que terão melhor desempenho per se ou proporcionarão melhores descendentes em gerações segregantes. Mais recentemente, o melhoramento genético tem agregado informações de grande valor para o estabelecimento de critérios de seleção. Informações massivas de marcadores moleculares, de espectroscopias, como do infravermelho próximo (NIRS), e de imagens têm permitido ajustar modelos bem mais parametrizados, com ganhos substanciais em medidas de qualidade de ajuste e de predição. Ademais, começa-se a entender que informações mais amplas do ambiente e de suas interações, do indivíduo e de seus aparentados são necessárias e requerem processamento apropriado.

Entretanto, o uso de modelagens com conjuntos de dados de grande dimensão demanda conhecimento de abordagens eficazes e, para tal propósito, alguns princípios básicos de predição por meio de paradigmas estatísticos, representados pelos modelos de regressão, e de inteligência computacional, representados pelo Perceptron Multicamadas e Redes de Base Radial, além de técnicas de Aprendizado de Máquinas, representadas pelas Árvores de Decisão, *Bagging*, *Boosting* bem

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

como teorias lógicas *Fuzzy*, precisam ser difundidos e utilizados rotineiramente pelos pesquisadores da área.

Assim, tem-se a perspectiva de que futuros profissionais agreguem aos seus conhecimentos básicos sobre o genótipo e o ambiente a incorporação de abordagens biométricas que promovam interpretações mais holísticas, como as preconizadas pelas técnicas de aprendizado de máquina e de inteligência computacional.

Em termos de avanços tecnológicos, o uso de tais abordagens estará, em breve, disponível para os melhoristas em equipamentos e aplicativos que permitam, por exemplo, avaliar severidade de doenças, produtividade e mensurar diversas características auxiliares que anteriormente podiam ser onerosas e, conseqüentemente, auxiliá-los na tomada de decisão em programas de melhoramento. Tais metodologias já estão sendo utilizadas para melhor entendimento do desempenho dos genótipos no ambiente e de suas interações, como na predição do desempenho de híbridos de milho em diferentes locais e condições.

Neste capítulo, foram apresentadas algumas das principais metodologias embasadas em Inteligência Artificial (RNA e lógica *fuzzy*) e Aprendizado de Máquinas (Árvores de Decisão, *Bagging* e *Random Forest*). Além disso, foram apresentados três exemplos de aplicação. De maneira geral, ficou evidenciado que tais metodologias vêm sendo utilizadas com sucesso na obtenção de soluções para o melhoramento genético. Especificamente, foram apresentadas soluções para problemas de predição genômica, nos quais informações de marcadores moleculares são adicionadas ao processo seletivo de genótipos e na organização da diversidade genética com base em valores fenotípicos. Diante das potencialidades das técnicas, os autores acreditam que os futuros profissionais devem, além de apresentar conhecimento em métodos rotineiramente utilizados no melhoramento, agregar aos seus conhecimentos metodologias que se baseiam em Inteligência Artificial e Aprendizado de Máquinas.

Referências

ALKIMIM, E. R.; CAIXETA, E. T.; SOUSA, T. V.; RESENDE, M. D. V.; SILVA, F. L.; SAKIYAMA, N. S.; ZAMBOLIM, L. Selective efficiency of genome-wide selection in *Coffea canephora* breeding. **Tree Genetics and Genomes**, v. 16, 41, p.1–11, Apr. 2020. DOI: <https://doi.org/10.1007/s11295-020-01433-3>.

BAHRAMMIRZAEI, A. A comparative survey of artificial intelligence applications in finance: artificial neural networks, expert system and hybrid intelligent systems. **Neural Computing and Applications**, v. 19, n. 8, p. 1165-1195, Nov. 2010. DOI: <https://doi.org/10.1007/s00521-010-0362-z>.

- BARBOSA, I. P.; SILVA, M. J.; COSTA, W. G.; SANT'ANNA, I. C.; NASCIMENTO, M.; CRUZ, C. D. Genome enabled prediction through machine learning methods considering different levels of trait complexity. **Crop Science**, v. 61, n. 3, p. 1890-1902, May 2021. DOI: <https://doi.org/10.1002/csc2.20488>.
- CARNEIRO, V. Q.; PRADO, A. L.; CRUZ, C. D.; CARNEIRO, P. C. S.; NASCIMENTO, M.; CARNEIRO, J. E. S. Fuzzy control systems for decision-making in cultivars recommendation. **Acta Scientiarum. Agronomy**, v. 40, e39314, 2018. DOI: <https://doi.org/10.4025/actasciagron.v40i1.39314>.
- CASTRO, C. A. O.; RESENDE, R. T.; KUKI, K. N.; CARNEIRO, V. Q.; MARCATTI, G. E.; CRUZ, C. D.; MOTOIKE, S. Y. High-performance prediction of macauba fruit biomass for agricultural and industrial purposes using Artificial Neural Networks. **Industrial Crops and Products**, v. 108, p. 806-813, Dec. 2017. DOI: <https://doi.org/10.1016/j.indcrop.2017.07.031>.
- COSTA, W. G.; CELERI, M. O.; BARBOSA, I. P.; SILVA, G. N.; AZEVEDO, C. F.; BOREM, A.; NASCIMENTO, M.; CRUZ, C. D. Genomic prediction through machine learning and neural networks for traits with epistasis. **Computational and Structural Biotechnology Journal**, v. 20, p. 5490-5499, 2022a. DOI: <https://doi.org/10.1016/j.csbj.2022.09.029>.
- COSTA, W. G.; VALADARES, C. B.; SILVA JUNIOR, A. C.; ARCANJO, E. S.; ALMEIDA, H. C. F.; CRUZ, C. D.; CARVALHO, I. R.; SILVA, J. A. G.; AZEVEDO, C. F.; NASCIMENTO, M.; NASCIMENTO, A. C. C. Genetic diversity and interaction between the maintainers of commercial soybean cultivars using self-organizing maps. **Crop Science**, v. 62, n. 6, p. 2233-2245, Nov./Dec. 2022b. DOI: <https://doi.org/10.1002/csc2.20816>.
- CRUZ, C. D.; NASCIMENTO, M. **Inteligência computacional aplicada ao melhoramento genético**. Viçosa, MG: Ed. UFV, 2018. 414 p.
- DALE, R. GPT-3: what's it good for? **Natural Language Engineering**, v. 27, n. 1, p. 113-118, Jan. 2021. DOI: <https://doi.org/10.1017/S1351324920000601>.
- DIAS, R.; TORKAMANI, A. Artificial intelligence in clinical and genomic diagnostics. **Genome Medicine**, v. 11, 70, Nov. 2019. DOI: <https://doi.org/10.1186/s13073-019-0689-8>.
- FUJIYOSHI, H.; HIRAKAWA, T.; YAMASHITA, T. Deep learning-based image recognition for autonomous driving. **IATSS Research**, v. 43, n. 4, p. 244-252, Dec. 2019. DOI: <https://doi.org/10.1016/j.iatssr.2019.11.008>.
- GAJJAR, R.; GAJJAR, N.; THAKOR, V. J.; PATEL, N. P.; RUPARELIA, S. Real-time detection and identification of plant leaf diseases using convolutional neural networks on an embedded platform. **The Visual Computer**, v. 38, p. 2923-2938, Aug. 2022. DOI: <https://doi.org/10.1007/s00371-021-02164-9>.
- GLÓRIA, L. S.; CRUZ, C. D.; VIEIRA, R. A. M.; RESENDE, M. D. V.; LOPES, P. S.; SIQUEIRA, O. H. G. B. D.; SILVA, F. F. Assessing marker effects and heritability estimates from genome prediction by Bayesian regularized neural networks. **Livestock Science**, v. 191, p. 91-96, Sept. 2016. DOI: <https://doi.org/10.1016/j.livsci.2016.07.015>.
- HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The elements of statistical learning: data mining, inference, and prediction**. 2nd ed. New York: Springer, 2009. 745 p.
- HO, T. K. Random decision forest. In: INTERNATIONAL CONFERENCE ON DOCUMENT ANALYSIS AND RECOGNITION, 3., 1995, Montreal. **Proceedings**. Montreal: IEEE, 1995. p. 278-282. DOI: <https://doi.org/10.1109/ICDAR.1995.601943>.
- JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. **An introduction to statistical learning with applications in R**. New York: Springer, 2013. 426 p.
- JANG, J.-S. R.; SUN, C.-T.; MIZUTANI, E. **Neuro-fuzzy and soft computing: a computational approach to learning and machine intelligence**. New Delhi: Pearson, 1997. 614 p.

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

- JIANG, Y.; LI, C. Convolutional neural networks for image-based high throughput plant phenotyping: a review. **Plant Phenomics**, v. 2020, 4152816, Apr. 2020. DOI: <https://doi.org/10.34133/2020/4152816>.
- KIM, S.-Y.; HONG, S.-J.; KIM, E.; LEE, C.-H.; KIM, G. Application of ensemble neural-network method to integrated sugar content prediction model for citrus fruit using Vis/NIR spectroscopy. **Journal of Food Engineering**, v. 338, 111254, Feb. 2023. DOI: <https://doi.org/10.1016/j.foodeng.2022.111254>.
- KOHONEN, T. The self-organizing map. **Proceedings of the IEEE**, n. 78, v. 9, p. 1464–1480, Sept. 1990. DOI: <https://doi.org/10.1109/5.58325>.
- KUMAR, A.; KAUR, A.; KUMAR, M. Face detection techniques: a review. **Artificial Intelligence Review**, v. 52, p. 927–948, Aug. 2019. DOI: <https://doi.org/10.1007/s10462-018-9650-2>.
- LEITE, L. C. M. **Geração e simplificação da base de conhecimento de um sistema híbrido fuzzy-genético**. 2009. 110 f. Dissertação (Mestrado em Engenharia Eletrônica) – Universidade do Estado do Rio de Janeiro, Rio de Janeiro.
- LIU, X.; WANG, H.; HU, X.; LI, K.; LIU, Z.; WU, Y.; HUANG, C. Improving genomic selection with quantitative Trait Loci and nonadditive effects revealed by empirical evidence in maize. **Frontiers in Plant Science**, v. 10, 1129, Sept. 2019. DOI: <https://doi.org/10.3389/fpls.2019.01129>.
- MAMDANI, E. H.; ASSILIAN, S. An experiment in linguistic synthesis with a fuzzy logic controller. **International Journal of Man-Machine Studies**, v. 7, n. 1, p. 1–13, Jan. 1975. DOI: [https://doi.org/10.1016/S0020-7373\(75\)80002-2](https://doi.org/10.1016/S0020-7373(75)80002-2).
- MCCULLOCH, W. S.; PITTS, W. H. A logical calculus of the ideas immanent in nervous activity. **The Bulletin of Mathematical Biophysics**, v. 5, p. 115–133, Dec. 1943. DOI: <https://doi.org/10.1007/BF02478259>.
- 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, Apr. 2001. DOI: <https://doi.org/10.1093/genetics/157.4.1819>.
- MONTESINOS-LÓPEZ, O. A.; MONTESINOS-LÓPEZ, J. C. C.; SINGH, P.; LOZANO-RAMIREZ, N.; BARRÓN-LÓPEZ, A.; MONTESINOS-LÓPEZ, A.; CROSSA, J. A. Multivariate poisson deep learning model for genomic prediction of count data. **G3 Genes|Genomes|Genetics**, v. 10, n. 11, p. 4177–4190, Nov. 2020. DOI: <https://doi.org/10.1534/g3.120.401631>.
- NASCIMENTO, M.; AZEVEDO, C. F.; NASCIMENTO, A. C. C.; CRUZ, C. D. **Abordagens biométricas para reconhecimento de padrões, classificação e predição nas ciências agrárias**. Maringá: Mecenaz, 2022. 282 p.
- ODA, M. C.; SEDIYAMA, T.; CRUZ, C. D.; NASCIMENTO, M.; MATSUO, E. Adaptability and yield stability of soybean genotypes by mean Eberhart and Russell methods, artificial neural networks and centroid. **Agronomy Science and Biotechnology**, v. 8, p. 1–13, Oct. 2021. DOI: <https://doi.org/10.33158/ASB.r142.v8.2022>.
- OLDEN, J. D.; JOY, M. K.; DEATH, R. G. An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. **Ecological Modelling**, v. 178, n. 3/4, p. 389–397, Nov. 2004. DOI: <https://doi.org/10.1016/j.ecolmodel.2004.03.013>.
- PÉREZ, P.; DE LOS CAMPOS, G. Genome-wide regression and prediction with the BGLR statistical package. **Genetics**, v. 198, n. 2, p. 483–495, Oct. 2014. DOI: <https://doi.org/10.1534/genetics.114.164442>.
- QI, S.; NING, X.; YANG, G.; ZHANG, L.; LONG, P.; CAI, W.; LI, W. Review of multi-view 3D object recognition methods based on deep learning. **Displays**, v. 69, 102053, Sept. 2021. DOI: <https://doi.org/10.1016/j.displa.2021.102053>.
- ROSENBLATT, F. **The perceptron, a perceiving and recognizing automaton (Project Para)**. Buffalo: Cornell Aeronautical Laboratory, 1957. 29 p. (Report n. 85-460-1).

RUMELHART, D.; HINTON, G.; WILLIAMS, R. Learning representations by back-propagating errors. **Nature**, v. 323, p. 533–536, Oct. 1986. DOI: <https://doi.org/10.1038/323533a0>.

SILVA JUNIOR, A. C.; SILVA, M. J.; CRUZ, C. D.; SANT'ANNA, I. C.; SILVA, G. N.; NASCIMENTO, M.; AZEVEDO, C. F. Prediction of the importance of auxiliary traits using computational intelligence and machine learning: A simulation study. **PLoS One**, v. 16, e0257213, Nov. 2021. DOI: <https://doi.org/10.1371/journal.pone.0257213>.

SINGH, D.; WANG, X.; KUMAR, U.; GAO, L.; NOOR, M.; IMTIAZ, M.; SINGH, R. P.; POLAND, J. High-throughput phenotyping enabled genetic dissection of crop lodging in wheat. **Frontiers in Plant Science**, v. 10, 394, Apr. 2019. DOI: <http://dx.doi.org/10.3389/fpls.2019.00394>.

SOUSA, I. C.; NASCIMENTO, M.; SANT'ANNA, I. C.; CAIXETA, E. T.; AZEVEDO, C. F.; CRUZ, C. D.; SILVA, F. L.; ALKIMIM, E. R.; NASCIMENTO, A. C. C.; SERÃO, N. V. L. Marker effects and heritability estimates using additive-dominance genomic architectures via artificial neural networks in *Coffea canephora*. **PLoS One**, v. 17, e0262055, Jan. 2022. DOI: <https://doi.org/10.1371/journal.pone.0262055>.

SOUSA, I. C.; NASCIMENTO, M.; SILVA, G. N.; NASCIMENTO, A. C. C.; CRUZ, C. D.; SILVA, F. F.; ALMEIDA, D. P.; PESTANA, K. N.; AZEVEDO, C. F.; CAIXETA, E. T. Genomic prediction of leaf rust resistance to Arabica coffee using machine learning algorithms. **Scientia Agricola**, v.78, e20200021, 2021. DOI: <https://doi.org/10.1590/1678-992X-2020-0021>.

ZADEH, L.A. Fuzzy sets. **Information and Control**, v. 8, n. 3, p. 338-353, June 1965. DOI: [https://doi.org/10.1016/S0019-9958\(65\)90241-X](https://doi.org/10.1016/S0019-9958(65)90241-X).

João Luis da Silva Filho, Marcos Santos, Gustavo Eduardo Marcatti, Rafael Tassinari Resende

Introdução

A Pesquisa Operacional (PO) é uma ciência interdisciplinar que fornece ferramentas quantitativas ao processo decisório. Por essa assertiva, muitos dos modelos que os melhoristas usam em suas pesquisas podem ser considerados como pertencentes à PO. A programação matemática é, reconhecidamente, a técnica mais popular de pesquisa operacional, e essa temática domina o conteúdo de livros que são considerados referência na área, tal como os de Hillier e Lieberman (2001) e Winston (2004). Porém, a PO vai muito além dos modelos determinísticos da programação e otimização matemática, abrangendo também modelos estocásticos, modelos não supervisionados, aprendizado de máquina e aqueles que também levam em consideração a subjetividade do tomador de decisão, inerentes a qualquer processo decisório.

O termo "pesquisa operacional" foi proposto em 1939, durante a Segunda Guerra Mundial, sendo que muitos dos ramos que hoje a compõe (e que já existiam) passaram a ser abrigados sob a mesma área do conhecimento. Por exemplo, modelos clássicos como as Cadeias de Markov (1856-1922), que atualmente são basilares para heurísticas modernas na área de Teoria dos Grafos, também são objeto de pesquisa da PO. Técnicas de pesquisa mais modernas, a exemplo da Lógica *Fuzzy*, também podem ser classificadas como PO. Já na subárea Apoio Multicritério à Decisão, em especial nos denominados Métodos de Agregação e Síntese (MAS), há técnicas que datam do século XVIII, como o índice Borda (em melhoramento conhecido como índice de *Mulamba* e *Mock*), além de técnicas desenvolvidas posteriormente, como o método *Analytic Hierarchy Process* (AHP) desenvolvido na década de 1970 por Thomas L. Saaty.

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

É possível afirmar que muitos dos modelos usados em melhoramento e em PO para descrever realidades diversas, aparentemente desconexas, são bem correlatos em suas essências. Em PO, a estratégia da modelagem ocorre após a definição da tipologia do problema: seleção, ordenação ou classificação (categorização). Qualquer associação dessa afirmação com o melhoramento de plantas não é mera coincidência: selecionar, ordenar e categorizar faz parte do dia a dia do melhorista, sendo abordados por diferentes modelos. Alguns modelos em PO apresentam motivação muito próxima aos modelos de Índices de Seleção usados no melhoramento vegetal. Uma vez que modelos clássicos de Índices de Seleção, como o de Smith (1936), são anteriores ao surgimento do termo "pesquisa operacional", pode-se afirmar que a PO se desenvolveu com roupagens diferentes em diversas áreas do conhecimento. Isso deixa evidente a interdisciplinaridade da PO com outras áreas de conhecimento.

Apesar do potencial da PO para auxiliar o melhoramento de plantas, sua aplicação ainda não é amplamente utilizada nesse campo. A implementação da PO poderá trazer uma série de vantagens, como a otimização do uso de recursos, melhoria da gestão nos programas de melhoramento, identificação de soluções mais eficientes e tomada de decisões mais acertadas, incluindo também a seleção de parentais, planejamento de cruzamentos e design de experimentos. Dessa forma, é fundamental que sejam realizados mais esforços para promover a utilização da PO no melhoramento de plantas.

Modelos Probabilísticos, Determinísticos e a Subjetividade

A Pesquisa Operacional visa encontrar a melhor ou melhores soluções para diversos tipos de problemas. A análise comparativa de dois cenários revela distintas características nas superfícies de resposta que representam as soluções de um problema matemático em função de dois parâmetros, θ e λ (Figura 1). No primeiro cenário, a superfície de resposta exibe clareza na definição da melhor solução, apresentando aparência limpa e plana (Figura 1A). Nesse caso, a solução ótima é facilmente identificável, evidenciada por um ponto de máximo global nitidamente destacado em relação às demais soluções, e a uniformidade da superfície contribui para a simplicidade da análise. Em contraste, um segundo cenário apresenta uma superfície de resposta altamente irregular, caracterizada por múltiplos pontos de máximo local e um ponto de máximo global inconclusivo (Figura 1B). A complexidade dessa superfície

dificulta a identificação da solução ótima, exigindo abordagens estratégicas para evitar convergências prematuras para os máximos locais, que obviamente são soluções não ótimas ou inverossímeis para o problema.

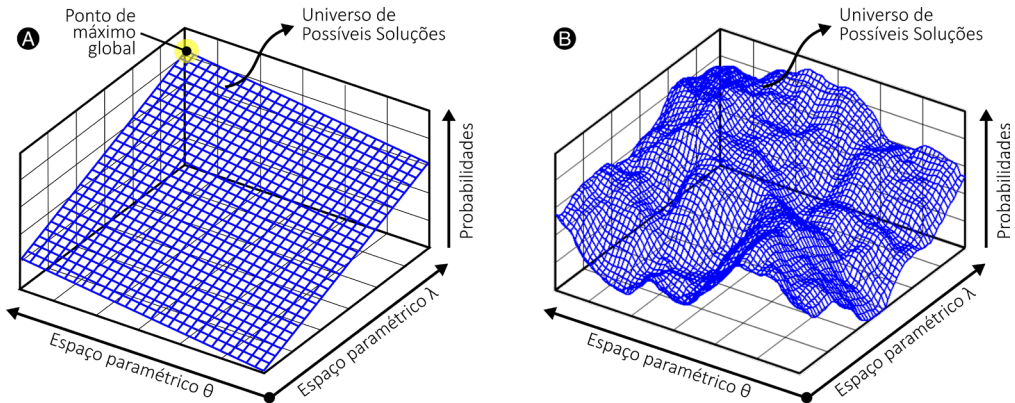


Figura 1. Comparação de cenários com soluções em superfícies de resposta de problemas matemáticos, em função dos parâmetros θ e λ . Na parte A, é observada uma superfície de resposta bem definida, onde a solução ótima é facilmente identificada. Um ponto de máximo global é claramente visível, representando a solução ideal, dentro de um espaço de soluções contido em uma área perfeitamente plana. Na parte B, a superfície de resposta apresenta uma configuração irregular, com múltiplos pontos de máximo local. A identificação do ponto de máximo global torna-se incerta.

Na rotina do melhorista, a tomada de decisão baseada em diferentes soluções é uma ocorrência frequente, variando em níveis de complexidade, como ilustrado nas Figuras 1-A e 1-B. Essas soluções abrangem desde a seleção dos melhores materiais genéticos para ciclos futuros de seleção até a combinação seletiva de características fenotípicas (Mullin; Belotti, 2016, Simiqueli et al., 2023), incluindo até a utilização de modelos preditivos para seleção genômica (Siragusa, 2019). Nesse tema, no próprio processo de convergência dos modelos mistos, muito utilizados no âmbito do melhoramento de plantas, tanto em suas versões frequentistas quanto bayesianas, a obtenção da solução ótima dos componentes de variância estimados pelo método de máxima verossimilhança restrita – REML (*Restricted Maximum Likelihood*) também está intrinsecamente ligada ao processo de encontrar a melhor solução global, como, por exemplo, pelo método *Average Information* (Gilmour et al., 1995).

Via de regra, as características de maior importância econômica no melhoramento genético de plantas são quantitativas, ou seja, possuem controle poligênico e/ou são muito influenciadas por fatores do ambiente. Nesses cenários, os melhoristas utilizam

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

modelos visando mensurar e compreender o quanto os componentes genéticos, não genéticos e os erros aleatórios influenciam a seleção e o seu progresso. Assim, na população de interesse do melhorista são considerados tanto aspectos genéticos como estatísticos para uma sólida tomada de decisão. Dada essa particularidade, os modelos usados em melhoramento são majoritariamente probabilísticos, em especial para as características em que o erro experimental é elevado, seja em estudos de controle genético e herança, programas de seleção recorrente, condução de populações segregantes e/ou estudos da interação genótipos \times ambientes. Subáreas da PO, como Simulação e Eventos Discretos, também fazem uso essencialmente de modelos probabilísticos.

Talvez por isso, com foco em uma formação essencialmente experimental, pouca ênfase tenha sido dada até agora, nos cursos de pós-graduação em melhoramento, aos modelos determinísticos, a exemplo dos modelos de programação matemática, seja essa linear ou inteira. Uma das características desses modelos é a necessidade da construção de uma função objetivo, que pode ser de custo (minimização) ou de lucro (maximização). Certamente, tais modelos poderiam ser muito úteis na gestão dos programas de melhoramento e/ou do portfólio dos produtos gerados, em especial no dimensionamento de sua logística, melhor aplicação dos recursos financeiros e humanos, posicionamento comercial de novas cultivares, dentre outras situações.

Menos ênfase ainda talvez seja dada a abordagens que considerem a subjetividade como parte inerente do processo decisório, a exemplo do modelo AHP. Em PO, a subjetividade não é tratada como algo que "contamina" o processo de tomada de decisão, mas como algo que não pode ser dissociado dele. Uma vertente da PO denominada *PO soft* se ocupa de uma etapa do processo decisório denominada "abstração de cenários", quando múltiplos pontos de vista sobre um problema podem surgir, sendo que essa etapa geralmente antecede a etapa de formalização matemática e lógica do problema (*PO hard*).

Grandes Áreas da Pesquisa Operacional

A finalidade principal deste capítulo é difundir as possibilidades de uso da PO no melhoramento genético de plantas, sem a intenção ser um material didático de consulta de como aplicá-las. Antes de ilustrar a aplicação de alguns modelos de decisão, serão brevemente descritas algumas das grandes áreas de atuação da PO.

Programação Matemática: refere-se ao estudo de problemas em que se busca otimizar (minimizar ou maximizar) uma função por meio da escolha sistemática dos valores de variáveis reais ou inteiras dentro de um conjunto viável. Basicamente, a

programação matemática tenta encontrar a melhor solução do problema, dentre as soluções alternativas viáveis, e assim alocar de forma adequada os recursos disponíveis. A programação matemática abrange tanto a Programação Linear como a Programação Inteira, que são utilizadas na otimização de processos. Para isso, uma função objetivo é construída, podendo essa ser de maximização (lucro ou rendimento) ou de minimização (custo ou recursos), sob um dado conjunto de restrições. Uma aplicação em agricultura, por exemplo, seria alocar as melhores cultivares em uma propriedade agrícola (Telles et al., 2021). Vale destacar que esse tipo de modelagem se aplica somente a problemas determinísticos. Em melhoramento genético de plantas, há trabalhos na literatura usando programação matemática em vários cenários, como escolha de genitores (Liu; Dauer, 1996), melhoramento de plantas perenes (Mattheiss; Land, 1984), genômica (Siragusa, 2019), maximização de diversidade genética na seleção (Mullin; Belotti, 2016), dentre outras aplicações (Miller et al, 2010; Voorrips et al., 2011; Kuska et al., 2015).

Simulação a Eventos Discretos: em melhoramento genético, em muitas ocasiões, simulações são usadas visando maior entendimento dos modelos genéticos-estatísticos sob vários cenários de seleção. Essa grande área da PO usa, também, simulação para vislumbrar os possíveis pontos fortes e/ou limitantes em processos produtivos de uma indústria, empresas de logística, etc. A Simulação a Eventos Discretos permite identificar, por exemplo, mão de obra ociosa e gargalos em um processo produtivo. Essa subárea da PO interage muito bem com a Teoria das Restrições e a Teoria das Filas, com algumas aplicações no melhoramento genético de plantas (Beukelaer, 2017).

Teoria dos Grafos: grande área da PO que permeia todas as outras e permite estruturar e visualizar problemas de forma muito elegante via estruturas denominadas grafos. Matematicamente, estuda as relações entre objetos de um dado conjunto. Os objetos desse conjunto são denominados vértices (ou nós), enquanto uma conexão entre qualquer par de objetos é denominada de aresta, e são essenciais em estudos de logística. Em um processo produtivo, todas as vezes que se tem um conjunto de elementos que se inter-relacionam de alguma maneira, esse fenômeno pode ser representado por meio de um grafo. Embora a topologia empregada seja relativamente simples, muitos problemas possuem estrutura matemática extremamente complexa, por se tratar de problemas não polinomiais completos ou *NP-Hard*, cuja solução ótima só pode ser encontrada por exaustão.

Lógica Fuzzy: também conhecida como *Lógica Nelubosa*, a Lógica Fuzzy foi proposta na década de 1960 por Lotfali Askar-Zadeh. É baseada no conceito de Variável Linguística e de Função de Pertinência. Pode ser entendida como uma

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

generalização da teoria clássica dos conjuntos. O valor dessa função de pertinência varia de 0 a 1 para uma dada variável linguística. No Brasil, um dos maiores nomes da Lógica *Fuzzy* é o Professor Carlos Alberto Nunes Cosenza, criador do método COPPE-COSENZA, que permite que as decisões sejam tomadas considerando vários objetivos ao mesmo tempo, levando em conta a incerteza e a subjetividade presentes em muitos problemas. É amplamente utilizada em vários setores, como administração, finanças, logística e produção, entre outros. No caso do melhoramento genético, há trabalhos na literatura considerando o uso da Lógica *Fuzzy*, como em estudos da interação Genótipos × Ambientes (Carneiro et al., 2019).

Análise Envoltória de Dados (DEA): é a área da PO que permite avaliar a eficiência relativa de diferentes alternativas em uma determinada tarefa ou processo. É uma abordagem não paramétrica que não requer suposições sobre a forma da função de produção ou relações lineares entre as entradas (*inputs*) e as saídas (*outputs*) do modelo. Ao invés disso, a DEA constrói fronteiras de eficiência que definem o máximo possível de *outputs* que podem ser obtidos dado um conjunto fixo de *inputs* ou o mínimo possível de *inputs* que são necessários para produzir uma determinada quantidade de *outputs*. A DEA pode ser orientada por diferentes critérios, tais como economia de insumos (*input-oriented* DEA) ou aumento de retornos (*output-oriented* DEA). Isso significa que a análise pode ser direcionada a identificar alternativas que usam menos *inputs* para produzir a mesma quantidade de *outputs* ou alternativas que produzem mais *outputs* com a mesma quantidade de *inputs*.

Apoio Multicritério à Decisão (AMD): como todo tomador de decisão, o melhorista decide quais genótipos (alternativas de decisão) selecionar com base em várias características (critérios de decisão). Alguns dos modelos dessa grande área da PO apresentam forte similaridade com os índices de seleção usados em melhoramento. Por exemplo, o índice de *Mulamba e Mock* usado em melhoramento é idêntico ao método de Borda, proposto por Jean Charles de Borda (1733-1799). Na literatura, há centenas desses métodos, desenvolvidos principalmente na década de 1960 e aplicados nos dias atuais. Basicamente, a utilização de um ou de outro modelo dependerá da natureza do problema e dos dados que se dispõem para a tomada de decisão.

Meta-Heurística: consiste em uma abordagem de resolução de problemas que envolve a utilização de uma combinação de técnicas heurísticas, buscando alcançar soluções ótimas, porém não únicas ou exatas. A meta-heurística é uma abordagem geral, que pode ser aplicada a uma ampla gama de problemas, incluindo problemas de otimização, roteamento, classificação e agrupamento. Algumas técnicas comumente usadas em meta-heurística incluem: *Algoritmos Genéticos*, *Simulated Annealing*,

Algoritmos de Busca Tabu, Algoritmos de Otimização por Colônia de Formigas, Algoritmos de Otimização por Enxame Particulado, entre outras.

Exemplos de Aplicação da PO na Agricultura e Melhoramento

A seguir são apresentados cinco exemplos ilustrativos de teoria dos grafos, programação linear, programação inteira, apoio multicritério à decisão genética e meta-heurística, todos aplicados ao melhoramento ou agricultura, visando explorar cenários potencialmente vividos pelos melhoristas cujas ferramentas quantitativas para atacar o problema não fazem parte de sua formação padrão. Essas abordagens proporcionam soluções analíticas e quantitativas para os desafios enfrentados pelos melhoristas e agricultores, ajudando-os a tomar decisões informadas e otimizadas em diversas áreas, como alocação de recursos, roteamento, seleção de variedades de plantas e planejamento de projetos agrícolas. Ao aplicar essas técnicas, os profissionais envolvidos no melhoramento ou agricultura poderão ampliar sua capacidade de análise e encontrar soluções eficientes e eficazes, maximizando a produtividade, reduzindo os custos e minimizando os riscos.

Teoria dos grafos, o problema do caixeiro viajante e a coleta de germoplasma

Grafos são estruturas que permitem estudar as relações entre objetos de um dado conjunto (vértices ou nós em uma representação gráfica) e pares não ordenados desses objetos (as arestas que conectam os nós e que definem quais são os pares de objetos). Do ponto de vista da álgebra linear, todo grafo está associado a uma matriz de adjacência, em que as linhas e as colunas são os vértices e os elementos da matriz caracterizam a relação entre cada par de vértices.

Os grafos têm aplicações em várias áreas, como engenharia de sistemas, logística, rede de computadores, dentre outras, permeando diferentes áreas da PO, como simulação a eventos discretos, programação linear e inteira, teoria das filas e apoio multicritério à decisão. Uma rede social, por exemplo, pode ser representada por um grafo, em que os vértices (nós) são as pessoas e as conexões entre os nós (arestas) representam a existência ou não de "amizade" entre essas pessoas (Figura 2-A). O grau de um vértice é definido como o número de arestas que incidem sobre ele e, nesse caso, representaria o número de amigos de uma pessoa na rede. No grafo abaixo, o vértice associado a "Maria" tem só uma aresta e, conceitualmente, seria considerado um vértice pendente.

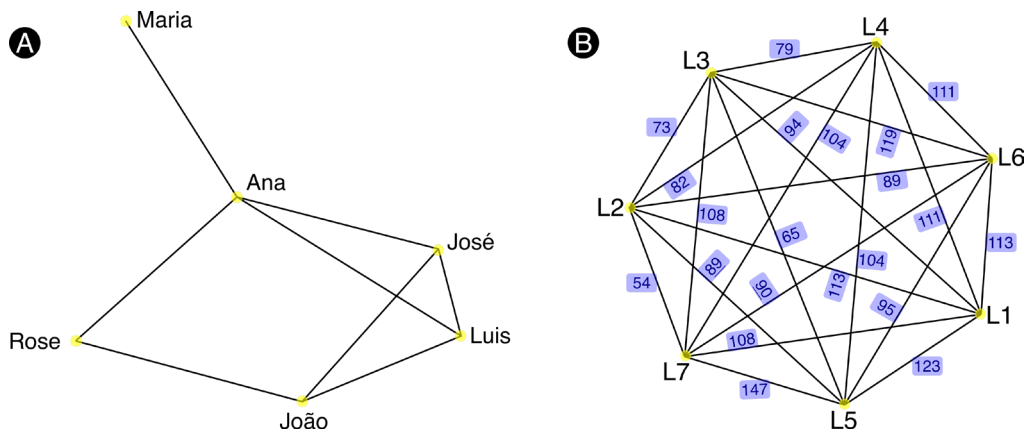


Figura 2. Representação de grafos em diferentes contextos. Parte A: Exemplo de uma rede social com seis pessoas (vértices) e a conexão de amizade entre elas (arestas). Parte B: Ilustração de um grafo em que os vértices representam os locais de coleta de germoplasma e as arestas as rotas existentes entre essas locais. Os valores sobre as arestas são as distâncias em km entre os locais (nem sempre lineares como mostrado no diagrama).

Mas, o que um grafo teria a ver com as atividades de um melhorista? Por exemplo, suponha-se que ao invés de uma rede social o grafo representasse locais escolhidos para se fazer uma coleta de germoplasma e que se deseja otimizar o deslocamento da equipe conforme as distâncias entre os pontos de coleta (Figura 2-B). Um grafo em que todos os nós estão conectados é chamado de grafo completo, em que o número total de arestas é dado pelo número de combinações entre os nós dois a dois.

Esse cenário é essencialmente um problema de logística e, nesse caso, poderia ser considerado como o problema do caixeiro viajante. Ao se desejar visitar cada local de coleta (cada nó do grafo) apenas uma vez, sem repetir nenhum, busca-se realizar o que em teoria dos grafos denomina-se ‘caminho hamiltoniano’. Se nesse percurso for possível voltar ao ponto de partida passa-se a ter um ‘ciclo hamiltoniano’.

Dado um grafo completo, há vários caminhos/ciclos hamiltonianos e o que se deseja é identificar o caminho/ciclo ótimo. Na verdade, considerando grafos com poucos nós, a solução pode ser encontrada por exaustão, ou seja, avaliando-se todas as possibilidades do espaço solução e escolher aquela que minimizaria o percurso. O número de rotas possíveis corresponde a $(n - 1)!$, em que n é o número de nós, desde que haja arestas entre todos os nós. Nesse exemplo, 720 rotas alternativas são possíveis ($6!$), o que do ponto de vista computacional é irrelevante, sendo viável buscar a solução ótima por exaustão.

Todavia, com o aumento do valor do número de nós há uma explosão exponencial no número de soluções possíveis e, nesse caso, heurísticas são usadas na busca de soluções. Heurísticas são algoritmos criados para se aproximar da solução de um problema, ainda que não haja garantia de que seja a ótima. No caso do problema do caixeiro viajante ainda não há algoritmo que garanta que a solução ótima seja atingida. Por exemplo, com $n > 20$ a busca da melhor solução por exaustão já seria inviável.

Para um outro exemplo da aplicação da teoria dos grafos, especialmente no contexto do melhoramento molecular, pode-se propor a modelagem do mapeamento genético em grafos de ligação. Nessa abordagem, os nós do grafo representam marcadores genéticos, como SNPs (polimorfismos de nucleotídeo único), e as arestas refletem a proximidade ou distância genética entre esses marcadores. Essa representação em grafo permite a identificação de regiões genômicas associadas a características desejáveis por meio de estudos de associação genômica ampla (GWAS). Ao analisar as conexões e padrões de ligação entre os marcadores genéticos, é possível identificar regiões cromossômicas que mostram uma associação significativa com características fenotípicas de interesse. Essas regiões são conhecidas como locos de características quantitativas (QTLs) e podem fornecer informações sobre os genes ou regiões do genoma que controlam os caracteres qualitativos (ou quantitativos?) de interesse.

Programação Linear (PL) – um exemplo ilustrativo sobre a produção de sementes

A Programação Linear talvez seja a técnica da PO mais conhecida e, de fato, é muito utilizada em processos de otimização, seja na maximização de lucro ou na minimização de custos em um dado cenário de condições (restrições). Por se tratar de um modelo determinístico, é necessária a construção de uma "função objetivo" que pode ser de custo ou de lucro, como já citado. Muitas vezes, o melhorista assume também a função de tomador de decisão em aspectos comerciais e, dessa forma, é importante saber da existência de técnicas que o auxiliem nas decisões, ainda que não as domine. Enfatizando mais uma vez, o objetivo aqui é antes difundir as possibilidades de uso das técnicas de PO do que ser um material de consulta didático para aplicação das técnicas.

Seja um exemplo hipotético quanto à produção de sementes de duas cultivares de uma cultura hipotética, em que se deseja maximizar o lucro com a venda, onde Q_{cx} e Q_{cy} são as quantidades de sementes em toneladas produzidas das cultivares X e Y. Sabe-se que cada tonelada vendida das cultivares X e Y geram lucros de R\$ 4.500,00 e R\$ 5.000,00, respectivamente. Assim, a função objetivo de maximização Z nesse cenário pode ser expressa por:

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

$$Z = 4.500 \times Q_{CX} + 5.000 \times Q_{CY}$$

Na ausência de restrições e/ou condições que viessem a afetar a decisão, apenas sementes da cultivar Y deveriam ser produzidas uma vez que geram lucro maior por tonelada. Entretanto, restrições ou condições limitantes sempre ocorrem no processo decisório, fato que pode alterar completamente a decisão a ser tomada. Por simplicidade, vamos considerar que a capacidade de produção de sementes esteja limitada a 2.000 toneladas e que os custos de produção por tonelada das cultivares X e Y sejam, respectivamente, R\$ 8.000,00 e R\$ 10.000,00, com capacidade de endividamento para a produção de sementes limitada a R\$ 18 milhões de reais. Matematicamente, essas duas restrições podem ser expressas por:

$$Q_{CX} + Q_{CY} \leq 2.000 \text{ toneladas} - \text{restrição da quantidade de produção de sementes}$$

$$8.000 \times Q_{CX} + 10.000 \times Q_{CY} \leq \text{R\$ } 18.000.000 - \text{restrição da capacidade de endividamento}$$

Em cenários mais simples, como esse exemplo ilustrativo, 'varrer' o cenário das possíveis soluções é uma forma de se aproximar da solução ótima. Isso seria obtido considerando todas as combinações das quantidades de sementes das cultivares X e Y que somassem 2.000 toneladas até o limite da capacidade de endividamento. Uma tentativa para isso está na tabela abaixo, onde se vê que a produção de 1.000 toneladas de cada cultivar respeita a capacidade de endividamento e proporciona o maior lucro (Tabela 1):

Tabela 1. Quantidade hipotética de sementes produzidas das cultivares X e Y e o respectivo lucro e endividamento.

QCX	QCY	Endividamento	Lucro
0	2.000	20.000.000	10.000.000
250	1.750	19.500.000	9.875.000
500	1.500	19.000.000	9.750.000
750	1.250	18.500.000	9.625.000
1.000	1.000	18.000.000	9.500.000
1.250	750	17.500.000	9.375.000
1.500	500	17.000.000	9.250.000
1.750	250	16.500.000	9.125.000
2.000	0	16.000.000	9.000.000

Uma outra forma de solucionar o problema é via programação linear, aqui realizada via abordagem gráfica (Figura 3). A área laranja representa e delimita o espaço das possíveis soluções dadas as restrições de limite de produção de sementes (linha azul) e de da capacidade de endividamento (linha vermelha). Graficamente, a produção de sementes da cultivar X está representada no eixo X, enquanto a da cultivar Y está representada no eixo Y. Por Z ser uma função a ser maximizada, a solução ótima é o ponto mais externo do espaço solução (mais afastado da origem o quanto possível) e que faz parte de uma possível função Z (que nesse caso é uma reta). As retas pretas em linhas tracejadas representam três possibilidades da função objetivo Z dentro do espaço solução (área laranja do gráfico). O ponto mais externo que faz parte de uma possível função objetivo Z e do espaço solução (área laranja) é o ponto (1.000, 1.000).

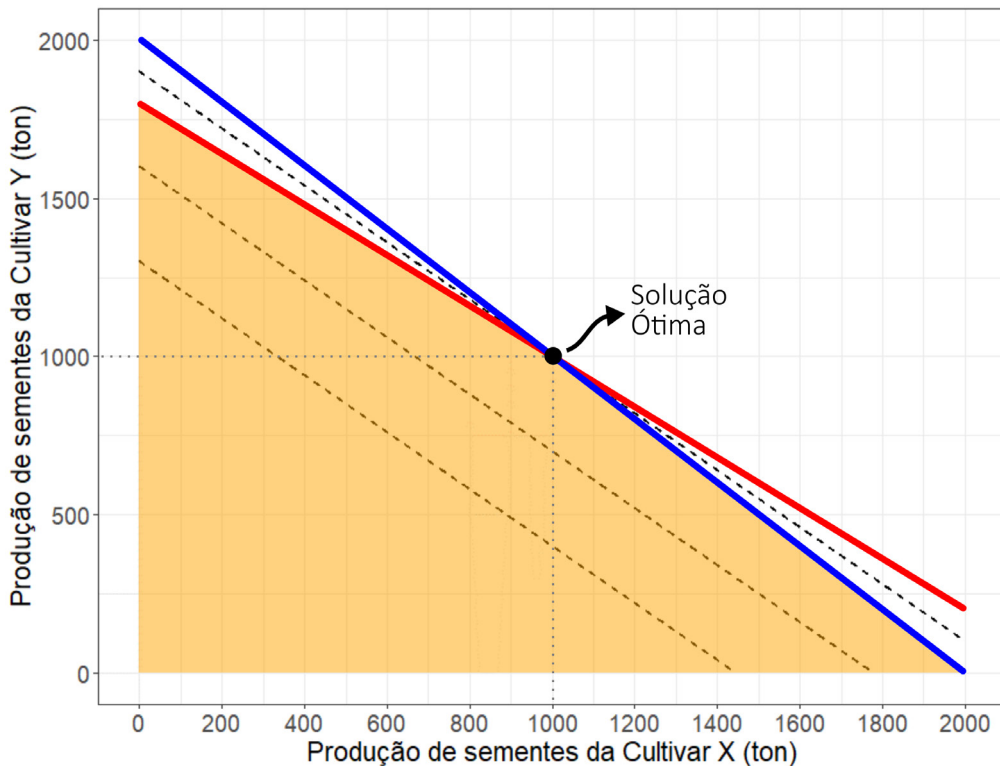


Figura 3. Representação gráfica da decisão de quantidade de sementes a serem produzidas, utilizando programação linear. A área laranja delimita o espaço das soluções viáveis, a linha azul é uma restrição de produção de sementes e a linha vermelha uma restrição de capacidade de endividamento. Os eixos X e Y representam a produção de sementes para as cultivares X e Y, respectivamente. A solução ótima é o ponto (1.000, 1.000), que é o ponto mais externo da área laranja e que faz parte de uma possível função objetivo Z.

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

Outros três exemplos de aplicações da Programação Linear no contexto do melhoramento genético de plantas são: i) a alocação de recursos na seleção de genótipos promissores, visando maximizar a eficiência na seleção considerando restrições de recursos limitados, como espaço de cultivo, mão de obra e orçamento disponíveis (Mattheiss; Land, 1984); ii) a formulação de cruzamentos ótimos, permitindo a identificação da melhor combinação de genótipos com diferentes características desejadas (como a produtividade, teores de aminoácidos, resistência a doenças, qualidade de grãos, tolerância à seca, dentre outros), maximizando, assim, a expressão dessas características desejáveis nas gerações subsequentes; e iii) o planejamento de espaço e rotação de culturas, em que a Programação Linear pode ser usada para determinar a sequência ideal de culturas em um sistema de rotação, considerando as necessidades nutricionais das plantas, as restrições de cultivo e as metas de produtividade, a fim de otimizar a alocação de espaço para cada cultura ao longo do tempo e maximizar o rendimento geral do sistema de cultivo.

Programação Inteira (PI) – Distribuindo a produção dos campos até unidades de beneficiamento

Em PO, muitas das decisões são referentes à questão de logística, sendo o problema dos transportes uma classe especial. Muitos agrônomos e/ou melhoristas, principalmente aqueles envolvidos em desenvolvimento de mercado, se deparam com situações em que o uso da programação inteira (PI) pode ser bastante útil na solução dos problemas. Na verdade, a PI é caso especial da PL em que as soluções são valores inteiros para as variáveis.

Considerando um determinado grupo empresarial produtor de algodão que tem fazendas arrendadas para cultivo, e que não possuam estrutura para beneficiamento nelas, a produção de algodão em caroço deve ser enviada para outras unidades (fazendas) do grupo onde há algodoeira (unidades que fazem a separação da pluma e do caroço), em que cada uma tenha uma dada capacidade de beneficiamento, considerando o prazo de entrega disponível. O grafo abaixo esquematiza uma situação fictícia do escoamento da produção das fazendas até as algodoeiras (Figura 4).

Cada retângulo no grafo é um vértice (nó) e as identificações das fazendas ou algodoeiras os rótulos desses vértices. Por sua vez, cada linha conectando uma fazenda a uma algodoeira é uma aresta do grafo. No cenário acima, apenas para simplificação, o somatório da capacidade de processamento das algodoeiras ($25000 + 13000 + 12000$) é igual à produção total das fazendas ($10000 + 18000 + 22000$). O desafio é encontrar a combinação ótima de escoamento da produção, dada a produção de cada fazenda (lado

esquerdo do grafo), a capacidade de beneficiamento de cada algodoeira (lado direito do grafo) e o custo do transporte de cada fazenda para cada algodoeira.

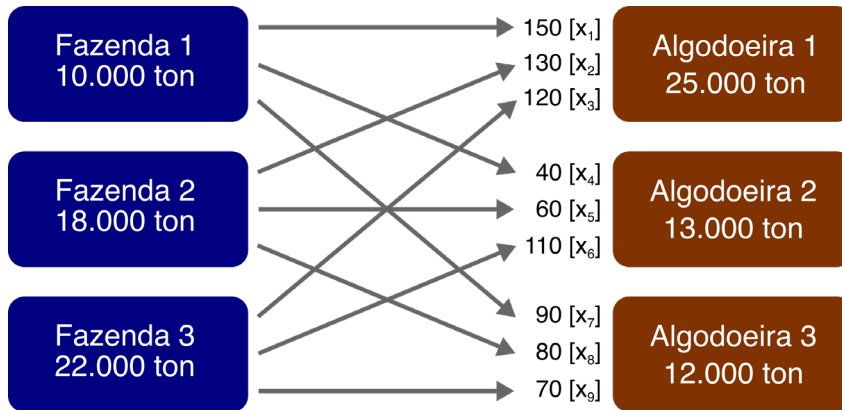


Figura 4. Grafo representando o problema do transporte de três fazendas produtoras de algodão até três possíveis unidades de beneficiamento (algodoeiras). São mostradas as respectivas capacidades de produção e beneficiamento em toneladas (t) em cada caixa. Os coeficientes (150,130,...,70) associados às variáveis x_i , com $i = \{1, 2, \dots, 9\}$, são as distâncias (em km) de uma fazenda produtora de algodão até uma algodoeira.

Assumindo que os valores descritos acima de cada aresta no grafo sejam as distâncias em km de uma dada fazenda até uma dada algodoeira (esses valores não estão descritos), admitindo por simplicidade que essa distância seja diretamente proporcional ao custo de transporte por tonelada de algodão, uma pergunta que pode ser feita é: quanto da produção de cada fazenda deve ser destinada a cada algodoeira para que os custos de transporte sejam minimizados?

Da mesma forma que no exemplo anterior, necessita-se identificar uma função objetivo, as variáveis do modelo e as restrições que devem ser obedecidas. Nesse problema do transporte, o número de variáveis é igual ao número de arestas, e o número de restrições é dado pelo número de vértices. Assim, tem-se nove variáveis por se ter nove caminhos possíveis (arestas) e seis restrições por se ter três fazendas e três algodoeiras. Assumindo como notação x_i para cada variável, portanto variando de x_1 a x_9 , em que cada uma delas é a respectiva produção a ser transportada de uma fazenda para uma algodoeira e cujo coeficiente é a respectiva distância (custo), tem-se que a função de minimização [min] de custos Z e as funções de restrições no espaço solução são dadas por:

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

$$Z = 150 x_1 + 130 x_2 + 120 x_3 + 40 x_4 + 60 x_5 + 110 x_6 + 90 x_7 + 80 x_8 + 70 x_9 \text{ [min]},$$

sujeito a:

$$x_1 + x_2 + x_3 \leq 10 \text{ mil t}$$

$$x_4 + x_5 + x_6 \leq 18 \text{ mil t}$$

$$x_7 + x_8 + x_9 \leq 22 \text{ mil t}$$

$$x_1 + x_4 + x_7 = 25 \text{ mil t}$$

$$x_2 + x_5 + x_8 = 13 \text{ mil t}$$

$$x_3 + x_6 + x_9 = 12 \text{ mil t}$$

$$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9 \geq 0 \text{ (restrição de não negatividade)}$$

No presente exemplo, as restrições "forçam" a procura da combinação que minimize a função Z (função de custo), desde que não sejam transferidas produções além daquelas produzidas em cada fazenda e que cada algodoeira trabalhe no limite do desempenho. As restrições de não negatividade indicam que os valores das variáveis x_j devem ser maiores ou iguais a zero.

Como já salientado, o escopo deste capítulo não é pormenorizar o algoritmo para solução desse tipo de problema, apenas ilustrar o potencial da pesquisa operacional na tomada de decisão. E ferramentas de planilhas eletrônicas, como o "Solver" do Excel, são suficientes para isso. Usando o "Solver" do "Excel", o cenário que minimizaria custos de transportes seria $x_2 = 10$, ou seja, enviar toda a produção da Fazenda 1 para a Algodoeira 2; $x_4 = 18$, ou seja, toda a produção da Fazenda 2 para a Algodoeira 1; e $x_7 = 7$, $x_8 = 3$, $x_9 = 12$, ou seja, 7 mil t da Fazenda 3 iriam para a Algodoeira 1, 3 mil t para a algodoeira 2 e 12 mil t para a Algodoeira 3. As variáveis x_1 , x_3 , x_5 e x_6 seriam iguais a zero.

Métodos de Apoio Multicritério à Decisão

Muitas das decisões que gravitam em torno da estratégia de pesquisa são tomadas apenas com base na experiência do pesquisador, sem o uso de nenhum tipo de ferramenta para apoio à decisão. Como exemplos, podem-se citar a escolha de uma casa vegetação, um telado, uma incubadora BOD (de "Demanda Bioquímica de Oxigênio"), uma área experimental, dentre outras. Normalmente, em cenários como esses, a escolha é predominantemente influenciada pelo menor preço e não necessariamente menor custo ou a que melhor atende as necessidades. Mesmo em cenários em que predomina a subjetividade, a PO pode ser muito útil no apoio à tomada de decisão.

Em PO, estabelecer quais são as "alternativas de decisão" e os 'critérios de decisão' é um passo fundamental. Em melhoramento, em geral, os dados usados em

pesquisa são oriundos de ensaios delineados. Considerando dados experimentais, o que os melhoristas designam como tratamentos dos ensaios (quase sempre genótipos) em PO seriam designados como "alternativas de decisão", enquanto que as variáveis respostas dos experimentos seriam denominadas "critérios de decisão". Em melhoramento, as variáveis respostas em geral são quantitativas ou têm escore ordinal. Uma tabela de dupla entrada que fosse formada por genótipos nas linhas e variáveis nas colunas é denominada "matriz de decisão" em PO (Tabela 2).

Tabela 2. Representação da disposição das variáveis de decisão, alternativas e critérios, em uma tabela de dupla entrada ou matriz de decisão.

	Critério 1	Critério 2	...	Critério 'n'
Alternativa 1	M_{11}	M_{12}	...	M_{1n}
Alternativa 2	M_{21}	M_{22}	...	M_{2n}
...
Alternativa m	M_{m1}	M_{m2}	...	M_{mn}

A seguir serão descritos alguns métodos que poderiam ser aplicados em diversas situações, inclusive naquelas em que as comparações entre alternativas de decisão são realizadas exclusivamente de forma subjetiva. Alguns desses métodos já fazem parte da rotina dos melhoristas, mas conhecidos por outros nomes, enquanto outros certamente são muito pouco difundidos dentre os profissionais da área.

Método Borda

Como já comentado, esse método é idêntico ao modelo de índice de postos de Mulamba e Mock (1978), muito usado em melhoramento de plantas. Na verdade, poderia ser usado para ordenar alternativas de decisão de qualquer natureza numa abordagem multicritério. Assim, dando-se o crédito a quem primeiro propôs o método, o uso do termo "Método Borda" deveria ser preferido à designação Mulamba e Mock (1978). Consiste em ordenar as alternativas em cada critério de decisão, atribuindo-lhes como pontuações os valores das respectivas posições ordinais que ocupam em cada critério. Em seguida, os valores das pontuações de cada alternativa são somados, sendo assumida como melhor alternativa a que obtiver menor soma. Em PO é considerado como um método ordinal compensatório, uma vez que a posição de uma alternativa em um critério pode ser compensada pela posição em uma outra. Uma vez que a literatura sobre esse método é vasta, ele não será pormenorizado aqui.

Método Condorcet

Esse método deriva da alcunha dada ao matemático e filósofo francês Jean-Marie Antoine Nicolas de Caritat (1743-1794), Marquês de *Condorcet*, considerado o precursor da escola francesa de métodos multicritérios. É considerado como um método de sobreclassificação e não de agregação de síntese. É também um método ordinal, em que a comparação entre alternativas em cada critério, duas a duas, depende do ordenamento relativo (posicional) e não da cardinalidade dos valores, atendo-se o julgador apenas às relações de superação entre as alternativas (tratamentos) nos diferentes critérios de avaliação (variáveis). Esse método, pelo menos em parte, é usado de forma intuitiva (mas incompleta!) pelos melhoristas em muitas ocasiões, quando, por exemplo, o melhorista observa a prevalência de vitórias de um genótipo em relação a outro ao longo dos ensaios multilocais. Esse método tem a vantagem de que a comparação relativa entre duas alternativas independe da posição relativa de uma terceira. Além disso, as relações de superioridade entre as alternativas nesse método podem ser representadas por um grafo. O grande problema do método de *Condorcet* é que, não raro, ele pode conduzir um problema a um ciclo de intransitividade. Um comentário sobre o que é transitividade e sua importância é feito no tópico seguinte.

***Analytic Hierarchy Process* (Método AHP) – uma forma elegante de tratar a subjetividade**

O método *Analytic Hierarchy Process* (AHP) foi desenvolvido na década de 1970 pelo então docente da Universidade da Pensilvânia, Thomas L. Saaty. Esse é um método muito versátil e pode ser aplicado tanto quando se dispõe apenas de elementos subjetivos para comparações, quanto quando são baseados na cardinalidade de variáveis quantitativas. Detalhes sobre o método podem ser vistos em Saaty (1980). De maneira mais simples, uma vez estabelecido um objetivo estratégico da pesquisa, esse procedimento permite ao pesquisador definir as prioridades dos critérios de decisão (conforme o objetivo estratégico) a partir de comparações par a par entre eles e, em uma segunda hierarquia, comparar par a par as alternativas de decisão em cada critério, caracterizando assim uma abordagem hierárquica das variáveis de decisão (Figura 5).

Nesse método, os critérios de decisão não precisam ser necessariamente variáveis aleatórias no sentido estatístico. Atributos qualitativos de natureza subjetiva também poderiam tê-los. Em adição, na comparação entre as alternativas dentro de um critério quantitativo, o pesquisador pode optar por fazê-la com base na cardinalidade dos valores ou de uma forma subjetiva com base na escala desenvolvida pelo proponente do método (Saaty, 1980).

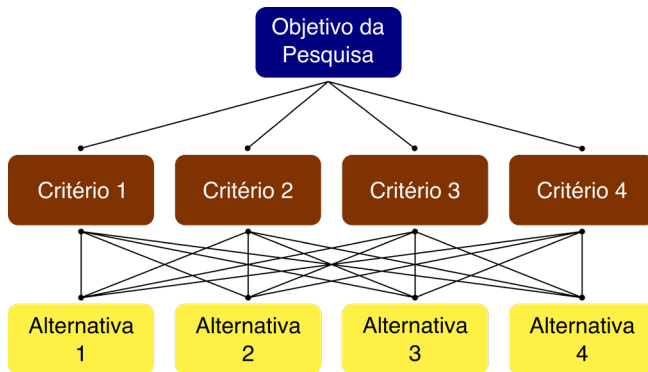


Figura 5. Variáveis de decisão associadas a um objetivo estratégico, incluindo quatro critérios e quatro possíveis alternativas.

Fonte: Adaptado de Saaty (1980).

A Escala Fundamental de Saaty e a estrutura axiomática do método AHP

A escala de Saaty varia de 1 a 9, cujos valores representam as prioridades entre os critérios de decisão (ou entre as alternativas dentro de cada critério) atribuídas pelo decisor. Vê-se que os valores pares são situações intermediárias entre as relações de importância propriamente ditas. Um valor 3 indica que um critério tem importância moderada em relação ao outro, enquanto um valor 9 indica um cenário em que um critério é extremamente mais importante do que outro (Tabela 3).

Tabela 3. Escala fundamental de Saaty para comparação entre critérios de decisão ou entre alternativas dentro de cada critério.

Relação de importância	Grau de importância	Recíproco
Igualdade	1	1
Intermediário	2	1/2
Importância moderada	3	1/3
Intermediário	4	1/4
Mais importante	5	1/5
Intermediário	6	1/6
Muito mais importante	7	1/7
Intermediário	8	1/8
Extremamente mais importante	9	1/9

Fonte: Adaptado de Saaty (1980).

A estrutura axiomática do método pode ser expressa pelo seguinte passo a passo, desde a estruturação do problema até a agregação final dos resultados para tomada de decisão:

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

- 1) definição do conjunto de critérios de decisão e alternativas de decisão: construção de uma matriz que tem nas linhas o conjunto de alternativas e nas colunas os critérios de decisão (similar à Tabela 2);
- 2) avaliação dos critérios de decisão par a par e definição dos pesos de cada critério: faz-se atribuição de julgamentos entre os critérios a partir dos quais será obtido o respectivo vetor de prioridade;
- 3) avaliação das alternativas em cada critério e definição dos respectivos pesos: nesse passo é feito o julgamento de importância entre as alternativas em cada critério de decisão, obtendo-se o vetor de prioridade das alternativas em cada critério;
- 4) processo de agregação das preferências: nesse passo obtém-se o escore final das alternativas, considerando-se os resultados obtidos nos passos anteriores.

A seguir, um exemplo ilustrativo de aplicação do método AHP, mas sem entrar em detalhes. Suponha-se que um melhorista esteja interessado em escolher uma área experimental dentre quatro opções disponíveis (alternativas de decisão) e dispõe de algumas informações (critérios de decisão) a respeito de cada estação para balizar sua escolha. Nesse exemplo, há variáveis de decisão quantitativas contínuas (distância e tempo de viagem), quantitativa discreta (número de viagens semanais) e qualitativas (experiência da equipe). A Tabela 4 a seguir representa a construção do Passo 1 apresentado acima.

Tabela 4. Dados hipotéticos para quatro estações experimentais considerando a distância, tempo de viagem, número de viagens semanais e equipe.

Estação Experimental	Distância (km)	Tempo Viagem (h)	Viagens Semanais Disponíveis	Equipe
Estação 1 (E1)	50	1	10	Inexperiente
Estação 2 (E2)	100	1,5	7	Treinada
Estação 3 (E3)	120	3	7	Experiente
Estação 4 (E4)	220	4	3	Muito Experiente

Usando-se os pesos da escala de Saaty (Tabela 3) obtém-se a matriz de comparação par a par entre os critérios de decisão (matriz de julgamentos) (Tabela 5). Ressalte-se que uma matriz de julgamento obtida pode variar de julgador para julgador, uma vez que as atribuições de pesos podem ser diferentes. No cenário hipotético considerou-se que o critério “Equipe” é “extremamente mais importante” que o critério “Distância” e, por essa razão, o valor 9 foi atribuído para o elemento que ocupa a linha “Equipe” e coluna “Distância”. No Método AHP, elementos simétricos em relação à diagonal

principal sempre apresentam valores recíprocos. Assim, o elemento que ocupa a linha “Distância” e a coluna “Equipe” apresenta valor $1/9$, indicando que o critério “Distância” tem, reciprocamente, importância muito menor do que o critério “Equipe”. O mesmo raciocínio se estende a todas as comparações par a par entre os critérios.

Tabela 5. Matriz de julgamento e totais das colunas para escolha de uma área experimental conforme os critérios de decisão do exemplo anterior.

	Equipe	Distância	Tempo de Viagem	Viagens/Semana
Equipe	1	9	5	3
Distância	$1/9$	1	$1/2$	$1/5$
Tempo de Viagem	$1/5$	2	1	$1/3$
Viagens/Semana	$1/3$	5	3	1
Soma	1,644	17	9,5	4,533

O procedimento seguinte é a padronização dos escores dessa matriz. Para isso, divide-se cada elemento da matriz pela soma dos elementos da respectiva coluna. Em seguida, visando a obtenção do “Vetor de Prioridade” dos critérios, obtém-se a média de cada linha (Tabela 6). Vê-se que o critério “Equipe” seria o mais importante (peso de 58,1%), seguido pelo critério “Viagens/Semana” (peso de 25,8%), “Tempo de Viagem” (peso de 10,5%) e “Distância” (peso de 5,6%). Ressalte-se que a soma dos pesos do vetor de prioridades será sempre a unidade.

Tabela 6. Valores padronizados da matriz de julgamento entre os critérios e o respectivo peso do critério no vetor de prioridade.

	Equipe	Distância	Tempo de Viagem	Viagens/Semana	Vetor de Prioridades
Equipe	0,608	0,529	0,526	0,662	0,581
Distância	0,068	0,059	0,053	0,044	0,056
Tempo de Viagem	0,122	0,118	0,105	0,074	0,105
Viagens/Semana	0,203	0,294	0,316	0,221	0,258

Definida a contribuição de cada critério, o próximo passo é comparar as alternativas de decisão dentro de cada um deles. Para isso, é preciso observar duas propriedades de cada variável de decisão: (i) se é quantitativa ou não; (ii) se é uma métrica que se deseja maximizar ou minimizar. No exemplo exposto, tem-se que “Distância”, “Tempo de Viagem” e “Viagens/Semana” são grandezas quantitativas, sendo que as duas primeiras são de minimização, uma vez que quanto menor o tempo de viagem e menor for a distância, melhor; já o critério “Viagens/Semana” é de maximização, uma vez

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

que quanto maior o número de viagens disponíveis melhor. Para o critério “Equipe”, como não há uma métrica quantitativa que a caracterize, serão atribuídos escores em que quanto maior a pontuação, melhor treinada é a equipe. É possível comparar as alternativas de decisão par a par em cada critério via escala de Saaty, conforme feito para os critérios de decisão, mas não será apresentado aqui.

Assim, é necessário um ajuste nos dados originais (Tabela 4), considerando as particularidades descritas acima. Assim, para os critérios “Distância” e “Tempo de Viagem” devem ser considerados os inversos dos valores originais, pois são critérios de minimização. Já para o critério Viagens/Semana, os valores originais são preservados, pois é um critério de maximização. Para o critério “Equipe” há um escore de notas (subjetivo) conforme o nível de treinamento das equipes de cada estação (Tabela 7).

Tabela 7. Valores da matriz de decisão ajustados conforme critérios de decisão de maximização ou minimização.

Estação Experimental	Equipe	Distância (km)	Tempo Viagem (h)	Viagens/Semanas Disponíveis
Estação 1 (E1)	1	1/50	1,0	10
Estação 2 (E2)	3	1/100	1/1,5	7
Estação 3 (E3)	5	1/120	1/3,0	7
Estação 4 (E4)	7	1/220	1/4,0	3
Soma	16	0,0428	2,25	27

Para proceder a normalização dos valores em cada critério, divide-se cada elemento pelo somatório da respectiva coluna (Tabela 8). Como nesse exemplo apenas as cardinalidades estão sendo usadas, esses valores obtidos já são os pesos dos vetores de prioridades das alternativas em cada critério.

Tabela 8. Matriz de decisão normalizada.

Estação Experimental	Equipe	Distância (km)	Tempo Viagem (h)	Viagens/Semanas Disponíveis
Estação 1 (E1)	0,0625	0,467	0,444	0,370
Estação 2 (E2)	0,1875	0,234	0,296	0,259
Estação 3 (E3)	0,3125	0,194	0,148	0,259
Estação 4 (E4)	0,4375	0,105	0,111	0,111

Por fim, tem-se o procedimento de agregação em que a matriz obtida anteriormente é multiplicada pelo vetor de prioridade dos critérios.

Equipe	Distância	Tempo de Viagem	Viagens/Semana		Prioridades Critérios	=	Prioridades Agregadas	
0,0625	0,467	0,444	0,370	×	0,581	=	0,2045	Estação 1
0,1875	0,234	0,296	0,259		0,056		0,2198	Estação 2
0,3125	0,194	0,148	0,259		0,105		0,2748	Estação 3
0,4375	0,105	0,111	0,111		0,258		0,3003	Estação 4

Vê-se que, pelo método AHP, as estações 4 e 3 seriam as mais adequadas para a instalação dos ensaios. Na prática, o pesquisador poderia ficar em dúvida sobre instalar o ensaio na estação 3 ou estação 4, mas descartaria as estações 1 e 2.

A consistência das atribuições no método AHP e o princípio da transitividade

No Método AHP é possível a avaliação da consistência dos julgamentos par a par entre as variáveis, sejam elas os critérios ou alternativas de decisão. Há alguma tolerância para inconsistências nas comparações, mas não para incoerências. Isso porque o princípio da transitividade deve ser sempre mantido o quanto possível nas comparações entre as variáveis. A propriedade da transitividade pode ser compreendida com um exemplo simples: se $A > B$ e $B > C$, então $A > C$, para todo A, B e C pertencente a \mathbb{R} . Assim, a comparabilidade “maior que” é sempre transitiva. Similarmente, uma relação de preferência P deve ser sempre transitiva: se $A P B$ e $B P C$, então $A P C$. Por outro lado, em uma relação de *não diferença I* (um teste de comparações múltiplas como o de Tukey, por exemplo) não necessariamente há transitividade, pois se A / B e B / C , não implica em A / C . O cálculo da razão de consistência é feito mediante a comparação dos índices de consistência das atribuições feito pelo tomador de decisão em relação a um índice de consistência de uma matriz aleatória. Aqui, não será apresentado tal procedimento, mas detalhes podem ser encontrados em Saaty (1980).

As técnicas de Apoio Multicritério à Decisão também podem desempenhar um papel na definição e implantação de experimentos genéticos no melhoramento de plantas. Ao considerar múltiplos critérios e vários objetivos, como produtividade, número de materiais genéticos e recursos disponíveis, os melhoristas podem priorizar características de interesse, selecionar genótipos promissores e otimizar

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

a alocação de recursos. Além disso, o Apoio Multicritério à Decisão oferece uma abordagem abrangente para a tomada de decisões, aumentando a eficiência do processo de seleção e contribuindo para o avanço de características desejáveis nas culturas agrícolas. Em projetos de processos baseados em simulação, a robustez dos parâmetros incertos do modelo também é importante. Um procedimento prático foi apresentado por Bortz et al. (2015), incorporando a robustez como um objetivo em um *framework* de otimização multicritério, aplicável tanto ao projeto de processos quanto ao projeto de experimentos com plantas.

Meta-Heurística

Foi comentado que em PO, em muitas ocasiões, é possível "varrer" o espaço de soluções em busca da solução ótima (como visto na Figura 1-A) e, por outras vezes, essa varredura é bastante dificultada (como visto na Figura 1-B). Os métodos ou algoritmos de meta-heurística são abordagens poderosas e versáteis na otimização de problemas complexos que se baseiam em uma série de técnicas e estratégias para explorar o espaço de busca do problema com o objetivo de aproximar-se da solução ótima, mas não garantindo encontrá-la. Algumas dessas técnicas incluem: "perturbação e diversificação", "busca local", "aceitação probabilística", entre outras. A escolha da meta-heurística a ser utilizada depende do problema a ser resolvido, sendo possível combinar várias delas para melhorar a eficiência de resolução.

No melhoramento genético de plantas, a meta-heurística pode ser usada para selecionar plantas com características desejadas, como resistência a doenças, produção de frutos, etc. Por exemplo, um algoritmo genético pode ser utilizado para selecionar as melhores combinações de genes para produzir plantas mais resistentes. Além disso, a meta-heurística também pode ser usada para determinar a melhor maneira de plantar e cultivar as plantas, levando em conta fatores como o clima, o solo e outras restrições. Em geral, a meta-heurística é uma ferramenta valiosa para aprimorar as práticas de melhoramento genético de plantas e produzir variedades mais resistentes e produtivas. Pode-se destacar algumas outras aplicações pontuais de meta-heurística no entendimento de processos combinatórios altamente complexos no melhoramento de plantas:

- Tamanho do genoma de uma espécie, comparado com número de combinações em cruzamentos biparentais, tanto para espécies vegetais diploides, como em um nível de complexidade exponencialmente maior para espécies poliploides.

É oportuno lembrar que populações de melhoramento são sistematicamente geradas por meio de múltiplos cruzamentos biparentais;

- Na seleção dos melhores materiais genéticos, evitando-se erosão genética ao longo dos ciclos de seleção recorrente (isto é, aumento dos ganhos genéticos, porém mantendo o tamanho efetivo populacional (N_e) dos indivíduos selecionados);
- Na seleção dos melhores materiais genéticos considerando-se multicares (isto é, visando obter ganhos genéticos simultâneos em todos os caracteres fenotípicos avaliados);
- Num ambiente de banco germoplasma, ou em uma população base, inúmeras combinações de cruzamento entre genitores são possíveis. Nesse contexto, algoritmos de pesquisa operacional, incluindo as meta-heurísticas, podem direcionar quais serão as melhores combinações de genitores, evitando-se genitores aparentados e, portanto, reduzindo a endogamia e maximizando efeitos de heterose;
- Coleta de sementes em árvores-matrizes baseado em distâncias geográficas, com objetivo de montar testes de progênies. Conforme já comentado, um problema clássico na PO é o do “caixeiro-viajante” (um impasse de otimização que busca encontrar a rota mais curta para visitar todas locais e retornar ao ponto de partida). Esse problema parece simples, porém para se visitar 10 cidades, são possíveis $9!$ (fatorial de 9) soluções diferentes, ou seja, cerca de 362.880 soluções são possíveis. No entanto, é importante lembrar que nem todas essas soluções serão ótimas, já que o objetivo é encontrar a rota mais curta;
- Como nos métodos exatos mostrados anteriormente neste capítulo, é possível otimizar o número de experimentos em diferentes locais pelas inferências sobre a interação Genótipos por Ambientes ($G \times A$), combinadas com o tamanho de cada experimento e, também, na otimização do melhor número de indivíduos por famílias e a melhor quantidade de repetições experimentais.

Algumas meta-heurísticas são baseadas em eventos observados na natureza ou no dia a dia, com o objetivo de inspirar soluções para problemas complexos. Essas meta-heurísticas são conhecidas como meta-heurísticas bioinspiradas. A meta-heurística Algoritmo Genético (do inglês *Genetic algorithms*) é baseada no processo de seleção natural darwiniano. O Algoritmo Genético simula a evolução

de uma população, em que cada indivíduo é uma possível solução para o problema. A seleção é realizada com base em critérios de avaliação, e a recombinação de soluções é realizada de forma semelhante à reprodução dos seres vivos.

É apresentado na Figura 6 um processo iterativo usando algoritmo genético, visando seleção simultânea de três caracteres fenotípicos quaisquer (A, B e C), em uma cultura florestal, combinado com a manutenção do tamanho efetivo (N_e) da população original (Simiqueli et al., 2023). Observa-se que o caráter “A” possui maior potencial para alcançar maiores ganhos, enquanto o “C”, menores ganhos. Isso é dependente da variabilidade genética do caráter, e não uma limitação do processo de otimização. É possível observar a convergência do processo iterativo em aproximadamente 1000 iterações.

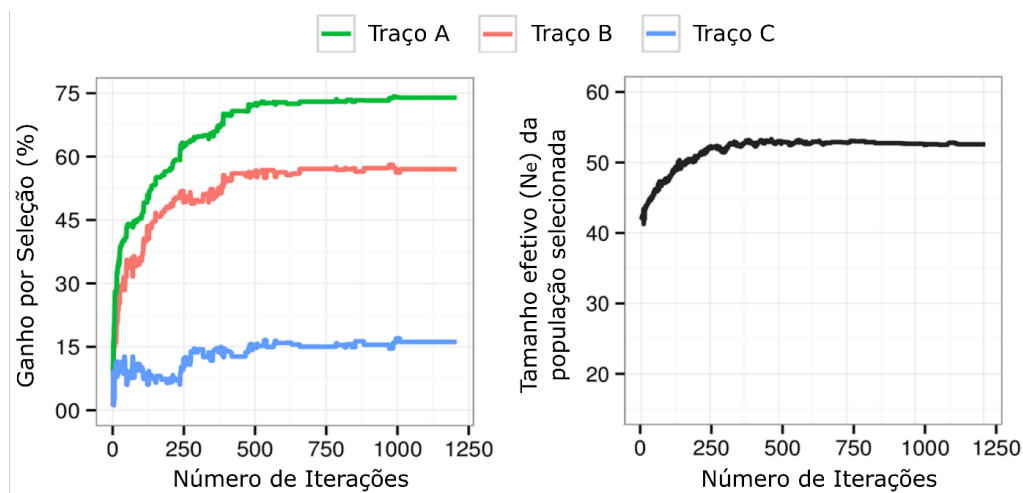


Figura 6. Processo iterativo para seleção múltipla de três caracteres fenotípicos, com manutenção otimizada do tamanho efetivo da população inicial, por meio de algoritmo genético.

Fonte: Adaptado de Simiqueli et al. (2023).

Outra meta-heurística que é baseada em eventos cotidianos é o Algoritmo de Arrefecimento Simulado (do inglês *Simulated Annealing*), que é inspirado no processo de endurecimento de metais. O algoritmo simula o processo de arrefecimento (ou resfriamento) de um metal, onde ocorrem mudanças aleatórias na solução, e a solução é aceita ou rejeitada com base na sua avaliação e na temperatura corrente. A temperatura é gradualmente diminuída, a fim de evitar que a solução fique presa em um mínimo local, resultando em uma solução global ótima.

Considerações Finais

Este capítulo teve o objetivo de apresentar conceitos introdutórios de Pesquisa Operacional e Tomada de Decisão aplicados ao melhoramento genético vegetal. Optou-se por exemplos simples que auxiliem melhoristas a refletirem o quanto a PO pode auxiliar na gestão dos programas e/ou tomadas de decisão que estejam além dos modelos genéticos-estatísticos. A ciência do melhoramento genético está muito bem fundamentada, tanto do ponto de vista teórico quanto operacional, no que tange a estratégia de pesquisa. Portanto, ela se torna um importante campo de aplicação ainda não totalmente explorado para a PO.

A PO é uma área consolidada em diversos setores, incluindo indústrias, serviços, logística, saúde, finanças, energia, meio ambiente, agricultura e transporte, tendo muito potencial para ser bastante aplicada ao melhoramento genético de plantas. No Brasil, poucos profissionais de melhoramento de plantas aplicam a PO no seu dia a dia e, na verdade, a sua maioria nem ao menos têm conhecimento da PO, ou a conhecem por meio de outras ferramentas correlatas (ver Capítulo 1). Portanto, uma primeira perspectiva, necessária, é a incorporação das técnicas da pesquisa operacional na rotina dos programas de melhoramento de espécies alógamas e autógamias no Brasil.

Pode-se elencar outras perspectivas, tanto a curto quanto médio prazo, como por exemplo: i) uso de técnicas avançadas de simulação e análise de dados para modelar e prever o desempenho de plantas em diferentes condições ambientais; ii) aplicação de técnicas de otimização para selecionar as melhores combinações de genes e características para a produção de plantas mais eficientes e adaptáveis; iii) desenvolvimento de sistemas inteligentes para ajudar na tomada de decisão em relação ao melhoramento genético de plantas, utilizando técnicas de inteligência artificial e aprendizado de máquina; iv) uso de tecnologias de edição de DNA para modificar genomas de plantas com mais precisão e rapidez; v) integração de técnicas de pesquisa operacional com outras áreas da biotecnologia, como a biologia sintética e a engenharia genética, para criar plantas mais resistentes e adaptáveis; vi) consideração de aspectos socioeconômicos e de sustentabilidade na seleção de plantas para o melhoramento genético.

Referências

BEUKELAER, H. de. **Discrete optimization algorithms for marker-assisted plant breeding**. 2017. 225 p. Thesis (Doctor of Computer Science) – Faculty of Sciences, Ghent University, Belgium.

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

BORTZ, M.; MAAG, V.; SCHWIENTEK, J.; BENFER, R.; BÖTTCHER, R.; BURGER, J.; VON HARBOU, E.; ASPRION, N.; KÜFFER, K.-H.; HASSE, H. Decision support by multicriteria optimization in process development: an integrated approach for robust planning and design of plant experiments. **Computer Aided Chemical Engineering**, v. 37, p. 2063-2068, 2015. DOI: <https://doi.org/10.1016/B978-0-444-63576-1.50038-8>.

CARNEIRO, A. R. T.; SANGULARD, D. A.; AZEVEDO, A. M.; SOUZA, T. L. P. O.; PEREIRA, H. S.; MELO, L. C. Fuzzy logic in automation for interpretation of adaptability and stability in plant breeding studies. **Scientia Agricola**, v. 76, n. 2, p. 123-129, mar./abr. 2019. DOI: <https://doi.org/10.1590/1678-992X-2017-0207>.

GILMOUR, A. R.; THOMPSON, R.; CULLIS, B. R. Average information REML: an efficient algorithm for variance parameter estimation in linear mixed models. **Biometrics**, v. 51, n. 4, p. 1440-1450, Dec. 1995. DOI: <https://doi.org/10.2307/2533274>.

HILLIER, F. S.; LIEBERMAN, G. J. **Introduction to operations research**. 7th ed. New York: McGraw-Hill, 2001. 1214 p.

KUSKA, M.; WAHABZADA, M.; LEUCKER, M.; DEHNE, H.-W.; KERSTING, K.; OERKE, E.-C.; STEINER, U.; MAHLEIN, A.-K. Hyperspectral phenotyping on the microscopic scale: towards automated characterization of plant-pathogen interactions. **Plant Methods**, v. 11, 28, Apr. 2015. DOI: <https://doi.org/10.1186/s13007-015-0073-7>.

LIU, Y.-H.; DAUER, J. P. A bicriteria linear programming model for determining linear utility functions in simultaneous multiple trait selection and classification. **Applied Mathematical Modelling**, v. 20, n. 8, p. 572-578, Aug. 1996. DOI: [https://doi.org/10.1016/0307-904X\(95\)00003-3](https://doi.org/10.1016/0307-904X(95)00003-3).

MATTHEISS, T. H.; LAND, S. B. A tree breeding strategy based on multiple objective linear programming. **Interfaces**, v. 14, n. 5, p. 96-104, Sept./Oct. 1984. DOI: <https://doi.org/10.1287/inte.14.5.96>.

MILLER, W.; WRIGHT, S. J.; ZHANG, Y.; SCHUSTER, S. C.; HAYES, V. M. Optimization methods for selecting founder individuals for captive breeding or reintroduction of endangered species. In: PACIFIC SYMPOSIUM ON BIOCOMPUTING, 15., 2010, Kamuela, HI. **Proceedings**. p. 43-53. DOI: https://doi.org/10.1142/9789814295291_0006.

MULAMBA, N. N.; MOCK, J. J. Improvement of yield potential of the ETO blanco maize (*Zea mays* L.) population by breeding for plant traits [Mexico]. **Egyptian Journal of Genetics and Cytology**, v. 1, p. 40-51, 1978.

MULLIN, T. J.; BELOTTI, P. Using branch-and-bound algorithms to optimize selection of a fixed-size breeding population under a relatedness constraint. **Tree Genetics & Genomes**, v. 12, 4, 4, Feb. 2016. DOI: <https://doi.org/10.1007/s11295-015-0961-z>.

SAATY, T. L. **The analytic hierarchy process**. New York: McGraw-Hill, 1980. 287 p.

SIMIQUელი, G.; RESENDE, R. T.; RESENDE, M. D. V. Maximizing multi-trait selection gain and diversity with genetic algorithms. **TreeDimensional Journal**, v. 10, e023001, Mar. 2023. DOI: <https://doi.org/10.55746/treed.2023.03.001>.

SIRAGUSA, E.; HAIMINEN, N.; FINKERS, R.; VISSER, R.; PARIDA, L. Haplotype assembly of autotetraploid potato using integer linear programming. **Bioinformatics**, v. 35, n. 18, p. 3279-3286, Sept. 2019. DOI: <https://doi.org/10.1093/bioinformatics/btz060>.

SMITH, H. F. A discriminant function for plant selection. **Annals of Eugenics**, v. 7, n. 3, p. 240-250, Nov. 1936. DOI: <https://doi.org/10.1111/j.1469-1809.1936.tb02143.x>.

TELLES, L. A. A.; ARROYO, J. E. C.; BINOTI, D. H. B.; LORENZON, A. S.; SANTOS, A. R.; DOMINGUES, G. F.; RESENDE, R. T.; MARCATTI, G. E.; GONZALES, D. G. E.; CASTRO, N. L. M.; MOTA, P. H. S.; OLIVEIRA, B. A.; SILVA, M. L. When, where and what cultivate: an optimization model for rural property planning. **Journal of Cleaner Production**, v. 290, 125741, Mar. 2021. DOI: <https://doi.org/10.1016/j.jclepro.2020.125741>.

VOORRIPS, R. E.; GORT, G.; VOSMAN, B. Genotype calling in tetraploid species from bi-allelic marker data using mixture models. **BMC Bioinformatics**, v. 12, 172, May 2011. DOI: <https://doi.org/10.1186/1471-2105-12-172>.

WINSTON, W. L. **Operations research: applications and algorithms**. 4th ed. Boston: Cengage Learning, 2004. 1418 p.

Marcelo Gonçalves Narciso, Claudio Brondani

Introdução

A Internet tem usos variados, não só para obter informações ou acesso a redes sociais, mas também para acesso a aparelhos, relógios, sensores, câmeras, e variados itens, os quais são chamados de “coisas” ou objetos. O termo “Internet das Coisas” ou Internet of Things (IoT) é usado para “coisas” que podem ser acessadas ou controladas pela Internet. Essa tecnologia é empregada em diferentes setores (educação, agricultura, indústria, etc.) e traz uma série de benefícios. Uma característica importante de um dispositivo IoT é que ele pode ser acessado por qualquer pessoa, a qualquer momento, de qualquer lugar que exista acesso à Internet. Segundo o site da revista da Fapesp (Fapesp, 2017), a ideia de conectar objetos à Internet é quase tão antiga quanto a própria rede mundial de computadores. O que mudou nos últimos anos foi o barateamento de tecnologias de microeletrônica e de sensores, e a gigantesca expansão da conectividade. Segundo dados da consultoria Gartner em 2021, havia no mundo 8,4 bilhões de “coisas” conectadas, como Smart TVs, automóveis, sistemas inteligentes de iluminação, equipamentos agrícolas, dentre outros (Gartner, 2021).

A IoT visa fundamentalmente facilitar a vida do ser humano. No domínio agrícola, o uso de sensores para obter dados de solo, clima, monitoramento de máquinas agrícolas e irrigação, e posterior envio de dados para análise para o aparelho celular ou computador do usuário, possibilita o acompanhamento da lavoura a todo instante. Desse modo, um gestor pode tomar decisões baseadas em informações provenientes de um banco de dados completo, aumentando as chances de ganho de eficiência na produtividade. Se os dados de uma fazenda forem capturados, seja por sensores no campo ou de sensoriamento remoto por satélite ou drone, eles podem ser combinados para o planejamento de safras de acordo com o clima esperado, com a cotação

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

de preços presumidos de determinada *commodity* e com a demanda do mercado. Adicionalmente, o planejamento pode incluir o período de semeadura com base no tipo de cultura e clima previsto, quando deverá ocorrer a irrigação, adubação, aplicação de defensivos, e quando proceder à colheita considerando os preços de mercado, previsão do tempo e custos de secagem e armazenamento.

As soluções inteligentes de IoT para agricultura são um sistema construído para monitorar o campo de cultivo com a ajuda de sensores (luz, umidade, temperatura, umidade do solo, sanidade das plantas, etc.). Por exemplo, se o nível de umidade do solo diminuir, o agricultor pode implantar sensores para iniciar a irrigação. A IoT pode fornecer dados relacionados à qualidade do solo em diferentes profundidades, como quantidade de nitrogênio, fósforo e potássio presentes no solo. Uma vez que os dados são coletados, são entregues à nuvem ou servidores de análise de dados por meio de uma rede gateway conectada à Internet por Wi-Fi ou outros meios de comunicação. Posteriormente, os dados são transferidos da nuvem ou servidores analíticos de dados para os dispositivos móveis inteligentes do agricultor ou dispositivos de computação portáteis, onde eles tomam decisões examinando esses dados analisados. Como os agricultores não podem estar fisicamente presentes no campo 24 horas por dia e devido a uma provável falta de habilidade em empregar várias tecnologias para avaliar as condições apropriadas para a colheita, a agricultura inteligente oferece soluções automatizadas que podem ser executadas sem supervisão humana e podem ajudá-los a tomar decisões adequadas para enfrentar os vários desafios que podem encontrar (Yadav et al., 2018; Thilakarathne et al., 2023). Alguns exemplos de aplicação da IoT estão descritos a seguir:

- A partir de sensores instalados em máquinas agrícolas, é possível gerenciar a quantidade de combustível gasta pelos tratores e outros veículos, analisar as rotações dos motores, etc.
- Sensores de imagens (câmeras) podem ser acoplados a drones que sobrevoam as plantações, buscando identificar ocorrências de insetos-praga e doenças. A partir das imagens colhidas, é possível identificar falhas no plantio ou doenças/pragas por meio de índices de vegetação, como o NDVI (*Normalized Difference Vegetation Index*), ou processamento de imagens nos domínios do visível (RGB, *red - green - blue*) e não visível (infravermelho próximo e médio e ultravioleta).
- Sensores de umidade do solo podem obter dados de sua condição hídrica, os quais são transmitidos através de IoT e, mediante sistemas de predição

de quantidade de água para irrigação, pode-se ter previsão da quantidade de lâmina de água necessária para a irrigação de determinada cultivar.

- A partir de dados climáticos, obtidos por meio de sensores, e transmitidos através de IoT, pode ser determinado se uma planta está com estresse hídrico ou não e, se for o caso, tomar medidas de manejo apropriadas para que a lavoura não tenha queda na produtividade.
- Seleção automatizada: Dados coletados por sensores IoT podem ser usados para selecionar automaticamente as plantas com as melhores características desejadas, como resistência a doenças, alta produtividade, ou adaptação ao ambiente. Isso pode ser feito por meio de algoritmos de inteligência artificial que analisam os dados coletados e selecionam as plantas com as melhores características.
- Monitoramento de progresso: Os sensores IoT podem ser usados para monitorar o progresso dos ganhos do melhoramento genético ao longo do tempo. Isso pode incluir o rastreamento de características de plantas individuais, bem como o progresso geral de um grupo de plantas ao longo de várias gerações.
- Recomendação de cultivares comerciais: Os sensores IoT podem ser usados para recomendar cultivares comerciais com base nas características das plantas. Isso pode incluir a utilização de algoritmos de inteligência artificial que analisam os dados coletados e recomendam as cultivares com as melhores características para determinadas condições climáticas e de cultivo.
- Descarte de materiais genéticos: Com base em características fenotípicas e genotípicas coletadas por sensores IoT, pode ser possível descartar materiais genéticos que não atendem aos critérios estabelecidos para melhoramento genético.
- Monitoramento da qualidade dos grãos: Sensores IoT podem ser usados para monitorar a qualidade dos grãos, como o teor de umidade, a presença de impurezas e a qualidade nutricional. Isso pode ser útil na seleção de plantas para o melhoramento genético de plantas com grãos de alta qualidade.
- Monitoramento de nutrição das plantas: Sensores IoT podem ser usados para monitorar a nutrição das plantas, incluindo o teor de nutrientes no solo e a sua absorção pelas plantas. Isso pode ser útil na seleção de plantas para o melhoramento genético de plantas que são mais eficientes na absorção de nutrientes.

IoT Aplicada ao Melhoramento Genético de Precisão

No melhoramento de plantas, a implementação da IoT visa aumentar a sua eficiência, fornecendo soluções que economizam tempo e que sejam economicamente viáveis. O uso de IoT no melhoramento está ligado ao uso de dispositivos e sensores de IoT para coletar dados e processá-los automaticamente, facilitando a tomada de decisão pelo melhorista. Como exemplo de aplicação, tem-se a automação de casas de vegetação para o controle da irrigação e fotoperíodo, além do envio das condições ambientais em tempo real via computação na nuvem. Sensores podem coletar dados de temperatura, fertilidade do solo e sanidade das parcelas experimentais, e também podem ser usados em combinações com outras soluções, como GPS, imagens aéreas e de satélite. Essas abordagens visam aumentar a precisão experimental, prever a produtividade, além de reduzir custos. O acúmulo de grande volume de dados (*Big Data*) permite que sejam realizadas extrapolações por meio da inteligência artificial, prevenindo-se o desempenho de linhagens em ambientes não avaliados e, com isso, aumentar as chances de sucesso de futuras cultivares comerciais. Em geral, os dados fenotípicos da planta têm sido coletados usando ferramentas de medição manuais tradicionais, mas esse método é demorado, ineficiente e tem baixa precisão, limitando assim a eficiência da seleção genética e, por consequência, estudos básicos de genômica que envolvem a identificação de genes de interesse. Atualmente, há grande necessidade de que pesquisadores possam realizar detecção rápida, precisa e sem erros de parâmetros de caracteres de plantas para entender suas condições de crescimento e desenvolvimento (Fan et al., 2021).

A fenotipagem precisa de plantas e a obtenção de dados climáticos estabelecem a base da análise de interação genótipo x ambiente (GxE), tão importante para a seleção genética de linhagens superiores em programas de melhoramento genético. A IoT permite automatizar tanto a fenotipagem quanto a obtenção dos dados climáticos. Além das vantagens da obtenção desses dados em tempo real, a IoT pode resolver uma questão importante para o melhoramento: a falta de mão de obra intensiva e treinada para as tarefas do dia a dia (Ali et al., 2018). Supondo que houvesse essa disponibilidade, os custos operacionais seriam muito elevados. Com o aprimoramento do desenvolvimento tecnológico nos últimos anos, ainda é tecnicamente desafiador calibrar, anotar e agregar *Big Data* de forma eficaz, especialmente quando os dados foram gerados em vários locais e, muitas vezes, em diferentes escalas (Reynolds et al., 2019). A maximização do desempenho das cultivares conta com o auxílio da metodologia de *Machine Learning* (ML), que é uma maneira inteligente dos computadores em simular atividades de

aprendizado das pessoas, adquirirem novos conhecimentos e melhorar continuamente o desempenho. A tecnologia de ML está desempenhando papel importante no melhoramento genético, identificação de doenças, previsão da ocorrência de insetos-praga e doenças, planejamento inteligente de irrigação, dentre outros. Por exemplo, a tecnologia ML pode analisar dados anteriores em determinada estação experimental, incluindo o desempenho de culturas sob diferentes condições climáticas e a herança de um determinado fenótipo. Com isso, a tecnologia ML pode explorar a associação genótipo:fenótipo e então construir um modelo de probabilidade para prever quais genes são mais propensos a participar da expressão de um importante caractere de interesse, o que pode ajudar geneticistas e melhoristas a realizarem melhoramento genético com elevada precisão (Shi et al., 2019).

O processo para uso no melhoramento genético se baseia na obtenção de dados e transmissão pela internet. Dispositivos instalados em parcelas experimentais devem coletar e processar os dados em um ciclo repetitivo que permita aos melhoristas avaliar o desempenho das linhagens em tempo real, seguindo ciclos que envolvem a observação (por sensores), diagnóstico (após um problema ser detectado por sensores em uma plataforma IoT, são identificados quais problemas específicos estão ocorrendo), e decisões/ações (sobre o tratamento específico necessário a ser adotado, ou simplesmente não adotar procedimento visando a seleção de genótipos superiores). Essas tecnologias podem acompanhar o processo seletivo das melhores linhagens com muita precisão, em tempo real, ao longo do ciclo da cultura. Um exemplo de módulo integrado de grande importância para a avaliação de genótipos é o monitoramento do clima, água e solo (Ali et al., 2018). Para o monitoramento do clima, um sensor DHT 22 (comentado mais adiante neste capítulo) mede a temperatura e a umidade do ambiente e, se ultrapassarem determinado limite, essa informação pode ser enviada ao melhorista. O mesmo sensor DHT 22 pode monitorar a umidade do solo, a fim de orientar a entrada da irrigação. Experimentos que envolvam diferentes quantidades de água precisamente fornecidas nas parcelas, após a detecção da necessidade de irrigação pelo sensor, são fundamentais para a seleção de genótipos tolerantes à deficiência hídrica (Dhanaraju et al., 2022).

Por meio da IoT aplicada ao melhoramento genético, ao associar as condições ambientais com os dados obtidos ao longo do desenvolvimento das plantas em um experimento, assegura-se que o sistema é capaz de fornecer uma base sólida para estudos GxE, como mencionado anteriormente. Adicionalmente, por meio da geolocalização, leituras de sensores remotos e transmissão automática de dados, em uma série histórica, pode-se fazer referência cruzada envolvendo vários locais e anos, tornando o processo de desenvolvimento e seleção de linhagens muito mais

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

preciso. O acesso em tempo real permite que melhoristas possam tomar decisões rápidas, levando em conta estágio de desenvolvimento dos genótipos em uma rede nacional de experimentos, além de poder decidir sobre tratamentos culturais e necessidade de irrigação, por exemplo. Além disso, a transmissão automática de dados permite a centralização dos dados e o gerenciamento de experimentos, o que significa que o sistema pode ser ampliado para a escala nacional. Por meio da interpolação de dados a partir de pontos específicos onde foram montadas estruturas IoT pode-se modelar a variação ambiental em todo o território nacional, auxiliando no processo de seleção de linhagens, futuras cultivares comerciais, que sejam específicas para determinada área geográfica (Reynolds et al., 2019).

Embora muitas plataformas automatizadas de fenotipagem sejam capazes de gerar grande volume de dados coletados de plantas e do ambiente, é ainda tecnicamente desafiador coletar, calibrar, anotar e agregar esses conjuntos de dados de forma eficaz, especialmente para experimentos realizados em vários locais e em diferentes escalas. Soluções comerciais já estão disponíveis envolvendo hardware e plataforma de processamento de dados que combinam coleta de imagens, associações de metadados e análise fenotípica para ambientes controlados. No entanto, esses sistemas comerciais requerem manutenção do licenciamento e custos adicionais para o desenvolvimento de novas funções (Reynolds et al., 2019). Portanto, é um desafio para a pesquisa desenvolver uma plataforma flexível e que atenda necessidades específicas de cada programa de melhoramento. Por esse motivo, este capítulo aborda a partir deste momento alguns conceitos para o desenvolvimento de uma plataforma básica que emprega IoT tanto para a pesquisa, quanto para o monitoramento de áreas de lavouras.

Componentes mais Comuns Usados em IoT

Para a aquisição de dados ou acionamento de algum dispositivo são usados sensores, câmeras, atuadores e outros componentes eletrônicos. Esses dispositivos são acionados ou controlados por alguma placa eletrônica, com software embarcado, com a finalidade de obter os dados e os enviar a algum servidor remoto para que determinado aplicativo possa exibir os resultados, ou ainda, acionar algum dispositivo, como um aparelho de ar condicionado, por exemplo.

Um sensor, basicamente, é um dispositivo que tem a função de detectar e responder com eficiência algum estímulo. Existem vários tipos de sensores para serem usados na agricultura, assim como em ambiente doméstico, na indústria, no comércio, em

hospitais, ruas, etc. Para obter dados climáticos sobre algum lugar são usados sensores de temperatura do ar, umidade do ar, pressão atmosférica, radiação solar, velocidade do vento, precipitação (mm de chuva), etc. O sensor de temperatura e umidade do ar DHT-11, por exemplo, pode ser acessado por meio de uma placa Arduino (Arduino, 2023), usada para acionar e ler o sensor, além de mostrar o resultado, tal como ilustra o circuito eletrônico da Figura 1. A placa Arduino possui um software embarcado que possibilita o acesso aos dados do sensor no tempo que for programado. Um *website* que explica como fazer a montagem e o código embarcado pode ser acessado em Makerhero (2013).

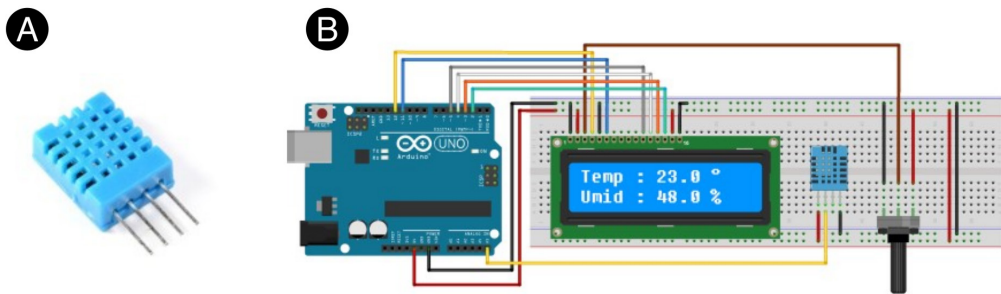


Figura 1. A) Sensor de temperatura e umidade do ar modelo DHT11. B) Placa Arduino, modelo Uno, à esquerda, usada para fornecer energia e acessar os dados obtidos pelo sensor DHT11.

Fonte: Makerhero (2013).

Outros sensores que podem ser citados são os de distância, presença, temperatura da água, pressão atmosférica, acelerômetro e vários outros (Figura 2). Esses sensores necessitam de alguma placa para fazer a leitura dos dados, como a Arduino, Raspberry Pi (Rasp-Pi, 2023), a ser vista mais adiante neste capítulo, Banana Pi (Banana-Pi, 2023), Orange Pi (Orange-Pi, 2023), dentre outras. A escolha de uma dessas placas depende da aplicação a que se destina.

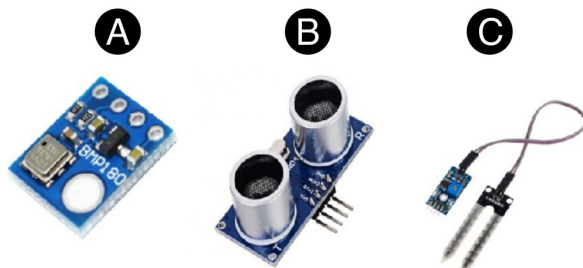


Figura 2. Exemplos de sensores de pressão atmosférica (A), distância (B), e umidade do solo (C).

Fonte: Baú da Eletrônica (2023).

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

A placa Arduino pode ser usada para o caso em que se deseja controlar sensores, motores de baixa corrente e tensão, válvulas, relés, etc. Um microcontrolador consiste em um único circuito integrado que reúne um núcleo de processador, memórias voláteis e não voláteis e diversos periféricos de entrada e de saída de dados (Achei Componentes, 2021). O microcontrolador do Arduino tem pouco espaço de memória e pouca velocidade, em comparação a processadores de outras placas como a Raspberry Pi 3 ou 4, ou similares, como Orange Pi 4 ou Banana Pi 4. Para exemplificar, considere o microcontrolador do Arduino Uno, Atmega 328P-U, o qual alcança 1 MIPS/MHz (1 milhão de instruções por segundo por megahertz), possui memória flash de 32 Kb, memória SRAM de 2 Kb e memória EEPROM de 1 Kb. O *clock* tem valor igual a 20 MHz e, com esse valor, pode-se ter até 20 MIPS (20 milhões de instruções por segundo). A memória tem pouco espaço e assim uma placa Arduino deve ser usado em situações nas quais não se necessita de muita memória RAM ou quando o tempo de processamento a ser exigido não seja tão importante.

A vantagem da placa Arduino em relação às placas com poder de processamento maior e memória RAM acima de 512 MB, como a Raspberry Pi 3, por exemplo, é o baixo custo, além da simplicidade de se programar e também à quantidade de potência a ser fornecida à placa, que é menor em comparação a uma placa Raspberry Pi 3 ou 4, por exemplo.

Wi-Fi é uma tecnologia de rede sem fio que permite que computadores (laptops e desktops), dispositivos móveis (como smartphones) e outros equipamentos (impressoras e câmeras de vídeo) se conectem à Internet. O Wi-Fi permite que esses e muitos outros dispositivos troquem informações entre si, criando uma rede. Contudo, mesmo que não haja sinal Wi-Fi, a transmissão de dados de um ponto a outro pode ser feita através do protocolo LoRa ou XBEE. Para locais onde exista Wi-Fi, podem ser usadas placas como a Arduino uno Wi-Fi Rev 2 (Embarcados, 2018) (Figura 3 A).

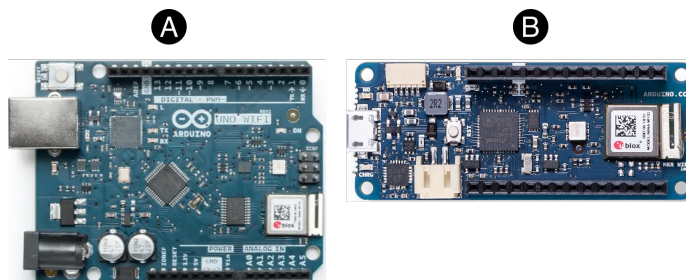


Figura 3. A) Placa Arduino Uno Wi-Fi. B) Placa Arduino MKR Wi-Fi.

Fonte: Fig. A: Arduino Uno Wi-Fi rev 2 (2023).

Fig. B: Arduino MKR WiFi 1010 (2021).

Outra placa, feita exclusivamente para IoT, é a Arduino MKR1010 (2021) (Figura 3 B). Os dados de sensores que forem adicionados à placa Arduino MKR1010 podem ser visualizados por meio de aplicativo para celular ou computador. Existem aplicativos prontos que podem ser acessados por meio do *website* da empresa Adafruit (2019) ou na página *web* do Arduino (Cloud, 2023).

Existem outras opções, similares ao Arduino, que contemplam o uso do Wi-Fi. Alternativamente, pode-se usar a placa ESP8266, que pode ser conectada à placa Arduino. Na Figura 4 são mostradas as placas Wemos (similar à placa Arduino Uno, mas com Wi-Fi) e a ESP8266 para ser conectada à placa Arduino (a qual não contém Wi-Fi).

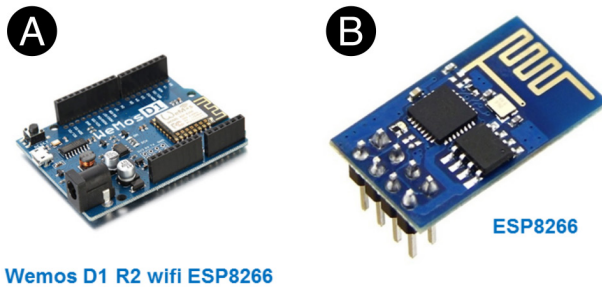


Figura 4. Placa Wemos com ESP8266 e placa ESP8266.

Fonte: Wemos (2023).

Um modo de conectar uma placa Arduino sem Wi-Fi com a placa ESP8266 está descrito na Figura 5, a qual indica uma alternativa para o envio de dados obtidos por sensores de temperatura (LM35) e de pulso acoplados a um hardware ou circuito.

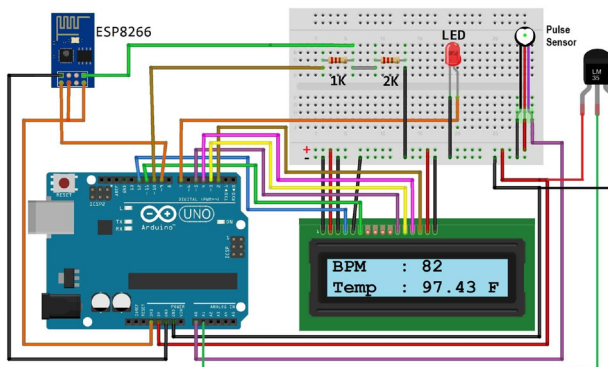


Figura 5. Placas Arduino Uno e ESP8266 para envio de dados coletados de sensores.

Fonte: Systemcap (2021).

Protocolo de Comunicação LoRa

Há situações onde não há Wi-Fi disponível para acesso à Internet e, considerando que a conectividade é a espinha dorsal da Internet das Coisas, é necessária outra forma de transmissão de dados. A tecnologia LoRaWAN, nome dado ao protocolo que define a arquitetura do sistema, bem como os parâmetros de comunicação usando a tecnologia LoRa, possibilita levar os dados dos sensores até onde possa ter um Wi-Fi e assim os dados serem acessíveis para gerar informação ao usuário. Atualmente existem inúmeras tecnologias em comunicação, com ou sem fio, usadas em aplicações IoT. Cada uma dessas tecnologias tem suas características, vantagens e desvantagens, e nenhuma delas pode ser dita como a melhor tecnologia, mas apenas como a mais adequada para uma aplicação específica. No contexto da agricultura, por exemplo, tecnologias LPWAN (*Low Power Wide Area Network*) são adequadas para aplicações que não requerem altas taxas de transmissão, são tolerantes a atrasos e possuem baixo consumo de energia quando comparadas a outras tecnologias de longa distância, como 3G/4G e Zigbee (Figura 6). Observe-se que LPWAN tem maior alcance de sinal, cobertura geográfica e latência de transmissão, enquanto que a tecnologia 3G/4G tem maior valor quanto ao consumo de energia e largura de banda.

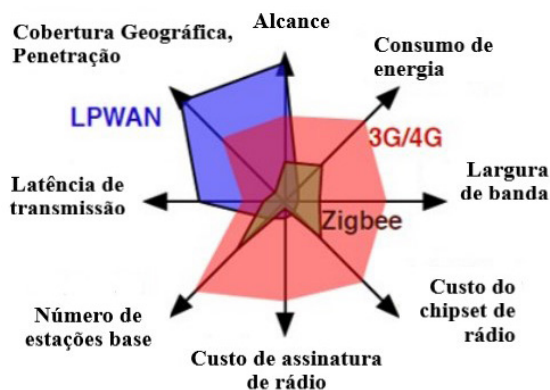


Figura 6. Comparação entre LPWAN e outras tecnologias sem fio (*wireless*). As áreas marrom, rosa e azul indicam o alcance em cada eixo (largura de banda, custo do chipset de rádio, etc.).

Fonte: Redes-IOT (2023).

LoRa (*Long Range Radio*) é a camada física de uma tecnologia de comunicação sem fio que pode atuar como parte de uma rede LPWAN. Seu esquema proprietário de modulação por espalhamento espectral é derivado da modulação CSS (*Chirp Spread Spectrum*). Redes com LoRa preenchem a lacuna tecnológica das redes celulares, Wi-Fi e BLE (*Bluetooth Low Energy*), que demandam alta largura de banda, podem

ter alcance limitado ou serem incapazes de penetrar ambientes. Isso a torna bastante atrativa para uma ampla variedade de setores, em especial para propriedades rurais, onde geralmente o sistema está localizado em áreas remotas, sem amplo acesso a redes de telefonia móvel.

A tecnologia LoRa é aplicada na rastreabilidade de animais, monitoramento da produção de leite, aquisição de dados climáticos e de solo em casa de vegetação, agricultura de precisão, monitoramento do solo, monitoramento de plantio e colheita, aquisição de dados climáticos no campo, irrigação inteligente, etc. Mais informações sobre aplicações de LoRa na agricultura ou pecuária podem ser obtidas em Mokolora (2023).

Em termos de hardware, há a opção de utilizar um dispositivo que já contenha um microprocessador a ser programado e que fique responsável pelo software de transmissão de dados. Há também a opção de usar um dispositivo que não contenha um microprocessador embutido, fazendo-se necessária a programação da transmissão dos dados em um outro microcontrolador ou microprocessador, como os já citados Arduino e Raspberry Pi.

No primeiro caso, pode-se citar como exemplo o ESP32 Wi-Fi LoRa 32 (Usinainfo, 2018), da empresa Heltec (Heltec, 2018), um dos dispositivos LoRa mais populares. Esse módulo conta com microprocessador ESP 32 (MCU dual-core de 32 bits mais um núcleo ULP), com chip de nó LoRa SX1276/SX1278 e regulador de tensão completo embutido, Wi-Fi integrado, três conexões de rede Bluetooth e é compatível com o ambiente de desenvolvimento Arduino (Figura 7 A).

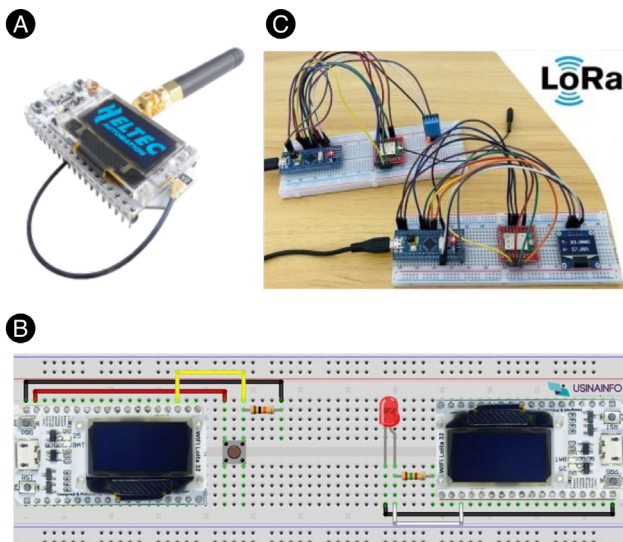


Figura 7. A) Módulo ESP32 Wi-Fi LoRa 32 (V2) da Heltec. B) Aplicação de comunicação LoRa utilizando dois módulos ESP32 Wi-Fi LoRa 32 (V2) da Heltec. C) Circuito de envio de dados (placa superior) e circuito de recepção de dados.

Fonte: Fig. A: Heltec (2018), Fig. B: Bauermeister (2018) e Fig. C: How to Electronics (2023).

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

Um exemplo de aplicação com essa placa de desenvolvimento pode ser visto na Figura 7 B. À esquerda, um dos módulos, denominado *Sender*, está conectado a um *push button* (botão) e a um resistor de 10 K ohms. O fio amarelo ilustra a conexão com a GPIO 12. O outro módulo, denominado *Receiver*, está conectado a um LED na GPIO 12 (ilustrada pelo fio branco) e um resistor de 150 ohms. Mais informações podem ser obtidas em Bauermeister (2018).

Além do protocolo LoRa, existe outro muito usado em IoT, o MQTT (*Message Queuing Telemetry Transport*). É um protocolo de transporte de mensagens de formato Cliente/Servidor, que é executado em ambiente TCP/IP, e que possibilita a comunicação entre máquinas (*Machine to Machine* – M2M). Esse protocolo é muito usado em ambientes onde exista a comunicação entre computadores ou placas (Arduino ou Raspberry Pi, por exemplo) por cabeamento ou Wi-Fi. Possui várias aplicações de IoT na indústria automotiva, manufatura, petrolífera e, também, em residências, telecomunicações e comércio (Kalatec, 2023). O protocolo MQTT, que pode ser utilizado em conjunto com o LoRa, facilita a configuração dos sensores, atuadores e, também, o envio de requisições aos sensores e recebimento de respostas (mensagens ou resultados). Mais informações sobre o protocolo MQTT podem ser obtidas em Hivemq (2023).

Exemplo de Aplicação na Agricultura

Neste tópico será apresentado o exemplo de um sistema IoT que obtém dados de campo de temperatura e umidade do ar, e que depois envia os resultados sobre o risco de aparecimento da antracnose (doença que ataca o feijoeiro) em função desses dados climáticos, para ser visualizado em um *website* ou smartphone. No exemplo, serão considerados todos os municípios do Estado de Goiás.

Para o sistema de alerta para predição do aparecimento da doença antracnose em plantações de feijão, será considerado um conjunto de regras em função da temperatura (T) e da umidade relativa do ar (UR), quais sejam:

Se $15^{\circ}\text{C} < T < 25^{\circ}\text{C}$ e $UR > 70\%$, então a probabilidade de aparecimento é muito favorável.

Se $15^{\circ}\text{C} < T < 25^{\circ}\text{C}$ e $50\% < UR < 70\%$, então a probabilidade de aparecimento é favorável.

Se $25^{\circ}\text{C} < T < 30^{\circ}\text{C}$ e $UR < 70\%$, então a probabilidade de aparecimento é pouco favorável

Qualquer outro intervalo de valores, além dos que foram citados acima, indicará que a probabilidade de aparecimento de antracnose é nula.

Para coletar dados para as regras acima é necessário um sensor para obter a temperatura e umidade relativa do ar. Para isso podem ser utilizados os sensores DHT 22 ou DHT 11, esse último citado anteriormente. Esse sensor deverá estar ligado a uma placa Arduino e a uma placa LoRa SX1276, conforme pode ser visto em (How to Electronics, 2023). Para este projeto, será utilizado o Arduino modelo nano, uma placa Arduino mais barata e com menos possibilidades de conexão de componentes eletrônicos, visto que serão obtidos apenas os dados dos sensores e realizadas as avaliações de temperatura e umidade (Figura 7 C).

Na Figura 7C, o circuito superior contém o sensor, que irá coletar os dados de temperatura e umidade relativa do ar, e enviará os dados coletados para o circuito inferior. Na prática, o circuito na parte superior da Figura 7C estará no campo e o circuito da parte inferior que irá receber esses dados estará no escritório, a quilômetros de distância. Após o circuito inferior ter recebido os dados, eles são enviados a uma base de dados, por Wi-Fi, ou os dados são enviados para um local, via placa LoRa SX1276 (análoga à placa ESP 32), onde tenha Internet. A partir daí, os dados são enviados à base de dados no ambiente *web*. A base de dados é então acessada por um sistema de software que obtém os valores de temperatura e umidade do ar para então fazer o cálculo, cujo resultado será acessado pelo usuário no próprio *website* ou por meio de aplicativo de smartphone. Imagine agora que cada cidade do Estado de Goiás tenha um cálculo para a probabilidade de aparecimento de antracnose. A partir dos dados coletados em cada cidade, em tempo real, um mapa de risco de aparecimento de antracnose para o Estado de Goiás é então gerado (Figura 8).

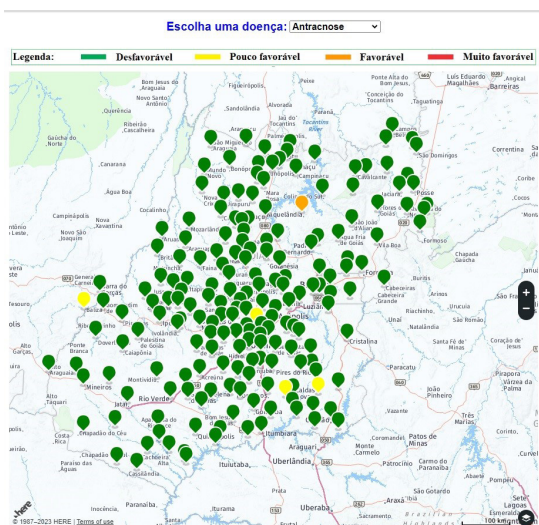


Figura 8. Mapa do risco de aparecimento da antracnose no Estado de Goiás em 07/01/23.

Fonte: Infodoenças (2023).

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

Observe que esse sistema de alerta de risco de antracnose pode ser visto a qualquer momento, em qualquer lugar, por qualquer pessoa, tal como está ilustrado na Figura 8. De forma similar, outros sistemas podem ser construídos, usando basicamente a forma de transmissão de dados pelo protocolo LoRa. As mudanças do projeto seriam nos sensores, para obter os dados desejados, no software de análise, e no acesso a essas informações por meio de um aplicativo com funcionalidades específicas.

Existem demandas que necessitam de maior poder de processamento, para as quais uma placa Arduino (Uno, Nano, Mega) não seria adequada. Um exemplo seria a captura de imagens de determinado objeto para seu posterior processamento. Para esse caso, placas Raspberry Pi, Orange Pi, Banana Pi ou similares seriam mais adequadas. Essas placas possuem versões de 2 Gb, 4 Gb, 6 GB ou 8 Gb de RAM e também CPU de 1.4 GHz ou superior. Uma placa Raspberry Pi modelo 4, por exemplo, pode ter 8 Gb de RAM e 1.5 GHz com 4 núcleos, e, portanto, é possível utilizá-la sem dificuldades para processamentos que exijam mais CPU e RAM.

Um exemplo de aplicação é a detecção de insetos-praga em armadilhas por meio do processamento de imagens, como a presença do inseto mosca-branca (*Bemisia tabaci*) em lavouras de feijão e tomate, cuja tecnologia está em desenvolvimento na Embrapa Arroz e Feijão. Existe uma armadilha já validada para vários tipos de insetos e que consiste de um papel colante amarelo. Desse modo, é possível monitorar a presença desses insetos nas armadilhas de papel amarelo por meio da tomada e processamento de imagens. Caso seja detectada a presença do inseto-alvo, um sinal de alerta é enviado ao agricultor para que ele possa tomar as decisões sobre o combate a essa praga (Figura 9). Em cada armadilha (contendo uma placa amarela) existe uma câmera acoplada e controlada por uma placa Raspberry Pi (Figura 10). Admitindo que o processamento da imagem para detectar a mosca-branca na armadilha seja feito nessa placa, o resultado será então enviado para um local onde há sinal de Internet através do ESP 32, tal como foi feito para o sistema de alerta para antracnose, descrito anteriormente.

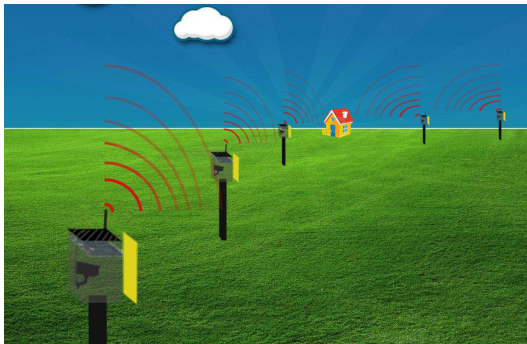


Figura 9. Esquema de armadilhas para monitoramento e detecção de insetos-praga utilizando IoT.

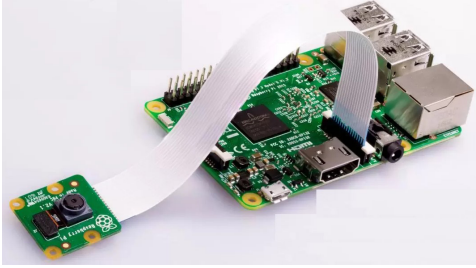


Figura 10. Placa Raspberry Pi com uma câmera acoplada para monitoramento da armadilha.

Fonte: Rasp-PI (2023).

Para uma placa Arduino, ESP 32, Raspberry Pi ou similar poderem controlar ou acionar algum sensor, é necessário que exista um software embarcado. Para as placas Arduino ou ESP 32 existe uma interface com o usuário chamada de Arduino IDE, que pode ser baixada a partir da página <https://www.arduino.cc/en/software>. Com esse software é possível programar o Arduino ou ESP 32 para diversos fins. Existe um cabo que liga a placa Arduino ou placa ESP 32 ao microcomputador, permitindo que a programação realizada na interface Arduino IDE seja carregada no microcontrolador do Arduino ou ESP 32 (Figura 11).

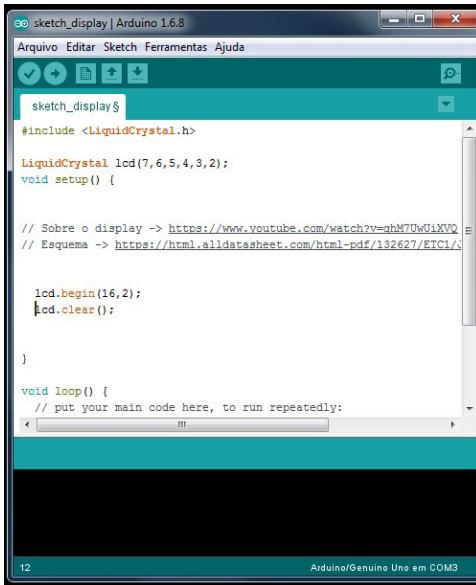


Figura 11. Interface de programação para Arduino ou ESP 32.

Fonte: Arduino (2023).

A programação para Arduino, ESP 32 ou similar é feita em uma linguagem baseada em C++. Para alguém iniciar na programação para Arduino não é necessário conhecer detalhadamente essa linguagem de computador, mas uma parte dela, isto é, conhecer

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

as funcionalidades envolvidas no acesso a sensores, motores, atuadores, etc. Não são muitas as funcionalidades necessárias e, assim, após a consulta a alguns exemplos para implementação de circuitos, o aprendizado da linguagem será facilitado.

Para o caso de programas para as placas Raspberry Pi ou similar, faz-se necessário saber alguma linguagem de programação, como C, C++, Java, Python ou outra que contenha funcionalidades para acessar os sensores e obter os dados necessários para que os mesmos sejam manipulados e convertidos em informação útil ao usuário. O Sistema operacional mais comum para a placa Raspberry Pi é uma distribuição Linux (Debian, Ubuntu, Raspbian). As placas Orange Pi e Banana Pi têm recomendações de quais distribuições Linux o hardware suporta. Na Internet existem programas prontos para várias linguagens aplicáveis a essas placas, como para Python (Gpiozero, 2023), Java (Makerhero, 2018) e C++/C (Raspberrypi, 2023).

Considerações Finais

A Internet das Coisas foi abordada neste capítulo sob o ponto de vista do hardware e aplicações, para facilitar o entendimento do leitor. Os casos enfocados são a base para fazer outros sistemas de IoT. Não foram destacados detalhes relativos à parte de eletrônica envolvida e nem o software embarcado, visto que seriam necessários vários capítulos para isso. Porém, o leitor pode facilmente encontrar cursos de programação de Arduino na Internet e, também, livros com exemplos diversos de como fazer circuitos eletrônicos. Os componentes de hardware podem ser obtidos tanto em lojas físicas como no exterior, por meio de compras na Internet.

A agricultura digital – ou seja, usar tecnologia e dados digitais para conduzir processos de tomada de decisão – pode ajudar todo o setor agrícola a ser mais eficiente, transparente, produtivo e lucrativo. Sensores no campo, combinados com equipamentos agrícolas automatizados e dados de drones e satélites, podem fornecer melhores orientações aos agricultores. Melhores dados e compreensão dos riscos da operação agrícola podem ajudar a melhorar o direcionamento de insumos. Algumas das principais tendências na agricultura digital incluem sensores (para medir dados climáticos, de solo, etc.), a conexão de sensores em plataformas IoT para monitoramento em tempo real, automação para o desenvolvimento de novas aplicações em semeadura, aplicação de produtos químicos, irrigação e capina. Veículos autônomos equipados com câmeras de alta resolução monitoram continuamente as plantações à medida que crescem por meio de um processo conhecido como fenotipagem rápida, que promete aumentar o ritmo de desenvolvimento de novas cultivares comerciais. O

sensoriamento remoto, que na agricultura geralmente se refere à análise de imagens de dados de satélite, existe há várias décadas. As novas tendências se concentram no uso de satélites em órbita baixa para acessar imagens com mais frequência ou no uso de imagens de alta resolução de drones - voando sob as nuvens e disponíveis sob demanda - para monitorar o estado sanitário das plantações ou surtos de insetos-praga e doenças. Os algoritmos estão melhorando rapidamente para aproveitar as maiores quantidades de imagens e fotos disponíveis. Um modelo de IA analisando uma fotografia de uma folha, por exemplo, pode identificar rapidamente insetos-praga e doenças específicas de plantas e fazer recomendações de tratamento em tempo real (Chandra; Collis, 2021).

O desenvolvimento de sensores deve se concentrar em novos materiais sensíveis, mecanismos, processos e metodologias, bem como baixo consumo de energia e baixo custo. Novos sensores permitirão que uma gama ainda mais abrangente de situações e eventos sejam detectados no campo. Os sensores atuais cairão de preço para tornarem-se mais acessíveis, e novos algoritmos surgirão para extrair e deduzir mais informações das tecnologias de sensores atuais (Planin, 2023). Com pesquisa colaborativa contínua em tecnologia de sensores, tecnologia de comunicação, ciência de plantas, ciência da computação e engenharia, a IoT na agricultura permitirá um salto tecnológico por meio de redes de sensores ambientais, tomada de imagens não destrutiva, análise espectral, robótica, dentre outros avanços. Com base em dados de características de plantas multiescala, multiambientais e multimodo que constituem *Big Data* de fenótipos de plantas, a relação entre genótipo, fenótipo e ambiente em análises de genoma poderão ser exploradas em detalhe (Fan et al., 2021).

Com relação à rede de transmissão de dados, devemos nos concentrar em dados de alta capacidade, tecnologias de comunicação de banda larga em tempo real para aplicações agrícolas complexas, que tenha alta confiabilidade, adaptável, e possua estratégias de gerenciamento de algoritmos (Shi et al., 2019). Com a tecnologia 5G é esperado crescimento de aplicações com IoT, devido ao maior alcance de sinal e aumento da velocidade de transmissão de dados. Isso possibilitará maior quantidade de sensores e atuadores no campo para facilitar o gerenciamento de colheitas, dados climáticos e análise de solos em tempo real, mapas de irrigação e semeadura, e monitoramento e controle remoto de máquinas e equipamentos (Tecnologia-5G, 2023). O aumento da cobertura móvel também permitirá amplo uso de smartphones, acesso à informações, uso de aplicativos e fluxos de dados bidirecionais (Chandra; Collis, 2021). Os desenvolvedores de software devem gerar grandes bancos de dados de código aberto e bibliotecas de algoritmos de processamento de sinal para diferentes áreas de agricultura de instalações. Para

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

universidades e institutos de pesquisa, as principais direções de trabalho para o futuro incluirão a agricultura inteligente baseada em nuvem, modelos de tomada de decisão, algoritmos de fusão de informações com base em dados de várias fontes, *Big Data* agrícola, tecnologias de mineração, sistemas de processamento inteligente, e autenticação por mecanismos de criptografia (Shi et al., 2019). Investimentos em segurança da informação em IoT deverão continuar a crescer. A ideia da ampla conexão é fascinante, mas os desafios de manter um sistema seguro são complexos e exigem investimentos em treinamento pessoal e equipamentos. Os dispositivos IoT geram grande quantidade de informações e elas devem ser armazenadas e o acesso deve ser restrito a quem realmente tem privilégio (Santos; Sales, 2015). A geração de empregos para profissionais da área de IoT para a área agrícola será cada vez maior. É uma grande oportunidade para profissionais de Engenharia, Tecnologia da Informação e áreas afins. Essas inovações tecnológicas também podem levar a novos modelos de negócios. Para que as tecnologias digitais sejam adotadas amplamente, mais pesquisas precisam ser conduzidas. Por exemplo, o desenvolvimento de sensores e hardware de baixo custo, novas técnicas de IA que permitam combinar diferentes dados para gerar resultados preditivos e novas interfaces de usuário usando técnicas de baixo custo permitirão que a agricultura digital seja acessível para todos os agricultores. A fim de lidar com um ambiente de cultivo cada vez mais complexo pelas mudanças climáticas, o melhoramento genético contará com o contínuo aperfeiçoamento da IoT. Dispositivos de hardware devem ser totalmente atualizados para aumentar ainda mais sua universalidade, confiabilidade, expansibilidade e resistência, reduzindo custos e dificuldades operacionais, além da otimização de modelos de suporte à decisão de *Big Data* (Shi et al., 2019).

O custo de desenvolvimento e implantação de dispositivos conectados diminuirá a cada ano e as alternativas de conectividade e integração dos dados irá aumentar, o que torna a automação do processo de coleta de dados no campo mais acessível, com baixo custo de manutenção para os programas de melhoramento e para os agricultores (Sigmais, 2023). O aumento do uso de conectividade LPWA (*Low Power Wide Area*), considerando a sua utilização nas propriedades rurais, trará o benefício do baixo custo de comunicação comparada com as alternativas móveis de dados e, também, ajudarão na redução de custo de infraestrutura, já que o dispositivo possui conectividade embarcada e bateria, sem a necessidade de ser energizado (Sigmais, 2023). Com o avanço da conectividade no campo, no futuro se terá muitas máquinas conectadas e enviando dados pela Internet a todo instante. As funções dos sensores agrícolas estão se tornando cada vez mais diversificadas, incluindo sensores de solo, sensores meteorológicos, sensores de água e sensores de plantas.

Plataformas de medição de fenótipos devem ser melhoradas e inovadas, ao mesmo tempo em que os dados massivos (*Big Data*) podem ser gerenciados e analisados para transformar fenótipos de linhagens superiores em modelo de parâmetros ao longo do ciclo de desenvolvimento de uma espécie vegetal de interesse. Com isso, é de se esperar que, em um futuro próximo, devido ao aprimoramento de sensores e ao progresso de ciência da computação, essas plataformas de medição de parâmetros fenotípicos multidimensionais introduzam uma nova era de fenômica e melhoramento.

Referências

- ACHEI COMPONENTES. **Você sabe o que é um microcontrolador e para que ele serve?** 2021. Disponível em: <https://www.acheicomponentes.com.br/loja/noticia.php?loja=648216&id=11>. Acesso em: 20 mar. 2023.
- ADAFRUIT. **Adafruit IO basics**. 2019. Disponível em: <https://learn.adafruit.com/series/adafruit-io-basics>. Acesso em: 10 jan. 2023.
- ALI, T. A. A.; CHOKSI, V.; POTDAR, M. B. Precision agriculture monitoring system using Internet of Things (IoT). **International Journal for Research in Applied Science and Engineering Technology**, v. 6, n. 4, p. 2961-2970, Apr. 2018. DOI: <http://doi.org/10.22214/ijraset.2018.4493>.
- ARDUINO. **What is Arduino?** Disponível em: <https://www.Arduino.cc>. Acesso em: 20 mar. 2023.
- ARDUINO. **MKR Wi-Fi 1010**. 2021. Disponível em: <https://store.arduino.cc/products/arduino-mkr-Wi-Fi-1010>. Acesso em: 10 jan. 2023.
- ARDUINO. **UNO WI-FI REV 2**. Disponível em: <https://store.arduino.cc/products/arduino-uno-Wi-Fi-rev2>. Acesso em: 20 mar. 2023.
- BANANA-PI. Disponível em: <https://www.banana-pi.org>. Acesso em: 20 mar. 2023.
- BAÚ DA ELETRÔNICA. Disponível em: <https://www.baudaeletronica.com.br/Arduino/sensores>. Acesso em: 20 mar. 2023.
- BAUERMEISTER, G. **Comunicação LoRa com Arduino**. 2018. Disponível em: <https://www.makerhero.com/blog/primeiros-passos-LoRa-com-Arduino/>. Acesso em: 10 jan. 2023.
- CHANDRA, R.; COLLIS, S. Digital agriculture for small-scale producers: challenges and opportunities. **Communications of the ACM**, v. 64, n. 12, p. 75-84, Dec. 2021. DOI: <https://doi.org/10.1145/3454008>.
- CLOUD. **The simplest, fastest, low-code way to get started with connected projects**. 2023. Disponível em: <https://cloud.arduino.cc/>. Acesso em: 20 mar. 2023.
- DHANARAJU, M.; CHENNIAPPAN, P.; RAMALINGAM, K.; PAZHANIVELAN, S.; KALIAPERUMAL, R. Smart farming: Internet of Things (IoT)-based sustainable agriculture. **Agriculture**, v. 12, 1745, Oct. 2022. DOI: <https://doi.org/10.3390/agriculture12101745>.
- EMBARCADOS. **Conheça o novo Arduino Uno Wi-Fi**. 2018. Disponível em: <https://embarcados.com.br/conheca-o-novo-Arduino-uno-Wi-Fi>. Acesso em: 10 jan. 2023.
- FAN, J.; ZHANG, Y.; WEN, W.; GU, S.; LU, X.; GUO, X. The future of Internet of Things in agriculture: plant high-throughput phenotypic platform. **Journal of Cleaner Production**, v. 280, 123651, Jan. 2021. DOI: <https://doi.org/10.1016/j.jclepro.2020.123651>.

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

FAPESP. Fundação de Amparo à Pesquisa do Estado de São Paulo. **O Brasil da Internet das Coisas**. 2017. Disponível em: <https://revistapesquisa.fapesp.br/o-brasil-da-Internet-das-coisas/>. Acesso em: 10 jan. 2023.

GARTNER. 2021. Disponível em: <https://www.gartner.com.br>. Acesso em 10 jan. 2023.

GPIOZERO. 2023. Disponível em: <https://gpiozero.readthedocs.io/en/stable/>. Acesso em: 10 jan. 2023.

HELTEC. **Wi-Fi LoRa 32 (V2.1) Phaseout**. 2018. Disponível em: <https://heltec.org/project/wifi-lora-32/>. Acesso em: 10 jan. 2023.

HIVEMQ. **LoRaWAN and MQTT integration for IoT application design**. 2023. Disponível em: <https://www.hivemq.com/blog/LoRawan-and-mqtt-integrations-for-iot-applications-design/>. Acesso em: 10 jan. 2023.

HOW TO ELECTRONICS. **Interfacing LoRa SX1276 with STM32 microcontroller**. 2023. Disponível em: <https://how2electronics.com/interfacing-lora-sx1276-with-stm32-microcontroller-lr1276-915mhz/>. Acesso em: 07 mai. 2023.

INFODOENÇAS. **Sistema de alerta de risco de doenças para o feijoeiro no Estado de Goiás**. 2023. Disponível em: <https://www.cnpaf.embrapa.br/infodoencas/>. Acesso em: 20 mar. 2023.

KALATEC. **Como o Protocolo MQTT funciona e quais são as suas vantagens?** 2023. Disponível em: <https://blog.kalatec.com.br/protocolo-mqtt/>. Acesso em: 20 mar. 2023.

MAKERHERO. 2013. Disponível em: <https://www.makehero.com/blog/monitorando-temperatura-e-umidade-com-o-sensor-dht11/>. Acesso: 20 mar. 2023.

MAKERHERO. **Programação da Raspberry Pi com Java: primeiros passos**. 2018. Disponível em: <https://www.makehero.com/blog/programacao-da-raspberry-pi-com-java/>. Acesso em: 20 mar. 2023.

MOKOLORA. **Top 9 LoRa smart agriculture applications to watch out for**. 2023. Disponível em: <https://www.mokoLoRa.com/LoRa-smart-agriculture-applications/>. Acesso em: 10 jan. 2023.

ORANGE-PI. Disponível em: <https://www.orangepi.org>. Acesso em: 20 mar. 2023.

PLANIN. **Gartner anuncia 10 principais tendências estratégicas para Internet das Coisas até 2023**. Disponível em: <https://planin.com/gartner-anuncia-10-principais-tendencias-estrategicas-para-Internet-das-coisas-ate-2023/>. Acesso em: 10 jan. 2023.

RASP-PI. Disponível em: <https://www.raspberrypi.org>. Acesso em: 10 jan. 2023.

RASPBERRYTIPS. **Getting started with C++ on Raspberry Pi (guide & examples)**. 2023. Disponível em: <https://raspberrytips.com/use-c-on-raspberry-pi/>. Acesso em: 20 mar. 2023.

REDES-IOT. **Overview of IoT networks**. Disponível em: <https://iofactory.eu/iot-knowledge-center/overview-of-iot-networks/>. Acesso em: 20 mar. 2023.

REYNOLDS, D.; BALL, J.; BAUER, A.; DAVEY, R.; GRIFFITHS, S.; ZHOU, J. CropSight: a scalable and open-source information management system for distributed plant phenotyping and IoT-based crop management. **GigaScience**, v. 8, n. 3, p. 1–11, Mar. 2019. DOI: <https://doi.org/10.1093/gigascience/giz009>.

SANTOS, C. C.; SALES, J. D. A. O desafio da privacidade na internet das coisas. **Revista Gestão.Org**, v. 13, p. 282-290, maio 2015.

SHI, X.; AN, X.; ZHAO, Q.; LIU, H.; XIA, L.; SUN, X.; GUO, Y. State-of-the-art Internet of Things in protected agriculture. **Sensors**, v. 19, 1833, Apr. 2019. DOI: <https://doi.org/10.3390/s19081833>.

SIGMAIS. **Tendências de IoT para 2023: o que você precisa saber**. 2023. Disponível em: <https://sigmais.com.br/tendencias-de-iot-para-2023-o-que-voce-precisa-saber/>. Acesso em: 10 jan. 2023.

SYSTEMCAP. **Monitoramento de frequência de pulso na Internet usando ESP8266 e ThingSpeak.** 2021. Disponível em: <https://capsistema.com.br/index.php/2021/02/04/monitoramento-de-frequencia-de-pulso-na-Internet-usando-esp8266-e-thingspeak/>. Acesso: 10 jan. 2023.

TECNOLOGIA-5G. **Internet das coisas:** como o 5G poderá contribuir para o setor agrícola? Disponível em: <https://pt.linkedin.com/pulse/internet-das-coisas-como-o-5g-poder%C3%A1-contribuir-para-setor->. Acesso em 26 jan. 2023.

THILAKARATHNE, N. N.; BAKAR, M. S. A.; ABAS, P. E.; YASSIN, H. Towards making the fields talks: A real-time cloud enabled IoT crop management platform for smart agriculture. **Frontiers in Plant Science**, v. 13, 1030168, Jan. 2023. DOI: <https://doi.org/10.3389/fpls.2022.1030168>.

USINAINFO. **ESP32 LoRa Wi-Fi SX1278.** 2018. Disponível em: <https://www.usinainfo.com.br/lora/esp32-lora-wifi-sx1278-433mhz-de-longo-alcance-com-display-oled-e-bluetooth-5517.html>. Acesso em: 10 jan. 2023.

YADAV, Y.; MAHJABIN, M.; KUMAR, L.; CHAUDHARY, S.; MALIK, S. An overview on Internet of things (IoT) as a technological boon for sustained agriculture farming. **International Journal of Agriculture Innovations and Research**, v. 7, n. 1, p. 129-133, July 2018.

WEMOS. **Home boards.** Disponível em: <https://www.wemos.cc/en/latest/>. Acesso em: 10 jan. 2023.

Germano Martins Ferreira Costa-Neto, Rafael Tassinari Resende, Roberto Fritsche-Neto, Alexandre Bryan Heinemann

Introdução

A tipagem de ambientes, ou simplesmente *ambientipagem*, consiste na ciência de dados que abrange a coleta, o processamento e a interpretação de informações ambientais, assim como a quantificação de seu efeito em processos biológicos de um dado organismo ou de múltiplos organismos, ou ainda, na interação entre eles (Costa-Neto et al., 2021a; Costa-Neto; Fritsche-Neto, 2021; Resende et al., 2021). O termo *ambientipagem* consiste numa tradução direta da palavra “*envirotyping*” (Cooper et al., 2014; Xu, 2016), do inglês *environment* (ambiente) e de seu sufixo *typing* (tipificação, tipagem). Pode ser aplicado em diversos níveis do Dogma Central da Biologia (DNA → RNA → Proteína → Fenótipo), cobrindo também diversas dimensões físicas, tais como o ambiente nuclear e intracelular, passando pelo nível de tecido e órgão vegetal (por exemplo, raiz, caule ou folhas), até atingir o nível de organismo (planta). Tais níveis contemplam a fisiologia do organismo e como ele interage com o ambiente, em termos de expressão gênica, epigenética e os campos vinculados a produção de metabólitos, hormônios e possíveis interações com outros organismos (como por exemplo, a microbiota do solo).

Apesar de válidas, tais tipificações adentram outros campos de estudo, sobretudo na ótica das “multiômicas”, as quais serão contempladas em outros capítulos deste livro. Neste capítulo, o foco será na identificação e análise dos componentes ambientais que afetam o conjunto de plantas em que são realizados experimentos agrícolas ou observações em populações naturais, tais como florestas e lavouras. Na prática, o manejo da terra e das condições do solo, assim como o rendimento da colheita e outros fatores relacionados à produção agrícola são amplamente estudados no campo da Agricultura de Precisão. Isso envolve o uso de mapas de solo e de dados históricos

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

de cultivo para entender a variabilidade espacial dentro de determinada região ou lavoura. No contexto da ambientômica, o endereçamento geográfico também está presente nas inferências, sendo um dos principais objetivos o mapeamento dos melhores materiais genéticos para cada local (ou pixel), levando em consideração toda gama de características ambientais que influenciam na expressão fenotípica das culturas.

No melhoramento de plantas, a caracterização dos padrões ambientais em redes de ensaios experimentais tem sido o maior gargalo – mas também o campo com maiores avanços. Nesse contexto, pesquisadores buscam utilizar a ciência de dados ambientais para estabelecer relações entre os resultados de seus experimentos, assim como sua inter-relação com as possíveis populações de ambientes-alvo de um programa de melhoramento. É importante destacar que realizar técnicas ambientômicas com poucos locais é o mesmo que realizar análise de genômica amostrando-se poucos materiais genéticos (Resende et al., 2022). Para se capturar boa parte da variação ambiental incidente sob os genótipos, é importante se lançar mão de uma boa gama de locais com diferentes gradientes ambientais e suas interações, proporcionando melhor compreensão das interações Genótipos-Ambientes ($G \times A$) e, conseqüentemente, maior precisão na seleção de materiais genéticos com desempenho superior em cada ambiente-alvo. Além disso, a inclusão de mais locais e gradientes ambientais também pode revelar a existência de respostas adaptativas específicas a cada ambiente, o que pode ser explorado pelos programas de melhoramento para aumentar a eficiência e efetividade da seleção de genótipos superiores (Costa-Neto et al. 2021b). A tipagem ambiental não deve ser vista como uma etapa isolada, mas sim como uma parte integrante desde a fase inicial de planejamento até a etapa final de lançamento das novas cultivares.

A ambientômica é uma ferramenta interessante para entender a interação entre organismos e seu ambiente em diferentes níveis. Embora as tipificações de ambiente sejam importantes, elas estão se tornando cada vez mais complexas, com a crescente demanda por estudos multiômicos (Sarzaeim et al., 2023). Neste capítulo, o foco será na tipagem e análise de condições ambientais em populações de plantas e em estudos de agricultura de precisão e redes de ensaios experimentais. Pesquisadores estão usando ciência de dados ambientais para estabelecer relações entre os resultados dos experimentos e as possíveis populações de ambientes-alvo de um programa de melhoramento. Com o desenvolvimento contínuo da tecnologia e das técnicas de análise de dados, a ambientômica continuará sendo

uma ferramenta essencial para entender a interação entre materiais genético e ambientes.

Termos da “Ambientômica”

Para um melhoramento genético de plantas eficaz, é fundamental entender a interação entre fatores genéticos e ambientais. O ambiente fornece as condições necessárias para a expressão adequada dos genes de uma planta, assim como a genética é responsável por sua constituição. Fazendo uma analogia, a ambientômica é o estudo de toda gama de fatores ambientais que afetam o crescimento e o desenvolvimento das plantas, enquanto a genômica se concentra no conjunto completo de genes de uma planta. A ambientipagem busca identificar os fatores ambientais mais importantes para o desenvolvimento da planta, enquanto a genotipagem identifica seus genes. Assim, o ambientoma se refere ao conjunto completo de fatores ambientais que influenciam uma planta, assim como o genoma é o conjunto completo dos genes responsivos pelo seu desenvolvimento.

Na Figura 1 é mostrada uma visualização compreensiva dos conceitos de Ambientômica, em que o ambiótipo é a constituição ambiental do local de desenvolvimento da planta, enquanto a combinação de múltiplos ambiótipos no contexto da interação Genótipos por Ambientes ($G \times A$) irá formar um “polimorfismo” ambientípico influente na plasticidade fenotípica dos materiais genéticos plantados. A ambientômica e a ambientipagem são ferramentas essenciais para o estudo e a melhoria das condições ambientais de crescimento das plantas, enquanto a genotipagem e o genoma permitem o manejo dos genes importantes para o melhoramento genético.

Pode-se, então, definir o conceito de *Ambientômica* (do inglês *Enviromics*), primeiramente introduzido no melhoramento genético por Resende et al. (2019), no preambular *preprint* postado na bioRxiv, mais tarde publicado na *Theoretical and Applied Genetics* – Resende et al. (2021), como a análise da ambientipagem em larga escala (de alto desempenho), uma ciência também pioneiramente trabalhada por Costa-Neto et al. (2021a) em abordagens inéditas de ambientômica + genômica. Pode também ser vista como a ciência que estuda o *ambientoma* (conjunto de fatores ambientais) e, sob a ótica de pertencer a uma “ômica”, expande-se as perspectivas, a fim de desempenhar uma caracterização ambiental visando estabelecer uma associação global entre o ambiente atual, presente e o futuro, considerando múltiplas localidades, manejos, épocas de plantio, cenários ambientais e genótipos.

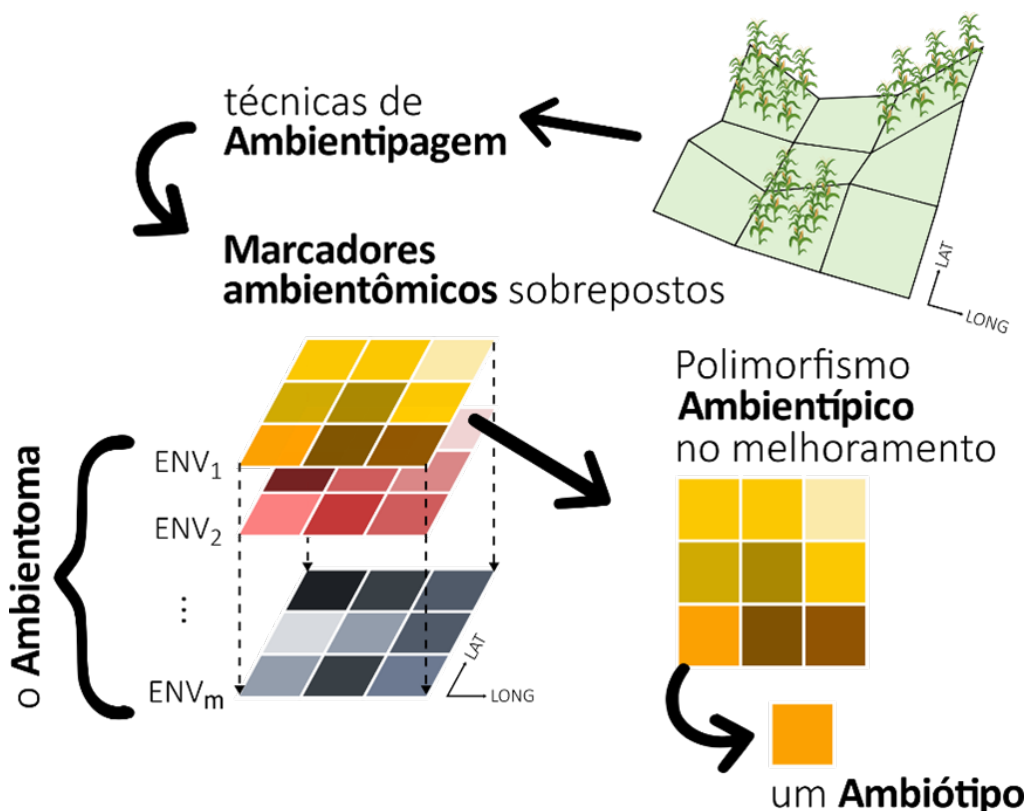


Figura 1. Termos de Ambientômica ancorados em um ambiente de geoprocessamento (nesta figura, é mostrada uma área terrestre com nove pixels ilustrativos). Marcadores ambientômicos podem amplamente ser obtidos por meio de modernas técnicas de ambientipagem (por exemplo, tendência temporal do clima, características físico-químicas do solo, variações da paisagem ou tratamentos fitotécnicos) e, posteriormente, gerados por meio de geoinformática. Um conjunto exaustivo desses marcadores compõe o ambientoma. Um possível valor que um marcador ambientômico pode assumir é chamado de ambiótipo, e tais ambiótipos combinados compõem a variação polimórfica ambientípica do marcador ambientômico.

Fonte: Adaptado de Resende et al. (2021).

No entanto, apesar de ser baseada em um conjunto de ciências aplicadas, tais como sensoriamento (*in situ* ou remoto), sistemas de informação geográfica (SIG) e modelagem estatístico-computacional, a ambientômica é um campo cujo desenvolvimento teórico ainda se faz necessário, baseando-se em aspectos interdisciplinares que englobam os campos da (eco)fisiologia vegetal, ciência de solos, agrometeorologia, agronomia e fitopatologia, entre outros (Costa-Neto et al., 2021a; Costa-Neto; Fritsche-Neto, 2021; Crossa et al., 2021).

O Que é “Ambiente”?

Na seção anterior abordou-se os conceitos gerais que caracterizam o campo da Ambientômica, mas, no entanto, não se especificou um dos conceitos mais importantes: *o que é ambiente?* Sob o ponto de vista da (eco)fisiologia vegetal, sabe-se que todos os processos de crescimento, desenvolvimento e reprodução vegetal são direcionados pelo ambiente (Bilings, 1952). No entanto, de forma genérica, o conceito de ambiente pode ser detalhado como o *conjunto de fatores* nos quais o indivíduo (ser vivo) está inserido. Esses fatores (biofísicos) definem as *circunstâncias* nas quais está sujeito um dado organismo. Tais circunstâncias podem ser descritas como um *saldo do balanço entre a demanda e a disponibilidade de recursos* para execução dos processos biológicos, assim como seu consumo para a constituição dos componentes celulares e ativação de diversas funções moleculares, como controle da expressão gênica (Shelford, 1931; Taylor, 1934). Por serem *organismo-específico*, são, portanto, dependentes de fatores do próprio indivíduo (sua genética e epigenética), dos quais sua regulação ao longo de seu ciclo de vida varia em magnitude entre os diversos estágios de desenvolvimento.

Em termos práticos (e experimentais), o termo “ambiente” também tem sido usado para designar uma combinação de um certo local (geográfico), época de plantio (ano, dia, mês) e manejo cultural (por exemplo, nível de fertilizante, irrigação). Tal definição, embora simplista, resume o conjunto de fatores ambientais numa única “variável” categórica, não discriminando possíveis variações temporais ao longo do ciclo de crescimento. Maiores detalhes sobre esse tópico serão dados nas próximas seções.

No campo de estudo da genética e evolução, é possível apontar que o estudo da influência do ambiente sobre o fenótipo tem sido reconhecido desde o trabalho de Jean-Baptiste Lamarck (1744-1829). Inicialmente, acreditou-se que as variações ambientais forçariam os organismos a expressarem características que lhes confeririam adaptação e sobrevivência. Atualmente, no advento da epigenética, ecofisiologia e ambientômica, sabe-se que tais afirmações não estão longe da realidade. Outra reinterpretção de conceitos sob a ótica da ambientômica pode ser obtida olhando os trabalhos de Charles Darwin (1809-1882), que demonstrou que as modificações fenotípicas que diferenciam espécies são consequências do processo de evolução e adaptação em nível populacional. Em outras palavras: pressões ambientais implicariam em diferenciação (que mais tarde seria reconhecida como genética), priorizando a sobrevivência dos indivíduos portadores de genes com maior adaptação a tais pressões ambientais – processo denominado de “seleção natural”. Esse processo atua em conjunto com a “seleção artificial” conduzida pelos seres humanos

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

desde os primórdios da domesticação das espécies. Por exemplo, as “assinaturas de seleção natural” são evidentes em espécies como milho (*Zea Mays* L), cujo processo de adaptação e modificação de características estruturais do genoma (e.g., número de cromossomos, presença de elementos transponíveis, mudanças nos números de cópias de dado alelo) foram claramente moldadas por variações ambientais, tais como altitude e temperatura do ar (Romero Navarro et al., 2017; Barnes et al., 2022).

Genética e Diversidade Ambiental

Para melhor compreender as fontes de variação ambiental, pode-se assumir que a *qualidade de um dado ambiente* é definida por um balanço de fatores ambientais. Consequentemente, a relevância dos diversos fatores ambientais (por exemplo, temperatura do ar, umidade disponível no solo, déficit de pressão de vapor), residirá em *como*, *quando* e *quanto* esses fatores podem limitar a expressão do potencial genético. Nesse contexto, a relevância de um único fator para restringir a expressão fenotípica pode ser contemplada pelo enunciado da “Lei do Mínimo”, baseada no trabalho de nutrição mineral em plantas, de Justus Von Liebig (1803-1873): sob multiplicidade de fatores, aquele mais próximo do limite da tolerância será o fator limitante (Taylor, 1934). De forma mais ampla, Shelford (1931) relaciona esse conceito à capacidade adaptativa das espécies vegetais, sobretudo a nível populacional – isto é, a variação ambiental pode ditar diversas zonas de adaptação populacional. Portanto, sob essa ótica, a *adaptação* é contemplada como limites genéticos de tolerância frente aos diversos gradientes de qualidade ambiental (Costa-Neto; Fritsche-Neto, 2021).

Dentro dos limites da tolerância, o genótipo possui determinada *flexibilidade fenotípica* (Nicotra et al., 2010; Nicoglou, 2015). Essa flexibilidade permite a sobrevivência do indivíduo quando os fatores do ambiente forem limitantes, tais como drásticas mudanças no *status* hídrico ou térmico. A relação entre fatores ambientais e limites genéticos das espécies vegetais integram os preceitos da “Lei da Tolerância”. Dentro de determinado nível de variação ambiental, os indivíduos são capazes de ‘ajustar’ ou ‘aclimatizar’ suas respostas fenotípicas. Tal comportamento foi denominado por Nilsson-Ehle, em 1914, como plasticidade (Nicotra et al., 2010).

Os determinantes genéticos desse fenômeno são associados a efeitos de pleiotropia, dada a sensibilidade diferencial entre as interações de alelos e fatores ambientais limitantes. Assim, essas respostas diferenciais entre propriedades hereditárias sob certas condições ambientais condicionam a expressão de ampla gama de fenótipos.

Isso constitui os fenômenos biológicos denominados como “norma de reação” (do alemão, *reaktionsnorm* = curva de padrão de resposta) por Richard Woltereck, em 1909, cuja variação é descrita pelo componente interação “Genótipos × Ambientes” (G × A), resultando, portanto, na plasticidade fenotípica (Bradshaw, 1965).

A plasticidade fenotípica é contemplada pela expressão de estruturas morfológicas (e.g., mudanças permanentes nos órgãos vegetativos) e fisiológicas (e.g., alterações metabólicas reversíveis) (Bradshaw, 1965; Sánchez Alvarado; Yamanaka, 2014). Nesse sentido, surgem também interações epistáticas (e.g., alocação de biomassa entre órgãos de armazenamento e componentes de rendimento). Em termos de adaptação fenotípica, isso contempla o fenômeno denominado plasticidade de desenvolvimento (Nicoglou, 2015).

A importância dessas propriedades genético-adaptativas sobre a diferenciação fenotípica foi avaliada cultivando clones de uma planta da espécie *Achillea* sp. em três ambientes distintos, na Califórnia-EUA (Clausen; Hiesey, 1958). Entre outros fatores ambientais ocorrentes nos ambientes, tais como tipo de solo ou temperatura, o mais diferencial era a altitude – estando esse último diretamente vinculado com as variações nas condições edafoclimáticas. O primeiro ambiente era localizado ao nível do mar (0 m de altitude), o segundo a 1.400 m e o terceiro a 3.000 m. Em cada local, diferentes fenótipos foram expressos, demonstrando que o mesmo genótipo reage de formas diferentes conforme a mudança ambiental.

Por esse motivo, pode-se começar a afirmar que as relações ambiente-fenótipo podem ser compreendidas quando o ambiente é detalhado em *fatores ambientais*. Tisdale et al. (1993) descrevem o ambiente sob a ótica do sistema solo-planta-atmosfera, considerando a capacidade fisiológica, particular de cada genótipo, em responder a atuação dos fatores do meio. Para esses autores, três grupos principais de fatores podem ser elucidados: climáticos (por exemplo, precipitação, temperatura do ar, umidade relativa e ventos), edáficos (e.g., material de origem, estrutura, textura, profundidade, temperatura, potencial hidrogeniônico) e biológicos (e.g., constituição gênica dos materiais, eficiência no uso de recursos ambientais e a competição da cultura com insetos-pragas, patógenos e plantas daninhas).

Para o melhoramento genético de plantas, os fatores ambientais podem ser classificados conforme sua previsibilidade de ocorrência no meio. Os fatores ambientais *previsíveis* (fixos ou repetíveis) são aqueles de natureza determinística, tais como o tipo de solo e a posição geográfica. Já os *imprevisíveis* (aleatórios ou não repetíveis) constituem as variações sazonais, em intensidade e frequência, dos fatores de propriedade estocástica (Allard; Bradshaw, 1964). Tais definições são

fundamentais para desenvolver um diagnóstico mais acurado sobre a adaptação fenotípica de genótipos avaliados em ampla gama de ambientes, indicando possíveis causas ambientais relacionadas à sua instabilidade (Heinemann et al., 2019).

Ambientômica e as Multi-Interações Genótipos × Ambientes

Conforme descrito na seção anterior, as condições ambientais exercem papel-chave, embora discreto e quase “invisível”, tanto na seleção natural como naquela conduzida artificialmente, conforme os interesses dos seres humanos. Em melhoramento genético de plantas é possível o estabelecimento dos chamados “ensaios multiambientais” (MET, do inglês *Multi-Environment Trials*), concebidos visando quantificar aspectos agronômicos, tais como a estabilidade e a adaptabilidade dos genótipos candidatos sob múltiplas condições de clima, solo, manejo varietal e pressão de doenças. Tais ensaios objetivam orientar esforços na difusão de tecnologia agropecuária, sobretudo na recomendação das melhores cultivares e/ou práticas de manejo (por exemplo, a época de plantio) para uma dada região-alvo (Costa-Neto et al., 2020a). Quando dois ou mais fatores estão envolvidos nessas avaliações, as respostas diferenciais de cada genótipo frente às variações ambientais podem ser compreendidas como um efeito não aditivo denominado interação Genótipos × Ambientes ($G \times A$). Esse efeito gera implicações logísticas sobre a avaliação de genótipos, podendo inflacionar ou minimizar estimativas de valor fenotípico e genético, limitar ganhos de seleção e dificultar a recomendação de cultivares (Marcatti et al., 2017). Contudo, também denota a oportunidade de selecionar genótipos adaptados e produtivos para condições ambientais específicas. Diante dessas incertezas, estatísticas são utilizadas para diagnosticar a interação $G \times A$, visando identificar padrões agronômicos úteis para capitalizar esforços e orientar a tomada de decisão.

A interação $G \times A$ é um fenômeno universal entre os seres vivos, pois todo fenótipo possui influência ambiental (em menor ou maior escala). Isso decorre da interação multifatorial e não aditiva entre o genótipo (conjunto de genes, mas também epigenética) e o meio no qual se encontra (conjunto de fatores ambientais). Em biometria, pode ser mensurada como a diferença entre os valores fenotípicos (conjunto de respostas observadas), ambientais e genotípicos. Por esse motivo, constitui um fenômeno estatístico-biológico de grande importância no processo de melhoramento genético de plantas, pois desde a seleção de linhagens até a estimação do valor de cultivo

e uso (VCU), para fins de recomendação de cultivares, a interação $G \times A$ dificulta a interpretação do fenótipo observado e subseqüentes inferências sobre o valor genético (Duarte; Vencovsky, 1999).

Para melhor compreender a relação entre diversidade ambiental e interação $G \times A$, é preciso primeiro entender três conceitos principais: (i) plasticidade fenotípica; (ii) norma-reação; (iii) inter-relação entre população-alvo de ambientes (TPE, do inglês *Target Population of Environments*) e ensaios multiambientais. Na Figura 2 é ilustrado que a interação $G \times A$ não é um fenômeno *per se*, mas sim uma propriedade emergente da diversidade ambiental *observada dentro dos limites* de uma rede experimental. A rede experimental é, por sua vez, uma amostra aleatória das possíveis condições ambientais de uma TPE e, por esse motivo, se mal dimensionada, pode falhar em cobrir a diversidade ambiental da TPE. Considerando genótipos distintos (G1 e G2), eles terão potencial variação fenotípica ao longo de um gradiente ambiental teórico (por exemplo, do frio ao calor). Essa curva segue a Lei de Shelford (Figura 2), ou seja, existem zonas nas quais os genótipos apresentam máxima adaptação (máximo valor fenotípico) e zonas nas quais o aumento/decréscimo do recurso ambiental (neste caso, ilustrado por temperatura) irá propiciar um decaimento do valor fenotípico, que em prática é um indicativo de uma zona de “estresse ambiental”. Uma outra forma de expor esse conceito é a de que o excesso ou deficit de insumos ambientais é um fator estressante que limita o potencial de expressão dos genótipos, sendo esse nível de estresse também definido por fatores genéticos. Ainda na Figura 2, as linhas sólidas verticais indicam o intervalo das condições da rede experimental: o nível de entrada máximo e mínimo que os genótipos experimentaram em cada condição MET. De acordo com esses limites ambientais, cada genótipo expressa uma diferente *norma de reação*, ou seja, sua resposta linear ao gradiente *observado*. Portanto, a norma de reação é uma reta que interpola os limites ambientais, sendo uma simplificação da não linearidade inerente a plasticidade fenotípica.

O conceito de norma de reação é uma ideia antiga, mas reflete sensibilidades específicas de genótipos particulares para fatores ambientais chave – o que define a adaptabilidade dos genótipos para a condição vigente na rede experimental. Finalmente, ainda na Figura 2, nas linhas sólidas (verde e roxa) é possível visualizar as diferentes normas de reação para os mesmos genótipos, na mesma condição ambiental. Essa diferença é traduzida, em termos práticos, no que se chama de interação $G \times A$ observada. Portanto, para cada gradiente ambiental particular de cada rede experimental, é possível visualizar também um padrão particular de interação $G \times A$, ainda que para os mesmos genótipos. Por exemplo, na rede experimental

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

1, o padrão é o cruzado (complexo), ou seja, há mudança na classificação dos genótipos entre os ambientes – também conhecida como mudança qualitativa. Por outro lado, se as condições de crescimento do intervalo diferirem, o padrão pode ser *non-crossover* (ou simples), como observado na rede experimental 2, 3 e 4 – mudanças quantitativas na expressão fenotípica entre os ambientes sem mudanças na classificação dos genótipos. Assim, considerando esta ilustração para um cenário de programa real de melhoramento, é impossível inferir como diferentes caracteres, em diferentes redes experimentais, irão se comportar para cada germoplasma. Por isso, os melhoristas devem definir os limites ambientais da TPE para realizar ensaios de campo que representem uma amostra dessas condições – e depois explorar o $G \times A$ para selecionar as cultivares mais adaptadas. O processo de identificar TPE e relacioná-la com as condições MET só pode ser feito usando um conjunto diversificado de ferramentas analíticas de ambientômica.

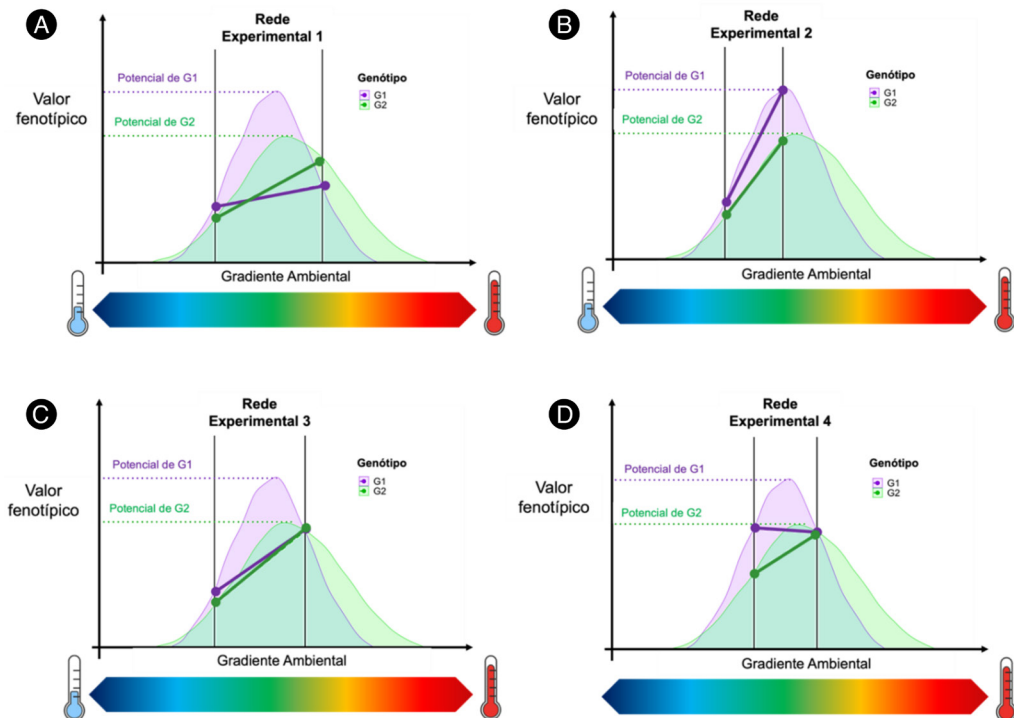


Figura 2. A inter-relação entre plasticidade fenotípica (curvas) e norma de reação (retas entre pontos) em função da diversidade ambiental de cada rede experimental e seu restrito gradiente ambiental (linhas verticais pretas).

Fonte: Adaptado de Costa-Neto et al. (2021b).

Desenvolvimento de uma Rotina de Ambientômica

Nesta seção serão abordados alguns pontos essenciais para o desenvolvimento e a implementação de uma rotina de ambientômica em ciência de dados. Em síntese, existem três passos gerais a serem seguidos.

Passo 1: Coleta de dados ambientais

A primeira etapa consiste na coleta de dados ambientais, sendo um dos pontos-chave para o sucesso na implementação de qualquer rotina de análise de dados em ambientômica. Em analogia à fenotipagem (caracterização de fenótipos), em que é possível mensurar os fenótipos com base em medidas (de peso, altura, comprimento, entre outros), na ambientômica são mensuradas características biofísicas de uma determinada *unidade de análise* (planta, parcela, pixel, área experimental, lavoura, fazenda, município, etc.), sob uma determinada *escala temporal* (e.g., minuto, hora, dia, mês, ano, época de plantio, estágio fenológico). Tais medidas biofísicas são, então, assumidas como *descritores ambientais* (Costa-Neto et al., 2021b), também denominados de *marcadores ambientômicos*. Contudo, ao contrário das características fenotípicas, tais marcadores descrevem diferentes atributos vinculados às condições atmosféricas, fertilidade e umidade dos solos e microbiota do solo.

Conforme a natureza de cada atributo ambiental, diferentes sensores (em campo ou remotos) podem ser utilizados, assim como medidas tomadas no campo, como amostras de solo ou mensuração da penetrância dos solos (por meio dos chamados penetrômetros). Além disso, fenótipos secundários das próprias plantas, como a ocorrência de doenças, e anotações vinculadas ao histórico das práticas culturais empregadas, como culturas que antecederam o plantio e tipo de manejo da cobertura do solo, também podem ser considerados. Nesse sentido, a coleta de dados pode ser resumida em pelo menos quatro tecnologias: (1) coleta por sensores instalados no campo, como estações meteorológicas, sensores de umidade no solo ou amostragem de campo; (2) dados de sensoriamento remoto, ou seja, obtidos de plataformas remotas, como satélites, aviões e drones; (3) bancos de dados globais, que reúnem informações previamente coletadas e interpoladas das categorias 1 e 2, sendo totalmente virtuais; e (4) anotações realizadas em campo, como práticas de manejo cultural. Nesta seção, será dado ênfase às categorias (2) e (3), por serem inovações tecnológicas gratuitas, de fácil acesso e implementação. As categorias (1) e (4) são, de certo modo, intuitivas e se sobrepõem conceitualmente.

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

Vários tipos de imagens de satélite podem ser usados para caracterizar ambientes, podem-se citar:

- *Imagens de visível e infravermelho*: consistem em imagens capturadas em diferentes comprimentos de onda dentro da faixa do espectro eletromagnético. Em geral, são usadas para classificar diferentes tipos de cobertura do solo, como florestas, áreas agrícolas, pastagens, entre outros.
- *Imagens de radar (ondas de rádio)*: consistem em imagens do uso de ondas de rádio e são úteis para caracterizar a presença de nuvens. Também utilizadas para identificar diferentes tipos de topografia, como montanhas e vales. Desde os anos 2000, a NASA (Agência Espacial Norte-Americana) tem utilizado radares no projeto SRTM (*Shuttle Radar Topography Mission*) para mapear a topografia do planeta. Tais dados podem ser utilizados, por exemplo, para definir a altitude exata de uma dada localidade.
- *Imagens termográficas*: assim como as imagens do grupo I, as imagens termográficas capturam diferentes comprimentos de onda dentro da faixa do espectro eletromagnético. Porém, capturam a faixa vinculada as ondas de infravermelho, a qual possibilita mensurar a emissão/absorção de temperatura dos corpos. Por esse motivo, são empregadas na caracterização de áreas de calor, como incêndios florestais, erupções vulcânicas, fontes geotérmicas, entre outros.
- *Imagens multiespectrais*: tais imagens combinam informações de diferentes bandas espectrais para produzir uma representação mais completa da superfície dos corpos. São amplamente utilizadas na identificação de áreas de cultivo, florestas, áreas urbanas, entre outros.

Tais imagens podem ser obtidas gratuitamente em diversas agências governamentais, tais como a NASA (<https://www.earthdata.nasa.gov/>), NOAA (<https://www.noaa.gov/>) e o INPE (<http://www.dgi.inpe.br/CDSR/>), além de empresas privadas que cobram por esse serviço, tais como DigitalGlobe (<https://evwhs.digitalglobe.com/myDigitalGlobe/>) e Maxar Technologies (<https://www.maxar.com/>), entre outras. Após o download dessas imagens, diversas práticas de geoprocessamento (processamento de imagens geográficas) são empregadas, como o uso de máscaras (*shapefiles*) para “cortar” a imagem de acordo com a unidade de análise de interesse, ou técnicas para computação de índices combinando diferentes camadas (*layers*) de imagens. Interessante notar que tais conceitos também são aplicados ao uso de drones para coleta de imagens em parcelas ou áreas agrícolas. Na seção de perspectiva será

discutido como a fenômica e a ambientômica podem trabalhar em conjunto para caracterização de parcelas experimentais em melhoramento de precisão. Quanto as técnicas de geoprocessamento das imagens, recomenda-se a leitura de materiais bibliográficos da área de geoprocessamento e sensoriamento remoto.

Talvez um usuário mais leigo no campo do geoprocessamento faça melhor uso dos bancos de dados. Tais bancos de dados muitas vezes são baseados no processamento das imagens de satélite (categoria 2), em extensas mensurações de estações meteorológicas (categoria 1), ou em extensas e acuradas interpolações globais que combinam tais imagens com mensurações das categorias 1 e 2. No Brasil, Xavier et al. (2016) disponibilizaram um banco de dados climáticos que pode ser acessado de maneira simplificada em <https://www.cnpaf.embrapa.br/infoclima/>. Já nos EUA e Canadá existe atualmente a plataforma CLIM4OMICS (Sarzaeim et al., 2023). Contudo, atualmente existem outras opções que operam em nível global e que são facilmente implementadas em softwares de código-aberto, como o R.

A primeira opção é o *WorldClim* (<https://www.worldclim.org/>), que possui dados em formato de *raster* (arquivo digital que armazena informações espaciais em formato de pixel), com diferentes resoluções (dimensão do pixel), envolvendo normais climatológicas de diversos fatores ambientais (isto é, média das condições meteorológicas observadas durante um longo período de referência, geralmente de 30 a 50 anos), assim como projeções de cenários futuros de mudanças climáticas. A FAO (*Food and Agriculture Organization*), o órgão das Organização Nações Unidas (ONU) voltado para agricultura e segurança alimentar, oferta um banco de dados climáticos, zoneamento agrícola e predições de rendimento dos mais variados cultivos. Esse banco é chamado de GAEZ (*Global Agro-Ecological Zones*, <https://gaez.fao.org/pages/data-viewer>), disponível gratuitamente também em formato *raster*. O banco de dados de solos SoilGrids (<https://soilgrids.org/>) é também uma opção de alta resolução espacial e que oferece ampla variabilidade de atributos do solo (e.g., carbono orgânico, densidade do solo, teor de argila). Baseado no banco mundial de solos, estimativas das características físico-químicas podem ser facilmente obtidas para qualquer posição geográfica no planeta por meio do pacote em R SoilType (<https://github.com/rfn-qt/SoilType>) (Fritsche-Neto, 2023). Mas não menos importante, existe a iniciativa NASA POWER (<https://power.larc.nasa.gov/>), assim como seu API (Interface de Programação de Aplicações, em português (<https://cran.r-project.org/web/packages/nasapower>) que pode ser acessado de dentro do ambiente computacional em linguagem R (popular e mais acessível). Algumas dessas plataformas, incluindo o NASAPOWER, podem ser encontradas dentro de uma mesma plataforma de ambientipagem implementada no pacote EnvRtype (<https://github.com/allogamous/EnvRtype>). Nesse software é

possível fazer o download de informações climáticas em escala diária, correspondendo a um período compreendido entre 01 de janeiro de 1981 até os dias atuais (na ocasião da redação deste capítulo, trata-se do dia 04 de fevereiro de 2023).

Passo 2: Controle de Qualidade

Em ciência de dados, a etapa de controle de qualidade é um processo essencial para garantir a integridade, consistência, precisão e a confiabilidade dos dados coletados. Em termos gerais, consiste no conjunto de técnicas que visam verificar e corrigir possíveis erros ou inconsistências nos dados antes de serem utilizados para análise ou tomada de decisão. O uso de estatísticas gráficas, tais como box-plot e análises da distribuição dos dados, pode auxiliar na verificação de consistência interna, como valores incorretos ou inválidos (exemplos: -999, -99, NA, NaN, Inf, etc.), percentual de dados ausentes e a detecção de valores extremos (*ouliers*). É fundamental para garantir a qualidade dos resultados e a confiabilidade das conclusões baseadas em dados. Em ambientômica, outros aspectos devem ser considerados, tais como a padronização dos dados coletados de diferentes fontes, o que inclui a checagem da resolução temporal (e.g., horária, diária, semanal, mensal) e espacial (por exemplo, a área coberta pelo sensor).

A resolução de um modelo ambientômico dependerá, principalmente, do tamanho da unidade de amostra disponível (pixel). Por exemplo, ao usar informações geográficas, o tamanho do pixel delimitará o nível de refinamento da avaliação genotípica (Marcatti et al., 2017). Informações ambientais com alta resolução espacial, especialmente adequadas para contemplar áreas florestais, plantações agrícolas e fazendas de criação, devem ser preferidas para melhorar a precisão dos modelos.

Segundo Resende et al. (2021), assim como as métricas de “*Call Rate*” e Frequência do Alelo Menor (MAF, do inglês: “*Minor Allele Frequency*”) ao usar dados de marcadores de DNA, as variáveis ambientais também requerem medidas de controle de qualidade. O número de valores ausentes descritos pelo parâmetro *Call Rate* pode ser resolvido por meio da adoção de duas estratégias. A primeira seria aumentar o tamanho de todos os pixels na área e utilizar as informações médias disponíveis no entorno. Embora isso aumente a área de manipulação, diminuindo assim a precisão da recomendação ou predição, ainda pode ser uma alternativa adequada. A segunda estratégia, provavelmente a mais viável, seria imputar os valores dos pixels ausentes por meio de *krigagem* (do inglês *kriging*) ou, ainda, por meio de técnicas de inteligência artificial, como o *Random Forest*, na qual tanto os valores do entorno quanto os valores de outros marcadores ambientômicos podem ser usados. Por exemplo, valores de

temperatura semelhantes entre localizações próximas são mais prováveis do que entre aquelas distantes e, em escala local, a temperatura se correlaciona bem com a elevação do terreno ou com a latitude global. Assim como a MAF para um marcador genético, um marcador ambientômico com pequena variância teria uma baixa Escala de Variação Ambiental (SEV, do inglês: *Scale of Environmental Variation*), de modo que as covariáveis ambientais teriam uma baixa variância nos pixels subamostrados em comparação com todos os pixels na área. Nesse caso, marcadores ambientômicos com baixo SEV, ou seja, baixo conteúdo de informação, podem ser descartados. Estudos adicionais de acompanhamento podem ser realizados para definir os limiares ótimos para as métricas de controle de qualidade dos marcadores ambientômicos.

Passo 3: Transformando dados em informação

Conforme mencionado anteriormente, os dados coletados e processados são assumidos como *marcadores ambientômicos* (descritores ambientais). Analogamente, tratando-se de marcadores genômicos, pode-se facilmente modelá-los com efeitos lineares com relação ao caráter fenotípico, tanto para efeitos de aditividade como para efeitos de dominância (Figura 3-A). No entanto, para os marcadores ambientômicos, essa relação pode nem sempre ser diretamente compreendida como linear (Figura 3-B). Uma estratégia é linearizar as covariáveis ambientais (que poderão ter diferentes relações não lineares com os valores fenotípicos) em marcadores ambientômicos, sendo que, esses sim, terão relações lineares com os caracteres fenotípicos quantitativos. Pode-se realizar essa linearização a partir da decomposição dos efeitos não lineares em diversos parâmetros lineares, como por exemplo no clássico modelo sigmoide logístico $f(x) = c/(1 + \alpha x e^{-bx})$, que contém três parâmetros (a , b , c). Nesse caso, pode-se converter essa variável ambiental, que tem relação não linear logística com o caráter fenotípico, em três marcadores ambientômicos, com relação linear com o caráter fenotípico. Outras estratégias contemplam técnicas de Inteligência Artificial e aprendizado de máquina para realizar essa decomposição linear (Resende et al., 2021). Essa abordagem pode também incluir uma análise das matrizes de relacionamento ambiental, obtidas por correlações fenotípicas entre ambientes (Costa-Neto et al., 2023).

Os descritores são utilizados para caracterizar os ambientes, ou seja, descrever as condições que circundam a *unidade de análise*. Em termos práticos, a caracterização ambiental é resultante do estudo dos padrões presentes numa matriz contendo os descritores ambientais (cada coluna), para cada unidade de análise (cada linha da matriz). O estudo dos padrões de similaridade/redundância entre as unidades de

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

análise, além de quantificar o impacto dos fatores ambientais na variação de uma dada característica em estudo (e.g., produção de grãos, altura de plantas, expressão gênica, ocorrência de uma dada espécie, refletância da folha, entre outros). Nesse sentido, pode-se dividir em duas categorias: (1) estudo dos padrões ambientais analisando apenas os marcadores ambientômicos e (2) uso de modelos estatísticos e de aprendizado de máquina para modelar a associação entre os marcadores ambientômicos e a variação fenotípica observada no campo. Na próxima seção se discutirá mais sobre esses tópicos.

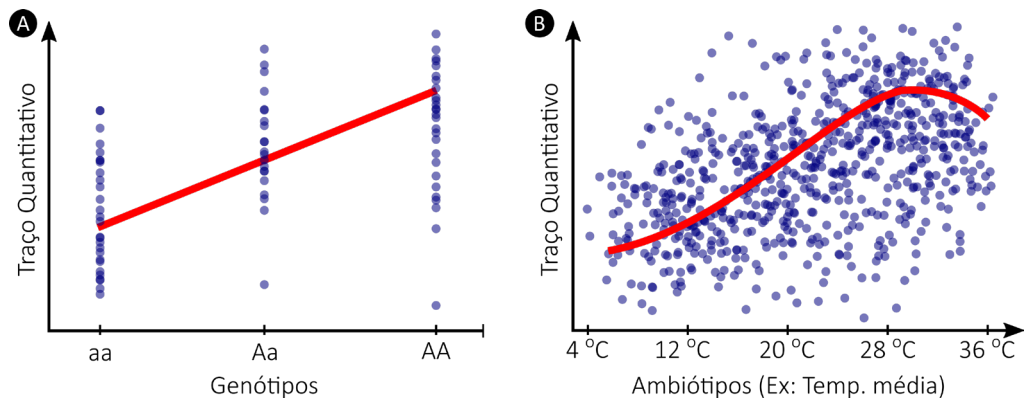


Figura 3. Comparação entre um marcador genômico, com um dado ambiental. Na parte “A”, vê-se três genótipos possíveis (aa, Aa e AA) e suas possíveis relações com diferentes fenótipos (caráter quantitativo). Na parte “B”, vê-se inúmeros ambiótipos, distribuídos quantitativamente, e suas possíveis relações com os fenótipos (caráter quantitativo).

Uma forma comum de transformação dos dados em informações ambientômicas é a computação de *frequências de eventos*, muito útil para variáveis que apresentam sazonalidade climática (Costa-Neto et al., 2021b). Por exemplo, os dados de temperatura média do ar, em escala diária, podem ser convertidos em frequências de temperaturas baixas (estresse por frio), frequências de temperaturas ótimas e frequências de temperaturas excessivas (estresse por calor). Contudo, nesse cenário, é preciso definir o que são temperaturas baixas e o que são temperaturas altas. Curiosamente, esses conceitos seguem a Lei de Shelford e podem ser úteis para descrever a plasticidade fenotípica esperada para uma dada cultura, frente a um gradiente de variação ambiental. Esse gradiente de variação ambiental também pode ser “categorizado”, isto é, dividido em classes de categorias. Em seguida, são contadas as frequências de cada evento (classe). Os diversos ambientes são analisados com base nas similaridades/diferenças entre as classes predominantes. Tal abordagem é

comum para computar a frequência de eventos de seca (Heinemann et al., 2019), calor (Carcedo et al., 2022), frio (Zheng et al., 2015), entre outros. Tais aplicações podem ser implementadas no pacote EnvRtype (Costa-Neto et al., 2021a).

Outra forma de transformar dados em informação é por meio do emprego de modelos de simulação do crescimento, desenvolvimento e produtividade de uma determinada cultura, denominado aqui como modelos de simulação de culturas (MSC). MSC vem sendo desenvolvido desde os anos 1970, tendo sua origem nos modelos agrometeorológicos clássicos. Em termos práticos, o MSC consiste em uma ferramenta de aprendizado de máquina supervisionado, pois são concebidos com funções que descrevem processos biológicos e respostas ambientais, de forma não linear, e respeitando as diferentes fases de desenvolvimento vegetal. Um exemplo interessante é o uso de MSC para simular o ambiente de múltiplos genótipos (Heinemann et al., 2015, 2016, 2019) ou para entender o comportamento dos cultivos em cenários futuros, podendo ser útil para o desenho de ideótipos para condições ambientais de interesse específico do melhorista de plantas (Heinemann et al., 2017; Tao et al., 2017; Ramirez-Villegas et al., 2018). A utilização de MSC possibilita a obtenção de dados de produtividades simulados para uma determinada cultura, abrangendo toda a região de produção. Já, a utilização de ensaios multiambientais dificilmente permite essa abrangência, devido aos custos elevados da realização de ensaios de campo. MSC conseguem capturar algumas interações ($G \times A \times M$, genótipo \times ambiente \times manejo) fenotípicas com robustez e, dentre elas, destacam-se o ciclo da cultura, o florescimento e estresses hídricos e de temperatura.

Exemplos de Aplicação

Apesar das enormes perspectivas de aplicação de ambientômica em melhoramento de plantas, ela tem se restringido à: (1) caracterização ambiental, que pode envolver a compreensão dos fatores ambientais determinantes da interação Genótipo \times Ambiente \times Manejo ($G \times A \times M$), como também a descrição das similaridades ambientais; (2) formatação de zonas de melhoramento, ou macroambientes que minimizem a $G \times A$ dentro delas; (3) estudo da Norma de Reação de genótipos, isto é, da sua responsividade a diversos fatores ambientais; (4) predição de fenótipos, podendo integrar modelos estatísticos e de aprendizado de máquina visando a predição de fenótipos sob os mais variados contextos, seja em nível de propriedade rural ou área experimental (e.g., ensaios de melhoramento); e (5) cálculo de parâmetros de melhoramento como ganhos de seleção, herdabilidades, acurácias preditivas, representatividade da predição pixel-a-pixel na(s) área(s) de estudo.

Caracterização Ambiental

A interação Genótipo x Ambiente x Manejo ($G \times A \times M$) dificulta o entendimento de combinações desejáveis ($G \times M$) para determinada cultura. Além disso, a interação pode também desacelerar o progresso no melhoramento de plantas para a adaptação ampla (largamente utilizado no Brasil) e/ou para adaptação a condições específicas dentro de uma região. A identificação dos fatores causais da interação $G \times A \times M$ e a quantificação sobre a variação residual não explicada pelos efeitos principais são elementos-chave para o entendimento da adaptação fenotípica e para maximizar a produção em circunstâncias ambientais específicas (Costa-Neto et al., 2020a). A análise da interação $G \times A \times M$ permite melhor compreensão do(s) fator(es) que estão por trás das diferenças regionais nas diferenças de produtividade, ou seja; identificar o(s) fator(es) limitante(s) que estão penalizando a produtividade. Essa análise, também denominada como “Caracterização Ambiental”, pode ser realizada de duas maneiras: a) utilizando modelos estatísticos, por meio de dados observados, proveniente dos ensaios multiambientais realizados por um programa de melhoramento de plantas e relacionando esses dados a covariáveis ambientais (dados climáticos, de solo, etc.); ou b) análise *in silico*, por meio de modelos de simulação de culturas.

Caracterização ambiental utilizando modelos estatísticos

Os modelos estatísticos que consideram informações ambientais podem ser modelos de regressão linear simples (regressão fatorial, FR, por exemplo (Van Eeuwijk et al., 1996), quadrados parciais mínimos – PLS (Porker et al., 2020; Costa-Neto et al., 2023) entre outros, integrando métodos não lineares e sistemas de informação geográfica (Costa-Neto et al., 2020b; Bustos-Korts et al., 2022; Heinemann et al., 2022). Ainda, tais estudos permitem identificar os padrões de interação dentro da população do ambiente-alvo do programa de melhoramento genético (Comstock, 1977).

Como exemplo, citam-se dois trabalhos recentes envolvendo as culturas do arroz e do feijão. Por meio do modelo FR-GIS foi possível observar que apenas variáveis geográficas (latitude, longitude e altitude) são capazes de prever até 59% dos padrões da interação $G \times A$ em arroz de terras altas, no Brasil (Costa-Neto et al., 2020b). Por meio do mapeamento da adaptação fenotípica foi possível detectar zonas em que o programa de melhoramento ainda é carente na alocação de genótipos com maior adaptação, ou seja, nas quais há maior dificuldade em desenvolver novas cultivares. Já Heinemann et al. (2022) identificaram, para as regiões de produção do feijoeiro,

variáveis climáticas significativas e seus respectivos valores ótimos. Além disso, foi identificado os “limites climáticos” para as fases de desenvolvimento da cultura para as respectivas regiões e épocas de semeadura. Isso permite aos melhoristas desenvolverem ideótipos, o que por sua vez dará suporte ao processo de seleção de novas variedades. As limitações desse método são: a) número, localização e épocas de semeadura dos METs, o que restringe o alcance temporal e espacial das análises $G \times A \times M$ (muitas vezes os METs também não são representativos da região de produção) e b) a dificuldade em parametrizar modelos estatísticos complexos.

Caracterização ambiental *in silico*

A análise *in silico* (no computador), realizada por meio de modelos de simulação de culturas baseados em processo, supera a limitação dos problemas espaciais e temporais, pois havendo disponibilidade de dados de entrada (clima, solo e características genotípicas) pode-se definir a escala desejável (Antolin et al., 2021; Hajjarpoor et al., 2022). No Brasil, alguns exemplos de caracterização ambiental por meio de MSC incluem a determinação dos padrões de deficiência hídrica para o arroz de terras altas (Heinemann et al., 2015), para o feijoeiro (Heinemann et al., 2016), para a soja (Battisti; Sentelhas, 2019), e projeções de mudanças climáticas para o arroz (Ramirez-Villegas et al., 2018) e para o feijoeiro (Antolin et al., 2021). Como exemplo de aplicação, na análise do arroz de terras altas para a região central do Brasil (Rondônia, Mato Grosso, Goiás e Tocantins), sugeriu-se que a região de produção pode ser dividida em três grupos ambientais homogêneos de acordo com a deficiência hídrica: altamente favorável (AF, 19% da área), favorável (F, 44%) e pouco favorável (PF, 37%). Também foram determinados os principais padrões de deficiência hídrica para o arroz de terras altas e suas respectivas probabilidades. Para o feijoeiro no estado de Goiás, ambas as épocas de semeadura, águas e seca, podem ser divididas em dois grupos ambientais, denominados altamente favorável (AF) e favorável (F). Para essas épocas de semeadura, o ambiente altamente favorável representa 44 e 58% da região de produção, respectivamente. Para ambas as épocas de semeadura há ocorrência da deficiência hídrica terminal e/ou reprodutiva na região de estudo (Goiás), sendo que a deficiência hídrica é mais limitante na época da seca. Já a região de produção da soja no Brasil pode ser dividida em dez ambientes homogêneos. As maiores produtividades ($> 2800 \text{ kg ha}^{-1}$) ocorreram na região Centro-Oeste, cobrindo a maior parte das áreas de cultivo de soja no Brasil. As menores produtividades foram observadas no Extremo Sul ($< 1170 \text{ kg ha}^{-1}$). Esse método restringe as interações que os MSC conseguem capturar.

Estudos das Normas de Reação

A identificação dos fatores causais da interação $G \times A$ e a sua quantificação sobre a variação residual não explicada pelos efeitos principais são elementos-chave para o entendimento da adaptação fenotípica (Charmet et al., 1993; Costa-Neto et al., 2020b). A interpretação desse fenômeno pode ser realizada sob as perspectivas empírica, a partir do valor fenotípico observado *in situ*, e/ou analítica, fundamentada em informações adicionais, tais como variáveis ambientais ou efeito de caracteres secundários (Richards, 1982). Historicamente, a perspectiva empírica tem sido priorizada pelos pesquisadores devido às limitações, como o processamento computacional, a dificuldade de obtenção de informações ambientais e os paradigmas relacionados à compreensão do que seja “ambiente” (Xu, 2016). Contudo, essas dificuldades já não mais justificam o desuso de métodos analíticos, uma vez que as plataformas de ambientômica se tornaram democráticas e acessíveis. Mais recentemente, os métodos de norma de reação ganharam força ao serem acoplados aos modelos modernos de Ambientômica (Resende et al., 2021).

Métodos analíticos possibilitam ampla gama de inferências úteis à tomada de decisão em programas de melhoria, tais como: *i*) identificação de fatores ambientais e genotípicos determinantes da interação $G \times A$ (Heinemann et al., 2019; Porker et al., 2020); *ii*) aumento da acurácia na predição dos valores genéticos (ou genômicos) sob múltiplos ambientes (Heslot et al., 2014; Jarquín et al., 2014; Messina et al., 2018; Morais Junior et al., 2018; Millet et al., 2019); *iii*) adaptabilidade e estabilidade contempladas em função de fatores climáticos, edáficos, bióticos, localização geográfica e manejo cultural (Marcatti et al., 2017); *iv*) perspectiva de avaliação genotípica integrada a sistemas de informação geográfica (SIG), otimizando o processo de recomendação por meio de interface simplificada e dinâmica, baseada em mapas (Costa-Neto, 2017; Marcatti et al., 2017; Costa-Neto et al., 2020b); *v*) identificação indireta de cultivares passíveis de utilização como genitores em programas de melhoria para estresses bióticos e abióticos (Ortiz et al., 2007); *vi*) cálculo de parâmetros genéticos em gradiente ambiental (Resende et al., 2018); entre outras aplicações, como o estudo da interação entre QTL (*quantitative trait loci*) e ambientes (Van Eeuwijk et al., 2007).

O início do uso de regressões analíticas (RA) para avaliação genotípica pode ser atribuído aos trabalhos de Freeman; Perkins (1971), Hardwick; Wood (1972) e Perkins (1972). Esses autores fundamentaram suas propostas em discussões anteriores (Finlay; Wilkinson, 1963; Eberhart; Russell, 1966; Perkins; Jinks, 1968) visando elaborar índices ambientais independentes da média fenotípica e biologicamente informativos. Contudo, avanços consideráveis no desenvolvimento desses métodos

foram feitos (Denis, 1988), com a consolidação da regressão fatorial e da regressão por quadrados mínimos parciais, respectivamente. Deste então, é notória a contribuição da abordagem RA em diversas culturas agrícolas como: milho (Romay et al., 2010), soja (Cardoso Júnior, 2013), eucalipto (Resende et al., 2018), trigo (Heslot et al., 2014), cana-de-açúcar (Oliveira, 2013) e espécies olerícolas (Ortiz et al., 2007; Nunes et al., 2011).

Recentemente, Costa-Neto et al. (2023) definiram o conceito de “associação ambiental ampla”, consistindo no uso do maior número possível de marcadores ambientômicos para descrever, simultaneamente, os padrões de norma de reação dos genótipos e as assinaturas ambientais em uma dada localidade através dos anos. Tal abordagem permite que os melhoristas computem dois tipos de fenômenos vinculados a $G \times A$. O primeiro fenômeno é uma expansão da modelagem analítica previamente descrita, focada na compreensão da norma de reação. Contudo, agora os coeficientes de norma de reação para os mais variados fatores ambientais são assumidos como “marcadores fenotípicos-ambientípicos” e usados nos mais variados propósitos, como associação genômica (Mu et al., 2022) e modelagem de matrizes de relacionamento genético capazes de descrever as interações $G \times A$ (Costa-Neto et al., 2023). Com essa última, é possível verificar o grau de “similaridade entre as normas de reação” de indivíduos avaliados no passado, enquanto a primeira possibilita a modelagem das interações entre o ambiente e regiões específicas do genoma que controlam uma dada característica.

O segundo fenômeno vinculado a $G \times A$ é a modelagem das similaridades ambientais. Por meio da associação ambiental ampla, é possível computar matrizes de relacionamento ambiental, corrigindo a variação climática/edáfica esperada com a correlação fenotípica observada entre diferentes ensaios de campo. Em síntese, é possível computar um “peso” (ou fator de correção) para cada fator com base em seu impacto histórico na variação fenotípica em uma dada localidade. Nesse sentido, o real efeito ambiental sobre as correlações fenotípicas é representado por: *Efeito ambiental* = *Fator Ambiental* \times *Peso* (impacto no fenótipo), em que em uma dada localidade onde não há registros de avaliações fenotípicas para se poder computar esse peso, ele é, portanto, assumido como *Peso* = 1, logo, o *Efeito Ambiental* = *Fator Ambiental*. Além de corrigir as estruturas de relacionamento ambiental, tais pesos ambientais podem também serem usados como covariáveis em algoritmos de agrupamento (análise de componentes principais, *K-means*, *Autoencoders*, dentre outros), assim como em outras abordagens de aprendizado de máquina, tais como algoritmos genéticos visando assim identificar o número mínimo de localidades capaz de representar a diversidade ambiental de uma rede experimental (Gevartosky et

al., 2023). Em associação ambiental ampla, tanto a abordagem 1 como a 2 podem também ser utilizadas para melhorar os modelos de predição de fenótipos, auxiliando na triagem de indivíduos sob condições ambientais conhecidas ou desconhecidas (e.g., próximo ano de cultivo).

Recomendação de cultivares em ambientes de geoprocessamento

A integração de Sistemas de Informação Geográfica (SIG) e genética/genômica quantitativa pode proporcionar resultados acurados ao nível de pixel (tal como nas fotografias digitais). Na Figura 4 são retratados os resultados potenciais que podem ser obtidos mediante a aplicação de técnicas de análise ambientômica em uma área georreferenciada. A representação mostra pontos de coleta de dados fenotípicos marcados por “•”, que podem ser experimentos de melhoramento genético ou simplesmente dados coletados por intermédio da colaboração de produtores locais (*on-farm data*), onde é necessário registrar informações sobre os materiais genéticos utilizados, como cultivares comerciais, variedades, híbridos, clones, linhagens, entre outros (Schmidt et al., 2018). Nesse contexto, é de suma importância trabalhar-se com dados desbalanceados, isto é, nem todos os locais de coleta de dados fenotípicos têm todos os materiais genéticos. Resende et al. (2021) demonstraram que é muito mais prejudicial para captura da interação $G \times A$ a utilização de poucos locais, mesmo que todos eles tenham todos os genótipos (dados balanceados), do que utilizar muitos locais, tendo eles números diferentes de genótipos (dados desbalanceados). Portanto, trabalhar com dados desbalanceados é uma maneira eficiente de otimizar a captura do máximo de informação ao nível “ômico-ambiental” da área.

A seção da Figura 4 identificada como “A” exibe a recomendação dos melhores materiais genéticos de forma detalhada, considerando cada pixel da área georreferenciada. Além disso, na figura também é apresentada uma representação da aplicação desses resultados, destacando a operacionalização para melhor aplicação prática, suavizando-se o efeito difuso do primeiro resultado de recomendação. Na seção “B” da Figura 4, são delimitadas quatro zonas de melhoramento, também chamadas de mega-ambientes (González-Barríos et al., 2019). Essas zonas visam minimizar a interação entre genótipo e ambiente ($G \times A$). A interação $G \times A$ ocorre quando o desempenho de um determinado genótipo varia em diferentes ambientes, ou seja, um mesmo material genético pode apresentar resultados diferentes em diferentes regiões. As zonas de melhoramento auxiliam na identificação de áreas onde os materiais genéticos exibem maior estabilidade e consistência de desempenho, facilitando a seleção e recomendação de cultivares adaptadas a cada mega-ambiente.

Na seção “C” da figura são evidenciadas regiões que maximizam o ganho de seleção. O ganho de seleção refere-se ao aumento no desempenho dos materiais genéticos selecionados em comparação com a geração anterior. Essas regiões destacadas indicam áreas onde a escolha de materiais genéticos resulta em maior ganho de produtividade ou outras características desejáveis, portanto maximizando-se assim a prática do melhoramento genético nesses locais.

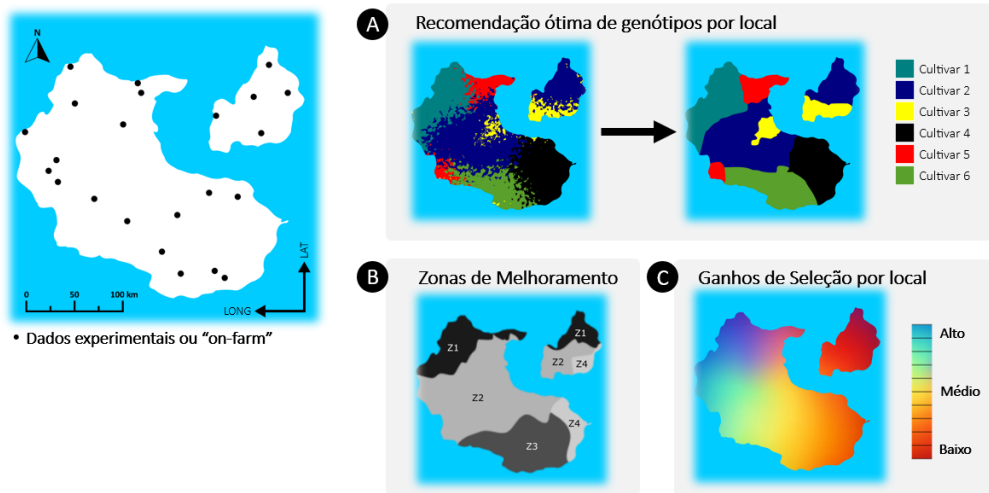


Figura 4. Resultados possíveis a partir da aplicação das técnicas de Ambientômica em uma área georreferenciada qualquer. Nesta área são mostrados pontos de coleta de dados fenotípicos (*), que devem ter a anotação dos materiais genéticos (que podem ser cultivares comerciais, variedades, híbridos, clones, linhagens, entre outros). Na parte “A” é mostrada a recomendação dos melhores materiais genéticos pixel-a-pixel, e uma respectiva operacionalização desses resultados. Na parte “B”, observam-se quatro zonas de melhoramento (ou mega-ambientes), que são delimitações que minimizam a interação $G \times A$. Na parte “C”, vê-se regiões que maximizam o ganho de seleção.

Tanto a identificação das zonas de melhoramento como a atribuição de ganhos de seleção potenciais por pixel irão auxiliar na alocação de novos experimentos ou pontos de coleta de dados fenotípicos nos próximos ciclos de melhoramento. Ademais, pode-se realizar a exclusão de experimentos que não são representativos ou são “redundantes” para a finalidade de aplicação de Ambientômica. Por exemplo, pela Figura 4-B observa-se que a zona de melhoramento “Z4” poderia ter mais pontos de coleta de dados fenotípicos, enquanto que nas demais zonas pode-se pensar na

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

redução do quantitativo de experimentos. Ademais, observa-se que a região noroeste do mapa tende a apresentar ganhos de seleção mais elevados, possibilitando que se diferencie os melhores materiais genéticos dos menos desejáveis.

Perspectivas

Os programas de melhoramento de plantas enfrentam um enorme desafio: continuar a desenvolver variedades que se adequem aos sistemas agrícolas/florestais, os quais sejam capazes de atender as demandas de uma crescente população global, enquanto lida com cenários nada otimistas de mudanças climáticas. Além disso, há necessidade de melhor gerir os insumos e reduzir impacto nas áreas de cultivo, reduzindo assim os efeitos da degradação ambiental. Em termos gerais, buscam-se variedades produtivas, resilientes a fatores bióticos e abióticos, e que sejam eficientes no uso dos recursos ambientais. Para atender a essas expectativas, o campo da ambientômica pode fornecer um olhar que conecta o passado, presente e futuro (Figura 5). Nesta seção serão abordadas duas perspectivas promissoras para auxiliar o melhoramento genético de precisão.

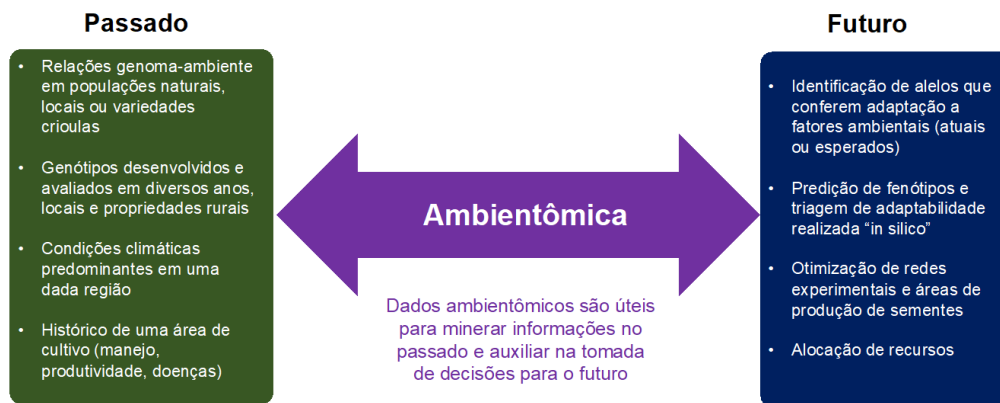


Figura 5. Ambientômica como uma ponte entre o passado e o futuro, auxiliando na tomada de decisões em melhoramento de plantas.

Perspectiva 1. Conectando dados do ambiente celular às informações de satélite

Talvez não seja ousado afirmar que, com o avanço da ciência de dados, os melhoristas de plantas atualmente conseguem minerar informações variando desde

o nível celular (proteínas, transcritos) até o nível de satélite (imagens de satélite usadas para caracterizar regiões de cultivo ou áreas experimentais). Nesse contexto, a ambientômica assume um papel-chave, conectando as áreas de conhecimento da bioquímica, ecofisiologia, fenotipagem e genética/genômica quantitativa. Os modelos multimodais (múltiplas fontes de dados) envolvendo ambientômica poderão ser cada vez mais presentes na realidade do melhorista de plantas, auxiliando no desenvolvimento (e teste) de hipóteses que contemplem os diversos níveis do Dogma Central da Biologia de forma mais ampla e mecanista, inclusive entre diversas espécies com algum grau de parentesco filogenético. Nesse sentido, espécies cujos recursos genômicos ainda são escassos (por exemplo, ausência de sequenciamento genômico) poderão usufruir de dados de outras espécies. Atualmente diversos grupos de pesquisa trabalham nesse tópico, o que faz dessa perspectiva algo tangível nos próximos anos.

Perspectiva 2. Desenho de ideótipos combinando genômica e ambientômica

Em melhoramento de plantas, um “ideótipo” refere-se a um ideal (ou modelo) de planta a ser buscado, por meio da seleção de características desejáveis, para uma dada condição de cultivo. Em verdade, o melhoramento de plantas é fundamentado em ideótipos, pois cada programa de melhoramento busca uma planta ideal para atender as necessidades da sociedade. Por exemplo, nos primórdios da Revolução Verde (anos 1960), os melhoristas passaram a perseguir um ideótipo de plantas adequado para condições de menores espaçamentos (maior densidade de plantio), maior responsividade à aplicação de insumos (fertilizantes e defensivos) e menor ciclo de cultivo (em especial em cultivos de cereais e leguminosas). Tais plantas-modelo são concebidas de forma subjetiva, embora considerando necessidades objetivas dos atuais sistemas de cultivo. Frente a um cenário de mudanças climáticas e as atuais necessidades de estabelecer sistemas agrícolas e florestais mais sustentáveis, a ambientômica pode auxiliar no desenho de ideótipos orientado por dados. Tal abordagem, de certo modo, já é utilizada com modelos de simulação de culturas (Tao et al., 2017) e genômica (Barnes et al., 2022). Essa última abordagem tem sido cada vez mais usada, sendo útil para revelar os aspectos ambientais vinculados à adaptação do genoma das espécies, isto é, as assinaturas ambientais que a seleção natural deixou nas populações naturais e/ou variedades locais (Romero Navarro et al., 2017; De La Torre et al., 2019; López-Hernández; Cortés, 2019). Por meio dessas ferramentas, é possível identificar alelos responsivos ao ambiente, o que combinado

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

com robustas ferramentas estatístico-computacionais de aprendizado de máquina possibilitará prever as melhores combinações genômicas para uma dada condição ambiental. Ainda é possível considerar a fenômica, usada como ferramenta auxiliar na mensuração do maior número possível de características, a fim de auxiliar na compreensão da plasticidade fenotípica sob múltiplas condições ambientais.

Considerações Finais

A ambientômica é um dos campos mais promissores da ciência de dados em melhoria de precisão. O profissional interessado nesse campo poderá se especializar nas mais diversas aplicações, navegando em outras ciências como estatística, fisiologia, física e fertilidade dos solos, irrigação, agrometeorologia, fitopatologia, geoprocessamento, sensoriamento remoto, agricultura de precisão e, até mesmo, nas áreas de ecologia, filogenia e evolução genética. Por esse motivo, possivelmente todo grupo científico e empresa de pesquisa terá um ou mais profissionais dessa área atuando de forma conjunta para mineração de dados úteis ao melhoramento. Para os desenvolvedores interessados nesse campo, ainda há muitos aspectos a serem aprimorados, como o desenvolvimento de rotinas com maior resolução temporal (dados climáticos em escala horária, previsões climáticas para dados futuros ou de um passado geológico distante) e espacial (diferentes profundidades do solo, pixels com dimensões cada vez menores a nível de planta ou parcela experimental), além dos novos modos de processar, interpolar e interpretar os dados dos marcadores ambientômicos. Portanto, espera-se que este capítulo sirva de inspiração para que futuros trabalhos sejam executados.

Referências

- ALLARD, R. W.; BRADSHAW, A. D. Implications of genotype-environmental interactions in applied plant breeding. **Crop Science**, v. 4, n. 5, p. 503-508, Sept./Oct. 1964. DOI: <https://doi.org/10.2135/cropsci1964.0011183X000400050021x>.
- ANTOLIN, L. A. S.; HEINEMANN, A. B.; MARIN, F. R. Impact assessment of common bean availability in Brazil under climate change scenarios. **Agricultural Systems**, v. 191, 103174, June 2021. DOI: <https://doi.org/10.1016/j.agsy.2021.103174>.
- BARNES, A. C.; RODRÍGUEZ-ZAPATA, F.; JUARÉZ-NÚÑEZ, K. A.; RELLÁN-ÁLVAREZ, R. An adaptive teosinte *mexicana* introgression modulates phosphatidylcholine levels and is associated with maize flowering time. **PNAS**, v. 119, e2100036119, June 2022. DOI: <https://doi.org/10.1073/pnas.2100036119>.
- BATTISTI, R.; SENTELHAS, P. C. Characterizing Brazilian soybean-growing regions by water deficit patterns. **Field Crops Research**, V. 240, p. 95–105, July 2019. DOI: <https://doi.org/10.1016/j.fcr.2019.06.007>.

- BILINGS, W. D. The environmental complex in relation to plant growth and distribution. **The Quarterly Review of Biology**, v. 27, n. 3, p. 251–265, Sept. 1952. DOI: <https://doi.org/10.1086/399022>.
- BRADSHAW, A. D. Evolutionary significance of phenotypic plasticity in plants. **Advances in Genetics**, v. 13, p. 115-155, 1965. DOI: [https://doi.org/10.1016/S0065-2660\(08\)60048-6](https://doi.org/10.1016/S0065-2660(08)60048-6).
- BUSTOS-KORTS, D.; BOER, M. P.; LAYTON, J.; GEHRINGER, A.; TANG, T.; WEHRENS, R.; MESSINA, C.; DE LA VEGA, A.; VAN EEUWIJK, F. Identification of environment types and adaptation zones with self-organizing maps; applications to sunflower multi-environment data in Europe. **Theoretical and Applied Genetics**, v. 135, n. 6, p. 2059-2082, June 2022. DOI: <https://doi.org/10.1007/s00122-022-04098-9>.
- CARCEDO, A. J. P.; MAYOR, L.; DEMARCO, P.; MORRIS, G. P.; LINGENFELSER, J.; MESSINA, C. D.; CIAMPITTI, I. A. Environment characterization in sorghum (*Sorghum bicolor* L.) by modeling water-deficit and heat patterns in the Great Plains Region, United States. **Frontiers in Plant Science**, v. 13, 768610, Mar. 2022. DOI: <https://doi.org/10.3389/fpls.2022.768610>.
- CARDOSO JÚNIOR, L. A. **Uso de ferramentas de geoprocessamento para obtenção de mapas de recomendação de cultivares para a macrorregião sojícola 3 do Brasil**. 2013. 101 f. Dissertação (Mestrado em Genética e Melhoramento de Plantas) - Universidade Federal de Goiás, Goiânia.
- CHARMET, G.; BALFOURIER, F.; DENIS, J.-B. Genotype x environment interactions in a core collection of French perennial ryegrass populations. **Theoretical and Applied Genetics**, v. 86, n. 6, p. 731–736, July 1993. DOI: <https://doi.org/10.1007/BF00222663>.
- CLAUSEN, J.; HIESEY, W. M. **Experimental studies on the nature of species**. IV. Genetic structure of ecological races. Washington: Carnegie Institution of Washington, 1958. 312 p. (Publication, 615).
- COMSTOCK, R. E. Quantitative genetics and the design of breeding programs. In: INTERNATIONAL CONFERENCE ON QUANTITATIVE GENETICS, 1976, Ames, Iowa. **Proceedings**. Ames: Iowa State University, 1977. p. 705–718.
- COOPER, M.; MESSINA, C. D.; PODLICH, D.; TOTIR, D.; BAUMGARTEN, A.; HAUSMANN, N. J.; WRIGHT, D.; GRAHAM, G. Predicting the future of plant breeding: Complementing empirical evaluation with genetic prediction. **Crop and Pasture Science**, v. 65, n. 4, p. 311-336, 2014. DOI: <https://doi.org/10.1071/CP14007>.
- COSTA-NETO, G. M. F. **Integrando covariáveis ambientais e mapas temáticos na análise da interação “genótipo x ambiente” em arroz de terras altas**. 2017. 125 f. Dissertação (Mestrado em Genética e Melhoramento de Plantas) - Universidade Federal de Goiás, Goiânia.
- COSTA-NETO, G.; FRITSCHÉ-NETO, R. Enviromics: bridging different sources of data, building one framework. **Crop Breeding and Applied Biotechnology**, v. 21, e393521S12, Aug. 2021. DOI: <https://doi.org/10.1590/1984-70332021v21Sa25>.
- COSTA-NETO, G.; CRESPO-HERRERA, L.; FRADGLEY, N.; GARDNER, K.; BENTLEY, A. R.; DREISIGACKER, S.; FRITSCHÉ-NETO, R.; MONTESINOS-LÓPEZ, O. A.; CROSSA, J. Envirome-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>.
- COSTA-NETO, G. M. F.; DUARTE, J. B.; CASTRO, A. P.; HEINEMANN, A. B. **Uso de informações ambientais na modelagem e interpretação da interação genótipo x ambiente: revisão bibliográfica**. Santo Antônio de Goiás: Embrapa Arroz e Feijão, 2020a. 46 p. (Embrapa Arroz e Feijão. Boletim de pesquisa e desenvolvimento, 56). <http://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/1124389>.
- COSTA-NETO, G. M. F.; MORAIS JÚNIOR, O. P.; HEINEMANN, A. B.; CASTRO, A. P.; DUARTE, J. B. A novel GIS-based tool to reveal spatial trends in reaction norm: upland rice case study. **Euphytica**, v. 216, 37, Feb. 2020b. DOI: <https://doi.org/10.1007/s10681-020-2573-4>.

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

COSTA-NETO, G.; GALLI, G.; CARVALHO, H. F.; CROSSA, J.; FRITSCHÉ-NETO, R. EnvRtype: a software to interplay enviromics and quantitative genomics in agriculture. **G3 Genes|Genomes|Genetics**, v. 11, jkab040, Apr. 2021a. DOI: <https://doi.org/10.1093/g3journal/jkab040>.

COSTA-NETO, G.; CROSSA, J.; FRITSCHÉ-NETO, R. Enviromic assembly increases accuracy and reduces costs of the genomic prediction for yield plasticity in maize. **Frontiers in Plant Science**, v. 12, 717552, Oct. 2021b. DOI: <https://doi.org/10.3389/fpls.2021.717552>.

CROSSA, J.; FRITSCHÉ-NETO, R.; MONTESINOS-LOPEZ, O. A.; COSTA-NETO, G.; DREISIGACKER, S.; MONTESINOS-LOPEZ, A.; BENTLEY, A. R. The modern plant breeding triangle: optimizing the use of genomics, phenomics, and enviromics data. **Frontiers in Plant Science**, v. 12, 651480, Apr. 2021. DOI: <https://doi.org/10.3389/fpls.2021.651480>.

DE LA TORRE, A. R.; WILHITE, B.; NEALE, D. B. Environmental genome-wide association reveals climate adaptation is shaped by subtle to moderate allele frequency shifts in loblolly pine. **Genome Biology and Evolution**, v. 11, n. 10, p. 2976–2989, Oct. 2019. DOI: <https://doi.org/10.1093/gbe/evz220>.

DENIS, J. B. Two way analysis using covarites. **Statistics**, v. 19, n. 1, p. 123–132, 1988. DOI: <https://doi.org/10.1080/02331888808802080>.

DUARTE, J. B.; VENCOSKY, R. **Interação genótipos x ambientes**: uma introdução à análise “AMMI”. Ribeirão Preto: Sociedade Brasileira de Genética, 1999. 60 p. (Monografias, n. 9).

EBERHART, S. A.; RUSSELL, W. A. Stability parameters for comparing varieties. **Crop Science**, v. 6, n. 1, p. 36–40, Jan./Feb. 1966. DOI: <https://doi.org/10.2135/cropsci1966.0011183X000600010011x>.

FINLAY, K. W.; WILKINSON, G. N. The analysis of adaptation in a plant breeding programme. **Australian Journal of Agricultural Research**, v. 14, n. 6, p. 742–754, 1963. DOI: <https://doi.org/10.1071/AR9630742>.

FREEMAN, G. H.; PERKINS, J. M. Environmental and genotype-environment components of variability: VIII. Relations between genotypes grown in different environments and measures of these environments. **Heredity**, v. 27, p. 15–23, Aug. 1971. DOI: <https://doi.org/10.1038/hdy.1971.67>.

FRITSCHÉ-NETO, R. SoilType: an R package to interplay soil characterization in plant science. **Agronomy Journal**, May 2023. DOI: <https://doi.org/10.1002/ajg2.21383>.

GEVARTOSKY, R.; CARVALHO, H. F.; COSTA-NETO, G.; MONTESINOS-LÓPEZ, O. A.; CROSSA, J.; FRITSCHÉ-NETO, R. Enviromic-based kernels may optimize resource allocation with multi-trait multi-environment genomic prediction for tropical Maize. **BMC Plant Biology**, v. 23, 10, Jan. 2023. DOI: <https://doi.org/10.1186/s12870-022-03975-1>.

GONZÁLEZ-BARRIOS, P.; DÍAZ-GARCÍA, L.; GUTIÉRREZ, L. Mega-environmental design: Using genotype × environment interaction to optimize resources for cultivar testing. **Crop Science**, v. 59, n. 5, p. 1899-1915, Sep./Oct. 2019. DOI: <https://doi.org/10.2135/cropsci2018.11.0692>.

HAJJARPOOR, A.; NELSON, W. C. D.; VADEZ, V. How process-based modeling can help plant breeding deal with G x E x M interactions. **Field Crops Research**, v. 283, 108554, July 2022. DOI: <https://doi.org/10.1016/j.fcr.2022.108554>.

HARDWICK, R. C.; WOOD, J. T. Regression methods for studying genotype-environment interactions. **Heredity**, v. 28, p. 209–222, Apr. 1972. DOI: <https://doi.org/10.1038/hdy.1972.26>.

HEINEMANN, A. B.; BARRIOS-PEREZ, C.; RAMIREZ-VILLEGAS, J.; ARANGO-LONDOÑO, D.; BONILLA-FINDJI, O.; MEDEIROS, J. C.; JARVIS, A. Variation and impact of drought-stress patterns across upland rice target population of environments in Brazil. **Journal of Experimental Botany**, v. 66, n. 12, p. 3625-3638, June 2015. DOI: <https://doi.org/10.1093/jxb/erv126>.

HEINEMANN, A. B.; COSTA-NETO, G.; FRITSCHÉ-NETO, R.; MATTA, D. H.; FERNANDES, I. K. Enviromic prediction is useful to define the limits of climate adaptation: a case study of common bean in Brazil. **Field Crops Research**, v. 286, 108628, Oct. 2022. DOI: <https://doi.org/10.1016/j.fcr.2022.108628>.

- HEINEMANN, A. B.; RAMIREZ-VILLEGAS, J.; SOUZA, T. L. P. O.; DIDONET, A. D.; DI STEFANO, J. G.; BOOTE, K. J.; JARVIS, A. Drought impact on rainfed common bean production areas in Brazil. **Agricultural and Forest Meteorology**, v. 225, p. 57-74, Sept. 2016. DOI: <https://doi.org/10.1016/j.agrformet.2016.05.010>.
- HEINEMANN, A. B.; RAMIREZ-VILLEGAS, J.; STONE, L. F.; DIDONET, A. D. Climate change determined drought stress profiles in rainfed common bean production systems in Brazil. **Agricultural and Forest Meteorology**, v. 246, p. 64-77, Nov. 2017. DOI: <https://doi.org/10.1016/j.agrformet.2017.06.005>.
- HEINEMANN, A. B.; RAMIREZ-VILLEGAS, J.; REBOLLEDO, M. C.; COSTA NETO, G. M. F.; CASTRO, A. P. Upland rice breeding led to increased drought sensitivity in Brazil. **Field Crops Research**, v. 231, p. 57-67, 2019. DOI: <https://doi.org/10.1016/j.fcr.2018.11.009>.
- HESLOT, N.; AKDEMIR, D.; SORRELLS, M. E. JANNINK, J.-L. Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. **Theoretical and Applied Genetics**, v. 127, n. 2, p. 463-480, Feb. 2014. DOI: <https://doi.org/10.1007/s00122-013-2231-5>.
- JARQUÍN, D.; CROSSA, J.; LACAZE, X.; CHEYRON, P. D.; DAUCOURT, J.; LORGEOU, J.; PIRAUX, F.; GUERREIRO, L.; PÉREZ, P.; CALUS, M.; BURGUEÑO, J.; DE LOS CAMPOS, G. A reaction norm model for genomic selection using high-dimensional genomic and environmental data. **Theoretical and Applied Genetics**, v. 127, n. 3, p. 595-607, Mar. 2014. DOI: <https://doi.org/10.1007/s00122-013-2243-1>.
- LÓPEZ-HERNÁNDEZ, F.; CORTÉS, A. J. Last-generation genome-environment associations reveal the genetic basis of heat tolerance in common bean (*Phaseolus vulgaris* L.). **Frontiers in Genetics**, v. 10, 954, Nov. 2019. DOI: <https://doi.org/10.3389/fgene.2019.00954>.
- MARCATTI, G. E.; RESENDE, R. T.; RESENDE, M. D. V.; RIBEIRO, C. A. A. S.; SANTOS, A. R.; CRUZ, J. P.; LEITE, H. G. GIS-based approach applied to optimizing recommendations of Eucalyptus genotypes. **Forest Ecology and Management**, v. 392, p. 144-153, May 2017. DOI: <https://doi.org/10.1016/j.foreco.2017.03.006>.
- MESSINA, C. D.; TECHNOW, F.; TANG, T.; TOTIR, R.; GHO, C.; COOPER, M. Leveraging biological insight and environmental variation to improve phenotypic prediction: Integrating crop growth models (CGM) with whole genome prediction (WGP). **European Journal of Agronomy**, v. 100, p. 151-162, Oct. 2018. DOI: <https://doi.org/10.1016/j.eja.2018.01.007>.
- MILLET, E. J.; KRUIJER, W.; COUPEL-LEDRU, A.; PRADO, S. A.; CABRERA-BOSQUET, L.; LACUBE, S.; CHARCOSSET, A.; WELCKER, C.; VAN EEUWIJK, F.; TARDIEU, F. Genomic prediction of maize yield across European environmental conditions. **Nature Genetics**, v. 51, p. 952-956, June 2019. DOI: <https://doi.org/10.1038/s41588-019-0414-y>.
- MORAIS JÚNIOR, O. P.; DUARTE, J. B.; BRESEGHELLO, F.; COELHO, A. S. G.; MORAIS, O. P.; MAGALHÃES JÚNIOR, A. M. Single-step reaction norm models for genomic prediction in multi-environment recurrent selection trials. **Crop Science**, v. 58, n. 2, p. 592-607, Mar./Apr. 2018. DOI: <https://doi.org/10.2135/cropsci2017.06.0366>.
- MU, Q.; GUO, T.; LI, X.; YU, J. Phenotypic plasticity in plant height shaped by interaction between genetic loci and diurnal temperature range. **New Phytologist**, v. 233, n. 4, p. 1768-1779, Feb. 2022. DOI: <https://doi.org/10.1111/nph.17904>.
- NICOGLU, A. The evolution of phenotypic plasticity: genealogy of a debate in genetics. **Studies in History and Philosophy of Biological and Biomedical Sciences**, v. 50, p. 67-76, Apr. 2015. DOI: <https://doi.org/10.1016/j.shpsc.2015.01.003>.
- NICOTRA, A. B.; ATKIN, O. K.; BONSER, S. P.; DAVIDSON, A. M.; FINNEGAN, E. J.; MATHESIUS, U.; PURUGGANAN, M. D.; RICHARDS, C. L.; VALLADARES, F.; VAN KEUNEN, M. Plant phenotypic plasticity in a changing climate. **Trends in Plant Science**, v. 15, n. 12, p. 684-692, Dec. 2010. DOI: <https://doi.org/10.1016/j.tplants.2010.09.008>.

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

NUNES, G. H. S.; ANDRADE NETO, R. C.; COSTA FILHO, J. H.; MELO, S. B. Influência de variáveis ambientais sobre a interação genótipos x ambientes em meloeiro. **Revista Brasileira de Fruticultura**, v. 33, n. 4, p. 1194–1199, dez. 2011. DOI: <https://doi.org/10.1590/S0100-29452011000400018>.

OLIVEIRA, R. S. **Fatores edafoclimáticos determinantes da interação de genótipos com ambientes em cana-de-açúcar**. 2013. 75 f. Dissertação (Mestrado em Genética e Melhoramento de Plantas) - Universidade Federal de Goiás, Goiânia.

ORTIZ, R.; CROSSA, J.; VARGAS, M.; IZQUIERDO, J. Studying the effect of environmental variables on the genotype x environment interaction of tomato. **Euphytica**, v. 153, n. 1, p. 119–134, Jan. 2007. DOI: <https://doi.org/10.1007/s10681-006-9248-7>.

PERKINS, J. M. The principal component analysis of genotype-environmental interactions and physical measures of the environment. **Heredity**, v. 29, p. 59–70, Aug. 1972. DOI: <https://doi.org/10.1038/hdy.1972.64>.

PERKINS, J. M.; JINKS, J. L. Environmental and genotype-environmental components of variability. III. Multiple lines and crosses. **Heredity**, v. 23, p. 339–356, Aug. 1968. DOI: <https://doi.org/10.1038/hdy.1968.48>.

PORKER, K.; COVENTRY, S.; FETTELL, N. A.; COZZOLINO, D.; EGLINTON, J. Using a novel PLS approach for envirotyping of barley phenology and adaptation. **Field Crops Research**, v. 246, 107697, Feb. 2020. DOI: <https://doi.org/10.1016/j.fcr.2019.107697>.

RAMIREZ-VILLEGAS, J.; HEINEMANN, A. B.; CASTRO, A. P.; BRESEGHELLO, F.; NAVARRO-RACINES, C.; LI, T.; REBOLLEDO, M. C.; CHALLINOR, A. J. Breeding implications of drought stress under future climate for upland rice in Brazil. **Global Change Biology**, v. 24, n. 5, p. 2035-2050, May 2018. DOI: <https://doi.org/10.1111/gcb.14071>.

RAMIREZ-VILLEGAS, J.; MOLERO MILAN, A.; ALEXANDROV, N.; ASSENG, S.; CHALLINOR, A. J.; CROSSA, J.; VAN EEUWIJK, F.; GHANEM, M. E.; GRENIER, C.; HEINEMANN, A. B.; WANG, J.; JULIANA, P.; KEHEL, Z.; KHOLOVA, J.; KOO, J.; PEQUENO, D.; QUIROZ, R.; REBOLLEDO, M. C.; SUKUMARAN, S.; VADEZ, V.; WHITE, J. W.; REYNOLDS, M. CGIAR modeling approaches for resource-constrained scenarios: I. Accelerating crop breeding for a changing climate. **Crop Science**, v. 60, n. 2, p. 547-567, Mar./Apr. 2020. DOI: <https://doi.org/10.1002/csc2.20048>.

RESENDE, R. T.; CHENU, K.; RASMUSSEN, S. K.; HEINEMANN, A. B.; FRITSCHÉ-NETO, R. Editorial: Enviromics in plant breeding. **Frontiers in Plant Science**, v. 13, 935380, June 2022. DOI: <https://doi.org/10.3389/fpls.2022.935380>.

RESENDE, R. T.; PIEPHO, H.-P.; ROSA, G. J. M.; SILVA JUNIOR, O. B.; SILVA, F. F.; RESENDE, M. D. V.; GRATTAPAGLIA, D. Enviromics in breeding: applications and perspectives on envirotypic-assisted selection. **Theoretical and Applied Genetics**, v. 134, n. 1, p. 95-112, Jan. 2021. DOI: <https://doi.org/10.1007/s00122-020-03684-z>.

RESENDE, R. T.; PIEPHO, H.-P.; SILVA JUNIOR, O. B.; SILVA, F. F.; RESENDE, M. D. V.; GRATTAPAGLIA, D. *Enviromics* in breeding: applications and perspectives on envirotypic-assisted selection. **bioRxiv**, Aug. 2019. DOI: <https://doi.org/10.1101/726513>.

RESENDE, R. T.; SOARES, A. A. V.; FORRESTER, D. I.; MARCATTI, G. E.; SANTOS, A. R.; TAKAHASHI, E. K.; SILVA, F. F.; GRATTAPAGLIA, D.; RESENDE, M. D. V.; LEITE, H. G. Environmental uniformity, site quality and tree competition interact to determine stand productivity of clonal Eucalyptus. **Forest Ecology and Management**, v. 410, p. 76-83, Feb. 2018. DOI: <https://doi.org/10.1016/j.foreco.2017.12.038>.

RICHARDS, R. A. Breeding and selecting for drought resistance in wheat. In: DROUGHT resistance in crops with emphasis on rice. Los Baños: International Rice Research Institute, 1982. p. 303–316.

- ROMAY, M. C.; MALVAR, R. A.; CAMPO, L.; ALVAREZ, A.; MORENO-GONZÁLEZ, J.; ORDÁS, A.; PRECIADO, P. Climatic and genotypic effects for grain yield in maize under stress conditions. **Crop Science**, v. 50, n. 1, p. 51–58, Jan. 2010. DOI: <https://doi.org/10.2135/cropsci2008.12.0695>.
- ROMERO NAVARRO, J. A.; WILLCOX, M.; BURGUEÑO, J.; ROMAY, C.; SWARTS, K.; TRACHSEL, S.; PRECIADO, E.; TERRON, A.; DELGADO, H. V.; VIDAL, V.; ORTEGA, A.; BANDA, A. E.; MONTIEL, N. O. G.; ORTIZ-MONASTERIO, I.; SAN VICENTE, F.; ESPINOZA, A. G.; ATLIN, G.; WENZL, P.; HEARNE, S.; BUCKLER, E. S. A study of allelic diversity underlying flowering-time adaptation in maize landraces. **Nature Genetics**, v. 49, p. 476–480, Mar. 2017. DOI: <https://doi.org/10.1038/ng.3784>.
- SÁNCHEZ ALVARADO, A.; YAMANAKA, S. Rethinking differentiation: Stem cells, regeneration, and plasticity. **Cell**, v. 157, n. 1, p. 110–119, Mar. 2014. DOI: <https://doi.org/10.1016/j.cell.2014.02.041>.
- SARZAEIM, P.; MUNOZ-ARRIOLA, F.; JARQUIN, D.; ASLAM, H.; GATTI, N. L. **CLIM4OMICS**: a geospatially comprehensive climate and multi-OMICS database for Maize phenotype predictability in the U.S. and Canada. *Earth System Science Data Discussions*, Feb. 2023. DOI: <https://doi.org/10.5194/essd-2023-11>.
- SCHMIDT, P.; MÖHRING, J.; KOCH, R. J.; PIEPHO, H.-P. More, larger, simpler: How comparable are on-farm and on-station trials for cultivar evaluation? **Crop Science**, v. 58, n. 4, p. 1508-1518, July/Aug. 2018. DOI: <https://doi.org/10.2135/cropsci2017.09.0555>.
- SHELFORD, V. E. Some concepts of bioecology. **Ecology**, v. 12, n. 3, p. 455–467, July 1931. DOI: <https://doi.org/10.2307/1928991>.
- TAO, F.; RÖTTER, R. P.; PALOSUO, T.; DÍAZ-AMBRONA, C. G. H.; MÍNGUEZ, M. I.; SEMENOV, M. A.; KERSEBAUM, K. C.; NENDEL, C.; CAMMARANO, D.; HOFFMANN, H.; EWERT, F.; DAMBREVILLE, A.; MARTRE, P.; RODRÍGUEZ, L.; RUIZ-RAMOS, M.; GAISER, T.; HÖHN, J. G.; SALO, T.; FERRISE, R.; BINDI, M.; SCHULMAN, A. H. Designing future barley ideotypes using a crop model ensemble. **European Journal of Agronomy**, v. 82, p. 144–162, Jan. 2017. DOI: <https://doi.org/10.1016/j.eja.2016.10.012>.
- TAYLOR, W. P. Significance of extreme or intermittent conditions in distribution of species and management of natural resources, with a restatement of Liebig's Law of Minimum. **Ecology**, v. 15, n. 4, p. 374–379, Oct. 1934. DOI: <https://doi.org/10.2307/1932352>.
- TISDALE, S. L.; NELSON, W. L.; BEATON, J. D.; HAVLIN, J. L. **Soil fertility and fertilizers**. 5th ed. London: Macmillan, 1993. 648 p.
- VAN EEUWIJK, F. A.; DENIS, J.-B.; KANG, M. S. Incorporating additional information on genotypes and environments in models for two-way genotype by environment tables. In: KANG, M. S.; GAUCH, H. G. (ed.). **Genotype-by-environment interaction**. Boca Raton: CRC Press, 1996. p. 15-49.
- VAN EEUWIJK, F. A.; MALOSETTI, M.; BOER, M. P. Modelling the genetic basis of response curves underlying genotype x environment interaction. In: SPIERTZ, J. H. J.; STRUIK, P. C.; VAN LAAR, H. H. (ed.). **Scale and complexity in plant systems research: gene-plant-crop relations**. Dordrecht: Springer, 2007. p. 115–126.
- XAVIER, A. C.; KING, C. W.; SCANLON, B. R. Daily gridded meteorological variables in Brazil (1980-2013). **International Journal of Climatology**, v. 36, n. 6, p. 2644-2659, May 2016. DOI: <https://doi.org/10.1002/joc.4518>.
- XU, Y. Envirotyping for deciphering environmental impacts on crop plants. **Theoretical and Applied Genetics**, v. 129, n. 4, p. 653-673, Apr. 2016. DOI: <https://doi.org/10.1007/s00122-016-2691-5>.
- ZHENG, B.; CHAPMAN, S.; CHRISTOPHER, J.; FREDERIKS, T.; CHENU, K. Frost trends and their estimated impact on yield in the Australian wheatbelt. **Procedia Environmental Sciences**, v. 29, p. 171-172, 2015. DOI: <https://doi.org/10.1016/j.proenv.2015.07.244>.

Filipe Inácio Matias, Leonardo Volpato, João Otavio Firigato

Introdução

O melhoramento de variedades adaptadas a diferentes ambientes, juntamente com a utilização da agricultura de precisão, são ferramentas fundamentais para enfrentar a demanda alimentícia global. Nesse cenário, o incremento da produção agrícola das principais culturas econômicas tornou-se um fator chave para assegurar a segurança alimentar no longo prazo (Moreira et al., 2020). Logo, o papel do melhoramento de plantas é fundamental para atender essa crescente demanda, da mesma forma que a utilização de novas ferramentas de agricultura de precisão tende a maximizar o uso de recursos não renováveis no campo. Os programas de melhoramento atuais são limitados por custo, tempo, trabalho humano, espaço e outros recursos para avaliar eficientemente uma grande população de progênies, o que limita a intensidade da seleção, interferindo na acurácia genética e pode resultar na redução do avanço do ganho genético. O desenvolvimento e aplicação de ferramentas de fenotipagem de baixo custo e alto rendimento permitem a realocação de recursos para gerenciar populações maiores, possibilitando aumento na intensidade de seleção dentro de um orçamento pré-estabelecido. Logo, ganhos reais nos programas de melhoramento são perfeitamente viáveis adotando novas práticas de fenotipagem em larga escala (HTP - *high-throughput phenotyping*). O termo "fenotipagem" começou a ser usado na década de 1960 e mais tarde foi referido como o conjunto de metodologias e protocolos usados para medir com precisão o crescimento, arquitetura e composição das plantas em diferentes escalas (Fiorani; Schurr, 2013; Walter et al., 2015). Tradicionalmente, para selecionar progênies superiores ou identificar regiões no genoma que controlam uma característica, os fenótipos de plantas são medidos pelos melhoristas utilizando métodos de baixo rendimento em laboratório, observações visuais e ferramentas

manuais. No entanto, muitos dos métodos tradicionais de fenotipagem de plantas tendem a ser trabalhosos, demorados, subjetivos e frequentemente destrutivos para as plantas.

Existe uma pressão do mercado para que programas de melhoramento convencionais se transformem gradativamente em programas de melhoramento modernos mais eficientes por meio da integração de tecnologias emergentes, com destaque para a HTP, a qual impacta diretamente na escolha dos melhores fenótipos. A evolução da genômica, transcriptômica e proteômica alcançada nas últimas duas décadas permitiu a investigação em larga escala da base genética de características agrônômicas relevantes ao incremento da produção agrícola (Roitsch et al., 2019). Nos últimos anos, a HTP está estreitando a distância entre estudos relacionados a genômica com a fenômica de plantas devido, principalmente, a implementação de novas ferramentas de hardwares e softwares empregados diretamente no melhoramento. Portanto, as contribuições do melhoramento genético baseadas na tecnologia de HTP possuem grande potencial para aumentar os ganhos atuais de produtividade, reduzir o custo da operação, assim como melhorar o bem-estar do melhorista (ou do colaborador responsável pelas avaliações no campo) e, ainda, tendem a garantir ganhos significativos da produção agrícola a longo prazo. Dito isso, neste capítulo veremos a importância do uso de tecnologias digitais para a fenotipagem de plantas, explicando conceitos, mostrando o passo a passo do uso de drones em áreas experimentais, bem como demonstrando várias aplicações da fenômica nas atividades diárias de um programa de melhoramento de plantas.

Fenotipagem de Alto Rendimento na Produção Agrícola

As tecnologias de fenotipagem HTP surgiram na última década graças a modernização tecnológica, promovendo a redução no custo de implementação de sensores, o implemento da visão computacional, a automação de fluxo de trabalho e tecnologias avançadas de aprendizado de máquina. Um sistema HTP abrangente é composto por hardware de suporte (sensores e plataformas) e componente de computação (processamento e análise de dados). Os sensores amplamente utilizados na tecnologia HTP são principalmente sensores que não possuem contato diretamente com o alvo (plantas) e não invasivos, como câmeras digitais - por exemplo, visível ou RGB (vermelho, verde e azul), multiespectral, hiperespectral e câmeras térmicas - e sensores de profundidade tridimensionais (LiDAR). Esforços têm sido empregados para

integrar os avanços na tecnologia de sensores aos sistemas de HTP desenvolvidos, principalmente pelo setor industrial. As plataformas de automação comumente usadas incluem sistemas de automação baseados em trilhos (Zhou et al., 2018), sistemas robóticos internos e externos (Zhao et al., 2019; Yang et al., 2020), e aeronaves remotamente pilotadas (RPA, *Remotely Piloted Aircraft*) (Yang et al., 2017).

Os sensores HTP podem ser classificados com base na fonte da radiação detectada (ativa ou passiva), a porção do espectro eletromagnético utilizada e, ainda, subdivididos pelo tipo de dados que são retornados (por exemplo, imagem ou valor digital). Ambos os sensores ativos e passivos medem a energia refletida. O sensor do tipo ativo emite energia que é então detectada, enquanto que o passivo depende da radiação externa (solar, por exemplo) refletida em um objeto (Barmeier; Schmidhalter, 2017). Sensores ativos populares usados em HTP, como LiDAR e sensores ultrassônicos, são úteis para gerar nuvens de pontos 3D, que podem ser usadas para extração da arquitetura do dossel (Conn et al., 2017), altura da planta (Wang et al., 2018) e biomassa acima do solo (Pittman et al., 2015). Os sensores passivos consistem em multiespectrais, hiperespectrais, RGB, fluorescência e infravermelho térmico. Os sensores espectrais geram informações de características fenotípicas, utilizando a relação entre a refletância medida de um dossel ou outro tecido vegetal com as várias características fisiológicas que controlam a resposta fenotípica observada. As características fisiológicas são derivadas de valores de refletância de regiões do espectro eletromagnético (Parmley et al., 2019) e índices de vegetação que estão associados a processos bioquímicos na planta (Xu et al., 2017). As câmeras RGB são sensores acessíveis e escaláveis que têm sido usados extensivamente em pesquisas para medir a cobertura do dossel (Xavier et al., 2017), biomassa acima do solo (Ballesteros et al., 2018), detecção e estimativa de estresse (Dobbels; Lorenz, 2019), altura de plantas (Volpato et al., 2021b), avaliação de doenças (Nagasubramanian et al., 2018) e muitos outros propósitos. Diversos sensores estão disponíveis para aplicações HTP, no entanto, sua utilidade depende da versatilidade, significância das correlações entre características, acessibilidade e rendimento (Tao et al., 2022).

Um dos componentes mais importantes de um sistema HTP é o processamento e análise dos dados coletados pelos sensores de fenotipagem. Os sistemas HTP atuais, especialmente os sistemas de imagem de alta resolução, estão projetados para coletar dados de alta dimensão em larga escala a nível de campo. No entanto, um dos principais gargalos da HTP é a manipulação de enormes conjuntos de dados ou *big data*, os quais muitas vezes não são satisfatoriamente explorados devido à falta de capacidade operacional disponível (Blumenthal et al., 2020; Yang et al., 2020). Portanto, uma das tarefas urgentes para implementação dos métodos de HTP é desenvolver estruturas

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

ou esquemas de trabalho para o processamento e análise de dados de forma mais eficiente, e que se possa traduzir os dados coletados pelos sensores para características de interesse das culturas trabalhadas em programas de melhoramento. Mais esforços devem ser feitos para desenvolver e integrar tecnologias emergentes, como computação em nuvem e o uso de tecnologia computacional de ponta em sistemas HTP, entre eles a inteligência artificial (IA) e suas subáreas: aprendizado de máquina (ML – *Machine Learning*), e aprendizado profundo (DL – *Deep Learning*). Os algoritmos de ML e DL são abordagens promissoras para análises mais rápidas, mais eficientes e melhores para descobertas significativas. Os métodos de aprendizado de máquina são hábeis em extrair padrões e características de uma grande quantidade de dados, inerentes às plataformas HTP. O aprendizado de máquina trouxe melhorias rápidas na capacidade de identificar, classificar, quantificar e prever características (Singh et al., 2021). Com os esforços contínuos da comunidade científica e tecnológica, HTP pode potencialmente ser o componente chave para resolver a equação de ganho genético do melhoramento de plantas e acelerar o processo de lançamento de novas variedades com características de interesse agrônomo, resultando assim em maiores ganhos de produtividade no campo e menor uso de recursos não renováveis, tais como os insumos agrícolas.

Aplicação da Fenotipagem de Alto Rendimento no Melhoramento de Plantas

A fenotipagem de alto rendimento vem se transformando cada vez mais desde a última década, permitindo a transição da fenotipagem manual tradicional para sistemas totalmente automatizados. O progresso na área de fenotipagem de plantas está se tornando viável devido ao rápido desenvolvimento tecnológico de sensores e da capacidade computacional, assim como o emprego de algoritmos de processamento de dados mais eficientes. Nesse sentido, diversos parâmetros fenotípicos podem ser estimados ou preditos, incluindo taxa de crescimento da planta, altura, data de maturidade, acamamento, tolerância a estresses abióticos e bióticos, entre outros (Tao et al., 2022). Fenotipagem de boa qualidade tende a produzir resultados desejáveis, portanto, os pesquisadores que pretendem utilizar a HTP em seus esquemas de desenvolvimentos de cultivares devem seguir com critério os protocolos de sensoriamento remoto, de qualidade na obtenção das imagens e extração de dados e, principalmente, ter a clara definição das estratégias de análises estatísticas.

Os sistemas HTP de campo podem medir diferentes parâmetros fenotípicos no nível do dossel ao longo do desenvolvimento da cultura. Por exemplo, a dinâmica da

cobertura vegetal, altura média, e índices de vegetação podem ser medidos diariamente ou semanalmente. Embora os melhoristas de plantas usem parâmetros de média de parcela para avaliar linhagens de melhoramento, informações mais detalhadas podem ser obtidas ou recuperadas se os sistemas HTP forem capazes de realizar a medição ao nível de planta e, até mesmo, obter dados unitários de folhas. No entanto, ainda é um desafio medir com precisão os parâmetros fenotípicos a nível de planta na condição de campo com precisão e alta resolução espaçotemporal. Os fatores ambientais, como a perturbação do vento, a mudança da radiação solar recebida e o posicionamento dos sensores são as principais restrições.

As plataformas HTP de campo podem ser categorizados em instalações terrestres fixas de grande escala e plataformas móveis. Os sistemas HTP mais comum utilizados atualmente incluem plataformas móveis terrestre (robôs) e aéreas (drones). Os veículos terrestres móveis geralmente têm uma capacidade de carga e tempo de bateria superior em comparação com as plataformas aéreas. As plataformas terrestres podem variar de veículos empurrados manualmente de baixo custo a robôs de campo totalmente automatizados especificamente projetados. As plataformas aéreas, por sua vez, têm maior rendimento de fenotipagem e acessibilidade de campo durante todo o estágio de desenvolvimento da cultura (Li et al., 2019). Os sistemas terrestres fixos de grande escala podem desempenhar um papel vital no teste de diferentes combinações de sensores sob diferentes condições de trabalho, devido às suas vantagens de carga de sensores, suplemento energético e precisão de posicionamento. Entretanto, o custo do sistema de larga escala geralmente é muito maior do que o das plataformas móveis (Kirchgessner et al., 2016). Um sistema HTP fixo de grande escala a campo tem a capacidade de realizar a coleta de dados em resoluções espaçotemporais com alta repetibilidade e com o mínimo de perturbação de campo. Um exemplo é o programa *Renewable Agriculture Phenotyping Reference Platform* (TERRA-REF), localizado na estação de Maricopa-AZ, da Universidade do Arizona (EUA), que visa transformar o melhoramento de plantas usando sensoriamento remoto para quantificar características como arquitetura da planta, absorção de carbono, química dos tecidos, uso da água e outros recursos, com o objetivo de prever o potencial de produtividade e resistência/tolerância a estresses (<https://terraef.org/>). Assim, em termos gerais, pode-se classificar as plataformas de fenotipagem de plantas em três tipos básicos: (1) plataformas terrestres, (2) plataformas aéreas e (3) plataformas espaciais (Li et al., 2022).

(1) Plataformas terrestres: (a) **Plataformas fixas do tipo correia transportadora:** adequado para uso interno, possui a vantagem em coletar dados em tempo real devido à facilidade de transporte e substituir amostras dentro da área experimental (plantas/

vasos). (b) **Plataformas fixas do tipo cavalete/pórtico/guindaste**: elevada precisão e potencial para coleta de medidas repetidas ao longo do dia. Construído por meio de um dispositivo de caminhada sentido XY que facilita sua adaptação a diferentes tipos de amostra. De custo elevado, pode ser construída em casas de vegetação ou no campo de forma fixa, o que limita a área observada (ex. TERRA-REF). (c) **Plataformas do tipo suspensão**: operação contínua, boa repetibilidade e alta precisão de medição. Normalmente são de custo elevado e cobrem um número limitado de áreas. (d) **Plataformas acopladas em veículos**: geralmente são sensores em barras/suportes associadas a máquinas agrícolas (tratores, por exemplo). Ampla adaptação a diferentes culturas agrícolas e ambientes, facilidade em cobrir grandes áreas devido ao baixo custo, e facilidade de instalação e operação. No entanto, limitações como a roda larga de um trator, peso do equipamento, velocidade de movimentação e baixa altura do corpo do veículo restringem a coleta de dados às plantas de porte baixo e/ou em estágios iniciais de desenvolvimento.

(2) Plataformas aéreas: (a) **Helicópteros/aviões tripulados**: coleta de dados em alturas elevadas, maior eficiência em cobrir grandes áreas, e pode carregar equipamentos e sensores pesados. No entanto, possui custo elevado e geralmente menor resolução nas imagens. (b) **Aeronaves Remotamente Pilotadas (RPA) ou Veículo Aéreo Não Tripulado (VANT)**: também conhecidos como drones, podem ser de asas fixas ou multirrotor. Se destaca como a metodologia mais amplamente utilizada em melhoramento de plantas devido ao baixo custo, coleta de dados em baixas alturas, captura de imagens com qualidade e de alta resolução, além da facilidade/precisão na aquisição de detalhes (adequado para avaliação de parcelas experimentais). Como desvantagens, pode-se citar a necessidade de elevado número de baterias para cobrir grandes áreas de fenotipagem, e a capacidade máxima de carga para realizar os voos, pois alguns sensores possuem peso além do limite de carga de voo, o qual restringe a utilização de modelos de baixo custo.

(3) Plataformas espaciais: (a) **Plataformas de satélite**: adequada para monitorar grandes áreas de cultivo. Portam inúmeros sensores como frequências ópticas, térmicas, micro-ondas e de fluorescência, permitindo a coleta de grandes quantidades de dados em curto período de tempo. Apesar da menor precisão e qualidade das imagens, o fato de fornecerem informações recorrentes permite sua ampla utilização na avaliação das condições agrícolas em diferentes momentos do ciclo de crescimento dos cultivos. Suas aplicações incluem a previsão de rendimento, preparação do campo, monitoramento da sanidade da cultura, irrigação e manejo específico do local. Com o desenvolvimento das tecnologias e a elevação da resolução e qualidade das imagens, sua utilização vai se tornar cada vez mais frequente no melhoramento de plantas.

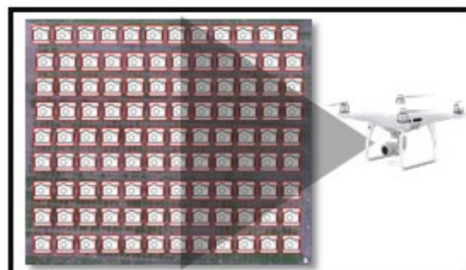
Elaborando o Pipeline de Fenotipagem para o Uso de Drones

Na Figura 1 são apresentadas as principais etapas a serem realizadas no programa de melhoramento utilizando RPA ou VANT para a coleta e extração de dados. A seguir, discutiremos com mais detalhes as etapas básicas desse esquema de trabalho que é adotado pela maioria dos programas de melhoramento de plantas, o qual se inicia pelas normas e regulamentações de voos até a extração e coleta dos dados das imagens. No entanto, antes de se realizar qualquer atividade por meio de RPA, o piloto ou o responsável pelos voos deve atender algumas demandas regulamentares da legislação nacional para preparar e ajustar o campo experimental antes do período de voos.

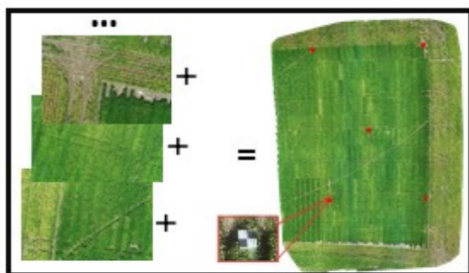
A Plano de voo



B Captura das imagens



C Agrupamento das imagens



D Segmentação das parcelas

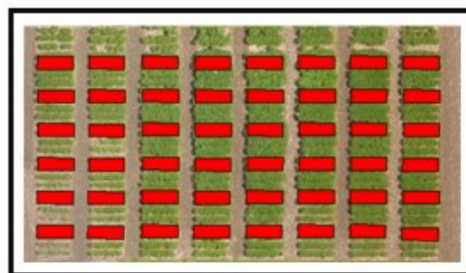


Figura 1. Principais etapas de fenotipagem de um esquema de trabalho referente a um programa de melhoramento de plantas utilizando imagens coletadas via drone. A) Plano de voo, B) Captura das imagens por meio de sensores acoplados. C) Organização e agrupamento das imagens. D) Segmentação das parcelas experimentais.

Pré-requisitos de voo

Após a escolha da plataforma de fenotipagem, o pesquisador necessita realizar alguns pré-ajustes e atender algumas normas vigentes para iniciar a coleta de dados. O responsável por realizar os voos no campo experimental deve atentar, primeiramente, para as normas e condições regulamentares de cada região. Entender como registrar os equipamentos junto aos órgãos brasileiros competentes e suas principais regras de uso é fundamental antes mesmo da aquisição do equipamento de fenotipagem. Para fins de demonstração didática, vamos abordar mais detalhadamente as normas brasileiras da Agência Nacional de Aviação Civil (ANAC, <https://www.gov.br/anac/pt-br/assuntos/drones>) aplicadas a RPAs. É importante também estar atento ao Regulamento Brasileiro de Aviação Civil Especial nº 94/2017 (RBAC-E nº 94/2017), e às normas de operações de drones estabelecidos pelo Departamento de Controle do Espaço Aéreo (DECEA) e pela Agência Nacional de Telecomunicações (ANATEL).

Para planejar a aquisição do equipamento de fenotipagem, o drone, por exemplo, o pesquisador deve atentar não somente para o tipo de dado fenotípico a ser coletado ou sensor disponível, mas também para as normas vigentes de voos para uma determinada área de melhoria. Locais experimentais ideais para uso de drones na coleta de dados devem estar distantes de um aeródromo, aeroporto, grupo de pessoas e rodovias principais, de acordo com as recomendações vigentes da ANAC. Além disso, áreas planas e sem relevo evidente devem ser preferidas, pois podem favorecer o processamento e acurácia na análise dos dados das imagens e evitar possíveis ruídos nas imagens. Não será abordado neste capítulo, mas atualmente existe um grande entrave para o uso de drones comerciais de pulverização agrícola, pois devido ao seu alto peso de decolagem (geralmente superior a 25 kg) e baixa altura de voo, as restrições de localidade e segurança podem inabilitar a utilidade do equipamento para um determinado local experimental.

Ao realizar a aquisição de um drone, o responsável necessita cadastrar o equipamento junto a ANAC (plataforma SISANT). Geralmente esses equipamentos para uso não recreativo possuem peso máximo de decolagem acima de 250 g e menor que 25 kg (somando o peso do equipamento, de sua bateria e de eventual carga), o que os classifica como RPA de Classe 3. Outro ponto importante é estar atento à altura de voo, que pode variar consideravelmente, dependendo de vários fatores, como modelo do drone, tipo de sensor utilizado, resolução da imagem desejada e objetivo da pesquisa. Atualmente, os modelos mais comuns de drones disponíveis no mercado são drones de asas fixas e drones de multirrotores ou multi-hélices, conforme descrito anteriormente. No entanto, essa última classe de drones é a mais acessível e utilizada

na fenotipagem de plantas, na qual a altura de voo recomendada geralmente não ultrapassa os 120 m (400 pés) de altitude acima do solo. Casos em que o pesquisador necessita realizar voos superiores a 120 m acima do nível do solo, o piloto remoto do aeromodelo deverá possuir licença e habilitação (certificado de aeronavegabilidade), além do equipamento ser registrado na ANAC. Adicionalmente, para voos acima de 120 m de altitude do solo o operador também deve ficar atento a diversos outros requisitos de documentos da aeronave e autorização de projeto de voos. Neste capítulo, iremos abordar um estudo de caso no qual o RPA pertence a Classe 3, ou seja, não é necessária a obtenção do Certificado de Aeronavegabilidade Especial para RPA – CAER, segundo as normas e diretrizes da ANAC. Nesse tipo de classe, o drone deve ser operado na linha visual (VLOS) e até 400 pés (aproximadamente 120 m) acima do nível do solo e ser cadastrado no Sistema de Aeronaves não Tripuladas (SISANT – <https://www.gov.br/anac/pt-br/assuntos/drones/cadastro-de-drones>). As regras e normas para operação de um RPA nos EUA, por exemplo, são diferentes das exigidas no Brasil. Para essa classe de drones, o piloto remoto responsável deve possuir o certificado de Piloto Remoto de Pequenas Aeronaves Não Tripuladas (Part 107) emitido pela FAA (*Federal Aviation Administration*). Além disso, dependendo da instituição onde irá realizar as atividades de voo, o piloto deve obter aprovação legal da própria instituição para a coleta de dados utilizando os RPA.

Preparo do campo experimental

Após a aquisição do equipamento e da obtenção de todos os documentos legais exigidos para realizar os voos, o próximo passo é o preparo do campo experimental de melhoramento. Como já mencionado anteriormente, deve-se preferir locais mais planos, distante de aeródromo ou aeroporto e com condições climáticas favoráveis. Áreas com histórico de ventos fortes constantes e de excesso de chuvas e más condições de drenagem tendem a dificultar o planejamento do número de voos, assim como aumentar o ruído no processamento das imagens. Após delimitadas as áreas dos ensaios experimentais, é importante o pesquisador estar atento à distribuição dos Pontos de Controle Terrestres (GCPs – *Ground Control Points*) ao longo do campo. Os GCPs são pontos cujas coordenadas são conhecidas e podem ser facilmente identificadas nas imagens, cujo objetivo é georreferenciar os projetos de fenotipagem a fim de reduzir o ruído nas imagens. Portanto, os GCPs podem ser quaisquer marcas no campo experimental feitas com algum objeto ou placas, de forma que seu centro possa ser visto nas imagens aéreas e possuírem coordenadas geográficas capturadas por algum instrumento acurado de georreferenciamento, como por exemplo utilizando o RTK (*real*

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

time kinematic). Recomenda-se inserir os GCPs nas extremidades de cada ensaio e ter alguns localizados no seu centro, de modo que, no mínimo, três a seis GCPs devam ser posicionados de forma equidistantes ao longo de toda extensão do campo. No passo seguinte, essas marcações serão utilizadas para corrigir o georreferenciamento pixel a pixel do projeto, e a maioria dos softwares que realizam essa etapa aceitam, no mínimo, três marcações. O incremento do número de GCPs distribuídos ao longo do campo experimental tende a aumentar consideravelmente a acurácia de georreferenciamento e redução do ruído, principalmente em terrenos com alta desuniformidade do solo e em ensaios experimentais consideravelmente grandes. Os GCPs devem permanecer no campo durante todo o desenvolvimento da cultura e coleta de dados, e devem ser fotografados em todos os voos durante as coletas de imagens no campo. Além disso, uma grande vantagem do uso dos GCPs é permitir melhor sobreposição das delimitações das parcelas experimentais, facilitando o desenho de polígonos ou *grid shapefile*, os quais são gerados apenas uma vez para todos os projetos (voos).

Plano de voo e preparo

O plano de voo é uma etapa fundamental durante a etapa de coleta de dados, pois pode determinar diretamente a qualidade das imagens e influenciar nos resultados da fenotipagem (Figura 1A). O plano de voo consiste no mapeamento da área de voo, devendo indicar as configurações utilizadas durante o voo pelo RPA. O software será responsável por enviar esses dados e configurações para o drone, o qual irá executar a missão. Existem, atualmente, diversos softwares que fazem o plano de voo, entre eles *DJI Go*, *Pix4D Capture*, *DJI GS Pro*, etc. Como cada um tem suas peculiaridades, se fará uma abordagem mais abrangente, considerando as configurações principais: *grid* de mapa, altitude, sobreposição, e tempo estimado de voo.

O plano de voo deve ser executado anteriormente à ida ao campo, no entanto, para maior assertividade do local de voo, recomenda-se a confirmação das extremidades do *grid* da missão percorrendo o campo. Essa tarefa pode ser feita apenas uma vez, pois os demais voos de uma determinada área experimental podem adotar as mesmas configurações e ajustes. No plano de voo, o percurso do drone deve cobrir com sobras laterais o ensaio experimental, recomendando-se ao menos 75 % de sobreposições laterais e verticais das imagens. O pesquisador deve atentar para o tempo de voo, uma vez que a vida útil da bateria durante o voo costuma ser curta e, em grandes áreas, a quantidade de baterias disponíveis pode ser um problema para a coleta das imagens planejadas no plano de voo. Por fim, a altura de voo é fundamental para se obter uma resolução de imagem desejável. Essa configuração vai depender principalmente

da característica que se deseja avaliar via imagens, mas também da quantidade de baterias (tempo de voo) e dos recursos computacionais para o processamento das imagens disponíveis. Um exemplo prático é a fenotipagem em larga escala para realizar a contagem de estande. Nesse caso, o voo deve ser feito em baixas altitudes a fim de aumentar a resolução das imagens (diminuir o tamanho dos pixels), pois o objeto alvo (planta em estádios iniciais de desenvolvimento) é bem pequeno. Logo, a quantidade de imagens do projeto aumenta consideravelmente, requisitando maiores recursos computacionais para o processamento da grande quantidade de imagens obtidas durante o voo. Por outro lado, características que não exigem alta resolução de imagens (altura de planta, por exemplo), tendem a utilizar muito menos recursos para o processamento das imagens.

Por fim, para preparar todos os componentes necessários para executar a missão de voo, o pesquisador deve estar atento a algumas condições técnicas antes mesmo de ir ao campo, como por exemplo: atualizar todos os softwares de voo e pilotagem, realizar a carga das baterias do drone, do controle e do tablet ou iPad. Além disso, as condições climáticas devem ser verificadas com antecedência por meio de aplicativos específicos que recomendam o melhor dia e horário de voos.

Captura das imagens

Uma vez checado todos os requisitos legais de voo e estabelecidas as configurações do plano de voo para a decolagem do drone, o voo pode ser realizado com qualidade e segurança (Figura 1B). Também não se irá abordar neste capítulo os aspectos técnicos de como realizar os voos, no entanto se recomenda que cada grupo de pesquisa ou equipe tenha seu próprio protocolo de voo elaborado e que todos estejam familiarizados em realizar todas as etapas de forma mais uniforme possível. Um ponto crucial são os horários de voos, pois as condições de luminosidade podem influenciar muito a refletância do objeto alvo em estudo e os resultados serem bastante influenciados pelo sombreamento. Sempre que possível, evitar voos com céu nublado e ventos fortes, a fim de reduzir o ruído do efeito do sombreamento e das movimentações das plantas, respectivamente. Em geral, recomenda-se a coleta das imagens ao meio-dia solar da região para reduzir o sombreamento entre plantas e garantir o máximo de iluminação do dia.

Dependendo do tipo de sensor (câmera) a ser utilizado durante o voo, o pesquisador deve adotar algumas práticas de qualidade para o ajuste do equipamento antes da decolagem, entre elas a calibração radiométrica. As correções radiométricas são usadas para melhorar a qualidade dos dados, permitindo corrigir a refletância

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

da imagem, levando em consideração a iluminação do local e a influência do sensor (câmera multiespectral, por exemplo). Essa etapa é necessária para poder comparar imagens tiradas em diferentes pontos no tempo sob diferentes condições climáticas e calcular valores de índice confiáveis. Os valores de pixel das imagens dependem das condições de iluminação e das configurações do sensor da câmera, entre outras variáveis, as quais geralmente mudam entre os voos. A calibração radiométrica resolve esse problema, compensando variáveis externas para estimar uma propriedade física do campo. Logo, essa correção permite comparar dados de diferentes dias ou locais, o que é muito comum em programas de melhoria de plantas.

Durante o voo, o drone irá seguir as configurações adotadas no plano de voo, em que a quantidade de fotos tiradas será dada em função basicamente da altura de voo e da porcentagem de sobreposição de imagens. As imagens geralmente são salvas no formato JPEG ou PNG para futuras análises. Dependendo do tipo de sensor utilizado, as imagens podem ser obtidas no formato de cores que são visíveis ao espectro de luminosidade sensível humano, conhecidas como RGB. Neste caso, apenas uma imagem possui as três bandas (vermelho, verde, e azul), sendo que esse tipo de sensor produz o conjunto de dados que comumente chamamos de ortomosaicos. No caso de câmeras que possuem vários canais de bandas, como as câmeras multiespectrais, cada banda é salva separadamente, e, nesse caso, são utilizadas para a geração dos mapas de refletâncias, que por sua vez são utilizados para a obtenção dos índices vegetativos.

Agrupamento das imagens

Considerando o esquema de trabalho mais comumente utilizado na fenotipagem atualmente, as imagens obtidas pelas câmeras RGB podem gerar os ortomosaicos (Figura 1C). Ortomosaico é um produto de imagem ortorretificado fotogrametricamente de uma coleção de imagens, em que a distorção geométrica foi corrigida e as imagens foram balanceadas em cores para produzir um conjunto de dados de mosaico contínuo, ou seja, as imagens em formato original são utilizadas para se agruparem e são corrigidas pixel a pixel para formar a imagem completa da área experimental sobrevoada. Esse método remove as distorções de perspectiva das imagens usando o modelo de superfície digital ou superfície de terreno e consegue trabalhar com todos os tipos de superfícies, bem como grandes conjuntos de dados (Ruzgienie et al., 2015). As distâncias são preservadas e, portanto, o ortomosaico pode ser usado para medições. Dentre as diversas funcionalidades, pode-se utilizar

as informações obtidas pelo ortomosaico para extrair informações de características morfológicas das plantas em avaliação, assim como gerar os índices vegetativos oriundos das bandas RGB.

Atualmente, diversos softwares de fotogrametria podem ser utilizados na obtenção do ortomosaico e, dentre os mais conhecidos, tem-se *Pix4D*, *Agisoft PhotoScan* e o *Open-Drone-Map* (ODM). Durante essa etapa do processamento das imagens, os GCPs são utilizados para a realizar a georreferência das imagens pixel a pixel, como citado anteriormente. Diferentemente dos sensores de profundidade tridimensionais, como o LiDAR, esses softwares de fotogrametria se baseiam nos algoritmos de *Structure from Motion* (SfM). Esse método SfM é uma técnica que utiliza uma série de imagens bidimensionais para reconstruir a estrutura tridimensional de uma cena ou objeto. O SfM pode produzir modelos 3-D baseados em nuvem de pontos semelhantes ao LiDAR, e pode ser usado para criar modelos de superfície digital de alta resolução (incluindo modelos digitais de elevação) e modelos de objetos utilizando câmeras digitais. Os avanços da tecnologia na área de computação gráfica, juntamente com o desenvolvimento de novas câmeras digitais e de aeronaves remotamente pilotadas nos últimos anos, possibilitaram o uso dessa tecnologia para uma ampla gama de usuários gerarem modelos 3-D de forma rápida e acurada.

Segmentação das parcelas

A etapa de segmentação e identificação das parcelas experimentais, a qual é comumente conhecida por *field-grid* ou *shapefile*, pode ser elaborada de diversas maneiras, entre elas por meio de bibliotecas do software R (*UAStools* e *FIELDimageR*), Python e ferramentas de GIS como o QGIS e ArcGIS, por exemplo (Figura 1D). Essa etapa, apesar de ser relativamente simples, tem um peso muito grande no resultado final das análises. Para o incremento da acurácia da característica em avaliação, a assertividade da delimitação das parcelas deve ser alta, uma vez que os dados são formados pelos valores dos pixels referentes ao alvo em estudo. Portanto, somente pixels pertencentes a esse alvo (neste caso a parcela experimental de melhoramento) devem ser utilizados para contabilizar uma dada informação da parcela.

Em análises de dados coletados ao longo do tempo por vários voos de uma mesma área experimental, os GCPs têm uma função fundamental na elaboração dos *grid-shapefiles*. A elaboração da segmentação das parcelas é uma etapa relativamente simples, mas pode se tornar uma tarefa muito dispendiosa, dependendo do tamanho do campo experimental e das dimensões das parcelas. Portanto, manter padronização

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

entre tamanho de parcelas, enquadramento experimental bem feito, e a mesma posição geográfica dos diversos projetos (ortomosaicos) ao longo dos voos, reduzem o tempo gasto para a elaboração das segmentações das parcelas, pois, nesse caso, essa etapa pode ser elaborada uma única vez.

Análise de Dados da Fenotipagem de Alto Rendimento na Prática

Para ser válida em um programa de melhoramento de plantas, a Fenômica precisa beneficiar ao menos um dos parâmetros de ganho de fenotipagem, quais sejam: (1) custo da operação, (2) tempo de realização, (3) qualidade de dados e (4) bem-estar do colaborador.

Custo é um fator limitante para a fenotipagem de muitas características, como as relacionadas a análises laboratoriais químicas. Alguns exemplos são as características relacionadas à qualidade nutricional (ex. concentração de proteína bruta, celulose, fibra, etc.), compostos químicos primários e secundários, teor de açúcares, tipos de lipídeos, entre muitos outros. Nesse caso, tecnologia como a *Near Infrared Reflectance Spectroscopy* (NIRS) vem sendo excelente opção para o ganho em rendimento de trabalho e redução de custos, permitindo, por exemplo, ampliar o número de parcelas experimentais e, conseqüentemente, a variabilidade genética avaliada por geração de seleção (Molano et al., 2016). Um exemplo é relatado no melhoramento de forrageiras tropicais, em que a tecnologia NIRS permitiu ampliar o tamanho experimental de centenas de parcelas para milhares de parcelas em fases iniciais (Mazabel et al., 2020).

Em outra esfera, indo além dos drones, com base em modelos estatísticos e algoritmos de aprendizado de máquina, também é possível estabelecer relações entre os espectros obtidos por NIRS e as variáveis de interesse, permitindo a predição de características ou propriedades que seriam difíceis de medir diretamente, como inúmeras características da madeira em *Eucalyptus* spp., tal como a densidade básica e o rendimento de celulose (Castro et al., 2019), prever a qualidade em grãos (Barbon-Junior et al., 2020) e na determinação do teor de açúcares solúveis (como o BRIX) em frutas (Walsh et al., 2020).

Tempo e custo estão fortemente associados e, assim, a redução do tempo para obtenção dos dados pode acelerar a tomada de decisão e, potencialmente, aumentar o ganho genético. Segundo a famosa equação do melhorista, o tempo, por estar

no denominador, provoca a redução do ganho genético e, assim, ao reduzir esse tempo, o melhorista consegue maximizar os potenciais ganhos com a seleção. Um grande exemplo é a possibilidade de desenvolver modelos preditivos de produtividade e crescimento, podendo associá-los a marcadores moleculares (Toda et al., 2021) ou mesmo praticar a seleção em fases iniciais, em que normalmente os dados de produtividade não são coletados devido ao grande número de parcelas, como por exemplo, no melhoramento da batata. Informações podem ser coletadas com sensores embarcados em drones para avaliação da curva de crescimento esperada dos genótipos e, nesse caso, as testemunhas são fundamentais para o ajuste do modelo. Um exemplo é o uso de (1) altura de plantas estimada e (2) índices vegetativos coletados durante o ciclo da cultura da batata para predição da produtividade final (Figura 2A). Os valores preditos por parcela são então colocados em aplicativos digitais (celulares ou tablets, por exemplo) e usados pelo melhorista para dar suporte no momento da seleção visual. Normalmente, em fases iniciais do melhoramento, somente características de alta herdabilidade são selecionadas, como cor da casca, e formato e deformações no tubérculo. Dessa forma, por antecipar um ciclo na seleção para produtividade, o melhorista consegue elevar o ganho de seleção esperado.

Se você trabalha com análise de dados, provavelmente já deve ter ouvido a frase em inglês “*garbage in, garbage out*” (lixo entra, lixo sai), que literalmente define uma coleta de dados malconduzida e sem protocolo. Qualidade de dados é um dos principais gargalos para uma seleção efetiva que tem sido relatado por melhoristas e pesquisadores. A qualidade na coleta de dados é a chave para o sucesso de um programa de melhoramento, e pode ser um grande problema quando resultar na eliminação de um genótipo superior, e/ou seleção de um ruim. Bons exemplos do uso de imagens para seleção são os casos de características categóricas ou que seguem escalas, como avaliação de doenças e maturidade (Matias et al., 2020). Por exemplo, avaliações de maturidade são normalmente baseadas no nível ou intensidade de verde em que os olhos do colaborador/melhorista estão diagnosticando, levando em conta a iluminação presente em determinado momento do dia. Os olhos são os sensores que levam à tomada da decisão em classificar um genótipo como precoce, moderado ou tardio. Conforme o esperado, os olhos de pessoas diferentes possuem morfologias diferentes e podem acessar a intensidade de verde de forma diferente. Assim, pensando em uma escala de 1 (precoce) até 9 (tardio) é muito fácil confundir notas próximas, podendo um colaborador/melhorista dar nota 5 para um genótipo que foi classificado como 7 por outro (Figura 2B). O uso de um único sensor embarcado em uma plataforma de fenotipagem auxilia na padronização da coleta de dados em todo o experimento, permitindo o cálculo de índices vegetativos associados à intensidade de

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

verde no campo. Isso promove a (1) redução do tempo da coleta de dados em áreas grandes, o que facilita manter a mesma iluminação entre leituras, e (2) transforma uma característica categórica em quantitativa, sendo assim muito mais acurada a tomada de dados para essa característica.

O bem-estar do colaborador cada vez mais tem sido considerado como um parâmetro decisivo no melhoramento moderno. Algumas atividades de fenotipagem podem ser perigosas ou mesmo morosas e repetitivas, podendo levar ao desgaste das principais engrenagens do programa, os colaboradores. Assim, a inclusão de ferramentas ou mesmo novas formas de coleta de dados que facilitem e tornem mais prazeroso realizar uma atividade devem ser consideradas. Um grande exemplo é a contagem de número de plantas em campo, que demanda muitas pessoas e horas em pé, o que se torna exaustivo em dias ensolarados, comuns nas regiões tropicais. Reconhecimento digital das plantas seguido da contagem automatizada por parcela baseado em algoritmos de inteligência artificial e visão computacional têm sido cada vez mais implementados nos programas de melhoramento de todas as espécies (Karami et al., 2020; Wang et al., 2021). Outros exemplos são a contagem de quaisquer características cuja imagem segue um padrão passível de treinar uma máquina para o reconhecimento, como número de flores, esporos de fungos, sementes, grãos de pólen (Figura 2C), entre outros.

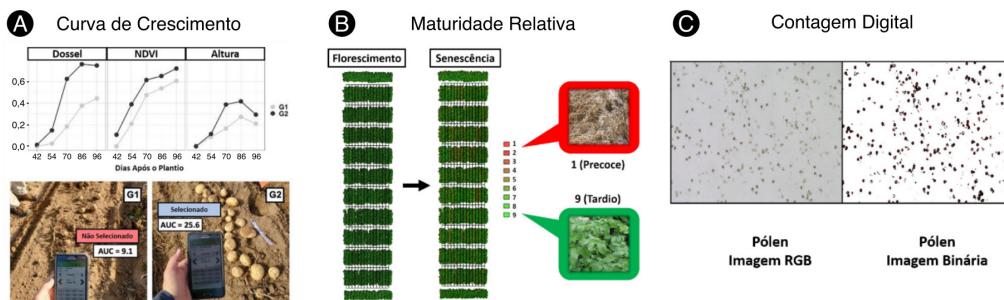


Figura 2. A) Avaliação da curva de crescimento de dois genótipos de batata (G1 e G2) durante 100 dias após o plantio por meio de características coletadas/calculadas com o uso de imagens aéreas multiespectrais (Dossel, NDVI e Altura). B) Ilustração de dois mosaicos mostrando o período de florescimento (todas as plantas verdes) e senescência em uma população de melhoramento de batata para avaliar maturidade relativa que tradicionalmente é avaliada por nota visual, variando de 1 (precoce) até 9 (tardio). O objetivo é fazer a avaliação por meio de índices vegetativos relacionados com a intensidade da cor verde. C) Uso de segmentação de imagens para diferenciar pixels de pólen do plano de fundo em uma imagem RGB e fazer a contagem dos objetos (pólen) na imagem binária (pixels com valor 1 referente ao pólen e 0 referente ao plano de fundo).

Exemplos de como Aplicar Fenômica para Avaliar Plantas

A aplicação de tecnologias digitais e técnicas de Fenômica para avaliar plantas no campo ou ambientes controlados é baseada na criatividade e inovação. Diferentes métodos e equipamentos podem ser utilizados em função da (1) disponibilidade, (2) custo, (3) praticidade, e (4) nível de exatidão/proximidade com o real desejado. Em muitos casos os valores oriundos de características digitais não são parecidos em termos de unidade com as características agronômicas tradicionais, no entanto, o *rank* (classificação) dos tratamentos é o mesmo. Neste caso, o pesquisador deve medir bem o custo-benefício da tecnologia utilizada para fazer a fenotipagem e os resultados/objetivos da atividade.

Altura de Plantas e Acamamento

Em Fenômica, a altura de plantas é facilmente medida usando o modelo de altura do dossel (MAD) ou a altura estimada da planta (AEP) calculada por meio do modelo digital de superfície (MDS) ou por meio de um sensor baseado em laser, como LiDAR, a partir do qual os dados brutos podem ser transformados em uma nuvem de pontos tridimensional (3D) (Walter et al., 2019). O MDS é um produto da etapa de montagem do mosaico (*orthomosaic*) quando as imagens possuem informações georreferenciadas ou também chamadas de mapa de elevação de superfície. O método começa por coletar imagens antes das plantas começarem a emergir no solo, momento também chamado de “Imagem da Base do Solo” (Matias et al., 2020). As próximas imagens são coletadas mais tarde no ciclo de crescimento, quando as plantas já atingiram a altura desejada para ser avaliada no programa de melhoramento. No exemplo da Figura 3A, o MAD foi utilizado para avaliação de plantas de batata a 40 e 70 dias após o plantio (DAP). É possível identificar que alguns genótipos de maior ou menor estatura já eram visíveis em fases iniciais de crescimento, fato que se manteve ao longo do tempo. O melhorista pode, por exemplo, utilizar essa informação para antecipar a seleção de genótipos desejados e reduzir o tempo para o próximo ciclo.

Outro exemplo do uso desta técnica é a avaliação do acamamento. Acamamento é quando as plantas perdem sua posição vertical, tornando-as difíceis de colher e reduzindo drasticamente a produtividade. No exemplo da Figura 3B, o primeiro conjunto de imagens foi coletado durante o florescimento, antes de um evento de

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

chuva forte que provocou o acamamento de alguns indivíduos na população de melhoramento de trigo (Matias et al., 2022). Outro conjunto de imagens foi coletado no dia seguinte para investigar a habilidade dessa abordagem na identificação de indivíduos acamados.

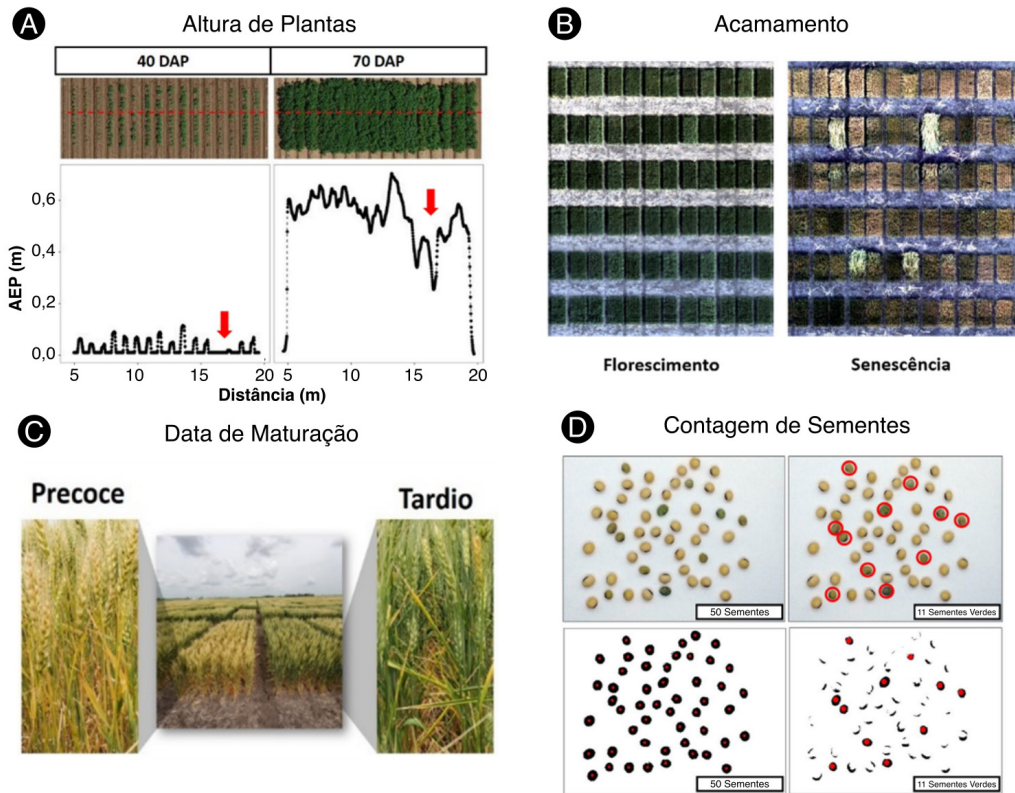


Figura 3. A) Altura estimada de plantas (AEP) medida em 40 e 70 dias após o plantio (DAP) de batatas por meio do modelo de altura de dossel (MAD). B) Comparação das parcelas de uma população de melhoramento de trigo antes (florescimento) e depois de um evento de acamamento. C) Genótipo precoce de trigo ao lado de um genótipo tardio facilmente diferenciados por meio de índices vegetativos digitais. D) Contagem de sementes verdes em soja avaliadas por meio de segmentação de imagens.

Ao comparar a altura estimada de plantas (AEP) pelo modelo de dossel foi possível verificar que os genótipos com altas pontuações de acamamento também estavam entre as plantas mais altas aos 73 DAP (conjunto superior de boxplot, Figura 4A). Os

mesmos genótipos tiveram uma redução de AEP aos 74 DAP após a chuva, indicando que esses genótipos acamaram (conjunto central de boxplot, Figura 4A). Um terceiro voo para coletar imagens dessa população foi feito alguns dias mais tarde, aos 80 DAP durante a fase de maturação. Ao calcular a AEP, foi possível diagnosticar que alguns dos genótipos que acamaram apresentaram elevada plasticidade fenotípica e retornaram à posição vertical (conjunto inferior de boxplot, Figura 4A). Essa informação pode ser usada pelo melhorista para inferir a capacidade de recuperação e selecionar potenciais genótipos de alto rendimento.

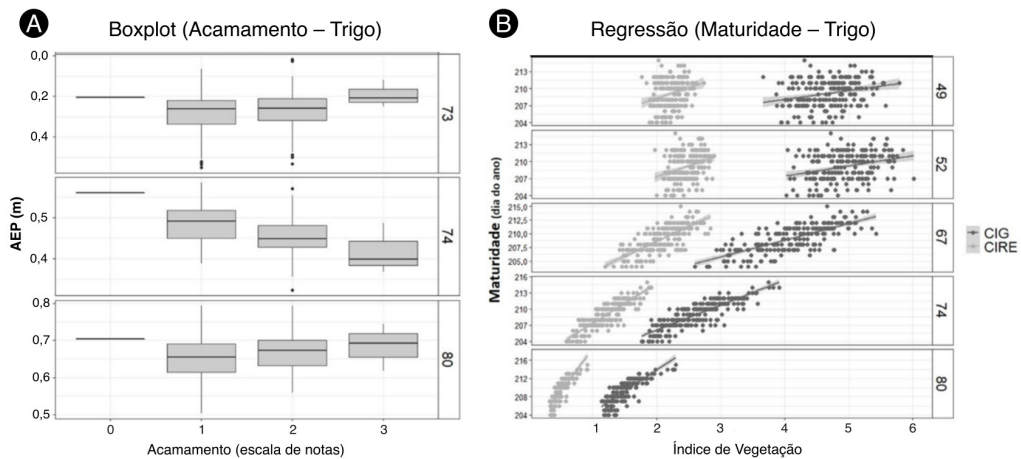


Figura 4. A) Efeito da altura estimada de plantas (AEP) com base no modelo digital de superfície (MDS) em relação à nota visual de acamamento. As imagens foram coletadas um dia antes (73 DAP), um dia após (74 DAP) e uma semana após (80 DAP) um evento de acamamento causado por tempestade e ventos fortes. B) Regressão entre índices vegetativos e data de maturidade de trigo coletados entre 49 e 80 dias após o plantio (DAP). CIG (*Chlorophyll Index – Green*) e CIRE (*Chlorophyll Index – Red Edge*).

Maturidade relativa

A maturidade é outra característica importante no melhoramento de plantas e tem grande potencial para ser medida usando sensores embarcados em drones, pois está relacionada à intensidade da cor verde. Por exemplo, os índices de vegetação são ótimas ferramentas para identificar quando as plantas estão fisiologicamente prontas para serem colhidas (Volpato et al., 2021a). No exemplo da Figura 4B, uma regressão linear simples foi usada para avaliar a relação entre índices de vegetação (CIG e CIRE) e data de maturidade em trigo (Matias et al., 2022). É possível identificar um ótimo momento para coleta de dados com drones (dia de voo) por volta dos 65 DAP,

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

quando já é possível observar alta correlação entre as medidas coletadas no campo e o valor dos índices extraídos das imagens por parcela. No caso do trigo, o genótipo é considerado fenotipicamente maduro quando 50% dos pedúnculos das plantas na parcela estão amarelos, conforme representado à esquerda na imagem da Figura 3C. Essa técnica pode ser utilizada também para avaliação de doenças ao nível de campo, deficiência ou intoxicação de nutrientes de solo, tolerância à deficiência hídrica, entre várias outras características medidas por meio da intensidade de cores na parcela (Shakoor et al., 2017).

Contagem de sementes verdes

Sementes de soja esverdeadas ocasionam deterioração na qualidade fisiológica do lote e impactam no estande final de plantas. Da mesma forma, na indústria, grãos verdes elevam o custo de processamento por reduzirem a estabilidade oxidativa do óleo e retardarem o processo de hidrogenação (França-Neto et al., 2012). Em geral, os atuais métodos de avaliação de qualidade de lotes de sementes são realizados de forma manual, sendo considerados de baixa efetividade e alta subjetividade. Como essa característica é avaliada por meio da coloração e da contagem das sementes/grãos, pode ser facilmente avaliada por processos digitais de identificação e quantificação. No Figura 3D, um sensor RGB comum de celular foi utilizado para identificar e contar sementes verdes em um lote de soja (Antunes Neto et al., 2020). O método de segmentação de imagens chamado *thresholding* ou limiar de corte foi utilizado primeiro para separar o que era semente do fundo da imagem e segundo o que era semente verde da comum. Esse método consiste em calcular um índice usando as diferentes camadas da imagem RGB (vermelho, verde e azul) para criar uma camada única em que o objeto de interesse se diferencie dos demais. Com isso, é fácil escolher um ponto de corte e informar os algoritmos para classificar os pixels da imagem em verdadeiro (1) ou falso (0), criando uma imagem binária que é utilizada para a identificação dos objetos isolados. Para identificação de objetos, diferentes algoritmos podem ser utilizados, no entanto, o algoritmo Linhas Divisoras de Água ou *Watershed* se destaca por sua praticidade e simplicidade. Ele calcula a matriz de distância entre cada pixel e utiliza esse critério para decidir se um pixel pertence a um objeto ou ao outro, determinando assim as bordas dos objetos. Esse método identificou 50 sementes na imagem como um todo e, em seguida, 11 sementes verdes (Figura 3D). A mesma técnica pode ser utilizada para contagem de frutos, pólen, flores, manchas de doenças, entre inúmeras outras aplicações em que a segmentação de imagem possa ser aplicada.

Estudo de caso em detecção de plantas usando *deep learning*

Na fase inicial do ciclo de crescimento das plantas à campo pode se realizar voos para analisar/avaliar se houve germinação uniforme, e contagem de plantas que germinaram. A partir dessas informações, é possível realizar acompanhamento durante todo o ciclo fenológico de cada planta. Além da posição geográfica obtida por meio da identificação automática por algoritmos de visão computacional, o uso de imagens de sensores RGB ou multiespectral pode ser muito útil para análise de diversos fatores fenológicos. Todas os três tipos básicos de plataformas de fenotipagem de plantas descritos anteriormente (plataformas terrestres, plataformas aéreas e plataformas espaciais) podem ser usados para coleta das imagens em função da espécie (arquitetura e tamanho das plantas), área de avaliação (campo experimental ou produção) e tempo/custo de aquisição.

Multilayer Perceptrons (MLP), CNN, *Recurrent Neural Networks* (RNN) e *Long Short-Term Memory* (LSTM) são alguns dos modelos de DL comumente usados em aplicações de fenotipagem em culturas agrônômicas (Singh et al., 2018; Niazian; Niedbała, 2020). Avanços recentes em automação, computação e tecnologia de sensores permitiram a coleta de dados fenotípicos em larga escala de alta resolução temporal. Essa grande quantidade de dados permitiu expandir algoritmos de DL com sucesso frente a uma ampla variedade de tarefas de fenotipagem de plantas. Existem diversas maneiras de identificar objetos em imagens usando algoritmos de visão computacional. Vale destacar que tanto os algoritmos tradicionais quanto os mais avançados necessitam de treinamento de redes neurais convolucionais. Nos últimos anos, com o avanço desses algoritmos baseados em redes neurais, houve grande aumento na precisão e confiabilidade dos resultados obtidos. A seguir, será descrito o passo a passo para desenvolver um algoritmo de contagem de plantas baseado em DL utilizando como exemplo um conjunto de dados público (Tseng et al., 2022).

Conjunto de dados (*Dataset*)

Para este exemplo, o conjunto de dados utilizado foi o *RiceSeedling detection dataset*, disponibilizado pelo Taiwan Agriculture Research Institute, (Yang et al., 2021). Ele é formado por 600 imagens de 320 x 320 pixels e as anotações das localizações de cada planta (Figura 5A). As anotações estão em um formato XML com as informações dos *bound boxes* (caixas delimitadoras) dos objetos e a classe. Cada anotação compreende o identificador (ID) da imagem, a classe do objeto, e os quatro valores de coordenadas xmin, ymin, xmax e ymax, correspondente à posição do objeto na imagem (Figura 5B).

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

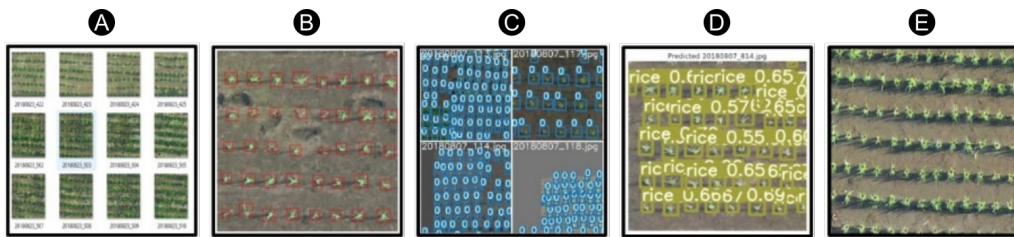


Figura 5. Passo a passo do desenvolvimento de um algoritmo de contagem de plantas. (A) Coleta de imagens, (B) identificação do objeto alvo, (C) treinamento do modelo com o máximo de imagens possível, (D) validação das imagens e verificação da acurácia do modelo, e (E) aplicação e contagem final.

Metodologia

Diversas arquiteturas de detecção de objetos foram criadas, e uma das mais utilizadas é a família YOLO (Redmon et al., 2016). Para este exemplo vamos utilizar a YOLOR, que é uma rede unificada que integra conhecimentos implícito e explícito. De forma resumida, pois não é o foco deste capítulo, esse método começa com o pré-treinamento de uma rede de conhecimento implícito com todas as tarefas presentes no conjunto de dados COCO para aprender uma representação geral, ou seja, conhecimento implícito. Para otimizar tarefas específicas, YOLOR treina outro conjunto de parâmetros que representam conhecimento explícito. Tanto o conhecimento implícito quanto o explícito são usados para inferência.

Treinamento do modelo

Para treinar a arquitetura YOLOR foi utilizado o ambiente de computação em nuvem Google Colab, que possibilita o acesso a uma GPU para aceleração do treinamento. As imagens de entrada e as anotações foram separadas em dois conjuntos de dados, um para treinar o modelo, outro para testar o modelo (Figura 5C). Foi usado o *framework* *Pytorch* e posteriormente o clone do GitHub da implementação da YOLO. Posteriormente, os dados de treinamento foram carregados e treinados por 10 épocas. Maiores informações estão disponíveis no GitHub do projeto em <https://github.com/aipal-nchu/RiceSeedlingDataset>.

Resultados

Após o treinamento, os dados de teste foram utilizados para realizar a detecção das plantas. O modelo recebe a imagem de teste e retorna os *bound boxes* e a

classe do objeto (Figura 5D). Com o modelo treinado, é possível realizar um fluxo para prever os objetos em um ortomosaico. Primeiro se divide o ortomosaico em imagens com o mesmo tamanho daquelas que foram usadas no treinamento. Essas imagens continuam georreferenciadas e depois servirão para obter as coordenadas geográficas dos objetos detectados. A partir das previsões de cada uma das imagens, obtêm-se os pontos centrais dos objetos detectados e converte-se em coordenadas geográficas. Assim é possível obter um mapa de pontos de plantas detectadas para um campo inteiro (Figura 5E).

Considerações Finais

Fenômica é um antigo campo da ciência ao pensar em medidas manuais de fenotipagem, porém, é uma ciência nova e promissora em termos de dados digitais a serem explorados, e que podem ser aplicados não somente no melhoramento de plantas, mas também está se tornando proeminente em outras áreas de pesquisa, como fitotecnia, produção agrícola e agricultura de precisão. Os avanços tecnológicos do sensoriamento remoto, juntamente com o desenvolvimento de novos softwares de análise de dados e inteligência artificial, tendem a melhorar significativamente os resultados obtidos na fenotipagem e, conseqüentemente, de todo o esquema de trabalho do programa de pesquisa. A fenotipagem é mais uma ferramenta disponível ao melhorista dentre o seu conjunto de ferramentas que podem ser utilizadas no melhoramento de plantas. No entanto, metodologias sofisticadas de engenharia e softwares avançados para a análise de dados da Fenômica não substituem a necessidade de experimentação adequada e escolha apropriada dos métodos de melhoramento e análises estatísticas. A implementação adequada da HTP visando ganhos no melhoramento de plantas depende muito da validação metodológica e do esforço contínuo da comunidade científica. Plataformas de fenotipagem de baixo custo e a disponibilidade de códigos abertos devem ser incentivados e promovidos, o que exige esforços cooperativos entre as comunidades internacionais de fenotipagem de plantas. A rápida adoção dessas técnicas nos setores público e privado é facilitada pela capacidade de comparar essas medições com dados reais, fornecendo confiabilidade na tomada de decisão do pesquisador. Colaborações entre diferentes disciplinas são necessárias para realizar o potencial oferecido pelo HTP. Certas características complexas podem ser dissecadas por meio do estudo da Fenômica de plantas, como por exemplo, respostas que regulam os estresses abiótico e biótico, e a produtividade. Portanto, a HTP tem um grande potencial para beneficiar o melhoramento de plantas, contribuindo, deste modo, para a nova revolução digital no campo.

Referências

- ANTUNES NETO, A.; MATIAS, F. I.; PASSOS, A. M. A.; ROCHA, N. M. **Determinação de sementes de soja esverdeadas por meio de análise de imagens**. Sete Lagoas: Embrapa Milho e Sorgo, 2020. 21 p. (Embrapa Milho e Sorgo. Boletim de Pesquisa e Desenvolvimento, 212). <http://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/1126547>.
- BALLESTEROS, R.; ORTEGA, J. F.; HERNANDEZ, D.; MORENO, M. A. Onion biomass monitoring using UAV-based RGB imaging. **Precision Agriculture**, v. 19, n. 5, p. 840–857, out. 2018. DOI: <https://doi.org/10.1007/s11119-018-9560-y>.
- BARMEIER, G.; SCHMIDHALTER, U. High-throughput field phenotyping of leaves, leaf sheaths, culms and ears of spring barley cultivars at anthesis and dough ripeness. **Frontiers in Plant Science**, v. 8, 1920, Nov. 2017. DOI: <https://doi.org/10.3389/fpls.2017.01920>.
- BARBON-JUNIOR, S.; MASTELINI, S. M.; BARBON, A. P. A. C.; BARBIN, D. F.; CALVINI, R.; LOPES, J. F.; ULRICI, A. Multi-target prediction of wheat flour quality parameters with near infrared spectroscopy. **Information Processing in Agriculture**, v. 7, n. 2, p. 342-354, June 2020. DOI: <https://doi.org/10.1016/j.inpa.2019.07.001>.
- BLUMENTHAL, J.; MEGHERBI, D. B.; LUSSIER, R. Unsupervised machine learning via Hidden Markov Models for accurate clustering of plant stress levels based on imaged chlorophyll fluorescence profiles & their rate of change in time. **Computers and Electronics in Agriculture**, v. 174, 105064, July 2020. DOI: <https://doi.org/10.1016/j.compag.2019.105064>.
- CASTRO, C. A. O.; NUNES, A. C. P.; ROQUE, J. V.; TEÓFILO, R. F.; SANTOS, O. P.; SANTOS, G. A.; GALLO, R.; PANTUZA, I. B.; RESENDE, M. D. V. Optimization of *Eucalyptus benthamii* progeny test based on Near-Infrared Spectroscopy approach and volumetric production. **Industrial Crops and Products**, v. 141, 111786, Dec. 2019. DOI: <https://doi.org/10.1016/j.indcrop.2019.111786>.
- CONN, A.; PEDMALE, U. V.; CHORY, J.; NAVLAKHA, S. High-resolution laser scanning reveals plant architectures that reflect universal network design principles. **Cell Systems**, v. 5, n. 1, p. 53- 62.e1-e3, July 2017. DOI: <https://doi.org/10.1016/j.cels.2017.06.017>.
- DOBBELS, A. A.; LORENZ, A. J. Soybean iron deficiency chlorosis high-throughput phenotyping using an unmanned aircraft system. **Plant Methods**, v. 15, 97, Aug. 2019. DOI: <https://doi.org/10.1186/s13007-019-0478-9>.
- FIORANI, F.; SCHURR, U. Future scenarios for plant phenotyping. **Annual Review of Plant Biology**, v. 64, p. 267–291, Apr. 2013. DOI: <https://doi.org/10.1146/annurev-arplant-050312-120137>.
- FRANÇA-NETO, J. B.; PÁDUA, G. P.; KRZYZANOWSKI, F. C.; CARVALHO, M. L. M.; HENNING, A. A.; LORINI, I. **Semente esverdeada de soja: causas e efeitos sobre o desempenho fisiológico - Série Sementes**. Londrina: Embrapa Soja, 2012. 15 p. (Embrapa Soja. Circular técnica, 91). <http://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/924805>.
- KARAMI, A.; CRAWFORD, M.; DELP, E. J. Automatic plant counting and location based on a few-shot learning technique. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 13, p. 5872–5886, 2020. DOI: <https://doi.org/10.1109/JSTARS.2020.3025790>.
- KIRCHGESSNER, N.; LIEBISH, F.; YU, K.; PFEIFER, J.; FRIEDLI, M.; HUND, A.; WALTER, A. The ETH field phenotyping platform FIP: a cable-suspended multi-sensor system. **Functional Plant Biology**, v. 44, n. 1, p. 154-168, Oct. 2016. DOI: <https://doi.org/10.1071/FP16165>.
- LI, B.; XU, X.; HAN, J.; ZHANG, L.; BIAN, C.; JIN, L.; LIU, J. The estimation of crop emergence in potatoes by UAV RGB imagery. **Plant Methods**, v. 15, 15, Feb. 2019. DOI: <https://doi.org/10.1186/s13007-019-0399-7>.

- LI, X.; XU, X.; CHEN, M.; XU, M.; WANG, W.; LIU, C.; YU, L.; LIU, W.; YANG, W. The field phenotyping platform's next darling: Dicotyledons. **Frontiers in Plant Science**, v. 13, p. 935748, Aug. 2022. DOI: <https://doi.org/10.3389/fpls.2022.935748>.
- MATIAS, F. I.; CARAZA-HARTER, M. V.; ENDELMAN, J. B. FIELDimageR: an R package to analyze orthomosaic images from agricultural field trials. **The Plant Phenome Journal**, v. 3, e20005, May 2020. DOI: <https://doi.org/10.1002/ppj2.20005>.
- MATIAS, F. I.; GREEN, A.; LACHOWIEC, J. A.; LEBAUER, D.; FELDMAN, M. Bison-Fly: an open-source UAV pipeline for plant breeding data collection. **The Plant Phenome Journal**, v. 5, e20048, July 2022. DOI: <https://doi.org/10.1002/ppj2.20048>.
- MAZABEL, J.; WORTHINGTON, M.; CASTIBLANCO, V.; PETERS, M.; ARANGO, J. Using near infrared reflectance spectroscopy for estimating nutritional quality of *Brachiaria humidicola* in breeding selections. **Agrosystems, Geosciences & Environment**, v. 3, e20070, July 2020. DOI: <https://doi.org/10.1002/agg2.20070>.
- MOLANO, M. L.; CORTÉS, M. L.; ÁVILA, P.; MARTENS, S. D.; MUÑOZ, L. S. Ecuaciones de calibración en espectroscopía de reflectancia en el infrarrojo cercano (NIRS) para predicción de parámetros nutritivos en forrajes tropicales. **Tropical Grasslands-Forrajes Tropicales**, v. 4, n. 3, p. 139-145, 2016. DOI: [https://doi.org/10.17138/tgft\(4\)139-145](https://doi.org/10.17138/tgft(4)139-145).
- MOREIRA, F. F.; OLIVEIRA, H. R.; VOLENEC, J. J.; RAINEY, K. M.; BRITO, L. F. Integrating high-throughput phenotyping and statistical genomic methods to genetically improve longitudinal traits in crops. **Frontiers in Plant Science**, v. 11, 681, May 2020. DOI: <https://doi.org/10.3389/fpls.2020.00681>.
- NAGASUBRAMANIAN, K.; JONES, S.; SARKAR, S.; SINGH, A. K.; SING, A.; GANAPATHYSUBRAMANIAN, B. Hyperspectral band selection using genetic algorithm and support vector machines for early identification of charcoal rot disease in soybean stems. **Plant Methods**, v. 14, 86, Oct. 2018. DOI: <https://doi.org/10.1186/s13007-018-0349-9>.
- NAZIAN, M.; NIEDBAŁA, G. Machine learning for plant breeding and biotechnology. **Agriculture**, v. 10, 436, Sept. 2020. DOI: <https://doi.org/10.3390/agriculture10100436>.
- PARMLEY, K.; NAGASUBRAMANIAN, K.; SARKAR, S.; GANAPATHYSUBRAMANIAN, B.; SING, A. K. Development of optimized phenomic predictors for efficient plant breeding decisions using phenomic-assisted selection in soybean. **Plant Phenomics**, v. 2019, 5809404, July 2019. DOI: <https://spj.science.org/doi/10.34133/2019/5809404>.
- PITTMAN, J. J.; ARNALL, D. B.; INTERRANTE, S. M.; MOFFET, C. A.; BUTLER, T. J. Estimation of biomass and canopy height in bermudagrass, alfalfa, and wheat using ultrasonic, laser, and spectral sensors. **Sensors**, v. 15, n. 2, p. 2920–2943, Jan. 2015. DOI: <https://doi.org/10.3390/s150202920>.
- REDMON, J.; DIVVALA, S.; GIRSHICK, R.; FARHADI, A. You only look once: unified, real-time object detection. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2016, Las Vegas. **Proceedings**. Disponível em: <http://ieeexplore.ieee.org/document/7780460/>. Acesso em: 27 nov. 2022.
- ROITSCH, T.; CABRERA-BOSQUET, L.; FOURNIER, A.; GHAMKHAR, K.; JIMÉNEZ-BERNI, J.; PINTO, F.; OBER, E. S. Review: new sensors and data-driven approaches – a path to next generation phenomics. **Plant Science**, v. 282, p. 2–10, May 2019. DOI: <https://doi.org/10.1016/j.plantsci.2019.01.011>.
- RUZGIENÉ, B.; BERTESKA, T.; GECYTE, S.; JAKUBAUSKIENE, E.; AKAMITAUSKAS, V. C. The surface modelling based on UAV photogrammetry and qualitative estimation. **Measurement**, v. 73, p. 619–627, Sept. 2015. DOI: <https://doi.org/10.1016/j.measurement.2015.04.018>.
- SHAKOOR, N.; LEE, S.; MOCKLER, T. C. High throughput phenotyping to accelerate crop breeding and monitoring of diseases in the field. **Current Opinion in Plant Biology**, v. 38, p. 184–192, Aug. 2017. DOI: <https://doi.org/10.1016/j.pbi.2017.05.006>.

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

SINGH, A.; JONES, S.; GANAPATHYSUBRAMANIAN, B.; SARKAR, S.; MUELLER, D.; SANDHU, K.; NAGASUBRAMANIAN, K. Challenges and opportunities in machine-augmented plant stress phenotyping. **Trends in Plant Science**, v. 26, n. 1, p. 53–69, Jan. 2021. DOI: <https://doi.org/10.1016/j.tplants.2020.07.010>.

SINGH, A. K.; GANAPATHYSUBRAMANIAN, B.; SARKAR, S.; SINGH, A. Deep learning for plant stress phenotyping: trends and future perspectives. **Trends in Plant Science**, v. 23, n. 10, p. 883–898, Oct. 2018. DOI: <https://doi.org/10.1016/j.tplants.2018.07.004>.

TAO, H.; XU, S.; TIAN, Y.; LI, Z.; GE, Y.; ZHANG, J.; WANG, Y.; ZHOU, G.; DENG, X.; ZHANG, Z.; DING, Y.; JIANG, D.; GUO, Q.; JIN, S. Proximal and remote sensing in plant phenomics: 20 years of progress, challenges, and perspectives. **Plant Communications**, v. 3, 100344, Nov. 2022. DOI: <https://doi.org/10.1016/j.xplc.2022.100344>.

TODA, Y.; KAGA, A.; KAJIYA-KANEGAE, H.; HATTORI, T.; YAMAOKA, S.; OKAMOTO, M.; TSUJIMOTO, H.; IWATA, H. Genomic prediction modeling of soybean biomass using UAV-based remote sensing and longitudinal model parameters. **The Plant Genome**, v. 14, e20157, Nov. 2021. DOI: <https://doi.org/10.1002/tpg2.20157>.

TSENG, H.-H.; YANG, M.-D.; SAMINATHAN, R.; HSU, Y.-C.; YANG, C.-Y.; WU, D.-H. Rice seedling detection in UAV images using transfer learning and machine learning. **Remote Sensing**, v. 14, 2837, June 2022. DOI: <https://doi.org/10.3390/rs14122837>.

VOLPATO, L.; DOBBELS, A.; BORÉM, A.; LORENZ, A. J. Optimization of temporal UAS-based imagery analysis to estimate plant maturity date for soybean breeding. **The Plant Phenome Journal**, v. 4, e20018, May 2021a. DOI: <https://doi.org/10.1002/ppj2.20018>.

VOLPATO, L.; PINTO, F.; GONZÁLEZ-PÉREZ, L.; THOMPSON, I. G.; BORÉM, A.; REYNOLDS, M.; GÉRARD, B.; MOLERO, G.; RODRIGUES JUNIOR, F. A. High throughput field phenotyping for plant height using UAV-based RGB imagery in wheat breeding lines: feasibility and validation. **Frontiers in Plant Science**, v. 12, 591587, Feb. 2021b. DOI: <https://doi.org/10.3389/fpls.2021.591587>.

WALSH, K. B.; MCGLONE, V. A.; HAN, D. H. The uses of near infra-red spectroscopy in postharvest decision support: a review. **Postharvest Biology and Technology**, v. 163, 111139, May 2020. DOI: <https://doi.org/10.1016/j.postharvbio.2020.111139>.

WALTER, A.; LIEBISCH, F.; HUND, A. Plant phenotyping: from bean weighing to image analysis. **Plant Methods**, v. 11, 14, Mar. 2015. DOI: <https://doi.org/10.1186/s13007-015-0056-8>.

WALTER, J. D. C.; EDWARDS, J.; MCDONALD, G.; KUCHEL, H. Estimating biomass and canopy height with LiDAR for field crop breeding. **Frontiers in Plant Science**, v. 10, 1145, Sept. 2019. DOI: <https://doi.org/10.3389/fpls.2019.01145>.

WANG, L.; XIANG, L.; TANG, L.; JIANG, H. A convolutional neural network-based method for corn stand counting in the field. **Sensors**, v. 21, 507, Jan. 2021. DOI: <https://doi.org/10.3390/s21020507>.

WANG, X.; SINGH, D.; MARLA, S.; MORRIS, G.; POLAND, J. Field-based high-throughput phenotyping of plant height in sorghum using different sensing technologies. **Plant Methods**, v. 14, 53, July 2018. DOI: <https://doi.org/10.1186/s13007-018-0324-5>.

XAVIER, A.; HALL, B.; HEARST, A. A.; CHERKAUER, K. A.; RAINEY, K. M. Genetic architecture of phenomic-enabled canopy coverage in *Glycine max*. **Genetics**, v. 206, n. 2, p. 1081–1089, June 2017. DOI: <https://doi.org/10.1534/genetics.116.198713>.

XU, Y.; LI, P.; ZOU, C.; LU, Y.; XIE, C.; ZHANG, X.; PRASANNA, B. M.; OLSEN, M. S. Enhancing genetic gain in the era of molecular breeding. **Journal of Experimental Botany**, v. 68, n. 11, p. 2641–2666, May 2017. DOI: <https://doi.org/10.1093/jxb/erx135>.

YANG, G.; LIU, J.; ZHAO, C.; LI, Z.; HUANG, Y.; YU, H.; XU, B.; YANG, X.; ZHU, D.; ZHANG, X.; ZHANG, R.; FENG, H.; ZHAO, X.; LI, Z.; LI, H.; YANG, H. Unmanned aerial vehicle remote sensing for field-based

crop phenotyping: current status and perspectives. **Frontiers in Plant Science**, v. 8, 1111, June 2017. DOI: <https://doi.org/10.3389/fpls.2017.01111>.

YANG, M.-D.; TSENG, H.-H.; HSU, Y.-C.; YANG, C.-Y.; LAI, M.-H.; WU, D.-H. A UAV open dataset of rice paddies for deep learning practice. **Remote Sensing**, v. 13, 1358, Apr. 2021. DOI: <https://doi.org/10.3390/rs13071358>.

YANG, W.; FENG, H.; ZHANG, X.; ZHANG, J.; DOONAN, J. H.; BATCHELOR, W. D.; XIONG, L.; YAN, J. Crop phenomics and high-throughput phenotyping: past decades, current challenges, and future perspectives. **Molecular Plant**, v. 13, n. 2, p. 187–214, Feb. 2020. DOI: <https://doi.org/10.1016/j.molp.2020.01.008>.

ZHAO, C.; ZHANG, Y.; DU, J.; GUO, X.; WEN, W.; GU, S.; WANG, J.; FAN, J. Crop phenomics: current status and perspectives. **Frontiers in Plant Science**, v. 10, 714, June 2019. DOI: <https://doi.org/10.3389/fpls.2019.00714>.

ZHOU, J.; CHEN, H.; ZHOU, J.; FU, X.; YE, H.; NGUYEN, H. T. Development of an automated phenotyping platform for quantifying soybean dynamic responses to salinity stress in greenhouse environment. **Computers and Electronics in Agriculture**, v. 151, p. 319–330, Aug. 2018. DOI: <https://doi.org/10.1016/j.compag.2018.06.016>.

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.

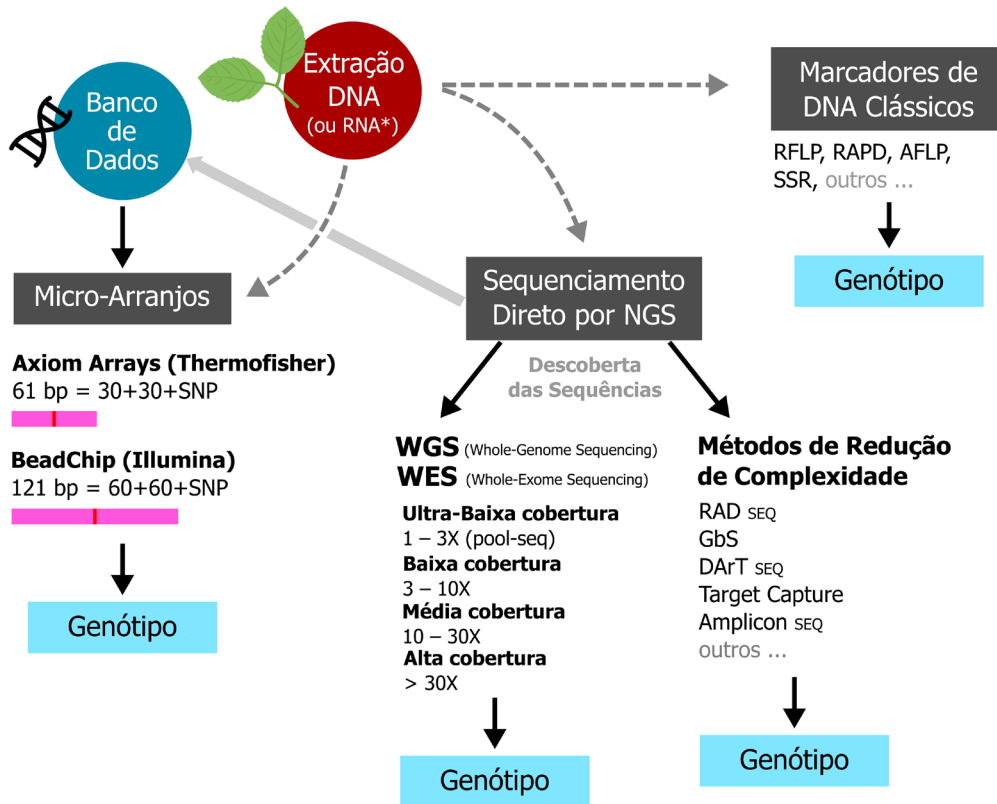


Figura 1. Organograma simplificado dos processos de genotipagem, considerando as novas tecnologias e marcadores moleculares clássicos. Em todas elas, “Genótipo” representa a disponibilização de marcadores genéticos para as amostras (Com colaboração do Dr. Orzenil Silva-Júnior).

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.

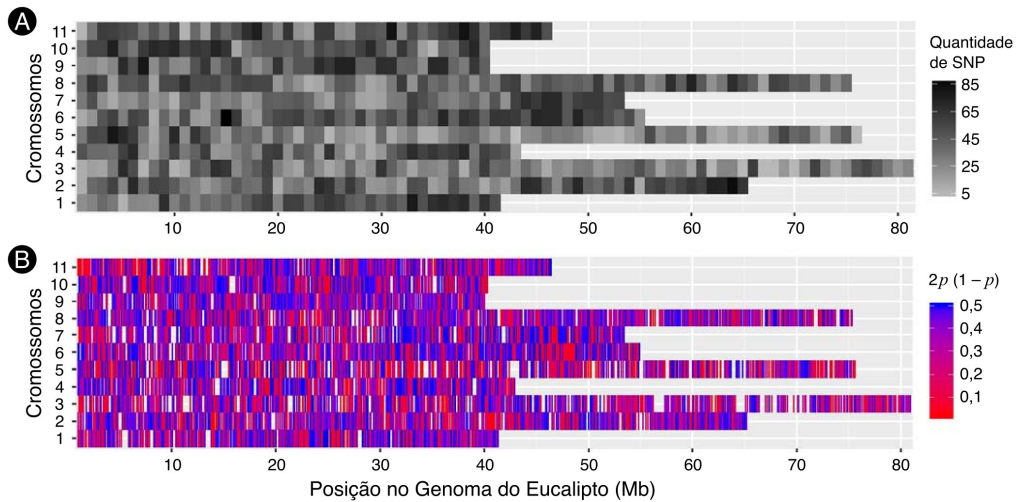


Figura 2. Distribuição de 24.806 marcados SNP polimórficos ao longo dos onze cromossomos de uma população de eucalipto. A parte “A” da figura mostra a concentração de SNPs por janela de 1 Mb. A parte “B” da figura mostra a heterozigose média dos SNPs em uma janela de 100 kb.

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

■ Melhoria 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.

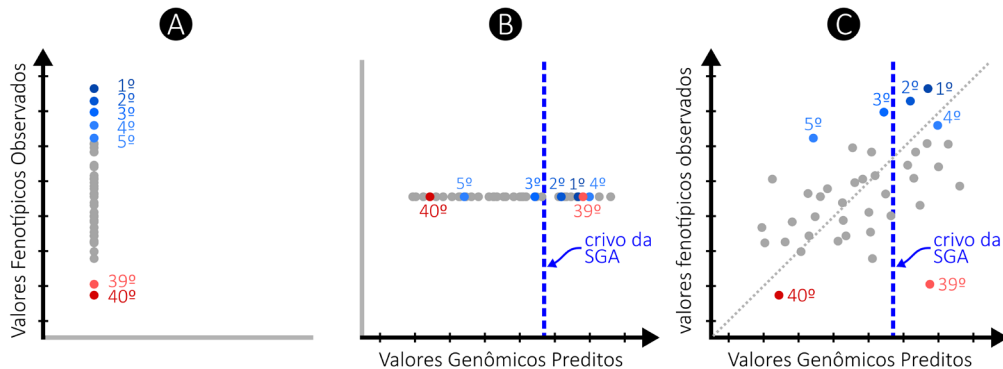


Figura 3. Relacionamento entre valores fenotípicos (médias ou BLUP de quarenta materiais genéticos) observados nos experimentos versus os valores preditos genomicamente. Na parte “A” é mostrado apenas o ranqueamento dito “verdadeiro”, a partir dos valores mensurados em campo. Na parte “B” são mostrados 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...

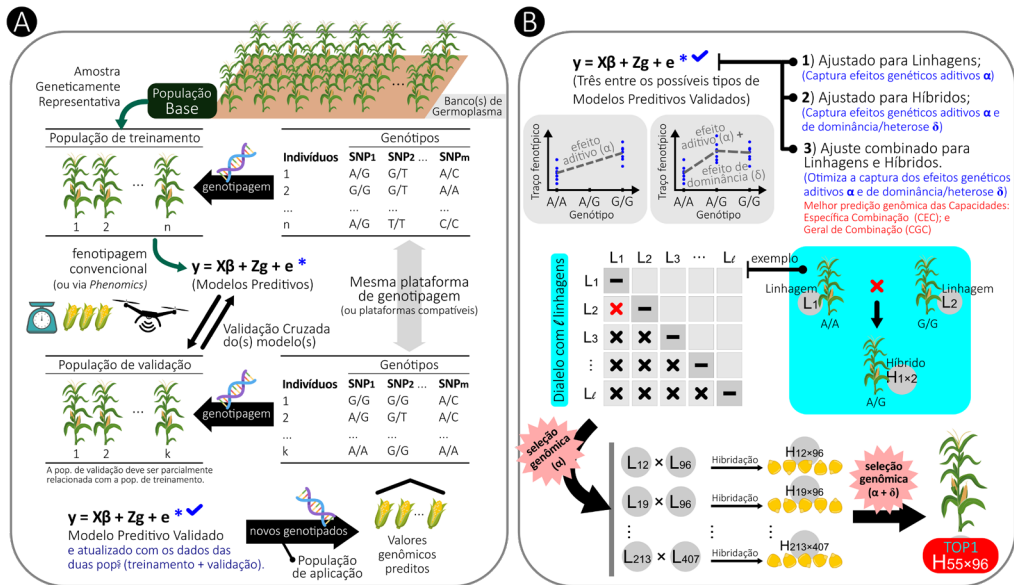


Figura 4. Esquema reduzido de um processo de Seleção Genômica Ampla (SGA), utilizando o milho como exemplo. Na parte A da figura é mostrado o processo de ajuste e validação de modelos genômicos preditivos. Na parte B da figura é mostrado possíveis esquemas de utilização, tanto para predição de linhagens 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 predizer 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>.

Roberto Tarazi, Magno Antonio Patto Ramalho, Flavia Maria Avelar Gonçalves

Introdução

A produção de grãos, frutos e fibras e outros produtos da agricultura para atender demanda da população mundial, que deverá superar os 10 bilhões de habitantes em 2050, é um dos maiores desafios da humanidade nas próximas décadas. O Brasil, com quase 216 milhões de habitantes, ocupa a 7ª posição em termo dos países mais populosos do planeta (United Nations, 2022). O Brasil é um dos únicos países do mundo a ter um vasto território com todas as condições necessárias para produção agrícola, principalmente em relação à água e terra agricultável, sem a necessidade de abrir novas fronteiras agrícolas, especialmente na região amazônica. A produção agrícola no Brasil, além de atender a demanda interna, deve ter também excedentes para a exportação, visando ter superávit crescente na balança de pagamento. O Brasil já alimenta 800 milhões de pessoas no planeta. Espera-se que nas próximas décadas o Brasil produza alimentos para atender dois bilhões de pessoas, ou seja, número quase dez vezes superior à sua população (Contini; Aragão, 2021).

A denominada Revolução Verde, que ocorreu na década de 60 do século passado, quando a humanidade apresentou a maior taxa de crescimento populacional, foi assentada em três pilares: o emprego mais acentuado de fertilizantes; a disponibilidade de inúmeros defensivos agrícolas; e o melhoramento genético das plantas, disponibilizando novas cultivares. Não existe esperança de que os dois primeiros pilares possam continuar contribuindo para a maior produtividade da agricultura, como ocorreu antes, principalmente devido às questões ambientais e ao incremento no custo de produção. Assim, a expectativa é que a obtenção de novas cultivares seja a protagonista responsável pelo aumento da produção dos produtos agrícolas (Ramalho et al., 2021).

Contudo, além da pressão do crescimento populacional, há previsões que os estresses bióticos e abióticos sejam crescentes. Em um cenário como esse, o melhoramento genético das plantas deve utilizar todas as ferramentas possíveis, algumas são comentadas nos demais capítulos deste livro, para ampliar a eficiência dos processos seletivos, ou seja, aumentar o ganho genético com a seleção, por unidade de tempo. Uma das possibilidades, entre as já comentadas, seria o emprego das estratégias visando a aceleração na obtenção de novas cultivares. Na Austrália, em 2003, na University of Queensland, surgiu o termo Speed Breeding (SB) para denominar o conjunto de processos que visavam acelerar os programas de melhoramento de trigo (Watson et al., 2018). Esse termo passou, então, a ser amplamente utilizado. O foco deste capítulo é discutir alternativas de SB, como ênfase na interação dos princípios biológicos e a manipulação de fatores ambientais; as experiências brasileiras no emprego de SB ao longo do tempo; e as perspectivas do processo nas próximas décadas.

Alternativas para acelerar os ganhos genéticos

Todos os caracteres das plantas de importância agrônômica e econômica, tais como aqueles envolvidos na produção de grãos, frutos e fibras, são controlados por inúmeros genes. A genética mendeliana mostra que a probabilidade de se acumular todos os alelos favoráveis dos diferentes genes, para uma dada condição, em um ou poucos ciclos seletivos, é infinitamente pequena. Desse modo, o melhoramento genético de plantas é um acúmulo de alelos favoráveis ao longo do tempo, realizado consciente ou inconscientemente, por meio da seleção recorrente (SR). Ou seja, ciclo de seleção seguido da recombinação/intercruzamento das melhores progênies/linhagens/indivíduos (Rutkoski, 2019; Bernardo, 2020). Esse processo, sem dúvida nenhuma, começou com a domesticação das plantas e continua até hoje.

Como já comentado, a sociedade humana necessita que cada ciclo seletivo seja o mais rápido e eficiente possível, pois, só assim, toda a demanda dos produtos da agricultura será atendida. Para visualizar como a eficiência da SR pode ser ampliada, é válido comentar a respeito da “equação do melhorista” e seus componentes (Cobb et al., 2019; Sinha et al., 2021). Ela possibilita estimar o ganho esperado com a seleção (GS) para qualquer caráter submetido a SR, obtido pelo estimador:

$$GS = \frac{\sigma_A^2}{t} ic \frac{\sigma_F}{\sigma_F}$$

Nessa equação, t representa o número de anos necessário para completar um ciclo de SR. A eficiência, portanto, é avaliada pelo maior GS por unidade de tempo, que é o foco deste capítulo – *Speed Breeding* (SB).

Desse modo, existem várias alternativas para se estimar a eficiência do GS, as quais podem ser inferidas a partir da “equação do melhorista”. Resumidamente, tem-se i , por exemplo, que representa a intensidade de seleção padronizada. O valor é tabelado em função da proporção de indivíduos/progênes selecionada. Quanto menor a proporção, maior é o ganho esperado. Todavia, a SR exige que a variabilidade genética seja mantida para se ter a continuidade do sucesso a longo prazo. Em princípio, uma intensidade de seleção alta – pequena proporção de indivíduos selecionados - pode ser obtida avaliando maior número de genótipos diferentes a cada ciclo seletivo. Contudo, esse número tem um limite, que depende do seu custo-benefício. Isto é, o GS dividido pelo custo de cada progênie/indivíduo avaliado. Infelizmente, essa relação não é linear (Rutkoski, 2019).

Outro componente da equação é o c , o qual se refere à proporção da variância aditiva (σ_a^2) que é explorada com a seleção, quando ela envolve a reprodução sexuada. Por exemplo, pode-se avaliar progênes não endogâmicas e endogâmicas. Na primeira, estão as progênes de meios-irmãos (HS) ou irmãos completos (FS), as quais exploram menor proporção de σ_a^2 (menor c) que as endogâmicas, aquelas submetidas a uma ou mais gerações de autofecundação (S_1, S_2, S_g). O c depende, também, do tipo de progênes que serão inter cruzadas para se obter a população melhorada e possibilitar a continuação do processo. Suponha-se que fossem avaliadas progênes de HS, sendo a recombinação também com sementes remanescentes de HS, o c será igual a 1/4, indicando que aquele ciclo SR explorou 1/4 de σ_a^2 . Por outro lado, se na recombinação forem utilizadas progênes S_1 da mesma planta em que se obteve o HS ou clone dela, o c passa a ser 1/2. Maiores detalhes a esse respeito podem ser obtidos em Souza Junior (2001), Hallauer et al. (2010) e Resende (2015). Assim, pode-se visualizar com facilidade que o ganho por unidade de tempo (SB) irá depender do tipo de progênes avaliadas e recombinadas.

Ainda nesse contexto, é preciso comentar o que ocorre quando se trata de plantas tipicamente autógamas, em que predomina a autofecundação natural. Nesse caso, o avanço das gerações é predominantemente realizado por meio de progênes endogâmicas. Quando isso ocorre, dois aspectos são fundamentais, o primeiro é o método de condução da população segregante. Nesse contexto, na literatura existem algumas opções. Uma delas, é o método de descendentes de uma única semente (SSD), o qual consiste basicamente no avanço das gerações endogâmicas a partir

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

de uma única semente de cada planta da geração F_2 . A partir de então, o processo é repetido por gerações sucessivas. Especialmente em regiões de inverno rigoroso, em que para se obter mais de uma geração por ano é necessário o emprego de ambientes controlados - casa de vegetação -, o SSD é o escolhido porque ocupa menos espaço que os outros métodos. Em função desse fato, o SSD é considerado como uma técnica de SB em muitas publicações. Todavia, sob condições tropicais, em que é possível mais de um cultivo por ano no campo, evidentemente o SSD não tem a vantagem preconizada. Deve ser salientado que, nesse método, a velocidade de se conseguir a homozigose é igual ao dos demais. Além do mais, esse procedimento acarreta grande perdas de genótipos, devido à amostragem ineficiente. Destaca-se, também, que o fato de se tomar uma amostra a partir de uma planta da geração F_2 ou de uma outra geração não indica necessariamente que essa amostra represente a variabilidade daquela geração amostrada. Considerando o que foi comentado anteriormente, com SSD é impossível prever com precisão o valor do c - valor da proporção da σ_A^2 - que é utilizada na seleção em determinada geração, como ocorre com os demais métodos.

No estimador do GS, a σ_A^2 é, como já foi mencionado, a variância aditiva, sendo parte da variabilidade genética da população. Quanto maior σ_A^2 maior é o ganho por unidade de tempo. Assim, na SR deve-se utilizar de estratégias para manter um valor expressivo de σ_A^2 ao longo do tempo. Porém, esse assunto não será o foco deste capítulo. Finalmente, o desvio fenotípico (σ_F) da unidade de seleção, o qual pode ser entre indivíduos ou média de progênies. Em síntese, o GS é função da habilidade ou ferramentas que os melhoristas possuem para identificar, pelo fenótipo, os melhores genótipos – maior número de alelos favoráveis possíveis entre os indivíduos/progênies avaliados. Isso depende de muitos fatores, dentre eles, a precisão experimental. Nesse contexto, algumas alternativas têm sido empregadas, uma delas, a seleção genômica, discutida em detalhes no capítulo 10. Seu objetivo principal é o uso da informação existente nos ácidos nucleicos, a fim de identificar os melhores genótipos/fenótipos. Outra opção, a qual tem sido muito difundida, é a fenotipagem de alto rendimento, apresentada no capítulo 9 desta publicação. Deve ser mencionado, também, o envolvimento da biometria, por meio de inúmeras técnicas, visando a aproveitar melhor a fenotipagem, a partir de outras informações relacionadas, como fatores ambientais e ambientômica, abordados no capítulo 8 e em outras partes desta publicação.

Adicionalmente, deve-se enfatizar que, para reduzir o t da equação, existem alternativas para acelerar a SR. Todas elas são focadas em obter, o mais rapidamente

possível, as progênes/linhagens a serem avaliadas e/ou intercruzadas. Algumas dessas estratégias são o foco deste capítulo.

O processo de Speed Breeding

Para acelerar o ciclo de vida das plantas é importante saber que o crescimento e desenvolvimento dependem de sinais ambientais (abióticos ou exógenos), principalmente presença de água, nutrição, luz, fotoperíodo e temperaturas. Esses, por sua vez, induzem a síntese de sinais endógenos, os hormônios vegetais ou mensageiros primários. Os hormônios vegetais atuam no controle do ciclo celular, da expressão gênica e da ativação enzimática. A “ativação” ou “repressão” de genes causam alteração na produção de proteínas, sendo que a maioria dará origem as enzimas. Quando se pretende trabalhar com a aceleração de ciclo nas plantas, tenta-se acelerar ao máximo a mitose e a diferenciação celular. Assim, tem-se procurado atuar nos processos metabólicos nas diferentes células que constituem os tecidos dos órgãos das plantas, com foco em acelerar a obtenção da geração seguinte (Figura 1).

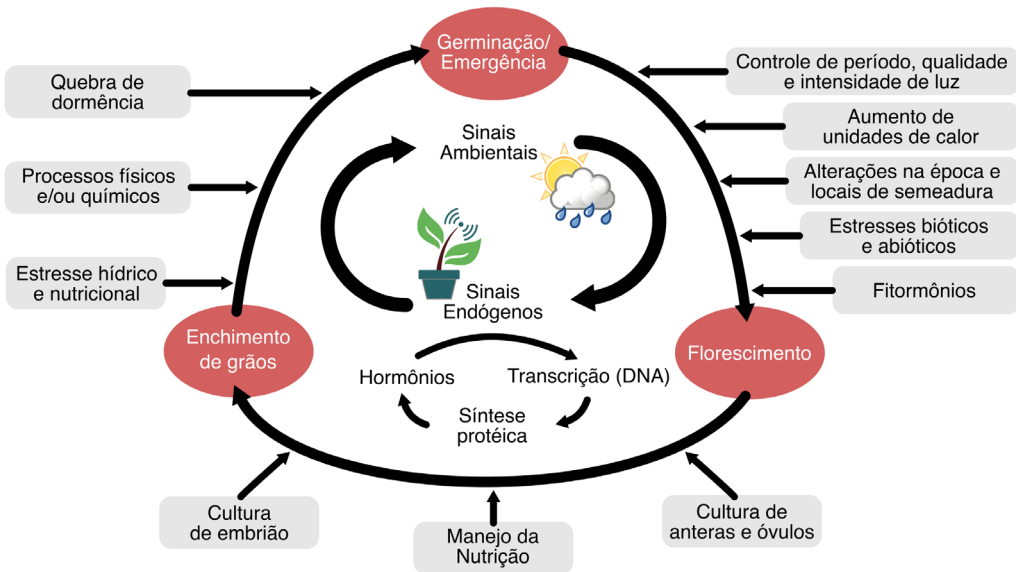


Figura 1. Esquema dos três principais momentos do ciclo de vida de uma planta relacionados à aceleração de ciclo. As técnicas de Speed Breeding intensificam os sinais ambientais que induzem a síntese de sinais endógenos acelerando o ciclo de vida das plantas.

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

A aceleração do ciclo das plantas, especialmente as anuais, pode, em princípio, ser obtida em três momentos. O primeiro é com relação a velocidade de germinação/emergência. A eficácia dos processos de SB com foco nessa etapa, para grande número das plantas cultivadas, deve ser pequena. Uma das razões é que a maioria das espécies, durante a domesticação, já foram selecionadas visando à redução na dormência das sementes (Soltani et al., 2021). Se não ocorrer dormência, a variabilidade para velocidade de germinação/emergência deve ser muito pequena. Todavia, em algumas plantas perenes cultivadas existem problemas de dormência das sementes, que atrasam o processo de obtenção de novos clones.

Sementes dormentes podem ser tratadas com estratificação a frio, imersão das sementes em água ou hormônios promotores da germinação, como giberelinas. Combinar a cultura de embriões com a quebra da dormência das sementes economizará mais tempo no ciclo de geração (Pandey et al., 2022). No caso da oliveira, os clones são obtidos a partir de populações segregantes oriundas de plantas seminais. A germinação das sementes, normalmente utilizando sementeiras com solo, demora alguns meses, devido à dormência das sementes, atrasando o processo de seleção. Montilon et al. (2022) descrevem um protocolo para acelerar a germinação de sementes híbridas de oliveira que utiliza cultura de embriões para superar os inconvenientes relacionados à longa dormência das sementes e, adicionalmente, contribui para maior uniformidade e percentagem de germinação que no processo tradicional.

O segundo momento é focado na redução do tempo da emergência ao pleno florescimento das plantas (Figura 1). Essa etapa tem recebido grande atenção dos pesquisadores de SB. Em várias espécies anuais cultivadas existem informações do controle genético do número de dias para o início do florescimento, como em milho (Buckler et al. 2009) e soja (Lin et al., 2021), por exemplo. Em todos esses trabalhos, têm sido identificados inúmeros genes envolvidos nesse processo. Contudo, existem muitas evidências do efeito do ambiente na expressão do caráter número de dias para o florescimento. Nesse contexto é que se concentrou as principais ações do Speed Breeding que serão comentadas neste tópico.

Para reduzir o número de dias para o florescimento, há a necessidade de compreender que, em muitas plantas, mudança no fotoperíodo pode desencadear uma transição da fase vegetativa para o florescimento (fase reprodutiva). As plantas são classificadas em três grupos de acordo com a duração crítica da noite que desencadeia a floração (Shafiq et al., 2021):

- 1) Plantas de dia curto (PDCs) requerem duração noturna mais longa que a crítica;
- 2) Plantas de dia longo (PDLs) requerem duração noturna mais curta que a crítica;

3) Plantas neutras ao dia (PNDs) não são reguladas pela duração da noite.

Utilizando o conceito proposto pelos pesquisadores da University of Queensland, o SB acelera a floração, explorando o fotoperíodo das plantas, a qualidade e intensidade de luz, além de fatores como temperatura e umidade (Hickey et al., 2017).

Independentemente se a aceleração de ciclo vai ser realizada no campo ou em ambientes artificiais, a luz é um fator crítico ao crescimento das plantas, em que a intensidade e a qualidade são importantes. A intensidade da luz é medida pela densidade de fluxo de fótons fotossintéticos, mais conhecida como PPFD, ou seja, os fótons que realmente atingem a planta. A título de curiosidade, o sol do meio-dia nos trópicos tem um valor de aproximadamente 110.000 lux, que convertido para PPFD é 1342 $\mu\text{mol}/\text{m}^2/\text{s}$. Com essa intensidade de luz, principalmente nas regiões Centro-Oeste e Nordeste do Brasil, a aceleração de ciclo pode ocorrer o ano inteiro no campo, diferentemente com o que ocorre em países de clima temperado, onde há necessidade de investir em ambientes confinados e com luz artificial. É importante ressaltar que em ambientes confinados, dependendo do tipo de cobertura, a intensidade de luz cai mais do que pela metade em relação à intensidade de luz de locais descobertos, por isso necessita-se investir em sistemas de iluminação artificial para ambientes confinados, mesmo nos trópicos (Figura 2).

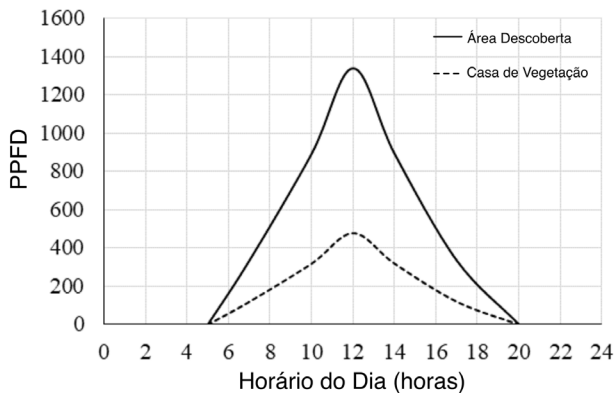


Figura 2. A linha contínua representa os valores de densidade de fluxo de fótons fotossintéticos (PPFD) em área descoberta e a linha tracejada os valores de PPFD dentro de uma casa de vegetação com teto de policarbonato translúcido. Ambas medições foram realizadas dia 7 de janeiro de 2022, na UFRJ, Rio de Janeiro, RJ.

Enquanto a intensidade de luz é medida pela PPFD, a qualidade da luz é medida por meio da radiação fotossinteticamente ativa, ou PAR, que é a faixa de comprimento de onda entre 400 nm e 700 nm. As plantas usam essa faixa de radiação eletromagnética para realizar a fotossíntese. Estudos relatam que a luz azul (425–490 nm) e a luz vermelha (610–700 nm) têm os melhores comprimentos do espectro de luz para a fotossíntese das plantas (Figura 3) (Wanga et al., 2021). Já a proporção de luz vermelha (R) e vermelha distante (FR) é crítica e responsável pela floração da planta.

Essa proporção R/FR varia com a estação do ano e a latitude, sendo assim, é possível utilizar diferentes genótipos em épocas de semeadura e latitudes diferentes para acelerar a floração. No caso de ambientes controlados, como casas de vegetação, é possível ajustar essa proporção para induzir a floração (Li et al., 2018).

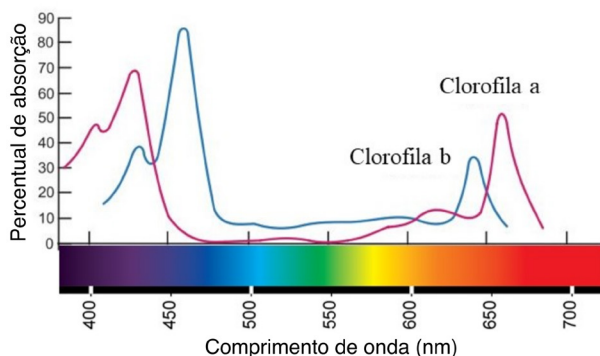


Figura 3. Radiação fotossinteticamente ativa (PAR) com a faixa de comprimento de onda entre 400 nm e 700 nm e percentual de absorção pelas clorofilas a e b de acordo com o comprimento de onda.

A nova tecnologia de iluminação LED permite a entrega precisa de comprimentos de onda de luz para manipular o crescimento e desenvolvimento das plantas. Isso foi aplicado com sucesso para promover o florescimento e ao mesmo tempo suprimir o crescimento da planta para culturas em que se utiliza o SB, como tremoço e soja (Wanga et al., 2021). Desse modo, aproveitar os comprimentos de onda para controlar o crescimento da planta é um campo promissor, ainda a ser mais explorado ou otimizado para SB em muitas espécies. Teoricamente, adaptando a composição do comprimento de onda ao longo do ciclo de vida da planta é possível gerenciar a atividade dos fitormônios visando aumentar a eficiência do uso de SB (Watson et al., 2018). Por exemplo, no arroz, o controle dinâmico dos sistemas de iluminação pode permitir comprimentos de onda enriquecidos com vermelho distante (FR) para promover o florescimento precoce, seguido por comprimentos de onda vermelho-azul mais baixos, que podem suprimir o alongamento do caule e reduzir a altura da planta.

O desenvolvimento fenológico também é impulsionado pela temperatura. No entanto, a relação entre fotoperíodo e temperatura não é linear. A taxa de crescimento geralmente aumenta com temperaturas mais altas e, na maioria das espécies, o desenvolvimento vegetativo ocorre em uma temperatura ótima mais alta do que o desenvolvimento reprodutivo. A ocorrência do florescimento em muitas culturas é sensível a altas temperaturas, afetando particularmente a meiose e a viabilidade do pólen. Para otimizar os protocolos de SB, podem ser usadas temperaturas mais altas, com o intuito de acelerar o crescimento vegetativo, e temperaturas mais baixas, a partir do florescimento (Ghosh et al., 2018; Li et al., 2018). Contudo, o que realmente

conta para o número de dias para o início do florescimento são as “unidades de calor”, que variam de acordo com a espécie envolvida.

A aplicação de tratamentos de estresse é outra abordagem utilizada em SB para acelerar o ciclo. A maioria das espécies cultivadas submetidas a estresse aceleram o início do florescimento. Esses protocolos de estresse podem ser estabelecidos com alta ou baixa temperatura, deficiência de nutrientes, excesso ou falta de água, ou regulando a intensidade luminosa. Adicionalmente, densidade populacional, tratamentos com UV e infecção por patógenos podem também acelerar o florescimento (Bhatta et al., 2021). Um exemplo de uso de estresse abiótico foi apresentado por Zheng et al. (2013), no qual o trigo foi cultivado em recipientes com 80 cm³ e as plantas foram submetidas a estresse hídrico. Com esse protocolo, foi possível obter oito gerações por ano, todavia o número de sementes produzido por planta foi pequeno, duas a três.

Para alguns cereais, a semeadura pode ser realizada tanto na primavera como no inverno, em muitas regiões do mundo. Diferentemente das de primavera, as cultivares de inverno precisam ser tratadas em baixa temperatura por longos períodos para acelerar o florescimento. Cha et al. (2022) aplicaram o conhecimento de SB existente para as de primavera nas de inverno. O sucesso foi grande quando se associou o protocolo aplicado às linhagens de primavera com a vernalização de mudas de embriões jovens a 10°C, conseguindo, assim, até sete gerações em cevada, seis em trigo, cinco em aveia e triticale por ano.

A terceira possibilidade de acelerar os programas de melhoramento é no período denominado de enchimento de grão. Existem duas fases importantes após a obtenção da célula ovo ou zigoto. A primeira fase é o período de intensa divisão celular para se obter as células que irão constituir as sementes, já na segunda, ocorre o acúmulo de carboidratos nessas células, que irão propiciar as reservas necessárias para germinação e emergência das plantas. Esse acúmulo de matéria seca nas células irá cessar quando a maturação fisiológica estiver completa (Lee; Tollenaar, 2007). As tentativas de SB são realizadas para “pular” essa etapa, por meio de cultivo de embriões sobre condições controladas. Essa é uma alternativa de SB que será comentada a seguir.

A cultura de embriões é uma das técnicas para a obtenção de plântulas o mais precoce possível, utilizando células embrionárias, ou seja, aquelas obtidas a partir das células derivadas do ovo ou zigoto. O tecido é isolado e colocado para o cultivo direto no meio de cultura em condições estéreis, para obter plantas viáveis. O sucesso do processo depende das condições que se tem no meio de cultura, de modo a promover o crescimento e desenvolvimento até a obtenção da plântula, que será posteriormente transplantada para um meio apropriado, visando completar o ciclo da planta. Essas

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

técnicas têm sido amplamente utilizadas com algumas finalidades além da SB, como por exemplo, na recuperação de híbridos interespecíficos. A dificuldade do processo está associada ao manejo das células embrionárias em laboratórios e a identificação do meio de cultura apropriado para o crescimento e desenvolvimento, o que provavelmente deve variar com a espécie. É importante salientar que essa estratégia de SB, pelo menos em princípio, é a que mais pode acelerar os programas de melhoramento, uma vez que o período entre a fertilização e a final da maturação fisiológica é a fase mais demorada da vida de um vegetal. Ela tem sido utilizada com sucesso para acelerar o ciclo do milho, trigo, aveia, soja, entre outras culturas (McCaw et al., 2021).

Alguns trabalhos de SB estão relacionados na Tabela 1. Nessa tabela foi incluída as principais estratégias de SB utilizadas e o número de gerações por ano que foi possível conduzir.

Tabela 1. Algumas espécies em que se aplicou o Speed Breeding para acelerar o ciclo das plantas.

Espécie	Estratégia de SB	Gerações/ano	Fonte
Aveia (<i>Avena sativa</i>) ¹	Cultura de embrião	7	Liu et al. (2016)
Cevada (<i>Hordeum vulgare</i>) ¹	Desenvolvimento das plantas por meio do controle de luz em casa de vegetação	6	Hickey et al. (2017)
Trigo (<i>Triticum aestivum</i>) ¹	Desenvolvimento das plantas por meio do controle de luz em casa de vegetação	6	Ghosh et al. (2018)
Trevo (<i>Trifolium subterraneum</i>) ¹	Cultura de embrião seguido de crescimento em casa de vegetação, regulando temperatura e estendendo o fotoperíodo	6	Pazos-Navarro et al. (2017)
Ervilha ¹ (<i>Pisum sativum</i>)	Cultura de embrião	6,8	Ribalta et al. (2017)
Arroz (<i>Oryza sativa</i>) ²	Luz vermelha para indução de florescimento	5	Jähne et al. (2020)
Sorgo (<i>Sorghum bicolor</i>) ²	Cultura de embrião seguido de crescimento em casa de vegetação, regulando temperatura e estendendo o fotoperíodo	6	Forster et al. (2014)
Soja (<i>Glycine max</i>) ²	Luz vermelha para indução de florescimento	5	Jähne et al. (2020)
Amendoim (<i>Arachis hypogaea</i>) ²	Crescimento em ambiente controlado, regulando temperatura e estendendo o fotoperíodo	4	Rowell et al. (1999)
Amaranto (<i>Amaranthus</i> spp.) ²	Luz vermelha para indução de florescimento	4	Jähne et al. (2020)

¹Espécies de dias longos; ²Espécies de dias curtos

O emprego de duplo-haploide para acelerar programas de melhoramento

Em plantas autógamas e em plantas alógamas que utilizam sementes híbridas o que se almeja, como já comentado em outras técnicas de SB anteriormente apresentadas, é obter linhagens o mais rápido possível. Com essa finalidade, o emprego de duplo haploides (DH) possibilita obter rapidamente indivíduos 100% homocigotos, o que não é possível com nenhuma outra técnica. As etapas de obtenção do DH envolvem a indução e identificação dos indivíduos haploides, duplicação cromossômica dos selecionados e multiplicação das sementes DH. A etapa de indução de haploidia pode ser realizada de duas formas, *in vitro* ou *in vivo* (Niazian; Shariatpanahi, 2020). O sistema de indução *in vitro* utiliza técnicas da cultura de tecidos com os gametas masculinos- cultura de anteras - ou feminino - óvulos (Figura 1). A indução de haploides *in vivo* pode ser realizada por meio da polinização com linhagens indutoras, como tem sido muito utilizada na cultura do milho. Nesse caso, pode-se obter haploides ginogenéticos, quando a linhagem indutora é a doadora do grão de pólen, ou androgenéticos, a indutora é o genitor feminino. O primeiro indutor identificado foi a linhagem Stock 6, na década de 1950. Todavia, o gene indutor tem uma penetrância incompleta, pois a taxa de indução não é de 100%. Inicialmente, ela era de apenas de 2,5%. Com a seleção, tem sido possível aumentar a taxa de indução para valores em torno de 15% (Andorf et al., 2019). Outra forma de se obter indivíduos haploides *in vivo* é o cruzamento interespecífico entre espécies relacionadas, como ocorre em cevada, batata, trigo e tabaco.

Como mencionado, após a indução da haploidia é necessária a identificação e seleção dos indivíduos haploides. Essa é uma das etapas mais importante e constitui um gargalo no processo de obtenção de DH em algumas espécies. Ela pode ser realizada, principalmente, por marcadores fenotípicos, por contagem cromossômica, pela citometria de fluxo ou pelo uso de marcadores moleculares. Maiores comentários a respeito desses procedimentos podem ser obtidos em Lemos (2021). Após a identificação dos haploides, é necessário a duplicação cromossômica. O processo envolve a aplicação de agentes antimitóticos, que atuam durante a formação das fibras do fuso na mitose. Com essa finalidade, a colchicina foi muito utilizada no passado, entretanto, alternativas com tratamentos químicos têm sido usadas, entre elas citam-se a trifluralina, amiprofos-metilo e gás óxido e, também, alguns físicos, como choque de temperatura. Em algumas culturas, pode ocorrer também a duplicação espontânea do haploide, como tem sido observado no milho, embora com baixa porcentagem (Boerman et al., 2020).

Experiência no Brasil Visando a Aceleração dos Programas de Melhoramento

Embora o melhoramento genético das plantas no Brasil tenha iniciado há muito tempo, os maiores avanços ocorreram a partir dos anos 1970 do século passado. Sobretudo, com a criação dos programas de pós-graduação, da Empresa Brasileira de Pesquisa Agropecuária (Embrapa), de empresas estaduais de pesquisa e das empresas produtoras de sementes e outros produtos agrícolas. O progresso genético obtido nesse período foi enorme (Ramalho et al., 2021). Grande parte desse progresso foi obtido por meio de estratégias que aceleraram os programas de melhoramento de plantas anuais, bem como perenes.

No caso da cultura do milho, até próximo de 1980, as cultivares disponíveis, seja de híbrido duplos ou de variedades, eram todas excessivamente altas e de ciclo longo. As primeiras tentativas de alterar o porte das plantas ocorreram, sem muito êxito, por meio da utilização de alelos de genes braquíticos visando a redução, sobretudo, do comprimento dos entrenós das plantas. O sucesso ocorreu posteriormente por meio da utilização de linhagens de clima temperado nos programas de melhoramento das empresas privadas em condições tropicais, o que era considerado inviável na época. Evidentemente, tiveram enorme êxito na redução na altura e, conseqüentemente, no ciclo das plantas. Outra estratégia que contribuiu enormemente para o SB do milho foi a migração das estações de pesquisa, que se concentravam mais para o sul do País, com a possibilidade da condução de apenas uma geração por ano, para outras regiões em que a oscilação de temperatura, durante o ano, não é grande, permitindo duas ou até três gerações de melhoramento por ano. Deve-se salientar que essa migração para as regiões Centro-Oeste e Norte coincidiu também com a extensão das épocas de semeadura do milho, principalmente em sucessão da cultura da soja. O sucesso foi enorme, tanto é que a produção do milho, que antes se concentrava na safra de verão (semeadura de setembro a novembro no Sul e Sudeste) passou a ser predominante obtida na segunda safra, em sucessão ao cultivo da soja (Conab, 2023).

Em outras plantas anuais, como o feijão, o programa de melhoramento foi acelerado por meio de até três safras por ano, em algumas delas usando irrigação. No caso da soja, como ocorreu com o milho, muitas empresas migraram as suas estações experimentais para regiões principalmente do Centro-Oeste. Fato que contribuiu e muito para a redução do ciclo da soja, o que tornou mais viável o cultivo em sucessão do milho após essa leguminosa. Deve ser enfatizado que, anteriormente, praticamente todas as cultivares de soja precoce eram de crescimento determinado, com menor produtividade. A obtenção de plantas de crescimento indeterminado e precoces, com maior potencial produtivo, tem sido a principal razão do maior sucesso.

No caso das plantas perenes, o exemplo mais notável tem sido com a cultura do eucalipto. O sucesso do melhoramento dessa espécie possibilitou ao Brasil ser um dos maiores produtores de celulose, carvão vegetal e outros produtos derivados da madeira, no mundo. Para isso, algumas estratégias que se enquadram na categoria de Speed Breeding tiveram que ser adotadas antes mesmo da criação desse termo. De início, é importante salientar que os plantios de eucalipto, até próximo do final dos anos 60 do século XX, eram com mudas seminais. Após o domínio da técnica de clonagem, o uso de clones em escala comercial tornou-se viável.

O emprego de clones tem a vantagem de possibilitar utilizar na seleção todos os tipos de interação alélica, dominância e epistasia, e não só a aditiva. Dizendo de outro modo, possibilita a perpetuação de boas combinações híbridas, híbrido simples, sem a necessidade de obter linhagens (Rezende et al., 2014). Como mencionado, no eucalipto, os plantios anteriores eram de mudas seminais oriundas de sementes derivadas de populações obtidas de poucos ciclos de seleção massal ou de meios-irmãos. No início, o progresso genético com a seleção de clones dos plantios comerciais foi enorme, pois permitia a aplicação de grande intensidade de seleção, em função da existência de alguns milhares de plantas para serem selecionadas. Contudo, com a ampliação das áreas de plantio clonais reduziu-se a área seminal, e a possibilidade de selecionar nas extensas áreas de plantios comerciais. Para continuar tendo sucesso com a seleção, foi necessário acelerar os programas SR nas populações e ampliar o número de clones a serem avaliados (Nunes et al., 2002). Para condução de SR com maior eficiência, foi necessário buscar alternativas para ampliar o progresso genético por unidade de tempo.

Para isso, foram desenvolvidas algumas tecnologias. Uma delas foi com relação a necessidade de realização da hibridação artificial. As dificuldades de cruzar duas árvores de eucalipto sob condição de campo eram enormes. A primeira era com relação a não coincidência de florescimento das árvores. Para solucionar o problema, foi pesquisada a possibilidade de armazenamento dos pólenes, o que se mostrou viável usando algumas tecnologias (Pereira et al., 2002). Outra grande dificuldade prática era como realizar hibridação manualmente em função da altura das plantas no campo. Os melhoristas tiveram que desenvolver estratégias que atenuassem o problema. O sucesso foi grande e rápido. Para isso, utilizaram uma substância aplicada no solo que estimulou o florescimento em plantas bem jovens. Essa tecnologia foi associada a enxertia de topo (garfagem), no qual utilizam como cavalo uma planta clonada da mesma espécie que já tivesse passado pelo período juvenis, e como enxerto o ramo da planta em que se pretendia ter o florescimento. Com essa estratégia, foi possível realizar as hibridações em telados ou viveiros, com a altura das plantas de eucalipto semelhante à de uma planta anual (Figura 4).

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

Para complementar o processo, foi desenvolvida técnicas de hibridação artificial, entre elas a de Assis et al. (2005), que eles denominaram de Protoginia Artificialmente Induzida (PAI). O PAI envolve o corte da extremidade de um botão maduro. Nessa condição, o estigma está receptivo, e ainda não ocorreu a antese. Nessa operação, o estigma também é cortado, e a fertilização artificial é realizada em seguida colocando o pólen (genitor masculino) na parte cortada do estilete (Figura 5). Segundo os autores, esse procedimento, além de mais fácil e rápido, reduz a contaminação em relação ao que era tradicionalmente realizado, “abrindo” o botão floral e retirando as anteras. Todas essas técnicas contribuíram para acelerar a seleção recorrente, tanto na etapa de obtenção das progênies, no início do processo, como ao proceder o intercruzamento dos melhores indivíduos/progênies.

Como o desempenho dos indivíduos/árvores nos testes de progênies e o clone deles derivados não estavam tendo alta associação, aventou-se a possibilidade da utilização do teste de progênies clonadas. Nesse caso, ao mesmo tempo em que se avalia o indivíduo no teste de progênie, já se está avaliando os clones (cópias) deles. Esse procedimento melhora a eficiência da avaliação do indivíduo, porque ele está sendo repetido e, conseqüentemente, da progênie. Evidentemente, com enorme redução do tempo de obtenção de novos clones. Esse é um claro exemplo de SB que associa a aceleração do processo de melhoramento sem perder a oportunidade de avaliação a indivíduos/clones, o que normalmente não ocorre na maioria dos procedimentos preconizados anteriormente.

Foto: Flávia Maria Avelar Gonçalves



Figura 4. Indução do florescimento precoce em plantas de eucalipto cultivadas em vaso.

Foto: Flávia Maria Avelar Gonçalves

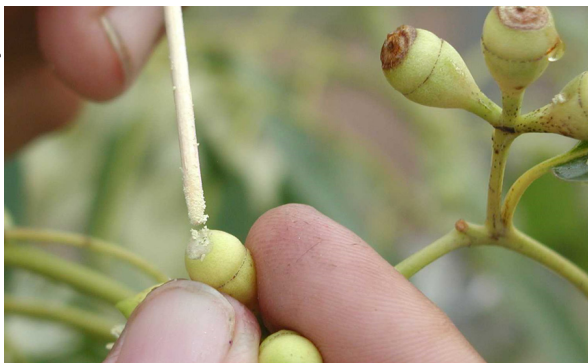


Figura 5. Realização da hibridação artificial utilizando a técnica da Protoginia Artificialmente Induzida (PAI).

A obtenção de linhagens visando a produção de sementes híbridas de eucalipto era inimaginável devido ao tempo demandado para realizar quatro a cinco gerações de autofecundação. Contudo, Castro et al. (2021) propuseram uma tecnologia de acelerar o florescimento. O procedimento proposto pelos autores preconiza que as mudas obtidas via sementes, com aproximadamente 150 a 180 dias após a semeadura, serão enxertadas em um ramo de uma planta que já passou o período juvenil. Detalhes do processo (Figura 6) como um todo é apresentado por Castro et al. (2021). Considerando que, com 90 dias após a enxertia o florescimento nos ramos enxertados deve ocorrer, a autofecundação pode ser realizada e as sementes da geração F_9 são obtidas com, aproximadamente, nove meses após a realização da enxertia das mudas advindas com sementes da geração anterior. Utilizando esse procedimento, a obtenção de linhagens torna-se factível, uma vez que demandaria tempo inferior a oito anos (Ramalho et al., 2022).

Fotos: Carla Castro



Figura 6. Procedimento para a obtenção de uma geração de eucalipto em um ano. A- Mudas obtidas a partir das sementes híbridas; B- Os ramos das mudas de plantas híbridas são enxertados via garfagem em plantas que já passaram pelo período juvenil; C - Crescimento e desenvolvimento da parte enxertada e D- Aparição dos primeiros botões florais nos ramos da planta.

Considerações Finais

Um complicador no emprego do SB é que ele é genotipicamente dependente. Ou seja, nem todos os genótipos respondem aos fatores ambientais que contribuem para a aceleração do crescimento e desenvolvimento. Evidentemente, esse é um problema que

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

pode prejudicar a adoção da tecnologia para algumas espécies. Outra limitação é o custo de implantação, na maioria dos casos, de todo o processo envolvido. Muitos equipamentos são caros, além do mais, no caso específico do Brasil, iria exigir aquisição deles, muitas vezes no exterior, o que demandaria um pouco mais de tempo e custo. Também não se deve esquecer do custo de materiais de consumo, como alguns reagentes e outros produtos. Muitos desses equipamentos consomem muita energia elétrica (Samantara et al., 2022), o que pode ser uma restrição do emprego de SB, sobretudo em algumas espécies que são melhoradas apenas em empresas públicas e universidades.

Destaques especiais devem ser dados quando o objetivo for reduzir o tempo para o início do florescimento. Assim, casas de vegetação são indispensáveis. Evidentemente, o tamanho de casa irá depender da espécie e da magnitude do programa de melhoria. Deve-se destacar que o custo por metro quadrado de implantação é elevado, mesmo para os padrões dos países em que restrição de recurso para pesquisa é bem menor que no Brasil. Dessa forma, é indispensável estudar alternativas de aceleração de ciclo, sem a necessidade de casas de vegetação. Todavia, em sendo necessário o seu emprego, algumas observações são necessárias:

a) Tentar fazer adaptação de uma casa de vegetação que já existe não é uma boa alternativa. O ajuste pode ficar mais caro do que fazer uma nova.

b) Evidentemente, deve-se preocupar em uniformizar ao máximo, em toda sua extensão, temperatura e umidade. Essa observação é particularmente importante para regiões tropicais.

c) Sempre que possível, deve-se racionalizar o uso de mão de obra. Deve-se considerar espaço para mecanização e automação, pois a colocação e retirada de vasos deve ser feita de maneira rápida a fim de contemplar o cronograma do SB. Além dos vasos, existe toda a questão da automação para fenotipagem e controle ambiental.

d) Atenção especial deve ser dada ao planejamento do uso racional de energia, que como já mencionado, é de custo elevado, e de água. Nesse contexto, procurar utilizar materiais de alta transparência e poder de difusão para reduzir o consumo de luz durante o dia. Os materiais existentes no mercado (policarbonato, vidro e plástico) reduzem pela metade o PPFd da luz solar.

e) Outro detalhe importante é tentar alocar todos os equipamentos, de modo a não prejudicar a colocação e retirada das plantas e a realização das hibridações artificiais quando for necessária.

Técnicas de SB que envolvam acelerar o período entre o ovo ou zigoto e obtenção de uma plântula, como já mencionado, requerem a necessidade de laboratórios e

pessoal qualificado em cultura de tecidos vegetais. Evidentemente que o protocolo pode variar de uma espécie para outra e o número de gerações a ser obtido por ano é também variável com a espécie. Uma dificuldade adicional, é o número de descendentes – sementes – que pode ser obtido de cada planta nessas condições totalmente artificiais.

Um comentário que serve para outros procedimentos de SB, incluindo o DH, é que a segregação esperada no final do processo é enorme. Seja, por exemplo, a obtenção de RIL – linhagens endogâmicas recombinantes - a partir da geração F_1 do cruzamento de duas linhagens que estejam segregando para n genes. Nessa condição, é esperado que o número de RILs genotipicamente diferentes seja de 2^n . Imagine-se que estejam segregando um número relativamente pequeno, por exemplo, dois em cada 1000 genes da espécie. Tendo como referência uma planta como o milho, cujo número de genes estimados é próximo de 30 mil, n será 60. Assim, serão esperados $2^{60} = 1,15 \times 10^{18}$ linhagens genotipicamente diferentes. Para se imaginar o que representa esse número, considere-se a semeadura de 70 mil plantas por hectare, com esses RILs. Nessa condição, seriam necessários 16,47 milhões de hectare para se ter todas as combinações possíveis. Ou seja, área superior à cultivada com milho, em um ano, no Brasil. Certamente, durante o processo muitas linhagens frágeis serão eliminadas, mas ainda permanece um número excessivamente grande, como ocorre com qualquer outro método de avanço das gerações segregantes até se obter linhagens.

Deve ser enfatizado que o importante é não ter um grande número de linhagens e sim ter boas linhagens ao final do processo. Nesse contexto, o fato de acelerar o programa de melhoramento para atingir rapidamente a homozigose muitas vezes diminui as chances de selecionar as plantas/progênies durante o avanço das gerações, pois as condições são totalmente artificiais. Nessa situação, não há outra alternativa a não ser de avaliar as linhagens ao final do processo de SB, sob condições de cultivo. O que evidentemente irá demandar tempo, reduzindo a vantagem da economia de tempo da SB. Esse fato foi comentado por Lemos et al. (2022), empregando duplo haploides no programa de SR com tabaco.

Um fato que não deve ser esquecido, e que nos programas de melhoramento já estabelecidos e bem planejados é perfeitamente possível para todas as espécies, quer sejam anuais ou até mesmo perene, é a obtenção de novas cultivares todos os anos. Assim, a vantagem de aceleração dos programas de seleção pode ser reduzida. Contudo, certamente irão ocorrer casos especiais, como por exemplo, a ocorrência de um novo patógeno, cuja fonte de resistência só é encontrada em germoplasma não adaptado. Nessa condição, certamente técnicas de SB poderão ser indispensáveis

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

visando a solução do problema o mais rápido possível. Esse fato foi discutido por Samantara et al. (2022), associando o SB a aplicações de outras ferramentas modernas que podem ajudar os melhoristas.

Não há dúvida que, nas próximas décadas, a sociedade irá exigir dos programas de melhoria maior agilidade na obtenção de novas cultivares melhores que as já existentes na produção de grãos, frutos, fibras e outros produtos da agricultura. Assim, novas tecnologias de SB serão sempre bem-vindas. A decisão ou não de sua utilização será evidentemente em função do seu benefício em relação ao custo de sua aplicação.

Referências

- ANDORF, C.; BEAVIS, W. D.; HUFFORD, M.; SMITH, S.; SUZA, W. P.; WANG, K.; WOODHOUSE, M.; YU, J.; LÜBBERSTEDT, T. Technological advances in maize breeding: past, present and future. **Theoretical and Applied Genetics**, v. 132, p. 817-849, Mar. 2019. DOI: <https://doi.org/10.1007/s00122-019-03306-3>.
- ASSIS, T.; WARBURTON, P.; HARWOOD, C. Artificially induced protogyny: an advance in the controlled pollination of *Eucalyptus*. **Australian Forestry**, v. 68, n. 1, p. 27–33, 2005. DOI: <https://doi.org/10.1080/00049158.2005.10676223>.
- BERNARDO, R. **Breeding for quantitative traits in plants**. 3rd ed. Woodbury: Stemma Press, 2020. 422 p.
- BHATTA, M.; SANDRO, P.; SMITH, M. R.; DELANEY, O.; VOSS-FELS, K. P.; GUTIERREZ, L.; HICKEY, L. T. Need for speed: manipulating plant growth to accelerate breeding cycles. **Current Opinion in Plant Biology**, v. 60, 101986, Apr. 2021. DOI: <https://doi.org/10.1016/j.pbi.2020.101986>.
- BOERMAN, N. A.; FREI, U. K.; LÜBBERSTEDT, T. Impact of spontaneous haploid genome doubling in maize breeding. **Plants**, v. 9, 369, Mar. 2020. DOI: <https://doi.org/10.3390/plants9030369>.
- BUCKLER, E. S.; HOLLAND, J. B.; BRADBURY, P. J.; ACHARYA, C. B.; BROWN, P. J.; BROWNE, C.; ERSOZ, E.; FLINT-GARCIA, S.; GARCIA, A.; GLAUBITZ, J. C.; GOODMAN, M. M.; HARJES, C.; GUILL, K.; KROON, D. E.; LARSSON, S.; LEPAK, N. K.; LI, H.; MITCHELL, S. E.; PRESSOIR, G.; PEIFFER, J. A.; ROSAS, M. O.; ROCHEFORD, T. R.; ROMAY, M. C.; ROMERO, S.; SALVO, S.; SANCHEZ VILLEDA, H.; SILVA, H. S.; SUN, Q.; TIAN, F.; UPADYAYULA, N.; WARES D.; YATES, H.; YU, J.; ZHANG, Z.; KRESOVICH, S.; MCMULLEN, M. D. The genetic architecture of maize flowering time. **Science**, v. 325, n. 5941, p. 714–718, Aug. 2009. DOI: <https://doi.org/10.1126/science.1174276>.
- CASTRO, C. A. O.; SANTOS, G. A.; TAKAHASHI, E. K.; NUNES, A. C. P.; SOUZA, G. A.; RESENDE, M. D. V. Accelerating *Eucalyptus* breeding strategies through top grafting applied to young seedlings. **Industrial Crops and Products**, v. 171, 113906, Nov. 2021. DOI: <https://doi.org/10.1016/j.indcrop.2021.113906>.
- CHA, J. K.; O'CONNOR, K.; ALAHMAD, S.; LEE, J. H.; DINGLASAN, E.; PARK, H.; LEE, S. M.; HIRSZ, D.; KWON, S. W.; KWON, Y.; KIM, K. M.; KO, J. M.; HICKEY, L. T.; SHIN, D.; DIXON, L. E. Speed vernalization to accelerate generation advance in winter cereal crops. **Molecular Plant**, v. 15, n. 8, p. 1300–1309, Aug. 2022. DOI: <https://doi.org/10.1016/j.molp.2022.06.012>.
- COBB, J. N.; JUMA, R. U.; BISWAS, P. S.; ARBELAEZ, J. D.; RUTKOSKI, J.; ATLIN, G.; HAGEN, T.; QUINN, M.; NG, E. H. Enhancing the rate of genetic gain in public-sector plant breeding programs: lessons from the breeder's equation. **Theoretical and Applied Genetics**, v. 132, n. 3, p. 627–645, Mar. 2019. DOI: <https://doi.org/10.1007/s00122-019-03317-0>.
- CONAB. Companhia Nacional de Abastecimento. Boletim da safra de grão: levantamento - safra 2022/23 Disponível em: <https://www.conab.gov.br/info-agro/safras/graos/boletim-da-safra-de-graos>. Acesso em: 20 maio 2023.

CONTINI, E.; ARAGÃO, A. **O Agro brasileiro alimenta 800 milhões de pessoas.**

Brasília, DF: Embrapa, 2021. Disponível em: <https://www.embrapa.br/busca-de-noticias/-/noticia/59784047/o-agro-brasileiro-alimenta-800-milhoes-de-pessoas-diz-estudo-da-embrapa>.

FORSTER, B. P.; TILL, B. J.; GHANIM, A. M. A.; HUYNH, H. O. A.; BURSTMAYR, H.; CALIGARI, P. D.S. Accelerated plant breeding. **CAB Reviews**, v. 9, 2014, Nov. 2014. DOI: <https://doi.org/10.1079/PAVSNNR20149043>.

GHOSH, S.; WATSON, A.; GONZALEZ-NAVARRO, O. E.; RAMIREZ-GONZALEZ, R. H.; YANES, L.; MENDOZA-SUÁREZ, M.; SIMMONDS, J.; WELLS, R.; RAYNER, T.; GREEN, P.; HAFEEZ, A.; HAYTA, S.; MELTON, R. E.; STEED, A.; SARKAR, A.; CARTER, J.; PERKINS, L.; LORD, J.; TESTER, M.; OSBOURN, A.; MOSCOU, M. J.; NICHOLSON, P.; HARWOOD, W.; MARTIN, C.; DOMONEY, C.; UAUY, C.; HAZARD, B.; WULFF, B. B. H.; HICKEY, L. T. Speed breeding in growth chambers and glasshouses for crop breeding and model plant research. **Nature Protocols**, v. 13, n. 12, p. 2944–2963, Nov. 2018. DOI: <https://doi.org/10.1038/s41596-018-0072-z>.

HALLAUER, A. R.; CARENA, M. J.; MIRANDA FILHO, J. B. **Quantitative genetics in maize breeding.** New York: Springer, 2010. 663 p.

HICKEY, L. T.; GERMÁN, S. E.; PEREYRA, S. A.; DIAZ, J. E.; ZIEMS, L. A.; FOWLER, R. A.; PLATZ, G. J.; FRANCKOWIAK, J. D.; DIETERS, M. J. Speed breeding for multiple disease resistance in barley. **Euphytica**, v. 213, 64, Mar. 2017. DOI: <https://doi.org/10.1007/s10681-016-1803-2>.

JÄHNE, F.; HAHN, V.; WÜRSCHUM, T.; LEISER, W. L. Speed breeding short-day crops by LED-controlled light schemes. **Theoretical and Applied Genetics**, v. 133, no. 8, p. 2335–2342, May 2020. DOI <https://doi.org/10.1007/s00122-020-03601-4>.

LEE, E. A.; TOLLENAAR, M. Physiological basis of successful breeding strategies for maize grain yield. **Crop Science**, v. 47, n. S3, p. S202-S215, Dec. 2007. DOI: <https://doi.org/10.2135/cropsci2007.04.0010IPBS>.

LEMOS, R. C. **Performance of doubled haploid lines in relation to those obtained by the conventional breeding method in tobacco.** 2021. 63 p. Tese (Doutorado em Genética e Melhoramento de Plantas) – Universidade Federal de Lavras, Lavras.

LEMOS, R. C.; PÁDUA, J. M. V.; BRUZI, A. T.; OLIVEIRA, V. B.; RAMALHO, M. A. P. Comparison between doubled haploid lines and lines obtained via the bulk method in tobacco. **Crop Breeding and Applied Biotechnology**, v. 22, e42992249, 2022. DOI: <https://doi.org/10.1590/1984-70332022v22n4a44>.

LI, H.; RASHEED, A.; HICKEY, L. T.; HE, Z. Fast-forwarding genetic gain. **Trends in Plant Science**, v. 23, n. 3, p. 184–186, Mar. 2018. DOI: <https://doi.org/10.1016/j.tplants.2018.01.007>.

LIN, X.; LIU, B.; WELLER, J. L.; ABE, J.; KONG, F. Molecular mechanisms for the photoperiodic regulation of flowering in soybean. **Journal of Integrative Plant Biology**, v. 63, n. 6, p. 981–994, June 2021. DOI: <https://doi.org/10.1111/jipb.13021>.

LIU, H.; ZWER, P.; WANG, H.; LIU, C.; LU, Z.; WANG, Y.; YAN, G. A fast generation cycling system for oat and triticale breeding. **Plant Breeding**, v. 135, n. 5, p. 574–579, Oct. 2016. DOI: <https://doi.org/10.1111/pbr.12408>.

MCCAW, M. E.; LEE, K.; KANG, M.; ZOBRIST, J. D.; AZANU, M. K.; BIRCHLER, J. A.; WANG, K. Development of a transformable fast-flowering mini-maize as a tool for maize gene editing. **Frontiers in Genome Editing**, v. 2, 622227, Jan. 2021. DOI: <https://doi.org/10.3389/fgeed.2020.622227>.

MONTILON, V.; SUSCA, L.; POTERE, O.; ROSETI, V.; CAMPANALE, A.; SAPONARI, A.; MONTEMURRO, C.; FANELLI, V.; VENERITO, P.; BOTTALICO, G. Embryo culture, in vitro propagation, and molecular identification for advanced olive breeding programs. **Horticulturae**, v. 8, 36, Jan. 2022. DOI: <https://doi.org/10.3390/horticulturae8010036>.

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

NIAZIAN, M.; SHARIATPANAH, M. E. In vitro-based doubled haploid production: recent improvements. **Euphytica**, v. 216, 69, Apr. 2020. DOI: <https://doi.org/10.1007/s10681-020-02609-7>.

NUNES, G. H. S.; REZENDE, G. D. S. P.; RAMALHO, M. A. P.; SANTOS, J. B. Implicações da interação genótipos x ambientes na seleção de clones de eucalipto. **Cerne**, v. 8, n. 1, p. 49–58, 2002.

PANDEY, S.; SINGH, A.; PARIDA, S. K.; PRASAD, M. Combining speed breeding with traditional and genomics-assisted breeding for crop improvement. **Plant Breeding**, v. 141, n. 3, p. 301–313, June 2022. DOI: <https://doi.org/10.1111/pbr.13012>.

PAZOS-NAVARRO, M.; CASTELLO, M.; BENNETT, R. G.; NICHOLS, P.; CROSER, J. In vitro-assisted single-seed descent for breeding-cycle compression in subterranean clover (*Trifolium subterraneum* L.). **Crop and Pasture Science**, v. 68, n. 10/11, p. 958-966, 2017. DOI: <https://doi.org/10.1071/CP17067>.

PEREIRA, R. C.; DAVIDE, L. C.; RAMALHO, M. A. P.; ANDRADE, H. B. Alternativas para aumentar a eficiência dos cruzamentos em programas de melhoramento de *Eucalyptus*. **Cerne**, v. 8, n. 2, p. 60-69, 2002.

RAMALHO, M. A. P.; MARQUES, T. L.; LEMOS, R. C. Plant breeding in Brazil: retrospective of the past 50 years. **Crop Breeding and Applied Biotechnology**, v. 21, e383021S3, maio/jun. 2021. 2021. DOI: <https://doi.org/10.1590/1984-70332021v21Sa16>.

RAMALHO, M. A. P.; SANTOS, H. G.; SOUZA, T. S. Eucalyptus breeding programs: a proposal for the use of inbred progênies. **Cerne**, v. 28, e-103049, 2022. DOI: <https://doi.org/10.1590/01047760202228013049>.

RESENDE, M. D. V. **Genética quantitativa e de populações**. Viçosa, MG: Suprema, 2015. 463 p.

REZENDE, G. D. S. P.; RESENDE, M. D. V.; ASSIS, T. F. Eucalyptus breeding for clonal forestry. In: FENNING, T. (ed.). **Challenges and opportunities for the World's Forests in the 21st Century**. Dordrecht: Springer, 2014. p. 393-424.

RIBALTA, F. M.; KAREN, M. P.; KYLIE, N.; ROSS, J. J.; BENNETT, R. G.; MUNDAY, C.; ERSKINE, W.; OCHAT, S. J.; CROSER, J. S. Precocious floral initiation and identification of exact timing of embryo physiological maturity facilitate germination of immature seeds to truncate the lifecycle of pea. **Plant Growth Regulation**, v. 81, n. 2, p. 345–353, 2017. DOI: <https://doi.org/10.1007/s10725-016-0211-x>.

ROWELL, T.; MORTLEY, D. G.; LORETAN, P. A.; BONSI, C. K.; HILL, W. A. Continuous daily light period and temperature influence peanut yield in nutrient film technique. **Crop Science**, v. 39, n. 4, p. 1111–1114, July/Aug. 1999. DOI: <https://doi.org/10.2135/cropsci1999.0011183X003900040026x>.

RUTKOSKI, J. E. A practical guide to genetic gain. **Advances in Agronomy**, v.157, p. 217-249, 2019. DOI: <https://doi.org/10.1016/bs.agron.2019.05.001>.

SAMANTARA, K.; BOHRA, A.; MOHAPATRA, S. R.; PRIHATINI, R.; ASIBE, F.; SINGH, L.; REYES, V. P.; TIWARI, A.; MAURYA, A. K.; CROSER, J. S.; WANI, S. H.; SIDDIQUE, K. H. M.; VARSHNEY, R. K. Breeding more crops in less time: a perspective on speed breeding. **Biology**, v. 11, 275, Feb. 2022. DOI: <https://doi.org/10.3390/biology11020275>.

SHAFIQ, I.; HUSSAIN, S.; RAZA, M. A.; IQBAL, N.; ASGHAR, M. A.; RAZA, A.; FAN, Y.-F.; MUMTAZ, M.; SHOAI, M.; ANSAR, M.; MANAF, A.; YANG, W.-Y.; YANG, F. Crop photosynthetic response to light quality and light intensity. **Journal of Integrative Agriculture**, v. 20, n. 1, p. 4–23, Jan. 2021. DOI: [https://doi.org/10.1016/S2095-3119\(20\)63227-0](https://doi.org/10.1016/S2095-3119(20)63227-0).

SINHA, P.; SINGH, V. K.; BOHRA, A.; KUMAR, A.; REIF, J. C.; VARSHNEY, R. K. Genomics and breeding innovations for enhancing genetic gain for climate resilience and nutrition traits. **Theoretical and Applied Genetics**, v. 134, n. 6, p. 1829–1843, May 2021. DOI: <https://doi.org/10.1007/s00122-021-03847-6>.

SOLTANI, A.; WALTER, K. A.; WIERSMA, A. T.; SANTIAGO, J. P.; QUIQLEY, M.; CHITWOOD, D.; PORCH, T. G.; MIKLAS, P.; MCCLEAN, P. E.; OSORNO, J. M.; LOWRY, D. B. The genetics and physiology of seed dormancy, a crucial trait in common bean domestication. **BMC Plant Biology**, v. 21, 58, Jan. 2021. DOI: <https://doi.org/10.1186/s12870-021-02837-6>.

SOUZA JUNIOR, C. L. Melhoramento de espécies alógamas. In: NASS, L. L.; VALOIS, A. C. C.; MELO, I. S. de; VALADARES-INGLIS, M. C. (ed.). **Recurso genético e melhoramento**. Rondonópolis: Fundação Mato Grosso, 2001. p. 159-199.

UNITED NATIONS. Department of Economic and Social Affairs. **World population prospects 2022: summary of results**. UN DESA/POP/2022/TR/NO. 3. Disponível em: https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/wpp2022_summary_of_results.pdf. Acesso em: 20 maio 2023.

WANGA, M. A.; SHIMELIS, H.; MASHILO, J.; LAING, M. D. Opportunities and challenges of speed breeding: a review. **Plant Breeding**, v. 140, n. 2, p. 185–194, Apr. 2021. DOI: <https://doi.org/10.1111/pbr.12909>.

WATSON, A.; GHOSH, S.; WILLIAMS, M. J.; CUDDY, W. S.; SIMMONDS, J.; REY, M. D.; HATTA, M. A.; HINCHLIFFE, A.; STEED, A.; REYNOLDS, D.; ADAMSKI, N. M.; BREAKSPEAR, A.; KOROLEV, A.; RAYNER, T.; DIXON, L. E.; RIAZ, A.; MARTIN, W.; RYAN, M.; EDWARDS, D.; BATLEY, J.; RAMAN, H.; CARTER, J.; ROGERS, C.; DOMONEY, C.; MOORE, G.; HARWOOD, W.; NICHOLSON, P.; DIETERS, M. J.; DELACY, I. H.; ZHOU, J.; UAUY, C.; BODEN, S. A.; PARK, R. F.; WULFF, B. B. H.; HICKEY, L. T. Speed breeding is a powerful tool to accelerate crop research and breeding. **Nature Plants**, v. 4, n. 1, p. 23–29, Jan. 2018. DOI: <http://dx.doi.org/10.1038/s41477-017-0083-8>.

ZHENG, Z.; WANG, H. B.; CHEN, G. D.; YAN, G. J.; LIU, C. J. A procedure allowing up to eight generations of wheat and nine generations of barley per annum. **Euphytica**, v. 191, n. 2, p. 311–316, Mar. 2013. DOI: <https://doi.org/10.1007/s10681-013-0909-z>.

Renata Fuganti Pagliarini¹, Mayla Daiane Correa Molinari¹,
Fabricio Barbosa Monteiro Arraes, Alexandre Lima
Nepomuceno, Hugo Bruno Correa Molinari

Introdução

Nos últimos anos, o surgimento de tecnologias de edição de genoma, também chamadas novas tecnologias de melhoramento (NBTs – *New Breeding Technologies*), altamente versáteis, forneceram aos pesquisadores a capacidade de introduzir, rápida e economicamente, modificações específicas em sequências genômicas de um amplo espectro de tipos de células e organismos. Em particular, a facilidade com que a metodologia CRISPR/Cas (*Clustered Regularly Interspaced Short Palindromic Repeats* - repetições palindrômicas curtas agrupadas regularmente e inter espaçadas) pode ser utilizada para reconhecer novas sequências impulsionou uma revolução na edição de genomas, o que acelerou avanços científicos e descobertas em disciplinas diversas como biologia sintética, terapia genética humana, modelagem de doenças, descoberta de medicamentos, neurociência, e as ciências agrárias (Gaj et al., 2016).

Na agricultura, o uso de novas tecnologias de melhoramento genético abre possibilidades para o desenvolvimento de novas cultivares que podem superar as limitações do melhoramento clássico. Essas tecnologias eliminam ou limitam os riscos associados à primeira geração de ferramentas que resultaram em organismos geneticamente modificados (OGMs), por meio de diferentes técnicas de transgenia. No setor agrícola, espera-se que as barreiras ideológicas e de aceitação pelo público consumidor sejam superadas à medida que a informação científica se torne mais amplamente divulgada, permitindo que os benefícios dos produtos derivados dessas novas tecnologias cheguem aos consumidores, às lavouras e às prateleiras

¹Ambos os autores contribuíram igualmente.

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

dos supermercados. Embora muitas das aplicações das novas tecnologias de melhoramento genético estejam em fase de pesquisa, alguns produtos já são uma realidade comercial, como será apresentado no decorrer do texto.

Uma característica notável dos produtos resultantes das NBTs é a potencialidade de contornar o laborioso e oneroso processo de desregulamentação que atualmente os OGMs estão sujeitos. Diversos países produtores e consumidores de NBTs, incluindo o Brasil, estão discutindo e implantando novas legislações para regular esses produtos. Nos novos termos, a maioria desses produtos, dependendo do tipo de alteração genética realizada, será considerada não geneticamente modificada (GM), o que facilitará e tornará mais barato o processo de desregulamentação. Tal fato amplia a competição entre os diferentes agentes do cenário científico, permitindo a emergência de novos “players”, além das grandes empresas multinacionais, e acelerando o tempo em que esses produtos chegarão aos produtores e aos consumidores finais.

As NBTs estão sendo amplamente utilizadas, tanto no setor público quanto no privado, para edição de genomas. No setor público, as NBTs são aplicadas para pesquisas básicas de cultivo com o objetivo de ampliar o conhecimento sobre a interação entre as culturas e o meio ambiente (Wessberg et al., 2021). Além disso, pesquisadores de institutos públicos, como a Embrapa (Empresa Brasileira de Pesquisa Agropecuária), também estão envolvidos na introdução e avaliação de características orientadas para o mercado consumidor.

No setor privado, muitas empresas têm intensificado a pesquisa, o desenvolvimento e as atividades de melhoramento genético, utilizando a edição de genomas como uma ferramenta adicional às técnicas de melhoramento clássico (Jorasch, 2020). Empresas de P&D de maior porte têm diversificado as abordagens para diferentes cultivares, com o objetivo de torná-las mais adaptadas aos mercados específicos, levando em conta a regulamentação e aceitação pelos consumidores de produtos editados geneticamente. Esse cenário tem permitido o surgimento de diversas startups, como Benson Hill, Calyxt, Inari, Pairwise, Plantedit, SolEdits e Tropic Bioscience, que aplicam especificamente a tecnologia de edição de genomas para melhorar cultivares de diferentes espécies, impulsionando a economia tanto dos países de origem quanto mundial (European Parliament, 2022). Países em desenvolvimento, como o Brasil, estão avançando rapidamente no uso das NBTs para melhorar a agricultura nacional. Empresas privadas nacionais, como a SEMPRE AgTech/WIN, dentre outras, vem sendo motivadas a investir e desenvolver tecnologias inovadoras que visam transformar o setor agrícola. Essas tecnologias

são projetadas para ajudar os agricultores a superar os desafios únicos que enfrentam, como a ocorrência de insetos-pragas e doenças, deficiência hídrica e os efeitos das mudanças climáticas. Com o aumento do investimento e aplicação em NBTs, espera-se que o Brasil possa aumentar a produtividade e a sustentabilidade da agricultura, além de gerar benefícios econômicos significativos para a sociedade como um todo.

Neste capítulo serão discutidas as principais técnicas de edição de genomas e exemplos de sua aplicação em culturas de interesse econômico, tanto em fases de pesquisa quanto de comercialização. Também serão abordados os aspectos regulatórios dos produtos obtidos por essas técnicas, bem como a percepção do público em relação a eles e o que se espera para o futuro dessa tecnologia.

Edição Gênica CRISPR/Cas

O sistema CRISPR/Cas tem papel importante na imunidade adaptativa em bactérias. Nesses organismos, o sistema CRISPR tipo II fornece proteção contra o DNA de vírus invasores e plasmídeos por meio da clivagem do DNA guiada por proteínas Cas (Wiedenheft et al., 2012; Sorek et al., 2013). Segmentos curtos de DNA estranho são integrados no *locus* CRISPR e transcritos em CRISPR RNA (crRNA), que então se ligam ao crRNA transativador (tracrRNA - *trans-activating crRNA*) para direcionar a degradação específica da sequência do DNA patogênico pela proteína Cas (Jinek et al., 2012). No ano de 2012, Emmanuelle Charpentier, Jennifer Doudna e colaboradores relataram que o reconhecimento do alvo pela proteína Cas9 requer apenas uma sequência específica dentro do crRNA e um motivo adjacente ao protoespaçador conservado (PAM - *Protospacer-Adjacent Motif*) localizada a montante do local de ligação do crRNA (Jinek et al., 2012). Desde então esse sistema foi simplificado para ser utilizado na engenharia de genomas e atualmente consiste apenas na nuclease Cas9, e suas variantes, e um único RNA guia (gRNA – *guide RNA*) contendo os elementos essenciais de crRNA e tracrRNA (Gaj et al., 2016).

Resumidamente e de forma simplificada: o sistema CRISPR/Cas9 consiste em duas moléculas-chave que introduzem uma mudança no DNA. Uma enzima chamada Cas9 atua como um par de “tesouras moleculares” que cortam as fitas duplas de DNA em um local específico do genoma para que segmentos de DNA possam ser adicionados ou removidos; e um segmento de RNA chamado RNA guia (gRNA) pré-desenhado (cerca de 20 bases de comprimento), que apresenta bases complementares aos da sequência de DNA alvo no genoma e está localizado dentro de uma sequência de RNA

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

mais longa, que “guia” a Cas9, assegurando que a enzima corte no ponto desejado do genoma. Nessa fase, a célula reconhece que o DNA está danificado e tenta repará-lo. Com essa descoberta, os cientistas puderam usar a maquinaria de reparo do DNA para introduzir alterações em um ou mais genes no genoma de uma célula de interesse, editando-o.

A diversidade de resultados genéticos possibilitados por essas tecnologias se deve, em grande parte, de sua capacidade de induzir eficientemente quebras na fita dupla de DNA de forma direcionada (DSBs – *Double Strand Breaks*). Essas quebras de DNA então conduzem a ativação das vias de reparo do DNA celular e facilitam a introdução de modificações genômicas específicas no local (Rouet et al., 1994; Choulifa et al., 1995). Esse processo é mais frequentemente usado para obter o nocaute (silenciamento) de genes por meio de inserções de base aleatórias e/ou deleções que podem ser introduzidas por junção de extremidade não homóloga (NHEJ - *Nonhomologous End Joining*). Alternativamente, na presença de um modelo de doador com homologia com o sítio cromossômico alvo, pode ocorrer a integração gênica ou correção de base via reparo direcionado por homologia (HDR - *Homology-Directed Repair*) (Gaj et al., 2016).

De fato, a simplicidade da programação CRISPR/Cas, juntamente com um mecanismo específico de clivagem de DNA, a capacidade de reconhecimento de alvo e a existência de muitas variantes do sistema, permitiram desenvolvimentos notáveis, aplicando esse método econômico e fácil de usar para direcionar, editar, modificar, regular e marcar com precisão e eficiência os loci genômicos de uma ampla gama de células e organismos (Doudna; Charpentier, 2014).

Na Figura 1 é apresentada uma visualização esquemática do sistema CRISPR/Cas, destacando similaridades e diferenças entre as endonucleases Cas9 e Cas12a (Cpf1), as quais têm sido amplamente empregadas em programas de edição genômica por diversas instituições, abrangendo empresas públicas, privadas e universidades. Características notáveis das enzimas Cas9 e Cas12a são destacadas, incluindo a diferença nos modos de corte genômico (um corte abrupto para Cas9 e um corte coesivo para Cas12a), a necessidade de gRNA de tamanho variável para cada enzima, e as dimensões distintas das próprias enzimas. Adicionalmente, enfatiza-se o processo subsequente de reparação de DNA, incorporando elementos como a recombinação homóloga e a junção de extremidades não homólogas.

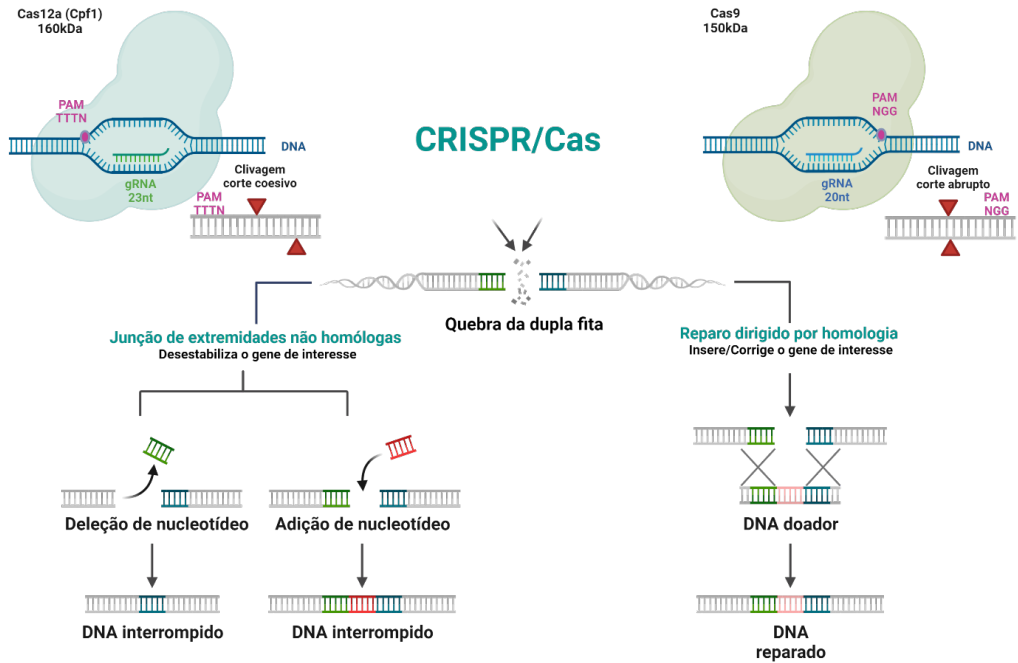


Figura 1. Comparação esquemática do sistema CRISPR/Cas, enfatizando as endonucleases Cas9 e Cas12a e a subsequente reparação do DNA (Criado com BioRender.com).

Exemplos de Aplicação das NBTs

As NBTs estão se consolidando, e a tendência é que haja presença cada vez mais frequente de produtos desenvolvidos com essas tecnologias no mercado agrícola e nas prateleiras dos supermercados. É evidente que o uso das NBTs é irreversível e, para garantir o sucesso desses produtos, é essencial regulamentação confiável, bem como rotulagem precisa e eficaz para transmitir informações aos consumidores sobre as culturas editadas e seus produtos relacionados. Abaixo estão alguns exemplos bem-sucedidos de aplicação das NBTs no desenvolvimento de produtos agrícolas, tanto em fase de pesquisa quanto em fase comercial.

Mediante busca no banco de dados EU-SAGE (<https://www.eu-sage.eu/genome-search>), foram identificadas até abril de 2023, 614 plantas editadas utilizando a técnica CRISPR/Cas, na grande maioria modificadas por meio do silenciamento (SDN1; site direct nuclease 1 - sítio de mutação direta da nuclease). A aplicação das NBTs foi realizada em mais de 60 espécies vegetais distintas. É importante ressaltar que o número de publicações que utilizam as NBTs aumentou significativamente a partir de 2015, com grande aumento nos anos subsequentes (Dima et al., 2022; European Parliament, 2022;

EU-SAGE, 2023). O EU-SAGE é banco de dados atualizado continuamente e novos trabalhos são inseridos à medida que são publicados em revistas científicas.

Dados mostram que no mundo, até o ano de 2023, as aplicações orientadas para o mercado incluíram plantas cultivadas e ornamentais, frutíferas e flores (EU-SAGE). A maioria dessas aplicações foram realizadas em arroz, tomate, milho, trigo e soja (Figura 2). No entanto, outras culturas com menor área de plantio, tais como canola, batata, tabaco e cevada, entre muitas outras espécies, também são alvos de edição gênica. Encontram-se ainda no EU-SAGE informações sobre diversas espécies frutíferas com menores números de publicações, como morango, kiwi, laranja, toranja e cacau, além de flores (dente de leão, glória da manhã, orquídea, torênia e petúnia). Resumidamente, a maioria das características modificadas está em categorias agrônômicas, tais como desempenho no crescimento, aumento de rendimento, seguidas de qualidade do produto, usos industriais, caracteres relacionados a cor e sabor, tolerância a herbicidas, aumento de tempo de prateleira, tolerância aos estresses bióticos, tais como fungos, bactérias e vírus, entre outros, e estresses abióticos tais como calor e salinidade, entre outros, além das características combinadas (Menz et al., 2020; EU-SAGE, 2023). Quanto à metodologia de edição gênica utilizada, a maioria das culturas editadas foram geradas via sistema CRISPR/Cas9, com poucas sendo geradas utilizando outras enzimas, como Cas13 e Cas12a/Cpf1. China e Estados Unidos (EUA) estão na liderança do número de produtos de edição gênica. Vários países da Europa, tais como Bélgica, Holanda, Reino Unido, Suécia, Alemanha, Espanha, Hungria, Itália e França também figuram nesse cenário. Brasil, Israel, Arábia Saudita, Coreia do Sul, Nova Zelândia, Quênia, Rússia e Turquia também vêm desenvolvendo pesquisas com NBTs (Menz et al., 2020; Parisi; Rodríguez-Cerezo, 2021; EU-SAGE, 2023).

Mais especificamente, algumas das características trabalhadas em culturas de grande importância econômica incluem floração precoce em arroz, com maior tamanho e número de grãos, enriquecido com carotenoides, e com resistência a doenças; camelina com composição de óleo alterada; soja com maiores teores de óleo e proteína; milho com maior produtividade de grãos em condições normais de cultivo e sob déficit hídrico; cana-de-açúcar com aumento no teor de sacarose e processo de sacarificação da biomassa; tomate com floração precoce, com alto teor de licopeno e tempo de prateleira mais longo; trigo rico em fibras, com baixo teor de glúten (redução de fatores alergênicos), e resistente a fungos; mostarda com sabor melhorado; amendoim com composição de óleo alterada, e batata sem glicoalcaloides e teores de amilopectina alterados, dentre outros. Para as espécies frutíferas, destacam-se características como videiras resistentes a fungos, com remoção do vírus da raia da banana, e morangos que florescem várias vezes (EU-SAGE, 2023).

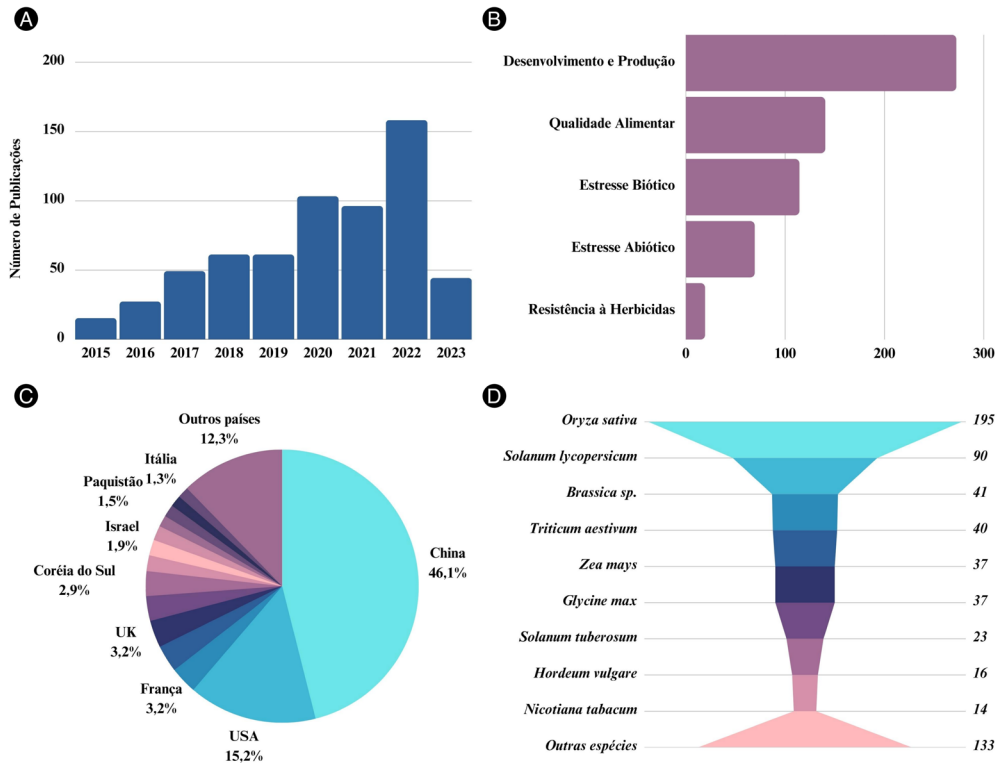


Figura 2. CRISPR/Cas na agricultura mundial, a partir da compilação de dados de 614 publicações selecionadas na rede EU-SAGE (European Sustainable Agriculture Through Genome Editing; <https://www.eu-sage.eu/>), disponíveis no mês de abril 2023. (A) Distribuição das publicações entre os anos de 2015-2023, mostrando tendência de aumento anual; (B) Distribuição das publicações em cinco classes temáticas; (C) Países líderes mundiais em pesquisas na área; (D) Espécies vegetais representadas nesses estudos.

Apesar de todos os esforços na pesquisa, as aplicações nas fases pré-comerciais e de mercado ainda são escassas (Parisi; Rodríguez-Cerezo, 2021). As razões para isso podem estar relacionadas ao fato de que novas técnicas genômicas (especialmente aquelas baseadas em CRISPR) ainda são recentes e/ou ainda residem incertezas regulatórias sobre essas técnicas em vários países. No entanto, espera-se que muitos outros produtos apareçam no futuro e beneficiem o consumidor final (Metje-Sprink et al., 2020; European Parliament, 2022).

Nos EUA, está disponível comercialmente uma cultivar de soja com óleo contendo alto teor de ácido oleico. Essa variedade contém mais ácidos graxos oleicos e menos linolênicos, resultando em maior estabilidade oxidativa do óleo quando submetido

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

ao calor (Demorest et al., 2016). No Japão, uma variedade de tomate com níveis aumentados de ácido gama-aminobutírico (GABA) chegou ao mercado em setembro de 2021. O tomate japonês, da empresa Sanatech Seed's Sicilian Rouge, aprovado para comercialização em dezembro de 2020, promete que a ingestão de GABA pode ajudar a diminuir a pressão arterial, promover o relaxamento e facilitar a concentração (Nonaka et al., 2017; European Parliament, 2022; Waltz, 2022). Ainda nesse país, foi aprovada, em 2021, a venda de duas espécies de peixes editados via CRISPR: um baiacu-tigre e outro dourado, ambos desenvolvidos pela startup *Regional Fish Institute*, com sede em Kyoto, em parceria com as Universidades de Kyoto e Kindai. As espécies editadas, que em breve estarão disponíveis no mercado, foram melhoradas para crescer mais do que as convencionais. No baiacu-tigre os pesquisadores conseguiram interromper o gene receptor de leptina, que controla o apetite, fazendo com que os peixes comam mais, aumentando a velocidade com que ganham peso. Os peixes editados crescem 1,9 vezes mais que os convencionais, permitindo-lhes atingir o tamanho de mercado mais cedo. Para o baiacu-dourado, os pesquisadores desativaram a proteína miostatina, que suprime o crescimento muscular, permitindo o peixe crescer cerca de 1,2 vezes mais, consumindo a mesma quantidade de alimento. Com a introdução dessas características espera-se reduzir os custos de produção dos peixes que serão cultivados em sistema de piscicultura (Japan..., 2022).

Uma curiosidade interessante é que pesquisadores já foram capazes de editar até 35 alelos de uma única vez via CRISPR/Cas9. Sánchez-León et al. (2018) editaram 35 de 45 cópias do gene α -gliadina simultaneamente. A gliadina é um componente das proteínas do glúten no trigo que desencadeia, a partir de sua ingestão, um distúrbio autoimune chamada doença celíaca em indivíduos geneticamente predispostos. A mutação das 35 cópias reduziu a imunorreatividade ao trigo em 85% (Sánchez-León et al., 2018).

No Brasil, diversos organismos foram geneticamente editados, dentre eles, eventos para culturas economicamente importantes, inclusive com aprovação para uso comercial (CTNBIO, 2022). Nessa lista encontram-se três cultivares de soja (duas desenvolvidas pela GDM Genética do Brasil S.A. e uma pela Embrapa Soja), duas variedades de cana de açúcar pela Embrapa Agroenergia e uma linhagem de milho pela Corteva. As cultivares de soja da empresa GDM apresentam níveis mais baixos de rafinose e tolerância à deficiência hídrica. O evento que teve o gene da rafinose editado apresentou redução de 75% de rafinose e 50% de estaquiose nas sementes, agregando valor para a cadeia produtiva da soja. A utilização da cultivar editada na avicultura e suinocultura deve aumentar a qualidade nutricional das rações, reduzindo custos com engorda e permitindo ganho de peso mais rápido nos animais (CTNBIO, 2022). Por sua importância na indústria de rações, a manipulação de fatores antinutricionais foi

também o alvo de pesquisa da Embrapa Soja. A cultivar editada possui um dos fatores antinutricionais desativado, a lectina, e foi obtida pela tecnologia CRISPR/Cas9. Cabe ressaltar que esses fatores antinutricionais dificultam a digestão da soja por animais monogástricos, que possuem estômago com capacidade de armazenamento reduzida, causando gases, desconforto intestinal e, conseqüentemente, redução no ganho de peso. Assim, o uso da soja convencional para esses fins depende de processamento térmico para inativar esses fatores antinutricionais, aumentando os custos de produção. Esse evento editado poderá reduzir os custos da utilização da soja para alimentação animal. É importante destacar ainda que, por serem consideradas não transgênicas, a liberação comercial dessas cultivares é mais rápida, reduzindo custos e facilitando a entrada de produtos no mercado com biossegurança assegurada (CTNBIO, 2022).

Duas variedades de cana-de-açúcar editadas foram desenvolvidas pela Embrapa Agroenergia, a partir do silenciamento dos genes *BAHD1* e *BAHD5* (acil transferases). Ambas variedades foram obtidas pelo método de Ribonucleoprotein (RNP) via CRISPR/Cas9, e apresentam maior sacarificação da biomassa vegetal e maior acúmulo de sacarose, permitindo ganhos incrementais na produção convencional de etanol de 2ª geração (2G), ração animal e produção de compostos químicos de alto valor agregado (CTNBIO, 2022).

A tecnologia CRISPR/Cas também foi usada para deletar parte do gene *waxy - Wx1* em linhagens de milho dentado da DuPont Pioneer (atual Corteva), resultando no fenótipo de milho ceroso, que reduziu a produção de enzimas funcionais, interrompendo assim a rota de síntese da amilose, e conduzindo a produção de amido com quase 100% de amilopectina. Esse milho oferece vantagens industriais, melhorando a uniformidade, estabilidade e textura em vários produtos alimentícios, além de apresentar melhores qualidades ligantes no processo de fabricação de papel, na indústria têxtil e na produção de adesivos (Grain, 2021).

Para os próximos anos, espera-se que mais cultivares geneticamente editadas sejam disponibilizadas aos agricultores para plantio e que novos produtos contendo diferentes características cheguem ao consumidor final. Em 2023, por exemplo, espera-se que uma variedade de mostarda enriquecida com nutrientes seja introduzida no mercado americano (European Parliament, 2022).

Biossegurança dos Produtos Gerados pelas NBTs

A regulação dos produtos gerados pelas NBTs ainda é assunto de muita discussão mundialmente. O debate principal foi e ainda é sobre se os organismos geneticamente

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

editados (GEDs – *Genetically Edited Organisms*) devem ser considerados no mesmo nível de escrutínio que os organismos geneticamente modificados (GMOs - *Genetically Modified Organisms*). A seguir, serão abordados os princípios básicos de regulamentação de culturas editadas em alguns países.

Em 2023, o mais alto tribunal da União Europeia, o Tribunal de Justiça da União Europeia (ECJ – *European Court of Justice*), resolveu, em uma decisão histórica, que as plantas geneticamente editadas não estão sujeitas ao escopo regulatório aplicado aos organismos geneticamente modificados (OGMs). Essa decisão legal com base científica está em alinhamento com a legislação já em vigor em países líderes nesse contexto, como Brasil e Estados Unidos. Esse marco representa um grande avanço para a biotecnologia vegetal Europeia e mundial, expandindo as possibilidades de mercado no setor agrícola. Esse é um passo significativo para a comunidade científica, revolucionando a maneira como cultivamos nossa comida e contribuimos para alimentar o mundo (Chee, 2023).

Novos esforços estão sendo realizados na tentativa de se estabelecer uma legislação única aos produtos NBTs. Em linhas gerais, duas direções estão sendo consideradas: a primeira é estabelecer tratamento especial para os OGMs, restringindo a natureza de um OGM à descrição genética da nova característica reivindicada. Não importaria qual método foi utilizado, mas sim outras alterações genéticas e epigenéticas que a metodologia pode causar, avaliando-se o produto e não a história de sua obtenção. A segunda direção seria criar tratamento especial para os OGMs que contribuam para as metas de sustentabilidade, onde um OGM rotulado e “sustentável” passaria por processo mais leve de avaliação. No entanto, um OGM “não sustentável” não seria autorizado (Krinke; Meunier, 2022).

Nos EUA, as novas variedades de plantas produzidas por NBTs são reguladas por três agências diferentes: no Departamento de Agricultura dos EUA, o Serviço de Inspeção de Saúde Animal (USDAAPHIS - *US Department of Agriculture, Animal Plant Health Inspection Service*) regula o risco de pragas de plantas sob a Lei de Proteção de Plantas (*Plant Protection Act*); a Agência de Proteção Ambiental (*Environmental Protection Agency*) regula os protetores incorporados às plantas sob a Lei Federal de Inseticidas, Fungicidas e Rodenticidas (*Federal Insecticide, Fungicide, and Rodenticide Act*); e a FDA (*Food and Drug Administration*) regulamenta a segurança alimentar sob a Lei Federal de Alimentos, Medicamentos e Cosméticos (*Federal Food, Drug, and Cosmetic Act*).

Essas três agências regulam as características dos produtos em si e não o processo para desenvolvê-los. O documento intitulado “SECURE” (*Sustainable,*

Ecological, Consistent, Uniform, Responsible, Efficient - Sustentável, Ecológico, Consistente, Uniforme, Responsável, Eficiente) (Detalhes disponíveis em <https://www.federalregister.gov/documents/2019/06/06/2019-11704/movement-of-certain-genetically-engineered-organisms>), compilado pelo APHIS, em 2020, isentaria (não regularia) plantas editadas por genes que poderiam ter sido desenvolvidas por metodologias convencionais, sob a mesma regulamentação de OGMs já existentes. Essas isenções têm como foco tornar a regulamentação de plantas editadas mais alinhada com as diretrizes para culturas criadas convencionalmente, que, embora sejam “isentas de risco”, têm riscos determinados como “gerenciáveis por padrões aceitos” (Global Gene Editing Regulation Tracker, 2022).

No Canadá, a Agência Canadense de Inspeção de Alimentos (CFIA- *Canadian Food Inspection Agency*) é responsável por regular as plantas geneticamente modificadas (chamadas PNTs – *Plants with novel traits*) para a segurança ambiental. O departamento de saúde (*Health Canada*) é responsável pela segurança de alimentos e rações e deve avaliar e aprovar as PNTs usadas para alimentação humana ou animal e notificar o público antes que possam ser vendidas no país. A CFIA avalia os produtos caso a caso para decidir se eles contêm novas características, com foco no produto final, não no processo usado para criar o produto. Quaisquer plantas, alimentos ou rações que contenham características novas requerem avaliações ambientais e de segurança para serem aprovadas. Embora o Canadá pareça estar caminhando para regular com menos restrições as culturas editadas por genes, permanece a incerteza sobre quais tipos e quantos produtos de edição de genes irão necessitar a supervisão e qual será esse nível de supervisão. Nesse contexto, atualmente, ambas as agências estão propondo novos documentos de “orientação regulatória” com foco em PNTs desenvolvidas por meio da tecnologia de edição de genes (Smyth, 2021; Global Gene Editing Regulation Tracker, 2022).

Na China, o Ministério da Agricultura e Assuntos Rurais (*Ministry of Agriculture and Rural Affairs*) publicou em janeiro de 2022 novas regras sobre testes de campo para plantas obtidas por NBTs, abrindo caminho para um melhoramento mais rápido das culturas, visando a segurança alimentar. As diretrizes preliminares de edição de genes indicam que, uma vez que testes-piloto com as plantas editadas tenham sido concluídos, um certificado de produção pode ser solicitado, eliminando a necessidade de longos testes de campo aplicados para plantas GM (ISAAA, 2022).

O Japão possui regras consideradas leves para culturas editadas, que devem ser registradas, mas não requerem testes de segurança ou ambientais, a menos que a planta contenha DNA exógeno. As culturas e alimentos obtidos por NBTs são avaliados

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

caso a caso e exigem uma notificação ao governo, que inclui informações sobre a técnica utilizada e os genes modificados. Se a cultura editada for cruzada com outra convencional ou editada para outros genes, um processo de notificação separado deve ocorrer. Quatro diferentes ministérios regulam as culturas e alimentos geneticamente modificados/editados: o Ministério da Agricultura, Silvicultura e Pesca (MAFF - *Ministry of Agriculture, Forestry and Fisheries*), o Ministério da Saúde, Trabalho e Bem-Estar (MHLW - *Ministry of Health, Labour and Welfare*), o Ministério do Meio Ambiente (MOE - *Ministry of Environment*) e o Ministério da Educação, Cultura, Esportes, Ciência e Tecnologia (MEXT - *Ministry of Education, Culture, Sports, Science and Technology*). A Comissão de Segurança Alimentar (FSC - *Food Safety Commission*), um órgão independente, realiza avaliação de risco de segurança de alimentos e rações para MHLW e MAFF. Adicionalmente, governos locais podem estabelecer requisitos regulatórios adicionais para culturas editadas. As recomendações, no entanto, não abordam os requisitos de rotulagem para alimentos obtidos via NBTs (Global Gene Editing Regulation Tracker, 2022).

Nos países da América Latina, não há consenso. Países como Belize, Bolívia, México, Peru e Venezuela possuem alta restrição aos GDEs, já Argentina, Brasil, Chile, Colômbia, Equador, Guatemala, Honduras e Paraguai apresentam níveis mais baixos de restrição. A definição das regras de regulamentação desses produtos está em andamento em El Salvador e Panamá. Costa Rica, República Dominicana, Nicarágua, Trindade e Tobago e Uruguai não possuem regulamentação definida, enquanto para países como Guiana Francesa, Guiana e Suriname não existe nenhuma informação disponível sobre o *status* da regulamentação de produtos obtidos via NBTs.

No Brasil, a Resolução Normativa nº 16 (RN16), publicada em 15 de janeiro de 2018, descreve todas as diretrizes e requisitos técnicos para apresentação de consulta à CTNBio sobre produtos derivados das NBTs. De forma resumida, esse documento determina, por meio de análise caso a caso, se um produto gerado por NBTs deve ou não ser classificado como OGM. Para essa consulta, a instituição desenvolvedora deve fornecer informações sobre o organismo original e o produto, incluindo os métodos utilizados para gerá-lo, e sua análise molecular. Em termos práticos, os produtos obtidos por mutação aleatória direcionada ao sítio que envolve a junção final não homóloga (NHEJ - *non-homologous end joining*) (considerado SDN1- *site direct nuclease 1*- sítio de mutação direta da nuclease) ou reparo homólogo direcionado (HRD - *homologous repair directed*) ao sítio que envolve poucos nucleotídeos (mutação tipo SDN2) e que atendem às condições estabelecidas na RN16, poderiam ser considerados como não OGM (Vieira et al., 2021; CTNBIO, 2022).

Em contraste, as inserções de transgenes direcionadas ao sítio (mutação tipo SDN3) serão provavelmente classificadas como OGM, de acordo com a RN16. Caso o produto seja designado como OGM, o desenvolvedor deverá cumprir com todos os requisitos de biossegurança e será aprovado somente após a avaliação de risco da CTNBio. O produto pode ser registrado usando os procedimentos existentes se não for classificado como OGM. A RN16 se aplica a todos os tipos de organismos, incluindo plantas, animais e microrganismos, em fase de pesquisa e/ou liberação comercial (Vieira et al., 2021; CTNBIO, 2022).

Redução do Impacto do Termo OGM

Segundo Tagliabue (2018), o termo “OGM” é prejudicial e uma bússola de direção enganosa. Muito da rejeição pública aos produtos e subprodutos GM deve-se ao argumento de que existe uma preocupação de que os OGMs representem uma tendência “*Frankenfoods*” que podem desencadear uma série de consequências não intencionais. No entanto, ao se questionar mais profundamente essa desaprovação, vê-se que a desinformação é o grande impulsionador desse cenário. Primeira pergunta: quem sente essa preocupação? Resposta: parte do público, mas certamente não a esmagadora maioria de cientistas envolvidos na pesquisa e obtenção desses produtos. Segunda questão: quem “alimenta” essa preocupação? Resposta: grupos antibioteclologia que lucram com esse alarmismo e a mídia, que tendenciosamente cria um falso entendimento ao comparar os resultados consistentes de pesquisa científica com alarmismo genérico e infundado. Terceira pergunta: qual é a base dessa preocupação? Resposta: não há, pois deriva de um grande erro epistemológico.

Ainda segundo esse mesmo autor, o erro que leva a controvérsias em relação aos OGMs é devido ao enfoque dado aos processos usados para desenvolver novas variedades de plantas, microrganismos e animais; quando se deveria limitar as considerações às avaliações de impacto ambiental e de segurança e às características fenotípicas dos produtos que são o resultado dos experimentos (Tagliabue, 2017). Na verdade, resultados insatisfatórios também podem derivar de técnicas tradicionais: inúmeras tentativas não OGM foram descartadas. De modo geral, aqueles que exigem testes intermináveis para OGMs, e que agora pressionam do mesmo modo para os produtos obtidos através das NBTs, são incapazes de apresentar qualquer fundamentação científica robusta para seus argumentos (Tagliabue, 2018).

No campo da biotecnologia agroalimentar, a objeção mais frequente é que os resultados das técnicas para se obter produtos GMs e, agora, os provenientes

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

das NBTs, são imprevisíveis; mas os críticos sempre falham em adicionar ao que a ciência já sabe: não existe uma bola de cristal para prever os resultados dessas tecnologias. No entanto, a ciência utiliza experiências anteriores, estudos, análises e conhecimento, e últimas avaliações a posteriori para determinar possíveis falhas e implicações, e também formular soluções. De fato, a imprevisibilidade é um padrão de qualquer processo de criação (Tagliabue, 2018).

Mesmo diante da inabilidade de se provar os “perigos” dos OGMs e agora dos produtos obtidos via NBTs, o obstáculo da opinião pública ainda existe e precisa ser superado. Mas há esperança para o futuro. Um estudo publicado em 2022, envolvendo 2.000 americanos, mostrou que a disposição dos indivíduos para comer e evitar propositalmente alimentos editados (GEFs - *gene-edited foods*) é impulsionada principalmente por seus valores sobre alimentação, ciência e tecnologia, confiança e conscientização sobre os alimentos. Surpreendentemente, as atitudes do público acerca das características tangíveis dos alimentos, tais como segurança, custo, sabor e aparência, não teve influência sobre as percepções de alimentos GM. Os dados mostraram ainda que as pessoas adeptas aos alimentos editados que expressam vontade de comer e não evitar GEFs, geralmente são pró-tecnologia, têm conhecimento sobre os alimentos GEFs e confiam no governo e na indústria para supervisionar os produtos à medida que entram no mercado. Por outro lado, os céticos da GEFs, aqueles que não querem comer e que planejam evitar GEFs, tendendo a ser guiados pela ética alimentar e crenças, são mais propensos a manter fortes crenças religiosas, e se identificam como politicamente mais conservadores (Cummings; Peters, 2022).

Esse grupo, na sua grande maioria, desconfia do governo e da indústria para supervisionar os produtos GEFs e, em vez disso, confiam mais nas organizações ambientais que podem representar melhor seus valores em relação à alimentação. Uma importante conclusão desse estudo foi a necessidade de melhor comunicação focada no consumidor. Os autores acreditam que à medida que as pessoas se familiarizarem com os processos de desenvolvimento, mais produtos entrarem no mercado comercial, e melhor comunicação atingindo públicos mais amplos seja implementada, a aceitação de produtos obtidos via NBTs pelo consumidor estará garantida (Cummings; Peters, 2022).

Outro estudo avaliou 59 manuscritos relacionados a atitudes dos consumidores e disposição de pagar por alimentos gerados a partir de NBTs, e apontou resultados similares. Os dados sugeriram que grandes segmentos de consumidores, mas não todos, estão dispostos a consumir e pagar por alimentos derivados das NBTs, especialmente se eles apresentarem características que os consumidores percebem como benéficas para a saúde humana e animal e o meio ambiente. Os autores

acreditam que devido à comercialização ainda limitada de alimentos derivados das NBTs, pesquisas constantes devem ser realizadas, à medida que novos alimentos estejam amplamente disponíveis no varejo (Beghin; Gustafson, 2021).

Nesse contexto, fica claro que a aceitação pública dos alimentos provenientes de NBTs deverá ficar menos conturbada, provavelmente, quando do lançamento de ampla gama de produtos derivados de OGMs no mercado consumidor, talvez pelo acesso à informação ou pela falta de alarde midiático. De qualquer maneira, esse cenário tornará os resultados dessas tecnologias mais rapidamente acessíveis ao grande público e aos produtores, contribuindo para a produção e segurança alimentar e economia dos países produtores.

Perspectivas

Considerando o cenário atual das pesquisas básicas e aplicadas no desenvolvimento de produtos via NBTs, para o futuro, pode-se vislumbrar um aumento de variedades sendo disponibilizadas aos produtores e mercadorias chegando nas prateleiras dos consumidores com diferentes características, abrindo-se um leque de possibilidades que vai desde um melhor manejo da lavoura e redução de custos, até produtos com maior valor agregado.

Fato importante que deve ser considerado para que no futuro esses produtos não encontrem a resistência e a rejeição que os OGMs tiveram no passado e ainda têm hoje, mas em menor intensidade, é como a “mensagem” será entregue pelos cientistas e como a mídia irá repassar estas informações ao grande público, para que um entendimento correto e claro sobre a ciência por trás desses produtos seja acessível, e compreendida sem medos e tabus, deixando claro a ciência, os pontos positivos, as vantagens desses produtos e garantindo sempre que a segurança das pessoas, dos animais e do meio ambiente foi observada.

Considerações Finais

A edição genômica via CRISPR/Cas veio como uma ferramenta de democratização da ciência, ampliando a possibilidade de startups e instituições públicas de pesquisa e universidades em competir na geração de produtos biotecnológicos com grandes multinacionais. Algumas dessas conquistas podem ser observadas no Brasil com as liberações de cultivares de soja e variedades de cana-de-açúcar por diferentes unidades da Embrapa, por exemplo.

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

Embora a tecnologia seja inovadora, de menor custo e altamente precisa ainda existem divergências nas legislações pelo mundo. Alguns países são mais rigorosos e relutantes para aceitar a liberação de produtos editados de forma diferente daqueles obtidos via transgenia.

Por se tratar de uma tecnologia bastante recente e envolver discussões sobre sua regulamentação, apenas dois produtos estão de fato no mercado, são eles um tomate japonês e uma soja americana. No entanto, espera-se que muitos outros produtos apareçam no futuro e, por fim, cheguem ao mercado.

Referências

BEGHIN, J. C.; GUSTAFSON, C. R. Consumer valuation of and attitudes towards novel foods produced with new plant engineering techniques: a review. **Sustainability**, v. 13, n. 20, p. 11348, Oct. 2021. DOI: <https://doi.org/10.3390/su132011348>.

CHEE, F. Y. **In-vitro plant gene editing technique excluded from GMO rules, EU court says**. Disponível em: <https://www.reuters.com/world/europe/in-vitro-plant-gene-editing-technique-excluded-gmo-rules-eu-court-says-2023-02-07/>. Acesso em: 20 maio 2023.

CHOUHIKA, A.; PERRIN, A.; DUJON, B.; NICOLAS, J. F. Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of *Saccharomyces cerevisiae*. **Molecular and Cellular Biology**, v. 15, n. 4, p. 1968-1973, Apr. 1995. DOI: <https://doi.org/10.1128/MCB.15.4.1968>.

CTNBIO. Comissão Técnica Nacional de Biossegurança. **Produtos avaliados - técnicas inovadoras de melhoramento de precisão**. 2022. Disponível em: <http://ctnbio.mctic.gov.br/documents/566529/2304555/Tabela+TIMP/8c4a7218-f810-405b-94bf-a352d849f3dc?version=1.1>. Acesso em: 16 jun. 2022.

CUMMINGS, C.; PETERS, D. J. Who trusts in gene-edited foods?: analysis of a representative survey predicting willingness to eat-and purposeful avoidance of gene edited foods in the United States. **Frontiers in Food Science and Technology**, v. 2, 858277, June 2022. DOI: <https://doi.org/10.3389/frfst.2022.858277>.

DEMAREST, Z. L.; COFFMAN, A.; BALTES, N. J.; STODDARD, T. J.; CLASEN, B. M.; LUO, S.; RETTERATH, A.; YABANDITH, A.; GAMO, M.E.; BISSEN, J.; MATHIS, L.; VOYTAS, D. F.; ZHANG, F. Direct stacking of sequence-specific nuclease-induced mutations to produce high oleic and low linolenic soybean oil. **BMC Plant Biology**, v. 16, 225, Oct. 2016. DOI: <https://doi.org/10.1186/s12870-016-0906-1>.

DIMA, O.; HEYVAERT, Y.; INZÉ, D. Interactive database of genome editing applications in crops and future policy making in the European Union. **Trends in Plant Science**, v. 27, n. 8, p. 746-748, Aug. 2022. DOI: <https://doi.org/10.1016/j.tplants.2022.05.002>.

DOUDNA, J.A.; CHARPENTIER, E. The new frontier of genome engineering with CRISPR-Cas9. **Science**, v. 346, n. 6213, 1258096, Nov. 2014. DOI: <https://doi.org/10.1126/science.1258096>.

EU-SAGE. European Sustainable Agriculture through Genome Editing. **The EC study on new genomic techniques**. Disponível em: <https://www.eu-sage.eu/>. Acesso em: 27 abr. 2023.

EUROPEAN PARLIAMENT. **Genome-edited crops and 21st century food system challenges**. 2022. Disponível em: [https://www.europarl.europa.eu/thinktank/en/document/EPRS_IDA\(2022\)690194](https://www.europarl.europa.eu/thinktank/en/document/EPRS_IDA(2022)690194). Acesso em: 20 abr. 2023.

GAJ, T.; SIRK, S. J.; SHUI, S. L.; LIU, J. Genome-editing technologies: principles and applications. **Cold Spring Harbor Perspectives in Biology**, v. 8, n. 12, a023754, Dec. 2016. DOI: <https://doi.org/10.1101/cshperspect.a023754>.

GLOBAL GENE EDITING REGULATION TRACKER. **United States: crops/food**: gene-edited crops are regulated as conventional plants, so they don't face the heavier restriction of GMOs. Disponível em: <https://crispr-gene-editing-regs-tracker.geneticliteracyproject.org/united-states-crops-food/>. Acesso em: 14 dez. 2022.

GRAIN. **GM waxy maize**: the gene edited Trojan Horse is moving through the gates. 2021. Disponível em: <https://grain.org/en/article/6640-gm-waxy-maize-the-gene-edited-trojan-horse-is-moving-through-the-gates>. Acesso em: 16 abr. 2022.

ISAAA. International Service for the Acquisition of Agri-biotech Applications. **China drafts new rules for gene-edited crops**. 2022. Disponível em: <https://www.isaaa.org/kc/cropbiotechupdate/article/default.asp?ID=19214>. Acesso em: 1 jan. 2023.

JAPAN embraces CRISPR-edited fish. **Nature Biotechnology**, v. 40, p. 10, Jan. 2022. DOI: <https://doi.org/10.1038/s41587-021-01197-8>.

JINEK, M.; CHYLINSKI, K.; FONFARA, I.; HAUER, M.; DOUDNA, J. A.; CHARPENTIER, E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. **Science**, v. 337, n. 6096, p. 816-821, Aug. 2012. DOI: <https://doi.org/10.1126/science.1225829>.

JORASCH, P. Potential, challenges, and threats for the application of new breeding techniques by the private plant breeding sector in the EU. **Frontiers in Plant Science**, v. 11, 582011, Sept. 2020. DOI: <https://www.doi.org/10.3389/fpls.2020.582011>.

KRINKE, C.; MEUNIER E. **Objective 2030**: deregulate most GMOs in Europe? Disponível em: <https://www.infogm.org/7512-objective-2030-deregulate-most-gmos-in-europe?lang=fr>. Acesso em: 16 maio 2022.

MENZ, J.; MODRZEJEWSKI, D.; HARTUNG, F.; WILHELM, R.; SPRINK, T. Genome edited crops touch the market: a view on the global development and regulatory environment. **Frontiers in Plant Science**, v. 11, 586027, Oct. 2020. DOI: <https://doi.org/10.3389/fpls.2020.586027>.

METJE-SPRINK, J.; SPRINK, T.; HARTUNG, F. Genome-edited plants in the field. **Current Opinion in Biotechnology**, v. 61, p. 1-6, Feb. 2020. DOI: <https://doi.org/10.1016/j.copbio.2019.08.007>.

NONAKA, S.; ARAI, C.; TAKAYAMA, M.; MATSUKURA, C.; EZURA, H. Efficient increase of γ -aminobutyric acid (GABA) content in tomato fruits by targeted mutagenesis. **Scientific Reports**, v. 7, 7057, Aug. 2017. DOI: <https://doi.org/10.1038/s41598-017-06400-y>.

PARISI, C.; RODRÍGUEZ-CEREZO, E. **Current and future market applications of new genomic techniques**. Luxembourg: Publications Office of the European Union, 2021. DOI: <https://doi.org/10.2760/02472>.

ROUET, P.; SMIH, F.; JASIN, M. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. **Molecular and Cellular Biology**, v. 14, n. 12, p. 8096-8106, Dec. 1994.

SÁNCHEZ-LEÓN, S.; GIL-HUMANES, J.; OZUNA, C. V.; GIMÉNEZ, M. J.; SOUSA, C.; VOYTAS, D. F.; BARRO, F. Low-gluten, nontransgenic wheat engineered with CRISPR/Cas9. **Plant Biotechnology Journal**, v. 16, n. 4, p. 902-910, Apr. 2018. DOI: <https://doi.org/10.1111/pbi.12837>.

SMYTH, S. J. Regulation of genome editing in plant biotechnology: Canada. In: DEDERER, H.-G.; HAMBURGER, D. (ed.). **Regulation of genome editing in plant biotechnology**: a comparative analysis of regulatory frameworks of selected countries and the EU. Cham: Springer, 2019. p. 111-135. DOI: https://doi.org/10.1007/978-3-030-17119-3_4.

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

SOREK, R.; LAWRENCE, C. M.; WIEDENHEFT, B. CRISPR-mediated adaptive immune systems in bacteria and archaea. **Annual Review of Biochemistry**, v. 82, p. 237-266, June 2013. DOI: <https://doi.org/10.1146/annurev-biochem-072911-172315>.

TAGLIABUE, G. Product, not process! Explaining a basic concept in agricultural biotechnologies and food safety. **Life Sciences, Society and Policy**, v. 13, n. 3, p. 1-9, Dec. 2017. DOI: <https://doi.org/10.1186/s40504-017-0048-8>.

TAGLIABUE, G. Product, not process. Reject the “GMO” fallacy, in terms of both safety concerns and socioeconomic issues: comments. **Geographical Review**, v. 108, n. 1, p. e1-e5, Jan. 2018. DOI: <https://doi.org/10.1111/gere.12293>.

VIEIRA, L. R.; FREITAS, N. C.; JUSTEN, F.; MIRANDA, V. J.; GARCIA, B. O.; NEPOMUCENO, A. L.; FUGANTI-PAGLIARINI, R.; FELIPE, M. S. S.; MOLINARI, H. B. C.; VELINI, E. D.; PINTO, E. R. C.; DAGLI, M. L. Z.; ANDRADE, G.; FERNANDES, P. M. B.; MERTZ-HENNING, L. M.; KOBAYASHI, A. K. Regulatory framework of genome editing in Brazil and worldwide. In: MOLINARI, H. B. C.; VIEIRA, L. R.; SILVA, N. V.; PRADO, G. S.; LOPES FILHO, J. F. (ed.). **CRISPR technology in plant genome editing: biotechnology applied to agriculture**. Brasília, DF: Embrapa, 2021. Cap. 5, p. 169-195. <https://ainfo.cnptia.embrapa.br/digital/bitstream/item/223607/1/Regulatory-framework-of-genome-CAP-5.pdf>.

WALTZ, E. GABA-enriched tomato is first CRISPR-edited food to enter market. **Nature Biotechnology**, v. 40, p. 9-11, Jan. 2022. DOI: <https://doi.org/10.1038/d41587-021-00026-2>.

WESSBERG, N.; LEHTINEN, S.; RITALA, A.; HÄKKINEN, S. T.; VILKKI, J.; SCHULMAN, A.; KORHONEN, S. **Utilisation of new genome editing techniques in Finland**. 2021. Disponível em: https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/163143/VNTEAS_2021_39.pdf?sequence=1. Acesso em: 20 fev. 2023.

WIEDENHEFT, B.; STERNBERG, S. H.; DOUDNA, J. A. RNA-guided genetic silencing systems in bacteria and archaea. **Nature**, v. 482, n. 7385, p. 331-338, Feb. 2012. DOI: <https://doi.org/10.1038/nature10886>.

Este livro apresenta 12 capítulos que fazem parte do que consideramos atualmente como "Melhoramento de Precisão". Esta obra foi inspirada na pioneira disciplina de mesmo nome do Programa de Pós-Graduação em Genética e Melhoramento de Plantas (PPGGMP) da UFG e foi concebida para atender tanto estudantes quanto profissionais atuantes na área de melhoramento genético. Esperamos contribuir para que os leitores possam adquirir e aplicar esse conhecimento em seu cotidiano, melhorando, assim, os resultados de seus trabalhos. Para atuar no campo do melhoramento de precisão, o profissional deve adquirir conhecimentos e habilidades relacionados à genômica, fenômica, ambientômica, ciência de dados, inteligência artificial, biotecnologia e dominar ferramentas avançadas de melhoramento genético, áreas contempladas neste livro. Além disso, estudantes que se familiarizarem com alguma dessas áreas podem, ao longo do tempo, se especializar e, futuramente, integrar uma nova geração de profissionais que contribuirão para o desenvolvimento do nosso país. Desejamos a todos uma ótima leitura e que obtenham excelentes resultados!

Os Editores

PARCERIA

PPGGMP-UFG
PROGRAMA DE PÓS-GRADUAÇÃO EM
GENÉTICA E MELHORAMENTO DE PLANTAS

EA
ESCOLA DE
AGRONOMIA



UFG
UNIVERSIDADE
FEDERAL DE GOIÁS

APOIO



MINISTÉRIO DA
AGRICULTURA E
PECUÁRIA



CGPE 18166