

# Modelos computacionais para a simulação do programa nacional de melhoramento de trigo

A modelação computacional aplicada ao melhoramento do trigo permite testar a viabilidade da introdução de novas metodologias no programa convencional, como a **Seleção Genômica**, através das projeções a médio/longo prazo dos ganhos de produção das novas variedades nos diferentes cenários.

Sílvia C. Alves . INIAV, I.P.



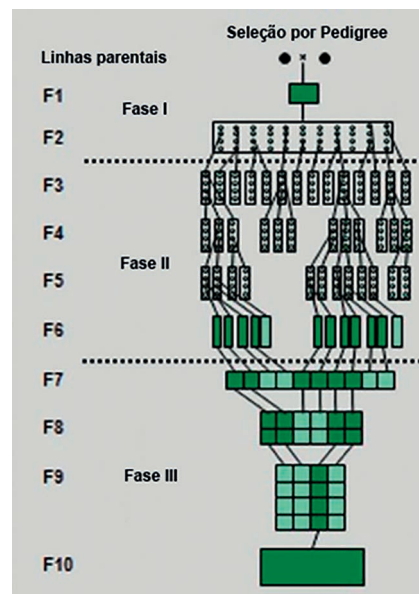
O melhoramento de plantas explora, através da seleção artificial, a variabilidade genética existente dentro de uma população – cruzando ao longo de várias gerações os melhores indivíduos (i.e., aqueles que são portadores de características desejadas do ponto de vista agronómico), e impedindo a reprodução entre os restantes. Como consequência desta metodologia, a frequência dos genes responsáveis pela transmissão dessas qualidades aumenta a cada geração, fazendo com que a semelhança genética entre os indivíduos seja cada vez maior e, no limite, tenda para a homozigocidade. A capacidade de aumentar a eficiência de um programa de melhoramento é, portanto, diretamente proporcional à variabilidade genética disponível (condicionada pela interação genética-ambiente) e à frequência de genes superiores nas populações segregantes, de onde se selecionam progenitores.

O processo de seleção e melhoramento convencional, contudo, não tem só aspetos positivos. Para que se possa obter uma nova variedade melhorada é necessário emasculiar e polinizar manualmente dezenas de plantas selecionadas. Posteriormente, é preciso colher, semear, observar e registar o comportamento agronómico de milhares de indivíduos, em diferentes contextos ambientais, ao longo de cerca de uma dezena de gerações (Figura 1).

Ainda assim, ao cruzar plantas geneticamente distintas, não são só as características desejadas que são propagadas à descendên-

cia. Genes diferentes que induzam comportamentos agronómicos opostos, mas que se encontrem localizados no mesmo cromossoma, e com uma forte **ligação genética**, ou seja, próximos fisicamente, são normalmente herdados em bloco. Isto acontece porque esta relação de proximidade entre genes é extremamente difícil de quebrar por recombinação espontânea/natural e, como tal, quando parte desses genes definem características indesejáveis ou têm o poder **pleiotrópico** de neutralizar a expressão de outros genes importantes, o ganho genético efetivo obtido através do melhoramento convencional é afetado. Este fenómeno é particularmente difícil de circunscrever quando associado ao melhoramento de um atributo sob regulação poligénica, e agrava-se com o aumento da **ploidia** (n, número de cromossomas homólogos característico de uma espécie). Como exemplos paradigmáticos desta condição destacam-se o trigo-duro (*T. turgidum* var. durum) e o trigo-mole (*T. aestivum*), duas espécies cuja base genética é constituída por dois (*Tt*:  $n=7_A+7_B$ ) e três (*Ta*:  $n=7_A+7_B+7_D$ ) subgenomas diplóides distintos, ou seja, para cada um dos seus cerca de 36 mil genes existem, respetivamente, quatro e seis cópias potencialmente diferentes [2].

Assim se percebe como é difícil e moroso fazer melhoramento genético de um atributo poligénico, como, por exemplo, a produtividade, e por que razão os melhoradores precisam de testar milhares de plantas, ao longo de vários anos e em condições ambientais diferentes, para que se possa obter uma nova variedade melhorada, porque: 1) para além de ser impossível reunir, com um único cruzamento, toda a superioridade genética de dois progenitores quando cada um deles apenas contribui para a fecundação com me-



**Figura 1** – Esquema do programa de melhoramento convencional – Seleção por Pedigree.

Fase I – Criação de variabilidade através de novos cruzamentos parentais (F1) e seleção de segregantes (F2);  
Fase II – Desenvolvimento e seleção de potenciais linhas varietais;  
Fase III – Ensaios de rendimento em diferentes condições ambientais.  
Adaptado de [1]

tade dos seus cromossomas homólogos; 2) é, também, preciso garantir que a qualidade do fundo genético de um progenitor elite não é prejudicada com a introdução de variantes alélicas indesejadas do outro progenitor.

O problema é de tal modo complexo que tem exigido da ciência e da comunidade científica um esforço constante para encontrar soluções inovadoras, de baixo custo, e numa conjuntura climática em mudança. No entanto, a introdução de modificações numa estrutura complexa como um programa de melhoramento, mesmo que feitas com a intenção de potenciar a sua eficiência, acar-

reta riscos, custos e tempo.

Neste contexto, a **simulação computacional** – um conjunto de instruções (modelos matemáticos), definidos por um programador e executadas por uma unidade computacional – oferece uma série de vantagens ao permitir testar, *in silico*, inúmeros cenários teóricos diferentes sem que haja a necessidade de fazer investimentos financeiros avultados na aquisição ou realocação de recursos, sem comprometer o normal funcionamento da estrutura já implementada, e a funcionar em velocidade de cruzeiro, e sem ter de aguardar anos pelos resultados de cada um dos cenários em teste para poder tomar decisões que levem ao incremento do rendimento do programa. Mais, no campo existe sempre a possibilidade de ocorrência de fenómenos meteorológicos extremos, incêndios difíceis de controlar, ou uma proliferação atípica de pragas/doenças que podem facilmente levar à devastação de parte significativa ou da totalidade dos ensaios em curso. Nesta situação, seria impossível fazer uma análise integral das condições em teste e retirar as conclusões respetivas, o que implicaria recomençar tudo de novo no ano seguinte. Assim, por meio de simulações computacionais é possível obter uma resposta prévia e independente, que analisa a validade intrínseca de cada estratégia, sem que o resultado da sua análise esteja subordinado à incerteza das condições em campo. Posteriormente, com base nessa informação, o melhorador pode ponderar e decidir qual das abordagens é mais útil implementar no terreno – poupando, assim, tempo e recursos. As simulações computacionais são, por-

tanto, ótimas ferramentas para descrever e comparar cenários teóricos. Contudo, para que possam recriar com precisão diferentes cenários reais, o modelo-base tem de obedecer a um conjunto de requisitos. Em primeiro lugar, é essencial que o modelo esteja calibrado para as condições em estudo. De seguida, este deverá ser submetido a uma fase de verificação que garanta que está a operar como o esperado. Posteriormente, quando o modelo já se encontra operacional, é preciso validá-lo. A validação passa pela simulação de vários casos conhecidos e pela comparação dos resultados assim obtidos com observações reais. Quanto maior for a semelhança entre os dados simulados e os dados reais, para diferentes cenários, mais preciso é o modelo. Durante a fase de validação, poderá ser necessário fazer vários ajustes na parametrização do modelo, para que os resultados simulados se aproximem cada vez mais dos valores reais. Este processo iterativo de ajustamento é fundamental para a produção de resultados realistas. Se estes passos não forem meticulosamente executados, é muito provável que a simulação produza resultados imprecisos e que as suas previsões sejam inúteis. Por fim, há que avaliar a reprodutibilidade dos resultados, ou seja, é preciso testar se a simulação, tal como foi delineada, produz resultados uniformes, independente das condições iniciais e do número de vezes que é executada. Tal é conseguido através do recurso a funções geradoras de números aleatórios, que produzem diferentes condições iniciais para um ou mais parâmetros, de cada vez que a simulação é executada. O *Burn-in* é um termo

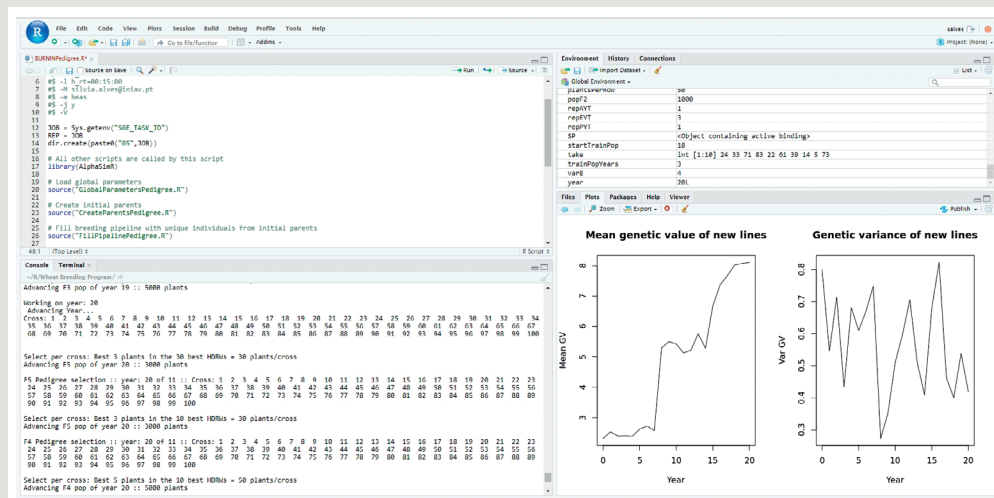
muito utilizado em computação para descrever as primeiras iterações de uma simulação, quando são usados geradores de números aleatórios. Enquanto não for ultrapassada a fase de *Burn-in*, os resultados da simulação estão ainda sob grande influência das condições aleatórias iniciais e, por isso, são altamente variáveis. À medida que as iterações se vão sucedendo, o modelo afasta-se cada vez mais das condições iniciais e começa a convergir para um valor de previsão uniforme (Figura 2). É nesta altura que é possível avaliar a reprodutibilidade do modelo.

Consoante a complexidade do modelo e a capacidade de processamento disponível, as simulações podem demorar alguns segundos, minutos, horas ou dias a correr, no entanto, desde que cumpridas todas as etapas de calibração, verificação e validação, a simulação computacional pode ser um grande aliado do melhorador.

Por esta razão, o INIAV está a implementar a construção de um modelo de simulação estocástica [3, 4, 5] do programa de melhoramento de trigo, sediado no polo de Elvas (antiga Estação Nacional de Melhoramento de Plantas), no âmbito do projeto FASTBREED (ALT20-03-0145-FEDER-000018) (Figura 2). Através desta nova ferramenta bioinformática, ir-se-á testar a introdução da Seleção Genómica em diferentes fases do programa de melhoramento e aferir qual a melhor estratégia a adotar, para otimizar a afetação de recursos, maximizar o ganho genético a médio/longo prazo e acelerar o desenvolvimento de novas variedades melhoradas. Esta será a estratégia do melhorador do futuro, que começa agora, ao serviço da garantia do alimento e da sustentabilidade. 🌾

## Bibliografia

- [1] Osei, M. *et al.* (2014). *Plant Breeding: A Tool for Achieving Food Sufficiency*. 10.1007/978-3-319-06904-3\_11.
- [2] Borrill, P. & Adamski, N. *et al.* (2015). Genomics as the key to unlocking the polyploid potential of wheat. *New Phytol.*, **208**:1008-1022.
- [3] R Core Team (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- [4] Chris Gaynor (2018). *AlphaSimR: Breeding Program Simulations*. <https://alphagenes.roslin.ed.ac.uk/wp/software/alphasimr/>, <https://bitbucket.org/hickey-johnnteam/alphasimr>.
- [5] Faux, A. *et al.* (2016). AlphaSim: Software for Breeding Program Simulation; *Plant Genome*, **9**.



**Figura 2** – Simulação do programa convencional de melhoramento de trigo do INIAV (Elvas). Fase de testes – *Burn-in*: Paineis superior esquerdo – ficheiro (*script*) central que compila todas as instruções necessárias à simulação. Paineis superior direito – conjunto de parâmetros da simulação. Paineis inferior esquerdo – monitor de controlo das etapas de simulação. Paineis inferior direito – resultados gráficos: projeções a 20 anos para os ganhos genéticos médios das novas variedades (esquerda) e a variância genética das novas variedades (direita). *Software* utilizado: R [3], AlphaSimR [4, 5]