Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# CrossingTools: Scalable cross evaluation and mating-plan optimisation for multiple traits

**Sven E. Weber**

Sven.E.Weber@agrar.uni-giessen.de

Justus-Liebig-Universität Gießen

**Dominic L. Waters**

University of New England

**Christian R. Werner**

Consultative Group of International Agricultural Research (CGIAR)

**Daniel J. Tolhurst**

University of Edinburgh

**Research Article**

**Keywords:** Index selection, Multi-trait selection, Genetic algorithm, Usefulness criterion, Superior progeny value

**Posted Date:** November 19th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-11431426/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Additional Declarations:** Competing interest reported. SW serves as Social Media Editor for *Theoretical and Applied Genetics*. The remaining authors declare no conflicts of interest.

# CrossingTools: Scalable cross evaluation and mating-plan optimisation for multiple traits

S. E. Weber[1*], D. L. Waters[2], C. R. Werner[3,4], D. J. Tolhurst[5]

[1*] Department of Plant Breeding, IFZ Research Centre for Biosystems, Land Use and Nutrition, Justus Liebig University Giessen, Giessen,Germany .

[2] School of Environmental and Rural Science, University of New England, Armidale, NSW,Australia .

[3] Breeding for Tomorrow (B4T), Consultative Group of International Agricultural Research (CGIAR), Texcoco,Mexico .

[4] International Maize and Wheat Improvement Center (CIMMYT), Texcoco,Mexico .

[5] The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, United Kingdom .

*Corresponding author(s). E-mail(s): Sven.Weber@agrar.uni-giessen.de;

**Abstract**

Selection and mate allocation shape performance and diversity in future breeding populations. Although numerous criteria for evaluating and selecting crosses exist, few tools offer a unified, scalable framework for different breeding schemes, multi-trait objectives, and diversity management. We present CrossingTools, an open source R package for predicting cross performance and optimising crossing plans using marker genotypes and genetic effects. Crosses can be evaluated based on their expected progeny mean, optimal haploid value, or superior progeny value using segregation (co)variance models for inbred and heterozygous parents with additive and dominance effects. Multi-trait objectives can be defined using economic weight or desired gains indices, while crossing plans can be optimised to balance predicted performance and genomic diversity through optimal cross selection. Efficient C++ subroutines enable large-scale cross evaluation, allowing the screening of entire breeding populations. To demonstrate its application, we simulated a line breeding programme with correlated traits and found that, under genomic selection, superior progeny value and optimal haploid value improved upper-tail progeny performance more strongly than population mean gain and

preserved more genetic variance than selection based on expected cross means. The criteria also affected genetic correlations differently as superior progeny value tended to shift correlations in directions favourable to the selection index, optimal haploid value largely maintained correlations, and selection on expected cross means often produced less favourable changes. Optimal cross selection reduced variance depletion at the cost of lower short-term gain. CrossingTools thus offers a framework for integrating progeny prediction, multi-trait selection, and genetic diversity management in plant breeding.

**Keywords:** Index selection, Multi-trait selection, Genetic algorithm, Usefulness criterion, Superior progeny value

**Key message**

CrossingTools provides a scalable and flexible framework to optimise crossing decisions for improving genetic gain and superior progeny performance while maintaining genetic variation across multiple traits.

# 1 Introduction

Plant breeding programmes aim to release improved varieties that primarily combine high yield, quality, and disease resistance across their target population of environments. Each breeding cycle involves selecting superior parents and crossing them to create the next generation of candidate varieties. While advances in genomic prediction have greatly increased the accuracy of identifying favourable parents, their allocation to optimal mating combinations remains a critical, often underutilised, step to achieve sustainable genetic improvement. The objective of this paper is to introduce the R package CrossingTools, which is a scalable and flexible application for optimising crossing decisions within breeding programmes.

Genomic selection (GS) has transformed breeding through the use of genomic estimated breeding values (GEBVs) derived from genome-wide marker data (Meuwissen et al. 2001; Bernardo 1994). GS has increased genetic gain by improving selection accuracy and shortening cycle times compared to traditional approaches based on pedigree and phenotypic data alone (Gaynor et al. 2017). In this context, parent selection is predominantly achieved by ranking individuals based on their GEBVs, while the genetic potential of their progeny is often overlooked. This provides motivation to develop practical tools which utilise GEBVs for calculating the potential of prospective crosses (Zhong and Jannink 2007; Osthushenrich et al. 2017; Allier et al. 2019; Lehermeier et al. 2017).

Mate allocation (Kinghorn 1987), often referred to as genomic mating when using marker data (Akdemir and Sánchez 2016), provides a complementary framework to selection by considering the genetic potential of a cross rather than that of the individual parents. The simplest measure of cross potential is the expected mean performance of the progeny, for example the average of the parental GEBVs. Selecting crosses with the highest expected mean performance maximises the expected response in the population mean. However, maximising the population mean does not necessarily maximise

the probability of creating extreme, high-performing individuals within that population through recombination and segregation. In plant breeding, commercial success is ultimately realised through the selection and release of superior varieties, which typically originate from the upper tail of the progeny distribution. Consequently, a cross with a slightly lower expected mean may still be preferred if it generates greater genetic variation among progeny and therefore has a higher probability of producing exceptional individuals. This provides the motivation to evaluate crosses not only by their expected mean performance, but also the distribution of genetic values expected within a family.

Following the classical definition of the usefulness criterion (Schnell and Utz 1975), the value of a cross can be expressed as the expected progeny mean together with the expected gain arising from genetic variation among progeny. The second component, often referred to as the segregation or family variance, reflects the opportunity for superior individuals to emerge through Mendelian sampling and therefore contributes to both short-term and long-term genetic gain (Allier et al. 2019). The availability of genome-wide marker data has made it possible to predict family means and family variances from parental genotypes while accounting for linkage, linkage disequilibrium, and recombination (Zhong and Jannink 2007; Lehermeier et al. 2017; Osthushenrich et al. 2017). Building on this, several genomic metrics have been proposed to evaluate the potential of a cross. For example, the superior progeny value (SPV) is a direct extension of the classical usefulness criterion to incorporate GEBVs (Zhong and Jannink 2007), with further extensions incorporating dominance effects and cross-specific dominance variances (Wolfe et al. 2021). Similarly, the optimal haploid value (OHV) quantifies the theoretical upper limit of progeny performance attainable through recombination of the parental haplotypes (Daetwyler et al. 2015; Cole and VanRaden 2011). Together, these approaches provide a framework for predicting the distribution of genetic values within a family and for identifying crosses with the greatest potential to generate superior offspring.

Although a number of software tools for mate allocation already exist (Peixoto et al. 2025; Mohammadi et al. 2015; Fernández-González et al. 2026), many of these have practical and computational limitations. For example, many approaches focus on a pre-selected subset of potential crosses to reduce the computational burden of evaluating all pairwise combinations (Lehermeier et al. 2017; Niehoff et al. 2024a; Peixoto et al. 2025). While this drastically reduces computational burden, it can overlook crosses that combine complementary alleles with many shared loci with little effect. There is also limited support for multi-trait breeding objectives, despite breeding programmes routinely targeting improvement across multiple correlated traits and environments. Although theoretical frameworks for multi-trait optimisation have long been established (Smith 1936; Hazel and Lush 1942; Pesek and Baker 1969), existing tools often rely on predefined selection indices rather than providing integrated functionality for their construction and use in cross evaluation and mate allocation.

Different breeding schemes, such as hybrid, line, and clonal breeding, require specific measures of cross performance. For example, in clonal breeding, within-family dominance variance can help identify superior crosses for variety development (Wolfe et al. 2021), whereas classical hybrid or line breeding typically involves the creation of

inbred lines prior to selection. Consequently, the appropriate recombination functions differ between breeding schemes (Lehermeier et al. 2017; Osthushenrich et al. 2017; Wolfe et al. 2021; Allier et al. 2019), yet these distinctions are not implemented in most available tools. Lastly, as widely recognised in animal breeding (Meuwissen 1997), managing inbreeding and maintaining genetic diversity are essential for preserving genetic variance and ensuring sustained long-term genetic gain. These considerations should also be incorporated into plant breeding optimisation frameworks (Gorjanc et al. 2018; Akdemir and Sánchez 2016).

The R package CrossingTools was developed to address current limitations by providing a scalable, flexible, and unified framework for optimising crossing decisions within hybrid, line, and clonal breeding schemes. CrossingTools enables exhaustive evaluation of all potential parent combinations, supports additive and dominance genetic effects, and extends mate allocation to multi-trait optimisation through index selection. The computer algorithm is implemented in R with C++ subroutines, combining a user-friendly scripting environment with high computational performance suitable for large populations and dense genomic datasets.

CrossingTools bridges the gap between theoretical mate allocation approaches and their practical application in modern breeding. Overall, the package provides a comprehensive suite of tools and resources to:

- Predict and evaluate cross performance using multiple criteria, including the expected mean, SPV, and OHV.
- Perform multi-trait optimisation using a Smith-Hazel index with economic weights or a Desired Gains index with predefined targets for individual traits.
- Design crosses which balance genetic gain and long-term genetic diversity using optimal cross selection (OCS), supporting sustainable improvement over multiple breeding cycles.

This paper presents the theoretical framework embedded within CrossingTools and showcases its utility for different selection criteria and optimisation approaches using a simulated line breeding programme with GS.

## 2 Materials and methods

This section presents the theoretical framework embedded within the R package CrossingTools, focusing on the prediction of progeny means and segregation (co)variances that form the foundation for evaluating cross potential. The framework is demonstrated here using a simulated working example, with supporting R code provided in the Supplementary Material.

The methods consist of three main parts:

1. General theory, which reviews the classical UC, SPV, and OHV, followed by the implemented procedures to manage genetic diversity in optimal cross selection (OCS).

2. Index selection, which extends the classical performance measures to multi-trait optimisation using weighted and Desired Gains indices.
3. Estimation, which derives expressions for all cross performance measures with estimation under the linear mixed model.

A stochastic simulation is then used to illustrate the different mate allocation strategies and evaluate their effects on genetic gain, diversity, and cross selection decisions under a genomic selection based line breeding programme

## 2.1 General theory

It is important to first distinguish between the two main components of a breeding programme: population improvement and product development. Population improvement involves the selection and crossing of superior parents for the purpose of moving the population mean of key traits in a desired direction over time. Product development then involves the extraction of genotypes from the breeding population for commercialisation. This process can be passive, where genotypes already existing within the population are selected for release, or active, where specific crosses are designed to create a desired product (Waters et al. 2026). While the focus of this paper is on population improvement crosses, the framework is general and can be applied to product development crosses, as noted throughout.

We consider three criteria for evaluating cross performance: cross expectation, defined as the mean genetic value of the progeny; usefulness criterion and SPV (Schnell and Utz 1975; Zhong and Jannink 2007), defined as the mean genetic value of progeny after within-family selection; and OHV (Daetwyler et al. 2015; Cole and VanRaden 2011), defined as the genetic value of the theoretically best progeny in a family of infinite size. In the following sections, these criteria are presented assuming an additive genetic model with known quantities. Estimation is then considered in Section 2.3.

### 2.1.1 Cross expectation

Under an additive genetic model, the expected genetic value of an individual is defined by its breeding value (BV). Consequently, the expected genetic value of the progeny from a cross between two individuals equals the average of the parental BVs. For a cross between parents $i$ and $j$, the expected progeny mean is therefore given by the mid-parent value:

$$\mu_{i\times j} = \frac{g_i + g_j}{2} \tag{1}$$

where $g_i$ and $g_j$ are the BVs of the two parents. This result follows directly from the definition of BVs as the additive genetic effects transmitted to the next generation, of which each parent contributes half to the expectation (Falconer and Mackay 2009).

Assuming the effects of all $r$ quantitative trait loci (QTL) are known, the BV of individual $i$ can be written as:

$$g_i = \sum_{k=1}^{r}(m_{ik} - 2p_k)\alpha_k \tag{2}$$

where $\alpha_k$ is the average allele-substitution effect of QTL $k$, $p_k$ is the frequency of the counted allele at locus $k$ in the reference population, and $m_{ik}$ is the allele count of individual $i$, such that $m_{ik} \in \{0, 1, 2\}$.

When dominance effects are present, breeding values alone do not capture the expected performance of specific crosses because complementary parental alleles may generate dominance deviations. Crosses can therefore instead be evaluated based on the expected total genetic value of their $F_1$ progeny (Werner et al. 2023). For a cross between individuals $i$ and $j$, this expectation is

$$\begin{aligned} \mu_{i\times j} = &\sum_{k=1}^{r} \left( \frac{m_{i,k} + m_{j,k}}{2} - 2p_k \right) \alpha_k \\ &- \frac{1}{2} \sum_{k=1}^{r} \left( m_{i,k} - 2p_k \right) \left( m_{j,k} - 2p_k \right) d_k, \end{aligned} \tag{3}$$

where $d_k$ is the statistical dominance effect at locus $k$. The first term represents the expected breeding value of the $F_1$ progeny, with the expected allele dosage $(m_{i,k} + m_{j,k})/2$ centred by $2p_k$. The second term represents the expected dominance deviation under the same statistical parameterisation. Thus, Eq. 3 gives the expected total genetic value as the sum of the expected breeding value and dominance deviation.

### 2.1.2 Usefulness criterion and superior progeny value

The UC evaluates the potential of a cross based on the expected mean of selected progeny after within-family selection (Schnell and Utz 1975). The UC for a cross between parents $i$ and $j$ is defined as

$$\mathcal{U}_{i\times j} = \mu_{\mathrm{p},i\times j} + \iota h_{i\times j} \sigma_{\mathrm{p},i\times j}, \tag{4}$$

where $\mu_{\mathrm{p},i\times j}$ is the expected progeny mean prior to selection, $\iota$ is the standardised selection differential corresponding to the within-family selection intensity, $h_{i\times j}$ is the square root of the narrow-sense heritability, and $\sigma_{\mathrm{p},i\times j}$ is the phenotypic standard deviation among progeny of cross $i \times j$.

Originally, the UC was formulated using phenotypic means and variances of parents and their progeny (Schnell and Utz 1975). In this formulation, the term $\iota\, h_{i\times j}\, \sigma_{\mathrm{p},i\times j}$ represents the expected response to selection within the family, where heritability links the phenotypic variance among progeny to the underlying additive genetic variance. Specifically, if $\sigma^2_{a_{i\times j}}$ denotes the additive genetic variance within cross $i \times j$, then $\sigma_{\mathrm{a},i\times j} = h_{i\times j} \sigma_{\mathrm{p},i\times j}$.

Consequently, the UC can be interpreted as combining the expected cross mean with the additive genetic standard deviation generated by segregation in the progeny. A general framework for deriving the segregation (co)variances is provided below and in Appendix A.

The SPV is analogous to the UC, but evaluates cross potential in terms of genotypic rather than phenotypic information. Following Zhong and Jannink (2007), the SPV for cross $i \times j$ is defined as

$$\mathcal{P}_{i\times j} = \mu_{i\times j} + \iota \sigma_{i\times j}, \tag{5}$$

where $\mu_{i\times j}$ is the expected genetic mean of the progeny from cross $i \times j$ (Eq. 1), $\iota$ is as defined above and $\sigma_{i\times j}$ is the additive genetic standard deviation among progeny from the cross. The UC is defined on a phenotypic scale, whereas the SPV is defined on the genetic scale. Hence the SPV corresponds to the UC under $h^2 = 1$.

Given average effects as defined above, the additive genetic variance among progeny from a cross can be written in quadratic form as

$$\begin{aligned}\sigma^2_{\mathrm{A},i\times j} &= \boldsymbol{\alpha}^\top \operatorname{Var}(\mathbf{M}_{i\times j})\boldsymbol{\alpha} \\ &= \sum_k \sum_l \alpha_k \operatorname{Cov}(m_{i\times j,k}, m_{i\times j,l})\alpha_l\end{aligned} \tag{6}$$

where $k, l$ sum over all locus pairs. Because centring locus genotypes only changes their means, the segregation variance can be calculated directly from the covariance of allele dosages. The expanded form highlights that segregation variance depends not only on individual locus effects, but also on pairwise covariances between loci, i.e. the linkage disequilibrium structure of the progeny. The key task is therefore to define $\operatorname{Var}(\mathbf{M}_{i\times j})$, or equivalently $\operatorname{Cov}(m_{i\times j,k}, m_{i\times j,l})$, which depends on the parental and progeny types (e.g., inbred lines or $F_1$).

For inbred parents, $\operatorname{Cov}(m_{i\times j,k}, m_{i\times j,l})$ can be obtained as

$$\operatorname{Cov}(m_{i\times j,k}, m_{i\times j,l}) = (m_{i,k} - m_{j,k})(m_{i,l} - m_{j,l})c_{kl},$$

where the first term represents the parental allele configuration (linkage phase) between QTL at loci $k$ and $l$, and $c_{kl}$ the recombination-dependent covariance coefficient between these QTL. Thus, the segregation variance depends on the parental allele configuration and its decay through recombination in the progeny. The construction of $c_{kl}$ depends on the offspring type (e.g., doubled haploid (DH) or recombinant inbred line (RIL)). For asymptotic RILs derived directly from an $\mathrm{F}_1$, Zhong and Jannink (2007), building on Haldane and Waddington (1931), derived

$$c_{kl} = 0.25\frac{1 - 2r_{kl}}{1 + 2r_{kl}},$$

where $r_{kl}$ is the recombination frequency between loci $k$ and $l$, derived from a mapping function such as Haldane (1919). This expression has an intuitive interpretation. The numerator $(1 - 2r_{kl})$ reflects the persistence of LD through recombination, whereas the denominator $(1 + 2r_{kl})$ accounts for the accumulation of recombination events across infinite generations of selfing. Analogous formulations exist for DH lines (Osthushenrich et al. 2017; Lehermeier et al. 2017); see Appendix A for the RIL and DH methods implemented in CrossingTools. Extensions of SPV that incorporate dominance effects through Eq. 3 and dominance segregation variance have been described by Wolfe et al. (2021) and Bonk et al. (2016) and their implementation in CrossingTools is summarised in Appendix A.

### 2.1.3 Optimal haploid value

The OHV (Daetwyler et al. 2015) or similarly upper bound selection limits (Cole and VanRaden 2011) evaluates the maximum attainable additive genetic value that could

be achieved from a cross, assuming that at each locus the favourable allele present in either parent can be fixed. Unlike the UC and SPV, which depend on the expected mean and variance of segregating progeny, the OHV represents a theoretical upper bound on cross potential under purely additive gene action.

Following Daetwyler et al. (2015), the OHV for a cross between two inbred lines $i$ and $j$ is defined as

$$\mathcal{H}_{i\times j} = \sum_{k=1}^{r} \max\left[\alpha_k(m_{i,k} - 2p_k), \alpha_k(m_{j,k} - 2p_k)\right]. \tag{7}$$

where $k$ sums over all loci. The OHV therefore assumes that, at each locus, the parental allele configuration with the larger additive contribution can be combined into an optimal genotype, ignoring linkage constraints among loci. For non-inbred parents, the definition extends naturally by considering the corresponding additive contributions of the parental gametes.

### 2.1.4 Optimal cross selection

Ranking crosses solely by predicted value with one of the criteria defined above may maximise short-term genetic gain, but ignores relationships among parents, potentially accelerating coancestry and reducing long-term genetic variance. Sustainable improvement therefore requires balancing expected progeny performance with the maintenance of genetic diversity.

For a crossing plan $P$ comprising $n_{\text{cross}}$ crosses, the average predicted value is

$$s_{(P)} = \frac{1}{n_{\text{cross}}} \sum_{(i,j)\in P} \Psi_{i\times j},$$

where $n_{\text{cross}} = |P|$ is the number of crosses in the plan and $\Psi_{i\times j}$ denotes the criterion value of cross $i \times j$. Diversity is quantified through group coancestry (Meuwissen 1997),

$$C_{(P)} = \mathbf{c}_{(P)}^{\top}\mathbf{K}\mathbf{c}_{(P)},$$

where $\mathbf{K}$ is the genomic relationship matrix and $\mathbf{c}_{(P)}$ contains proportional parental contributions, with each mating contributing $1/(2n_{\text{cross}})$ to each of its two parents.

Because $s_{(P)}$ and $C_{(P)}$ are measured on different scales, both quantities are normalised to $[0, 1]$, yielding $\tilde{s}_{(P)}$ and $\tilde{c}_{(P)}$, where smaller $\tilde{c}_{(P)}$ corresponds to greater diversity. Each crossing plan is therefore represented as

$$\mathbf{v}_{(P)} = \begin{pmatrix} \tilde{s}_{(P)} \\ \tilde{c}_{(P)} \end{pmatrix}$$

in gain–diversity space.

CrossingTools provides two optimisation strategies. The first is a target-angle approach, in which the user specifies a desired balance between gain and diversity through an angle $\theta \in [0°, 90°]$. optimisation then proceeds by maximising

$$\max_{P} \ \mathbf{a}_{\theta}^{\top}\mathbf{v}_{(P)}, \qquad \mathbf{a}_{\theta} = (\cos\theta, \sin\theta)^{\top},$$

where $\theta = 0°$ corresponds to pure gain maximisation and $\theta = 90°$ places maximum emphasis on diversity.

Alternatively, gain and diversity can be treated as two independent objectives and optimised simultaneously using the Non-dominated Sorting Genetic Algorithm II (NSGA-II) (Deb et al. 2002). Rather than specifying a desired trade-off *a priori*, NSGA-II identifies the set of non-dominated crossing plans forming an approximation to the Pareto frontier. These solutions represent alternative gain–diversity trade-offs, allowing breeders to subsequently choose the crossing plan that best matches their breeding objectives.

Because the number of feasible crossing plans grows combinatorially with the population size, both optimisation strategies are implemented using evolutionary algorithms. The target-angle approach employs a genetic algorithm (Storn and Price 1997), whereas Pareto optimisation uses NSGA-II (Deb et al. 2002) type genetic algorithm. In both cases, each candidate solution represents a feasible crossing plan that is iteratively improved through evolutionary operations like mating, mutation and selection. Practical breeding constraints, including mandatory crosses, prohibited crosses and restrictions on maximal or minimal contributions, can be incorporated directly into the optimisation procedure.

## 2.2 Index selection

Breeding objectives generally involve multiple traits, which can be combined into a single scalar objective using selection index theory. Within the cross evaluation framework developed above, this allows cross performance criteria to be extended directly to multi-trait breeding objectives. CrossingTools implements Smith–Hazel (economic-weight) and Desired Gains indices.

### 2.2.1 Smith–Hazel index

The Smith–Hazel Index (Smith 1936; Hazel and Lush 1942; Hazel 1943) combines trait values into an aggregate measure of genetic merit. For individual $i$,

$$\mathcal{I}_i = \sum_{t=1}^{T} \omega_t g_{i,t} = \boldsymbol{\omega}^\top \boldsymbol{g}_i, \tag{8}$$

where $\omega_t$ is the weight assigned to trait $t$, $g_{i,t}$ is the corresponding BV of individual $i$, and $T$ is the number of traits.

The same formulation can be applied to crosses by replacing $\boldsymbol{g}_i$ with a cross-specific measure of genetic merit, including the expected cross mean (Eq. 1), predicted $F_1$ performance (Eq. 3), or OHV (Eq. 7). Likewise, SPV can be extended to multiple traits by accounting for segregation covariances among traits. Following Wolfe et al. (2021),

$$\mathcal{P}_{i\times j} = \boldsymbol{\omega}^\top \boldsymbol{\mu}_{i\times j} + \iota\sqrt{\boldsymbol{\omega}^\top \boldsymbol{\Sigma}_{i\times j}\boldsymbol{\omega}}, \tag{9}$$

where $\boldsymbol{\mu}_{i\times j}$ is the vector of expected progeny means and $\boldsymbol{\Sigma}_{i\times j}$ the segregation (co)variance matrix among traits. The two terms therefore represent the expected

index mean and the response to within-family selection along the index direction, respectively.

Let $\mathbf{U}$ be an $r \times T$ matrix of average allele-substitution effect QTL, where $r$ is the number of QTL and $T$ is the number of traits, with element $u_{k,t}$ denoting the additive effect of QTL $k$ on trait $t$. From Eq. 6, the $T \times T$ segregation (co)variance matrix for cross $i \times j$ is

$$\boldsymbol{\Sigma}_{i\times j} = \mathbf{U}^\top \operatorname{Var}(\mathbf{M}_{i\times j})\mathbf{U}. \tag{10}$$

Consequently, the segregation variance of the index can be written as

$$\boldsymbol{\omega}^\top \boldsymbol{\Sigma}_{i\times j}\boldsymbol{\omega} = (\mathbf{U}\boldsymbol{\omega})^\top \operatorname{Var}(\mathbf{M}_{i\times j})(\mathbf{U}\boldsymbol{\omega}), \tag{11}$$

where $\mathbf{U}\boldsymbol{\omega}$ contains the QTL effects of the index itself. Thus, for selection based solely on a fixed linear index, segregation variance can be calculated directly from these transformed effects without constructing the complete trait-level segregation covariance matrix.

The full matrix $\boldsymbol{\Sigma}_{i\times j}$ is nevertheless required when trait-specific variances and covariances are of interest (Neyhart et al. 2019). For large numbers of traits, reduced-rank representations of $\mathbf{U}$ can provide a computationally efficient approximation; corresponding formulations and reconstruction of the full covariance matrix are described in Appendix B.

### 2.2.2 Desired gains index

An economic-weight index maximises response in the aggregate objective defined by $\boldsymbol{\omega}$, but does not directly control the relative response of individual traits. These responses also depend on the population genetic (co)variance structure and may therefore deviate from the desired breeding objective, particularly for unfavourably correlated traits. Desired Gains indices provide an alternative by defining the intended relative response directly (Pesek and Baker 1969). The Desired Gains Index derives the index weights from a vector of desired genetic gains rather than predefined economic weights. Let $\boldsymbol{\Delta}$ denote the Desired Gains vector and $\mathbf{G}$ the population-wide genetic covariance matrix. The weights in Eq. 8 are then

$$\boldsymbol{\omega} = \mathbf{G}^{-1}\boldsymbol{\Delta},$$

giving

$$\mathcal{I}_i = \boldsymbol{\Delta}^\top \mathbf{G}^{-1}\boldsymbol{g}_i. \tag{12}$$

The expected response is therefore proportional to $\boldsymbol{\Delta}$, allowing the relative improvement among traits to be predefined. The same weights can be applied to the cross evaluation criteria above to optimise crosses towards the specified multivariate response.

## 2.3 Estimation of marker effects and cross criteria

The preceding sections assume that QTL effects are known. In practice, genetic effects are estimated from marker data using mixed linear models. Let $\mathbf{y}$ denote the stacked

vector of $n_{\text{obs}}$ phenotypic observations recorded on $n_{\text{ind}}$ individuals across $T$ traits. The multivariate linear mixed model assuming only additive effects is given by

$$\mathbf{y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{Z}\boldsymbol{g} + \mathbf{e},$$

where $\boldsymbol{\tau}$ is a vector of fixed effects with design matrix $\mathbf{X}$, $\mathbf{Z}$ is the incidence matrix linking observations to individual-by-trait genetic effects, $\boldsymbol{g}$ is the vector of genetic effects, and $\mathbf{e}$ is the residual vector.

Assuming that genetic effects are stacked by trait,

$$\boldsymbol{g} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{G} \otimes \mathbf{K}\right), \qquad \mathbf{e} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{R}\right).$$

where $\mathbf{G}$ is the additive genetic (co)variance matrix among traits, $\mathbf{K}$ is the genomic relationship matrix (GRM), and $\mathbf{R}$ is the residual (co)variance matrix.

For example, following VanRaden (2008), the GRM can be constructed as

$$\mathbf{K} = \frac{\mathbf{W}\mathbf{W}^{\top}}{2\sum_{k=1}^{m} p_k(1-p_k)},$$

where $\mathbf{W} = \mathbf{M} - 2\boldsymbol{p}$ is the centred marker genotype matrix, $\mathbf{M}$ contains allele dosages coded as 0, 1, or 2, and $\boldsymbol{p}$ is the vector of reference allele frequencies.

The resulting GEBVs can be used to backsolve estimated marker effects. Let $\widehat{\mathbf{g}}_t$ denote the vector of estimated GEBVs for trait $t$. The corresponding marker-effect estimates are

$$\widehat{\boldsymbol{\alpha}}_t = \frac{1}{2\sum_{k=1}^{r} p_k(1-p_k)} \mathbf{W}^{\top}\mathbf{K}^{-1}\widehat{\mathbf{g}}_t. \tag{13}$$

The cross performance criteria defined above are then estimated by replacing the unknown QTL effects with their estimated marker effects and using marker genotypes in place of QTL genotypes. In CrossingTools Economic weights can be applied directly to the resulting multi-trait BLUPs, whereas the desired gains weights implemented in CrossingTools are derived using $\text{Var}(\widehat{\boldsymbol{g}})$ following Werner et al. (2024).

## 2.4 Overview of the CrossingTools workflow

This section demonstrates the implementation of the proposed framework using the R package CrossingTools. The accompanying supplementary R script presents a line-breeding example, while the general workflow comprises four stages: prediction, objective definition, cross evaluation, and crossing-plan optimisation (Figure 1).

First, marker effects and genetic values are estimated using an external genomic prediction model (Figure 1A). Second, the breeding objective is defined using selection index theory, with multiple traits combined through economic weights, i.e. Smith–Hazel index or Desired Gains (Figure 1B). Expected responses to index selection can be visualised to assess whether the resulting objective reflects the breeding goals. Third, all potential parental combinations are evaluated using the expected cross mean, SPV, OHV, or corresponding dominance-based extensions (Figure 1C). Finally, predicted cross values are combined with genomic diversity to optimise crossing plans using either a predefined target-angle or Pareto-based optimisation, with practical breeding constraints incorporated where required (Figure 1D).

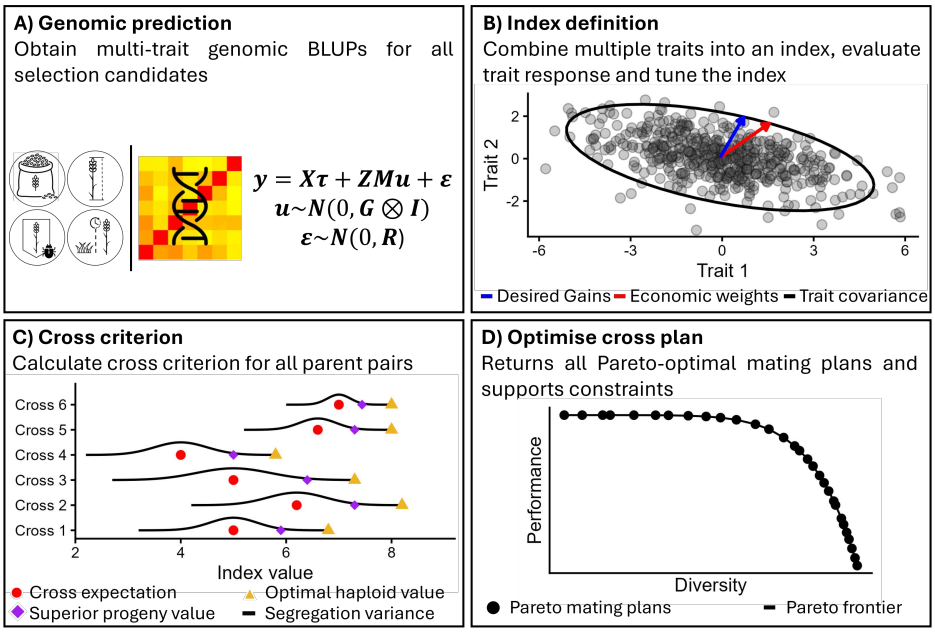



**Fig. 1** Graphical overview of the workflow implemented in CrossingTools. **(A)** Marker effects and genetic values are estimated using external genomic prediction software and imported into CrossingTools. **(B)** Multi-trait breeding objectives are defined from the predicted genetic values and their genetic (co)variance structure. Grey points represent individuals in the multi-trait space, and the ellipse depicts their genetic covariance structure. The arrows indicate the directions of the Smith–Hazel (red) and Desired Gains (blue) selection indices. In this example, the Smith–Hazel index uses equal economic weights for the two traits, $a = (1, 1)$, whereas the Desired Gains index targets equal genetic improvement in both traits, corresponding to a desired-gain ratio of 1:1. **(C)** Cross performance criteria are predicted for each potential parental combination. Curves represent the expected offspring distributions, with their spread reflecting the predicted within-cross segregation variance. Crosses can be evaluated using the expected cross mean (red circles), superior progeny value (purple diamonds), and optimal haploid value (yellow triangles). **(D)** Cross performance is combined with genomic diversity to optimise mating plans using OCS. Each point represents a candidate mating plan, and the frontier illustrates the trade-off between expected performance and diversity.

The demonstration considers fully inbred parents and DH progeny, but the framework can be adapted to alternative breeding schemes through the choice of progeny type, recombination model, and cross performance criterion. For example, dominance effects and within-family dominance variance can be incorporated when evaluating crosses in clonal breeding programmes.

## 2.5 Simulation study

Stochastic simulations were conducted to demonstrate the utility and practical application of CrossingTools for multi-trait cross selection. We simulated a line breeding programme with two correlated traits. We then compared alternative index-based cross selection strategies under cross evaluation criteria as implemented in CrossingTools.

### 2.5.1 Simulation of the founder population

Whole-genome sequences were simulated for a founder population of 1,000 fully inbred lines using the runMacs function implemented in AlphaSimR (Gaynor et al. 2021). The simulated genome consisted of 10 homozygous chromosome pairs, each with a physical length of $1 \times 10^8$ base pairs (bp) and a genetic length of 100 cM, resulting in a total genome length of 1000 cM.

Along each chromosome, 500 additive QTL and 500 SNP markers were randomly sampled, yielding 5,000 QTL and 5,000 SNP markers across the genome. All founder lines were fully homozygous.

### 2.5.2 Simulation of Genetic and Phenotypic Values

Two quantitative traits were simulated under purely additive gene action. Additive genetic values were obtained by summing the effects of the 5,000 QTL for each trait. QTL effects were sampled from a bivariate normal distribution with mean zero and covariance matrix

$$\mathbf{\Sigma} = \begin{bmatrix} 2 & \sigma_{12} \\ \sigma_{12} & 1 \end{bmatrix},$$

where $\sigma_{12}$ was set to 0.71 or $-0.71$, corresponding to a genetic correlation of 0.5 or $-0.5$ between the two traits, respectively.

Phenotypic values were generated by adding residual errors to the additive genetic values. Residuals were sampled from a bivariate normal distribution with mean zero and diagonal covariance matrix $\mathrm{diag}(\sigma^2_{e1}, \sigma^2_{e2})$ and scaled such that both traits have a heritability of 0.5 in the founder population. Heritability was defined as

$$h^2 = \frac{\sigma^2_a}{\sigma^2_a + \sigma^2_e},$$

where $\sigma^2_a$ and $\sigma^2_e$ denote the additive genetic and residual variances, respectively.

### 2.5.3 Cross selection strategies and breeding programme designs

The breeding programme aimed to improve both traits simultaneously, with cross selection performed among all possible biparental combinations. Two alternative multi-trait breeding objectives were considered:

(i) Smith–Hazel index with economic weights $(1, 1)^\top$,
(ii) Desired Gains index targeting proportional improvement of $(1, 1)^\top$.

For each index, crosses were evaluated using three alternative cross-level metrics:

(a) Expected cross mean,
(b) Superior Progeny Value,
(c) Optimal Haploid Value.

Crosses were selected either by truncation (top 50 crosses) or by OCS using the gain–diversity framework with target angles $\theta = 15°$ or $\theta = 30°$ to select the 50 crosses. The performance of the different selection methods was assessed using both true genetic values and estimated marker effects.

In the line breeding programme, selected crosses were executed and from each cross 50 DH lines were generated. These DH lines formed the candidate population for the subsequent cycle. GEBVs were obtained using the multivariate ridge regression BLUP model described in Section 2.3 with the R package asremlR (Butler et al. 2023). Marker effects were re-estimated in each cycle using only the DH lines produced in that cycle. Updated estimates were then used to evaluate all possible crosses in the next generation. This process of cross evaluation, DH production, and model retraining was repeated for 10 breeding cycles.

### 2.5.4 Evaluation

Genetic gain was assessed as the change in mean additive genetic value of the DH population across breeding cycles. Commercialisable gain was recorded as the maximum genetic value observed in the population in each generation. In addition, genetic (co)variances were monitored. All scenarios were replicated 50 times.

## 2.6 Prediction accuracy of cross criteria

Using the founder population described in Section 2.5, 5,000 biparental crosses were randomly sampled and 50 DH progeny were generated for each cross. Based either on the true QTL effects or estimated marker effects as described above, the expected cross mean, segregation variance, SPV, and OHV were calculated for both traits and for the two selection indices. The realised family mean and segregation variance were calculated directly from the true genetic values of the simulated progeny. Likewise, the realised SPV was calculated from the realised family mean and segregation standard deviation. The predicted cross mean, segregation variance, and SPV were then compared with their realised counterparts using Pearson correlations across the 5,000 crosses. As OHV represents a theoretical upper limit rather than the expected value of a finite family, it was instead compared with the highest genetic value realised among the 50 progeny of each cross. Correlations were calculated separately for predictions based on true QTL effects and estimated marker effects to assess both how well the different criteria represented realised family performance and how this was affected by marker-effect estimation.

## 2.7 Comparing CrossingTools with other software

The calculation of the expected cross mean (Eq. 1) or the expected $F_1$ performance (Eq. 3) is computationally straightforward. In contrast, the calculation of segregation variance requires evaluating covariance structures across markers. For example, with 5,000 markers this involves operations on matrices of dimension $5000 \times 5000$. Because the number of potential crosses increases quadratically with the number of candidate parents, segregation variance calculations must be repeated many times

when evaluating large crossing sets. To assess computational performance, we compared CrossingTools with several existing open-source R packages for cross evaluation: SimpleMating (Peixoto et al. 2025), PopVar (Mohammadi et al. 2015), and genomicSelectR (Wolfe et al. 2021). Using the founder population described in Section 2.5, a subset of 5000 biparental crosses was randomly sampled, and the segregation variance for a single trait was predicted using each package. This benchmarking procedure was repeated 100 times. For each replicate, the time required to compute segregation variances was recorded in seconds. Where applicable, multi-threading was disabled to ensure a comparable runtime measurement across implementations.

# 3 Results

## 3.1 Selection on true genetic values

### 3.1.1 Genetic gain

In the simulated breeding programme under the Smith–Hazel index, the realised pattern of genetic gain largely reflected the simulated variance structure. Across scenarios, the trait with the larger genetic variance exhibited the stronger response to selection, whereas response in the lower-variance trait was more limited (Fig. 2; Supplementary Fig. S1–S2). This asymmetry was particularly pronounced when the two traits were unfavourably correlated, as selection for improvement in one trait increasingly constrained response in the other (Fig. 2; Supplementary Fig. S1).

Without explicit control of genetic diversity through OCS, there were only minor differences in population mean genetic gain between mate allocation based on average BV and SPV. In contrast, mate allocation based on OHV resulted in lower population mean gain across the examined timeframe. Incorporating OCS primarily constrained short-term gain; however, unlike scenarios without OCS, no saturation of genetic progress was observed within the 10 breeding cycles. As expected, this effect was more pronounced at a target angle of 30° than at 15° (Fig. 2; Supplementary Fig. S1–S2).

Using the Desired Gains index, the response to selection was balanced across both traits, resulting in similar genetic gain irrespective of differences in trait-specific variance or the underlying genetic correlation (Fig. 2; Supplementary Fig. S1–S2). In contrast to the Smith–Hazel index, no asymmetry in response between traits was observed, and both short- and long-term genetic gain followed comparable trajectories. Differences among mate allocation criteria and the impact of OCS followed the same general pattern as observed under the Smith–Hazel index (Fig. 2; Supplementary Fig. S1–S2).

### 3.1.2 Commercialisable gain

In the simulated breeding programme, evaluation of commercialisable gain under truncation-based crossing decisions without OCS revealed two main patterns. First, ranking crosses by SPV consistently produced the highest maximum values across both traits, followed by average BV, while OHV-based selection generally resulted in the lowest gains. The difference between mean and maximum population performance was larger for the trait with greater genetic variance, although differences between

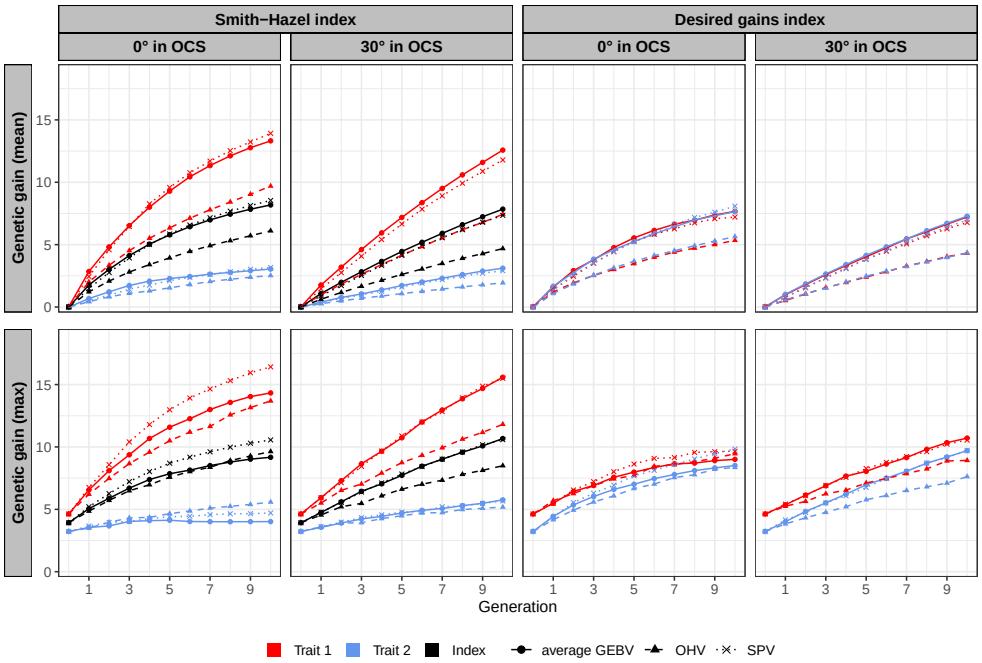



**Fig. 2** Mean (top row) and commercialisable/maximum (bottom row) genetic gain over ten generations of mating based on true QTL effects under the Smith–Hazel index (left block) and Desired Gains index (right block) with the simulated line breeding programme. Columns show OCS angles of 0° and 30°. Genetic gain is shown for Trait 1 (red), Trait 2 (blue), and for the Smith–Hazel index (black). Selection was based on average BVs (average GEBV; circles and solid lines), OHVs (OHV; triangles and dashed lines), or SPV (SPV; crosses and dotted lines). Results are shown for an initial genetic correlation of -0.5.

OHV and the other criteria became smaller at the level of maximum performance. Differences between SPV and average BV largely disappeared once OCS was introduced, whereas OHV continued to produce lower maximum values across scenarios (Fig. 2; Supplementary Fig. S1–S2).

Second, under negative trait correlations and the Smith–Hazel index, the maximum value of the lower-variance trait stagnated after approximately five generations under average BV-based mate allocation (Fig. 2; Supplementary Fig. S1). Despite this stagnation, overall index gain continued to increase throughout the observed time frame (Fig. 2; Supplementary Fig. S1 ).

Under the Desired Gains index, ranking crosses by SPV again produced the highest maximum values, followed by average BV, while OHV-based mate allocation consistently resulted in the lowest commercialisable gain. In contrast to the Smith–Hazel index, no stagnation in the maximum performance of the lower-variance trait was observed. As before, differences between SPV and average BV were substantially reduced when coancestry was controlled through OCS (Fig. 2; Supplementary Fig. S1–S2).

### 3.1.3 Genetic variance

In the simulated breeding programme, both indices resulted in a gradual depletion of genetic variance across breeding cycles. The rate of decline was consistently greater for the trait with the larger initial variance, and this pattern was observed across both simulated levels of trait correlation and with both indices. Across all scenarios, genetic variance was best preserved when crosses were selected based on OHV, followed by SPV, while selection based on average BVs led to the fastest depletion, particularly in the absence of diversity control (Fig. 3; Supplementary Fig. S3–S4).

Incorporating OCS reduced the rate of variance depletion for all selection criteria. This effect was most pronounced for average BV-based selection, whereas SPV- and OHV-based selection maintained relatively higher levels of variance irrespective of diversity control and were less affected by its implementation. Differences between target angles of 15° and 30° did not substantially alter these patterns (Fig. 3; Supplementary Fig. S3–S4).

### 3.1.4 Genetic covariance and correlation

In the simulated breeding programme, distinct trends in genetic covariance were observed depending on the sign of the initial trait correlation. Under positive trait correlations, the covariance between traits declined over breeding cycles but remained higher under SPV- than under average BV-based mate allocation, resulting in a higher realised genetic correlation. OHV-based selection followed a similar pattern, although to a lesser extent (Supplementary Fig. S3–S4).

Under negative trait correlations, the covariance between traits increased over time under both SPV- and average BV-based selection, whereas it remained approximately constant under OHV. Due to the slower depletion of genetic variance under SPV, the realised genetic correlation increased relative to the initial value, while it decreased under average BV-based selection and remained largely unchanged under OHV (Fig. 3; Supplementary Fig. S3).

Implementation of OCS reduced differences between mate allocation criteria. In particular, the increase in genetic correlation under SPV was less pronounced, the decline under average BV-based selection was moderated, and OHV-based selection continued to maintain correlations close to their initial values (Fig. 3; Supplementary Fig. S3–S4).

## 3.2 Genomic selection

Overall, patterns observed under GS were largely consistent with those obtained under selection on true genetic values. However, differences between mate allocation criteria were generally smaller under GS.

### 3.2.1 Genetic gain

Differences in population mean genetic gain between SPV- and average GEBV-based mate allocation were generally smaller under GS than under selection on true genetic values. Across breeding programmes and selection indices, OHV-based selection still

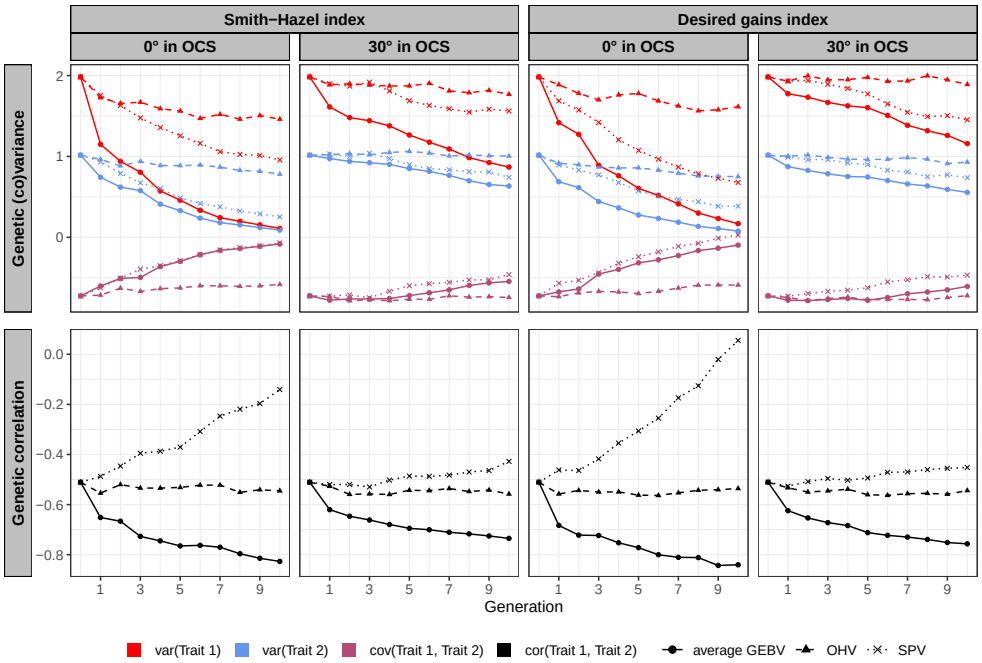



**Fig. 3** Genetic (co)variances (top row) and genetic correlation (bottom row) over ten generations of mating based on true QTL effects under the Smith–Hazel index (left block) and Desired Gains index (right block) with the simulated line breeding programme. Columns show OCS angles of 0° and 30°. Genetic variance is shown for Trait 1 (red), Trait 2 (blue), and their covariance (purple) and correlation (black). Selection was based on average BVs (GEBV; circles and solid lines), OHVs (OHV; triangles and dashed lines), or SPV (SPV; crosses and dotted lines). Results are shown for an initial genetic correlation of -0.5.

resulted in the lowest overall index gain in most scenarios, although the reduction in gain relative to SPV and average GEBV was less pronounced (Fig. 4; Supplementary Fig. S5–S6). Patterns under the Smith–Hazel index were otherwise consistent with those observed under selection on true genetic values, including the stronger response of the higher-variance trait and the effect of OCS in reducing short-term gain while maintaining continued genetic progress over time (Fig. 4; Supplementary Fig. S5–S6).

Using the Desired Gains index, response to selection remained balanced across both traits irrespective of differences in trait-specific variance or genetic correlation, and differences among mate allocation criteria followed the same general pattern as under the Smith–Hazel index (Fig. 4; Supplementary Fig. S5–S6).

### 3.2.2 Commercialisable gain

At the level of commercialisable gain, SPV- and OHV-based mate allocation produced the highest maximum values in most scenarios, although differences relative to average GEBV were generally smaller than under selection on true genetic values (Fig. 4; Supplementary Fig. S5–S6). Under negative trait correlations and using the Smith–Hazel index, stagnation or decline in commercialisable gain for the lower-variance trait was

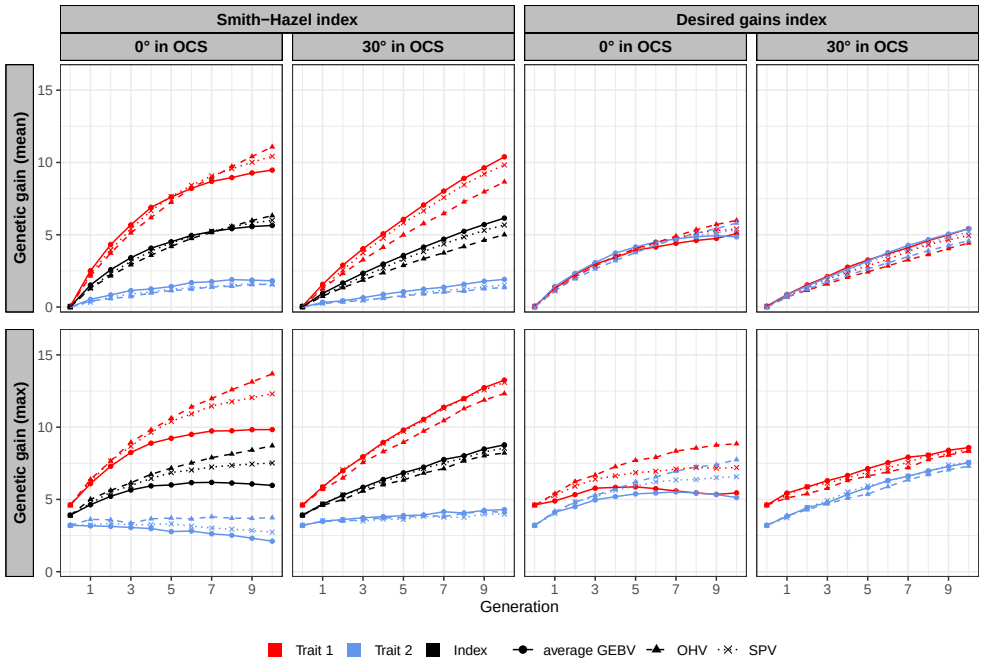



**Fig. 4** Mean (top row) and commercialisable/maximum (bottom row) genetic gain over ten generations of mating based on estimated marker effects under the Smith–Hazel index (left block) and Desired Gains index (right block). Columns show OCS angles of 0° and 30°. Genetic gain is shown for Trait 1 (red), Trait 2 (blue), and for the Smith–Hazel index (black). Selection was based on average GEBVs (average GEBV; circles and solid lines), OHVs (OHV; triangles and dashed lines), or SPV (SPV; crosses and dotted lines). Results are shown for an initial genetic correlation of -0.5.

again observed under SPV- and average GEBV-based mate allocation and occurred earlier than under selection on true genetic values, while gain in the overall selection decreased only with mate allocation based on average GEBV (Fig. 4; Supplementary Fig. S5). OHV and SPV, on the other hand, did not result in negative gain for the lower variance trait, leading to continued commercialisable gain in the index with OHV performing slightly better than SPV based selection (Fig. 4; Supplementary Fig. S5). Introducing OCS again reduced differences among the cross criteria (Fig. 4; Supplementary Fig. S5–S6), with OHV being more strongly affected by it, leading to lower gains compared to the other two strategies.

Under the Desired Gains index without control of genetic diversity, OHV-based selection performed comparably to SPV under negative trait correlations, while differences between mate allocation criteria remained smaller than under selection on true genetic values. Introducing control of genetic diversity had a stronger effect on OHV which achieved lower gain rates and total gain compared to SPV and parental average, which performed equally.

### 3.2.3 Genetic variance

Patterns of genetic variance under GS closely mirrored those observed under selection on true genetic values (Fig. 5; Supplementary Fig. S7–S8). Across breeding programmes, selection indices, and trait correlations, genetic variance was best preserved under OHV-based mate allocation, followed by SPV, whereas average GEBV-based selection resulted in the fastest depletion. OCS reduced the rate of variance depletion for all mate allocation criteria, with the strongest effect observed for average GEBV-based selection.

Compared with selection on true genetic values, differences between mate allocation criteria were generally smaller under GS, particularly when OCS was implemented.

### 3.2.4 Genetic covariance and correlation

Patterns of genetic covariance and correlation under GS were also consistent with those observed under selection on true genetic values (Fig. 5; Supplementary Fig. S7–S8). Under positive trait correlations, covariance declined over breeding cycles but remained highest under SPV-based mate allocation, whereas under negative trait correlations it increased under SPV- and average GEBV-based selection but remained approximately constant under OHV.

Implementation of OCS reduced differences between mate allocation criteria, resulting in smaller changes in realised genetic correlations while maintaining the same overall ranking of selection strategies (Fig. 5; Supplementary Fig. S7–S8). As for genetic variance, differences between mate allocation criteria were generally less pronounced under GS.

## 3.3 Accuracy of cross criteria

Prediction accuracies for the average (GE)BV, segregation variance, and SPV were generally higher when true QTL effects were used than when marker effects were estimated, with segregation variance and SPV showing the largest reductions under genomic prediction (Supplementary Fig. S9). Based on true QTL effects, accuracy was highest for the average BV, followed by segregation variance and SPV. Under genomic prediction, the same general pattern was observed, although SPV was more strongly affected by marker-effect estimation. In contrast, the correlation between OHV and the realised family maximum increased when estimated marker effects were used compared with true QTL effects (Supplementary Fig. S9).

## 3.4 Comparing CrossingTools

The computational performance of CrossingTools was evaluated by comparing runtimes for the prediction of segregation variances across different software packages. For inbred genotypes, CrossingTools required on average 3.47 s to compute segregation variances, compared to 49.03 s for SimpleMating and 90.77 s for PopVar (Fig. 6). This corresponds to a speed increase of 14.13-fold relative to SimpleMating and 26.16-fold relative to PopVar. For outcrossing genotypes, CrossingTools required

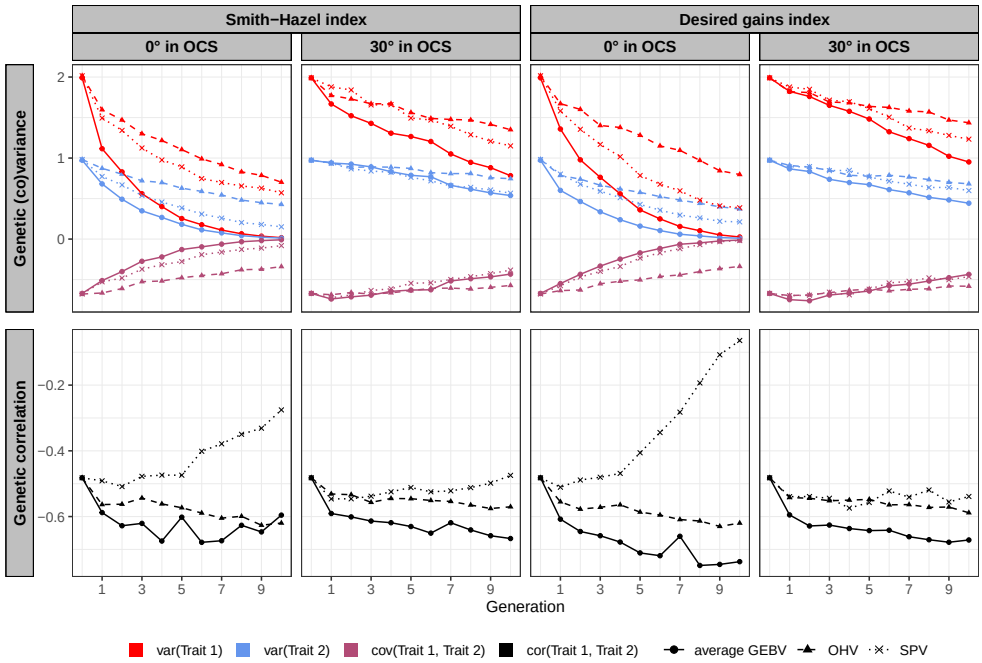



**Fig. 5** Genetic (co)variances (top row) and genetic correlation (bottom row) genetic gain over ten generations of mating based on estimated marker effects under the Smith–Hazel index (left block) and Desired Gains index (right block). Columns show OCS angles of 0° and 30°. Genetic variance is shown for Trait 1 (red), Trait 2 (blue), and their covariance (purple) and correlation (black). Selection was based on average GEBVs (average GEBV; circles and solid lines), OHVs (OHV; triangles and dashed lines), or SPV (SPV; crosses and dotted lines). Results are shown for an initial genetic correlation of -0.5.

on average 29.04 s, whereas the implementation by Wolfe required 2492.55 s (Fig. 6). This corresponds to an 85.83-fold reduction in computation time.

# 4 Discussion

The objective of this study was to introduce CrossingTools as a unified and computationally efficient framework for mate allocation in breeding programmes. CrossingTools integrates alternative cross evaluation criteria, including expected cross mean, SPV, and OHV, with multi-trait selection indices and gain–diversity optimisation through OCS. Using stochastic simulations of line breeding programmes under GS, we compared these criteria and optimisation strategies with respect to genetic gain, genetic variance, and mating decisions.

## 4.1 Relationships among cross evaluation criteria

Under purely additive gene action, the cross evaluation criteria considered here can be interpreted within a common framework based on expected progeny mean and segregation variance. The expected cross mean corresponds to a special case of the

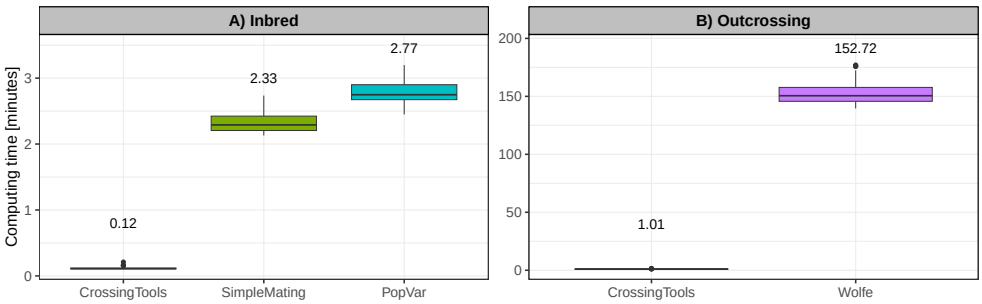



**Fig. 6** Computing time for segregation variance prediction across different software implementations. Boxplots show runtimes for evaluating 5,000 biparental crosses for a single trait across 100 benchmark replicates. For inbred genotypes, CrossingTools was compared with SimpleMating (Peixoto et al. 2025) and PopVar (Mohammadi et al. 2015), while for outcrossing genotypes it was compared with the implementation in genomicSelectR by Wolfe (Wolfe et al. 2021). Computing time is shown in minutes, with values above the boxplots indicating the mean runtime.

SPV with a standardised selection differential of $\iota = 0$, such that selection is based solely on the progeny mean. Likewise, the OHV can be related to the SPV framework by defining

$$\mathcal{H}_{i\times j} = \mu_{i\times j} + \iota\sigma_{i\times j}$$

and solving for $\iota$:

$$\iota = \frac{\mathcal{H}_{i\times j} - \mu_{i\times j}}{\sigma_{i\times j}}.$$

This formulation shows that the OHV can be interpreted as an extreme form of SPV associated with an implicit standardised selection differential $\iota$ that differs among crosses. Consequently, the magnitude of this cross specific $\iota$ provides an indication of how attainable the OHV is relative to the expected distribution of progeny values. In practice, the OHV therefore represents an optimistic upper bound on cross potential, corresponding to the assumption that recombination can assemble all favourable alleles into a single genotype (Daetwyler et al. 2015; Cole and VanRaden 2011). A further practical consideration is that certain plant-breeding procedures specifically based on the OHV criterion of Daetwyler et al. (2015) are patented in several jurisdictions, potentially limiting commercial use where relevant claims remain in force (Daetwyler et al. 2023). These limitations can be addressed by alternative criteria based on order statistics and finite family sizes, including the expected maximum among a specified number of progeny and the probability of exceeding a predefined threshold (Müller et al. 2018; Danguy des Déserts et al. 2023; Bijma et al. 2020). Such criteria may also support the allocation of progeny numbers among crosses when family size is treated as an optimisation variable (Okoye et al. 2026); corresponding formulations together with required family sizes to achieve a given value are summarised in Appendix C.

The assumptions underlying OHV are also less appropriate when dominance, particularly overdominance, contributes substantially to genetic value. In such cases, the optimal genotype may be heterozygous rather than a homozygous combination of

favourable alleles. Criteria based on predicted total genetic value or SPV extensions incorporating dominance variance may therefore be more appropriate for evaluating crosses in outcrossing and clonal breeding schemes (Wolfe et al. 2021; Werner et al. 2023; Bonk et al. 2016).

## 4.2 Commercialisable genetic gain versus average gain

Differences in population mean gain between selection based on expected cross means and SPV were generally small, consistent with previous studies (Niehoff et al. 2024b,a; Peixoto et al. 2025), although in some cases the effects of segregation variance and diversity control through OCS are not clearly separated. In contrast, SPV-based selection produced higher maximum progeny values, i.e. commercialisable gain, under truncation selection for both true and estimated effects, in agreement with Allier et al. (2019) and Lehermeier et al. (2017), while OHV did so only when GS was employed. This distinction is relevant because variety development depends on identifying a limited number of superior individuals rather than solely on increasing the population mean. The advantage of SPV was reduced when OCS constrained the set of selected crossing plans.

More generally, Wang et al. (2025) argued "why usefulness is rarely useful", highlighting the correlation between SPV and expected cross mean as one reason for the often limited differences between the criteria. This correlation arises because both the usefulnes criterion and SPV explicitly include the mid-parent component. However, the correlation between SPV and expected cross mean depends on the standardised selection differential $\iota$, and for the UC additionally on $\sqrt{h^2}$. Hence, as $\iota$ increases, the relative contribution of segregation variance increases, allowing SPV to distinguish crosses with similar means but different potential to generate extreme progeny. Accordingly, differences between SPV and expected cross mean were more apparent for commercialisable than for average population gain.

## 4.3 Accuracy of cross criteria

The accuracy results showed that the expected cross mean was predicted most accurately, whereas segregation variance and SPV were more strongly affected by the use of estimated marker effects. This is consistent with previous concerns regarding the practical accuracy of segregation variance prediction (Lehermeier et al. 2017; Osthushenrich et al. 2018). In rapid-cycling schemes, marker-effect accuracy can decline across generations (Gaynor et al. 2017; Melchinger and Frisch 2023), which may affect segregation variance more strongly than predictions of the family mean (Osthushenrich et al. 2018; Wang et al. 2025). Additional uncertainty arises for heterozygous parents, where segregation variance depends on phased parental haplotypes (Appendix A), although the effect of phasing errors appears comparatively small (Browning and Browning 2007; Wang et al. 2025). Importantly, the realised values used here and in other studies are themselves estimates from the progeny generated per cross rather than parameters of the complete progeny distribution. Sampling variation in the realised family mean and variance therefore also contributes to imperfect prediction accuracies, even when true QTL effects are used. Thus, absolute (co)variance

prediction accuracies should be interpreted cautiously, while realised values remain useful for validating general patterns across crosses.

OHV showed a distinctly different pattern, with its correlation with realised family maxima increasing when estimated rather than true QTL effects were used. This coincided with OHV producing substantially lower mean and maximum gain when based on true effects, whereas its performance became more similar to SPV and expected cross mean under genomic prediction. One possible explanation is that shrinkage and estimation error reduce the influence of extreme individual marker effects and thereby moderate the optimistic selection limit represented by OHV. Estimated marker effects also capture local LD structure (Lehermeier et al. 2017). Importantly, OHV represents a theoretical upper selection limit and is not intended to predict the maximum of a finite family. Its correlation with the realised maximum should therefore be interpreted as an indication of how well OHV ranks crosses for realised upper-tail performance rather than as prediction accuracy in a strict sense.

## 4.4 Multi-trait crossing decisions through selection indices

Although multi-trait extensions of cross evaluation criteria are straightforward (Allier et al. 2019; Wolfe et al. 2021), few available tools combine the derivation of selection index weights with their direct application to cross evaluation and mating-plan optimisation. CrossingTools integrates both steps for economic-weight and Desired Gains indices, thereby linking multi-trait breeding objectives directly to cross selection.

In the present simulations, the two indices produced distinct trait-specific responses consistent with their respective objectives. Under the Smith–Hazel index with equal economic weights, the trait with greater genetic variance showed the stronger response to selection. Under negative genetic correlation, however, commercialisable gain in the lower-variance trait stagnated, even though the overall index continued to improve. This pattern coincided with depletion of genetic variance in the affected trait, although this reduced when OCS was applied.

In contrast, the Desired Gains index produced similar trajectories of mean genetic gain for both traits, consistent with the specified (1:1) target ratio, irrespective of their initial variances or genetic correlation. Trait differences remained at the level of commercialisable gain, for which the higher-variance trait attained larger maximum values. These results illustrate the distinction between the two indices: the Smith–Hazel index maximises response in an aggregate BV defined by economic weights, whereas the Desired Gains index directs the multivariate response towards a predefined ratio. However, neither approach guarantees that the best individual satisfies minimum requirements for every trait. Index selection may therefore need to be complemented by independent culling or explicit trait constraints during product advancement (Werner et al. 2024).

## 4.5 Maintenance of genetic variance and trait correlations

Genetic variance is a key determinant of sustained response to selection in breeding programmes. Similar to Daetwyler et al. (2015) and Niehoff et al. (2024a), across all scenarios, mate allocation based on SPV or OHV resulted in a slower depletion of

genetic variance compared to selection based on expected cross mean, both with and without explicit control of coancestry. Consequently, higher levels of genetic variance were maintained over breeding cycles under both SPV- and OHV-based selection. This pattern was strengthened when combined with OCS, which reduced the rate of variance loss across both traits. However, the effect was smaller than for average GEBV-based selection, which benefited most from OCS. For SPV, these differences arise from its formulation, which incorporates segregation variance and therefore favours crosses between parents that differ at QTL. As a result, SPV-based selection tends to allocate matings among more genetically diverse parents, whereas selection based on expected mean alone more frequently selects closely related individuals with high BVs (Woolliams et al. 2015). However, in the absence of additional constraints, SPV-based selection can still lead to repeated use of certain parents, highlighting the role of OCS in regulating parental contributions where necessary.

In a multi-trait context, changes in genetic (co)variances translated into systematic shifts in trait correlations over breeding cycles. Under SPV-based mate allocation, the genetic correlation between traits generally moved towards more favourable values as defined by the index, regardless of whether the initial correlation was positive or negative. In contrast, selection based on OHV or expected cross mean tended to maintain or slightly shift correlations in a less favourable direction which is in line with previous findings (Batista et al. 2021). When OCS was applied, these trends were not as pronounced, but the overall pattern remained. Because the index segregation variance contains the cross-specific segregation covariance between traits, SPV favours families whose covariance structure increases variation in the index direction. Repeated selection of such families can alter the realised covariance and correlation structure of the breeding population. Similar patterns have been reported in Neyhart et al. (2019), although candidate preselection in that study may have restricted the range of cross-specific covariance structures explored. While these results were consistent across the simulated scenarios, further work is required to assess this under different index weightings and alternative sources of genetic correlation, such as pleiotropy or LD. Nevertheless, this pattern suggests a potential avenue for addressing unfavourable trait correlations in breeding programmes.

## 4.6 Role of Optimal Cross Selection

As discussed above, genetic variance is a key determinant of sustained genetic gain, yet intensive selection, particularly under GS, can rapidly deplete genetic variation within a few generations (Gaynor et al. 2017; Melchinger and Frisch 2023; Werner et al. 2023; Akdemir and Sánchez 2016). Optimal contribution selection (Meuwissen 1997), originally developed to control inbreeding in animal breeding, has therefore been extended to mate allocation, where coancestry is controlled to manage genetic diversity and limit variance loss (Gorjanc and Hickey 2018; Akdemir and Sánchez 2016). Consistent with previous studies (Gorjanc et al. 2018; Akdemir and Sánchez 2016), OCS reduced variance depletion and supported higher long-term genetic gain in the present simulations, at the cost of lower gain in early generations. Its effect, however, depended strongly on the cross criterion. Variance loss was substantially reduced when crossing plans were evaluated using average GEBV, whereas the additional benefit was

smaller under SPV and OHV because these criteria already preserved more variance without explicit diversity control. Consequently, increasing the target angle produced comparatively small changes in gain and variance trajectories under SPV- and OHV-based selection. For SPV, this may reflect its explicit preference for crosses generating greater trait-relevant segregation variance, such that part of the information captured through genome-wide diversity control overlaps with that already contained in SPV. One possibility to reduce this redundancy would be to regress pairwise coancestry on expected segregation variance and use the residual coancestry as a diversity measure, thereby capturing the component not explained by segregation variance.

The importance and interpretation of diversity control also depend on the breeding system. Optimal contribution and cross selection were developed primarily to control inbreeding while maximising genetic gain (Meuwissen 1997). In animal and clonal breeding, increasing inbreeding can have pronounced negative effects through inbreeding depression caused by deleterious recessive alleles. In contrast, many plant breeding programmes operate with fully or highly inbred lines, often derived from large families, facilitating the purging of deleterious alleles during inbreeding. Moreover, selection is typically conducted across multiple traits (Werner et al. 2024; Pesek and Baker 1969; Brascamp 1984; Smith 1936; Hazel 1943; Hallauer et al. 2010), reducing the effective selection intensity on any single trait. Changing disease pressures and environmental conditions further diversify breeding objectives and reduce the likelihood of rapid fixation of unfavourable alleles. Genotype-by-environment interactions add another source of variation because breeding targets can differ among environments and over time (Technow et al. 2026). In contrast, many simulation studies evaluate single-trait scenarios under purely additive gene action (Gaynor et al. 2017; Melchinger and Frisch 2023; Werner et al. 2023; Akdemir and Sánchez 2016). Although useful for comparing breeding strategies, such scenarios do not capture the full complexity of practical breeding programmes, where multiple traits, non-additive effects, and environmental interactions jointly influence genetic gain, genetic variance, and diversity. These factors may also contribute to the sustained genetic gain observed over long periods in crops such as wheat (Voss-Fels et al. 2019) and maize (Dudley and Lambert 2003; Duvick 2005). Thus, particularly in highly inbred crop breeding, preservation of usable genetic variance may be a more directly relevant objective than diversity or inbreeding per se.

Beyond diversity management, optimisation frameworks such as those implemented here and in other software packages (Kinghorn 2011; Akdemir and Sánchez 2016; Fernández-González et al. 2026; Gorjanc and Hickey 2018; Peixoto et al. 2025) provide a flexible framework for incorporating practical crossing constraints, including limits on parental contributions, mandatory or prohibited crosses, seed availability, and requirements for particular alleles or QTL. Different breeding-programme components may also require different objectives; population improvement may place greater emphasis on maintaining diversity and future genetic variation, whereas product development may prioritise the probability of producing a small number of exceptional genotypes. Crossing-plan optimisation therefore extends beyond coancestry control and provides a general framework for integrating biological and operational constraints.

## 4.7 Computational scalability of cross evaluation

Compared to the relatively simple operations required for calculating the mid-parent value or OHV, which primarily involve summations over marker effects, the calculation of the SPV and usefulness and extensions thereof (Müller et al. 2018; Danguy des Déserts et al. 2023) additionally requires the calculation of segregation variance. Early approaches addressed this by simulating offspring populations (Mohammadi et al. 2015; Bernardo 2014), which becomes computationally demanding as the number of candidate crosses increases. The size of the search space grows approximately quadratically with the number of parents, following $\frac{n(n-1)}{2}$, such that even moderate population sizes result in a large number of potential crosses. More scalable solutions were enabled by algebraic formulations of segregation variance, first introduced by Zhong and Jannink (2007) and later by Lehermeier et al. (2017) and Osthushenrich et al. (2017). Although these approaches avoid stochastic simulation, they still require evaluating quadratic forms for each cross, which can remain computationally intensive for large-scale applications. As observed in the comparisons presented here, existing implementations are often not sufficiently scalable when the full space of potential crosses is considered. To reduce computational burden, many studies apply preselection of candidate parents or crosses, typically based on parental GEBVs or coancestry (Niehoff et al. 2024a; Neyhart et al. 2019; Lehermeier et al. 2017; Allier et al. 2019). Such filtering may exclude crosses combining moderate expected means with high segregation variance, because segregation variance depends on trait-relevant allelic differences between parents rather than genome-wide relatedness alone. The C++ implementation in CrossingTools substantially reduced computation time in the present benchmark, enabling exhaustive evaluation of large candidate-cross sets without prior filtering.

For multi-trait objectives, computational demand can be reduced by calculating the segregation variance of the selection index directly, thereby avoiding explicit construction of the complete trait-by-trait segregation covariance matrix (Niehoff et al. 2025). However, the full matrix is required when cross-specific variances and covariances among individual traits are of interest (Neyhart et al. 2019). CrossingTools implements both approaches. For applications involving many traits, the reduced-rank formulation described in appendix B provides a further approximation, provided that the retained dimensions adequately capture the underlying covariance structure.

## 4.8 Current limitations and future extensions

A limitation of CrossingTools is that it currently supports at most four parental haplotypes. This includes biparental crosses in inbred lines (Osthushenrich et al. 2017; Lehermeier et al. 2017; Zhong and Jannink 2007), four- and three-way crosses (Allier et al. (2019), four-way extension of Osthushenrich et al. (2017) given in appendix A), or crosses between heterozygous individuals (Wolfe et al. 2021). More complex crossing designs, such as eight-way crosses, have been described in the literature (Niehoff et al. 2024b), but their implementation requires substantially more complex and computationally demanding formulations of segregation variance, which are not currently supported in CrossingTools. The current framework is also restricted to diploid species.

Extensions to autopolyploids require modelling segregation and recombination among larger numbers of homologous chromosomes, for which cross-specific variance formulations have only recently been proposed (Fernández-González et al. 2026; Wu et al. 2026). Support for autopolyploid crops therefore represents an important direction for future development.

# 5 Conclusion

Cross evaluation criteria incorporating segregation variance produced similar population mean gain to selection based on expected cross means but improved upper-tail performance, which is particularly relevant for variety development. SPV-based selection preserved more genetic variance and, under the simulated multi-trait scenarios, shifted genetic correlations in directions more favourable to the selection objective. These results demonstrate that breeding strategies should be evaluated using both population-level and commercialisable gain, because selection indices that effectively direct mean response may produce less desirable changes in the best-performing individuals for particular traits. OCS further reduced genetic variance depletion, although at the cost of lower short-term gain. These analyses were enabled by CrossingTools, a scalable R package integrating alternative cross performance criteria, multi-trait index selection, gain–diversity optimisation, and practical mating constraints. By supporting exhaustive evaluation of large candidate-cross sets, CrossingTools provides a flexible framework for translating genomic predictions into crossing decisions across a broad range of plant breeding schemes.

# Appendix A Segregation (co)variance models implemented in CrossingTools

CrossingTools predicts the genetic variance and covariance generated among progeny from parental marker genotypes, estimated marker effects, and pairwise recombination frequencies. The implemented calculations follow previously published segregation variance models for biparental and four-way crosses and for crosses between phased heterozygous parents.

## A.1 Crosses between inbred parents

For a cross between two fully inbred parents $i$ and $j$, let $m_{i,k}, m_{j,k} \in \{0, 2\}$ denote their allele dosages at locus $k$, and let $m_{i\times j,k}$ denote the allele dosage of a random progeny individual. The corresponding centred additive genotype codes are

$$\begin{aligned} w_{i,k} &= m_{i,k} - 2p_k, \\ w_{j,k} &= m_{j,k} - 2p_k, \\ w_{i\times j,k} &= m_{i\times j,k} - 2p_k. \end{aligned}$$

where $p_k$ is the frequency of the counted allele at locus $k$ in the reference population. With average allele-substitution effects $\boldsymbol{\alpha}$, the breeding value of a progeny individual is

$$BV_{i\times j} = \mathbf{W}_{i\times j}\boldsymbol{\alpha}.$$

The additive segregation variance, i.e. the within-family variance of breeding values, is therefore

$$\begin{aligned}\sigma^2_{\mathrm{A},i\times j} &= \mathrm{Var}(BV_{i\times j}) = \boldsymbol{\alpha}^\top \mathrm{Var}(\mathbf{W}_{i\times j})\boldsymbol{\alpha} \\ &= \sum_k \sum_l \alpha_k \ \mathrm{Cov}(w_{i\times j,k}, w_{i\times j,l})\, \alpha_l.\end{aligned}$$

Because centring by $2p_k$ only subtracts a constant, the within-family covariance is unchanged,

$$\mathrm{Cov}(w_{i\times j,k}, w_{i\times j,l}) = \mathrm{Cov}(m_{i\times j,k}, m_{i\times j,l}).$$

Furthermore, the difference between the parental centred genotype codes is

$$w_{i,k} - w_{j,k} = m_{i,k} - m_{j,k}.$$

Accordingly,

$$\mathrm{Cov}(w_{i\times j,k}, w_{i\times j,l}) = (m_{i,k} - m_{j,k})(m_{i,l} - m_{j,l})c_{kl}, \tag{A.1}$$

where $c_{kl}$ is a covariance-transmission coefficient determined by the recombination frequency $r_{kl}$ and the offspring type.

For RIL progeny, the formulation of Osthushenrich et al. (2017) accounts for $n_{\mathrm{RI}}$ generations of random intermating before line development through the covariance coefficient

$$c_{kl}^{\mathrm{RIL}(n_{\mathrm{RI}})} = \frac{1}{4}\frac{1-2r_{kl}}{1+2r_{kl}}(1-r_{kl})^{n_{\mathrm{RI}}}.$$

Alternatively, CrossingTools implements the formulation of Lehermeier et al. (2017), which explicitly models $g$ selfing steps, with coefficient

$$c_{kl}^{\mathrm{RIL}(g)} = \frac{1}{4}\frac{1-2r_{kl}}{1+2r_{kl}}\left[1 - 2^{-g}(1-2r_{kl})^g\right].$$

For DH progeny, analogous formulations are implemented. Following Osthushenrich et al. (2017), the coefficient after $n_{\mathrm{RI}}$ generations of random intermating is

$$c_{kl}^{\mathrm{DH}(n_{\mathrm{RI}})} = \frac{1}{4}(1-2r_{kl})(1-r_{kl})^{n_{\mathrm{RI}}}.$$

Following Lehermeier et al. (2017), the coefficient for DH production in generation $g$ is

$$c_{kl}^{\mathrm{DH}(g)} = \frac{1}{4}\left[1 - \frac{4r_{kl}}{1+2r_{kl}}\left(1 - 2^{-g}(1-2r_{kl})^g\right)\right].$$

Thus, for both RIL and DH progeny, users can choose between the Osthushenrich et al. (2017) ($n_{\mathrm{RI}}$ random-intermating generations) and Lehermeier et al. (2017) ($g$ rounds of selfing), depending on how progeny are intended to be produced.

## A.2 Four-way crosses

CrossingTools also predicts segregation variances and covariances for four-way crosses of the form

$$(P_1 \times P_2) \times (P_3 \times P_4).$$

Following Allier et al. (2019), covariance in a four-way population is decomposed into two within-block components, originating from the initial biparental crosses $P_1 \times P_2$ and $P_3 \times P_4$, and one across-block component associated with the subsequent cross between their $F_1$ progeny.

Let $m_{P_a,k} \in \{0, 2\}$ denote the allele dosage of parent $P_a$, with $a \in \{1, 2, 3, 4\}$, at locus $k$, and let $X_k$ denote the allele dosage of a random four-way progeny individual at locus $k$. The corresponding centred additive genotype code is

$$W_k = X_k - 2p_k,$$

where $p_k$ is the frequency of the counted allele at locus $k$ in the reference population. With average allele-substitution effects $\boldsymbol{\alpha}$, the breeding value of a random four-way progeny individual is

$$BV = \mathbf{W}\boldsymbol{\alpha}.$$

The additive segregation variance is therefore

$$\sigma^2_{\mathrm{A}} = \mathrm{Var}(BV) = \boldsymbol{\alpha}^\top \mathrm{Var}(\mathbf{W})\boldsymbol{\alpha}.$$

Because centring only subtracts constants,

$$\mathrm{Cov}(W_k, W_l) = \mathrm{Cov}(X_k, X_l).$$

Thus, the covariance between centred progeny genotype codes can be obtained directly from the covariance of the corresponding allele dosages:

$$\begin{aligned} \mathrm{Cov}(W_k, W_l) = \mathrm{Cov}(X_k, X_l) = \\ \frac{1}{4}\Big\{ c^{\mathrm{W}}_{kl}\Big[(m_{P_1,k} - m_{P_2,k})(m_{P_1,l} - m_{P_2,l}) \\ +(m_{P_3,k} - m_{P_4,k})(m_{P_3,l} - m_{P_4,l})\Big] \\ +c^{\mathrm{B}}_{kl}\Big[m_{P_1,k} + m_{P_2,k} - m_{P_3,k} - m_{P_4,k}\Big] \\ \times\Big[m_{P_1,l} + m_{P_2,l} - m_{P_3,l} - m_{P_4,l}\Big]\Big\}. \end{aligned} \tag{A.2}$$

Here, $c^{\mathrm{W}}_{kl}$ and $c^{\mathrm{B}}_{kl}$ are the recombination-dependent transmission coefficients for the within-block and across-block components, respectively.

For RIL progeny obtained after $g$ generations of selfing, CrossingTools uses the coefficients of Allier et al. (2019), extending Lehermeier et al. (2017). The across-block coefficient is

$$c_{kl}^{\mathrm{B,RIL}(g)} = \frac{1}{4}\frac{1-2r_{kl}}{1+2r_{kl}}\left[1-2^{-g}(1-2r_{kl})^{g}\right].$$

The coefficient for both within-block components is

$$c_{kl}^{\mathrm{W,RIL}(g)} = \frac{1}{4}\frac{1-2r_{kl}}{1+2r_{kl}}\left[2-2^{-g}(1-2r_{kl})^{g+1}\right].$$

For progeny development with $n_{\mathrm{RI}}$ generations of random intermating, following Osthushenrich et al. (2017), the across-block coefficient is

$$c_{kl}^{\mathrm{B,RIL}(n_{\mathrm{RI}})} = \frac{1}{4}\frac{1-2r_{kl}}{1+2r_{kl}}(1-r_{kl})^{n_{\mathrm{RI}}}.$$

The coefficient for both within-block components is

$$c_{kl}^{\mathrm{W,RIL}(n_{\mathrm{RI}})} = \frac{1}{2}\frac{1-2r_{kl}}{1+2r_{kl}}(1-r_{kl})^{n_{\mathrm{RI}}}.$$

For DH lines derived after $g-1$ generations of selfing, with $g=1$ corresponding to DH production directly from the four-way $F_1$, CrossingTools uses the coefficients of Allier et al. (2019), extending Lehermeier et al. (2017). The across-block coefficient is

$$c_{kl}^{\mathrm{B,DH}(g)} = \frac{1}{4}\frac{1-2r_{kl}}{1+2r_{kl}}\left[1+4r_{kl}2^{-g}(1-2r_{kl})^{g-1}\right].$$

The coefficient for both within-block components is

$$c_{kl}^{\mathrm{W,DH}(g)} = \frac{1}{2}\frac{1-2r_{kl}}{1+2r_{kl}}\left[1+2r_{kl}2^{-g}(1-2r_{kl})^{g}\right].$$

Extending Osthushenrich et al. (2017), the corresponding coefficients after $n_{\mathrm{RI}}$ generations of random intermating are defined analogously. The across-block coefficient is

$$c_{kl}^{\mathrm{B,DH}(n_{\mathrm{RI}})} = \frac{1}{4}(1-2r_{kl})(1-r_{kl})^{n_{\mathrm{RI}}}.$$

The coefficient for both within-block components is

$$c_{kl}^{\mathrm{W,DH}(n_{\mathrm{RI}})} = \frac{1}{2}(1-2r_{kl})(1-r_{kl})^{n_{\mathrm{RI}}+1}.$$

Depending on the selected progeny type and progeny development framework, the corresponding coefficients are substituted for $c_{kl}^{\mathrm{W}}$ and $c_{kl}^{\mathrm{B}}$ in Eq. A.2.

## A.3 Heterozygous parents and dominance effects

For crosses between two phased heterozygous parents, CrossingTools predicts the within-family variance of breeding values, dominance deviations, and total genetic values following Bonk et al. (2016).

Let $G_{1,k}$ and $G_{2,k}$ denote the two haplotypes of the first parent at locus $k$, and $G_{3,k}$ and $G_{4,k}$ those of the second parent, with $G_{h,k} \in \{0,1\}$. Let $A_k$ and $B_k$ denote the alleles transmitted by the first and second parent, respectively. The offspring allele dosage and heterozygosity indicator are

$$X_k = A_k + B_k, \qquad H_k = A_k + B_k - 2A_kB_k.$$

As for crosses between inbred parents, the centred additive genotype code is

$$W_k = X_k - 2p_k,$$

where $p_k$ is the frequency of the counted allele at locus $k$ in the reference population and $q_k = 1 - p_k$. The corresponding statistical dominance deviation code is

$$Z_k = H_k - (1 - 2p_k)W_k - 2p_kq_k.$$

With average allele-substitution effects $\boldsymbol{\alpha}$ and dominance effects $\mathbf{d}$, the breeding value, dominance deviation, and total genetic value of a progeny individual are

$$BV = \mathbf{W}\boldsymbol{\alpha}, \qquad DD = \mathbf{Z}\mathbf{d}, \qquad TGV = BV + DD.$$

Consequently, prediction of the within-family variance of total genetic values requires the additive variance, dominance variance, and additive–dominance covariance derived below.

### A.3.1 Additive–additive covariance

The within-parent gametic covariances are

$$\begin{aligned}
\operatorname{Cov}(A_k, A_l) &= \frac{1 - 2r_{kl}}{4}(G_{1,k} - G_{2,k})(G_{1,l} - G_{2,l}),\\
\operatorname{Cov}(B_k, B_l) &= \frac{1 - 2r_{kl}}{4}(G_{3,k} - G_{4,k})(G_{3,l} - G_{4,l}).
\end{aligned}$$

Because gametes transmitted by the two parents are independent, the covariance between offspring allele dosages is the sum of the two within-parent gametic covariances. Because centring only subtracts constants, this is also the covariance between the centred additive genotype codes,

$$\begin{aligned}
\operatorname{Cov}(W_k, W_l) &= \operatorname{Cov}(X_k, X_l)\\
&= \frac{1 - 2r_{kl}}{4}\Big[(G_{1,k} - G_{2,k})(G_{1,l} - G_{2,l})\\
&\qquad + (G_{3,k} - G_{4,k})(G_{3,l} - G_{4,l})\Big].
\end{aligned} \tag{A.3}$$

The within-family variance of breeding values is therefore

$$\begin{aligned}
\sigma^2_{\mathrm{A}} &= \operatorname{Var}(BV) = \boldsymbol{\alpha}^{\top} \operatorname{Var}(\mathbf{W})\boldsymbol{\alpha}\\
&= \sum_k \sum_l \alpha_k \operatorname{Cov}(W_k, W_l)\, \alpha_l.
\end{aligned}$$

### A.3.2 Dominance–dominance covariance

The expected transmitted alleles are

$$\mathbb{E}(A_k) = \frac{G_{1,k} + G_{2,k}}{2}, \qquad \mathbb{E}(B_k) = \frac{G_{3,k} + G_{4,k}}{2}.$$

Using $H_k = A_k + B_k - 2A_kB_k$ and the independence of the alleles transmitted by the two parents, the covariance between heterozygosity indicators at loci $k$ and $l$ is

$$\begin{aligned}\operatorname{Cov}(H_k, H_l) = {} & \operatorname{Cov}(A_k, A_l)\,[1 - 2\mathbb{E}(B_k)]\,[1 - 2\mathbb{E}(B_l)] \\ & + \operatorname{Cov}(B_k, B_l)\,[1 - 2\mathbb{E}(A_k)]\,[1 - 2\mathbb{E}(A_l)] \\ & + 4\operatorname{Cov}(A_k, A_l)\operatorname{Cov}(B_k, B_l).\end{aligned}$$

Defining

$$s_k = 1 - 2p_k,$$

and noting that the constant term $2p_kq_k$ in $Z_k$ does not contribute to covariances, the covariance between dominance deviation is

$$\begin{aligned}\operatorname{Cov}(Z_k, Z_l) = {} & \operatorname{Cov}(H_k, H_l) - s_l \operatorname{Cov}(H_k, W_l) \\ & - s_k \operatorname{Cov}(W_k, H_l) + s_k s_l \operatorname{Cov}(W_k, W_l).\end{aligned} \tag{A.4}$$

The required covariances between the centred additive genotype code and the heterozygosity indicator are

$$\begin{aligned}\operatorname{Cov}(W_k, H_l) = {} & \operatorname{Cov}(A_k, A_l)\,[1 - 2\mathbb{E}(B_l)] \\ & + \operatorname{Cov}(B_k, B_l)\,[1 - 2\mathbb{E}(A_l)]\,,\end{aligned}$$

and

$$\begin{aligned}\operatorname{Cov}(H_k, W_l) = {} & \operatorname{Cov}(A_k, A_l)\,[1 - 2\mathbb{E}(B_k)] \\ & + \operatorname{Cov}(B_k, B_l)\,[1 - 2\mathbb{E}(A_k)]\,.\end{aligned}$$

Using the dominance deviation covariance in Eq. A.4, the within-family dominance deviation variance is consequently

$$\begin{aligned}\sigma^2_{\mathrm{D}} = \operatorname{Var}(DD) &= \mathbf{d}^\top \operatorname{Var}(\mathbf{Z})\mathbf{d} \\ &= \sum_k \sum_l d_k \operatorname{Cov}(Z_k, Z_l)\, d_l.\end{aligned}$$

### A.3.3 Additive–dominance covariance

The statistical additive and dominance components are not necessarily uncorrelated within a biparental family. From the definition of the statistical dominance deviation code,

$$Z_l = H_l - s_l W_l - 2p_lq_l, \qquad s_l = 1 - 2p_l,$$

and because the final term is constant,

$$\operatorname{Cov}(W_k, Z_l) = \operatorname{Cov}(W_k, H_l) - s_l \operatorname{Cov}(W_k, W_l).$$

Using the gametic covariances defined above, this can be written explicitly as

$$\begin{aligned}\operatorname{Cov}(W_k, Z_l) = &\operatorname{Cov}(A_k, A_l)\,[1 - 2\mathbb{E}(B_l) - s_l] \\ &+ \operatorname{Cov}(B_k, B_l)\,[1 - 2\mathbb{E}(A_l) - s_l]\,.\end{aligned}$$

Since $s_l = 1 - 2p_l$, an equivalent expression is

$$\begin{aligned}\operatorname{Cov}(W_k, Z_l) = &\,2\operatorname{Cov}(A_k, A_l)\,[p_l - \mathbb{E}(B_l)] \\ &+ 2\operatorname{Cov}(B_k, B_l)\,[p_l - \mathbb{E}(A_l)]\,.\end{aligned}$$

Similarly,

$$\begin{aligned}\operatorname{Cov}(Z_k, W_l) = &\operatorname{Cov}(A_k, A_l)\,[1 - 2\mathbb{E}(B_k) - s_k] \\ &+ \operatorname{Cov}(B_k, B_l)\,[1 - 2\mathbb{E}(A_k) - s_k]\,,\end{aligned} \tag{A.5}$$

where $s_k = 1 - 2p_k$. Equivalently,

$$\begin{aligned}\operatorname{Cov}(Z_k, W_l) = &\,2\operatorname{Cov}(A_k, A_l)\,[p_k - \mathbb{E}(B_k)] \\ &+ 2\operatorname{Cov}(B_k, B_l)\,[p_k - \mathbb{E}(A_k)]\,.\end{aligned} \tag{A.6}$$

The within-family covariance between breeding value and dominance deviation is therefore

$$\begin{aligned}\sigma_{\mathrm{AD}} = \operatorname{Cov}(BV, DD) &= \boldsymbol{\alpha}^{\top} \operatorname{Cov}(\mathbf{W}, \mathbf{Z})\mathbf{d} \\ &= \sum_k \sum_l \alpha_k \operatorname{Cov}(W_k, Z_l)\, d_l.\end{aligned} \tag{A.7}$$

Although statistical additive and dominance effects may have been defined orthogonal in the reference population under the statistical parameterisation, this does not generally imply that breeding values and dominance deviations are uncorrelated within a particular cross.

### A.3.4 Total genetic variance

Combining the three components above gives the within-family segregation variance of total genetic values,

$$\begin{aligned}\sigma^2_{\mathrm{TGV}} = \operatorname{Var}(TGV) &= \sigma^2_{\mathrm{A}} + \sigma^2_{\mathrm{D}} + 2\sigma_{\mathrm{AD}} \\ &= \operatorname{Var}(BV) + \operatorname{Var}(DD) + 2\operatorname{Cov}(BV, DD).\end{aligned} \tag{A.8}$$

The total genetic segregation variance in Eq. A.8, together with the expected total genetic value of the cross, is used to calculate the SPV implemented in CrossingTools following Bonk et al. (2016).

# Appendix B Segregation variances under a reduced-rank model

Modern breeding programmes routinely optimise multiple traits across multiple environments, leading to situations where the dimensionality of the trait space becomes a computational bottleneck. In particular, the computation of segregation (co)variance matrices scales quadratically with the number of traits, and some applications require access to the full segregation covariance matrix, as in Neyhart et al. (2019). To reduce the computational burden, one strategy is to operate in a reduced-rank space by representing trait-specific marker effects through a smaller set of latent components.

One way to obtain such a reduced representation is through the singular value decomposition (SVD) of the marker-effect matrix,

$$\mathbf{U} = \mathbf{L}\mathbf{D}\mathbf{R}^\top, \tag{B.1}$$

where $\mathbf{L}$ contains the left singular vectors, $\mathbf{R}$ contains the right singular vectors, and $\mathbf{D} = \text{diag}(\rho_1, \rho_2, \ldots, \rho_S)$ is the diagonal matrix of singular values, with $S = \min(m, T)$.

Substituting Eq. B.1 into Eq. 10 yields

$$\text{Var}(\mathbf{M}_{i\times j}\mathbf{U}) = (\mathbf{L}\mathbf{D}\mathbf{R}^\top)^\top \text{Var}(\mathbf{M}_{i\times j})(\mathbf{L}\mathbf{D}\mathbf{R}^\top).$$

Rearranging gives

$$\text{Var}(\mathbf{M}_{i\times j}\mathbf{U}) = \mathbf{R}\mathbf{D}\mathbf{L}^\top \text{Var}(\mathbf{M}_{i\times j})\mathbf{L}\mathbf{D}\mathbf{R}^\top. \tag{B.2}$$

To obtain a reduced-rank approximation, let $\mathbf{U}_q = \mathbf{L}_q\mathbf{D}_q\mathbf{R}_q^\top$ denote the truncated SVD retaining the top $q$ singular components, where $q < S$. The corresponding segregation covariance matrix under the rank-$q$ approximation is

$$\text{Var}(\mathbf{M}_{i\times j}\mathbf{U}_q) = \mathbf{R}_q\mathbf{D}_q\mathbf{L}_q^\top \text{Var}(\mathbf{M}_{i\times j})\mathbf{L}_q\mathbf{D}_q\mathbf{R}_q^\top. \tag{B.3}$$

Eq. B.3 shows that, once segregation (co)variances are calculated in the reduced latent space defined by $\mathbf{L}_q$, an approximation to the full trait-space segregation covariance matrix can be reconstructed using $\mathbf{R}_q$ and $\mathbf{D}_q$ together with the segregation covariance of the reduced components.

The truncated SVD provides an exact reconstruction of Eq. B.2 when the true effect matrix $\mathbf{U}$ has rank no greater than $q$. In that case, $\mathbf{U}_q = \mathbf{U}$ and the reduced formulation spans the same column space as the full model, ensuring that

$$\text{Var}(\mathbf{M}_{i\times j}\mathbf{U}_q) = \text{Var}(\mathbf{M}_{i\times j}\mathbf{U}).$$

If, however, $\mathbf{U}$ has rank greater than $q$, then $\mathbf{U}_q$ spans only a subspace of the original column space. The reconstructed matrix in Eq. B.3 is therefore an approximation whose accuracy depends on how well the retained $q$ singular components capture variation in the QTL effects.

An analogous approach could be applied using factor-analytic models, in which segregation (co)variances are computed in the space of common factors and subsequently reconstructed using the factor loadings. As with the SVD-based approximation, this relies on the retained factors adequately capturing the genetic effects. If substantial

trait-specific variation remains unexplained by the common factors, the factor-analytic representation does not span the full column space of $\mathbf{U}$, and the reconstructed segregation covariance matrix will not capture the complete trait-space segregation covariance structure.

# Appendix C Threshold probabilities, expected maxima, and family size

Under a normal approximation to the distribution of genetic values within a family, the expected cross mean and segregation variance can be used to derive several family-size-dependent measures of cross potential. These include the probability of recovering at least one progeny exceeding a predefined threshold, the expected value of the best progeny, and the family size required to achieve a desired probability of exceeding that threshold.

Let the genetic value of a randomly sampled progeny from cross $i \times j$ be denoted by $X$, with

$$X \sim \mathcal{N}\left(\mu_{i\times j}, \sigma^2_{i\times j}\right),$$

where $\mu_{i\times j}$ is the expected cross mean and $\sigma^2_{i\times j}$ is the segregation variance generated by the cross. The probability that a single progeny exceeds a predefined threshold $x$ follows directly from the cumulative distribution function of the normal distribution and is given by

$$\Pr(X > x) = 1 - \Phi(z),$$

where $\Phi$ is the cumulative distribution function of the standard normal distribution and

$$z = \frac{x - \mu_{i\times j}}{\sigma_{i\times j}}.$$

For a family of $N$ independently sampled progeny, the probability of obtaining at least one individual with a genetic value exceeding $x$ is given by Danguy des Déserts et al. (2023) as

$$\Pr\left(\max(X_1, \ldots, X_N) > x\right) = 1 - \Phi(z)^N. \tag{C.1}$$

Similarly, the expected maximum genetic value among $N$ family members is given by Müller et al. (2018) as

$$\mathbb{E}\left[\max(X_1, \ldots, X_N)\right] = \mu_{i\times j} + \sigma_{i\times j}\iota_N, \tag{C.2}$$

where $\iota_N$ denotes the expected maximum among $N$ standardised normal outcomes,

$$\iota_N = \mathbb{E}\left[W_{(N)}\right] = \mathbb{E}\left[\max(W_1, \ldots, W_N)\right],$$
$$W_1, \ldots, W_N \sim \mathcal{N}(0, 1).$$

Thus, $\iota_N$ represents the expected number of standard deviations by which the best individual among $N$ family members exceeds the family mean and depends only on the family size $N$.

It may also be useful to determine the family size required to obtain at least one individual exceeding the predefined threshold $x$. For a desired probability $\pi$, this is obtained by setting

$$\Pr\left(\max(X_1, \ldots, X_N) > x\right) = \pi.$$

Since

$$\Pr\left(\max(X_1, \ldots, X_N) > x\right) = 1 - \Phi(z)^N,$$

solving for $N$ yields

$$N = \frac{\log(1-\pi)}{\log\left[\Phi(z)\right]} = \frac{\log(1-\pi)}{\log\left[\Phi\left(\frac{x-\mu_{i\times j}}{\sigma_{i\times j}}\right)\right]}. \tag{C.3}$$

Because family size must be an integer, the resulting value of $N$ should be rounded up to the next whole number.

# References


Allier, A., Lehermeier, C., Charcosset, A., Moreau, L., Teyssèdre, S.: Improving Short- and Long-Term Genetic Gain by Accounting for Within-Family Variance in Optimal Cross-Selection. Frontiers in Genetics **10** (2019) https://doi.org/10.3389/fgene.2019.01006

Allier, A., Moreau, L., Charcosset, A., Teyssèdre, S., Lehermeier, C.: Usefulness Criterion and Post-selection Parental Contributions in Multi-parental Crosses: Application to Polygenic Trait Introgression. G3 Genes|Genomes|Genetics **9**(5), 1469–1479 (2019) https://doi.org/10.1534/g3.119.400129

Akdemir, D., Sánchez, J.I.: Efficient Breeding by Genomic Mating. Frontiers in Genetics **7** (2016) https://doi.org/10.3389/fgene.2016.00210

Browning, S.R., Browning, B.L.: Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype Clustering. The American Journal of Human Genetics **81**(5), 1084–1097 (2007) https://doi.org/10.1086/521987

Butler, D.G., Cullis, B.R., Gilmour, A.R., Gogel, B.J., Thompson, R.: ASReml-R Reference Manual Version 4.2. VSN International Ltd., Hemel Hempstead, UK (2023). VSN International Ltd.

Bernardo, R.: Prediction of Maize Single-Cross Performance Using RFLPs and Information from Related Hybrids. Crop Science **34**(1), 1994–0011183003400010003 (1994) https://doi.org/10.2135/cropsci1994.0011183X003400010003x

Bernardo, R.: Genomewide Selection of Parental Inbreds: Classes of Loci and Virtual Biparental Populations. Crop Science **54**(6), 2586–2595 (2014) https://doi.org/10.2135/cropsci2014.01.0088

Batista, L.G., Gaynor, R.C., Margarido, G.R.A., Byrne, T., Amer, P., Gorjanc, G., Hickey, J.M.: Long-term comparison between index selection and optimal independent culling in plant breeding programs with genomic prediction. PLOS ONE **16**(5), 0235554 (2021) https://doi.org/10.1371/journal.pone.0235554

Brascamp, E.W.: Selection indices with constraints. Animal Breeding Abstracts **52**, 645–654 (1984)

Bonk, S., Reichelt, M., Teuscher, F., Segelke, D., Reinsch, N.: Mendelian sampling covariability of marker effects and genetic values. Genetics Selection Evolution **48**(1), 36 (2016) https://doi.org/10.1186/s12711-016-0214-0

Bijma, P., Wientjes, Y.C.J., Calus, M.P.L.: Breeding Top Genotypes and Accelerating Response to Recurrent Selection by Selecting Parents with Greater Gametic Variance. Genetics **214**(1), 91–107 (2020) https://doi.org/10.1534/genetics.119.302643

Cole, J.b., VanRaden, P.m.: Use of haplotypes to estimate Mendelian sampling effects and selection limits. Journal of Animal Breeding and Genetics **128**(6), 446–455 (2011) https://doi.org/10.1111/j.1439-0388.2011.00922.x

Danguy des Déserts, A., Durand, N., Servin, B., Goudemand-Dugué, E., Alliot, J.-M., Ruiz, D., Charmet, G., Elsen, J.-M., Bouchet, S.: Comparison of genomic-enabled cross selection criteria for the improvement of inbred line breeding populations. G3 Genes|Genomes|Genetics **13**(11), 195 (2023) https://doi.org/10.1093/g3journal/jkad195

Daetwyler, H.D., Hayes, B.J., Robbins, K., Hayden, M.J., Spangenberg, G.: Selection Based on Optimal Haploid Value to Create Elite Lines. US Patent US11744199B2. Assignee: Corteva Agriscience LLC (2023). https://patents.google.com/patent/US11744199B2/en

Daetwyler, H.D., Hayden, M.J., Spangenberg, G.C., Hayes, B.J.: Selection on Optimal Haploid Value Increases Genetic Gain and Preserves More Genetic Diversity Relative to Genomic Selection. Genetics **200**(4), 1341–1348 (2015) https://doi.org/10.1534/genetics.115.178038

Dudley, J.W., Lambert, R.J.: 5. 100 Generations of Selection for Oil and Protein in Corn, pp. 79–110. John Wiley & Sons, Ltd, ??? (2003). https://doi.org/10.1002/9780470650240.ch5 . https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470650240.ch5

Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation **6**(2), 182–197 (2002) https://doi.org/10.1109/4235.996017

Duvick, D.: Genetic progress in yield of United States maize (Zea mays L.). Maydica

(2005)

Falconer, D.S., Mackay, T.: Introduction to Quantitative Genetics, 4. ed., [16. print.] edn. Pearson, Prentice Hall, Harlow (2009)

Fernández-González, J., Metwally, S.M., Isidro y Sánchez, J.: MateR: A novel genomic mating framework. Genetics **232**(4), 013 (2026) https://doi.org/10.1093/genetics/iyag013

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 **57**(5), 2372–2386 (2017) https://doi.org/10.2135/cropsci2016.09.0742

Gorjanc, G., Gaynor, R.C., Hickey, J.M.: Optimal cross selection for long-term genetic gain in two-part programs with rapid recurrent genomic selection. Theoretical and Applied Genetics **131**(9), 1953–1966 (2018) https://doi.org/10.1007/s00122-018-3125-3

Gaynor, R.C., Gorjanc, G., Hickey, J.M.: AlphaSimR: An R package for breeding program simulations. G3 Genes|Genomes|Genetics **11**(2), 017 (2021) https://doi.org/10.1093/g3journal/jkaa017

Gorjanc, G., Hickey, J.M.: Alphamate: a program for optimizing selection, maintenance of diversity and mate allocation in breeding programs. Bioinformatics **34**(19), 3408–3411 (2018)

Haldane, J.: The combination of linkage values, and the calculation of distances between the loci of linked factors. Journal of Genetics **8**, 299 (1919)

Hazel, L.N.: The genetic basis for constructing selection indexes. Genetics **28**(6), 476–490 (1943)

Hallauer, A.R., Carena, M.J., Miranda Filho, J.B.: Quantitative Genetics in Maize Breeding, 3rd ed. edn. Handbook of Plant Breeding, vol. v. 6. Springer, New York London (2010)

Hazel, L.N., Lush, J.L.: THE EFFICIENCY OF THREE METHODS OF SELECTION*. Journal of Heredity **33**(11), 393–399 (1942) https://doi.org/10.1093/oxfordjournals.jhered.a105102

Haldane, J.B.S., Waddington, C.H.: INBREEDING AND LINKAGE. Genetics **16**(4), 357–374 (1931) https://doi.org/10.1093/genetics/16.4.357

Kinghorn, B.: On computing strategies for mate allocation. Journal of Animal Breeding and Genetics **104**(1-5), 12–22 (1987) https://doi.org/10.1111/j.1439-0388.1987.tb00104.x

Kinghorn, B.P.: An algorithm for efficient constrained mate selection. Genetics Selection Evolution **43**(1), 4 (2011) https://doi.org/10.1186/1297-9686-43-4

Lehermeier, C., de los Campos, G., Wimmer, V., Schön, C.-C.: Genomic variance estimates: With or without disequilibrium covariances? Journal of Animal Breeding and Genetics **134**(3), 232–241 (2017) https://doi.org/10.1111/jbg.12268

Lehermeier, C., Teyssèdre, S., Schön, C.-C.: Genetic Gain Increases by Applying the Usefulness Criterion with Improved Variance Prediction in Selection of Crosses. Genetics **207**(4), 1651–1661 (2017) https://doi.org/10.1534/genetics.117.300403

Meuwissen, T.: Maximizing the response of selection with a predefined rate of inbreeding. Journal of animal science **75**(4), 934–940 (1997)

Melchinger, A.E., Frisch, M.: Genomic prediction in hybrid breeding: II. Reciprocal recurrent genomic selection with full-sib and half-sib families. Theoretical and Applied Genetics **136**(9), 203 (2023) https://doi.org/10.1007/s00122-023-04446-3

Meuwissen, T.H.E., Hayes, B.J., Goddard, M.E.: Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. Genetics **157**(4), 1819–1829 (2001) https://doi.org/10.1093/genetics/157.4.1819

Müller, D., Schopp, P., Melchinger, A.E.: Selection on Expected Maximum Haploid Breeding Values Can Increase Genetic Gain in Recurrent Genomic Selection. G3 Genes|Genomes|Genetics **8**(4), 1173–1181 (2018) https://doi.org/10.1534/g3.118.200091

Mohammadi, M., Tiede, T., Smith, K.P.: PopVar: A Genome-Wide Procedure for Predicting Genetic Variance and Correlated Response in Biparental Breeding Populations. Crop Science **55**(5), 2068–2077 (2015) https://doi.org/10.2135/cropsci2015.01.0030

Neyhart, J.L., Lorenz, A.J., Smith, K.P.: Multi-trait Improvement by Predicting Genetic Correlations in Breeding Crosses. G3 Genes|Genomes|Genetics **9**(10), 3153–3165 (2019) https://doi.org/10.1534/g3.119.400406

Niehoff, T.A.M., ten Napel, J., Bijma, P., Pook, T., Wientjes, Y.C.J., Hegedűs, B., Calus, M.P.L.: Improving selection decisions with mating information by accounting for Mendelian sampling variances looking two generations ahead. Genetics Selection Evolution **56**(1), 41 (2024) https://doi.org/10.1186/s12711-024-00899-2

Niehoff, T.A.M., ten Napel, J., Calus, M.P.L.: Prediction of additive genetic variances of descendants for complex families based on Mendelian sampling variances. G3 Genes|Genomes|Genetics **14**(11), 205 (2024) https://doi.org/10.1093/g3journal/jkae205

Niehoff, T.A.M., ten Napel, J., Calus, M.P.L.: Practical Considerations When Using

Mendelian Sampling Variances for Selection Decisions in Genomic Selection Programs. Journal of Animal Breeding and Genetics **142**(4), 419–437 (2025) https://doi.org/10.1111/jbg.12913

Osthushenrich, T., Frisch, M., Herzog, E.: Genomic selection of crossing partners on basis of the expected mean and variance of their derived lines. PLOS ONE **12**(12), 0188839 (2017) https://doi.org/10.1371/journal.pone.0188839

Okoye, U.J., Frisch, M., Herzog, E.: Optimization of family sizes in sets of crosses with a greedy allocation strategy based on automatic differentiation. Frontiers in Plant Science **17** (2026) https://doi.org/10.3389/fpls.2026.1727383

Osthushenrich, T., Frisch, M., Zenke-Philippi, C., Jaiser, H., Spiller, M., Cselényi, L., Krumnacker, K., Boxberger, S., Kopahnke, D., Habekuß, A., Ordon, F., Herzog, E.: Prediction of Means and Variances of Crosses With Genome-Wide Marker Effects in Barley. Frontiers in Plant Science **9** (2018) https://doi.org/10.3389/fpls.2018.01899

Peixoto, M.A., Amadeu, R.R., Bhering, L.L., Ferrão, L.F.V., Munoz, P.R., Resende Jr., M.F.R.: SimpleMating: R-package for prediction and optimization of breeding crosses using genomic selection. The Plant Genome **18**(1), 20533 (2025) https://doi.org/10.1002/tpg2.20533

Pesek, J., Baker, R.J.: Comparison of tandem and index selection in the modified pedigree method of breeding self-pollinated species. Canadian Journal of Plant Science **49**(6), 773–781 (1969) https://doi.org/10.4141/cjps69-132

Smith, H.F.: A discriminant function for plant selection. Annals of eugenics **7**(3), 240–250 (1936)

Storn, R., Price, K.: Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces. Journal of Global Optimization **11**(4), 341–359 (1997) https://doi.org/10.1023/A:1008202821328

Schnell, F., Utz, H.: Bericht über die arbeitstagung der vereinigung österreichischer pflanzenzüchter. Gumpenstein: BAL Gumpenstein, 243–8 (1975)

Technow, F., Podlich, D., Cooper, M.: Back to the future 2: The implications of germplasm structure on the balance between short- and long-term genetic gain in a changing target population of environments. G3 Genes|Genomes|Genetics **16**(4), 044 (2026) https://doi.org/10.1093/g3journal/jkag044

VanRaden, P.M.: Efficient Methods to Compute Genomic Predictions. Journal of Dairy Science **91**(11), 4414–4423 (2008) https://doi.org/10.3168/jds.2007-0980

Voss-Fels, K.P., Stahl, A., Wittkop, B., Lichthardt, C., Nagler, S., Rose, T., Chen, T.-W., Zetzsche, H., Seddig, S., Majid Baig, M., Ballvora, A., Frisch, M., Ross, E., Hayes, B.J., Hayden, M.J., Ordon, F., Leon, J., Kage, H., Friedt, W., Stützel,

H., Snowdon, R.J.: Breeding improves wheat productivity under contrasting agrochemical input levels. Nature Plants **5**(7), 706–714 (2019) https://doi.org/10.1038/s41477-019-0445-5

Woolliams, J.a., Berg, P., Dagnachew, B.s., Meuwissen, T.h.e.: Genetic contributions and their optimization. Journal of Animal Breeding and Genetics **132**(2), 89–99 (2015) https://doi.org/10.1111/jbg.12148

Wolfe, M.D., Chan, A.W., Kulakow, P., Rabbi, I., Jannink, J.-L.: Genomic mating in outbred species: Predicting cross usefulness with additive and total genetic covariance matrices. Genetics **219**(3), 122 (2021) https://doi.org/10.1093/genetics/iyab122

Wang, F., Feldmann, M.J., Runcie, D.E.: Why usefulness is rarely useful. G3 Genes|Genomes|Genetics **15**(3), 296 (2025) https://doi.org/10.1093/g3journal/jkae296

Waters, D., Finnegan, C., Weber, S., Guardunia, B., Tolhurst, D.: Designing High Performing and Stable Genotypes through Complementary Crossing Schemes. In Review (2026). https://doi.org/10.21203/rs.3.rs-9793436/v1

Werner, C.R., Gaynor, R.C., Sargent, D.J., Lillo, A., Gorjanc, G., Hickey, J.M.: Genomic selection strategies for clonally propagated crops. Theoretical and Applied Genetics **136**(4), 74 (2023) https://doi.org/10.1007/s00122-023-04300-6

Werner, C.R., Gardner, K.A., Tolhurst, D.J.: Reviving the Desired Gains Index: An Optimal Solution for Parent Selection in Public Plant Breeding Programs. bioRxiv (2024). https://doi.org/10.1101/2024.07.21.603926

Wu, P.-Y., Thelen, K., Hartje, S., Muders, K., Prigge, V., Stich, B., Van Inghelandt, D.: Assessment of segregation variance estimates from derivation, simulations, and empirical data in autotetraploid species exemplified in potato. Genetics **232**(4), 031 (2026) https://doi.org/10.1093/genetics/iyag031

Zhong, S., Jannink, J.-L.: Using Quantitative Trait Loci Results to Discriminate Among Crosses on the Basis of Their Progeny Mean and Variance. Genetics **177**(1), 567–576 (2007) https://doi.org/10.1534/genetics.107.075358

## Statements and Declarations

**Funding.** SW was supported by the Federal Ministry of Research, Technology and Space (BMFTR) under grant number 031B1539. Computational resources for the simulations were provided by the BMFTR-funded de.NBI Cloud within the German Network for Bioinformatics (de.NBI). DT acknowledges funding through his Edinburgh Innovations Fellowship at The Roslin Institute.

**Competing Interests.** SW serves as Social Media Editor for Theoretical and Applied Genetics. The remaining authors declare no conflicts of interest.

**Author Contributions.** SW developed the R package CrossingTools. SW, DT, and CW conceived and designed the study. SW, DJT, and DW drafted the manuscript. DW and CW contributed practical breeding perspectives and critically reviewed the manuscript.

**Data Availability.** Not applicable.

**Code Availability.** The R package CrossingTools is available on CRAN and GitHub (https://github.com/svenernstW/CrossingTools). Example R scripts demonstrating the methods and workflows presented in this study are available in the GitHub repository (https://github.com/svenernstW/CrossingTools/tree/main/code-papers/CrossingTools_TAG).

**Acknowledgements.** We thank Matthias Frisch, Carola Zenke-Philippi and Joschua Okoye for stimulating discussions and valuable perspectives on practical plant breeding and segregation variances. We also thank Brian Kinghorn and Andrew Kinghorn for insightful discussions on mate allocation strategies in animal breeding.

**Authors' Note.** For the purpose of open access, the authors have applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementaryfigures.pdf