



OPEN ACCESS

EDITED BY

Salvatore Ceccarelli,
Bioversity International, Italy

REVIEWED BY

Marta Malinowska,
Aarhus University, Denmark
Ram Kumar Basnet,
Rijk Zwaan, Netherlands

*CORRESPONDENCE

Reem Joukhadar
✉ r.joukhadar@sustatability.com

†PRESENT ADDRESS

Reem Joukhadar,
SuSTATability Statistical Solutions, Melbourne,
VIC, Australia

RECEIVED 13 November 2023

ACCEPTED 23 May 2024

PUBLISHED 24 June 2024

CITATION

Joukhadar R, Li Y, Thistlethwaite R, Forrest KL,
Tibbits JF, Trethowan R and Hayden MJ
(2024) Optimising desired gain indices to
maximise selection response.
Front. Plant Sci. 15:1337388.
doi: 10.3389/fpls.2024.1337388

COPYRIGHT

© 2024 Joukhadar, Li, Thistlethwaite, Forrest,
Tibbits, Trethowan and Hayden. This is an
open-access article distributed under the terms
of the [Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction
in other forums is permitted, provided the
original author(s) and the copyright owner(s)
are credited and that the original publication
in this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Optimising desired gain indices to maximise selection response

Reem Joukhadar^{1*†}, Yongjun Li¹, Rebecca Thistlethwaite²,
Kerrie L. Forrest¹, Josquin F. Tibbits¹, Richard Trethowan^{2,3}
and Matthew J. Hayden^{1,4}

¹Agriculture Victoria, Centre for AgriBioscience, AgriBio, Bundoora, VIC, Australia, ²School of Life and Environmental Sciences, Plant Breeding Institute, Sydney Institute of Agriculture, The University of Sydney, Narrabri, NSW, Australia, ³School of Life and Environmental Sciences, Plant Breeding Institute, Sydney Institute of Agriculture, The University of Sydney, Cobbitty, NSW, Australia, ⁴School of Applied Systems Biology, La Trobe University, Bundoora, VIC, Australia

Introduction: In plant breeding, we often aim to improve multiple traits at once. However, without knowing the economic value of each trait, it is hard to decide which traits to focus on. This is where “desired gain selection indices” come in handy, which can yield optimal gains in each trait based on the breeder’s prioritisation of desired improvements when economic weights are not available. However, they lack the ability to maximise the selection response and determine the correlation between the index and net genetic merit.

Methods: Here, we report the development of an iterative desired gain selection index method that optimises the sampling of the desired gain values to achieve a targeted or a user-specified selection response for multiple traits. This targeted selection response can be constrained or unconstrained for either a subset or all the studied traits.

Results: We tested the method using genomic estimated breeding values (GEBVs) for seven traits in a bread wheat (*Triticum aestivum*) reference breeding population comprising 3,331 lines and achieved prediction accuracies ranging between 0.29 and 0.47 across the seven traits. The indices were validated using 3,005 double haploid lines that were derived from crosses between parents selected from the reference population. We tested three user-specified response scenarios: a constrained equal weight (INDEX1), a constrained yield dominant weight (INDEX2), and an unconstrained weight (INDEX3). Our method achieved an equivalent response to the user-specified selection response when constraining a set of traits, and this response was much better than the response of the traditional desired gain selection indices method without iteration. Interestingly, when using unconstrained weight, our iterative method maximised the selection response and shifted the average GEBVs of the selection candidates towards the desired direction.

Discussion: Our results show that the method is an optimal choice not only when economic weights are unavailable, but also when constraining the selection response is an unfavourable option.

KEYWORDS

desired gain indices, genomic prediction, genomic estimated breeding values, selection indices, genomic selection

Introduction

Plant and animal breeding programmes aim to improve multiple commercially valuable traits. Traditional selection methods often overlook the correlation and varying heritabilities among traits, leading to suboptimal selection response. Selecting for traits with conflicting influences can be tricky, as progress in one might harm another (Smith, 1936; Hazel, 1943). Direct selection for traits with antagonistic correlations can lead to unfavourable responses for some of the correlated traits, while synergistic correlations between two or more traits can bias the selection against the remaining traits. Hence, selection indices that account for the architecture and the relationships among different traits are used to simultaneously select for several traits weighted by their importance to the objectives of the breeding programme to avoid unfavourable biases towards or against specific traits (Céron-Rojas and Crossa, 2018).

Selection indices are expected to achieve higher genetic gain compared to independent culling or sequential selection, especially when the targeted traits have variable heritabilities, phenotypic and genotypic correlations, and economic values (Hazel et al., 1994). Selection indices can be applied to phenotypic records, genomic estimated breeding values (GEBVs), or phenotypes and genomic variants (Céron-Rojas and Crossa, 2020). Despite their utility, selecting the most appropriate indices remains a challenge, requiring a deep understanding of trait genetics and pairwise correlation among the traits (Dekkers et al., 2021). Additionally, the correlation structure among the traits, whether it is due to genetic correlation or environmental correlation, also plays a crucial role (Fernandes et al., 2021). It is important to use a selection index that considers these complexities to mitigate their effects on the selection response (Richardson et al., 2021).

Breeders usually calculate selection indices using the economic values for their traits or through setting a desired gain threshold for each trait (Pesek and Baker, 1969). However, these values may not always be available. Desired gain index methods describe the extent of genetic improvement a breeder intends to achieve for different traits in their germplasm (Pesek and Baker, 1969). The most

important advantage of these methods is that they do not require economic weights to be estimated. Desired genetic gain can be determined from a breeder's knowledge of the genetic merit of their materials and traits, although this may not be possible for all breeding programmes or traits (Céron-Rojas and Crossa, 2018). Alternatively, breeders can sample arbitrary values for the desired gain and choose a sample that shifts all traits towards the desired direction (Li et al., 2017). However, the main problem associated with desired gain selection index methods is that they neither maximise the correlation between the net genetic merit of the individual and the selection index, nor maximise the selection response for different traits (Itoh and Yamada, 1986, 1988).

Here, we report the development of a new iterative method that optimises the choice of the input desired gain (Yamada et al., 1975) to achieve a user-specified selection response for different traits. We demonstrate our method using a large bread wheat population composed of 6,336 lines, of which 3,331 lines were used to develop the selection indices while the remaining 3,005 lines were used as independent selection candidates. Selection indices were developed using the GEBVs for seven traits, namely, grain yield, thousand kernel weight, protein content, screening percentage, and stem rust, leaf rust, and stripe rust disease resistance.

Materials and methods

Plant materials and phenotyping

A total of 3,331 wheat lines were used including 36 Australian checks and 2,824 lines developed at the Plant Breeding Institute, Cobbitty, NSW, Australia from different bread wheat germplasm pools and diverse exotic resources including emmer wheat, synthetic wheat, and landraces. More information can be found in Joukhadar et al. (2021a). The population was planted in 30 irrigated field trials between 2014 and 2020 at Narrabri in New South Wales, Cadoux, Merredin, and Geraldton in Western Australia and Horsham in Victoria with population sizes ranging between 195 and 1,956 lines. A subset of approximately 180 lines were selected based on their GEBVs for yield and genetic diversity (Joukhadar et al., 2021b) and replicated across all trials, and different trials shared different numbers of individuals. Materials were sown at two times of sowing (TOS) in randomised complete

Abbreviations: GEBVs, genomic estimated breeding values; Lr, leaf rust; Prot, protein content; Screen, percentage screenings; Sr, stem rust; TKW, thousand kernel weight; TOS, time of sowing; YLD, grain yield; Yr, stripe rust.

block designs of two replicates (each replicate as one block) in adjacent blocks. The only exception was Horsham in 2017, where three TOS were used. Details of the 30 trials can be found in [Supplementary Table 1](#). Another set of 3,005 double haploid lines were used to validate the selection indices results. These lines were developed from crosses between various parents selected from the reference population of 3,331 individuals.

Seven traits were evaluated, namely, grain yield (YLD), percentage screenings (Screen), protein content (Prot), thousand kernel weight (TKW), and stem rust (Sr), leaf rust (Lr), and stripe rust (Yr) disease resistance. Plots were harvested, and harvested grain weight was subsequently converted to kg/ha as a measure for YLD. Prot was determined using an Infratec TM 1241 Grain Analyser. A total of 400 g of seeds from each plot was screened after 40 machine movements through a screen of 2 mm. The remaining material above the screen was weighed to calculate the screening percentage. One hundred visibly viable seeds (excluding cracked, broken, or diseased seeds) were counted and weighed to determine the 1,000 kernel weight. The three rust diseases were scored for their resistance in a scale from 1 (complete resistance) to 9 (complete susceptibility). The number of trials ranged from 5 for the three rust diseases to 30 for YLD, while the total phenotypic records ranged from 4,566 for Lr and Yr to 11,053 for YLD ([Table 1](#)). Best linear unbiased estimation (BLUE) values were calculated for each trial independently by fitting a spatial linear mixed model considering the field layout (row and column) and the replications as random effects using ASReml-R ([Gilmour et al., 2009](#)). More details can be found in [Joukhadar et al. \(2021a\)](#). BLUE values were used to calculate GEBVs that were used to develop the selection indices.

The reference population of 3,331 individuals was genotyped with the 90K Infinium SNP array ([Wang et al., 2014](#)). SNPs were filtered to keep only the one with call rate > 60% and minor allele frequency (MAF) > 5%. Missing genotypes within the 90K SNPs were imputed using LinkImpute software ([Money et al., 2015](#)) and the previous studies showed that the accuracy of the *in silico* cross-

validation was larger than 0.99 ([Joukhadar et al., 2020](#)). The validation population of 3,005 individuals was genotyped with 40K wheat and barley Infinium SNP array ([Keeble-Gagnère et al., 2021](#)). For all materials, a single plant was genotyped given that all individuals were even double haploid or fixed lines. Both the reference population genotyped with 90K and the validation population genotyped with 40K Infinium SNP arrays were imputed to exome capture level ([He et al., 2019](#)). The imputation from low density to high density was previously reported to have an accuracy of 92.4% ([Joukhadar et al., 2021a](#)). A total of 218,092 were common across the reference and the validation imputed SNP sets, which were used for subsequent analyses ([Joukhadar and Daetwyler, 2022](#)).

Genetic correlation and genomic estimated breeding values

Trait variances and narrow-sense heritability (SNP based heritability) was calculated using the univariate restricted maximum likelihood (REML) analysis implemented in MTG2 software ([Lee and Van der Werf, 2016](#)) by fitting the phenotypic records as well as the genomic relatedness matrix. The genetic covariances and genetic correlations between each pair of traits were calculated using bivariate REML analysis implemented in the same software. For both analyses, the trials were fitted as a covariate. The genetic correlation analysis was performed with two aims: (1) to calculate the genetic correlation between the two times of sowing (TOS1 and TOS2) to assess the level of genotype by environment interaction between both TOSs; and (2) to calculate the genetic covariances among traits, which is required to develop the selection indices.

GEBVs were calculated using the BayesR software, previously described in [Breen et al. \(2022\)](#). BayesR models SNP effects from a mixture of four normal distributions ([Erbe et al., 2012](#)). The aim is to simultaneously assign the SNPs with large, medium, small, and no effect. Marker effects were sampled from one of the following four normal distributions: $N(0, 0 \sigma_g^2)$ zero effect, $N(0, 0.0001 \sigma_g^2)$ small effect, $N(0, 0.001 \sigma_g^2)$ medium effect, and $N(0, 0.01 \sigma_g^2)$ large effect, where σ_g^2 is the additive genetic variance. The genomic relatedness matrix was calculated following [VanRaden \(2008\)](#). The prior proportions of loci attributed to each marker effect distribution were 0.94, 0.049, 0.01, and 0.001, respectively. A total of 50,000 iterations for the BayesR analysis were used, of which the first half was considered as burn in. For cross-validation, half of the phenotypic records for each trait were randomly selected as a reference to predict the other half, which was considered the validation set. Prediction accuracy was calculated as the correlation coefficient between the GEBVs and the actual phenotypes in the validation set. The cross-validation strategy was repeated 100 times, wherein the population was randomly partitioned into reference and validation sets in each replicate. Prediction accuracies for each trait were averaged across the 100 replicates, and the standard deviation of the average accuracies was subsequently calculated.

TABLE 1 Number of trials, total number of phenotypic records, genetic correlation between TOS1 and TOS2 for the studied traits, and prediction accuracy using BayesR.

| Trait | Number of trials | Number of records | Genetic correlation | Prediction accuracy |
|--------|------------------|-------------------|---------------------|---------------------|
| YLD | 30 | 11,053 | 0.89 | 0.29 (\pm 0.03) |
| TKW | 21 | 9,872 | 0.91 | 0.39 (\pm 0.04) |
| Prot | 17 | 7,491 | 0.98 | 0.32 (\pm 0.03) |
| Screen | 23 | 8,557 | 0.92 | 0.30 (\pm 0.03) |
| Sr | 5 | 4,633 | . | 0.40 (\pm 0.04) |
| Yr | 5 | 4,566 | . | 0.47 (\pm 0.05) |
| Lr | 5 | 4,566 | . | 0.42 (\pm 0.04) |

Prediction accuracy columns show the average of the 100 cross-validation replicates and the value between the brackets represent the standard deviation of the average accuracies. Lr, leaf rust; Sr, stem rust; Yr, yellow rust; Prot, protein content; Screen, screening percentage; TKW, thousand kernel weight; and YLD, grain yield. Genetic correlation was not calculated for the rust diseases because they were only scored in TOS1.

Desired gain index

Selection indices were calculated for the following traits: grain yield, thousand kernel weight, screening percentage, protein content, and leaf, stem, and yellow rust disease resistance. We used the desired trait gain index method described in Yamada et al. (1975). The method uses the following equation:

$$b = P^{-1}G(GP^{-1}G)^{-1}d \quad (1)$$

where b is the final desired gain-based index; P is the phenotypic variance–covariance matrix (in our case, the correlation between GEBVs); G is the genetic variance–covariance matrix; and d is a vector of the desired gains. Variances in G were calculated using the previous univariate REML analysis for each trait, while the covariances were calculated by applying the bivariate REML model to analyse each pair of traits.

Yamada et al. (1975) suggested choosing arbitrarily values for d until the averages for all traits in the selected lines were shifted towards the desired direction (e.g., increasing yield and reducing screening percentage). Instead of choosing arbitrary values for d , we developed a new strategy that optimises the choice of d to calculate an optimal index (b) that can achieve a user-specific level of improvement (selection response) for each trait. The user-specified selection response (dg) is defined as the number of standard deviations (positive or negative) the average phenotypes of the selected lines can be shifted from the average of the whole population for each trait. The method samples values for d within an iterative process and selects the sample that develops an index that best matches the desired selection response. For each iteration, the method calculates b and use it to calculate the genetic gain g of the sampled d . Then, g is used to calculate the goodness of fit for the sampled d for each trait with the following equation:

$$gof = \frac{\log_{base} \left(\frac{base^2}{|g|} \right)}{\frac{base^2}{|g|}} \quad (2)$$

where “base” is the base of the logarithm that can be calculated with the following equation: $base = \sqrt{e \times dg}$, in which e is the Euler’s number. Specifically, the base determines the scaling of the logarithmic transformation applied to the ratio of g to dg . Therefore, under the proposed equation, gof gets maximised ($gofMAX$) when $g=dg$, which means that we achieved the desired selection response in the selection candidates. The penalty of the sampled d for all traits (θ) is given by the following equation:

$$\theta = \sum_i^n \frac{gofMAX_i - gof_i}{gofMAX_i} \quad (3)$$

where n is the number of traits. θ serves as a measure of deviation between the observed and desired selection responses across traits, guiding the iterative optimisation process. A lower value of θ indicates better alignment between g and dg . The value of θ from each iteration is compared to the sample of the previous iteration and is accepted if the selection response improves, i.e., getting a smaller θ . The sampling mean for each trait in each

iteration is updated from the most recently accepted sampled weights. The analysis was run for 1,000 iterations.

We used the GEBVs for the phenotyped reference population of 3,331 lines to develop genotypic and phenotypic covariance matrices. The calculated covariance matrices calculated on the reference population were then applied to the GEBVs of the 3,005 double haploid selection candidates. From the selection index ranked lines, a total of 100 lines were selected. Selection response for the selected lines was expressed as the number of standard deviations their average differed from the average of the whole population for each trait. Three different indices were calculated. The user-specified selection response was first set to +0.5 standard deviations for yield, TKW, and protein, and –0.5 standard deviations for screening the three rust diseases (INDEX1) to ensure equal weights for all traits. Another yield dominant user-specified selection response was used with +2 standard deviations for yield, +0.5 standard deviation for TKW and protein, and –0.5 standard deviation for the three rust diseases (INDEX2). A third index was calculated with high user-specified selection responses at +4 standard deviations for yield, TKW, and protein, and –4 standard deviations for screening the three rust diseases (INDEX3) to maximise the potential selection response for all traits. Each index was run for 20 replicates and the correlations among replicates were averaged to ensure that the different replicates produced comparable solutions.

Results

The analysis revealed high genetic correlations among all traits, indicating limited genotype by environment interactions for TOS1 and TOS2. Notably, the smallest correlation observed was 0.89 for grain yield while the largest correlation was equal to 0.98 for protein percentage (Table 1). This analysis was not possible for the three rust diseases given that the disease resistances were scored only in optimal sowing time (TOS1). Additionally, the BayesR model demonstrated varying levels of prediction accuracy for the studied traits, ranging from medium to low, with values between 0.29 for YLD and 0.47 for Yr (Table 1). Particularly noteworthy were the higher prediction accuracies observed for resistances to the three rust diseases, averaging at 0.43, compared to an average of 0.33 for the remaining four traits.

The narrow-sense heritability and genetic correlations among the studied traits were summarised in Table 2. Narrow-sense heritability values varied across the traits, ranging from 0.21 for YLD to 0.59 for Yr. Additionally, the examination of genetic correlations revealed generally low values among the seven traits, spanning from –0.21 to 0.28, underscoring the relative independence of these traits in terms of their genetic basis. However, a striking exception to this pattern was observed in the strong negative correlation of –0.61 between TKW and screening percentage, which is expected given that small, low weight seeds are the ones that pass through the screen. Overall, these findings contribute to a deeper understanding of the genetic relationships and heritability of the studied traits, providing valuable insights for future breeding efforts aimed at improving crop performance and resilience.

TABLE 2 Narrow-sense heritability (diagonal) and genetic correlations (off diagonal) between the studied traits.

| | YLD | TKW | Prot | Screen | SR | YR | LR |
|--------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| YLD | 0.21 (± 0.01) | -0.21 (± 0.07) | -0.03 (± 0.09) | 0.10 (± 0.09) | 0.16 (± 0.09) | 0.12 (± 0.08) | 0.16 (± 0.09) |
| TKW | -0.21 (± 0.07) | 0.50 (± 0.01) | -0.01 (± 0.07) | -0.61 (± 0.05) | -0.05 (± 0.07) | 0.14 (± 0.06) | 0.28 (± 0.07) |
| Prot | -0.03 (± 0.09) | -0.01 (± 0.07) | 0.30 (± 0.02) | 0.15 (± 0.08) | -0.10 (± 0.07) | -0.15 (± 0.07) | 0.13 (± 0.08) |
| Screen | 0.10 (± 0.09) | -0.61 (± 0.05) | 0.15 (± 0.08) | 0.49 (± 0.01) | -0.06 (± 0.08) | -0.14 (± 0.07) | -0.18 (± 0.08) |
| SR | 0.16 (± 0.09) | -0.05 (± 0.07) | -0.10 (± 0.07) | -0.06 (± 0.08) | 0.35 (± 0.02) | -0.07 (± 0.06) | 0.11 (± 0.08) |
| YR | 0.12 (± 0.08) | 0.14 (± 0.06) | -0.15 (± 0.07) | -0.14 (± 0.07) | -0.07 (± 0.06) | 0.59 (± 0.01) | 0.28 (± 0.07) |
| LR | 0.16 (± 0.09) | 0.28 (± 0.07) | 0.13 (± 0.08) | -0.18 (± 0.08) | 0.11 (± 0.08) | 0.28 (± 0.07) | 0.34 (± 0.02) |

Values between brackets represent the standard errors of the estimations. Lr, leaf rust; Sr, stem rust; Yr, yellow rust; Prot, protein content; Screen, screening percentage; TKW, thousand kernel weight; and YLD, grain yield.

Iterative index analysis (Equations 2, 3), run using the parameters of the three tested indices, identified a subset of samples with average GEBVs that were all shifted towards the desired direction of increased yield, TKW, and protein; reduced screenings; and low rust disease resistance scores (Tables 3–5). The iterative method showed better performance compared to the standard desired genetic gain method (Yamada et al., 1975) with no iterations. For INDEX1 (equal weight for all traits), the absolute standard deviation of the response ranged between 0.46 and 1.02 with an average of 0.68, which was very close to the targeted user-specified response of 0.5 for all traits. On the other hand, without iteration, the absolute standard deviation of the response ranged between 0.12 and 1.69. The iterative method also successfully developed an index when the targeted response was biased towards a subset of trait(s). INDEX2 (the yield dominant index) had a higher weight for yield and selected lines with the iterative method had, on average, standard deviations 1.71 higher than the whole population. The average of the remaining traits was higher at approximately 0.5 standard deviations, which equalled the user-specified targeted response (Table 4). However, protein content was negatively selected when the index was used without iterations, indicating the superiority of the new method.

The last index, INDEX3, had a targeted response of 4 standard deviations across all traits, which was impossible to achieve given

that only 100 lines were selected from a population of 3,005 individuals (~3.3%). This index was designed to maximise the genetic gain for all traits in the selected materials as, theoretically, only 0.05% of the whole population could be above or below 4 standard deviations for a single trait. The average GEBVs of the selected materials had an absolute average of 1.04 standard deviations from the average of the whole population across the seven traits, which ranged from 0.64 for TKW to 1.24 for screenings percentage (Table 5). Calculating this index without iterations resulted in the same answer as INDEX1 given that they both have equal weights across the traits and *d* is a scaler in Equation 1. Therefore, the ranking of individuals will be the same for both INDEX1 and INDEX3. Figure 1 shows the distribution of YLD for the selection candidate whole population as well as the selected lines using the three indices (INDEX1, INDEX2, and INDEX3).

The method was validated using 3,005 double haploid lines, which were independent from the reference set that was used to calculate the phenotypic and genotypic covariance matrices. The selection response for the independent validation set was very similar but slightly higher than that calculated on the reference population (Tables 3–5). In other words, the average phenotypes for the selection candidates in both the reference and the validation sets had an equivalent number of standard deviations above (for YLD, TKW, and protein) or below (for screening and the three rust

TABLE 3 Selection response from INDEX1.

| Trait | Sampled desired gain (<i>d</i>) | Index (<i>b</i>) | Targeted response (<i>dg</i>) | Response NoIteration | Response SD Ref | Response SD Cand |
|--------|-----------------------------------|--------------------|---------------------------------|----------------------|-----------------|------------------|
| YLD | 1.06 | 16.39 | 0.5 | 0.50 | 0.67 | 0.83 |
| TKW | 3.37 | 1.34 | 0.5 | 1.69 | 0.57 | 1.02 |
| Prot | 0.72 | 1.26 | 0.5 | 0.48 | 0.87 | 0.66 |
| Screen | -1.22 | 0.15 | -0.5 | -0.12 | -0.43 | -0.46 |
| SR | -0.55 | -2.91 | -0.5 | -0.31 | -0.55 | -0.56 |
| LR | 0.03 | -1.67 | -0.5 | -0.95 | -0.58 | -0.61 |
| YR | -0.60 | -1.14 | -0.5 | -0.29 | -0.53 | -0.62 |

The sampled desired gain (*d*), final index (*b*), user-specified targeted selection response (*dg*), and number of standard deviations for the genetic gain of the selected lines [NoIteration, using the standard desired gain index method of Yamada et al. (1975), our method with the reference, Ref, and candidate, Cand, populations] using INDEX1 (equal weights for all traits). Lr, leaf rust; Sr, stem rust; Yr, yellow rust; Prot, protein content; Screen, screening percentage; TKW, thousand kernel weight; and YLD, grain yield; SD Ref, standard deviation reference population; and SD Cand, standard deviation selection candidates.

TABLE 4 Selection response from INDEX2.

| Trait | Sampled desired gain (<i>d</i>) | Index (<i>b</i>) | Targeted response (<i>dg</i>) | Response NoIteration | Response SD Ref | Response SD Cand |
|--------|-----------------------------------|--------------------|---------------------------------|----------------------|-----------------|------------------|
| YLD | 1.61 | 19.70 | 2 | 2.01 | 1.33 | 1.71 |
| TKW | -0.28 | 0.58 | 0.5 | 0.74 | 0.33 | 0.46 |
| Prot | 0.48 | 1.27 | 0.5 | -0.35 | 0.44 | 0.51 |
| Screen | 0.01 | 0.04 | -0.5 | -0.08 | -0.45 | -0.50 |
| SR | 0.70 | 1.40 | -0.5 | -0.21 | -0.46 | -0.54 |
| LR | 0.53 | -0.01 | -0.5 | 0.13 | -0.44 | -0.54 |
| YR | 0.65 | -0.61 | -0.5 | -0.45 | -0.42 | -0.54 |

The sampled desired gain (*d*), final index (*b*), user-specified targeted selection response (*dg*), and number of standard deviations for the genetic gain of the selected lines [NoIteration, using the standard desired gain index method of Yamada et al. (1975), our method with the reference, Ref, and candidate, Cand, populations] using INDEX2 (yield dominant). Lr, leaf rust; Sr, stem rust; Yr, yellow rust; Prot, protein content; Screen, screening percentage; TKW, thousand kernel weight; and YLD, grain yield; SD Ref, standard deviation reference population; and SD Cand, standard deviation selection candidates.

diseases) the average of the whole reference or validation populations, respectively. Running 20 replicates for each analysis showed that the iterative method produced indices with high repeatability. The average correlation coefficients over all replicates for the three indices were 0.95, indicating robust divergence after 1,000 iterations as well as consistently highly optimised weights for different traits.

Discussion

We utilised a large population of Australian bread wheat cultivars and breeding lines that were phenotyped in diverse environments across the Australian wheatbelt, stretching from Narrabri in New South Wales to Geraldton in Western Australia. The materials were planted at both optimal (TOS1) and late (TOS2) times of sowing and irrigated to avoid drought stress so that heat tolerance during anthesis and grain filling periods could be assessed (Thistlethwaite et al., 2020).

To develop robust selection indices for different traits, it is important to quantify the level of genotype by environment

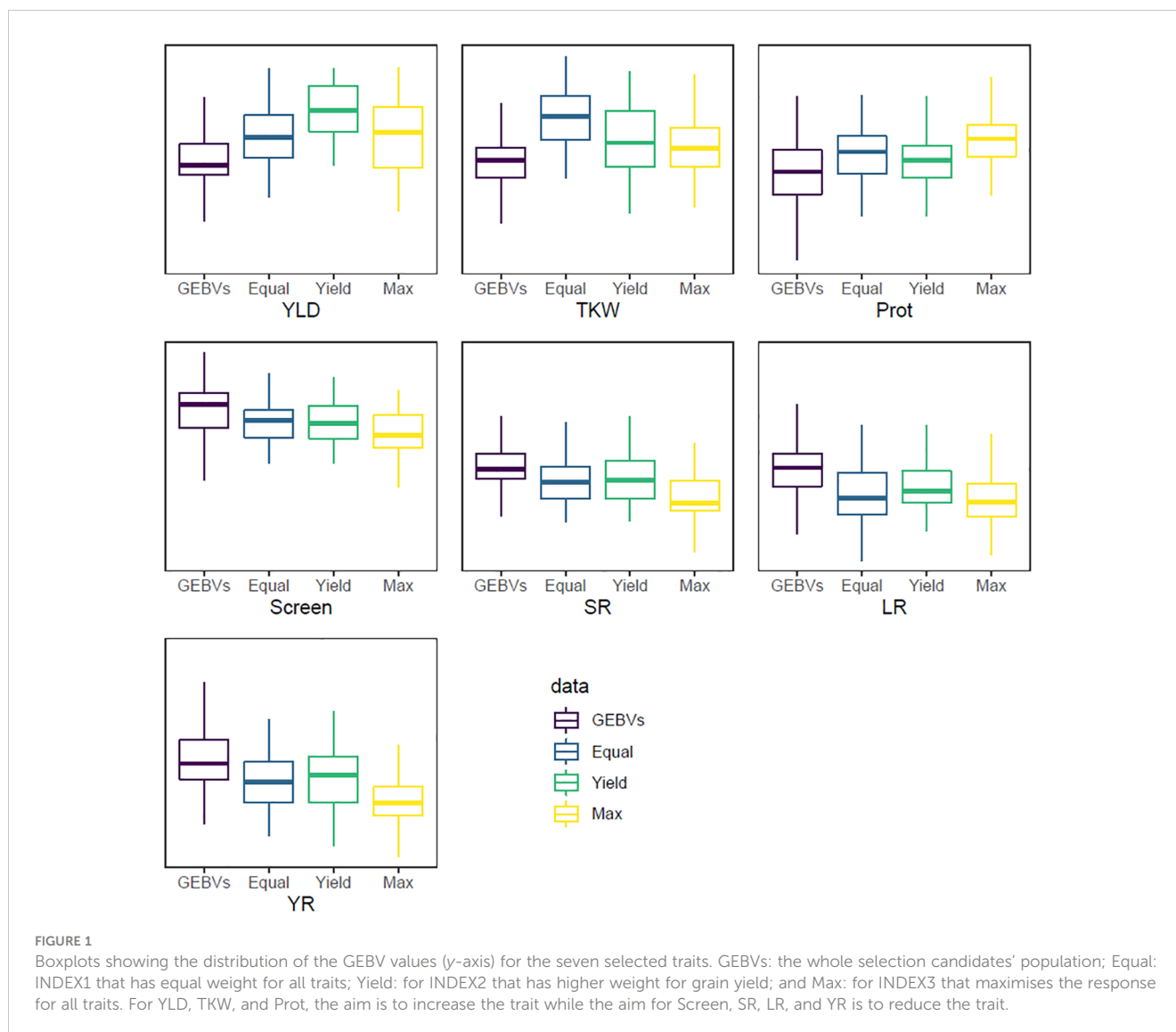
interaction between the optimal and the late sowing trials by calculating their genetic correlation. The high genetic correlation observed implied that the ranking of the lines had minimal change between the two sowing times for all traits. This is essential to decide whether a trait will need a single weight regardless of heat treatment or if two different weights will be needed, one for each treatment. All traits showed high correlations between both TOSs of which protein content had almost no genotype by environment interaction with a genetic correlation between TOS1 and TOS2 of 0.98 (Table 1). This was not unexpected given that protein content is affected more by nitrogen availability in the soil than weather (McDonald, 1992) and TOS1 and TOS2 were planted in adjacent fields in each environment and subjected to the same management practises. These results indicated that TOS1 and TOS2 did not need different weights when developing the selection indices and could be considered as a single trait.

The prediction accuracy for resistance to the three rust diseases was comparable to accuracies reported in previous studies that investigated diverse wheat germplasm (Daetwyler et al., 2014; Juliana et al., 2017). He et al. (2019) used a subset of the data used in the present study that included 10 trials conducted at

TABLE 5 Selection response from INDEX3.

| Trait | Sampled desired gain (<i>d</i>) | Index (<i>b</i>) | Targeted response (<i>dg</i>) | Response NoIteration | Response SD Ref | Response SD Cand |
|--------|-----------------------------------|--------------------|---------------------------------|----------------------|-----------------|------------------|
| YLD | 0.24 | 5.47 | 4 | 0.50 | 0.78 | 1.05 |
| TKW | 1.35 | 0.28 | 4 | 1.69 | 0.53 | 0.64 |
| Prot | 0.57 | 1.33 | 4 | 0.48 | 0.73 | 1.10 |
| Screen | -3.05 | -0.40 | -4 | -0.12 | -0.49 | -1.24 |
| SR | -0.62 | -2.53 | -4 | -0.31 | -0.50 | -1.09 |
| LR | -0.75 | -2.35 | -4 | -0.95 | -0.65 | -1.06 |
| YR | -0.78 | -0.74 | -4 | -0.29 | -0.60 | -1.11 |

The sampled desired gain (*d*), final index (*b*), user-specified targeted selection response (*dg*), and number of standard deviations for the genetic gain of the selected lines [NoIteration, using the standard desired gain index method of Yamada et al. (1975), our method with the reference, Ref, and candidate, Cand, populations] using INDEX3 (maximum gain). Lr, leaf rust; Sr, stem rust; Yr, yellow rust; Prot, protein content; Screen, screening percentage; TKW, thousand kernel weight; and YLD, grain yield; SD Ref, standard deviation reference population; and SD Cand, standard deviation selection candidates.



Narrabri, New South Wales between 2013 and 2017. They reported comparable prediction accuracies to our results for grain yield and screening percentage for models that did not fit genotype by environment interactions. However, their accuracies for protein content and TKW were much higher compared to our results. The latter could be a result of the inclusion of field trials with more diverse climates and soil types in the reference population used in our study. He et al. (2019) also showed that prediction accuracy was further improved for protein content and grain yield when fitting the genotype by environment interaction in the model. However, our research focussed on improving the development of indices to maximise the selection response regardless of the statistical model applied for genomic prediction.

The seven studied traits showed low to medium narrow-sense heritability (Table 2). Daetwyler et al. (2014) reported comparable heritability values to our estimates for resistance to the three rust diseases. Various studies have reported comparable heritability values for the remaining traits, especially low heritability for YLD

which is often highly influenced by environmental variability (Babar et al., 2007; Heffner et al., 2011; He et al., 2019; Watson et al., 2019). The genetic correlations among different traits were generally low and in most cases the value was smaller or equivalent to two standard errors, except for the negative correlation between TKW and screening percentage of -0.61 (Table 3). A negative correlation is expected given that a higher screening percentage implies smaller seeds. The remaining traits showed low genetic correlations between -0.21 and 0.28 , indicating that selection applied to a given trait will not significantly affect another trait.

Previously developed unconstrained linear phenotypic or genomic selection indices do not have control over the direction of the selection response; i.e., whether genetic gain is increased or decreased (Smith, 1936; Dekkers, 2007; Togashi et al., 2011; Cerón-Rojas and Crossa, 2019). For this reason, attempts have been made to constrain the genetic gain for a subset of traits. Kempthorne and Nordskog (1959) developed a method to prevent changing of the genetic gain for a subset of traits (in other words, the gain is equal to

zero), while others generalised this method to allow the setting of a predetermined level of genetic gain for a subset of traits (Mallard, 1972; Harville, 1975; Tallis, 1985). These methods improved the choice of selection candidates with better shifts in genetic gain in the desired direction. However, all these methods required an economic weight for each trait to be predefined, information that may not be always available for all traits or breeding programmes.

While the advantage of the desired gain indices is that economic weights are not required, their main disadvantage is inability to maximise the selection response (Itoh and Yamada, 1986, 1988). This results because of the subjective choice of the desired gain values that are usually approximated using the breeder's knowledge of the traits or arbitrarily sampled to achieve a proper selection response (Cérón-Rojas and Crossa, 2018). Cérón-Rojas and Crossa (2020) argued that the constrained linear phenotypic selection index is similar in principle to the desired gain indices with the advantage of maximising the selection response as well as the correlation between the selection index and the net genetic merit of each individual. Our iterative method optimises the choice of the desired gain to achieve a user-specified selection response for the targeted traits. Our results show that our method can efficiently move the genetic gain in the desired direction regardless of whether the breeding objective requires some traits to be constrained (INDEX1 and INDEX2) or unconstrained (INDEX3). Our method optimises the selection response not only when the economic weights are unavailable, but also when constraining a subset of traits is not the ideal option.

While our method succeeded in maximising the selection response for all traits in the desired direction, it is still not possible to determine if the method maximises the correlation between the selection index and net genetic merit. This is because the covariance between both is undefined given that it depends on the estimation of the economic weight (Cérón-Rojas and Crossa, 2020); thus, this parameter cannot be theoretically assessed using our method. However, as our validation population of 3,005 lines was developed from crosses between a selected subset of parents from the reference population of 3,331 individuals, the former materials can be considered as the next generation in a continuous breeding programme that was not included in the reference population. Hence, this population can empirically help assess the unobserved genetic merit. Similar results were previously obtained using unconstrained and constrained linear phenotypic or genomic selection indices using real data with two breeding cycles or simulated data with six cycles derived from the reference population (Cérón-Rojas and Crossa, 2019, 2020).

Our new approach holds significant practical implications for enhancing breeding efficiency and accelerating genetic gain in improving breeding programmes. Our method aids breeders in the selection of superior cultivars with broad adaptability simultaneously for multiple traits without biasing the selection towards or against a specific trait. Additionally, our iterative method for optimising selection responses, particularly in scenarios where economic weights are unavailable or when constrained selection is not ideal, offers a flexible and efficient approach to breeding programme management. Overall, our study contributes to advancing breeding methodologies tailored

to modern breeding objectives, ultimately facilitating the development of improved varieties with enhanced yield potential, stress resilience, and end-use quality.

Conclusion

We developed a new method that extends the application of desired trait gain selection indices to maximise the selection response for multiple constrained or unconstrained traits. We showed that our method shifts the selection response in non-reference individuals in the targeted direction to increase or decrease the trait average in the selection candidates. We demonstrated that our method has the power to maximise the genetic gain when using unconstrained weights and to achieve the targeted selection response set by the breeder for different traits in the appropriate direction. However, more empirical testing is required using multi-breeding cycle data to ensure that the calculated indices are sufficiently powerful to maximise the correlation between the index and the net genetic merit.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

RJ: Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. YL: Writing – review & editing, Software, Resources, Methodology, Investigation, Formal analysis, Conceptualization. RTh: Writing – review & editing, Resources, Funding acquisition, Data curation. KF: Writing – review & editing, Resources, Data curation. JT: Writing – review & editing, Resources, Project administration, Data curation. RTr: Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Data curation, Conceptualization. MH: Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The authors acknowledge the financial support of the University of Sydney, and Agriculture Victoria. The authors acknowledge the financial support of the Grains Research and Development Corporation. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The authors declare that this study received funding from Grains Research and Development Corporation. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2024.1337388/full#supplementary-material>

SUPPLEMENTARY TABLE 1

Summary of the field trials, number of phenotyped individuals, and traits. Trail names were recorded as (Location_year_TimeOfSowing). NoGeno: Number of Genotypes; Prot, protein%; Screen, screening%; TKW, thousand kernel weight (g); YLD, yield (t/ha); Lr, leaf rust; Sr, stem rust; Yr, stripe rust.

References

- Babar, M. A., Van Ginkel, M., Reynolds, M. P., Prasad, B., and Klatt, A. R. (2007). Heritability, correlated response, and indirect selection involving spectral reflectance indices and grain yield in wheat. *Aust. J. Agric. Res.* 58, 432–442. doi: 10.1071/AR06270
- Breen, E. J., MacLeod, I. M., Ho, P. N., Haile-Mariam, M., Pryce, J. E., Thomas, C. D., et al. (2022). BayesR3 enables fast MCMC blocked processing for largescale multi-trait genomic prediction and QTN mapping analysis. *Commun. Biol.* 5, 661.
- Céron-Rojas, J. J., and Crossa, J. (2018). "Constrained linear phenotypic selection indices in *Linear selection indices*," in *modern plant breeding*, eds. J. J. Céron-Rojas and J. Crossa (Cham, Switzerland: Springer), 43–69. doi: 10.1007/978-3-319-91223-3_3
- Céron-Rojas, J. J., and Crossa, J. (2019). Efficiency of a constrained linear genomic selection index to predict the net genetic merit in plants. *G3: Genes Genomes Genet.* 9, 3981–3994. doi: 10.1534/g3.119.400677
- Céron-Rojas, J. J., and Crossa, J. (2020). Expectation and variance of the estimator of the maximized selection response of linear selection indices with normal distribution. *Theor. Appl. Genet.* 133, 2743–2758. doi: 10.1007/s00122-020-03629-6
- Daetwyler, H. D., Bansal, U. K., Bariana, H. S., Hayden, M. J., and Hayes, B. J. (2014). Genomic prediction for rust resistance in diverse wheat landraces. *Theor. Appl. Genet.* 127, 1795–1803. doi: 10.1007/s00122-014-2341-8
- Dekkers, J. C. M. (2007). Prediction of response to marker-assisted and genomic selection using selection index theory. *J. Anim. Breed Genet.* 124, 331–341. doi: 10.1111/j.1439-0388.2007.00701.x
- Dekkers, J. C., Su, H., and Cheng, J. (2021). Predicting the accuracy of genomic predictions. *Genet. Selection Evol.* 53, 55. doi: 10.1186/s12711-021-00647-w
- Erbe, M., Hayes, B. J., Matukumalli, L. K., Goswami, S., Bowman, P. J., Reich, C. M., et al. (2012). Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. *J. Dairy Sci.* 95, 4114–4129. doi: 10.3168/jds.2011-5019
- Fernandes, G. M., Savegnago, R. P., Freitas, L. A., El Faro, L., Roso, V. M., and De Paz, C. C. P. (2021). Multi-trait selection index and cluster analyses in Angus cattle. *J. Agric. Sci.* 159, 455–462. doi: 10.1017/S0021859621000575
- Gilmour, A., Gogel, B., Cullis, B., and Thompson, R. (2009). *ASReml user guide release 3.0* (Hemel Hempstead, HP1 1ES, United Kingdom: VSN International Ltd).
- Harville, D. A. (1975). Index selection with proportionality constraints. *Biometrics* 31, 223–225. doi: 10.2307/2529722
- Hazel, L. N. (1943). The genetic basis for constructing selection indexes. *Genetics* 8, 476–490. doi: 10.1093/genetics/28.6.476
- Hazel, L. N., Dickerson, G. E., and Freeman, A. E. (1994). The selection index—Then, now, and for the future. *J. dairy Sci.* 77, 3236–3251. doi: 10.3168/jds.S0022-0302(94)77265-9
- He, S., Thistlethwaite, R., Forrest, K., Shi, F., Hayden, M. J., Trethowan, R., et al. (2019). Extension of a haplotype-based genomic prediction model to manage multi-environment wheat data using environmental covariates. *Theor. Appl. Genet.* 132, 3143–3154. doi: 10.1007/s00122-019-03413-1
- Heffner, E. L., Jannink, J. L., and Sorrells, M. E. (2011). Genomic selection accuracy using multifamily prediction models in a wheat breeding program. *Plant Genome* 4, 65–75. doi: 10.3835/plantgenome.2010.12.0029
- Itoh, Y., and Yamada, Y. (1986). Re-examination of selection index for desired gains. *Genet. Sel Evol.* 18, 499–504. doi: 10.1186/1297-9686-18-4-499
- Joukhadar, R., and Daetwyler, H. D. (2022). Data integration, imputation and meta-analysis meta-analysis for genome-wide association studies in *Genome-wide association studies* (Springer US, New York, NY), 173–183.
- Joukhadar, R., Thistlethwaite, R., Trethowan, R. M., Hayden, M. J., Stangoulis, J., Cu, S., et al. (2021b). Genomic selection can accelerate the biofortification of spring wheat. *Theor. Appl. Genet.* 134, 3339–3350. doi: 10.1007/s00122-021-03900-4
- Joukhadar, R., Thistlethwaite, R., Trethowan, R., Keeble-Gagnère, G., Hayden, M. J., Ullah, S., et al. (2021a). Meta-analysis of genome-wide association studies reveal common loci controlling agronomic and quality traits in a wide range of normal and heat stressed environments. *Theor. Appl. Genet.* 134, 2113–2127. doi: 10.1007/s00122-021-03809-y
- Joukhadar, R., Hollaway, G., Shi, F., Kant, S., Forrest, K., Wong, D., et al. (2020). Genome-wide association reveals a complex architecture for rust resistance in 2300 worldwide bread wheat accessions screened under various Australian conditions. *Theor. Appl. Genet.* 133, 2695–2717.
- Juliana, P., Singh, R. P., Singh, P. K., Crossa, J., Huerta-Espino, J., Lan, C., et al. (2017). Genomic and pedigree-based prediction for leaf, stem, and stripe rust resistance in wheat. *Theor. Appl. Genet.* 130, 1415–1430. doi: 10.1007/s00122-017-2897-1
- Keeble-Gagnère, G., Pasam, R., Forrest, K. L., Wong, D., Robinson, H., Godoy, J., et al. (2021). Novel design of imputation-enabled snp arrays for breeding and research applications supporting multi-species hybridization. *Front. Plant Sci.* 12, 756877. doi: 10.3389/fpls.2021.756877
- Kempthorne, O., and Nordskog, A. W. (1959). Restricted selection indices. *Biometrics* 15, 10–19. doi: 10.2307/2527598
- Lee, S. H., and Van der Werf, J. H. (2016). MTG2: an efficient algorithm for multivariate linear mixed model analysis based on genomic information. *Bioinformatics* 32, 1420–1422. doi: 10.1093/bioinformatics/btw012
- Li, Y., Dungey, H., Yanchuk, A., and Apiolaza, L. A. (2017). Improvement of non-key traits in radiata pine breeding programme when long-term economic importance is uncertain. *PLoS One* 12, e0177806. doi: 10.1371/journal.pone.0177806
- Mallard, J. (1972). The theory and computation of selection indices with constraints: a critical synthesis. *Biometrics* 28, 713–735. doi: 10.2307/2528758
- McDonald, G. K. (1992). Effects of nitrogenous fertilizer on the growth, grain yield and grain protein concentration of wheat. *Aust. J. Agric. Res.* 43, 949–967. doi: 10.1071/AR9920949
- Money, D., Gardner, K., Migicovsky, Z., Schwaning, H., Zhong, G. Y., and Myles, S. (2015). LinkImpute: fast and accurate genotype imputation for nonmodel organisms. *G3: Genes, Genomes, Genetics* 5, 2383–2390.
- Pesek, J., and Baker, R. J. (1969). Desired improvement in relation to selection indices. *Can. J. Plant Sci.* 49, 803–804. doi: 10.4141/cjps69-137
- Richardson, C. M., Sunduimijid, B., Amer, P., van den Berg, I., and Pryce, J. E. (2021). A method for implementing methane breeding values in Australian dairy cattle. *Anim. Product. Sci.* 61, 1781–1787. doi: 10.1071/AN21055
- Smith, H. F. (1936). A discriminant function for plant selection. *Ann. Eugen.* 7, 240–250. doi: 10.1111/j.1469-1809.1936.tb02143.x
- Tallis, G. M. (1985). Constrained selection. *Jpn. J. Genet.* 60, 151–155. doi: 10.1266/jgg.60.151

- Thistlethwaite, R. J., Tan, D. K., Bokshi, A. I., Ullah, S., and Trethowan, R. M. (2020). A phenotyping strategy for evaluating the high-temperature tolerance of wheat. *Field Crops Res.* 255, 107905. doi: 10.1016/j.fcr.2020.107905
- Togashi, K., Lin, C. Y., and Yamazaki, T. (2011). The efficiency of genome wide selection for genetic improvement of net merit. *J. Anim. Sci.* 89, 2972–2980. doi: 10.2527/jas.2009-2606
- VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91, 4414–4423. doi: 10.3168/jds.2007-0980
- Watson, A., Hickey, L. T., Christopher, J., Rutkoski, J., Poland, J., and Hayes, B. J. (2019). Multivariate genomic selection and potential of rapid indirect selection with speed breeding in spring wheat. *Crop Sci.* 59, 1945–1959. doi: 10.2135/cropsci2018.12.0757
- Wang, S., Wong, D., Forrest, K., Allen, A., Chao, A., Huang, B. E., et al (2014). Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. *Plant Biotechnol. J.* 12, 787–796.
- Yamada, Y., Yokouchi, K., and Nishida, A. (1975). Selection index when genetic gains of individual traits are of primary concern. *Jpn J. Genet.* 50, 33–41. doi: 10.1266/jjg.50.33