



Bayesian Optimization Approaches for Identifying the Best Genotype from a Candidate Population

Shin-Fu TSAI, Chih-Chien SHEN, and Chen-Tuo LIAO 

Bayesian optimization is incorporated into genomic prediction to identify the best genotype from a candidate population. Several expected improvement (EI) criteria are proposed for the Bayesian optimization. The iterative search process of the optimization consists of two main steps. First, a genomic BLUP (GBLUP) prediction model is constructed using the phenotype and genotype data of a training set. Second, an EI criterion, estimated from the resulting GBLUP model, is employed to select the individuals that are phenotyped and added to the current training set to update the GBLUP model until the sequential observed EI values are less than a stopping tolerance. Three real datasets are analyzed to illustrate the proposed approach. Furthermore, a detailed simulation study is conducted to compare the performance of the EI criteria. The simulation results show that one augmented version derived from the distribution of predicted genotypic values is able to identify the best genotype from a large candidate population with an economical training set, and it can therefore be recommended for practical use. Supplementary materials accompanying this paper appear on-line.

Key Words: Expected improvement; GBLUP model; Genomic prediction; Genomic selection; Linear mixed effects model.

1. INTRODUCTION

Accelerating the genetic improvement of crops through innovative breeding techniques and strategies is an urgent and important task when dealing with the problem of food security (2010). Owing to the long history of domestication and selection in agriculture, elite breeding populations or commercial cultivars have lost their genetic diversity (Reif et al. 2005). Thus, introgression of rich variation from wild, exotic, or indigenous germplasm has become essential to promote genetic diversity and to enhance the efficiency of plant breeding programs (McCouch et al. 2013). Plant breeders will therefore be faced with the

S.-F. Tsai · C.-C. Shen · C.-T. Liao (✉) Division of Biometry, Institute of Agronomy, National Taiwan University, No. 1, Sec. 4, Roosevelt Rd., Taipei 10617, Taiwan (E-mail: ctliao@ntu.edu.tw) .

© 2021 The Author(s)

Journal of Agricultural, Biological, and Environmental Statistics, Volume 26, Number 4, Pages 519–537
<https://doi.org/10.1007/s13253-021-00454-2>

need to identify superior accessions from germplasm collections. In this study, we attempt to provide a solution to this problem via genomic prediction (GP).

GP, which has become an increasingly popular approach in plant or animal breeding, takes advantage of dense DNA markers over an entire genome to predict genotypic values and then applies these to select superior individuals from genetic resources (Meuwissen et al. 2001). Typically, a training set with known genotype and phenotype data is employed to construct a GP model. The trained GP model is then used to predict genomic estimated breeding values (GEBVs) for the individuals of a breeding population with known genotype data. A linear regression model fitting the phenotypic values by marker effects is often implemented for GP. However, it is challenging to estimate all the regression coefficients in such a GP model, because the number of marker effects is usually much greater than the number of observed phenotypic values (the large- p -with-small- n problem). Therefore, shrinkage methods such as ridge regression, the least absolute shrinkage and selector operator (LASSO) (Tibshirani 1996), and elastic nets have been applied to estimate marker effects in the whole-genome regression model. Also, there are two kinds of linear mixed effects models for GP: (i) best linear unbiased prediction (BLUP) based on markers and (ii) BLUP based on a relationship matrix. In a marker-based model, marker effects are treated as random effects and GEBVs of individuals are obtained by multiplying their marker scores by these BLUP estimates. The ridge regression BLUP (rr-BLUP) model (Meuwissen et al. 2001) adopts this approach. In a relationship matrix-based model, genotypic values of individuals are treated directly as random effects and estimated through a genomic relationship matrix. The genomic BLUP (GBLUP) model (Habier et al. 2007) is in the spirit. Tempelman (2015) provided an in-depth review and discussion of the above GP models and noted that great care should be taken in correctly inferring hyperparameters using hierarchical Bayesian models in GP.

In the genomic era, genotyping costs have fallen dramatically, while phenotyping costs have remained relatively constant (Akdemir and Sanchez 2019). It would therefore be advantageous to sample individuals for selective phenotyping in more than one stage, because a sequential phenotyping approach can reduce the size of the required training set. Recently, Tanaka and Iwata (2018) proposed such an approach to discover the best genotype from a candidate population. They developed their procedure by incorporating Bayesian optimization into GP. Bayesian optimization is a powerful tool for solving black-box global optimization problems with computationally expensive function evaluation (Jones et al. 1998). The classical Bayesian optimization problem can be formulated as follows. The goal of the Bayesian optimization is to maximize or minimize an objective function $f(\mathbf{x})$ within a feasible space Ω . That is, let \mathbf{x}^* be the desired optimum; then,

$$\mathbf{x}^* = \operatorname{argmax}_{\mathbf{x} \in \Omega} f(\mathbf{x}).$$

Also, let y denote the response of an experiment conducted to evaluate the objective function; then

$$y = f(\mathbf{x}) + e,$$

where e denotes the random error. The experiment is said to be noisy if the response is measured with random error. Typically, the evaluation function is deterministic (fixed effect), and the feasible space is a continuous and compact region.

The optimization usually begins by evaluating a small number of randomly selected points and fitting a Gaussian process to the response. The Gaussian posterior provides an estimate of the object function at each point in the feasible space, as well as the uncertainty of this estimate (Letham et al. 2019). An acquisition function, such as the expected improvement (EI), is constructed based on the posterior estimation for determining new query points to evaluate the object function. The choice of the new query points should be a trade-off between exploration and exploitation so that one can optimize the objective function using as few query points as possible (Gong et al. 2019). For the classical Bayesian optimization problem of a continuous and compact feasible region, except in a few atypical cases, the EI has been proved to generally converge to the optimum for both noisy experiments (Vazquez et al. 2008; Letham et al. 2019) and noiseless function evaluation (Bull 2011). Additionally, the EI-based approach has been well developed to tackle various practical problems (Jordan et al. 2011; Shahriari et al. 2016).

Tanaka and Iwata (2018) considered the following GBLUP model:

$$\mathbf{y} = \mu \mathbf{1}_n + \mathbf{g} + \mathbf{e}, \tag{1}$$

where \mathbf{y} denotes the vector of phenotypic values, μ the general mean, $\mathbf{1}_n$ the unit vector of length n , \mathbf{g} the vector of genotypic values, and \mathbf{e} the vector of random errors. It is assumed that \mathbf{g} follows a multivariate normal distribution, denoted by $\mathbf{g} \sim MVN(\mathbf{0}, \sigma_g^2 \mathbf{K})$, where $\mathbf{0}$ is the zero vector, σ_g^2 is the additive genetic variance, and \mathbf{K} is a known genomic relationship matrix. Also, the vector of random errors $\mathbf{e} \sim MVN(\mathbf{0}, \sigma_e^2 \mathbf{I}_n)$, where σ_e^2 is the random error variance and \mathbf{I}_n is the identity matrix of order n . Moreover, \mathbf{g} and \mathbf{e} are assumed to be mutually independent. The evaluation function here is the unknown genotypic value for an individual, which is evaluated through its DNA marker scores (genotype). Interestingly, the evaluation function of genotypic value is stochastic (random effect) and the feasible space is a finite candidate population set in the GBLUP model, which are different from those in the classic Bayesian optimization problem. Tanaka and Iwata (2018) proposed an EI criterion and showed their Bayesian optimization approach to have advantage compared with the selection method based on posterior means and with a random sampling method. The aim of our study is to investigate potential EI-based Bayesian optimization approaches for identifying the best genotype from a large candidate population.

The rest of this article is organized as follows. In Sect. 2, we derive some modified versions of the EI criterion proposed by Tanaka and Iwata (2018). In Sect. 3, we present the search process of the Bayesian optimization approach and a stopping rule for real data analysis. We illustrate the proposed approach through analysis of two rice (*Oryza sativa* L.) and one wheat (*Triticum aestivum* L.) genome datasets in Sect. 4. We further conduct a detailed simulation study to compare the performance of the EI criteria in Sect. 5. Some concluding remarks are provided in Sect. 6.

2. THE EXPECTED IMPROVEMENT CRITERIA

It is known that the genomic relationship matrix \mathbf{K} plays an important role in building the GBLUP model (1). We considered the case where $\mathbf{K} = \mathbf{M}\mathbf{M}^\top/p$, where \mathbf{M} is the

standardized marker score matrix and p is the number of DNA markers. Let m_{ij} denote the (ij) th element of \mathbf{M} ; then $m_{ij} = (a_{ij} - \bar{a}_j)/s_j$, where a_{ij} is the (ij) th element of the original marker score matrix, and \bar{a}_j and s_j are, respectively, the sample mean and the sample standard deviation of the elements in column j . We first present the distributions of predicted phenotypic values (PPVs) and predicted genotypic values (PGVs), and then derive several EI criteria.

2.1. DISTRIBUTIONS OF PREDICTED PHENOTYPIC AND GENOTYPIC VALUES

Let \mathbf{y}_1 and \mathbf{g}_1 , respectively, denote the vectors of phenotypic and genotypic values for the training set, both of which are of size n_1 . Likewise, let \mathbf{y}_2 and \mathbf{g}_2 denote the corresponding vectors for the remaining n_2 individuals not chosen in the training set (the non-phenotyped set), where $n_1 + n_2 = n$. Therefore, the GBLUP model (1) can be partitioned as

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mu \mathbf{1}_{n_1} \\ \mu \mathbf{1}_{n_2} \end{bmatrix} + \begin{bmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix}$$

where

$$\begin{bmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \end{bmatrix} \sim MVN\left(\mathbf{0}, \sigma_g^2 \begin{bmatrix} \mathbf{K}_{11} & \mathbf{K}_{12} \\ \mathbf{K}_{21} & \mathbf{K}_{22} \end{bmatrix}\right).$$

We use the R package BGLR (Perez and de los Campos, G. 2014) to perform the Bayesian reproducing kernel Hilbert space (RKHS) method for estimating μ, σ_g^2 and σ_e^2 , and the BLUP of \mathbf{g}_1 using the training set. More precisely, the estimation is based only on the model

$$\mathbf{y}_1 = \mu \mathbf{1}_{n_1} + \mathbf{g}_1 + \mathbf{e}_1,$$

where

$$\mathbf{g}_1 \sim MVN(\mathbf{0}, \sigma_g^2 \mathbf{K}_{11}).$$

The resulting estimated values are denoted by $\hat{\mu}, \hat{\sigma}_g^2, \hat{\sigma}_e^2$, and $\hat{\mathbf{g}}_1$. We refer to Xavier et al. (2016) for more details regarding the Bayesian RKHS method.

Under the condition that $\hat{\mu}, \hat{\sigma}_g^2, \hat{\sigma}_e^2$, and $\hat{\mathbf{g}}_1$ are all assumed to be fixed and known values, the distribution of PPVs for the non-phenotyped set, i.e., the distribution of \mathbf{y}_2 conditioned by $\mathbf{y}_1 = \hat{\mathbf{y}}_1$, is given by

$$\tilde{\mathbf{y}}_2 | (\hat{\mu}, \hat{\sigma}_g^2, \hat{\sigma}_e^2, \hat{\mathbf{g}}_1) \sim MVN(\hat{\boldsymbol{\mu}}_p, \hat{\boldsymbol{\Sigma}}_p), \tag{2}$$

where $\hat{\boldsymbol{\mu}}_p = \hat{\mu} \mathbf{1}_{n_2} + \mathbf{K}_{21} \mathbf{K}_{11}^{-1} \hat{\mathbf{g}}_1$ and $\hat{\boldsymbol{\Sigma}}_p = \hat{\boldsymbol{\Sigma}}_{22} - \hat{\boldsymbol{\Sigma}}_{21} \hat{\boldsymbol{\Sigma}}_{11}^{-1} \hat{\boldsymbol{\Sigma}}_{12}$. Here, $\hat{\boldsymbol{\Sigma}}_{22} = \hat{\sigma}_g^2 \mathbf{K}_{22} + \hat{\sigma}_e^2 \mathbf{I}_{n_2}$, $\hat{\boldsymbol{\Sigma}}_{21} = \hat{\sigma}_g^2 \mathbf{K}_{21}$, $\hat{\boldsymbol{\Sigma}}_{11} = \hat{\sigma}_g^2 \mathbf{K}_{11} + \hat{\sigma}_e^2 \mathbf{I}_{n_1}$, and $\hat{\boldsymbol{\Sigma}}_{12} = \hat{\sigma}_g^2 \mathbf{K}_{12}$. Likewise, the distribution of PGVs for the non-phenotyped set, i.e., the distribution of \mathbf{g}_2 conditioned by $\mathbf{g}_1 = \hat{\mathbf{g}}_1$, is given by

$$\tilde{\mathbf{g}}_2 | (\hat{\sigma}_g^2, \hat{\mathbf{g}}_1) \sim MVN(\hat{\boldsymbol{\mu}}_g, \hat{\boldsymbol{\Sigma}}_g), \tag{3}$$

where $\hat{\boldsymbol{\mu}}_g = \mathbf{K}_{21} \mathbf{K}_{11}^{-1} \hat{\mathbf{g}}_1$ and $\hat{\boldsymbol{\Sigma}}_g = \hat{\sigma}_g^2 (\mathbf{K}_{22} - \mathbf{K}_{21} \mathbf{K}_{11}^{-1} \mathbf{K}_{12})$. Note that $\tilde{\mathbf{g}}_2 = \mathbf{K}_{21} \mathbf{K}_{11}^{-1} \tilde{\mathbf{g}}_1$, where $\tilde{\mathbf{g}}_1$ is the BLUP of \mathbf{g}_1 . From Henderson (1977), it follows that $\tilde{\mathbf{g}}_2$ is the BLUP of

\mathbf{g}_2 . Obviously, $\tilde{\mathbf{g}}_2$ differs from $\tilde{\mathbf{y}}_2$ in distribution by the term $\hat{\mu}$ in the mean vector and the term $\hat{\sigma}_e^2$ in the covariance matrix. That is, $\hat{\mu}_p = \hat{\mu}_g$ and $\hat{\Sigma}_p = \hat{\Sigma}_g$ if $\hat{\mu} = \hat{\sigma}_e^2 = 0$. In other words, the PPVs are affected by the random error, but the PGVs are free of this variation. The best genotype for a target trait is defined as the individual with the maximal or minimal genotypic value among the candidate population. Intuitively, an EI criterion derived from the distribution of PGVs rather than PPVs should be more efficient in identifying the desired best genotype. A construction method for the statistical tolerance interval for a distribution free of measurement errors like that of PGVs can be found in [Lin et al. \(2008\)](#).

2.2. PLUG-IN EXPECTED IMPROVEMENT CRITERIA

[Tanaka and Iwata \(2018\)](#) considered the EI criterion based on the distribution of PPVs in (2). Let $\hat{\mathbf{y}}_1 = \hat{\mu}\mathbf{1}_{n_1} + \hat{\mathbf{g}}_1$, representing the vector of estimated phenotyped values for the training set, and \hat{f}_{Mp} be the maximal value among $\hat{\mathbf{y}}_1$. Also, let the marginal distribution for each individual in (2) be denoted by

$$\tilde{y}_{2i} \sim N(\hat{\mu}_{pi}, \hat{\sigma}_{pi}^2).$$

The improvement function for \tilde{y}_{2i} is defined as

$$Im(\tilde{y}_{2i}) = \begin{cases} 0 & \text{if } \tilde{y}_{2i} < \hat{f}_{Mp}, \\ \tilde{y}_{2i} - \hat{f}_{Mp} & \text{otherwise.} \end{cases} \tag{4}$$

Here, $Im(\tilde{y}_{2i})$ is a random variable associated with the distribution of \tilde{y}_{2i} . From [Jones et al. \(1998\)](#), the expected improvement function has an elegant closed form given by

$$EI(\tilde{y}_{2i}) = (\hat{\mu}_{pi} - \hat{f}_{Mp})\Phi(Z_{pi}) + \hat{\sigma}_{pi}\phi(Z_{pi}), \tag{5}$$

where $Z_{pi} = (\hat{\mu}_{pi} - \hat{f}_{Mp})/\hat{\sigma}_{pi}$, $\Phi(\cdot)$ is the cumulative density function of the standard normal distribution, and $\phi(\cdot)$ is the probability density function of the standard normal distribution. The EI in (5) was referred as a ‘‘plug-in’’ strategy by [Picheny et al. \(2013\)](#) because \hat{f}_{Mp} is used for a noisy experiment. We simply call it EI-PPV. Alternatively, we consider the plug-in EI criterion based on the distribution of PGVs in (3). Likewise, we have the EI-PGV given by

$$EI(\tilde{g}_{2i}) = (\hat{\mu}_{gi} - \hat{f}_{Mg})\Phi(Z_{gi}) + \hat{\sigma}_{gi}\phi(Z_{gi}), \tag{6}$$

where $\tilde{g}_{2i} \sim N(\hat{\mu}_{gi}, \hat{\sigma}_{gi}^2)$ (the corresponding marginal distribution in (3)) and \hat{f}_{Mg} is the maximal value among $\hat{\mathbf{g}}_1$ and $Z_{gi} = (\hat{\mu}_{gi} - \hat{f}_{Mg})/\hat{\sigma}_{gi}$.

2.3. FORWARD EXPECTED IMPROVEMENT CRITERIA

Clearly, both of the above EI-PPV and EI-PGV ignore the genomic correlation among the candidate individuals. Thus, we take into account in the criterion the property that selection

of future individuals could be associated with the current selected individual. We rewrite the distribution of PPVs as

$$\tilde{y}_2 = \begin{bmatrix} \tilde{y}_* \\ \tilde{y}_c \end{bmatrix}, \quad \hat{\mu}_p = \begin{bmatrix} \hat{\mu}_* \\ \hat{\mu}_{*c} \end{bmatrix}, \quad \hat{\Sigma}_p = \begin{bmatrix} \hat{\sigma}_*^2 & \hat{\Sigma}_{p12} \\ \hat{\Sigma}_{p21} & \hat{\Sigma}_{p22} \end{bmatrix},$$

where \tilde{y}_* represents the PPV of the individual with the largest $EI(\tilde{y}_{2i})$ from (5), i.e., the individual with \tilde{y}_* is the first chosen from the non-phenotyped set. Subsequently, we search for the next individual with the largest $EI(\tilde{y}_{2i})$ among the remaining candidate individuals whose PPVs follow the conditional distribution

$$\tilde{y}_c | (\tilde{y}_* = \hat{\mu}_*) \sim MVN(\hat{\mu}_{cp}, \hat{\Sigma}_{cp}), \tag{7}$$

where $\hat{\mu}_{cp} = \hat{\mu}_{*c}$ and $\hat{\Sigma}_{cp} = \hat{\Sigma}_{p22} - \hat{\Sigma}_{p21}(\hat{\sigma}_*^2)^{-1}\hat{\Sigma}_{p12}$. We here adjust the distribution of \tilde{y}_c for the mean value of \tilde{y}_* . Let

$$\tilde{y}_{ci} \sim N(\hat{\mu}_{cpi}, \hat{\sigma}_{cpi}^2),$$

representing a marginal distribution for individual i in (7); then, the corresponding EI can be derived as

$$EI(\tilde{y}_{ci}) = (\hat{\mu}_{cpi} - \hat{f}_{Mp})\Phi(Z_{cpi}) + \hat{\sigma}_{cpi}\phi(Z_{cpi}), \tag{8}$$

where $Z_{cpi} = (\hat{\mu}_{cpi} - \hat{f}_{Mp})/\hat{\sigma}_{cpi}$. Similarly, we have the corresponding EI for the distribution of the PGVs as

$$EI(\tilde{g}_{ci}) = (\hat{\mu}_{cgi} - \hat{f}_{Mg})\Phi(Z_{cgi}) + \hat{\sigma}_{cgi}\phi(Z_{cgi}), \tag{9}$$

where $Z_{cgi} = (\hat{\mu}_{cgi} - \hat{f}_{Mg})/\hat{\sigma}_{cgi}$. We abbreviate the EI in (8) as EI-PPV-fwd, and the EI in (9) as EI-PGV-fwd. Note that ‘‘fwd’’ here means forward selection. For a specific candidate individual i , $\hat{\mu}_{cpi} = \hat{\mu}_{pi}$ and $\hat{\sigma}_{cpi} \leq \hat{\sigma}_{pi}$. That is to say, EI-PPV-fwd has a smaller prediction standard error than EI-PPV for a candidate individual. Similarly, the prediction standard error in EI-PGV-fwd is smaller than that in EI-PGV.

2.4. AUGMENTED EXPECTED IMPROVEMENT CRITERIA

Huang et al. (2006) proposed an augmented EI criterion to handle noisy experiments. We adopt such an augmented version for the above EI-PGV and EI-PGV-fwd. The augmented EI-PGV is modified from (6) and is given by

$$Aug.EI(\tilde{g}_{2i}) = [(\hat{\mu}_{gi} - \hat{f}_{Mg}^*)\Phi(Z_{gi}^*) + \hat{\sigma}_{gi}\phi(Z_{gi}^*)] \left(1 - \frac{\hat{\sigma}_e}{\sqrt{\hat{\sigma}_{gi}^2 + \hat{\sigma}_e^2}} \right), \tag{10}$$

and the augmented EI-PGV-fwd is modified from (9) and is given by

$$Aug.EI(\tilde{g}_{ci}) = [(\hat{\mu}_{cgi} - \hat{f}_{Mg}^*)\Phi(Z_{cgi}^*) + \hat{\sigma}_{cgi}\phi(Z_{cgi}^*)] \left(1 - \frac{\hat{\sigma}_e}{\sqrt{\hat{\sigma}_{cgi}^2 + \hat{\sigma}_e^2}} \right), \quad (11)$$

where \hat{f}_{Mg}^* denotes the current “effective best solution,” $Z_{gi}^* = (\hat{\mu}_{gi} - \hat{f}_{Mg}^*)/\hat{\sigma}_{gi}$, and $Z_{cgi}^* = (\hat{\mu}_{cgi} - \hat{f}_{Mg}^*)/\hat{\sigma}_{cgi}$. The current effective best solution \hat{f}_{Mg}^* is obtained from \hat{g}_1 and maximizes the utility function

$$u(\hat{g}_{1i}) = \hat{g}_{1i} - \gamma\hat{\sigma}_{1gi},$$

where γ is a constant that reflects the degree of risk aversion. That is,

$$\hat{f}_{Mg}^* = \operatorname{argmax}_{\hat{g}_{1i} \in \hat{g}_1} (\hat{g}_{1i} - \gamma\hat{\sigma}_{1gi}).$$

We specify $\gamma = 1$, which implies that 1 unit of \hat{g}_{1i} is traded for 1 unit of $\hat{\sigma}_{1gi}$. Following Huang et al. (2006), \hat{f}_{Mg}^* is used to reduce prediction uncertainty, and the heuristic multiplier of the EI accounts for the diminishing return of additional replicates as the prediction becomes more accurate. The EIs in (10) and (11) are abbreviated as aug-EI-PGV and aug-EI-PGV-fwd, respectively.

The above EI criteria are all developed for the situation in which the best genotype is assumed to give the maximal genotypic value for some target trait such as grain yield of a crop. The corresponding EI criteria for the situation in which the best genotype is desired to give the minimal genotypic value, such as plant height or flowering time of a crop, are provided in the Appendix.

3. SEARCH PROCESS

For a given candidate population in which all the individuals have been genotyped but not phenotyped, the iterative search process to identify the best genotype from the candidate population can be described as follows.

- Step 0:** Randomly choose n_0 individuals from the candidate population to be phenotyped for a target trait. The chosen individuals serve as the starting training set, denoted by S_0 . Initialize $n_{tr} \leftarrow n_0$ and $S_{tr} \leftarrow S_0$, where S_{tr} denotes the current training set and n_{tr} is its sample size.
- Step 1:** Perform the Bayesian RKHS method using the phenotype and genotype data of S_{tr} to construct the GBLUP model, and obtain the estimated values of $\hat{\mu}$, $\hat{\sigma}_g^2$, $\hat{\sigma}_e^2$, and \hat{g}_1 .
- Step 2:** Select n_{sel} individuals from the non-phenotyped set using each of the above EI criteria.

Step 3: If the average of the realized EI values among the n_{sel} selected individuals is less than a stopping tolerance δ , then STOP the search and GO TO Step 4. Otherwise, phenotype the selected n_{sel} individuals, denoted by \mathcal{S}_{sel} , and add them to the current training set. That is, the union of \mathcal{S}_{tr} and \mathcal{S}_{sel} forms the new training set, expressed as $\mathcal{S}_{tr} \leftarrow \mathcal{S}_{tr} \cup \mathcal{S}_{sel}$. Similarly, $n_{tr} \leftarrow n_{tr} + n_{sel}$. GO TO Step 1.

Step 4: Identify the individual with the maximal (or minimal) estimated genotypic value among the current selected individuals as the best genotype.

Since the candidate population is a finite set, it is easy to argue that all the EI criteria converge to 0. For example, from the definition of the improvement function in (4), the EI value becomes equal to 0 if all $\tilde{y}_{2i} < \hat{f}_{Mp}$, and this surely happens for a finite candidate set. In other words, \hat{f}_{Mp} definitely can be found after all candidate individuals have been evaluated.

4. ILLUSTRATIVE EXAMPLES

We use the following three genome datasets to illustrate the search process.

EXAMPLE 1: 44K RICE GENOME DATASET

This dataset was presented in Zhao et al. (2011) and contains 413 accessions and 44,100 single nucleotide polymorphism (SNP) markers. This dataset consists of five subpopulations and one admixed group. SNPs with missing rate > 0.05 and minor allele frequency < 0.05 were removed from the dataset, with the result that 36,901 out of the 44,100 SNPs were retained. The SNP genotype at each locus was coded as -1 , 0 , or 1 , where 1 indicates the homogeneous genotype of the major allele, -1 the homogeneous genotype of the minor allele, and 0 the heterogeneous genotype. After SNP coding, any missing locus was taken as 1 . Only 300 out of the 413 accessions with no missing phenotypic values on the flowering time in all three locations—Arkansas (FT-Ark), Faridpur (FT-Far), and Aberdeen (FT-Abe)—were used for the demonstration. We fixed the stopping tolerance as $\delta = 0.1$ for the above three target traits.

Example 2: Tropical Rice Genome Dataset

This dataset was presented in Spindle et al. (2015) and contains 73,147 SNP markers and 363 elite breeding lines belonging to indica or indica-admixed group. Phenotypic observations included 4 years (2009–2012, with two seasons (dry and wet) per year) of data on grain yield (GY), flowering time (FT), and plant height (PH), although PH data were not available for the wet season of 2009. Phenotypic values of 35 out of the 363 individuals were missing; therefore, adjusted means of only 328 individuals were used for this example. Additionally, only 10,772 out of the 73,147 SNPs were used in this study. One SNP marker was selected per 0.1 cM interval over each chromosome, because the chosen subset of the full marker set has been shown to be sufficient for genomic selection in this collection of

rice germplasm (Spindle et al. 2015). The SNP coding was the same as in Example 1. We set the stopping tolerance as $\delta = 0.01$ for this dataset.

Example 3: Wheat Genome Dataset

This dataset was used in Crossa et al. (2010) and contains 599 historical wheat lines from the International Maize and Wheat Improvement Center (CIMMYT) and 1279 diversity arrays technology (DArT) markers. The wheat dataset is from CIMMYT's Global Wheat Program, which conducted numerous international trials across a variety of producing environments. These environments were grouped into four mega-environments (E1 to E4) comprising four main agro-climatic regions. The phenotypic value was the average grain yield of the 599 wheat lines evaluated in each of these four mega-environments. The marker score for each DArT marker was coded as 1 for presence and 0 for absence. We set the stopping tolerance as $\delta = 0.001$ for all the target traits E1 to E4.

We set $n_0 = 10$ and $n_{sel} = 20$ in the search process for all three of the above datasets. The sample size of the training set when the search was stopped and the identified best genotype according to each of the EI criteria are displayed in Table 1, from which it can be seen that most of the EI criteria found a common best genotype for a specific target trait. However, the final sample sizes of the training set may differ from different EI criteria, and even from different traits within the same dataset. Certainly, the specification of the stopping tolerance and the convergence rate of the EI criteria can have a significant impact on the final sample size. The convergence process of the EI criteria for the three datasets is displayed in Figures S1 to S3 of the Supplementary Materials. The figures reveal that those EI criteria considering the random error variance, EI-PPV, EI-PPV-fwd, aug-EI-PGV, and aug-EI-PGV-fwd, have relatively quick convergence rates and smooth curves compared with EI-PGV, and EI-PGV-fwd, both of which don't take $\hat{\sigma}_e^2$ into account. In particular, the convergence process of aug-EI-PGV appears to be the fastest and stablest for most of the cases. An R code for the analysis of the tropical rice dataset by the aug-EI-PGV criterion is provided in the Supplementary Materials.

5. SIMULATION STUDY

We further conducted a simulation study to compare the EI criteria using the above three datasets as templates. The genomic relationship matrix \mathbf{K} was first calculated from all available genotype data for each dataset, and then the GBLUP model (1) was employed to generate the simulated datasets. Genotype data of 373, 328, and 599 individuals for the 44k rice, tropical rice, and wheat datasets, respectively, were used in this simulation study. The procedure of the simulation study can be described as follows.

- Step 1:** Generate the vector of genotypic values \mathbf{g} from $MVN(\mathbf{0}, \sigma_g^2 \mathbf{K})$, where σ_g^2 is set to be 25. The simulated \mathbf{g} is treated as the vector of true genotypic values, and the individual with the largest genotypic value is assumed to be the best genotype.

Table 1. Sample sizes of the required training sets and the identified best genotypes (in parentheses) for the three datasets

Dataset	Trait	EI criteria [†]					
		PPV	PGV	PPV-fwd	PGV-fwd	aPGV	aPGV-fwd
44k rice	FT-Ark	70 (#192)	70 (#192)	110 (#164)	110 (#192)	70 (#164)	70 (#164)
	FT-Far	70 (#161)	50 (#226)	70 (#226)	70 (#226)	50 (#161)	50 (#161)
	FT-Abe	110 (#186)	270 (#186)	170 (#186)	290 (#186)	110 (#161)	150 (#161)
Tropical rice	GY	328* (#94)	250 (#94)	328* (#94)	310 (#94)	230 (#94)	310 (#94)
	FT	210 (#92)	150 (#92)	150 (#92)	70 (#96)	250 (#92)	70 (#11)
	PH	110 (#261)	70 (#197)	50 (#261)	90 (#136)	50 (#197)	50 (#136)
Wheat	E1	90 (#367)	90 (#367)	110 (#367)	110 (#222)	90 (#367)	110 (#278)
	E2	130 (#463)	150 (#463)	310 (#463)	190 (#559)	110 (#463)	130 (#463)
	E3	110 (#594)	130 (#478)	150 (#594)	210 (#594)	110 (#594)	110 (#594)
	E4	70 (#537)	110 (#537)	170 (#549)	170 (#564)	70 (#537)	110 (#549)

[†] PPV denotes EI-PPV; PGV denotes EI-PGV; PPV-fwd denotes EI-PPV-fwd; PGV-fwd denotes EI-PGV-fwd; aPGV denotes aug-EI-PGV; aPGV-fwd denotes aug-EI-PGV-fwd.

* indicates the case in which all candidate individuals have been evaluated

Step 2: Generate the vector of random errors \mathbf{e} from $MVN(\mathbf{0}, \sigma_e^2 \mathbf{I}_n)$, where $\sigma_e^2 = \sigma_g^2(1 - h^2)/h^2$. Here, h^2 represents the trait heritability fixed at 0.2, 0.5, and 0.8 for low, intermediate, and high heritability, respectively.

Step 3: Repeat the above two steps 5,000 times to generate 5,000 simulated datasets of phenotypic values for each of the fixed values of h^2 . That is, $\mathbf{y} = \mu\mathbf{1} + \mathbf{g} + \mathbf{e}$, where μ is set to 100.

Step 4: Conduct steps 1 to 3 of the search process presented in Sect. 3 for each simulated dataset. Instead of the rule for stopping tolerance that was used in the real data analysis, we stop the search once the best genotype has been identified from a simulated dataset, and then record the number of the individuals that have been selected at this point. We set $n_0 = 10$ and $n_{sel} = 10, 20, 40$ for this simulation study.

The sample size of the training set required to discover the best genotype for each EI criterion obtained from Step 4 was used as a measure of the performance for the EI criterion. The mean and the standard deviation over the training set sizes from the 5,000 simulated datasets are reported in Table 2. As expected, the performance generally improves with increasing heritability h^2 and with decreasing size n_{sel} of the training set selected at each batch. Also, the EIs based on PGVs slightly outperform their counterparts based on PPVs. Interestingly,

Table 2. Mean and standard deviation (in parentheses) of the required training set sizes obtained from the 5,000 simulated datasets for the EI criteria

Dataset	n_{sel}	h^2	EI criteria [†]						
			PPV	PGV	PPV-fwd	PGV-fwd	aPGV	aPGV-fwd	
44k rice	10	0.2	80.03 (63.12)	76.02 (58.91)	97.63 (70.21)	94.95 (68.98)	76.19 (59.37)	96.60 (70.43)	
		0.5	68.56 (53.45)	63.90 (47.65)	84.06 (61.83)	82.55 (59.83)	62.79 (46.72)	81.25 (60.15)	
		0.8	59.42 (44.14)	55.60 (40.46)	73.28 (52.07)	70.41 (49.72)	54.26 (39.11)	69.69 (50.26)	
	20	0.2	83.13 (63.29)	79.53 (59.56)	106.14 (73.38)	101.73 (70.11)	79.22 (60.54)	103.35 (72.40)	
		0.5	72.58 (54.49)	67.94 (48.58)	91.11 (64.28)	89.55 (62.46)	66.24 (47.86)	88.35 (61.73)	
		0.8	63.89 (45.72)	60.39 (42.51)	81.11 (56.05)	78.91 (53.92)	58.24 (40.88)	78.23 (53.99)	
	40	0.2	90.31 (65.47)	85.32 (62.14)	112.48 (74.58)	109.09 (72.53)	83.79 (62.28)	110.38 (73.83)	
		0.5	79.04 (56.94)	74.21 (52.35)	100.26 (66.65)	97.70 (64.40)	71.41 (50.19)	97.55 (64.09)	
		0.8	70.55 (49.24)	66.53 (46.04)	90.71 (59.86)	88.06 (57.35)	64.53 (44.38)	87.87 (57.05)	
	Tropical rice	10	0.2	36.81 (43.66)	36.24 (42.47)	44.77 (50.99)	44.10 (49.43)	35.93 (41.78)	43.69 (48.88)
			0.5	34.49 (36.71)	34.15 (35.83)	42.77 (45.00)	42.68 (44.83)	34.01 (35.63)	42.18 (43.81)
			0.8	31.70 (31.88)	31.39 (31.01)	39.02 (39.75)	38.92 (38.98)	31.37 (30.88)	38.76 (38.92)
20		0.2	37.05 (44.04)	36.31 (42.31)	46.52 (53.35)	45.54 (51.42)	36.22 (42.23)	45.15 (50.89)	
		0.5	35.03 (37.20)	34.69 (36.67)	44.32 (47.48)	44.48 (47.45)	34.69 (36.80)	44.21 (47.44)	
		0.8	32.34 (32.82)	32.04 (32.05)	40.46 (41.36)	40.39 (41.19)	32.04 (32.04)	40.05 (40.97)	
40		0.2	37.56 (44.70)	36.72 (42.58)	46.50 (53.37)	46.23 (52.30)	36.22 (41.58)	45.59 (51.97)	
		0.5	36.16 (39.26)	35.70 (38.26)	45.47 (48.53)	45.73 (49.04)	35.72 (38.30)	45.19 (48.81)	
		0.8	33.56 (34.56)	33.32 (34.12)	41.69 (43.17)	41.85 (43.35)	33.21 (33.71)	41.44 (43.46)	
Wheat		10	0.2	105.12 (83.19)	103.98 (84.03)	132.40 (100.38)	131.26 (101.24)	106.17 (86.67)	134.34 (105.11)
			0.5	78.54 (59.93)	76.23 (58.44)	99.02 (75.69)	97.30 (75.74)	77.83 (60.40)	99.77 (76.64)
			0.8	67.99 (49.06)	65.86 (47.49)	84.63 (61.52)	81.63 (60.13)	66.34 (47.59)	84.05 (62.12)
	20	0.2	108.45 (83.83)	107.39 (84.52)	143.22 (107.05)	142.20 (106.97)	107.11 (85.75)	144.76 (109.48)	
		0.5	83.64 (61.99)	80.38 (59.92)	110.32 (81.07)	108.01 (80.21)	80.76 (59.80)	108.47 (80.97)	
		0.8	72.50 (50.81)	70.33 (49.01)	94.61 (67.41)	92.09 (65.62)	70.63 (49.09)	92.64 (65.56)	
	40	0.2	114.72	112.65	155.29	153.54	112.23	153.44	

Table 2. (Continued)

Dataset	n_{sel}	h^2	EI criteria [†]					
			PPV	PGV	PPV-fwd	PGV-fwd	aPGV	aPGV-fwd
			(86.83)	(85.23)	(111.78)	(111.44)	(87.12)	(112.23)
		0.5	90.07	86.96	122.92	121.13	87.08	120.33
			(64.21)	(61.57)	(85.71)	(85.56)	(62.33)	(85.24)
		0.8	79.69	77.52	105.59	105.14	77.42	105.47
			(54.12)	(52.62)	(72.10)	(71.81)	(52.58)	(71.00)

[†] PPV denotes EI-PPV; PGV denotes EI-PGV; PPV-fwd denotes EI-PPV-fwd; PGV-fwd denotes EI-PGV-fwd; aPGV denotes aug-EI-PGV; aPGV-fwd denotes aug-EI-PGV-fwd

Table 3. Proportions over the 5,000 simulation repetitions that the required training set size of one EI criterion is smaller than another under the setting of $n_0 = 10$ and $n_{sel} = 10$

Dataset	h^2	Pairwise comparisons [†]		
		$n(PPV) < n(PGV)$	$n(PPV) < n(aPGV)$	$n(PGV) < n(aPGV)$
44k	0.2	0.397	0.410	0.443
rice	0.5	0.377	0.392	0.435
	0.8	0.356	0.384	0.426
Tropical	0.2	0.190	0.204	0.211
rice	0.5	0.201	0.212	0.215
	0.8	0.195	0.212	0.222
Wheat	0.2	0.408	0.427	0.473
	0.5	0.383	0.418	0.465
	0.8	0.376	0.415	0.557

[†] $n(PPV)$, $n(PGV)$, and $n(aPGV)$ denote the numbers of individuals required to identify the true best genotype using EI-PPV, EI-PGV, and aug-EI-PGV, respectively

the forward EIs of EI-PPV-fwd, EI-PGV-fwd, and aug-EI-PGV-fwd have relatively unsatisfactory performance, which could be because their smaller prediction variances reduce the exploration capacity and hence increase the size of the required training set. The aug-EI-PGV performs quite similarly to EI-PGV, probably because the prediction standard errors are almost the same for all candidate individuals (the diagonal elements of the genomic relationship matrix are all equal to or close to 1). Thus, the current effective best solution \hat{f}_{Mg}^* and the heuristic multiplier in the aug-EI-PGV have little impact on the identification of the best genotype. We thus made a more detailed comparison between EI-PPV, EI-PGV, and aug-EI-PGV. Let $n(PPV)$, $n(PGV)$, and $n(aPGV)$ denote the numbers of individuals required to identify the true best genotype from one repetition of simulation using EI-PPV, EI-PGV, and aug-EI-PGV, respectively. Then, the proportions over the 5,000 repetitions that $n(PPV) < n(PGV)$, $n(PPV) < n(aPGV)$, and $n(PGV) < n(aPGV)$ under the setting of $n_0 = 10$ and $n_{sel} = 10$ are displayed in Table 3. From the table, both EI-PGV and aug-EI-PGV have better performance than EI-PPV because they often require less individuals to identify the true best genotype. Moreover, aug-EI-PGV slightly outperforms EI-PGV based on both the 44k rice and wheat templates, but shows a significant advantage over EI-PGV based on the tropical rice template. The corresponding scattered plots, displayed in Figures

Table 4. The resulting stopping tolerance (δ^*), the sample size average required to meet the standard set by the stopping tolerance ($\bar{n}_{tr}(\delta^*)$), and the sample size average required to identify the true best genotype (\bar{n}_{tr}) using aug-EI-PGV under the setting of $n_0 = 10$ and $n_{sel} = 10$

Dataset		h^2								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
44k rice	δ^*	10^{-3}	10^{-3}	10^{-3}	10^{-3}	10^{-3}	10^{-3}	10^{-3}	10^{-3}	10^{-3}
	$\bar{n}_{tr}(\delta^*)$	163.3	137.2	118.3	108.7	96.1	90.1	84.5	78.2	73.4
	\bar{n}_{tr}	82.2	76.5	68.4	66.1	61.3	60.2	57.8	54.8	52.6
Tropical rice	δ^*	10^{-2}	10^{-2}	10^{-2}	10^{-2}	10^{-2}	10^{-2}	10^{-2}	10^{-2}	10^{-2}
	$\bar{n}_{tr}(\delta^*)$	87.8	68.1	57.5	54.7	52.4	49.2	48.0	47.4	45.7
	\bar{n}_{tr}	37.5	35.7	35.5	34.3	33.6	33.0	32.0	31.7	31.1
Wheat	δ^*	10^{-3}	10^{-3}	10^{-3}	10^{-3}	10^{-3}	10^{-3}	10^{-3}	10^{-3}	10^{-3}
	$\bar{n}_{tr}(\delta^*)$	147.2	125.8	106.3	97.3	90.6	86.7	78.8	74.7	69.7
	\bar{n}_{tr}	125.3	105.3	91.7	84.7	78.3	75.0	69.5	65.9	61.3

†

S4 to S6 of the Supplementary Materials, reflect the consistent results. In consequence, aug-EI-PGV can be recommended for practical use.

In real data analysis, we propose a feasible stopping rule for the search process according to the fact that the empirical EI values are expected to converge to 0 as the number of selected individuals increases. However, the specification of stopping tolerance δ can significantly affect the final sample size of the required training set. It is therefore desirable to provide some tool to determine the stopping tolerance before performing the real data analysis. For a dataset with known genotype data, following the above simulation steps, we could similarly obtain the smallest sample size of the training set required to meet the standard set by a fixed δ . That is, the empirical EI value with the training set size is less than the fixed δ . Let $\bar{n}_{tr}(\delta)$ and \bar{n}_{tr} denote the sample size averages over the 5,000 simulation repetitions of meeting the standard set by the fixed δ and of identifying the true best genotype, respectively. Then, the largest stopping tolerance δ^* among those under consideration such that $\bar{n}_{tr}(\delta^*) \geq \bar{n}_{tr}$ is recommended for the dataset. Let δ be evaluated over 10^{-k} , for $k = 1, 2, \dots, 6$, and the trait heritability h^2 be set from 0.1 to 0.9 by increment of 0.1, then the resulting δ^* based on the three templates using aug-EI-PGV under the setting of $n_0 = 10$ and $n_{sel} = 10$ is displayed in Table 4. From Table 4, the stopping tolerance δ can be pre-specified as 10^{-3} , 10^{-2} , and 10^{-3} for the datasets of 44k rice, tropical rice, and wheat, respectively. An R code is provided in the Supplementary Materials for users to determine the stopping tolerance for their own datasets.

As mentioned earlier, the genomic relationship matrix \mathbf{K} plays an important role in construction of the GBLUP model, and might also affect the result of searching the best candidate in the current study. Again, following the above simulation steps, we could compare different relationship matrices in identifying the true best genotype for each simulated dataset generated according to the original relationship matrix described in Sect. 2. For the sake of clarity, let this original relationship matrix be denoted as \mathbf{K}_0 . Another three possible formulas for calculating the genomic correlation discussed in Rincint et al. (2012) are summarized in Eq. (A1)–(A3) of the Appendix. Let the relationship matrices corresponding

Table 5. Mean of the training set sizes required to discover the true best genotype from the 5,000 simulated datasets using the different relationship matrices by aug-EI-PGV under the setting of $n_0 = 10$ and $n_{sel} = 10$

Dataset	h^2	Relationship matrices [†]			
		M_0	M_1	M_2	M_3
44k rice	0.2	76.2	79.4	79.8	80.2
	0.5	62.8	67.9	67.2	68.8
	0.8	54.3	58.8	58.9	59.8
Tropical rice	0.2	35.9	54.0	54.2	52.5
	0.5	34.0	50.2	50.0	48.8
	0.8	31.4	46.9	47.3	45.9
Wheat	0.2	106.2	115.4	114.4	114.9
	0.5	77.8	84.9	85.6	85.0
	0.8	66.3	73.2	73.2	72.8

[†] M_0 denotes the original relationship matrix; and M_1 , M_2 , and M_3 are the relationship matrices corresponding to Eqs. (A1), (A2), and (A3), respectively

to Eq. (A1)–(A3) be denoted by K_1 , K_2 , and K_3 , respectively. The mean of the training set sizes required to discover the true best genotype from the 5,000 simulated datasets using the different relationship matrices by aug-EI-PGV under the setting of $n_0 = 10$ and $n_{sel} = 10$ is displayed in Table 5. From the table, it is interesting to find that K_1 , K_2 , and K_3 perform almost equally well in identifying the true best genotype. Notice that this study somewhat favors the original relationship matrix K_0 because the datasets were simulated based on it.

6. CONCLUDING REMARKS

The EI-based Bayesian optimization approach allows a trade-off between exploration and exploitation. From the definition of the improvement function, when there are genotypes better than the best identified at the current stage, the EI-based approach tends to select the genotypes with the large predicted means (exploitation). On the other hand, when there is no genotype better than the best, it tends to select the genotypes with large estimated standard deviations (exploration). The exploration here can enhance the genomic diversity of the training set and thereby increase the chance of discovering the true best genotype in the next search. The problem of interest in this study is actually a batch Bayesian optimization problem, i.e., one in which multiple query points are to be selected simultaneously. Some advanced approaches to such batch Bayesian optimization discussed in Gong et al. (2019) may be adapted for our current study.

To the best of our knowledge, there are no reports in the literature of sequential phenotyping experiments for GP studies in plant breeding. This could be due to the practical constraint that plant breeders are usually able to collect phenotypic data such as grain yield only after a whole growing season of a crop. This can make sequential phenotyping experiment unfeasible if the growing season is too long. From the simulation results, even though a scenario with a small batch size ($n_{sel} = 10$) at a specific heritability can be more efficient than a large one ($n_{sel} = 40$), the difference is not remarkable in practice. This indicates

that breeders can trade a little efficiency to phenotype a large batch of individuals in one stage instead of small batches in several stages. It is known that phenotype is affected by the genotype (G), environment (E), and G×E interaction. In reality, environment can have a significant impact on the performance of individuals selected at each batch. Thus, the proposed multi-stage strategy may need more extensive evaluation to make sure the method works.

Clearly, the application of Bayesian optimization in this study is not to build training and testing sets for GP, but only to identify the best candidate that may be used to serve as a commercial cultivar or to form the parents for the next breeding generation. The EI-based approach could be particularly useful when the phenotyping process is expensive or not very time-consuming, and in addition, the experimenter is able to phenotype individuals sequentially in the order that our proposed approach suggests. For example, for a stress-resistance breeding program, plant breeders need to identify superior accessions with the relevant resistance genes from a germplasm collection. They can then transfer the genes of the identified accessions to a cultivar, thereby enabling it to perform to a desired degree in spite of being subjected to stress (Acquaah 2007). Suppose that the phenotyping cost for a large number of candidate accessions is not affordable to breeders. In such a situation, our proposed sequential phenotyping strategy is a potentially useful choice. Recently, Wu et al. (2019) proposed a GP procedure to assess pumpkin hybrid performance and then identified potential F_1 hybrids with large fruit weight. There were 10,011 possible F_1 hybrids that needed to be evaluated. In this case, our proposed sequential phenotyping strategy might be recommended to the breeder if they are unable to phenotype a large number of the F_1 hybrids during a growing season.

Even though the vector of phenotypic values is assumed to follow a multivariate normal distribution in the GBLUP model (1), it seems not possible to assess the multivariate normality assumption based on such data with no replication on the phenotypic value vector. The multivariate normality assumption is actually imposed on the GBLUP model due to the genomic correlation between individuals, i.e., it is assumed that $\mathbf{g} \sim MVN(\mathbf{0}, \sigma_g^2 \mathbf{K})$. This single-trait GBLUP model is in essence a derivation from the univariate Gaussian regression model (Tempelman 2015), so it might be sufficient to make a univariate normality assessment using the phenotypic values of a single trait. Several graphic approaches or statistical tests can be applied for such normality assumption assessment (Garson 2012). For the proposed Bayesian optimization approach, the search process begins with a training set randomly selected from the candidate population. It would be beneficial to construct a better starting training set from some optimization algorithms (Ou and Liao 2019; Heslot and Feoktistov 2020). Also, from the results of the simulation study, it is reasonable to conclude that the selection efficiency generally improves with increasing trait heritability. The impact of other factors, such as trait genetic architecture, marker density, and linkage disequilibrium, needs further study. Another interesting issue for future study is the investigation of Bayesian optimization approaches with the aim of discovering a set of superior genotypes.

ACKNOWLEDGEMENTS

The authors thank the editor, an associate editor, and three reviewers for providing constructive comments, which helped us to improve both the content and presentation of the manuscript. This research was supported by the Ministry of Science and Technology, Taiwan (Grant No. MOST 108-2118-M-002-002-MY2).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

[Received June 2020. Revised March 2021. Accepted April 2021. Published Online May 2021.]

APPENDIX

THE EI CRITERIA WITH THE MINIMAL GENOTYPIC VALUE

The EI criteria for the situation in which the best genotype is assumed to give the minimal genotypic value are summarized below.

Plug-in Expected Improvement Criteria

$$\text{EI-PPV: } EI(\tilde{y}_{2i}) = (\hat{f}_{mp} - \hat{\mu}_{pi})\Phi(Z_{pi}) + \hat{\sigma}_{pi}\phi(Z_{pi}),$$

where $Z_{pi} = (\hat{f}_{mp} - \hat{\mu}_{pi})/\hat{\sigma}_{pi}$ and \hat{f}_{mp} is the minimal value of \hat{y}_1 ;

$$\text{EI-PGV: } EI(\tilde{g}_{2i}) = (\hat{f}_{mg} - \hat{\mu}_{gi})\Phi(Z_{gi}) + \hat{\sigma}_{gi}\phi(Z_{gi}),$$

where $Z_{gi} = (\hat{f}_{mg} - \hat{\mu}_{gi})/\hat{\sigma}_{gi}$ and \hat{f}_{mg} is the minimal value of \hat{g}_1 .

Forward Expected Improvement Criteria

$$\text{EI-PPV-fwd: } EI(\tilde{y}_{ci}) = (\hat{f}_{mp} - \hat{\mu}_{cpi})\Phi(Z_{cpi}) + \hat{\sigma}_{cpi}\phi(Z_{cpi}),$$

where $Z_{cpi} = (\hat{f}_{mp} - \hat{\mu}_{cpi})/\hat{\sigma}_{cpi}$;

$$\text{EI-PGV-fwd: } EI(\tilde{g}_{ci}) = (\hat{f}_{mg} - \hat{\mu}_{cgi})\Phi(Z_{cgi}) + \hat{\sigma}_{cgi}\phi(Z_{cgi}),$$

where $Z_{cgi} = (\hat{f}_{mg} - \hat{\mu}_{cgi})/\hat{\sigma}_{cgi}$.

Augmented Expected Improvement Criteria

$$\begin{aligned}
 \text{Aug-EI-PGV: } \quad & \text{Aug.EI}(\tilde{g}_{2i}) \\
 & = [(\hat{f}_{mg}^* - \hat{\mu}_{gi})\Phi(Z_{gi}^*) + \hat{\sigma}_{gi}\phi(Z_{gi}^*)] \left(1 - \frac{\hat{\sigma}_e}{\sqrt{\hat{\sigma}_{gi}^2 + \hat{\sigma}_e^2}}\right),
 \end{aligned}$$

$$\begin{aligned}
 \text{Aug-EI-PGV-fwd: } \quad & \text{Aug.EI}(\tilde{g}_{ci}) \\
 & = [(\hat{f}_{mg}^* - \hat{\mu}_{cgi})\Phi(Z_{cgi}^*) + \hat{\sigma}_{cgi}\phi(Z_{cgi}^*)] \left(1 - \frac{\hat{\sigma}_e}{\sqrt{\hat{\sigma}_{cgi}^2 + \hat{\sigma}_e^2}}\right),
 \end{aligned}$$

where $Z_{gi}^* = (\hat{f}_{mg}^* - \hat{\mu}_{gi})/\hat{\sigma}_{gi}$, $Z_{cgi}^* = (\hat{f}_{mg}^* - \hat{\mu}_{cgi})/\hat{\sigma}_{cgi}$, and \hat{f}_{mg}^* is obtained from \hat{g}_1 and minimizes the utility function, i.e.,

$$\hat{f}_{mg}^* = \operatorname{argmin}_{\hat{g}_{1i} \in \hat{g}_1} (\hat{g}_{1i} + \gamma \hat{\sigma}_{1gi}),$$

where $\gamma = 1$.

THE FORMULAS FOR CALCULATING THE GENOMIC CORRELATION

Let (a_1, a_2, \dots, a_p) and (b_1, b_2, \dots, b_p) be the marker score vectors for individuals A and B, respectively. Then, the genomic correlation between these two individuals can be calculated by the following three formulas.

$$\operatorname{corr}(A, B) = \frac{\sum_{i=1}^p (a_i b_i + (1 - a_i)(1 - b_i))}{p}, \tag{A1}$$

where the marker score is coded as 0, 0.5, or 1 for the genotype of M_1M_1 , M_1M_2 , or M_2M_2 . Here, M_1M_1 and M_2M_2 denote the two homogenous genotypes, and M_1M_2 denotes the heterogenous genotype.

$$\operatorname{corr}(A, B) = \frac{\sum_{i=1}^p (1 - 0.5|a_i - b_i|)}{p}, \tag{A2}$$

where the marker score is coded as 0, 1, or 2 for the genotype of M_1M_1 , M_1M_2 , or M_2M_2 .

$$\operatorname{corr}(A, B) = \frac{1}{p} \sum_{i=1}^p \frac{(a_i - q_i)(b_i - q_i)}{q_i(1 - q_i)}, \tag{A3}$$

where the marker score is coded as 0, 0.5, or 1 for the genotype of M_1M_1 , M_1M_2 , or M_2M_2 , and q_i is the proportion among the n individuals that the marker score is coded as 1 at marker i , for $i = 1, 2, \dots, p$.

REFERENCES

- Acquaah G (2007) Principles of plant genetics and breeding. Blackwell Publishing, Malden
- Akdemir D, Sanchez JI (2019) Design of training population for selective phenotyping in genomic prediction. *Sci Rep* 9:1446
- Bull AD (2011) Convergence rates of efficient global optimization algorithms. *J Mach Learn Res* 12:2879–2904
- Crossa J, Campos G, de los Pérez P (2010) Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics* 186:713–724
- Garson GD (2012) Testing statistical assumptions. Statistical Publishing Associate, Asheboro
- Gong C, Peng J, Liu Q (2019) Quantile Stein variational gradient descent for batch Bayesian optimization. In: Proceedings of the 36th international conference on machine learning, PMLR 97, pp 2347–2356
- Habier D, Fernando RL, Dekkers JCM (2007) The impact of genetic relationship information on genomic-assisted breeding values. *Genetics* 177:2389–2397
- Henderson CR (1977) Best linear unbiased prediction of breeding values not in the model for records. *J Dairy Sci* 60:783–787
- Heslot N, Feoktistov V (2020) Optimization of selective phenotyping and population design for genomic prediction. *J Agric Biol Environ Stat* 25:601–616
- Huang D, Allen TT, Notz WI, Zeng N (2006) Global optimization of stochastic black-box systems via sequential kriging meta-models. *J Global Optim* 34:441–466
- Jones DR, Schonlau M, Welch WJ (1998) Efficient global optimization of expensive black-box functions. *J Global Optim* 13:455–492
- Jordan DR, Mace ES, Cruickshank AW, Hunt CH, Henzell RG (2011) Exploring and exploiting genetic variation from unadapted sorghum germplasm in a breeding program. *Crop Sci* 51:1444–1457
- Letham B, Karrer B, Ottoni G, Bakshy E (2019) Constrained Bayesian optimization with noisy experiments. *Bayesian Anal* 14:495–519
- Lin TY, Liao CT, Iyer HK (2008) Tolerance intervals for unbalanced one-way random effects models with covariates and heterogeneous variances. *J Agric Biol Environ Stat* 13:221–241
- McCouch S, Baute GJ, Bradeen J, Bramel P, Bretting PK, Buckler E et al (2013) Agriculture: feeding the future. *Nature* 499:23–24
- Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829
- Ou JH, Liao CT (2019) Training set determination for genomic selection. *Theor Appl Genet* 132:2781–2792
- Perez P, de Los Campos G (2014) Genome-wide regression and prediction with the BGLR statistical package. *Genetics* 198:483–495
- Picheny V, Ginsbourger D, Richet Y, Caplin G (2013) Quantile-based optimization of noise computer experiments with tunable precision. *Technometrics* 55:2–13
- Reif JC, Zhang P, Dreisigacker S, Warburton ML, van Ginkel M et al (2005) Wheat genetic diversity trends during domestication and breeding. *Theor Appl Genet* 110:859–864
- Rincent R, Laloe D, Nicolas S, Altmann T, Brunel D et al (2012) Maximizing the reliability of genomic selection by optimizing the calibration set of reference individuals: comparison of methods in two diverse groups of maize inbreds (*Zea mays* L.). *Genetics* 192:715–728
- Shahriari B, Swersky K, Wang Z, Adams RP, De Freitas N (2016) Taking the human out of the loop: a review of Bayesian optimization. *Proc IEEE* 104:148–175
- Spindel J, Begum H, Akdemir D, Virk P, Collard B et al (2015) Genomic selection and association mapping in rice (*Oryza sativa*): Effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. *PLOS Genet* 11:e1005350
- Tanaka R, Iwata H (2018) Bayesian optimization for genomic selection: a method for discovering the best genotype among a large number of candidates. *Theor Appl Genet* 131:93–105

- Tempelman RJ (2015) Statistical and computational challenges in whole genome prediction and genome-wide association analyses for plant and animal breeding. *J Agric Biol Environ Stat* 20:442–466
- Tester M, Langridge P (2010) Breeding technologies to increase crop production in a changing world. *Science* 327:818–822
- Tibshirani R (1996) Regression shrinkage and selection via the LASSO. *J Royal Stat Soc Ser B Stati Methodol* 58:267–288
- Vazquez E, Villemonteix J, Sidorkiewicz M, Walter E (2008) Global optimization based on noisy evaluations: an empirical study of two statistical approaches. *J Global Optim* 43:373–389
- Wu PY, Tung CW, Lee CY, Liao CT (2019) Genomic prediction of pumpkin hybrid performance. *Plant Genome* 12:180082
- Xavier A, Muir WM, Craig B, Rainey KM (2016) Walking through the statistical black boxes of plant breeding. *Theor Appl Genet* 129:1933–1949
- Zhao K, Tung CW, Eizenga GC, Wright MH, Ali ML et al (2011) Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nat Commun* 2:467

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.