forests

MDPI

Review

Prediction of Genetic Gains from Selection in Tree Breeding

Zi-Han He [1,2], Yu Xiao [1,2], Yan-Wen Lv [1,2], Francis C. Yeh [3], Xi Wang [1,2,*] and Xin-Sheng Hu [1,2,*]

[1] College of Forestry and Landscape Architecture, South China Agricultural University, Guangzhou 510642, China
[2] Guangdong Key Laboratory for Innovative Development and Utilization of Forest Plant Germplasm, Guangzhou 510642, China
[3] Department of Renewable Resources, University of Alberta, 751 General Service Building, Edmonton, AB T6G 2H1, Canada
* Correspondence: xwangaga@gmail.com (X.W.); xinsheng@scau.edu.cn (X.-S.H.)

Abstract: The prediction of genetic gain from artificial selection in a trait is important in plant and animal breeding. Lush's classical breeder's equation (BE) is widely used for this purpose, although it is also applied to predicting evolution under natural selection. The current application of high throughput sequencing techniques potentially allows breeders at the individual gene level to capture both additive and non-additive genetic effects. Here, we provide a comprehensive evaluation of predicting genetic gains from the selection at multiple hierarchical levels of population structure (provenances, families within provenances, and individuals within families within provenances). We discuss the processes that could influence the power of prediction under the classical BE, including genetic drift, natural selection, and gene flow. We extend the classical BE to molecular breeding methods for improving the prediction of genetic gains; they include the conventional breeding approach, marker-assistant selection (MAS), genome-wide association study (GWAS), and genomic selection (GS). Lastly, we discuss the genetic gains from the selection using multi-omics traits, including gene expression and epigenetic traits. Our overall synthesis should contribute to a better understanding of predicting genetic gains from the artificial selection under classical and molecular breeding.

Keywords: breeder's equation; genetic gain; artificial selection; molecular breeding; marker-assistant selection

check for updates

Citation: He, Z.-H.; Xiao, Y.; Lv, Y.-W.; Yeh, F.C.; Wang, X.; Hu, X.-S. Prediction of Genetic Gains from Selection in Tree Breeding. *Forests* **2023**, *14*, 520. https://doi.org/10.3390/f14030520

Academic Editor: David Lee

Received: 1 February 2023
Revised: 27 February 2023
Accepted: 5 March 2023
Published: 7 March 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Both genetic gains from artificial selection in breeding populations and the genetic responses to selection in natural populations have long been studied since the publication of Darwin's [1] *On the Origin of Species*. How best to predict genetic gain from selection in a trait remains crucial in breeding. Lush [2] first proposed the classical breeder's equation (BE) to predict the genetic response *(R)* as the product of selection differential (*S*) and the narrow-sense heritability (h_N^2), i.e., $R = h_N^2 S$. Lush's BE was initially applied to predict the genetic response to artificial selection in animal breeding. This BE is now frequently expressed as $\Delta G = h_N^2 S$, where G is the mean of a trait in the population and ΔG is the change in the mean over one complete cycle of artificial selection, which is commonly termed as genetic gain in plant and animal breeding. The selection differential (S) is the difference between the mean of a trait of the selected parents and the mean of the whole parental population.

It is well known that artificial selection is a method of selecting individuals or populations according to the objectives that meet human demands (e.g., yields and quality), with the aim of producing genetically improved populations. It differs from natural selection in objective, the intensity of selection, and the rate of genetically changing populations. The target traits for genetic improvement often refer to those of economical values, although they infrequently refer to the fitness. The selection intensity (or selection differential) is

artificially controlled according to the genetic variation in a candidate population. Usually, a strong strength of selection is set to rapidly increase the frequencies of target genes. Analogous to natural selection, artificial selection may be conducted using different ways to change gene frequencies, such as directional, stable, or disruptive selection, depending on the breeding objectives. Multiple cycles of artificial selection are frequently needed to fulfil the breeding objectives.

Most traits studied in plant and animal breeding are quantitative in nature and controlled by many genes. The beauty of this BE is that all the complexities of multi-locus inheritance of a quantitative trait are condensed into one parameter, h_N^2. Griffing [3,4] extended this BE under the assumptions of additive and non-additive gene effects. However, these extensions were infrequently used in plant and animal breeding and the prediction of genetic gain remains mostly centered on Lush's BE. When the length of the selection cycle t (the generation time) is considered, BE is expressed as $\Delta G = h_N^2 S/t$ [5]. When S is expressed in terms of the standard deviation of a quantitative trait (σ_P), $S = i\sigma_P$, where i is the selection intensity, the BE is expressed as:

$$\Delta G = i h_N \sigma_A / t \quad (1)$$

where σ_A is the standard deviation of additive genetic effects. The assumptions underlying Equation (1) are (i) that the quantitative trait is controlled by many genes each with small effects and (ii) that both the quantitative trait and breeding value follow a normal distribution. When the quantitative trait is a fitness trait, Lush's BE under $t = 1$ has a similar inference as Fisher's fundamental theorem of natural selection [6]. The BE or Fisher's fundamental theorem of natural selection may be considered a special case of the Price or the Robertson–Price equation [7–12].

When multiple quantitative traits are considered simultaneously [13], the genetic gain ($\Delta\overline{\boldsymbol{G}}$) from a selection is expressed as

$$\Delta\overline{\boldsymbol{G}} = \boldsymbol{V_G V_P^{-1} S} \quad (2)$$

where $\overline{\boldsymbol{G}}$ is the vector of phenotypic means for multiple quantitative traits, $\boldsymbol{V_G}$ and $\boldsymbol{V_P}$ are the genetic and phenotypic variance–covariance matrices, respectively, and $\boldsymbol{S}$ is the vector of selection differentials of multiple traits. Equation (2) is often used in evolutionary biology to predict the evolution of multiple traits [14] but is infrequently used in plant and animal breeding. One possible reason is that a selection index is often constructed in breeding where multiple traits are combined by their relative weights. Selection is based on the values of individual selection indices [5].

In plant and animal breeding, selection often commences with populations of hierarchical structures. Tree breeding of a given species typically begins with natural populations across a geographical range. It is necessary to depict the genetic variations among provenances, among families within provenances, among individuals within populations within provenances, and the possible asexual lineages among them all. These characteristics of a species would allow breeders to exploit different components of the genetic variance at different levels of selection [5,15]. In addition, the means to predict genetic gains are modified accordingly to account for the different hierarchical levels of selection. Note that cases analogous to tree breeding in crop and animal breeding are when selection starts with artificial populations or lines derived from diverse mating designs. In these situations, the candidate populations may not have the typical hierarchical structure mentioned above and the genetic gain from an artificial selection is estimated using the BE with the heritability and selection differential at the correspondent level.

While the classical BE has been extensively deployed in plant and animal breeding, there are few reports in the literature that systematically reviewed the BE in terms of theory and methodology or assessed its limitation and extension in predicting genetic gains. Here, we begin with describing the genetic gains from a selection at different hierarchical levels of population structure and discuss factors that could bias the genetic gain predictions using

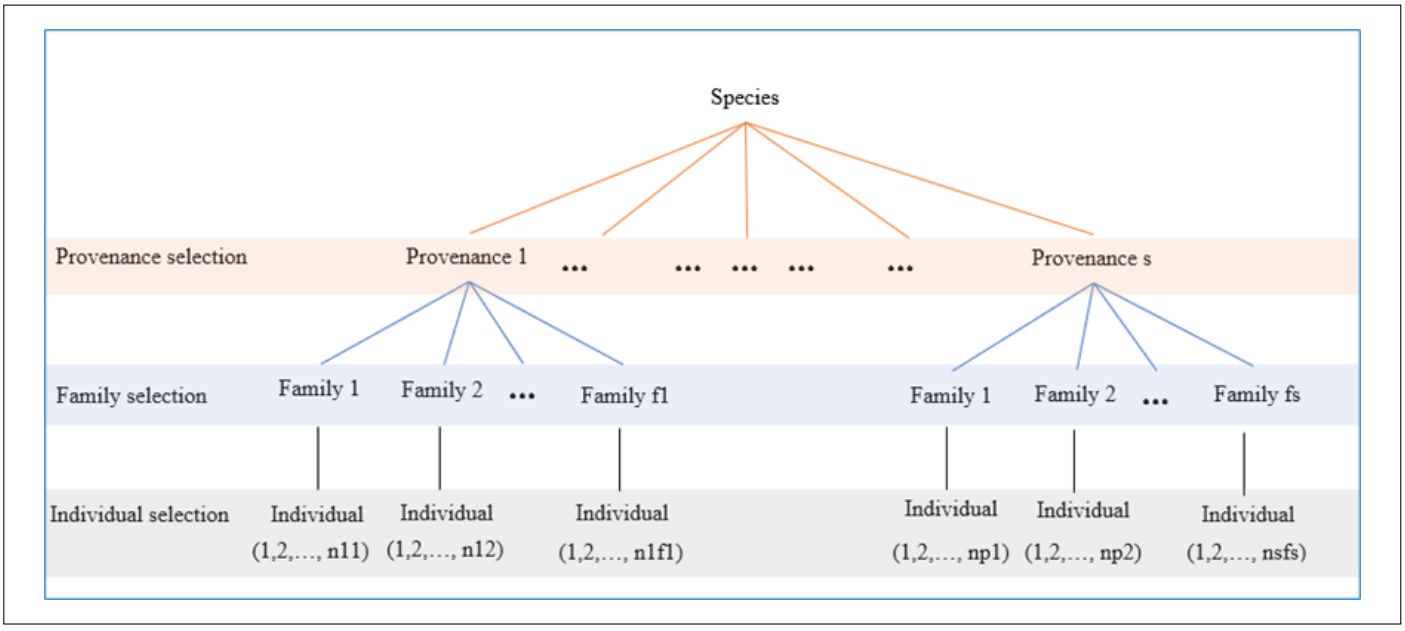
the classical BE. We then consider some molecular breeding approaches to advance the BE prediction of genetic gains. They include the conventional breeding approach, molecular marker-assisted selection (MAS), genome-wide association analysis (GWAS), and genomic selection (GS). Lastly, we discuss the genetic gains from a selection using multi-omics traits, including gene expression and epigenetic traits, which are potentially important tools in future molecular breeding. Our overall synthesis would aid in a better understanding of predicting genetic gains from artificial selection under classical and molecular breeding.

2. Genetic Gains from Selection at Different Levels

2.1. Selection with Populations of Hierarchical Structure

In this subsection, we illustrate the prediction of genetic gains from a selection at different hierarchical levels of population structure. A tree species typically occupies a certain geographical range in its spatial distribution. An appropriate breeding strategy considers its adaptation zones and reproductive system, which helps breeding design, such as the breeding populations created by hybridization or derived from open-pollinated progeny. The total genetic variation can be decomposed into the components at the levels of provenances, families within provenances (full- or half-sib families), and individuals within families within provenances (Box 1). To construct the breeding populations, a selection at different hierarchical levels of population structure produces different extents of genetic gains, which are estimated using different BEs. Box 2 shows a provenance trial at a single site with a random block design. A phenotypic observation is decomposed into different effects in a linear model. Different BEs are required to predict the genetic gains at different levels of selection. This breeding plan is applicable to other breeding designs for the prediction of different genetic gains from a selection.

Box 1. A framework of selection at different levels (provenance, family, and individual).



Selection at the provenance level requires an estimate of heritability at the provenance level to predict the genetic gain. Here, the heritability refers to the ratio of the genetic variance among provenances to the phenotypic variance among provenances (Box 2). The numerator of heritability is the genetic variance among provenance means. Heritability may also be viewed as a regression of the provenance trait breeding values on the provenance phenotypic values [15]. To gain insights into the genetic variance among provenances, assume that a quantitative trait is controlled by multiple QTLs (quantitative trait loci) of

additive and dominant effects and that each locus is independent of the other. The trait mean M_i in the ith provenance can be expressed as:

$$M_i = \sum_j a_{ij}(p_{ij} - q_{ij}) + 2\sum_j d_{ij}p_{ij}q_{ij}(1 - F_{IS}) \quad (3)$$

where a_{ij} and d_{ij} are the additive and dominant effects at the jth locus, respectively; p_{ij} and q_{ij} are the two allele frequencies at locus j ($p_{ij} + q_{ij} = 1$); F_{IS} is the inbreeding coefficient in the ith provenance [5]. The genetic variance among provenance means consists of both additive and dominant variances. If only additive effects are considered in the trait and each allele has stable additive effects ($V(a_{ij}) = 0$), the gene frequency at a locus, denoted by p, follows a beta distribution in each provenance [16]:

$$\phi(p) = \frac{\Gamma(4N_e m)}{\Gamma(4N_e mQ)\Gamma(4N_e m(1-Q))} p^{4N_e mQ-1}(1-p)^{4N_e m(1-Q)-1} \quad (4)$$

where N_e is the effective population size of each provenance, Q is the average gene frequency among all provenances or the initial gene frequency in the reference provenance before subdivision, and m is the per-generation migration rate among provenances. Thus, the genetic variance among provenance means, denoted by σ_S^2, is:

$$\begin{aligned} \sigma_S^2 &= V\left(\textstyle\sum_j a_{ij}(p_{ij} - q_{ij})\right) \\ &= \tfrac{1}{1+4N_e m}\textstyle\sum_j 4a_j^2 Q_j(1-Q_j) \\ &= 2F_{ST}\sigma_A^2 \end{aligned} \quad (5)$$

where F_{ST} is the provenance differentiation coefficient [16,17]. Population genetic structure shapes the genetic gains from a selection at the provenance level. Gene flow reduces selection efficacy at the provenance level. When QTLs with both additive and non-additive (dominance and epistatic) effects are in a trait, the genetic gains from a selection at the provenance level could be enhanced. This is because the non-additive effects also contribute to the genetic variance among provenance means.

Selection at the family level requires the estimation of genetic variance among family means. A practical operation may be conducted using entire families across all provenances or using families within individual provenances. The difference is that the former has more families for selection and provides us with a chance to set higher selection intensity compared with the latter. The genetic gains from selection vary with the type of family used. For open-pollinated families (half-sibs), such as seed samples collected from individual mother trees, the genetic variance is:

$$\sigma_{HS}^2 = \frac{1}{4}\sigma_A^2 + \frac{1}{16}\sigma_{AA}^2 + \frac{1}{64}\sigma_{AAA}^2 + \cdots \quad (6)$$

Thus, the numerator of the half-sib heritability includes partial additive and additive epistatic effects from parental population. For instance, with the artificial populations derived from diallel crosses, selection based on the general combining ability (GCA) of parents utilizes the half-sib families and the numerator of the half-sib heritability includes both additive and additive-by-additive variances.

When selection is based on full-sib families, such as the progeny populations derived from multiple artificial crosses, the numerator of the heritability at the full-sib family level includes variances of additive, dominance, additive-by-additive, and dominance-by additive effects:

$$\sigma_{FS}^2 = \frac{1}{2}\sigma_A^2 + \frac{1}{4}\sigma_D^2 + \frac{1}{4}\sigma_{AA}^2 + \frac{1}{8}\sigma_{AD}^2 \ldots \quad (7)$$

For instance, selection based on the specific combining ability (SCA) of parents makes use of full-sib genetic variances.

Box 2. A provenance trial at a single site and genetic gains from a selection at different levels.

Consider a provenance trial at a site. The linear model for any observation is:

$$x_{ikjl} = \mu + S_i + F_{j(i)} + B_k + SB_{ik} + FB_{j(i)k} + e_{ijkl}$$

where x_{ijkl} is the observed value of the lth individual in the jth family from the ith provenance in the kth block, S_i is the ith provenance effect ($i = 1, \ldots, s$), $F_{j(i)}$ is the jth family effect within the ith provenance ($j = 1, \ldots, f$), B_k is the kth block effect ($k = 1, \ldots, r$), SB_{ik} is the interaction effect between the ith provenance and the kth block, $FB_{j(i)k}$ is the interaction effect between the jth family within the ith provenance and the kth block, and e_{ijkl} is the residual error ($l = 1, \ldots, n$). Under the random model, the following relationships hold among effects: $\sum_i S_i = 0$, $\sum_j F_{j(i)} = 0$, $\sum_k B_k = 0$, $\sum_i SB_{ik} = \sum_k SB_{ik} = 0$, and $\sum_j FB_{j(i)k} = \sum_k FB_{j(i)k} = 0$. The analysis of variance is given below:

Source of Variation	Degree of Freedom	Expected Mean Square
Block	$r-1$	$\sigma_e^2 + nf\sigma_{SB}^2 + n\sigma_{FB}^2 + nsf\sigma_B^2$
Provenance	$s-1$	$\sigma_e^2 + nf\sigma_{SB}^2 + nrf\sigma_S^2$
Family within provenance	$s(f-1)$	$\sigma_e^2 + n\sigma_{FB}^2 + nr\sigma_F^2$
Provenance × Block	$(s-1)(r-1)$	$\sigma_e^2 + nf\sigma_{SB}^2$
Family × Block	$s(f-1)(r-1)$	$\sigma_e^2 + n\sigma_{FB}^2$
Error	$srf(n-1)$	σ_e^2

Genetic gains from selection at different levels:

Level of Selection	Selection Differential S	Genetic Gains
Provenance	Difference in trait mean between the selected provenances and the whole provenances investigated	$\Delta G = S\frac{\sigma_S^2}{\sigma_S^2+\sigma_{SB}^2/r+\sigma_E^2/nrf}$
Family	Difference in trait mean between the selected families and the whole families investigated	$\Delta G = S\frac{\sigma_F^2}{\sigma_F^2+\sigma_{FB}^2/r+\sigma_E^2/nr}$
Individual	Difference in trait mean between the selected individuals and the whole individuals investigated	$\Delta G = S\frac{4\sigma_F^2}{\sigma_S^2+\sigma_F^2+\sigma_B^2+\sigma_{SB}^2+\sigma_{FB}^2+\sigma_E^2}$ (half-sib family)

Selection at the individual level is flexible, irrespective of the population structure. It can be conducted within individual provenances or within individual families. The numerator of the heritability at the individual level includes all additive genetic variance (Box 2). Genetic gains from individual selection are different from those of selection at higher hierarchical levels of population structure. One specific situation is clonal selection, where heritability may be replaced with repeatability in predicting genetic gains [5].

When selection is conducted at the individual level but the genetic variance is derived from both within and between families, Falconer and Mackay [5] proposed an index to predict the comprehensive genetic gains. Genetic gain from each component is combined with its relative weight to construct the selection index. The total genetic gain from index selection is amplified, although the numerator of the heritability estimate consists of only additive genetic variance. This type of selection is more frequently applied to animal breeding than to plant breeding.

A more complex scheme is that both selection and genetic diversity are considered in breeding programs. For instance, when a balance between genetic gain and genetic diversity is proposed, then breeders must consider both the number of families [18,19] or the kinship relationship among individuals and genetic diversity in breeding populations during selection [20,21]. In this situation, genetic gains are predicted under different conditions. Nevertheless, only additive genetic variance in the numerator of heritability is emphasized in their simulative selection schemes.

For multiple cycles of selection, both selection differential and heritability are often altered between cycles and the genetic gains are separately estimated. The overall genetic gain is the sum of gains at each cycle of selection.

2.2. Factors Causing Biased Prediction of Genetic Gains

The classical BE only exploits additive genetic variance (σ_A^2). It is effective to predict genetic gains from individual selection for the traits with high narrow-sense heritability but inadequate for those traits with low heritability [5,15]. The causal factors could arise from the mode of gene action in quantitative traits, including the dominant effects within loci, epistasis between loci, and linkage disequilibrium (LD) between loci [22,23]. Here, we also discuss in theory a few processes that influence the prediction of genetic gains.

One dominating process is genetic drift, since breeding populations are established with finite sizes. Genetic drift is inevitably involved in shaping the genetic variance of a quantitative trait. These influences can be produced in two ways. One way is to change the allele frequencies at QTLs. From the population genetic theory [24], the additive genetic variance is given by:

$$\sigma_A^2(t) = \sigma_A^2(0)\left(1 - \frac{1}{2N_e}\right)^t \approx \sigma_A^2(0)e^{-t/2N_e} \tag{8}$$

$\sigma_A^2(t)$ decreases at a rate of $\frac{1}{2N_e}$, relative to the initial additive genetic variance $\sigma_A^2(0)$. As the gene frequency changes towards 1 (fixation) or 0 (loss) during the drift process, the additive genetic variance tends to approach zero. Similarly, the total genetic variance decreases,

$$\sigma_G^2 = \frac{1}{2N_e}\sigma_A^2 + \left(1 - \frac{1}{2N_e}\right)\sigma_D^2 + \left(\frac{1}{2N_e}\right)^2\sigma_{AA}^2 + \frac{1}{2N_e}\left(1 - \frac{1}{2N_e}\right)\sigma_{AD}^2 + \left(1 - \frac{1}{2N_e}\right)^2\sigma_{DD}^2 + \cdots \tag{9}$$

All components of the genetic variances converge to zero as the genetic drift proceeds [25]. In the presence of linkage disequilibria among loci, the joint effects of recombination and genetic drift further reduce the variance of epistatic effects.

The second way to influence the prediction of genetic gain is through joint effects of genetic drift and selection, which enhance the fixation of deleterious alleles. Kimura [26] showed that the fixation probability of a mutant allele with selection coefficient s:

$$u\left(\frac{1}{2N}\right) = \frac{1 - e^{-2N_e|s|/N}}{1 - e^{-2N_e|s|}} \tag{10}$$

where N is the actual population size. There is a certain probability of fixation for a deleterious allele ($s < 0$), while the fixation probability of an advantageous mutant ($s > 0$) is not equal to 1 in small populations. In relation to a mutant allele at a QTL, the selection coefficient (s_i) may be expressed as the product of selection intensity i and allelic additive effect (a): $s_i = ai/\sigma_P$ [3,27]. The fixation probability of a mutant allele at a QTL can be obtained by replacing s in Equation (10) with s_i. Thus, genetic drift enhances the maintenance of deleterious alleles and reduces the phenotypic values, which reduces the probability of generating superior genotypic combinations at multiple loci. The above two-way effects occur in multiple rounds of selection in breeding and naturally influence genetic gains from an individual selection.

The second process is concerned with the limit to artificial selection where alleles are fixed at most QTLs. This limit is positively correlated with the product of effective population size and selection coefficient ($N_e s$) or the product of N_e and the selection intensity ($N_e i/\sigma_P$) [28]. When favorable alleles are fixed, the genetic gain is given by:

$$\Delta G = 2N_e i h_N^2 \tag{11}$$

The ultimate genetic gain is proportional to the effective population size. Similarly, the ultimate genetic gain due to mutations is proportional to the effective population size [29]. All these joint effects of selection and genetic drift or mutation influence the genetic gain in the long-term selection.

The third process is an input of alien pollen to breeding populations, which could reduce or increase the genetic variance depending on the difference between the gene frequencies in migrating pollen and in breeding populations [30]. This will doubtlessly influence heritability at the population level and the genetic gain from provenance selection.

In addition, the mating system affects the prediction of genetic gain. When selfing or partial inbreeding occurs in breeding populations, non-additive genetic variances can be partially transmitted across generations, because the same genotypic combinations at multiple loci can occur in the next generation with a certain probability. In addition, the practical estimation of genetic parameters (e.g., heritability) is presumably under random mating, which is violated in breeding populations with partial selfing. The inclusion of non-additive effects owing to selfing could improve the prediction of genetic gains.

3. Improving Prediction of Genetic Gains

3.1. Conventional Breeding Approach

The classical BE implicates that genetic gain from selection can be increased by four ways: (i) increasing heritability through reducing the environmental variation and the genetic by environment interaction; (ii) increasing selection intensity; (iii) increasing genetic variation in breeding populations; (iv) shortening the length of breeding cycle [31]. Xu et al. [32] suggested that genetic gains from selection can be improved through improving crop and field management and optimizing socio-economic management, but such effects are indirect and unstable. There is no case report in crop breeding showing that genetic gain has been continuously and consistently improved through fine field management. In this subsection, we discuss the first three ways outlined above to increase the genetic gain from a selection. These three ways are based on the conventional tree breeding theory [15]. In the subsequent three subsections, we discuss the fourth way outlined above for molecular breeding.

To increase heritability, for a given breeding population, one can improve the experimental design that reduces environmental errors as well as the interaction between genotype and environment (G × E). Previous provenance trials showed that many tree species had significant interactions between genotype and environment [33]. Samples from the same provenance could have different breeding values under different environments. The additive or non-additive genetic effects may vary at different sites. Even if gene frequencies are insignificantly different between samples at two sites, their genetic variances may differ due to unstable additive effects; the variance of additive effects in Equation (3) is not zero ($V(a_{ij}) \neq 0$). Through controlling the environmental conditions at an experimental site, the denominator of the heritability estimate reduces, which indirectly increases heritability and improves the genetic gains. Note that practical genetic gains could be different when the breeding values of a genotype are different in breeding and productive populations [34]. Allwright and Taylor [35] confirmed that controlling environmental variables was effective in improving realized heritability and hence in increasing genetic gains.

The second way to improve genetic gain is easy to understand because ΔG is proportional to the selection differential (S). This can be obtained by increasing selection intensity i and expanding candidate populations. Endelman et al. [36] showed that genetic gains from selection was increased when a constant number of individuals was selected and the candidate population size was expanded.

The third way is to increase polymorphic loci since the additive genetic variance increases with the heterozygosity, i.e., $\sigma_A^2 = \sum_i 2p_i(1 - p_i)a^2$. Large segregating populations aid in increasing genetic variation, which then improves genetic gain.

3.2. Marker-Assisted Selection

MAS offers a method to advance the prediction of genetic gain from selection. The genetic basis of MAS relies on the assumption of linkage disequilibrium (LD) between

QTLs and markers in the breeding populations. Lande and Thompson [37] proposed a selection index by combining phenotypic traits with molecular marker scores:

$$I = b_z z + b_m m \tag{12}$$

here, the selection index I is assumed to follow a normal distribution, z is the phenotypic value of a single trait, m is the molecular marker score and equals the sum of the estimated additive effects of all marker genotypes, and b_z and b_m are the weights of the two components. The genetic gain from index selection is calculated by:

$$\Delta G = i r_{IA} \sigma_A \tag{13}$$

where r_{IA} is the correlation coefficient between the index and breeding value of the trait. The efficacy of MAS is determined by the degree of the correlation coefficient r_{IA}.

Lande and Thompson [37] used the relative efficiency (r) to describe MAS, which is termed as the rate of the genetic gain from MAS to that from phenotypic selection alone:

$$r = \sqrt{\frac{\rho}{h_N^2} + \frac{(1-\rho)^2}{1-\rho h_N^2}} \tag{14}$$

where ρ is the proportion of additive genetic variance explained by molecular markers. As ρ increases, the relative selection efficiency, r, increases. MAS is more effective for traits with low heritability than for traits with high heritability, which overcomes the weakness of the classical BE. The disadvantage of MAS is that the relative selection efficiency decreases with the inclusion of more information from relatives into the index. The use of full-sib information has a lower selection efficiency than the use of half-sib information. Moreover, the efficiency of selection decreases when the number of families increases.

Liu et al. [38] showed that the inclusion of epistatic effects into the selection index could further improve the efficiency within the short-term selection. Hu [39] provided a general framework for the selection index where epistatic effects between two loci were included. When the proportion of non-additive variance explained by the markers increases, a greater efficiency of selection is expected. However, LD between loci reduces the relative selection efficiency and should be avoided when choosing molecular markers.

MAS is effective in selection because the markers are not affected by the environmental factors or by the interaction between genotype and environment. This enhances early selection and shortens the period of selection [40], which enhances the genetic improvement of complex traits. Currently, most MAS cases are reported in crop and animal breeding [41–43]. MAS in tree species remains in the phase of searching for the markers associated with quantitative traits. For example, Feng et al. [44] applied MAS to a hybrid population of *Camellia oleifera* and realized the genetic improvement for oil production and some fruit characters. In studies with mulberry (*Morus* spp.), some markers were detected in association with important traits [45–47], such as leaf yield and biochemical traits. In general, MAS in tree breeding lags behind those of agricultural crop and animal species.

Advances in practical studies with MAS mainly confine to those traits controlled by a few major QTLs [48]. For those complex traits that are controlled by many QTLs each with small effects, such as wood volume and disease traits, it is often difficult to capture those markers associated with these QTLs. Interactions among QTLs further limit the power of detecting relevant markers. In breeding populations or natural populations, the extent of LD between loci is often low [49]. This limitation suggests that substantial genetic gains through conventional MAS are difficult to achieve for complex traits. To increase the efficiency of MAS, a genome-wide analysis is needed to search for the markers associated with QTLs of relatively small effects.

3.3. Genome-Wide Association Study

GWAS refers to the analysis of using genome-wide single nucleotide polymorphism (SNP) markers to search for those SNPs that are associated with the target traits. A certain sample size is required in the analysis. The genetic basis of GWAS remains as the presence of LD between the SNP and the QTL of a target trait. The preliminary objectives of GWAS are to map markers associated with the target traits and further to annotate the marker functions. However, GWAS also provides an approach to advancing the prediction of genetic gains in breeding. A selection index can be constructed using both phenotypes and all the markers that are screened from the genome-wide association analysis. It has an advantage over conventional MAS and is suitable for genetically improving complex traits. This is because the use of genome-wide SNPs increases the probability of finding SNPs associated with QTLs of small effects.

The genetic gain from index selection (individual) is estimated by:

$$\Delta G = i r_{GWAS,A} \sigma_A \quad (15)$$

The extent of improving genetic gain is determined by the correlation between breeding and index values, $r_{GWAS,A}$ This correlation coefficient increases when more related markers are incorporated into the selection index. A large sample size increases the probability to screen those SNPs associated with some QTLs of small additive effects.

When SNPs associated with non-additive effects (dominant or epistatic) are screened, the marker scores may also include non-additive effects. The selection index is re-constructed by maximizing the correlation between the genotype values (additive and some non-additive effects) and the marker scores [39]. The key point is that a higher correlation coefficient, $r_{GWAS,\,(A+NA)}$, could further increase the genetic gain, compared with the case where only additive effects are scored with the markers.

Recent advances in GWAS mainly confine to screening markers associated with target traits and to identifying the functions of markers [48]. For example, GWAS was used to search for the markers associated with wood properties and growth traits in *Populus deltoides*, *Populus trichocarpa*, and *Eucalyptus cladocalyx* [50–54] or with the fruit quality and yield in *Citrus sinensis* [55,56] and *Malus domestica* [57]. The effectiveness of applying GWAS to MAS is closely related to the number of related markers, the genetic model of marker effects, the type of breeding populations (e.g., provenances or multiple families/lines), and the sample size.

Another way to use GWAS in breeding is to combine GWAS results with the network of gene expression to further dissect the genetic variation of target traits. This will likely aid in capturing a greater proportion of heritability, although few reports show that multiple levels of interactions may transfer across generations [58]. For instance, Lamara [59] combined GWAS with gene co-expression networks to uncover 2652 candidate SNPs for wood quality of *Picea glauca* and constructed an expression network of two known transcription factors (MYB and NAC). They revealed the complex interactions and pleiotropic effects among the transcriptional regulator gene (PgNAC-8), wood hardness, and microfibril angle, as well as the complex genetic control and gene expression network of wood traits. Similar studies were reported in *Populus deltoides* [60] and *Eucalyptus cladocalyx* [54]. From the deployment perspective, all the genes or SNP markers expressed in the interaction network can be used to construct a selection index. The information of networks among genes is likely transmitted across generations, as implied from the evolution of the protein interaction network [61]. This approach helps to design complex molecular breeding for improving the genetic gain from an individual selection.

3.4. Genomic Selection

GS uses the genome-wide SNPs to estimate individual breeding values. The prediction model is derived from a training population but tested in a validation population. Selection is then based on the individual genomic estimates of breeding values (GEBVs).

An early-stage selection can be conducted without waiting for the mature stage of breeding populations. Up until now, many reports are available in the literature about GS principles, methods, examples, and reviews in plant and animal breeding [48,62–65]. For instance, the reports on tree breeding include species in genera *Populus*, *Eucalyptus*, *Elaeis*, *Hevea*, *Pinus*, and *Picea* [48,64,66]. The genetic gains from GS could be 1.5~3 times higher than those from phenotypic selection alone [67]. Here, we briefly review the theory of how genetic gains from GS could be enhanced, which has not been explicitly indicated in the literature.

As mentioned above, one approach of GS in improving the prediction of genetic gain is to reduce the cycle of selection time, t, in Equation (1). In theory, GS is more effective than the conventional early-selection approach, which is based on the correlation between early and late phenotypic values [48].

Another approach of GS in improving genetic gain is to increase the correlation coefficient between the GEBV and the breeding values, $r_{GEBV,A}$. The genetic gain is expressed as:

$$\Delta G = i r_{GEBV,A} \sigma_A \tag{16}$$

Currently, two algorithms are available to calculate this correlation coefficient. One is to accurately estimate the individual kinship $\boldsymbol{V}_g$ matrix using a mixed linear model. For instance, there are n phenotypic observations represented by the $\boldsymbol{y}_{n\times 1}$ vector, $p\boldsymbol{\beta}_{p\times 1}$ vector, a design matrix of $\boldsymbol{X}_{\mathrm{n}\times\mathrm{p}}$ for fixed effects that records the occurrence of each factor and level, and q random additive effects, which is specific to a genetic mating design matrix of $\mathbf{Z}_{n\times q}$ and represented by the $\boldsymbol{u}_{q\times 1}$ vector. The mixed linear model is expressed by:

$$\mathbf{y}_{n\times 1} = \boldsymbol{X}_{n\times p}\boldsymbol{\beta}_{p\times 1} + \mathbf{Z}_{n\times q}\boldsymbol{u}_{q\times 1} + \boldsymbol{e}_{n\times 1} \tag{17}$$

Vector $\boldsymbol{e}$ is the random errors, with $E(\boldsymbol{e}) = 0$. Expectations of other variables are $E(\boldsymbol{u}) = 0$ and $E(\boldsymbol{y}) = \boldsymbol{X\beta}$ and the variances are $\mathrm{V}(\boldsymbol{u}) = \boldsymbol{V}_g$ and $\mathrm{V}(\boldsymbol{e}) = \boldsymbol{R} = \sigma_E^2\boldsymbol{I}$, in which $\boldsymbol{I}$ is the identity matrix. All random effects follow a normal distribution: $\boldsymbol{u} \sim N(0, \boldsymbol{V}_g)$, $\boldsymbol{e} \sim N(0, \boldsymbol{R})$, and $\mathbf{y} \sim N(\boldsymbol{X\beta}, \boldsymbol{V})$, where $\boldsymbol{V} = \boldsymbol{Z}\boldsymbol{V}_g\boldsymbol{Z}' + \boldsymbol{R}$. The $\boldsymbol{V}_g$ matrix is estimated using genome-wide markers [68]:

$$\boldsymbol{V}_g = \frac{\boldsymbol{M}'\boldsymbol{M}}{2\sum_{i=1}^{m} p_i(1-p_i)} \tag{18}$$

$\boldsymbol{M}$ is the $m \times n$ genotype matrix, m is the total number of markers, and p_i is the minor allele frequency at the ith locus. The corresponding breeding value is estimated by:

$$\boldsymbol{u} = \left(\boldsymbol{Z}'\boldsymbol{Z} + V_g^{-1}.\sigma_e^2/\sigma_a^2\right)^{-1}\boldsymbol{Z}'(\boldsymbol{y} - \boldsymbol{X\beta}) \tag{19}$$

Use of the whole genome SNPs helps to capture Mendelian sampling variances caused by a free combination between gametes and the recombination between loci. This reduces the loss of genetic variation due to Mendelian sampling and makes the estimate of the $\boldsymbol{V}_g$ matrix be closer to the true value. As a result, more accurate estimates of breeding values ($\boldsymbol{u}$) and a higher correlation coefficient $r_{GEBV,\ A}$ are expected.

The second algorithm is to directly estimate individual GEBVs using a linear model:

$$\boldsymbol{Y} = \boldsymbol{\mu} + \boldsymbol{M\alpha} + \boldsymbol{e} \tag{20}$$

$\boldsymbol{\mu}$ is a vector of means, $\boldsymbol{M}$ is a SNP genotype or an allele occurrence matrix, $\boldsymbol{\alpha}$ is a vector of allelic effects at all loci, and $\boldsymbol{e}$ is a vector of residual errors. Based on the Bayesian models, individual GEBVs are estimated [69], e.g., $GEBV_i = \sum_j x_{ij}\hat{\alpha}_j$ for the ith individual. The use of genome-wide SNPs helps to estimate more accurate GEBVs and hence to improve the correlation coefficient $r_{GEBV,A}$.

Several factors influencing GS have been discussed in the literature [48,64], including model selection, marker density, LD, population size and structure, and the effects of genes controlling the traits. The second algorithm is more computationally demanding than the first one but has a greater scope for additional discoveries. Similarly, non-additive effects

could be included to predict genetic value (GV), which potentially increases $r_{GV,A+NA}$ and improves the genetic gain.

4. Advancing Prediction of Genetic Gains Using Multi-Omics Traits

Apart from selection using genomic and phenotypic traits, some omics traits are analogous to phenotypic quantitative traits and partially inherit across generations. When the target traits of breeding objectives are associated with the genes of epigenetic effects or with the genes whose expression can be detected, selection based on these omics traits is of significance in breeding. Note that the target traits of breeding objectives may also refer to the omics traits, such as the expression and protein contents of some specific genes. The main criteria for genetic gains are to genetically improve the populations by directly changing the frequencies of genes with detectable omics traits that are related to breeding objectives. This provides a way of designing molecular breeding to increase the frequencies of single or multiple target genes, which is complementary to the approach of gene editing in the laboratory. Here, we discuss two distinct ideas on how to predict genetic gains from a selection using gene expression (transcriptome) and epigenetic traits.

4.1. Gene Expression Traits

Under a certain environmental control, the expression (copy numbers of mRNA) of a particular gene varies among individuals in a population. The expression traits are usually transformed as normalized expressions, called RPKM (reads per kilobase million) or FPKM (fragments per kilobase million) values, which exhibit the characteristics of continuous variation. Extensive studies focus on the differential expression of genes under contrasting conditions or genetic backgrounds (e.g., different species). Studies are recorded in the literature on eQTL (expression quantitative trait loci) mapping of the expression traits and revealed gene regulatory interactions or gene-trait associations [70–73]. Studies on eQTL mapping mostly concentrate on humans [74–76] and animals [77], with few on plant species.

Theoretically, the expression trait can exhibit genetic variation both between and within populations. The expression trait, denoted by P_e, can be decomposed into genetic (G_e), environment (E_e), and genetic-by-environment ($G_e \times E_e$) effects. The linear model for expression trait of a single gene is written as:

$$P_e = G_e + E_e + G_e \times E_e \tag{21}$$

The subscript e stands for the gene expression trait, which is separated from our conventional symbols used for genetic and environment effects in quantitative genetics. When multiple genes are considered, Equation (21) is expressed in vectors. Different genetic parameters, such as genetic variances and heritability of expression traits, can be estimated using quantitative genetic methods, given a specific genetic mating and experimental design. The genetic gain from selection based on gene expression is estimated by applying the appropriate BE at different levels of selection. For example, the genetic gain from individual selection for the expression of a single gene is:

$$\Delta G_e = S_e h^2_{N(e)} \tag{22}$$

S_e is the selection differential and $h^2_{N(e)}$ is the narrow heritability for the gene expression trait.

When considering multiple genes simultaneously, we can construct the selection index as follows:

$$I_e = b_{e1}P_{e1} + b_{e2}P_{e2} + b_{e3}P_{e3} + \cdots \tag{23}$$

where b_{ej} and $P_{ej}(j = 1, 2, 3, \ldots)$ are the weight and expression value of the ith gene, respectively. The weight b_{ej} can be solved by the least square method (Min $\sum_l (I_{el} - A_{el})^2$ where l presents the lth individual). The genetic gain from an individual selection is calculated by:

$$\Delta G_e = i r_{I_e A(e)} \sigma_{A(e)} \tag{24}$$

where $\sigma_{A(e)}$ is the additive genetic variance of the expression trait and $r_{I_e A(e)}$ is the correlation coefficient between the selection index and the additive genetic effect [5].

Although few reports are available, the advantage of this approach lies in specifying which and how many genes are included in breeding. This provides a breeding strategy to modify the frequencies of a single gene or multiple specific genes, complementary to the way that uses molecular biotechniques to edit genes.

4.2. Epigenetic Traits

Epigenetic variation is another type of omics traits and can be used to map the relationship between genotype and phenotype. Epigenetic variation is caused by regulations in various phases during the formation of gene products (RNA and proteins), which leads to changes in cellular and individual phenotypes. Although DNA sequences are not altered, their structure is modified through the processes of DNA methylation, non-coding RNAs, histone modifications, and histone variant chromatin remodeling [78–81]. Epialleles at specific DNA loci can be generated [82]. Epigenetic study is an active research area and significant advances are reported in medicine [83], animal husbandry [84], aquaculture [85], and crops [86]. Reports are also available in tree species, including the studies focusing on the relationship of epigenetic variation and environmental adaptation in *Eucalyptus grandis* [80], *Picea abies* [87], and *Pinus pinea* [88] and the phenotypic plasticity or adaptation to local environments [89–92]. These studies provide evidence of phenotypic variation that is not explainable in terms of nuclear DNA sequence variation.

Epigenetic variation causes phenotypic variation through influencing transcriptomic, proteomic, or metabolomic activities or by regulating gene expression with positive and negative feedbacks in the intracellular environment. Such epigenetic effects could be transmitted to progeny and allow progeny to achieve higher phenotypic plasticity or adaptability to diverse environments, which provides a genetic basis for epigenetic selection and evolution [91]. Epigenetic loci also exhibit intra- and interpopulation variations [93].

Suppose that epigenetic effects can be separated from those caused by DNA sequence variation due to mutation in a quantitative trait. Following the quantitative genetic approach [5], the phenotypic value could be decomposed into genetic and epigenetic effects:

$$P = G + G_{epi} + E + G \times E + G_{epi} \times E \tag{25}$$

where G_{epi} and $G_{epi} \times E$ are the epigenetic effect and the interaction between epigenetic and environmental effects, respectively. If there are no interactions of genetic-by-environment and epigenetic-by-environment effects, the genetic gains from individual selection based on phenotypic traits, according to Lande and Arnold [94], can be expressed as

$$\Delta G = S\left(\sigma_G^2 + \sigma_{G(epi)}^2 + \sigma\left(G, G_{epi}\right)\right)/\sigma_P^2 \tag{26}$$

where $\sigma_{G(epi)}^2$ and $\sigma\left(G, G_{epi}\right)$ are the variance of epigenetic effects and the covariance between genetic and epigenetic effects, respectively. Epigenetic gain could account for a proportion of whole genetic gain to selection. Methodologically, a mixed linear model could be used to detect and estimate individual epigenetic effects. Variance and covariance components in Equation (26) can be estimated so that the epigenetic gains from selection are separated.

The significance of predicting epigenetic gain in breeding is to improve the adaptation of reproductive populations to specific environments, in addition to increasing the

frequency of genes associated with target traits. This also enhances the selection of suitable genetic materials for different environments or habitats so that the different breeding objectives can be realized. This line of breeding awaits data collection.

5. Conclusions and Perspective

Predicting genetic gain from an artificial selection is important for assessing genetic improvements in plant and animal breeding. The current prediction of genetic gains is mainly based on the classical BE. The limitation of the classical BE lies in the fact that it only exploits additive genetic effects. Several factors can cause biased predictions of genetic gains. They include (i) linkage disequilibria among loci and non-additive genetic effects that could partially be transmitted across generations, especially in populations with partially selfing or inbreeding; (ii) genetic drift that reduces the additive genetic variances or interacts with selection to fix the deleterious genes; (iii) input of alien pollen into the breeding populations that changes the additive genetic variance; (iv) the type of mating system that influences the estimates of heritability and genetic variances.

Conventional breeding methods can improve the prediction of genetic gains by reducing the environmental error in experimental design, by reducing the entry rates of selected individuals, and by expanding the genetic variation of candidate populations. Molecular breeding methods such as marker-assisted selection, genome-wide association analysis, and genomic selection permit the selection of desirable traits at the gene level. Here, the genetic gains are realized through increasing the correlation between the predicted genetic values and the breeding values.

It is of practical significance to advance genetic gains from selection using omics traits (gene expression and epigenetic traits). The breeder can specify which and how many genes are focused on breeding or on the selected individuals with adaptation to diverse habitats. From the practical perspective, bioinformatic and statistical methods are needed to screen markers associated with the omics traits, such as eQTL and epistatic allele mapping. A selection index can be constructed to simultaneously increase the frequencies of multiple genes.

Theoretically, partial inbreeding or selfing in breeding populations facilitates the transmission of non-additive effects across generations. Further, analysis of multiple omics traits will aid in capturing a greater proportion of the heritability. When GWAS and GS are performed, some epistatic effects, especially the additive–by–additive effects, can be potentially detected. All these genetic components may be recorded for use in practical selection but require theoretical extensions of the classical BE to include the non-additive genetic effects. Our comprehensive evaluation of genetic gain predictions is intended to provide a useful background for generating new research ideas and as a reference for what has been conducted.

Author Contributions: Conceptualization, X.-S.H. and F.C.Y.; investigation, Z.-H.H., Y.X. and X.W.; resources and data curation, Z.-H.H., Y.X., Y.-W.L. and X.W.; writing—original draft preparation, Z.-H.H. and X.W.; writing—review and editing, F.C.Y. and X.-S.H.; funding acquisition, X.-S.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was founded jointly by the 14th Five-Year Key Program of Research and Development of China, grant number 2022YFD2200205, and South China Agricultural University, grant number 4400-K16013.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data used in this article were included in the main text.

Acknowledgments: We appreciate four anonymous reviewers for constructive comments on this article.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Darwin, C. *On the Origin of Species by Means of Natural Selection, or Preservation of Favored Races in the Struggle for Life*, 1st ed.; J. Murray: London, UK, 1859.
2. Lush, J.L. *Animal Breeding Plans*; Iowa State College Press: Ames, IA, USA, 1937.
3. Griffing, B. Theoretical consequences of truncation selection based on the individual phenotype. *Aust. J. Biol. Sci.* **1960**, *13*, 307. [CrossRef]
4. Griffing, B. Selection in reference to biological groups. I. Individual and group selection applied to populations of unordered groups. *Aust. J. Biol. Sci.* **1967**, *20*, 127–139. [CrossRef] [PubMed]
5. Falconer, D.S.; Mackay, T.F.C. *Introduction to Quantitative Genetics*; Pearson: London, UK, 1996.
6. Fisher, R.A. *The Genetical Theory of Natural Selection*; Oxford University Press: Oxford, UK, 1930.
7. Robertson, A. A mathematical theory of the culling process in dairy cattle. *Anim. Sci.* **1966**, *8*, 95–108. [CrossRef]
8. Robertson, A. The spectrum of genetic variation. In *Population Biology and Evolution*; Lewontin, R., Ed.; Syracuse University Press: Syracuse, NY, USA, 1968; pp. 5–16.
9. Price, G.R. Selection and covariance. *Nature* **1970**, *227*, 520–521. [CrossRef]
10. Price, G.R. Extension of covariance selection mathematics. *Ann. Hum. Genet.* **1972**, *35*, 485–490. [CrossRef]
11. Morrissey, M.B.; Parker, D.J.; Korsten, P.; Pemberton, J.M.; Kruuk, L.E.B.; Wilson, A.J. The prediction of adaptive evolution: Em pirical application of the secondary theorem of selection and comparison to the breeder's equation. *Evolution* **2012**, *66*, 2399–2410. [CrossRef] [PubMed]
12. Queller, D.C. Fundamental theorems of evolution. *Am. Nat.* **2017**, *189*, 345–353. [CrossRef]
13. Lande, R. Quantitative genetic-analysis of multivariate evolution, applied to brain–body size allometry. *Evolution* **1979**, *33*, 402–416. [CrossRef] [PubMed]
14. Roff, D.A. A centennial celebration for quantitative genetics. *Evolution* **2007**, *61*, 1017–1032. [CrossRef]
15. Namkoong, G. *Introduction to Quantitative Genetics in Forestry*; U.S. Department of Agriculture: Washington, DC, USA, 1979.
16. Wright, S. *Evolution and the Genetics of Populations, Volume 2: Theory of Gene Frequencies*; University of Chicago Press: Chicago, IL, USA, 1969.
17. Yang, R.C.; Yeh, F.C.; Yanchuk, A.D. A comparison of isozyme and quantitative genetic variation in *Pinus contorta* ssp. latifolia by FST. *Genetics* **1996**, *142*, 1045–1052. [CrossRef]
18. Wei, R.-P. Response to Selection with restrictions while considering effective family number. *Hereditas* **1995**, *123*, 53–59. [CrossRef] [PubMed]
19. Wei, R.-P.; Lindgren, D. Selection effects on diversity and genetic gain. *Silva Fenn.* **1991**, *25*, 229–234. [CrossRef]
20. Lindgren, D.; Mullin, T.J. Balancing gain and relatedness in selection. *Silvae Genet.* **1997**, *46*, 124–129.
21. Wei, R.-P.; Yeh, F.C.; Lindgren, D. Expected gain and status number following restricted individual and combined-index selection. *Genome* **1997**, *40*, 1–8. [CrossRef]
22. Hill, W.G.; Robertson, A. The effect of linkage on limits to artificial selection. *Genet. Res.* **1966**, *8*, 269–294. [CrossRef]
23. Lynch, M.; Walsh, B. *Genetics and Analysis of Quantitative Traits*; Sinauer Associates Inc.: Sunderland, MA, USA, 1998.
24. Wright, S. The genetic structure of populations. *Ann. Eugen.* **1951**, *15*, 323–354. [CrossRef]
25. Walsh, B.; Lynch, M. *Evolution and Selection of Quantitative Traits*; Oxford University Press: Oxford, UK, 2018. [CrossRef]
26. Kimura, M. On the fixation probability of mutant genes in a subdivided population. *Genetics* **1962**, *47*, 713–719. [CrossRef]
27. Kimura, M.; Crow, J.F. Effect of overall phenotypic selection on genetic change at individual loci. *Proc. Natl. Acad. Sci. USA* **1978**, *75*, 6168–6171. [CrossRef]
28. Robertson, A. A theory of limits in artificial selection. *Proc. R. Soc. London. Ser. B Biol. Sci.* **1960**, *153*, 234–259. [CrossRef]
29. Hill, W.G. Predictions of response to artificial selection from new mutations. *Genet. Res.* **1982**, *40*, 255–278. [CrossRef]
30. Hu, X.S.; Zeng, W.; Li, B.L. Impacts of one-way gene flow on genetic variance components in a natural population. *Silvae Genet.* **2003**, *52*, 18–24.
31. Hu, X.S.; Chen, X.Y.; Yeh, F.C. *Forest Population Genetics*; China Forestry Publishers Education Branch: Beijing, China, 2021.
32. Xu, Y.B.; Li, P.; Zou, C.; Lu, Y.; Xie, C.; Zhang, X.; Prasanna, B.M.; Olsen, M.S. Article navigation enhancing genetic gain in the era of molecular breeding. *J. Exp. Bot.* **2017**, *68*, 2641–2666. [CrossRef]
33. Hu, X.S.; Wu, R.L.; Han, Y.F. Exploring the sustainable management of genetic resources of forest tree populations II. Analysis of natural and artificial population management of important native tree species in China. *For. Res.* **2001**, *14*, 1–7.
34. Mulder, H.A.; Bijma, P. Effects of genotype × environment interaction on genetic gain in breeding programs. *J. Anim. Sci.* **2005**, *83*, 49–61. [CrossRef] [PubMed]
35. Allwright, M.R.; Taylor, G. Molecular breeding for improved second generation bioenergy crops. *Trends Plant Sci.* **2016**, *21*, 43–54. [CrossRef] [PubMed]
36. Endelman, J.B.; Atlin, G.N.; Beyene, Y.; Semagn, K.; Zhang, X.; Sorrells, M.E.; Jannink, J. Optimal design of preliminary yield trials with genome-wide markers. *Crop Sci.* **2014**, *54*, 48–59. [CrossRef]

37. Lande, R.; Thompson, R. Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* **1990**, *124*, 743–756. [CrossRef]
38. Liu, P.; Zhu, J.; Lou, X.; Lu, Y. A method for marker-assisted selection based on QTLs with epistatic effects. *Genetica* **2003**, *119*, 75–86. [CrossRef]
39. Hu, X.-S. A general framework for marker-assisted selection. *Theor. Popul. Biol.* **2007**, *71*, 524–542. [CrossRef]
40. Wu, H.X. Study of early selection in tree breeding 4. efficiency of marker-aided early selection (MAES). *Silvae Genet.* **2002**, *51*, 261–269.
41. Wang, X.; Ma, P.; Liu, J.; Zhang, Q.; Zhang, Y.; Ding, X.; Jiang, L.; Wang, Y.; Zhang, Y.; Sun, D.; et al. Genome-wide association study in Chinese Holstein cows reveal two candidate genes for somatic cell score as an indicator for mastitis susceptibility. *BMC Genet.* **2015**, *16*, 111. [CrossRef]
42. Li, Y.-H.; Reif, J.C.; Hong, H.-L.; Li, H.-H.; Liu, Z.-X.; Ma, Y.-S.; Li, J.; Tian, Y.; Li, Y.-F.; Li, W.-B.; et al. Genome-wide association mapping of QTL underlying seed oil and protein contents of a diverse panel of soybean accessions. *Plant Sci.* **2018**, *266*, 95–101. [CrossRef] [PubMed]
43. Koochakpouk, Z.; Solouki, M.; Fakheri, B.A.; Aghnoum, R.; Nezhad, N.M. Genome wide association analysis of plant height, spike and awn length in barley (*Hordeum vulgare* L.) exposed to Mn stress. *J. Anim. Plant Sci.* **2020**, *30*, 384–390. [CrossRef]
44. Feng, J.L.; Jiang, Y.; Yang, Z.J.; Chen, S.P.; EI-Kassaby, Y.A.; Chen, H. Marker-assisted selection in *C. oleifera* hybrid population. *Silvae Genet.* **2020**, *69*, 63–72. [CrossRef]
45. Vijayan, K.; Srivastava, P.P.; Nair, C.V.; Awasthi, A.K.; Tikader, A.; Sreenivasa, B.; Urs, S.R. Molecular characterization and identification of markers associated with leaf yield traits in mulberry using ISSR markers. *Plant Breed.* **2006**, *125*, 298–301. [CrossRef]
46. Kar, P.K.; Srivastava, P.P.; Awasthi, A.K.; Urs, S.R. Genetic variability and association of ISSR markers with some biochemical traits in mulberry (*Morus* spp.) genetic resources available in India. *Tree Genet. Genomes* **2008**, *4*, 75–83. [CrossRef]
47. Vijayan, K.; Nair, C.V.; Chatterjee, S.N. Diversification of mulberry (*Morus indica* var. S36)—A vegetatively propagated tree species. *Casp. J. Env. Sci.* **2009**, *7*, 23–30.
48. Ahmar, S.; Ballesta, P.; Ali, M.; Mora-Poblete, F. Achievements and challenges of genomics-assisted breeding in forest trees: From marker-assisted selection to genome editing. *Int. J. Mol. Sci.* **2021**, *22*, 10583. [CrossRef]
49. Nagano, S.; Hirao, T.; Takashima, Y.; Matsushita, M.; Mishima, K.; Takahashi, M.; Iki, T.; Ishiguri, F.; Hiraoka, Y. SNP Genotyping with target amplicon sequencing using a multiplexed primer panel and its application to genomic prediction in Japanese Cedar, *Cryptomeria japonica* (L.f.) D.Don. *Forests* **2020**, *11*, 898. [CrossRef]
50. McKown, A.D.; Guy, R.D.; Quamme, L.; Klápště, J.; La Mantia, J.; Constabel, C.P.; El-Kassaby, Y.A.; Hamelin, R.C.; Zifkin, M.; Azam, M.S. Association genetics, geography and ecophysiology link stomatal patterning in Populus trichocarpa with carbon gain and disease resistance trade-offs. *Mol. Ecol.* **2014**, *23*, 5771–5790. [CrossRef]
51. Du, Q.; Gong, C.; Wang, Q.; Zhou, D.; Yang, H.; Pan, W.; Li, B.; Zhang, D. Genetic architecture of growth traits in Populus revealed by integrated quantitative trait locus (QTL) analysis and association studies. *N. Phytol.* **2016**, *209*, 1067–1082. [CrossRef] [PubMed]
52. Fahrenkrog, A.M.; Neves, L.G.; Resende, M.F.R.; Vazquez, A.I.; Campos, G.; Dervinis, C.; Sykes, R.; Davis, M.; Davenport, R.; Barbazuk, W.B.; et al. Genome-wide association study reveals putative regulators of bioenergy traits in Populus deltoides. *New Phytol.* **2017**, *213*, 799–811. [CrossRef] [PubMed]
53. Guerra, F.P.; Suren, H.; Holliday, J.; Richards, J.H.; Fiehn, O.; Famula, R.; Stanton, B.J.; Shuren, R.; Sykes, R.; Davis, M.F.; et al. Exome resequencing and GWAS for growth, ecophysiology, and chemical and metabolomic composition of wood of Populus trichocarpa. *BMC Genom.* **2019**, *20*, 875. [CrossRef] [PubMed]
54. Valenzuela, C.; Ballesta, P.; Ahmar, S.; Fiaz, S.; Heidari, P.; Maldonado, C.; Mora-Poblete, F. Haplotype- and SNP-based GWAS for growth and wood quality traits in Eucalyptus cladocalyx trees under arid conditions. *Plants* **2021**, *10*, 148. [CrossRef]
55. Yu, Y.; Chen, C.; Gmitter, F.G. QTL mapping of mandarin (Citrus reticulata) fruit characters using high-throughput SNP markers. *Tree Genet. Genomes* **2016**, *12*, 77. [CrossRef]
56. Minamikawa, M.F.; Nonaka, K.; Kaminuma, E.; Kajiya-Kanegae, H.; Onogi, A.; Goto, S.; Yoshioka, T.; Imai, A.; Hamada, H.; Hayashi, T.; et al. Genome-wide association study and genomic prediction in citrus: Potential of genomics-assisted breeding for fruit quality traits. *Sci. Rep.* **2017**, *7*, 4721. [CrossRef]
57. Duan, N.; Bai, Y.; Sun, H.; Wang, N.; Ma, Y.; Li, M.; Wang, X.; Jiao, C.; Legall, N.; Mao, L.; et al. Genome re-sequencing reveals the history of apple and supports a two-stage model for fruit enlargement. *Nat. Commun.* **2017**, *8*, 429. [CrossRef]
58. Manolio, T.A.; Collins, F.S.; Cox, N.J.; Goldstein, D.B.; Hindorff, L.A.; Hunter, D.J.; McCarthy, M.I.; Ramos, E.M.; Cardon, L.R.; Chakravarti, A.; et al. Finding the missing heritability of complex diseases. *Nature* **2009**, *461*, 747–753. [CrossRef]
59. Lamara, M.; Raherison, E.; Lenz, P.; Beaulieu, J.; Bousquet, J.; MacKay, J. Genetic architecture of wood properties based on association analysis and co-expression networks in white spruce. *N. Phytol.* **2016**, *210*, 240–255. [CrossRef]
60. Tan, B.; Ingvarsson, P.K. Integrating genome-wide association mapping of additive and dominance genetic effects to improve genomic prediction accuracy in Eucalyptus. *Plant Genome* **2022**, *15*, e20208. [CrossRef]
61. Fraser, H.B.; Hirsh, A.E.; Steinmetz, L.M.; Scharfe, C.; Feldman, M.W. Evolutionary rate in the protein interaction network. *Science* **2002**, *296*, 750–752. [CrossRef] [PubMed]
62. Goddard, M.E.; Hayes, B.J. Genomic selection. *J. Anim. Breed. Genet.* **2007**, *124*, 323–330. [CrossRef] [PubMed]

63. Yi, L.L.; Ma, Y.L.; Xing, T.; Zhu, M.J.; Yu, M.; Li, X.Y.; Liu, X.L.; Zhao, S.H. Progress and perspectives on genome-wide selection models. *J. Anim. Husb. Vet. Med.* **2019**, *50*, 233–242. [CrossRef]
64. Zhu, S.; Huang, M.R. Progress and prospects of genomic selection in genetic breeding research of forest trees. *Sci. Silvae Sin.* **2020**, *56*, 176–185.
65. McGowan, M.; Wang, J.; Dong, H.; Liu, X.; Jia, Y.; Wang, X.; Iwata, H.; Li, Y.; Lipka, A.E.; Zhang, Z. Ideas in genomic selection with the potential to transform plant molecular breeding: A review. *Plant Breed Rev.* **2021**, *45*, 273–319. [CrossRef]
66. Calleja-Rodriguez, A.; Chen, Z.; Suontama, M.; Pan, J.; Wu, H.X. Genomic predictions with nonadditive effects improved estimates of additive effects and oredictions of total genetic values in Pinus sylvestris. *Front. Plant Sci.* **2021**, *12*, 666820. [CrossRef] [PubMed]
67. Denis, M.; Favreau, B.; Ueno, S.; Camus-Kulandaivelu, L.; Chaix, G.; Gion, J.-M.; Nourrisier-Mountou, S.; Polidori, J.; Bouvet, J.-M. Genetic variation of wood chemical traits and association with underlying genes in *Eucalyptus urophylla*. *Tree Genet. Genomes* **2013**, *9*, 927–942. [CrossRef]
68. VanRaden, P. Efficient methods to compute genomic predictions. *J. Dairy Sci.* **2008**, *91*, 4414–4423. [CrossRef]
69. Meuwissen, T.H.E.; Hayes, B.J.; Goddard, M.E. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **2001**, *157*, 1819–1829. [CrossRef]
70. Sun, W.; Hu, Y. eQTL Mapping Using RNA-seq Data. *Stat. Biosci.* **2012**, *5*, 198–219. [CrossRef]
71. Wainberg, M.; Sinnott-Armstrong, N.; Mancuso, N.; Barbeira, A.N.; Knowles, D.A.; Golan, D.; Ermel, R.; Ruusalepp, A.; Quertermous, T.; Hao, K.; et al. Opportunities and challenges for transcriptome-wide association studies. *Nat. Genet.* **2019**, *51*, 592–599. [CrossRef] [PubMed]
72. Kang, M.; Gao, J. Integration of multi-omics data for expression quantitative trait loci (eQTL) analysis and eQTL epistasis. *Methods Mol. Biol.* **2020**, *2082*, 157–171. [CrossRef] [PubMed]
73. Võsa, U.; Claringbould, A.; Westra, H.J.; Bonder, M.J.; Deelen, P.; Zeng, B.; Kirsten, H.; Saha, A.; Kreuzhuber, R.; Yazar, S.; et al. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat. Genet.* **2021**, *53*, 1300. [CrossRef] [PubMed]
74. Petretto, E.; Mangion, J.; Dickens, N.J.; Cook, S.A.; Kumaran, M.K.; Lu, H.; Fischer, J.; Maatz, H.; Kren, V.; Pravenec, M.; et al. Heritability and tissue specificity of expression quantitative trait loci. *PLoS Genet.* **2006**, *2*, e172. [CrossRef]
75. Lloyd-Jones, L.R.; Holloway, A.; McRae, A.; Yang, J.; Small, K.; Zhao, J.; Zeng, B.; Bakshi, A.; Metspalu, A.; Dermitzakis, M.; et al. The genetic architecture of gene expression in peripheral blood. *Am. J. Hum. Genet.* **2017**, *100*, 228–237. [CrossRef]
76. Yao, D.W.; O'Connor, L.J.; Price, A.L.; Gusev, A. Quantifying genetic effects on disease mediated by assayed gene expression levels. *Nat. Genet.* **2020**, *52*, 626–633. [CrossRef]
77. Ponsuksili, S.; Oster, M.; Reyer, H.; Hadlich, F.; Trakooljul, N.; Rodehutscord, M.; Camarinha-Silva, A.; Bennewitz, J.; Wimmers, K. Genetic regulation and heritability of miRNA and mRNA expression link to phosphorus utilization and gut microbiome. *Open Biol.* **2021**, *11*, 200182. [CrossRef]
78. Matzke, M.A.; Mosher, R.A. RNA-directed DNA methylation: An epigenetic pathway of increasing complexity. *Nat. Rev. Genet.* **2014**, *15*, 394–408. [CrossRef]
79. Bei, X.; Shahid, M.Q.; Wu, J.; Chen, Z.; Wang, L.; Liu, X. Re-sequencing and transcriptome analysis reveal rich DNA variations and differential expressions of fertility-related genes in neo-tetraploid rice. *PLoS ONE* **2019**, *14*, e0214953. [CrossRef]
80. Hussey, S.G.; Mizrachi, E.; Groover, A.; Berger, D.K.; Myburg, A.A. Genome-wide mapping of histone H3 lysine 4 trimethylation in Eucalyptus grandis developing xylem. *BMC Plant Biol.* **2015**, *15*, 117. [CrossRef]
81. Kawashima, T.; Lorković, Z.J.; Nishihama, R.; Ishizaki, K.; Axelsson, E.; Yelagandula, R.; Kohchi, T.; Berger, F. Diversification of histone H2A variants during plant evolution. *Trends Plant Sci.* **2015**, *20*, 419–425. [CrossRef]
82. Hou, Q.; Wan, X. Epigenome and epitranscriptome: Potential resources for crop improvement. *Int. J. Mol. Sci.* **2021**, *22*, 12912. [CrossRef] [PubMed]
83. Yen, C.-Y.; Huang, H.-W.; Shu, C.-W.; Hou, M.-F.; Yuan, S.-S.F.; Wang, H.-R.; Chang, Y.-T.; Farooqi, A.A.; Tang, J.-Y.; Chang, H.-W. DNA methylation, histone acetylation and methylation of epigenetic modifications as a therapeutic approach for cancers. *Cancer Lett.* **2016**, *373*, 185–192. [CrossRef]
84. Ibeagha-Awemu, E.M.; Zhao, X. Epigenetic marks: Regulators of livestock phenotypes and conceivable provenances of missing variation in livestock improvement programs. *Front. Genet.* **2015**, *6*, 302. [CrossRef] [PubMed]
85. Granada, L.; Lemos, M.F.; Cabral, H.N.; Bossier, P.; Novais, S.C. Epigenetics in aquaculture—The last frontier. *Rev. Aquac.* **2018**, *10*, 994–1013. [CrossRef]
86. Yu, J.; Xu, F.; Wei, Z.; Zhang, X.; Chen, T.; Pu, L. Epigenomic landscape and epigenetic regulation in maize. *Theor. Appl. Genet.* **2020**, *133*, 1467–1489. [CrossRef]
87. Yakovlev, I.A.; Carneros, E.; Lee, Y.; Olsen, J.E.; Fossdal, C.G. Transcriptional profiling of epigenetic regulators in somatic embryos during temperature induced formation of an epigenetic memory in Norway spruce. *Planta* **2016**, *243*, 1237–1249. [CrossRef]
88. Sáez-Laguna, E.; Guevara, M.; Díaz, L.-M.; Sánchez-Gómez, D.; Collada, C.; Aranda, I.; Cervera, M.-T. Epigenetic variability in the genetically uniform forest tree species *Pinus pinea* L. *PLoS ONE* **2014**, *9*, e103145. [CrossRef]
89. Bräutigam, K.; Vining, K.J.; Lafon-Placette, C.; Fossdal, C.G.; Mirouze, M.; Marcos, J.G.; Fluch, S.; Fraga, M.F.; Guevara, M.Á.; Abarca, D.; et al. Epigenetic regulation of adaptive responses of forest tree species to the environment. *Ecol. Evol.* **2013**, *3*, 399–415. [CrossRef]

90. Sow, M.D.; Allona, I.; Ambroise, C.; Conde, D.; Fichot, R.; Gribkova, S.; Jorge, V.; Le-Provost, G.; Pâques, L.; Plomion, C.; et al. Epigenetics in forest trees: State of the art and potential implications for breeding and management in a context of climate change. *Adv. Bot. Res.* **2018**, *88*, 387–453. [CrossRef]
91. Amaral, J.; Ribeyre, Z.; Vigneaud, J.; Sow, M.D.; Fichot, R.; Messier, C.; Pinto, G.; Nolet, P.; Maury, S. Advances and promises of epigenetics for forest trees. *Forests* **2020**, *11*, 976. [CrossRef]
92. Klupczyńska, E.A.; Ratajczak, E. Can forest trees cope with climate change?—Effects of DNA methylation on gene expression and adaptation to environmental change. *Int. J. Mol. Sci.* **2021**, *22*, 13524. [CrossRef] [PubMed]
93. Baldanzi, S.; Watson, R.; McQuaid, C.; Gouws, G.; Porri, F. Epigenetic variation among natural populations of the South African sandhopper *Talorchestia capensis*. *Evol. Ecol.* **2017**, *31*, 77–91. [CrossRef]
94. Lande, R.; Arnold, S.J. The measurement of selection on correlated characters. *Evolution* **1983**, *37*, 1210–1226. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.