PROCEEDINGS B

royalsocietypublishing.org/journal/rspb

Commentary

Cite this article: McGaugh SE, Lorenz AJ, Flagel LE. 2021 The utility of genomic prediction models in evolutionary genetics. *Proc. R. Soc. B* **288**: 20210693. https://doi.org/10.1098/rspb.2021.0693

Received: 23 March 2021
Accepted: 15 July 2021

Subject Category:
Genetics and genomics

Subject Areas:
evolution, genomics, genetics

Keywords:
BLUP, quantitative genetics, breeding value

Author for correspondence:
Suzanne E. McGaugh
e-mail: smcgaugh@umn.edu

†Present address: Gencove Inc., 430 East 29th Street, New York, NY 10016, USA.

Special feature: Wild Quantitative Genomics: the genomic basis of fitness variation in natural populations edited by Susan Johnston, Nancy Chen, Emily Josephs.

THE ROYAL SOCIETY PUBLISHING

The utility of genomic prediction models in evolutionary genetics

Suzanne E. McGaugh[1], Aaron J. Lorenz[3] and Lex E. Flagel[2,4,†]

[1]Ecology, Evolution, and Behavior, and [2]Plant and Microbial Biology, University of Minnesota, 140 Gortner Lab, 1479 Gortner Avenue, Saint Paul, MN 55108, USA
[3]Agronomy and Plant Genetics, University of Minnesota, 411 Borlaug Hall, 1991 Upper Buford Circle, Saint Paul, MN 55108, USA
[4]Bayer Crop Science, 700 W Chesterfield Parkway, Chesterfield, MO 63017, USA

SEM, 0000-0003-3163-3436

Variation in complex traits is the result of contributions from many loci of small effect. Based on this principle, genomic prediction methods are used to make predictions of breeding value for an individual using genome-wide molecular markers. In breeding, genomic prediction models have been used in plant and animal breeding for almost two decades to increase rates of genetic improvement and reduce the length of artificial selection experiments. However, evolutionary genomics studies have been slow to incorporate this technique to select individuals for breeding in a conservation context or to learn more about the genetic architecture of traits, the genetic value of missing individuals or microevolution of breeding values. Here, we outline the utility of genomic prediction and provide an overview of the methodology. We highlight opportunities to apply genomic prediction in evolutionary genetics of wild populations and the best practices when using these methods on field-collected phenotypes.

1. Introduction

The goal of genomic prediction is to make a prediction of the additive genetic value of an individual using genome-wide molecular markers. Genomic prediction was described almost 20 years ago, and has been a revolution in plant and animal breeding [1–5]. Genomic prediction models, in conjunction with drastically reduced genotyping costs, have produced gains in key crop and livestock traits and shortened the time required for artificial selection [1,6–8]. Genomic prediction has served these practical needs well, and has attracted a considerable amount of research effort. A related field has sprung up in human genetics and personalized medicine, where genomic prediction (often called polygenic risk scores) is used to identify patient risk for disease or treatment options conditioned on their unique genotype [9]. However, only a small number of pioneering studies have applied genomic prediction methods in wild populations [10–18].

The power of genomic prediction is the ability to predict the breeding value of unobserved individuals or individuals who have not yet expressed the trait of interest. This capability could be used in a variety of applications in evolutionary biology, including understanding if there is an association between viability selection and breeding values, or to understand the microevolution of breeding values through time [15,19]. Further, with some care, formal tests of genetic architecture can be made [12] and the additive genetic value for historical populations can be predicted. Clearly, there is much untapped promise for genomic prediction in evolutionary studies.

In this review, we provide an introduction to genomic prediction, an overview of some of the models and how to improve their accuracy. We then shift to the challenges associated with bringing a tool developed for animal breeding to studies of evolutionary genetics and opportunities in which genomic prediction could be used in the field of evolutionary genetics. We envision genomic prediction aiding in selecting individuals to breed for conservation

genetics, wildlife disease resistance, and climate change resilience and to predict genetic component of phenotypes that are unobservable, such as for historical DNA samples.

2. Goals of genomic prediction

Genomic prediction models were developed to accelerate the breeding process by identifying individuals with high breeding merit for a particular trait (i.e. genomic selection, [2]). In the breeding literature, the additive genetic value an individual has for a phenotype is called the breeding value [20,21]. Estimating the breeding value of an organism can be especially useful when the phenotype is unobservable in that particular individual, such as the breeding value for milk production for a bull or when unmeasured individuals have undergone viability selection [15,19].

In genomic prediction, the breeding value is estimated by summing additive genetic effects of all genome-wide markers (e.g. single nucleotide polymorphisms (SNPs), indels) for a focal individual. Using appropriate training data (individuals whose genotypes and phenotypes have been observed), genomic prediction produces a prediction model where each genotype combination (e.g. aa, Aa, AA) is assigned a value indicating its contribution to the breeding value. Markers in regions of the genome with a large impact on phenotype will tend to have a large contrast among the estimated genotypic effects (e.g. aa < Aa < AA), whereas regions with little impact on phenotype will have small differences (e.g. aa ≈ Aa ≈ AA).

Significance thresholds are not applied to individual markers in genomic prediction as is commonly done in quantitative trait loci (QTL) mapping or genome-wide association studies (GWAS). Instead, all markers are allowed to contribute to the prediction. Additionally, in genomic prediction, all marker effects are estimated in a single model, whereas in GWAS, a model is typically fit for each marker (i.e. single-marker regression) [22]. In the following sections, we provide a brief overview of the basics, though, we encourage readers to examine previous work [23] as well as a curated virtual series highlighting genomic prediction https://www.genetics.org/content/genomic-prediction.

Box 1. Basic principles of GP: model fitting and cross-validation.

Estimating the phenotypic impact of a large number of markers simultaneously is a challenging task. The number of individuals (N) will frequently be less than the number of markers genotyped (P). When $N < P$, there is no unique least-squares solution, so multiple regression cannot be used. Even if N is greater than P, estimates will suffer from the multicollinearity problem of multiple regression, which can make marker effect estimates unstable. This problem is compounded by high-density genotyping which produces datasets were many markers are tightly linked. For these reasons, multiple regression is inappropriate for genomic prediction. One approach to overcome the $N < P$ and multicollinearity problems is to bound marker effects to a prior distribution with a shared variance, such as using a normal distribution prior for marker effects in ridge regression [2]. While this allows predictions to be made when $N < P$, this assumption (i.e. marker effects come from a normal distribution) also results in 'shrinkage'. Shrinkage causes marker effects to be drawn towards zero, and the degree to which this occurs is controlled by the 'shrinkage parameter' (a value that controls the variance of the prior distribution assumed to govern marker effects) [8]. Optimizing the shrinkage parameter can improve model performance [24]. Finally, when fitting the genomic prediction model, it is important to assess the model's accuracy. One common technique is cross-validation, where the full dataset (records of individuals with known genotypes and phenotypes) is subset into a training data sample (e.g. 80%) and a test data sample (e.g. 20%). Once the model is trained on the training set, breeding values are estimated for the test set and compared to observed phenotypes to assess model performance. This is repeated such that every individual is in the training set multiple times and the test set once [25,26]. This cross-validation approach is extensively employed across different genomic prediction methods to tune various model parameters and evaluate factors affecting accuracy such as marker number, model type, and training populations size and composition [26–28].

3. Beyond the basics: building model flexibility

Genomic prediction models fit all marker effects in a single model. This creates complications that make it unfavourable or often impossible to use standard least-squares methods like multiple regression (box 1) and has necessitated specialized statistical methods. A good starting place for discussing genomic prediction methods is ridge regression [2,29]. Ridge regression is similar to least-squares regression in that it estimates the marker effects such that they minimize the residual error in the model. However, ridge regression goes beyond least squares with one additional constraint, which is that all marker effects are assumed to have a normal distribution with a mean of zero and a specific variance as their prior (figure 1*a*). By sharing a common variance, the marker effects experience 'shrinkage' (box 1), which draws them towards zero. This helps prevent overfitting on the training data and conceptually, the normal prior mimics the assumptions of the quasi-infinitesimal model of quantitative genetics. There is an equivalency under certain conditions between ridge regression and a technique known as genomic best linear unbiased prediction (GBLUP). GBLUP uses genetic markers to compute a genomic relationship matrix between individuals and then uses this matrix as a covariate in a linear mixed model [24,30,31]. For this reason, ridge regression methods (such as rrBLUP) [24] and GBLUP methods produce similar results [31].

Genomic prediction has expanded to other statistical models that may better fit specific genetic architectures (i.e. number of causative loci and their effect sizes). For example, LASSO (least absolute shrinkage and selection operator) employs a double exponential distribution for the marker effects prior rather than the normal prior that is used in ridge regression (figure 1*a*). The double exponential model has thicker tails and a higher peak at zero compared to a normal distribution. This allows more markers to be estimated with larger effects and with zero effect. For some traits, this distribution more realistically represents the underlying genetic architecture [32–34].

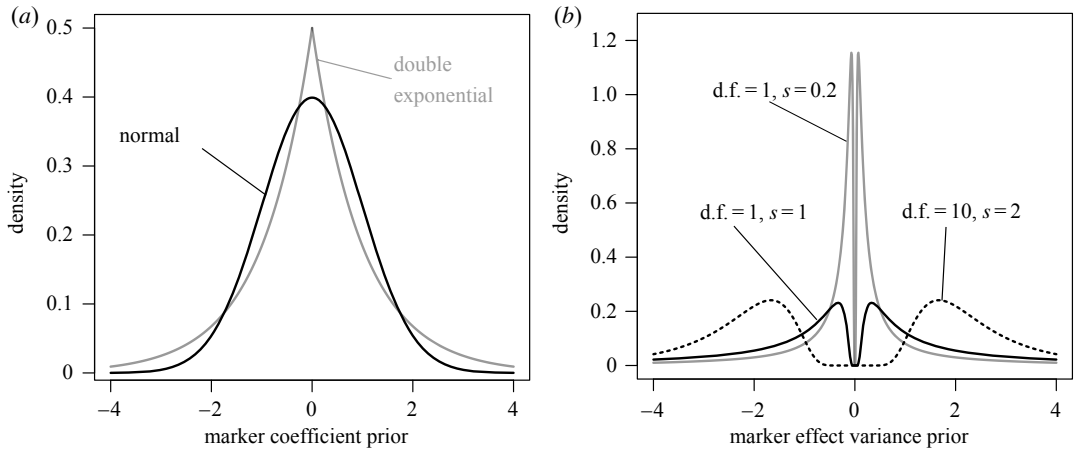

Figure 1. Marker effect coefficient prior and marker effect variance distributions used in genomic prediction models. (*a*) The double exponential (LASSO) and normal (ridge regression) marker effect coefficient prior distributions. (*b*) The marker effect variance prior for the scaled inverse χ^2 distribution (BayesA and BayesB). This distribution is governed by two parameters, degrees of freedom (d.f.) and a scaling factor (*s*). Three examples range from having most of the density near zero (d.f. = 1, $s = 0.2$) to most of the density spread away from zero (d.f. = 10, $s = 2$). These parameters make it possible to express priors corresponding to different genetic architectures. The scaled inverse χ^2 distribution has no density at zero. BayesB follows the scaled inverse χ^2 prior from BayesA, but adds a third parameter, π, which controls the proportion of marker effect variances that are zero.

LASSO and mixture models (models that allow marker effect sizes to be drawn from multiple distributions rather than a single distribution) are known as variable selection models. Variable selection models assume *a priori* that some markers have an effect size of zero on the phenotype, thus, can enable the fitting of large effect alleles better than GBLUP (box 1) [8,23,35]. There are multiple mixture models that differ in their prior distributions for marker effects and variances (e.g. BayesA, BayesB; figure 1*b*; and also BayesC, Bayesian LASSO and BayesR) [1,23]. For example, the prior in BayesC assumes that the marker effects either have no effect or follow a normal distribution estimated from the data, whereas Bayesian LASSO assumes that the marker effects either have no effect or follow a double exponential distribution (see comparable examples in figure 1*a*). The degree to which these methods shrink some marker effects to zero depends on the strength of the prior used and the data, and with some combinations, it may be possible to have no zero-effect markers [36].

In many breeding populations, the ridge regression model is as accurate as variable selection models since a normal distribution of marker effect sizes is a decent approximation of genetic architecture for many traits and can be computationally faster [4,30,37] (but see [35,38]). Breeding populations are special cases, as shared ancestry is often quite high and linkage disequilibrium (LD) is often extensive [37]. Moreover, breeding populations have typically experienced strong selection for economically valuable traits, which tends to fix beneficial large-effect alleles and purge detrimental large-effect alleles, leaving smaller phenotypic effects among the remaining segregating alleles [39,40]. In natural populations, variable selection models may be superior when individuals are unrelated and populations are diverse and polymorphic at some large-effect loci [37].

4. Striving for accuracy

Genomic prediction accuracy is a measure of how good the prediction equation is at predicting the breeding value of an individual solely from its genotype. Because breeding values are unknown, accuracy is often estimated as the correlation between the measured phenotypic value and the predicted breeding value (genomic estimated breeding value, GEBV). Accuracy, when estimated this way, is expected to be downwardly biased because the phenotype is an imperfect estimator of the breeding value. As a general rule-of-thumb, a correlation of 0.6 is considered moderate and many successful prediction equations obtain this level of accuracy [41,42]. The required genomic prediction accuracy varies across applications and should be determined within the context of how genomic selection will be used to increase the rate of genetic progress when compared with phenotypic selection.

There are several key characteristics that lead to more accurate prediction equations [43–46]:

(1) careful and accurate measurement of phenotypes (under relevant environments) [45],
(2) a highly heritable trait [47] (figure 2),
(3) a large training dataset [30] (figure 2),
(4) relatively close relationship between breeding population and training population [48,49],
(5) tight LD between the markers and the causative loci [50] and a number of markers sufficient to capture each unique haplotype [1,30],
(6) the genetic architecture of the trait [51] and how well it matches the model used [35,44,45].

An estimate of the accuracy, as measured by Pearson's correlation (*r*), between the GEBVs and the true breeding values for a population can be estimated *a priori* with the following equation:

$$r = \sqrt{\frac{Nh^2}{Nh^2 + M_e}},$$

where N is the training dataset size, h^2 is the trait heritability and M_e is the effective number of chromosome segments in the population [46]. M_e can be estimated by $M_e = 2N_eL$, where N_e is the effective population size and L is the length of the genetic map in Morgans [1], though there is debate

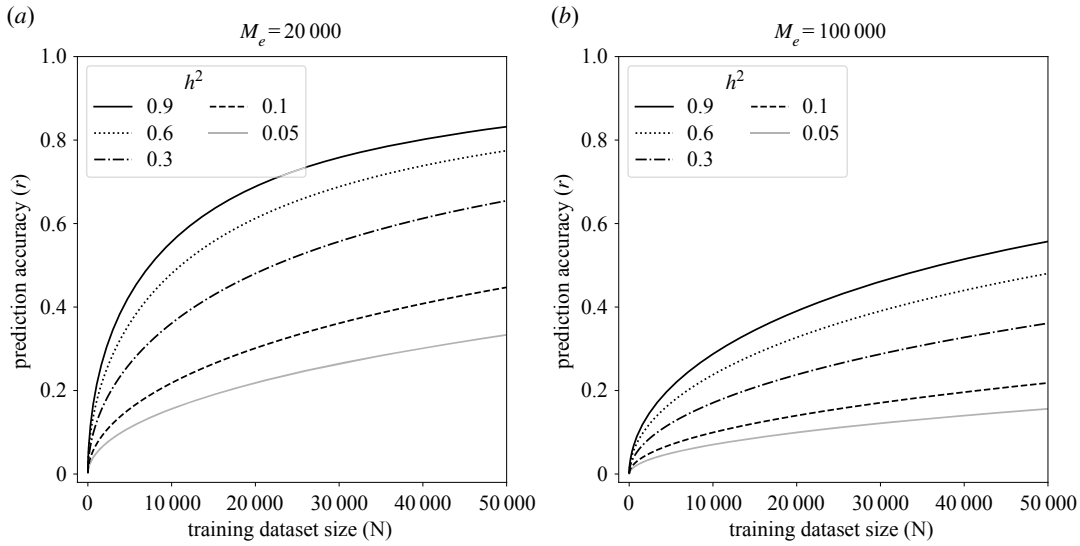

Figure 2. The relationship between training dataset size and prediction accuracy for five different heritability values. The effective number of chromosome segments (M_e) is 20 000 in *a* and 100 000 in *b*, corresponding to an N_e of 400 in *a* and 2000 in *b*.

on how M_e should be estimated [52]. In practice, this estimator represents the upper bound on accuracy, as it assumes marker saturation, where a marker tags every QTL in the population [30,46]. Figure 2 gives the relationship between the training dataset size (N) and the trait heritability (h^2) assuming $M_e = 20\,000$ (figure 2*a*) or $M_e = 100\,000$ (figure 2*b*). Given a 25 Morgan genetic map (as a general representative [53,54]), these values of M_e would be expected for a population with an N_e of approximately 400 and 2000, respectively. Reasonable prediction accuracy can be achieved for $M_e = 20\,000$ for high heritability traits with a moderate number of individuals. By contrast, accuracy is low when the trait heritability is low or when M_e is large (figure 2*b*). M_e is highly impacted by effective population size, and larger more diverse populations require more individuals in the training dataset and a higher marker density to reach high levels of accuracy [37]. This is an especially important consideration for natural populations.

More recent approaches to increase accuracy are centred on the development of models that better fit the genetic architecture or use prior functional information. For example, BayesRC includes a separate distribution for markers associated with known causal loci [8]. Another example is genomic feature BLUP (GFBLUP) which groups markers by gene function (e.g. GO annotation) or metabolic pathway information and allows that prior information to shape the effect size distributions [12,55]. Thus, the accuracy of genomic prediction models may be increased in natural populations for which a partial genetic basis is known, or by including additional biological knowledge from genomic features.

5. Considerations for applying genomic prediction to natural populations

Genomic prediction methods were developed for breeding programmes, and there are a few considerations regarding how they would perform in natural populations, chief among them are sample size and design of the training population. While we highlight considerations below, successful implementation of accurate genomic prediction in wild populations has been achieved [12,15,17].

(a) Sample size

Genomic prediction often requires large sample sizes and this scales with the diversity of the population being studied (figure 2). Recent livestock studies may incorporate 10 000+ animals, and sample sizes this large are often difficult to achieve in natural populations, or in fact may not exist for some species. The effective population size of some natural populations may also be several orders of magnitude higher than in domesticated species. This means training populations may need to be large to achieve reasonable accuracy [30] (figure 2). For example, a recent study of egg-laying date in approximately 2000 great tits found genomic prediction accuracy to be near 0.2, and this relatively low accuracy was attributed, in part, to the very large effective population size of that species [10].

Larger effective population sizes also mean that many more SNPs may be needed to capture all of the unique haplotypes associated with the trait [10,12,56]. Typically, SNP arrays are developed for genotyping agricultural species with tens of thousands of common SNPs across the populations (e.g. Illumina BovineSNP50 with approx. 50 000 SNPs, [57]). Natural populations may harbour more rare alleles and shorter LD blocks than domesticated species. Thus, marker density may need to be higher than domesticated species to ensure causal SNPs are linked with markers [37,56], and employing whole-genome sequencing may be helpful for some species [37].

(b) Training population

Thoughtfully designing a training population is key to the accuracy of the prediction model. For example, if the individuals in need of prediction are from a genetically less diverse population, the training population should be closely related to the prediction population. By contrast, to predict more diverse individuals, the training population will also need to be diverse and potentially quite large (table 1) [8]. If the training population is too distantly related to the prediction population, genomic prediction will be less accurate [59], as seen when prediction equations are applied across species, populations or breeds [60]. Newer models have shown that bridging information across breeds can be successful when focused on specific genomic regions that may have an outsized

Table 1. Sample sizes of breeding and evolutionary genetics studies and accuracies obtained.

study	species	trait	*N*	markers	h^2	accuracy	type of model
Ober *et al*. [13]	fruit fly (*Drosophila melanogaster*)	startle response	155	150 k–2.50 m	0.39[a]	0.239	GBLUP, BayesB
		starvation resistance	157		0.25[a]	0.230	GBLUP, BayesB
Beaulieu *et al*. [11]	white spruce (*Picea glauca*)	wood and growth	1694	6385	0.08–0.57	0.327–0.435	RRBLUP
Ober *et al*. [14]	fruit fly (*Drosophila melanogaster*)	chill coma recovery	176	1.87 m	0.35[b]	0–0.48[c]	GBLUP, trait-associated GBLUP
Edwards *et al*.[12]	fruit fly (*Drosophila melanogaster*)	startle response	13 276	1.73 m	0.47–0.49[d]	0.47–0.52[e]	GFBLUP
		starvation resistance	19 361	1.73 m	0.55–0.57[d]	0.37–0.43[e]	GFBLUP
		chill coma recovery	32 231	1.73 m	0.41–0.45[d]	0.32–0.37[e]	GFBLUP
Bosse *et al*. [18]	great tit (*Parus major*)	bill length	∼3000	376 k	DNS	DNS	BayesR
Gienapp *et al*. [10]	great tit (*Parus major)*	egg-laying date	2015	503 k	0.16–0.24	∼0.20	GBLUP
Stocks *et al*. [16]	European ash (*Fraxinus excelsior*)	ash dieback damage	150[f]	>10 k[g]	DNS	0.35	RRBLUP
Ashraf *et al*. [17][h]	Soay sheep (*Ovis aries*)	weight	1168	∼36 k	0.32	0.51	BayesR
		jaw length	897	∼36 k	0.59	0.38	BayesR
		foreleg length	1126	∼36 k	0.53	0.62	BayesR
		hindleg length	1139	∼36 k	0.50	0.59	BayesR
		metacarpal length	890	∼36 k	0.62	0.65	BayesR
		male horn length	472	∼36 k	0.42	0.67	BayesR
		coat colour	4737	∼36 k	DNS	1.00	BayesR
		coat pattern	4737	∼36 k	DNS	0.98	BayesR
Hunter *et al*. [15][i]	Soay sheep (*Ovis aries*)	adult body weight	1168	∼36 k	0.34–0.49 [58]	DNS	BayesR

[a]Heritability from random additive line effect g in ASReml; DNS, 'data not shown'.
[b]Broad-sense heritability.
[c]Depending on model, SNPs used and sex. This trait is heavily influenced by epistasis.
[d]Genomic heritability.
[e]These are the highest estimates for males and females given in the text on page 1876 of this publication.
[f]Thirty-one pools from pool seq used as training population and 150 individuals from moderate-coverage whole-genome sequencing as test dataset.
[g]Used only SNPs with most significant *p*-values in GWAS.
[h]Accuracies given for 50% training population and Bayes R taken from Table S2 in the manuscript. Accuracies vary based on traits and models and the manuscript reports results from GBLUP, LASSO, Bayes A, Bayes B and Bayes R.
[i]One thousand one hundred and sixty-eight genotyped and phenotyped individuals. Another 5627 animals were only phenotyped.
[101] Estimate taken from that citation because estimate was not provided in the manuscript.

impact on the trait of interest (GFBLUP [55]), and BayesRC [61] appears to be an advance in across-breed performance.

Similar to the issue of elevated effective population size, natural populations often exhibit genetic structure by distance or environment. Such issues plague GWAS, where they can cause spurious signatures of association [62]. Genomic prediction differs from GWAS in that it is generally not concerned with identifying causative loci. In genomic prediction, if the training and prediction populations share the same population structure and the trait of interest is associated with this structure, population structure can contribute beneficially to prediction accuracy, even if the alleles associated with this structure themselves are not causative [63]. One must use caution, though, if phenotypes are obtained from individuals where the direct effects of environment on phenotypes are associated with the genetic structure of the population.

An interesting example was found in a multi-breed sheep study on fleece weight. SNPs on chromosome 1 recovered 86% of the prediction accuracy of the full-genome model [63]. The authors hypothesized this is not because most fleece weight QTL reside on chromosome 1, but rather that a single chromosome captured much of the signal of population structure, and the structure itself was correlated with fleece weight. Next, the authors attempted to control for population structure (as one would in GWAS) and found that model accuracy decreased, a result that has also been shown in maize and rice [64].

In these studies, population structure was shared between the training and prediction populations. In natural populations, population structure differences may be detected between the training and prediction populations. As in GWAS, the goal of accounting for structure when it differs between training and prediction populations is to gain prediction accuracy by avoiding spurious associations due to structure alone [63]. Issues with population structure also need to be monitored over time, and updates to the prediction model and/or training population may need to be made if population structure changes (but see [37]).

To help infer the expected prediction accuracy for a planned training population size and composition, deterministic equations or simulations are often employed *a priori*, and cross-validation could be useful to give a quick assessment of empirical prediction accuracy [52,59,65,66].

(c) Genotype × environmental interactions and non-additive effects

In theory, the breeding value captures only the additive genetic component of the phenotype. This means that under ideal conditions, breeding values omit both non-additive genetic and environmental factors, which often interact with the genotype. Thus, the predicted breeding value is typically defined only for a target set of environmental conditions. One interesting example of how to deal with this complicating factor comes from [15,17] who opted to fit mixed models that incorporated fixed effects, non-genetic random effects and repeated measures. The random effect of individual identity was extracted and incorporated as the phenotype in genomic prediction. This resulted in a high-accuracy genomic prediction model.

There are extensions of genomic prediction models that incorporate non-additive effects and genotype-by-environmental interactions (GxE). Some models have been developed to explicitly capture epistasis [67–69] and dominance [70,71], while others, through the use of nonlinear functions relating genotype to phenotype, hold the potential to capture cryptic interactions [72]. Empirical studies on the superiority of these models in genomic prediction accuracy are inconsistent [4,73–75].

Similarly, genomic prediction models have been developed to better account for GxE effects [76]. These models would incorporate genomic, phenotypic and environmental data and allow for the prediction of unobserved genotypes in future, unobserved environments [77], though fully attaining this outcome is still a major challenge of the field [78,79].

Box 2. Superior progeny value.

One way to evaluate the value of a cross is to estimate the expected performance of the offspring. The superior progeny value (s) is defined as $s = \mu + i\sigma_g$, where μ is the progeny mean, i is the selection intensity and σ_g is the genetic standard deviation for a trait [80,81]. Superior progeny value estimates can be made by genotyping the parents and simulating a large number of offspring from those parents [82]. The genomic prediction model can be used to predict the expected breeding value of the simulated offspring (figure 3) and estimate μ and σ_g. This approach is limited by the accuracy of the underlying genomic prediction model.

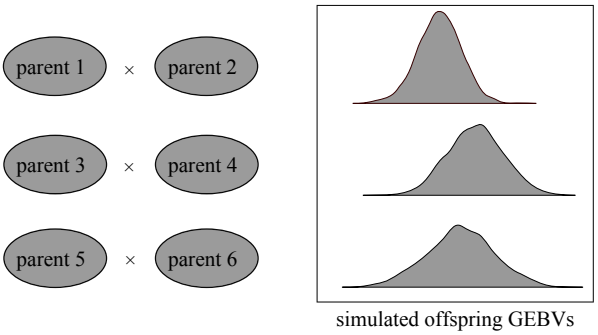

Figure 3. For each hypothetical cross, the distributions of GEBVs from simulated offspring are given to the right, made from the phenotypes and genotypes of the parents. The distributions have different means and variances, which can be used to determine the usefulness of each cross.

6. Uses of genomic prediction in evolutionary genetics

With some creativity, genomic prediction models can be an important tool for the evolutionary community. Genotyping and sequencing costs have fallen dramatically, making it possible to apply genomic prediction in ways that would have been cost-prohibitive a decade ago. Below, we explore several exciting uses for genomic prediction in evolutionary and conservation genetics.

(a) Mate choice: how good are individuals at picking the right mate?

Genomic prediction can predict the expected shape of the breeding value distribution among offspring for any pair of

genotyped individuals. Referred to as the superior progeny value of the cross (box 2) [80,81], this principle has been used extensively in breeding to identify crosses with the highest potential to produce exceptional offspring. For populations with extensive pedigree information (e.g. Soay sheep, scrub jays, fly catchers), one could determine how natural mating patterns correlate with the superior progeny value. This approach could offer an interesting contrast to the empirical approach of estimating the quality of mate choice based on the traits or fitness of observed offspring. In addition, one could use this approach to identify traits where mate choice correlates with the superior progeny value and traits where it does not.

(b) Testing the genetic architecture of traits

While the standard ridge regression methods work well for traits with many loci of small effect, variable selection models can better accommodate a genetic architecture that includes large effect loci [35] but see [36,83]. Such a formal test rejected the infinitesimal model for chill coma recovery, starvation resistance and startle response in *Drosophila melanogaster* [12].

This offers other exciting possibilities to learn about the genetic architecture of traits. For example, one could potentially fit models before and after an 'evolve-and-resequence' experiment to determine if variable selection models fit the data better after imposing specific evolutionary scenarios. This could offer a formal test of genomic architecture shifts that are theoretically expected under different forms of selection [84,85].

As with other uses of genomic prediction, this should be treated with awareness of limitations [86]. For example, it has been shown that when the number of markers is considerably greater than the number of individuals in the training data (i.e. $P \gg N$), the estimated marker effects will be strongly influenced by the prior distribution [36]. This has the potential to muddle the inference of the genetic architecture. However, adding more individuals than markers (i.e. $P < N$) can overcome this issue [36].

(c) Genome-wide association studies

Recent work has advocated for using variants of genomic prediction models for GWAS [8], as genomic prediction fits all SNPs simultaneously with different distributions of effect size. By contrast, most GWAS methods fit one SNP at a time, which can lead to large blocks of significant, linked SNPs in the region of a QTL. Originally, there was scepticism of this practice because if all SNPs are fit simultaneously with only one distribution of marker effect sizes, as in RRBLUP, these blocks of linked SNPs could be individually assigned smaller effects due to shrinkage, which may make it difficult to identify true QTL peaks [8]. However, newer genomic prediction methods show promise in more precisely and clearly identifying QTL [8]. Specifically, BayesRC [61] can be used to group markers by biological function or mutational class, allowing key markers within a large block of linked markers to assume much of the QTL effect size, thereby potentially increasing QTL resolution.

(d) Historical samples

Genomic prediction allows prediction of the breeding value of any sample where DNA is available, but the phenotype was not observed, so long as an appropriate training population exists. This opens up substantial potential for predicting the breeding value of historical samples and might be especially useful for populations under climate change. Simulations suggest that accuracies may be maintained for 10 generations and possibly beyond, when using BayesR and whole-genome sequence [37]. Thus, the training population could potentially be made from contemporary populations. A similar analysis, using polygenic risk scores, was undertaken to predict genomic health and attention-deficit/hyperactivity disorder in Neanderthals [87,88], though, as with all genomic prediction models, they must be applied with regard to environmental impacts on phenotypes and changes in LD between contemporary and past data, epistasis and mutation changes.

(e) Conservation genetics: which individuals should be bred in captive breeding programmes?

With the rise of wildlife infections such as sea-star wasting disease [89], white-nose syndrome in bats [90] and chytrid fungus in amphibians [91], the conservation community could potentially use genomic prediction models to better enable captive breeding strategies. For example, a genomic prediction training dataset could be built from resistant and susceptible individuals. Wild-collected individuals could be genotyped and genomic prediction used to select which individuals would be best to breed for resistance, accelerating the development of pathogen-resistant populations. Indeed, this has already been demonstrated in commercial fisheries settings [92], forestry [16] and in soya beans when searching for germplasm that was resistant to white mould [93]. Notably, in commercial fisheries and forestry settings, strategically reducing the genotyping effort to several thousand markers appears to be sufficient for achieving a decently accurate prediction in some cases [16,92]. In the same vein, genomic prediction could be used to accelerate the response to climate change [94].

7. Conclusion

While there are challenges associated with implementing genomic prediction models in evolutionary systems, these methods have been underused by the evolutionary genetics community. Above we offer five possible use cases for genomic prediction in evolutionary genetics, each of which could open up new opportunities. While genomic prediction arose in agricultural genetics, there are lessons to be taken from the advances in statistical methods derived for breeding purposes. Evolutionary biologists have already applied these models to learn more about selection over time in wild populations and to learn more about the genetic architecture of complex traits (table 1). As evolutionary biologists endeavour to collect ever-larger phenotypic datasets and use ever-improving sequencing approaches to quantify genomic variation, we believe genomic prediction methods will continue to grow as a tool.

Data accessibility. This article has no additional data.

Authors' contributions. S.M., A.L. and L.F. contributed equally to the writing of this manuscript. All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Competing interests. We declare we have no competing interests.

Funding. S.M. was supported by the National Science Foundation IOS-1546727, NSF EDGE Award IOS-1923372 and NIH 1R01GM127872-01 during the writing of this manuscript.

Acknowledgements. We thank Ruth Shaw, Yaniv Brandvain and the Brandvain and McGaugh labs for comments on the manuscript, three anonymous reviewers and Susan Johnson for helpful insights.

References

1. Georges M, Charlier C, Hayes B. 2019 Harnessing genomic information for livestock improvement. *Nat. Rev. Genet.* **20**, 135–156. (doi:10.1038/s41576-018-0082-2)
2. Meuwissen TH, Hayes BJ, Goddard ME. 2001 Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**, 1819–1829. (doi:10.1093/genetics/157.4.1819)
3. Hickey JM, Chiurugwi T, Mackay I, Powell W. 2017 Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. *Nat. Genet.* **49**, 1297–1303. (doi:10.1038/ng.3920)
4. Lorenzana RE, Bernardo R. 2009 Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. *Theor. Appl. Genet.* **120**, 151–161. (doi:10.1007/s00122-009-1166-3)
5. Bernardo R, Yu J. 2007 Prospects for genomewide selection for quantitative traits in maize. *Crop Sci.* **47**, 1082–1090. (doi:10.2135/cropsci2006.11.0690)
6. Veenstra LD, Poland J, Jannink JL, Sorrells ME. 2020 Recurrent genomic selection for wheat grain fructans. *Crop Sci.* **60**, 1499–1512. (doi:10.1002/csc2.20130)
7. Beyene Y *et al.* 2015 Genetic gains in grain yield through genomic selection in eight bi-parental maize populations under drought stress. *Crop Sci.* **55**, 154–163. (doi:10.2135/cropsci2014.07.0460)
8. Goddard M, Kemper K, MacLeod I, Chamberlain A, Hayes B. 2016 Genetics of complex traits: prediction of phenotype, identification of causal polymorphisms and genetic architecture. *Proc. R. Soc. B* **283**, 20160569. (doi:10.1098/rspb.2016.0569)
9. Wray NR, Kemper KE, Hayes BJ, Goddard ME, Visscher PM. 2019 Complex trait prediction from genome data: contrasting EBV in livestock to PRS in humans: genomic prediction. *Genetics* **211**, 1131–1141. (doi:10.1534/genetics.119.301859)
10. Gienapp P, Calus MP, Laine VN, Visser ME. 2019 Genomic selection on breeding time in a wild bird population. *Evol. Lett.* **3**, 142–151. (doi:10.1002/evl3.103)
11. Beaulieu J, Doerksen T, Clément S, MacKay J, Bousquet J. 2014 Accuracy of genomic selection models in a large population of open-pollinated families in white spruce. *Heredity* **113**, 343–352. (doi:10.1038/hdy.2014.36)
12. Edwards SM, Sørensen IF, Sarup P, Mackay TF, Sørensen P. 2016 Genomic prediction for quantitative traits is improved by mapping variants to gene ontology categories in *Drosophila melanogaster*. *Genetics* **203**, 1871–1883. (doi:10.1534/genetics.116.187161)
13. Ober U *et al.* 2012 Using whole-genome sequence data to predict quantitative trait phenotypes in *Drosophila melanogaster*. *PLoS Genet.* **8**, e1002685. (doi:10.1371/journal.pgen.1002685)
14. Ober U, Huang W, Magwire M, Schlather M, Simianer H, Mackay TF. 2015 Accounting for genetic architecture improves sequence based genomic prediction for a *Drosophila* fitness trait. *PLoS ONE* **10**, e0126880. (doi:10.1371/journal.pone.0126880)
15. Hunter DC, Ashraf B, Bérénos C, Johnston SE, Wilson AJ, Pilkington JG, Pemberton JM, Slate J. 2021 Using genomic prediction to detect microevolutionary change of a quantitative trait. bioRxiv 2021.01.06.425564. (doi:10.1101/2021.01.06.42556)
16. Stocks JJ, Metheringham CL, Plumb WJ, Lee SJ, Kelly LJ, Nichols RA, Buggs RJ. 2019 Genomic basis of European ash tree resistance to ash dieback fungus. *Nat. Ecol. Evol.* **3**, 1686–1696. (doi:10.1038/s41559-019-1036-6)
17. Ashraf B, Hunter DC, Bérénos C, Ellis PA, Johnston SE, Pilkington JG, Pemberton JM, Slate J. 2020 Genomic prediction in the wild: a case study in Soay sheep. bioRxiv 2020.07.15.205385. (doi:10.1101/2020.07.15.205385)
18. Bosse M *et al.* 2017 Recent natural selection causes adaptive evolution of an avian polygenic trait. *Science* **358**, 365–368. (doi:10.1126/science.aal3298)
19. Mojica JP, Kelly JK. 2010 Viability selection prior to trait expression is an essential component of natural selection. *Proc. R. Soc. B* **277**, 2945–2950. (doi:10.1098/rspb.2010.0568)
20. Lynch M, Walsh B. 1998 *Genetics and analysis of quantitative traits*. Sunderland, MA: Sinauer Associates, Inc.
21. Falconer D, Mackay T. 1996 *Introduction to quantitative genetics*, 4th edn. Essex, UK: Longman.
22. Burghardt LT, Young ND, Tiffin P. 2017 A guide to genome-wide association mapping in plants. *Curr. Protoc. Plant Biol.* **2**, 22–38. (doi:10.1002/cppb.20041)
23. de los Campos G, Hickey JM, Pong-Wong R, Daetwyler HD, Calus MP. 2013 Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics* **193**, 327–345. (doi:10.1534/genetics.112.143313)
24. Endelman JB. 2011 Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome* **4**, 250–255. (doi:10.3835/plantgenome2011.08.0024)
25. James G, Witten D, Hastie T, Tibshirani R. 2013 *An introduction to statistical learning*. Berlin, Germany: Springer.
26. Roberts DR *et al.* 2017 Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* **40**, 913–929. (doi:10.1111/ecog.02881)
27. Daetwyler HD, Calus MP, Pong-Wong R, de Los Campos G, Hickey JM. 2013 Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. *Genetics* **193**, 347–365. (doi:10.1534/genetics.112.147983)
28. Legarra A, Reverter A. 2018 Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. *Genet. Sel. Evol.* **50**, 1–18. (doi:10.1186/s12711-018-0426-6)
29. Whittaker JC, Thompson R, Denham MC. 2000 Marker-assisted selection using ridge regression. *Genet. Res.* **75**, 249–252. (doi:10.1017/S0016672399004462)
30. Hayes B, Visscher PM, Goddard ME. 2009 Increased accuracy of artificial selection by using the realized relationship matrix. *Genet. Res.* **91**, 47–60. (doi:10.1017/S0016672308009981)
31. Strandén I, Garrick D. 2009 Derivation of equivalent computing algorithms for genomic predictions and reliabilities of animal merit. *J. Dairy Sci.* **92**, 2971–2975. (doi:10.3168/jds.2008-1929)
32. Ogutu JO, Schulz-Streeck T, Piepho H-P. 2012 Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net and their extensions. In *Proc. of the 15th European workshop on QTL mapping and marker assisted selection (QTLMAS), Rennes, France 19–20 May. BMC Proc.* **6**(Suppl. 2), S10. (doi:10.1186/1753-6561-6-S2-S10)
33. Tsai H-Y, Janss LL, Andersen JR, Orabi J, Jensen JD, Jahoor A, Jensen J. 2020 Genomic prediction and GWAS of yield, quality and disease-related traits in spring barley and winter wheat. *Sci. Rep.* **10**, 1–15. (doi:10.1038/s41598-019-56847-4)
34. Gao H, Lund M, Zhang Y, Su G. 2013 Accuracy of genomic prediction using different models and response variables in the Nordic red cattle population. *J. Anim. Breed. Genet.* **130**, 333–340. (doi:10.1111/jbg.12039)
35. Wolc A, Arango J, Settar P, Fulton JE, O'Sullivan NP, Dekkers JC, Fernando R, Garrick DJ. 2016 Mixture models detect large effect QTL better than GBLUP and result in more accurate and persistent predictions. *J. Anim. Sci. Biotechnol.* **7**, 1–6. (doi:10.1186/s40104-016-0066-z)
36. Gianola D. 2013 Priors in whole-genome regression: the Bayesian alphabet returns. *Genetics* **194**, 573–596. (doi:10.1534/genetics.113.151753)
37. MacLeod IM, Hayes BJ, Goddard ME. 2014 The effects of demography and long-term selection on the accuracy of genomic prediction with sequence

data. *Genetics* **198**, 1671–1684. (doi:10.1534/genetics.114.168344)

38. Mehrban H, Lee DH, Moradi MH, IlCho C, Naserkheil M, Ibáñez-Escriche N. 2017 Predictive performance of genomic selection methods for carcass traits in Hanwoo beef cattle: impacts of the genetic architecture. *Genet. Sel. Evol.* **49**, 1. (doi:10.1186/s12711-016-0283-0)
39. Wisser RJ *et al.* 2019 The genomic basis for short-term evolution of environmental adaptation in maize. *Genetics* **213**, 1479–1494. (doi:10.1534/genetics.119.302780)
40. van Heerwaarden J, Hufford MB, Ross-Ibarra J. 2012 Historical genomics of North American maize. *Proc. Natl Acad. Sci. USA* **109**, 12 420–12 425. (doi:10.1073/pnas.1209275109)
41. Hayes B, Bowman PJ, Chamberlain AJ, Goddard ME. 2009 Invited review: Genomic selection in dairy cattle: progress and challenges. *J. Dairy Sci.* **92**, 433–443. (doi:10.3168/jds.2008-1646)
42. Lin Z, Hayes B, Daetwyler H. 2014 Genomic selection in crops, trees and forages: a review. *Crop Pasture Sci.* **65**, 1177–1191. (doi:10.1071/CP13363)
43. Bernardo R. 2016 Bandwagons I, too, have known. *Theor. Appl. Genet.* **129**, 2323–2332. (doi:10.1007/s00122-016-2772-5)
44. Momen M, Mehrgardi AA, Sheikhi A, Kranis A, Tusell L, Morota G, Rosa GJ, Gianola D. 2018 Predictive ability of genome-assisted statistical models under various forms of gene action. *Sci. Rep.* **8**, 1–11. (doi:10.1038/s41598-018-30089-2)
45. Daetwyler HD, Pong-Wong R, Villanueva B, Woolliams JA. 2010 The impact of genetic architecture on genome-wide evaluation methods. *Genetics* **185**, 1021–1031. (doi:10.1534/genetics.110.116855)
46. Goddard M. 2009 Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica* **136**, 245–257. (doi:10.1007/s10709-008-9308-0)
47. Heffner EL, Sorrells ME, Jannink JL. 2009 Genomic selection for crop improvement. *Crop Sci.* **49**, 1–12. (doi:10.2135/cropsci2008.08.0512)
48. Lorenz AJ, Smith KP. 2015 Adding genetically distant individuals to training populations reduces genomic prediction accuracy in barley. *Crop Sci.* **55**, 2657–2667. (doi:10.2135/cropsci2014.12.0827)
49. Clark SA, Hickey JM, Daetwyler HD, van der Werf JH. 2012 The importance of information on relatives for the prediction of genomic breeding values and the implications for the makeup of reference data sets in livestock breeding schemes. *Genet. Sel. Evol.* **44**, 1–9. (doi:10.1186/1297-9686-44-4)
50. Davoudi P, Abdollahi-Arpanahi R, Nejati-Javaremi A. 2018 The impact of QTL allele frequency distribution on the accuracy of genomic prediction. *Archiv Fuer Tierzucht* **61**, 207. (doi:10.5194/aab-61-207-2018)
51. Morgante F, Huang W, Maltecca C, Mackay TF. 2018 Effect of genetic architecture on the prediction accuracy of quantitative traits in samples of unrelated individuals. *Heredity* **120**, 500–514. (doi:10.1038/s41437-017-0043-0)
52. Brard S, Ricard A. 2015 Is the use of formulae a reliable way to predict the accuracy of genomic selection? *J. Anim. Breed. Genet.* **132**, 207–217. (doi:10.1111/jbg.12123)
53. Stapley J, Feulner PGD, Johnston SE, Santure AW, Smadja CM. 2017 Variation in recombination frequency and distribution across eukaryotes: patterns and processes. *Phil. Trans. R. Soc. B* **372**, 20160455. (doi:10.1098/rstb.2016.0455)
54. Fernandes JB, Séguéla-Arnaud M, Larchevêque C, Lloyd AH, Mercier R. 2018 Unleashing meiotic crossovers in hybrid plants. *Proc. Natl Acad. Sci. USA* **115**, 2431–2436. (doi:10.1073/pnas.1713078114)
55. Gebreyesus G, Bovenhuis H, Lund MS, Poulsen NA, Sun D, Buitenhuis B. 2019 Reliability of genomic prediction for milk fatty acid composition by using a multi-population reference and incorporating GWAS results. *Genet. Sel. Evol.* **51**, 1–14. (doi:10.1186/s12711-019-0460-z)
56. Lande R, Thompson R. 1990 Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* **124**, 743–756. (doi:10.1093/genetics/124.3.743)
57. Matukumalli LK *et al.* 2009 Development and characterization of a high density SNP genotyping assay for cattle. *PLoS ONE* **4**, e5350. (doi:10.1371/journal.pone.0005350)
58. Wilson A, Pemberton J, Pilkington J, Clutton-Brock T, Coltman D, Kruuk L. 2007 Quantitative genetics of growth and cryptic evolution of body size in an island population. *Evol. Ecol.* **21**, 337–356. (doi:10.1007/s10682-006-9106-z)
59. van den Berg I, Meuwissen T, MacLeod I, Goddard M. 2019 Predicting the effect of reference population on the accuracy of within, across, and multibreed genomic prediction. *J. Dairy Sci.* **102**, 3155–3174. (doi:10.3168/jds.2018-15231)
60. Erbe M, Hayes B, Matukumalli L, Goswami S, Bowman P, Reich C, Mason B, Goddard M. 2012 Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. *J. Dairy Sci.* **95**, 4114–4129. (doi:10.3168/jds.2011-5019)
61. MacLeod I, Bowman P, Vander Jagt C, Haile-Mariam M, Kemper K, Chamberlain A, Schrooten C, Hayes B, Goddard M. 2016 Exploiting biological priors and sequence variants enhances QTL discovery and genomic prediction of complex traits. *BMC Genomics* **17**, 144. (doi:10.1186/s12864-016-2443-6)
62. Lander ES, Schork NJ. 1994 Genetic dissection of complex traits. *Science* **265**, 2037–2048. (doi:10.1126/science.8091226)
63. Daetwyler HD, Kemper KE, van der Werf JHJ, Hayes BJ. 2012 Components of the accuracy of genomic prediction in a multi-breed sheep population. *J. Anim. Sci.* **90**, 3375–3384. (doi:10.2527/jas.2011-4557)
64. Guo Z, Tucker DM, Basten CJ, Gandhi H, Ersoz E, Guo B, Xu Z, Wang D, Gay G. 2014 The impact of population structure on genomic prediction in stratified populations. *Theor. Appl. Genet.* **127**, 749–762. (doi:10.1007/s00122-013-2255-x)
65. Wientjes YC, Bijma P, Veerkamp RF, Calus MP. 2016 An equation to predict the accuracy of genomic values by combining data from multiple traits, populations, or environments. *Genetics* **202**, 799–823. (doi:10.1534/genetics.115.183269)
66. Schopp P, Müller D, Wientjes YC, Melchinger AE. 2017 Genomic prediction within and across biparental families: means and variances of prediction accuracy and usefulness of deterministic equations. *G3 Genes Genomes Genet.* **7**, 3571–3586. (doi:10.1534/g3.117.300076)
67. Xu S. 2007 An empirical Bayes method for estimating epistatic effects of quantitative trait loci. *Biometrics* **63**, 513–521. (doi:10.1111/j.1541-0420.2006.00711.x)
68. Cai X, Huang A, Xu S. 2011 Fast empirical Bayesian LASSO for multiple quantitative trait locus mapping. *BMC Bioinf.* **12**, 211. (doi:10.1186/1471-2105-12-211)
69. Jiang Y, Reif JC. 2015 Modeling epistasis in genomic selection. *Genetics* **201**, 759–768. (doi:10.1534/genetics.115.177907)
70. Technow F, Schrag TA, Schipprack W, Bauer E, Simianer H, Melchinger AE. 2014 Genome properties and prospects of genomic prediction of hybrid performance in a breeding program of maize. *Genetics* **197**, 1343–1355. (doi:10.1534/genetics.114.165860)
71. González-Diéguez D, Legarra A, Charcosset A, Moreau L, Lehermeier C, Teyssèdre S, Vitezica ZG. 2021 Genomic prediction of hybrid crops allows disentangling dominance and epistasis. *Genetics* **218**, 1. (doi:10.1093/genetics/iyab026)
72. Morota G, Gianola D. 2014 Kernel-based whole-genome prediction of complex traits: a review. *Front. Genet.* **5**, 363. (doi:10.3389/fgene.2014.00363)
73. Wang D, Salah El-Basyoni I, Stephen Baenziger P, Crossa J, Eskridge KM, Dweikat I. 2012 Prediction of genetic values of quantitative traits with epistatic effects in plant breeding populations. *Heredity* **109**, 313–319. (doi:10.1038/hdy.2012.44)
74. Heslot N, Yang HP, Sorrells ME, Jannink JL. 2012 Genomic selection in plant breeding: a comparison of models. *Crop Sci.* **52**, 146–160. (doi:10.2135/cropsci2011.06.0297)
75. Bernardo R. 2014 *Essentials of plant breeding*. Woodbury, MN: Stemma Press.
76. Crossa J *et al.* 2017 Genomic selection in plant breeding: methods, models, and perspectives. *Trends Plant Sci.* **22**, 961–975. (doi:10.1016/j.tplants.2017.08.011)
77. Malosetti M, Bustos-Korts D, Boer MP, van Eeuwijk FA. 2016 Predicting responses in multiple environments: issues in relation to genotype × environment interactions. *Crop Sci.* **56**, 2210–2222. (doi:10.2135/cropsci2015.05.0311)
78. Jarquin D *et al.* 2020 Utility of climatic information via combining ability models to improve genomic prediction for yield within the genomes to fields maize project. *Front. Genet.* **11**, 1819.
79. Campos G. DL, Pérez-Rodríguez P, Bogard M, Gouache D, Crossa J. 2020 A data-driven simulation platform to predict cultivars' performances under

uncertain weather conditions. *Nat. Commun.* **11**, 1–10. (doi:10.1038/s41467-019-13993-7)

80. Zhong S, Jannink J-L. 2007 Using quantitative trait loci results to discriminate among crosses on the basis of their progeny mean and variance. *Genetics* **177**, 567–576. (doi:10.1534/genetics.107.075358)
81. Schnell F, Utz H. 1975 F1 Leistung und Elternwahl in der Zuchtung von Selbstbefruchtern. In *Ber Arbeitstag Arbeitsgem Saatzuchtleiter Bericht über die Arbeitstagung der Vereinigung Österreichischer Pflanzenzüchter*, pp. 234–258. Gumpenstein, Austria.
82. Mohammadi M, Tiede T, Smith KP. 2015 PopVar: a genome-wide procedure for predicting genetic variance and correlated response in biparental breeding populations. *Crop Sci.* **55**, 2068–2077. (doi:10.2135/cropsci2015.01.0030)
83. Karaman E, Cheng H, Firat MZ, Garrick DJ, Fernando RL. 2016 An upper bound for accuracy of prediction using GBLUP. *PLoS ONE* **11**, e0161054. (doi:10.1371/journal.pone.0161054)
84. Yeaman S, Whitlock MC. 2011 The genetic architecture of adaptation under migration-selection balance. *Evolution* **65**, 1897–1911. (doi:10.1111/j.1558-5646.2011.01269.x)
85. Yeaman S. 2015 Local adaptation by alleles of small effect. *Am. Nat.* **186**(S1), S74–S89. (doi:10.1086/682405)
86. Schrauf MF, Martini JW, Simianer H, de Los Campos G, Cantet R, Freudenthal J, Korte A, Munilla S. 2020 Phantom epistasis in genomic selection: on the predictive ability of epistatic models. *G3 Genes Genomes Genet.* **10**, 3137–3145. (doi:10.1534/g3.120.401300)
87. Berens AJ, Cooper TL, Lachance J. 2017 The genomic health of ancient hominins. *Hum. Biol.* **89**, 7–19. (doi:10.13110/humanbiology.89.1.01)
88. Esteller-Cucala P, Maceda I, Børglum AD, Demontis D, Faraone SV, Cormand B, Lao O. 2020 Genomic analysis of the natural history of attention-deficit/hyperactivity disorder using Neanderthal and ancient *Homo sapiens* samples. *Sci. Rep.* **10**, 8622. (doi:10.1038/s41598-020-65322-4)
89. Hewson I *et al.* 2014 Densovirus associated with sea-star wasting disease and mass mortality. *Proc. Natl Acad. Sci. USA* **111**, 17 278–17 283. (doi:10.1073/pnas.1416625111)
90. Blehert DS *et al.* 2009 Bat white-nose syndrome: an emerging fungal pathogen? *Science* **323**, 227–227. (doi:10.1126/science.1163874)
91. Garner TW, Schmidt BR, Martel A, Pasmans F, Muths E, Cunningham AA, Weldon C, Fisher MC, Bosch J. 2016 Mitigating amphibian chytridiomycoses in nature. *Phil. Trans. R. Soc. B* **371**, 20160207. (doi:10.1098/rstb.2016.0207)
92. Houston RD *et al.* 2020 Harnessing genomics to fast-track genetic improvement in aquaculture. *Nat. Rev. Genet.* **21**, 389–409. (doi:10.1038/s41576-020-0227-y)
93. de Azevedo Peixoto L, Moellers TC, Zhang J, Lorenz AJ, Bhering LL, Beavis WD, Singh AK. 2017 Leveraging genomic prediction to scan germplasm collection for crop improvement. *PLoS ONE* **12**, e0179191. (doi:10.1371/journal.pone.0179191)
94. Varshney RK, Singh VK, Kumar A, Powell W, Sorrells ME. 2018 Can genomics deliver climate-change ready crops? *Curr. Opin. Plant Biol.* **45**, 205–211. (doi:10.1016/j.pbi.2018.03.007)