The Journal of Agricultural Science

cambridge.org/ags

Genomic evaluation of threshold traits in different scenarios of threshold number using parametric and non-parametric statistical methods

M. Ghasemi, F. Ghafouri-Kesbi and P. Zamani

Department of Animal Science, Faculty of Agriculture, Bu-Ali Sina University, Hamedan, Iran

Modelling Animal Systems Research Paper

Cite this article: Ghasemi M, Ghafouri-Kesbi F, Zamani P (2023). Genomic evaluation of threshold traits in different scenarios of threshold number using parametric and non-parametric statistical methods. *The Journal of Agricultural Science* **161**, 109–116. https://doi.org/10.1017/S0021859623000072

Received: 9 April 2022
Revised: 5 December 2022
Accepted: 16 January 2023
First published online: 26 January 2023

Key words:
Chromosome; genomic breeding values; heritability; QTL; SNP

Author for correspondence:
F. Ghafouri-Kesbi,
E-mail: f.ghafouri@basu.ac.ir

Abstract

The aim was to study the effect of the threshold number on the accuracy of genomic evaluation of the threshold traits using support vector machine (SVM), genomic best linear unbiased prediction (GBLUP) and Bayesian method B (BayesB). For this purpose, a genome consisting of three chromosomes was simulated for 1000 individuals on which 3000 bi-allelic single nucleotide polymorphism markers were evenly distributed. Genomic breeding values were predicted in different scenarios of threshold number (1–6 thresholds), QTL number (30 and 300 QTLs) and heritability level (0.1, 0.3 and 0.5). By increasing the number of thresholds from 1 to 6 thresholds, especially at higher levels of heritability, the accuracy of genomic evaluation increased; however, the increase in accuracy was not linear so that it was much more noticeable when the number of thresholds increased from 1 to 2 thresholds. In the most studied scenarios, SVM showed a very poor performance compared to other methods. BayesB ranked first regarding prediction accuracy, though in some cases the observed differences with GBLUP was not significant. While increase in heritability increased the accuracy of genomic evaluation, change in the QTL number had a slight effect on the prediction accuracy. According to the results, the SVM is not recommended for genomic evaluation of threshold traits, especially those which have only one threshold and instead, use of GBLUP and BayesB is recommended. For traits with more than one threshold, fortunately we can achieve accuracy similar to continuous traits by applying traditional genomic evaluation methods.

Introduction

Animal and plant breeders are often concerned with the improvement of complex traits. A new approach called 'genome-wide selection' or 'genomic selection' (GS) (Meuwissen *et al.*, 2001), based on genome-wide marker profiling, can accelerate the genetic improvement of such traits. GS means using genomic information to evaluate and select potential candidates. A key feature of this method is that the entire genome is covered by dense markers. When several thousand markers are genotyped throughout the genome, it is assumed that the markers are next to the causal mutations and, therefore, capture and reflect causal effects. All genetic variance is justified by these markers and it is assumed that the markers are in linkage disequilibrium (LD) with quantitative trait loci (QTL) (Goddard and Hayes, 2007; de Roos *et al.*, 2008). Single nucleotide polymorphism markers (SNPs) are the most abundant type of DNA polymorphism in the genome, have lower mutation rates and are easily genotyped, and that is why they are used for GS. The individual effect of each SNP is calculated using both genotypic and phenotypic data with statistical methods and by summing up the effects of all SNPs, the genomic estimated breeding values (GEBVs) of individuals are estimated. Thus, QTL analysis for working out marker–trait associations is not needed (Kumar *et al.*, 2012). Although GS was introduced in 2001, its application delayed until availability of high-density SNP panels for genotyping of animals (Van Tassell *et al.*, 2008). GS has contributed significantly to increase genetic gain for a variety of economically important traits, both in animal (Van Raden 2008; Szyda *et al.*, 2013) and plant species (Kumar *et al.*, 2012; Brito *et al.*, 2017). In GS, increase in genetic gain arises from shorter generation interval, increased intensity of selection and greater precision in the selection of animals for breeding (Klímová *et al.*, 2020).

Many traits of biological and economic importance follow a discontinuous distribution, but their inheritance is not simply Mendelian such as susceptibility to disease with two phenotypic categories of affected and non-affected, degree of dystocia and the number of progenies in a delivery. These traits are termed 'threshold traits'. They are quantitative traits that are discretely expressed in a limited number of phenotypes (usually two), but which are based on an assumed continuous distribution of factors that contribute to the trait (latent variable, liability) (Falconer and Mackay, 1996; Roff *et al.*, 1997). At first, these traits were seemed a bit out of the

© The Author(s), 2023. Published by Cambridge University Press

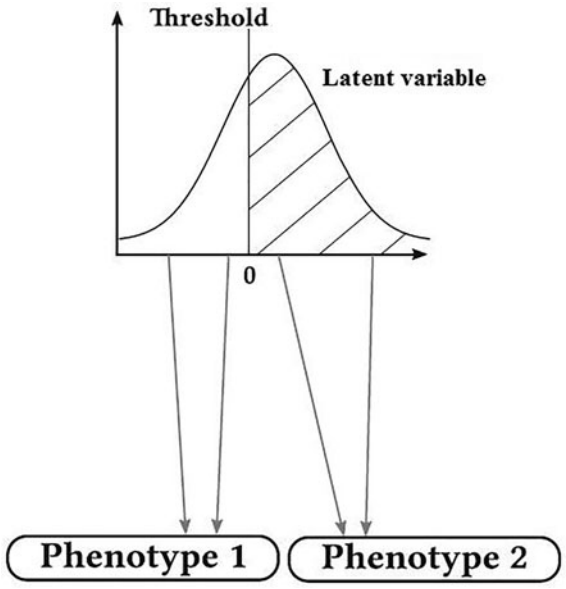


Fig. 1. Graphic representation of a trait with one threshold (de Villemereuil, 2018).

quantitative genetic theory, but when exposed to genetic analysis, it was revealed that their inheritance was similar to that of quantitative traits. These traits have inherently continuous changes, but due to having threshold, apparently they have discontinuous changes (Falconer and Mackay, 1996). Literally, threshold is a level, point or value above which something is true or will take place and below which it is not or will not (Fig. 1). This idea of applying a threshold to a Gaussian hypothetical trait can be traced back to even earlier work by Pearson (1900). When the latent variable (e.g. a biochemical material in the blood with normal distribution in the population) is below this threshold, the individuals show normal phenotype, and when the latent variable overrides the threshold, another phenotypic class is revealed (affected). Therefore, while latent variable has normal distribution, the observed variable follows a discrete distribution with a few phenotypic classes (de Villemereuil, 2018). The basis of threshold characters is a combination of several physiological and developmental processes (Gianola, 1982). Changes in these traits have both genetic and environmental origins and can be measured and studied as a quantitative trait in the routine way. González-Recio and Forni (2011) performed a genomic evaluation for scrotal hernia in pig with one threshold and reported that the accuracy of genomic evaluation was very low. The effect of the number of threshold on the accuracy of genomic evaluation has not been studied so far. Therefore, this study was conducted to study the effect of threshold number on the accuracy of genomic evaluation of discrete traits. In addition, the predictive performance of genomic best linear unbiased prediction (GBLUP) (Van Raden, 2008), support vector machine (SVM) (Boser *et al.*, 1992) and Bayesian method B (BayesB) (Meuwissen *et al.*, 2001) in genomic evaluation of discrete traits was also studied.

Materials and methods

Population and genome

Population and genome were simulated using *hypred* package (Technow, 2013) in R software (R Development Core Team, 2021). The focus of the package is on producing data for genomic applications in applied genetics, namely genomic prediction and selection. In a script, we listed the instructions that *hypred* should use for simulation of genome. Parameters such as number of chromosomes, length of each chromosome, number of SNPs and QTLs per chromosome and distribution of QTL effects were listed in the script. Hypred executed the instructions in the script and built the genome with several internal functions such as hypredGenome (used to define the genome parameters), hyprednewQTL (used to assign QTLs), hypredRecombine (used to simulate meiosis), hypredNewMap (used to modify the genetic map), etc.

A genome consisting of three chromosomes, each one Morgan length, was simulated and 3000 SNPs were uniformly distributed on it. Coding for each genotype with alleles A_1 and A_2 were, respectively, 2 for A_1A_1, 0 for A_2A_2 and 1 for A_1A_2 or A_2A_1. The mutation rate at the marker loci was 2.5×10^{-3} to provide a high probability of polymorphic marker loci. This was 2.5×10^{-5} per locus per generation for each QTL (Meuwissen *et al.*, 2001). Gamma distribution of QTL effects was considered, with shape (β) and scale parameters as 0.4 and 1.66, respectively (Meuwissen *et al.*, 2001).

The baseline population was simulated to be 100 individuals (50 males and 50 females) and randomly mated for 50 generations to create LD between the markers and QTLs. Because two progenies were born from both parents during 50 generations of random mating, the population size was constant throughout the generations of the historical population. In other words, the effective population size (N_e) was 100. The chromosomal compositions of the offspring were obtained by random sampling of the paternal and maternal chromosomes. In generation 51, the population was expanded to 1000 individuals and considered as the reference population. These individuals had both genotypic and phenotypic information. Thereafter, random mating was used for another generation. The animals in the generation 52 had known genotypes but without phenotypic records, which treated as validation population for which genomic breeding values had to be predicted. Thus the genotypic matrix included the genotypic information of 1000 individuals which were genotyped for 3000 SNPs. Parameters used for the simulation of genome are listed in Table 1. In order to convert the normal phenotype into a threshold, the Probit function (Gianola, 1982) was used. Using Probit function the phenotypic category of each individual is determined according to the individual's phenotypic value and the threshold points.

Scenarios under study

The main purpose of this study was to investigate the threshold number on the accuracy of genomic evaluation. Therefore, by

Table 1. Parameters used for simulation program

Genome size	300 cM
Number of chromosomes	3
Number of markers	3000
Distribution of additive QTL effects	Gamma
Number of QTL	30, 300
Effective population size (*Ne*)	100
Heritability	0.1, 0.3, 0.5

https://doi.org/10.1017/S0021859623000072 Published online by Cambridge University Press

applying 1, 2, 3, 4, 5 and 6 thresholds, traits with 2, 3, 4, 5, 6 and 7 phenotypic classes were simulated, respectively. The number of QTLs was defined as a ratio to the number of markers so that in two different scenarios, 1 and 10% of the number of markers were considered as QTL (30 and 300 QTLs, respectively). In addition, three levels of heritability (0.10, 0.30 and 0.50) were also considered for simulation of phenotypes.

Methods of genomic evaluation

Genomic best linear unbiased prediction (GBLUP)

The GBLUP was fitted as follows:

$$\boldsymbol{y} = 1\mu + \boldsymbol{Zg} + \boldsymbol{e}$$

where $\boldsymbol{y}$ is the vector of phenotypic observations, $\boldsymbol{Z}$ is the design matrix associating phenotypic observations to GEBVs, $\boldsymbol{g}$ is the vector of genomic breeding values and assumed that $\boldsymbol{g} \sim N(0, \boldsymbol{G}\delta_g^2)$ where δ_g^2 is the additive genetic variance, and $\boldsymbol{G}$ is the genomic relationship matrix whose elements estimated based on allelic similarity between individuals (Van Raden, 2008). The GBLUP was run using package BGLR in R (de los Campos and Perez Rodriguez, 2018).

Support vector machines (SVM)

One of the kernel methods is the SVM. Kernel methods can be thought of as instance-based learners. Rather than learning some fixed set of parameters corresponding to the features of their inputs, they 'remember' the *i-th* training example. In case of GS the input is genotypic and phenotypic information of animals in the reference population ($\boldsymbol{x_i}$, y_i) and SVM learns for it a corresponding weight w_i. Prediction of unlabelled inputs, i.e. those not in the training set (i.e. phenotypic information of candidate animals ($\hat{y}$)) is treated by the application of a kernel between the unlabelled input x' and each of the training inputs, $\boldsymbol{x_i}$. For quantitative responses, support vector regression (SVR) is used. The SVR uses linear models to implement non-linear regression by mapping the input space (the marker data set) to a feature space of a different dimension (lower in the case of GS) using a non-linear kernel function followed by linear regression in this feature space. In SVR, with input data set $G = \{(\boldsymbol{x_i}, d_i)\}_i^n$ (where $\boldsymbol{x_i}$ is the input vector, d_i is the desired real-valued labelling and n is the number of the input records), $\boldsymbol{x}$ is first mapped into a higher-dimension feature space F via a non-linear mapping Θ, then linear regression is performed in this space. In other words, SVR approximates a function using the following equation (Liu *et al.*, 2006; Hastie *et al.*, 2009):

$$y = f(x) = w\Theta(x) + b$$

The coefficients w and b are estimated by minimizing:

$$R(C) = \frac{1}{2}\|w\|^2 + C\frac{1}{n}\sum_{i=1}^{n} L_\varepsilon(d_i, y_i)*$$

where L_ε (d, y) is the empirical error measured by ε-insensitive loss function

$$L_\varepsilon(d, y) = \begin{cases} |d - y| - \varepsilon, \, if \, |d - y| \geq 0 \\ 0, \, otherwise \end{cases}$$

and the term $1/2||w||^2$ is a regularization term. The constant C is specified by the user, and it determines the trade-off between the empirical risk and the regularization term. The ε is also specified by the user, and it is equivalent to the approximation accuracy of the training data. The estimates of w and b are obtained by transforming Eqn (*) into the primal function:

$$R(w, \varepsilon^{(*)}) = \frac{1}{2}w^2 + C\sum_{i=1}^{n}(\varepsilon_i + \varepsilon_i^*)$$

By introducing Lagrange multipliers, the optimization problem can be transformed into a quadratic programming problem. The solution takes the following form:

$$y = f(x, \alpha_i, \alpha_i^*) = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)K(x, x_i) + b$$

where K is the kernel function $K(x, xi) = \Theta(x)^T \Theta(xi)$. By using a kernel function, we can deal with the problems of arbitrary dimensionality without having to compute the mapping Θ explicitly. Different kernel functions can be selected to map (or transform) input data to feature space. According to Kasnavi *et al.* (2018), we used radial kernel to construct SVM. The package *e1071* (Meyer *et al.*, 2013) was used for SVM analysis.

Bayesian method B (BayesB)

In this model, it is assumed that only part of the loci explains the entire genetic variance, and many loci do not play a role in genetic variance. BayesB can be written as follows:

$$y_i = \mu + \sum_{j=1}^{k} x_{ij}\beta_j\delta_j + e_i$$

where y is the phenotype of the animal i, μ is the mean, k is the number of marker loci, x is the genotype of the marker at the locus j (i_{th} allele) which is encoded as 0, 1 and 2 (number of copies of the SNP allele carried by the i_{th} animal). β_j is the effect of allelic substitution at position j and δ_j which is coded as 0 and 1 indicates the absence (with probability π) or the presence (with probability 1–π) of the locus j in the model.

$$\delta_i = 1 \Rightarrow \beta_j = N(0, \sigma_j^2)$$

$$\delta_i = 0 \Rightarrow \beta_j = 0$$

The main assumption of this method is that many SNPs are located in genomic regions that have no specific QTL association and have no effect on the trait and only a small part of SNPs are in LD with QTLs and therefore have an effect. In general, π represents the expected ratio of SNPs which are in LD with QTLs to the total SNP number. The effects of SNPs will be sampled from the t-distribution, but the variance of the effects will be sampled with probability π from a scaled inverse χ^2 distribution (Meuwissen *et al.*, 2001):

$$\beta_i|\nu_i = N(0, \sigma_j^2) = \beta_i|\sigma_j^2 = (1 - \pi)I_0 + \pi N(0, \sigma_{\beta_i}^2)$$

https://doi.org/10.1017/S0021859623000072 Published online by Cambridge University Press

$$(\beta_i|\pi, \sigma^2_{\beta_i})\begin{cases}\sigma^2_{\beta_i} = 0 \text{ with probability } \pi\\ \sigma^2_{\beta_i} = \mathcal{X}^{-2}(r, s) \text{with probability } 1-\pi\end{cases}$$

To implement BayesB, BGLR package (de los Campos and Perez Rodriguez, 2018) was used. Gibbs sampling algorithm was used to sample conditional posterior distribution of marker effects. Marker effects were inferred using 12 000 sample chains (2000 burning samples and the next 10 000 samples for posterior distribution inferences).

Accuracy of GEBV

The Pearson's correlation between the predicted genomic breeding values and the true genomic breeding values ($r_{p,t}$) was used as an indicator of the accuracy of genomic evaluation. Each scenario which was a combination of heritability level, QTL number and method used was analysed 100 times and the average accuracy of each scenario was presented.

Results

The effect of threshold number on the accuracy of genomic evaluation in different scenarios of QTL number and heritability level is shown in Fig. 2. As observed, by increasing the number of threshold from 1 to 6 thresholds, the accuracy of genomic evaluation increased. For SVM, by increasing the number of threshold from 1 to 6 thresholds, in different scenarios of heritability level (0.1, 0.3 and 0.5), the accuracy of prediction increased by 49, 29 and 24%, respectively. For GBLUP, in similar scenarios of heritability level, by increasing the number of thresholds from 1 to 6 thresholds, the accuracy of prediction increased by 27, 21 and 27%, respectively. Also, 22, 21 and 19% increase in accuracy was observed for BayesB following increase in the number of threshold from 1 to 6 thresholds. The increase in prediction accuracy was more noticeable when the number of threshold increased from 1 to 2 thresholds compared with scenarios in which the number of threshold increased from 2 to 3 and beyond. For example, for SVM and at the heritability level of 0.1, by increasing the number of threshold from 1 to 2 thresholds, the prediction accuracy increased by 47%, while by increasing the number of threshold from 2 to 3 thresholds, the prediction accuracy increased by only 6%.

The average prediction accuracy of the SVM, GBLUP and BayesB in different scenarios of heritability level and QTL number is shown in Fig. 3. On average, in the scenario of heritability = 0.1, the SVM showed a very poor performance, so that its prediction accuracy was significantly lower than the GBLUP and BayesB ($P < 0.05$). By increasing heritability to 0.3 and then to 0.5, the prediction accuracy of all three methods increased; however, even at higher levels of heritability, GBLUP and BayesB kept their distance with SVM. With one threshold, BayesB performed better than the SVM and GBLUP, though its difference with the GBLUP was not significant in most cases ($P > 0.05$) (Figs 2(*a*)–(*c*)). In general, in most studied scenarios, the BayesB and GBLUP had better performance than SVM.

The effect of heritability on the prediction accuracy is shown in Fig. 4. In all methods, the accuracy of genomic evaluation increased with increasing heritability. When the trait had one threshold and was controlled with 30 QTLs, by increasing heritability from 0.1 to 0.5, the prediction accuracy for SVM, GBLUP and BayesB increased by 62, 41 and 45%, respectively. It was 52, 44 and 49% when trait was controlled by 300 QTLs.

In most of the scenarios, following change in the number of QTLs from 30 to 300 QTLs, no significant change in the accuracy of genomic evaluation was observed. In some cases, with increasing QTL number, the accuracy of prediction decreased slightly, and in other cases, a slight increase was observed (Fig. 5).

Discussion

So far, most studies in the field of GS have been conducted on continuous traits and little efforts have been made for genomic evaluation of threshold traits, though many traits that significantly affect profitability belong to the threshold traits category. Gianola (1982) and Gianola and Foulley (1983) founded the mathematical theory for genetic analysis of threshold traits. Due to the fact that threshold traits occur discretely, the use of linear models cannot bring much genetic improvement for these traits. Deljoo-Issa-Lou (2013) reported that evaluation of threshold traits using pedigree-based threshold models cannot be highly reliable and suggested using genomic information to improve these traits. There are no previous reports on the effects of threshold number on the accuracy of genomic evaluation which makes comparison difficult. González-Recio and Forni (2011) simulated a discrete trait with a threshold and predicted genomic breeding values using different parametric and non-parametric methods. The accuracy of genomic evaluation for all methods was low with a maximum value of 0.41, which is in consistent with the results of the present study. However, they did not evaluate traits with threshold number higher than one. As a result, for traits with one threshold such as durability and liability to disease, the accuracy of genomic evaluation would be low. Therefore, methods with maximum prediction accuracy should be used for genomic evaluation of such traits. For traits with more than one threshold such as littler size in sheep, degree of calving difficulty, conformation and type scores, fortunately we can achieve accuracy similar to continuous traits by applying traditional genomic evaluation methods.

A result we noticed was that in some cases (mostly at heritability = 0.1, Fig. 2(*a*)), by increasing threshold number, the accuracy decreased (e.g. comparing accuracy of BayesB in 5 and 6 thresholds scenarios, the accuracy was higher in the 5 thresholds scenario). At low levels of heritability, greater environmental noises affect the power of models to extract small additive genetic effects leading to fluctuation in the estimates of genomic breeding values (Kasnavi *et al.*, 2018). It can result in random changes in the accuracy of GEBVs. In such a situation, higher accuracy at lower number of threshold would be expected.

Most studies have focused on Bayesian models to analyse threshold traits. Wang *et al.* (2012) by analysing threshold traits by Bayesian models reported that the accuracy of the BayesB and BayesC was almost similar and was higher than BayesA. Villanueva *et al.* (2011) reported that genomic evaluation of threshold traits with BayesB significantly increases the accuracy of genomic breeding values compared to linear models. The rate of increase in the accuracy of genomic breeding values obtained by Bayesian model compared to linear models for threshold traits ranged from 4% (at heritability 0.3) to 16% (at heritability 0.1). Baneh *et al.* (2017) also compared Bayesian methods including Ridge regression, BayesA, BayesB, BayesC and BayesL in genomic evaluation of threshold traits. Their results

https://doi.org/10.1017/S0021859623000072 Published online by Cambridge University Press

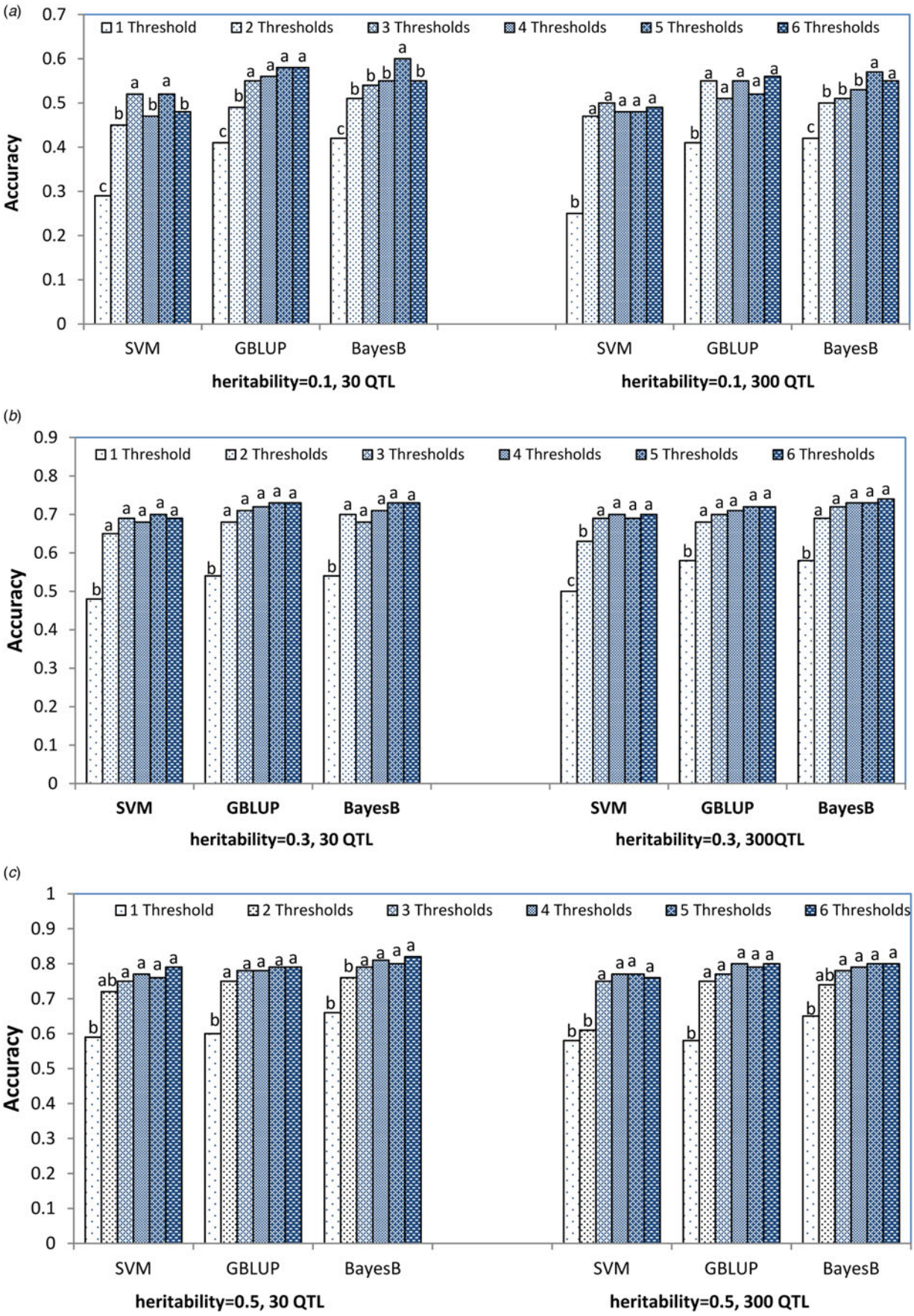


Fig. 2. The effect of threshold number on the accuracy of genomic evaluation.

https://doi.org/10.1017/S0021859623000072 Published online by Cambridge University Press

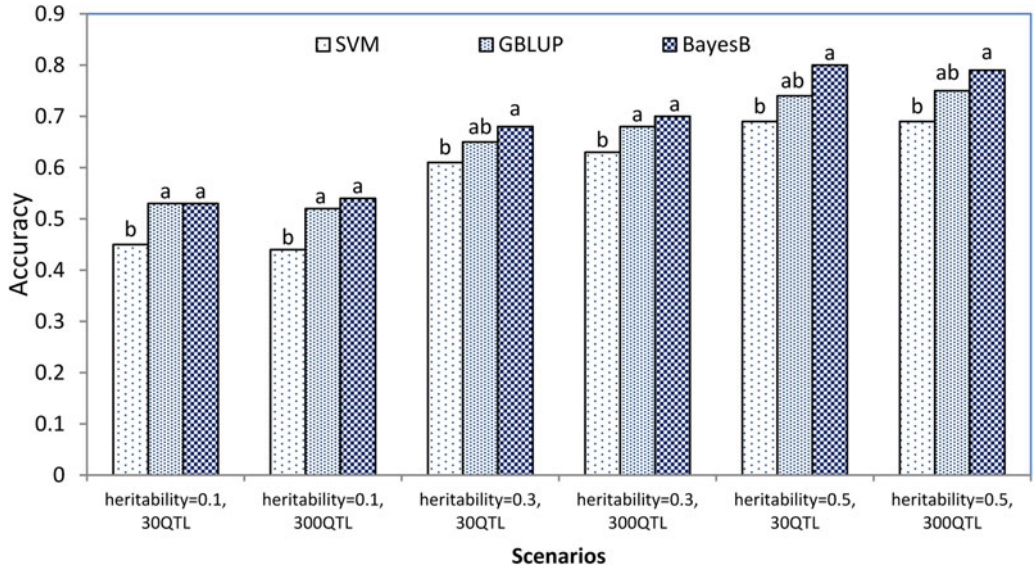


Fig. 3. Comparison of methods in different scenarios of heritability level and QTL number (the accuracy of each method is the average of the results of the 1, 2, 3, 4, 5 and 6 thresholds scenarios).

showed that the accuracy of prediction of all studied methods (due to similarity of computational nature) was close to each other, but in the meantime, BayesA and BaysB were able to estimate SNP effects slightly better (3–7%) than other methods. A Bayesian association mapping for threshold traits using a threshold model was also proposed by Iwata *et al.* (2009) and their approach could reduce both false-positive and false-negative rates in detecting QTL to reasonable levels.

Recent studies have shown the significant effect of heritability on the accuracy of genomic evaluation. For example, Hayes *et al.* (2010) examined the effect of different levels of heritability on the accuracy of genomic evaluation and reported that at heritability levels of 0.1, 0.3, 0.5, 0.7 and 0.9, the accuracy of genomic evaluation was 0.35, 0.5, 0.60, 0.65 and 0.72, respectively. Mohammadi Chamachar *et al.* (2015) reported that accuracy of genomic evaluation of a trait with heritability of 0.05, 0.1 and 0.25 was 0.79, 0.82 and 0.87, respectively. Naderi (2018) reported that for production traits with heritability of 0.3, higher prediction accuracy (0.67) was obtained than a trait with 0.05 heritability (0.41). Zhang *et al.* (2017) by studying the effect of marker density, reference population size and trait heritability reported that among the studied factors, heritability had the greatest impact on the accuracy of genomic evaluation. High heritability shows a higher ratio of genetic variance to phenotypic variance and means a smaller role of environmental noises in the phenotypic variation of the trait. Therefore, the additive genetic effect which is captured by each marker increases. In such a situation, the power of model to extract such greater individual additive effects increases leading to increased accuracy (Ahmadi *et al.*, 2021; Ashoori-Banaei *et al.*, 2021). Goddard (2009) showed that in order to achieve a certain

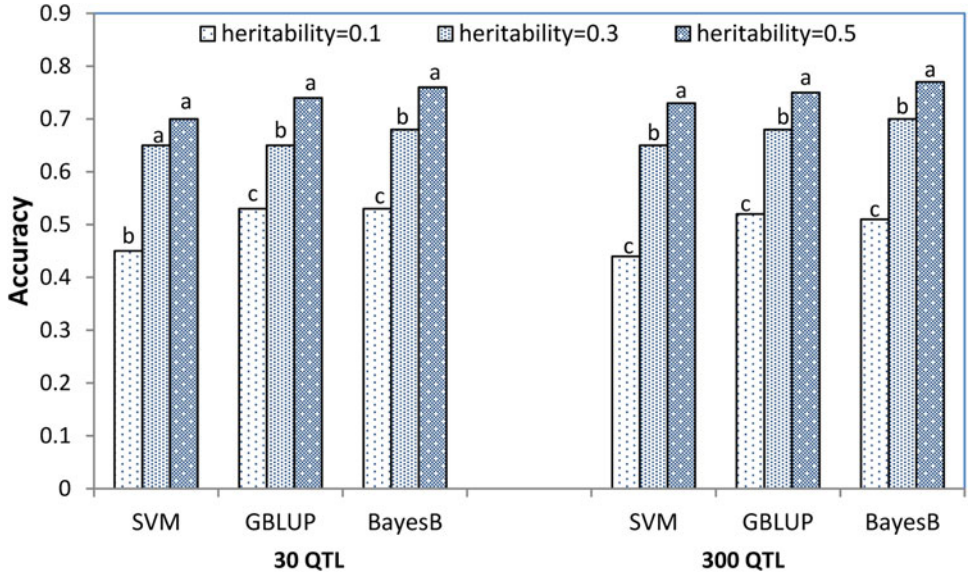


Fig. 4. The effect of heritability on the accuracy of genomic evaluation.

https://doi.org/10.1017/S0021859623000072 Published online by Cambridge University Press

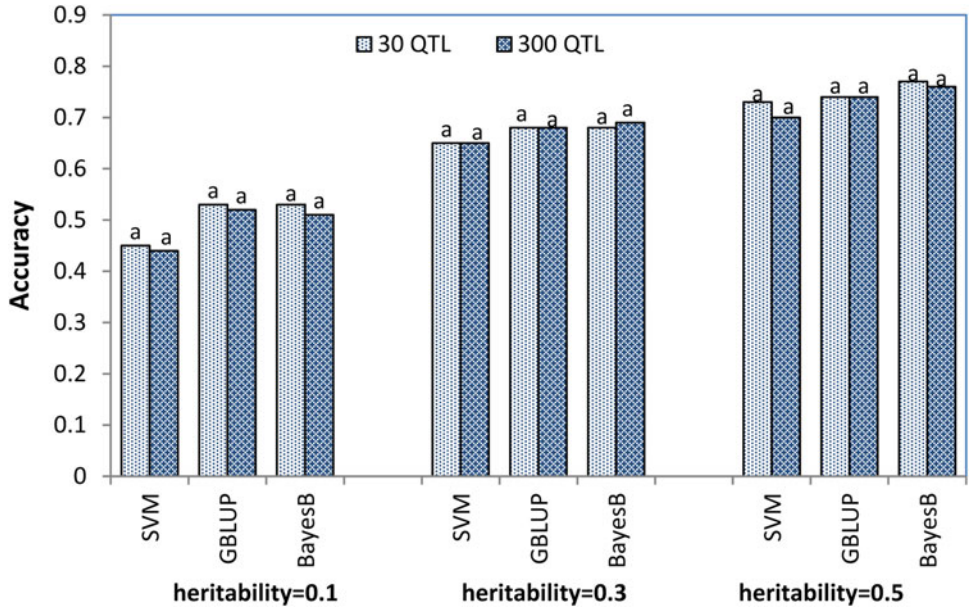


Fig. 5. The effect of QTL number on the accuracy of genomic evaluation.

degree of accuracy, traits with lower heritability require more phenotypic records in the reference group, but this trend was not linear. In other words, the effect of doubling the number of phenotypic records for low heritability traits is greater than the effect of doubling the number of phenotypic records for high heritability traits.

Foroutanifar (2017) reported that the prediction accuracy of BayesA, BayesB and BayesC methods was higher than BayesL and BayesR in the scenarios of small number of QTLs, but with increasing the number of QTLs to 150 and more, this advantage was completely disappeared. Coster *et al.* (2010) showed that when Bayesian regression and Bayesian LASSO fitted to the data, with decreasing QTL number, higher accuracy was obtained but the accuracy of least squares method (Partial Least Squares [PLS]) was not affected by QTL number. Comparing different Bayesian methods for genomic evaluation of threshold traits, Wang *et al.* (2012) observed that Bayesian methods B and C were sensitive to the number of QTLs, and the accuracy of estimates decreased with increasing the number of QTLs from 20 to 500. Also, by comparing Bayesian methods B and GBLUP, Daetwyler *et al.* (2010) observed that the accuracy of GBLUP was constant in different scenarios of number of QTLs, but the accuracy of BayesB was higher in scenarios of small number of QTLs. Assuming a constant total genetic variance, by increasing the number of QTLs, total genetic variance is distributed to a large number of QTLs. In other words, the contribution of each QTL in the total genetic variance decreases and in such a situation, the efficiency of models for estimating such small effects decreases. In addition, as the number of QTLs increases, more markers are needed to capture the effects of all QTLs (Habier *et al.*, 2009). Therefore, increase in the number of QTLs can lead to increase in the accuracy of genomic evaluation if the number of markers increases as well.

In conclusion, genomic evaluation of traits with one threshold had low accuracy and with increasing the number of thresholds, the accuracy of genomic prediction increased. The SVM showed a poor performance in predicting genomic breeding values, especially when the studied trait had only one threshold. GBLUP and specially BayesB showed better performance in genomic evaluation of threshold traits compared to SVM. While increase in the level of heritability increased the prediction accuracy, change in the QTL number had little effect on the accuracy of genomic evaluation.

Author contributions. M. Ghasemi performed the formal analyses. F. Ghafouri-Kesbi designed the study and wrote the original draft. P. Zamani contributed to formal analyses.

Financial support. This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

Conflict of interest. The authors declare no conflicts of interest between authors and other people, institutions or organizations.

Ethical standards. Not applicable.

References

Ahmadi A, Ghafouri-Kesbi F and Zamani P (2021) Assessing the performance of a novel method for genomic selection: rrBLUP-method6. *Journal of Genetics* **1**, 24.

Ashoori-Banaei S, Ghafouri-Kesbi F and Ahmadi A (2021) Comparison of regression tree-based methods in genomic selection. *Journal of Genetics* **100**, 85.

Baneh H, Nejati Javaremi A, Rahimi-Mianji GH and Honarvar M (2017) Genomic evaluation of threshold traits with different genetic architecture using Bayesian approaches. *Research on Animal Production* **15**, 149–154.

Boser B, Guyon I and Vapnik V (1992) An training algorithm for optimal margin classifiers. In *Proceedings of the 5th Annual Workshop on Computational Learning Theory, Pittsburgh, USA*, Vol. **1**, pp. 37–54.

Brito AC, Oliveira SAS and Oliveira EJ (2017) Genome-wide association study for resistance to cassava root rot. *Journal of Agricultural Science* **155**, 1424–1441.

Coster A, Bastiaansen JWM, Calus MPL, van Arendonk JAM and Bovenhuis H (2010) Sensitivity of methods for estimating breeding values using genetic markers to the number of QTL and distribution of QTL variance. *Genetic Selection Evolution* **42**, 9.

https://doi.org/10.1017/S0021859623000072 Published online by Cambridge University Press

Daetwyler H, Hickey JM and Henshal JM (2010) Accuracy of estimated genomic breeding values for wool and meat traits in a multi-breed sheep population. *Animal Production Science* **50**, 1004–1010.

Deljoo-Issa-Lou H (2013) Comparison of linear and threshold models in estimation of genetic and phenotypic parameters of some reproductive traits in Moghani sheep. *Animal Science and Research* **13**, 12–21.

de los Campos G and Perez Rodriguez P (2018) Bayesian generalized linear regression. Available at https://cran.r-project.org/web/packages/BGLR/index.html (Accessed 15 November 2021).

de Roos A, Hayes B, Spelman R and Goddard ME (2008) Linkage disequilibrium and persistence of phase in Holstein–Friesian, Jersey and Angus cattle. *Genetics* **179**, 1503–1512.

de Villemereuil P (2018) Quantitative genetic methods depending on the nature of the phenotypic trait. *Annals of the New York Academy of Sciences* **1422**, 291–47.

Falconer DS and Mackay TFC (1996) *Introduction to Quantitative Genetics*, 4th Edn. Harlow, Essex, UK: Longmans Green.

Foroutanifar S (2017) Effect of QTL number and distribution effects on some statistical methods genomic prediction of a threshold trait. *Iranian Journal of Animal Science Research* **9**, 221–228.

Gianola D (1982) Theory and analysis of threshold characters. *Journal of Animal Science* **54**, 1079–1096.

Gianola D and Foulley JL (1983) Sire evaluation for ordered categorical data with a threshold model. *Genetic Selection Evolution* **15**, 201–224.

Goddard M (2009) Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica* **136**, 245–257.

Goddard M and Hayes B (2007) Genomic selection. *Journal of Animal Breeding and Genetics* **12**, 323–330.

González-Recio O and Forni S (2011) Genome-wide prediction of discrete traits using Bayesian regressions and machine learning. *Genetics Selection Evolution* **43**, 7.

Habier D, Fernando RL and Dekkers JCM (2009) Genomic selection using low-density marker panels. *Genetics* **182**, 343–353.

Hastie TJ, Tibshirani R and Friedman J (2009) *The Elements of Statistical Learning*. New York, USA: Springer. 560 p.

Hayes BJ, Daetwyler HD, Bowman P, Moser G, Tier B, Crump R, Khatkar M, Raadsma HW and Goddard ME (2010) Accuracy of genomic selection: comparing theory and results. In *Proceedings of the 18th Conference of the Association for the Advancement of Animal Breeding and Genetics. Barossa Valley, Australia*.

Iwata H, Ebana K, Fukuoka S, Jannink JL and Hayashi T (2009) Bayesian multilocus association mapping on ordinal and censored traits and its application to the analysis of genetic variation among. *Oryza sativa L. germplasms. Theoretical and Applied Genetics* **118**, 865–880.

Kasnavi SA, Aminafshar M, Shariati MM and Emam Jomeh Kashan N (2018) The effect of kernel selection on genome wide prediction of discrete traits by support vector machine. *Gene Reports* **11**, 279–282.

Klímová A, Kašná E, Machová K, Brzáková M, Přibyl J and Vostrý L (2020) The use of genomic data and imputation methods in dairy cattle breeding. *Czech Journal of Animal Science* **65**, 136–145.

Kumar J, Pratap A, Solanki RK, Gupta DS, Goyal A, Chaturvedi SK, Nadarajan N and Kumar S (2012) Genomic resources for improving food legume crops. *Journal of Agricultural Science* **150**, 289–318.

Liu W, Meng X, Xu O, Flower DR and Li T (2006) Quantitative prediction of mouse class I MHC peptide binding affinity using support vector machine regression (SVR) models. *BMC Bioinformatics* **7**, 182.

Meuwissen THE, Hayes BJ and Goddard ME (2001) Prediction of total genetic value using genome wide dense marker maps. *Genetics* **157**, 1819–1829.

Meyer D, Dimitriadou E, Hornik K, Weingessel A and Leisch K (2013) Misc functions of the department of statistics (e1071), TU Wien. Available at http://cran.r-project.org/web/packages/e1071/index.html (Accessed 15 November 2021).

Mohammadi Chamachar N, Hafezian SH, Honarvar M and Farhadi A (2015) Effects of heritability and number of quantitative trait loci (QTL) on accuracy of genomic estimated breeding value. *Journal of Ruminant Research* **3**, 111–124.

Naderi Y (2018) Evaluation of genomic prediction accuracy in different genomic architectures of quantitative and threshold traits with the imputation of simulated genomic data using random forest method. *Research on Animal Production* **9**, 129–138.

Pearson K (1900) Mathematical contributions to the theory of evolution. VII. On the correlation of characters not quantitatively measurable. *Philosophical Transactions of the Royal Society A* **195**, 1–504.

R Development Core Team (2021) A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at http://www.R-project.org/.

Roff DA, Stirling G and Fairbairn DJ (1997) The evolution of threshold traits: a quantitative genetic analysis of the physiological and life-history correlates of wing dimorphism in the sand cricket. *Evolution* **51**, 1910–1919.

Szyda J, Żukowski K, Kamiński S and Żarnecki A (2013) Testing different single nucleotide polymorphism selection strategies for prediction of genomic breeding values in dairy cattle. *Czech Journal of Animal Science* **58**, 136–145.

Technow F (2013) hypred: Simulation of genomic data in applied genetics. Available at http://cran.r-project.org/web/packages/hypred/index.html (Accessed 20 October 2013).

Van Raden PM (2008) Efficient methods to compute genomic predictions. *Journal of Dairy Science* **91**, 4414–4423.

Van Tassell CP, Smith TPL and Matukumalli LK (2008) SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nature Methods* **5**, 247–252.

Villanueva B, García-Cortés LA, Toro MA, Varona L and Daetwyler HD (2011) Accuracy of genome-wide evaluation for diease resistance in aquaculture breeding programs. *Journal of Animal Science* **89**, 3433–42.

Wang H, Woodward B, Bauck S and Rekaya R (2012) Imputation of missing SNP genotypes using low density panels. *Livestock Science* **146**, 80–83.

Zhang A, Wang H, Beyene Y and Semagn K (2017) Effect of trait heritability, training population size and marker density on genomic prediction accuracy estimation in 22 bi-parental tropical maize populations. *Frontiers in Plant Science* **8**, 1916.

https://doi.org/10.1017/S0021859623000072 Published online by Cambridge University Press