frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 10 February 2025
DOI 10.3389/fgene.2025.1513615

OPEN ACCESS

EDITED BY
Kui Zhang,
Michigan Technological University,
United States

REVIEWED BY
Zitong Li,
Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia
Cheng Gao,
University of Florida, United States

*CORRESPONDENCE
Thinh Tuan Chu,
chu.thinh@qgg.au.dk

RECEIVED 18 October 2024
ACCEPTED 17 January 2025
PUBLISHED 10 February 2025

CITATION
Chu TT and Jensen J (2025) ADAM-multi: software to simulate complex breeding programs for animals and plants with different ploidy levels and generalized genotypic effect models to account for multiple alleles. *Front. Genet.* 16:1513615. doi: 10.3389/fgene.2025.1513615

ADAM-multi: software to simulate complex breeding programs for animals and plants with different ploidy levels and generalized genotypic effect models to account for multiple alleles

Thinh Tuan Chu[1,2]* and Just Jensen[1]

[1]Center for Quantitative Genetics and Genomics, Aarhus University, Aarhus, Denmark, [2]Faculty of Animal Science, Vietnam National University of Agriculture, Hanoi, Vietnam

Stochastic simulation software, ADAM, has been developed for the purpose of breeding optimization in animals and plants, and for validation of statistical models used in genetic evaluations. Just like other common simulation programs, ADAM assumed the bi-allelic state of quantitative trait locus (QTL). While the bi-allelic state of marker loci is due to the common choice of genotyping technology of single nucleotide polymorphism (SNP) chip, the assumption may not hold for the linked QTL. In the version of ADAM-Multi, we employ a novel simulation model capable of simulating additive, dominance, and epistatic genotypic effects for species with different levels of ploidy, providing with a more realistic assumption of multiple allelism for QTL variants. When assuming bi-allelic QTL, our proposed model becomes identical to the model assumption in common simulation programs, and in genetic textbooks. Along with the description of the updated simulation model in ADAM-Multi, this paper shows two small-scale studies that investigate the effects of multi-allelic versus bi-allelic assumptions in simulation and the use of different prediction models in a single-population breeding program for potatoes. We found that genomic models using dense bi-allelic markers could effectively predicted breeding values of individuals in a well-structure population despite the presence of multi-allelic QTL. Additionally, the small-scale study indicated that including non-additive genetic effects in the prediction model for selection did not lead to an improvement in the rate of genetic gains of the breeding program.

KEYWORDS

stochastic breeding program, genotypic model, polyploidy, dominance, epistasis

1 Introduction

Stochastic simulation is a cost-effective and powerful tool to optimize breeding programs with reduced experimental costs. Such a tool unlocks possibilities for investigating alternative breeding schemes, in order to maximizing genetic gains of the breeding program at a given input of resources. Software package, ADAM (Pedersen et al., 2009; Liu et al., 2019), has been developed for the purpose of breeding optimization in pig,

cattle, fish and plants (Bengtsson et al., 2022; Tessema et al., 2020; Zaalberg et al., 2022; Chu et al., 2020). The tool is also very useful in validating statistical genetic prediction models (Romé et al., 2023; Chu et al., 2021) and in studying methods for preserving genetic diversity in breeding programs (Henryon et al., 2015). Over time the software has been further developed and updated with many features including extension to non-additive genetic models (Chu et al., 2024), indirect genetic effects (Chu et al., 2021), categorical traits (Gebreyesus et al., 2020), definition of true inbreeding (Henryon et al., 2019), or extension to plant breeding (Liu et al., 2019).

ADAM (Pedersen et al., 2009) simulates genotypic effects for each allele in a quantitative trait locus (QTL). This model is unique from other software like AlphaSim (Gaynor et al., 2021; Faux et al., 2016), ChromaX (Younis et al., 2023), MoBPS (Pook et al., 2020), SeqBreed (Pérez-Enciso et al., 2020) and XSim (Chen et al., 2022) that use substitution genotypic effects of QTL. However, all of these tools including ADAM assume a bi-allelic state of each segregating QTL. This assumption might come from current, common genetic models (Christensen et al., 2012; VanRaden, 2008; Falconer and Mackay, 1996) that assume substitution effects of one allele to its alternative. While the bi-allelic state of marker loci is due to the common choice of genotyping technology that yield single nucleotide polymorphism (SNP). The assumption of bi-allelic QTL may not true for all QTL. Multiple alleles have been shown in numerous QTL (Biová et al., 2024; Jiang et al., 2020). In addition, multi-allelic models of QTL are more reasonable explanations for different functional genetic effects from unrelated populations (González-Diéguez et al., 2021).

Multi-allelic models have been developed for genomic prediction in diploid species (Álvarez-Castro and Crujeiras, 2019; Álvarez-Castro and Yang, 2011; Yang and Álvarez-Castro, 2008; Da, 2015). Relevant model based on haplotype blocks also have been shown for genomic prediction (Weber et al., 2023). Thérèse Navarro et al. (2022) has developed a package for genome wide association studies (GWAS) of polyploid populations with multi-allelic models, but only the additive genetic effects were included in the model. For the purpose of simulation, however, we are not aware of any studies that have used multi-allelic models with additive, dominance and epistatic genetic effects, or accounting for different levels of ploidy.

In addition to bi-allelic assumptions, earlier version of ADAM (Pedersen et al., 2009; Liu et al., 2019) use was limited to diploid species only. Many economically important species like potato, banana, sugar cane and some fish orders of salmonids and common carps are polyploid. Extension of simulation models to different ploidy levels is necessary for ADAM (Pedersen et al., 2009; Liu et al., 2019) to design complex breeding schemes for these species.

This paper will describe new features of software package, now called ADAM-Multi, for simulating breeding programs for plants and animals. The focus will be on description of genotypic models for simulating traits with genotypic effects of additive, dominance, and epistatic genetics for species with different ploidy levels including extensions to multi-allelic assumptions. The methods implemented are illustrated in two examples that study the effects of different assumptions on number of alleles, ploidy level and different prediction models used in selection.

2 Materials and methods

2.1 Genotypic models for simulation

We aim to simulate genotypic effects that are generalized to multi-allelic QTL with number (n_B) of alleles, and the genome with ploidy level of n_{ploidy}. Assuming a QTL with alleles B_1, B_2,. . ., B_{n_B}, ADAM-Multi uses following model to simulate additive genotypic value (a) at one locus:

$$a = \sum_{i_B}^{n_B} t^a_{i_B} a_{i_B} = t^a_1 a_1 + t^a_2 a_2 + \cdots + t^a_{n_B} a_{n_B} \tag{1}$$

where a is the additive genotypic value of a QTL; a_{i_B} is genotypic additive effect of allele B_{i_B} (or called i_B for short) at the QTL; $t^a_{i_B}$ is the additive covariate for allele i_B that is scaled genotype dosage calculated as in the AlphaSimR software (Gaynor et al., 2021):

$$t^a_{i_B} = \left(t_{i_B} - \frac{n_{ploidy}}{2}\right)\left(\frac{2}{n_{ploidy}}\right) \tag{2}$$

where t_{i_B} is a raw genotype dosage, or number of copies of allele i_B at the locus, n_{ploidy} is the ploidy level of genome. Key notations are defined in Table 1.

Similarly, the genotypic model for dominance value (d) of a QTL at the locus level is:

$$d = \sum_{i_B}^{n_B} t^d_{i_B} d_{i_B} = t^d_1 d_1 + t^d_2 d_2 + \cdots + t^d_{n_B} d_{n_B} \tag{3}$$

where d_{i_B} is dominance genotypic effect of allele i_B; $t^d_{i_B}$ is the dominance covariate for allele i_B that is scaled genotype dosage calculated as in AlphaSimR (Gaynor et al., 2021):

$$t^d_{i_B} = t_{i_B}\left(n_{ploidy} - t_{i_B}\right)\left(\frac{2}{n_{ploidy}}\right)^2 \tag{4}$$

This simulation model assumes digenic dominance for each allele, i.e., each allele has the same dominance effect with all other alleles. Table 2 shows examples of t, t^a and t^d for diploid and tetraploid genome assuming $n_B = 2$ with allele B_1 and B_2.

The simulation model for additive × additive genotypic value $(aa)^{kl}$ of the two-locus epistatic interaction between the pair of loci k and l:

$$(aa)^{kl} = \left\{\begin{bmatrix} t^{a,l}_1 & t^{a,l}_2 & \cdots & t^{a,l}_{n_B}\end{bmatrix} \otimes \begin{bmatrix} t^{a,k}_1 & t^{a,k}_2 & \cdots & t^{a,k}_{n_B}\end{bmatrix}\right\} \cdot \begin{bmatrix} (aa)^{kl}_1 \\ (aa)^{kl}_2 \\ \cdots \\ (aa)^{kl}_{n_B \times n_B}\end{bmatrix} \tag{5}$$

where $[\, t^{a,x}_1 \;\; t^{a,x}_2 \;\; \cdots \;\; t^{a,x}_{n_B} \,]$ is a vector of additive covariates for locus x (k or l) with n_B elements; $\otimes$ denotes the Kronecker product; $\cdot$ denotes the symbol for matrix multiplication; $\begin{bmatrix} (aa)^{kl}_1 \\ (aa)^{kl}_2 \\ \cdots \\ (aa)^{kl}_{n_B \times n_B}\end{bmatrix}$ is vector of additive × additive genotypic effects that have $n_B \times n_B$ elements. In Equation 5, the number of alleles in loci k and l are the same, and equal to n_B. Simulation models in this paper consider a fixed number

TABLE 1 List of key symbols.

Symbol	Definition
Simulation models at one-locus level	
n_B	Defined number of alleles in a QTL. It is also the maximum number of segregating alleles for all QTL.
B_1, B_2, B_{i_B},. . ., B_{n_B}	Allele 1st, 2nd, i_B th, . . . , n_B th of B. Allele i_B is referred to B_{i_B}
t_1, t_2, t_{i_B},. . ., t_{n_B}	A raw genotype dosage, or number of copies of allele B_1, B_2, B_{i_B},. . ., B_{n_B} at the locus
n_{ploidy}	Ploidy level of genome
t_1^a, t_2^a, $t_{i_B}^a$,. . ., $t_{n_B}^a$	Additive effect covariates, or scaled genotype dosages for additive genetic effect corresponding to t_1, t_2, t_{i_B},. . ., t_{n_B} at the locus for allele B_1, B_2, B_{i_B},. . ., B_{n_B}
a_1, a_2, a_{i_B},. . ., a_{n_B}	Additive genotypic effect of allele B_1, B_2, B_{i_B},. . ., B_{n_B} at a locus
t_1^d, t_2^d, $t_{i_B}^d$,. . ., $t_{n_B}^d$	Dominance effect covariates, or scaled genotype dosages for dominance effect corresponding to t_1, t_2, t_{i_B},. . ., t_{n_B} at the locus for allele B_1, B_2, B_{i_B},. . ., B_{n_B}
d_1, d_2, d_{i_B},. . ., d_{n_B}	Dominance genotypic effect of allele B_1, B_2, B_{i_B},. . ., B_{n_B} at the locus
a, d	Additive, dominance genotypic value at the locus level
Simulation models of two-loci epistatic interactions	
kl	The epistatic interaction pair between loci k and l
$t_{i_B}^{a,x}$, $t_{i_B}^{d,x}$	Additive, and dominance covariates for the effect of allele i_B th at locus x. Here, x represents k and l in two-loci epistasis
$(aa)^{kl}$, $(dd)^{kl}$	Additive × additive, and dominance × dominance genotypic values of the epistatic interaction between the pair of loci k and l
$(ad)^{kl}$	Additive-dominance genotypic values of the epistatic interaction that includes both additive × dominance and dominance × additive between the pair of loci k and l
$(aa)_{i_B^k i_B^l}^{kl}$, $(ad)_{i_B^k}^{kl}$, $(da)_{i_B^l}^{kl}$, $(dd)_{i_B^k i_B^l}^{kl}$	Additive × additive, additive × dominance, dominance × additive, dominance × dominance genotypic effects of the epistatic interaction between allele i_B^k of locus k and allele i_B^l of locus l
Simulation models at individual level	
g_i	Total genetic value of individual i
n_{qtl}, n_{ep}	Number of QTL, and number of epistatic interactions
a, **d**	$n_B \times n_{qtl}$ matrices of additive and dominance genotypic effects of n_B alleles for n_{qtl} QTL
$\mathbf{t_i^a}$, $\mathbf{t_i^d}$	$n_{qtl} \times n_B$ matrices of additive and dominance covariates of n_B alleles for n_{qtl} QTL of individual i
(**aa**), (**ad**), (**da**), (**dd**)	Matrices of additive × additive, additive × dominance, dominance × additive, dominance × dominance epistatic effects with a dimension of $(n_B \times n_B)$ rows and n_{ep} columns
$\mathbf{t_i^{aa}}$, $\mathbf{t_i^{ad}}$, $\mathbf{t_i^{da}}$, $\mathbf{t_i^{dd}}$	Matrices of additive × additive, additive × dominance, dominance × additive, dominance × dominance epistatic covariates with a dimension of n_{ep} rows and $(n_B \times n_B)$ columns for individual i
a*, **d***, (**aa**)*, (**ad**)*, (**da**)*, (**dd**)*	Starting (prior) values of **a**, **d,** (**aa**), (**ad**), (**da**) and (**dd**)
σ_A^2, σ_D^2, σ_{AA}^2, σ_{AD}^2, σ_{DD}^2	Functional variances of additive, dominance, additive × additive, additive-dominance, and dominance × dominance for simulation
Prediction models based on population level	
u, **v**, (**uu**), (**uv**), (**vv**)	Vectors of individuals' additive, dominance, and epistatic effects
$\mathbf{G}_u$, $\mathbf{G}_v$, $\mathbf{G}_{uu}$, $\mathbf{G}_{uv}$, $\mathbf{G}_{vv}$	Genomic relationship matrices constructed based on bi-allelic markers
σ_u^2, σ_v^2, σ_{uu}^2, σ_{uv}^2, σ_{vv}^2	Statistical variances of additive, dominance, additive × additive, additive-dominance, and dominance × dominance

n_B for all QTL, even if not all alleles in a QTL are segregating. Genotypic models in case of arbitrary number of alleles for QTL that set the effects of non-segregating alleles to zero can be found in Supplementary Appendix 1.

The simulation model for additive-dominance genotypic value $(ad)^{kl}$ of the epistatic interaction is the sum of additive × dominance and dominance × additive interaction between the pair of loci k and l:

TABLE 2 Conversion from raw genotype dosages (t_1, t_2) to additive (t_1^a, t_2^a) and dominance (t_1^d, t_2^d) covariates when assuming bi-allelic loci.

Genotype	t_1	t_2	t_1^a	t_2^a	t_1^d	t_2^d
Diploid						
B_1B_1	2	0	1	−1	0	0
B_1B_2	1	1	0	0	1	1
B_2B_2	0	2	−1	1	0	0
Tetraploid						
$B_1B_1B_1B_1$	4	0	1	−1	0	0
$B_1B_1B_1B_2$	3	1	$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{3}{4}$	$\frac{3}{4}$
$B_1B_1B_2B_2$	2	2	0	0	1	1
$B_1B_2B_2B_2$	1	3	$-\frac{1}{2}$	$\frac{1}{2}$	$\frac{3}{4}$	$\frac{3}{4}$
$B_2B_2B_2B_2$	0	4	−1	1	0	0

$$(ad)^{kl} = \left\{\left[\, t_1^{d,l} \;\; t_2^{d,l} \;\; \ldots \;\; t_{n_B}^{d,l} \,\right] \otimes \left[\, t_1^{a,k} \;\; t_2^{a,k} \;\; \ldots \;\; t_{n_B}^{a,k} \,\right]\right\} \cdot \begin{bmatrix} (ad)_1^{kl} \\ (ad)_2^{kl} \\ \ldots \\ (ad)_{n_B \times n_B}^{kl} \end{bmatrix} + \left\{\left[\, t_1^{a,l} \;\; t_2^{a,l} \;\; \ldots \;\; t_{n_B}^{a,l} \,\right] \otimes \left[\, t_1^{d,k} \;\; t_2^{d,k} \;\; \ldots \;\; t_{n_B}^{d,k} \,\right]\right\} \cdot \begin{bmatrix} (da)_1^{kl} \\ (da)_2^{kl} \\ \ldots \\ (da)_{n_B \times n_B}^{kl} \end{bmatrix} \tag{6}$$

where $[\, t_1^{d,x} \;\; t_2^{d,x} \;\; \ldots \;\; t_{n_B}^{d,x} \,]$ is a vector of dominance covariates (Equation 4) for locus x with n_B elements; $\begin{bmatrix} (ad)_1^{kl} \\ (ad)_2^{kl} \\ \ldots \\ (ad)_{n_B \times n_B}^{kl} \end{bmatrix}$ and $\begin{bmatrix} (da)_1^{kl} \\ (da)_2^{kl} \\ \ldots \\ (da)_{n_B \times n_B}^{kl} \end{bmatrix}$ are vectors of additive × dominance $((ad)_x^{kl})$, and dominance × additive $((da)_x^{kl})$ genotypic effects that have $n_B \times n_B$ elements. The value $(ad)_x^{kl}$ is different from $(da)_x^{kl}$.

The simulation model for dominance × dominance genotypic value $(dd)^{kl}$ of the epistatic interaction between the pair of loci k and l:

$$(dd)^{kl} = \left\{\left[\, t_1^{d,l} \;\; t_2^{d,l} \;\; \ldots \;\; t_{n_B}^{d,l} \,\right] \otimes \left[\, t_1^{d,k} \;\; t_2^{d,k} \;\; \ldots \;\; t_{n_B}^{d,k} \,\right]\right\} \cdot \begin{bmatrix} (dd)_1^{kl} \\ (dd)_2^{kl} \\ \ldots \\ (dd)_{n_B \times n_B}^{kl} \end{bmatrix} \tag{7}$$

where $\begin{bmatrix} (dd)_1^{kl} \\ (dd)_2^{kl} \\ \ldots \\ (dd)_{n_B \times n_B}^{kl} \end{bmatrix}$ is vector of dominance × dominance genotypic effects that have $n_B \times n_B$ elements.

So far, the simulated genotypic values were presented at the levels of locus and loci pairs. Here, we present the model for simulating genotypic value at the individual level, which is the sum effects of all QTLs and epistatic pair interactions. The model in a matrix form for total genotypic value g_i of individual i is:

$$g_i = \mathrm{tr}\left(\mathbf{t}_\mathbf{i}^\mathbf{a} \cdot \mathbf{a}\right) + \mathrm{tr}\left(\mathbf{t}_\mathbf{i}^\mathbf{d} \cdot \mathbf{d}\right) + \mathrm{tr}\left[\mathbf{t}_\mathbf{i}^\mathbf{aa} \cdot (\mathbf{aa})\right] + \mathrm{tr}\left[\mathbf{t}_\mathbf{i}^\mathbf{dd} \cdot (\mathbf{dd})\right] + \mathrm{tr}\left[\mathbf{t}_\mathbf{i}^\mathbf{ad} \cdot (\mathbf{ad})\right] + \mathrm{tr}\left[\mathbf{t}_\mathbf{i}^\mathbf{da} \cdot (\mathbf{da})\right] \tag{8}$$

where $\mathbf{a}$ is a $n_B \times n_{qtl}$ matrix:

$$\mathbf{a} = \begin{bmatrix} a_1^{j_1} & a_1^{j_2} & \ldots & a_1^{n_{qtl}} \\ a_2^{j_1} & a_2^{j_2} & \ldots & a_2^{n_{qtl}} \\ \ldots & \ldots & \ldots & \ldots \\ a_{n_B}^{j_1} & a_{n_B}^{j_2} & \ldots & a_{n_B}^{n_{qtl}} \end{bmatrix},$$

where n_{qtl} is the number of QTL; $a_{i_B}^{j_{qtl}}$ is the additive genotypic effect of allele i_B at QTL j_{qtl}; $\mathbf{t}_\mathbf{i}^\mathbf{a}$ is a $n_{qtl} \times n_B$ matrix:

$$\mathbf{t}_\mathbf{i}^\mathbf{a} = \begin{bmatrix} t_{1,i}^{a,1} & t_{2,i}^{a,1} & \ldots & t_{n_B,i}^{a,1} \\ t_{1,i}^{a,2} & t_{2,i}^{a,2} & \ldots & t_{n_B,i}^{a,2} \\ \ldots & \ldots & \ldots & \ldots \\ t_{1,i}^{a,n_{qtl}} & t_{2,i}^{a,n_{qtl}} & \ldots & t_{n_B,i}^{a,n_{qtl}} \end{bmatrix},$$

where $t_{i_B,i}^{a,j_{qtl}}$ is the additive covariate of allele i_B at QTL j_{qtl} of individual i. $t_{i_B,i}^{a,j_{qtl}}$ can be calculated based on the genotype of individual i at locus j_{qtl} using Equation 2; tr() is the trace of a matrix. Similarly, $\mathbf{d}$ is a $n_B \times n_{qtl}$ matrix of dominance genotypic effects; $\mathbf{t}_\mathbf{i}^\mathbf{d}$ is a $n_{qtl} \times n_B$ matrix of dominance covariate of individual i. Matrix $(\mathbf{aa})$ has a dimension of $(n_B \times n_B)$ rows and n_{ep} columns:

$$(\mathbf{aa}) = \begin{bmatrix} (aa)_{1,1}^{kl} & (aa)_{2,1}^{kl} & \ldots & (aa)_{n_{ep},1}^{kl} \\ (aa)_{1,2}^{kl} & (aa)_{2,2}^{kl} & \ldots & (aa)_{n_{ep},2}^{kl} \\ \ldots & \ldots & \ldots & \ldots \\ (aa)_{1,(n_B \times n_B)}^{kl} & (aa)_{2,(n_B \times n_B)}^{kl} & \ldots & (aa)_{n_{ep},(n_B \times n_B)}^{kl} \end{bmatrix},$$

where n_{ep} is number of epistatic interactions between pairs of loci; $(aa)_{j_{ep},(i_B^k,i_B^l)}^{kl}$ is the epistatic additive × additive effect at the interaction pair j_{ep} between allele i_B^k of locus k and allele i_B^l of locus l. $\mathbf{t}_\mathbf{i}^\mathbf{aa}$ is a matrix with n_{ep} rows and $(n_B \times n_B)$ columns.Row j_{ep} th of $\mathbf{t}_\mathbf{i}^\mathbf{aa}$ is set up as in Equation 5. For example,

$$\mathbf{t}_{\mathbf{jep,i}}^\mathbf{aa} = \left[\, t_{j_{ep},1,i}^{a,l} \;\; t_{j_{ep},2,i}^{a,l} \;\; \ldots \;\; t_{j_{ep},n_B,i}^{a,l} \,\right] \otimes \left[\, t_{j_{ep},1,i}^{a,k} \;\; t_{j_{ep},2,i}^{a,k} \;\; \ldots \;\; t_{j_{ep},n_B,i}^{a,k} \,\right],$$

where $t_{j_{ep},i_B^x,i}^{a,x}$ is the additive covariate of allele i_B^k of QTL k or allele i_B^l of QTL l at the interaction pair j_{ep} for individual i.

Similarly, $(\mathbf{dd})$ is a $(n_B \times n_B) \times n_{ep}$ matrix of dominance × dominance genotypic effects; $\mathbf{t}_\mathbf{i}^\mathbf{dd}$ is a $n_{ep} \times (n_B \times n_B)$ matrix of dominance × dominance covariates of individual i. The genotypic additive-dominance effects consist of two components: additive × dominance $\mathrm{tr}[\mathbf{t}_\mathbf{i}^\mathbf{ad} \cdot (\mathbf{ad})]$ and dominance × additive $\mathrm{tr}[\mathbf{t}_\mathbf{i}^\mathbf{da} \cdot (\mathbf{da})]$. Similar to $(\mathbf{dd})$, matrices $(\mathbf{ad})$ and $(\mathbf{da})$ have dimension of $(n_B \times n_B) \times n_{ep}$. $\mathbf{t}_\mathbf{i}^\mathbf{ad}$ and $\mathbf{t}_\mathbf{i}^\mathbf{da}$ are matrices with n_{ep} rows and $(n_B \times n_B)$ columns. Row j_{ep} th of $\mathbf{t}_\mathbf{i}^\mathbf{ad}$ is set up as in Equation 6. For example,

$$\mathbf{t}_{\mathbf{jep,i}}^\mathbf{ad} = \left[\, t_{j_{ep},1,i}^{d,l} \;\; t_{j_{ep},2,i}^{d,l} \;\; \ldots \;\; t_{j_{ep},n_B,i}^{d,l} \,\right] \otimes \left[\, t_{j_{ep},1,i}^{a,k} \;\; t_{j_{ep},2,i}^{a,k} \;\; \ldots \;\; t_{j_{ep},n_B,i}^{a,k} \,\right],$$

where $t^{d,x}_{j_{ep},i^x_B,i}$ is the dominance covariate of allele i^k_B of QTL k or allele i^l_B of QTL l at the interaction pair j_{ep} for individual i. Row j_{ep} th of $\mathbf{t}^{\mathbf{da}}_{\mathbf{i}}$ is:

$$\mathbf{t}^{\mathbf{da}}_{\mathbf{j_{ep},i}} = \left[\, t^{a,l}_{j_{ep},1,i} \;\; t^{a,l}_{j_{ep},2,i} \;\; \cdots \;\; t^{a,l}_{j_{ep},n_B,i} \,\right] \otimes \left[\, t^{d,k}_{j_{ep},1,i} \;\; t^{d,k}_{j_{ep},2,i} \;\; \cdots \;\; t^{d,k}_{j_{ep},n_B,i} \,\right].$$

2.2 Stochastic simulation

Details of simulation steps and theoretical principles of ADAM-Multi can be found in previous version by Pedersen et al. (2009) and Liu et al. (2019). These principles are also similar to those in AphaSim (Gaynor et al., 2021). Simulation of genomic models with ADAM first starts with founder haplotypes of a defined genome structure. To create linkage disequilibrium (LD) between QTL and markers, ADAM-Multi can be used in case of multi-allelism. Other packages such as QMSIM (Sargolzaei and Schenkel, 2009) and AlphaSim (Gaynor et al., 2021) do not support multi-allelic models, but they can be used to generate the genome with a specified degree of LD in case of bi-allelic loci. The genotypic effects of alleles in QTLs are sampled, and then centered and scaled to user-defined parameters using the founders' QTL haplotypes (Chu et al., 2024). Steps for generating additive effects of alleles $\mathbf{a}$ ($n_B \times n_{qtl}$ matrix) in ADAM-Multi are:

- Sampling: Each element of matrix $\mathbf{a}^s$ is sampled from a user-defined normal distribution, e.g., mean of zero and additive variance σ^2_A [or N(0, σ^2_A)]. Matrix $\mathbf{a}^s$ with the same dimension as $\mathbf{a}$ contains starting values of additive effects.
- Centering: Based on matrix $\mathbf{a}^s$ and genotypes of a founder population, we can calculate population mean at each QTL locus. Additive effects of alleles within each QTL (each row of $\mathbf{a}^s$) are centered to achieve population mean of zero. For example,: $a^{j_{qtl}\star}_{i_B} = a^{j_{qtl}s}_{i_B} - \mu^{j_{qtl}s}$, where $\mu^{j_{qtl}s}$ is the population mean at locus j_{qtl} given $\mathbf{a}^s$ (before being centered) and genotypes of the founder population; $a^{j_{qtl}\star}_{i_B}$ is the prior value after centering. Matrix $\mathbf{a}^\star$ (same dimension as $\mathbf{a}$) with elements of $a^{j_{qtl}\star}_{i_B}$ is the prior values of additive effects after centering.
- Rescaling: Prior variance $\sigma^2_{A^\star}$ can also be calculated as we know all individuals' genotype in the population and functional effects of QTL ($\mathbf{a}^\star$). Calculation of the variance can be done by different methods including variance by locus, by chromosome, or by individual as in Chu et al. (2024). Additive effects $\mathbf{a}$ are calculated by rescaling prior effects $\mathbf{a}^\star$ to achieve the user-defined variance input σ^2_A for the founder population. For example:

$$\mathbf{a} = \mathbf{a}^\star \times \sqrt{\frac{\sigma^2_A}{\sigma^2_{A^\star}}}$$

The calculated variances of σ^2_A and $\sigma^2_{A^\star}$ in the rescaling step are functional, biological or genotypic parameters, which are different from classical, statistical quantitative parameters. The differences between functional and statistical variances are detailed and explained in Álvarez-Castro and Carlborg (2007), Álvarez-Castro and Yang (2011), Chu et al. (2024), and Vitezica et al. (2017). Functional effects of dominance and epistasis are also generated based on user-defined inputs of functional variances. Steps for generating dominance effects of alleles $\mathbf{d}$ ($n_B \times n_{qtl}$ matrix) are:

- Sampling: A dominance degree $\delta^{d,j_{qtl}}_{i_B}$ for allele i_B of QTL j_{qtl} is sampled from a user-defined normal distribution $N(\mu_\delta, \sigma^2_\delta)$, where μ_δ is the mean of dominance degree, and σ^2_δ is the variance of dominance degree. If a user does not provide this distribution, ADAM-Multi default values are $N(0.19, 0.097)$ as in Wellmann and Bennewitz, (2011). Each element of matrix $\mathbf{d}^\star$ (starting (or prior) values of dominance effects) is generated as:

$$d^{j_{qtl}\star}_{i_B} = \delta^{d,j_{qtl}\star}_{i_B} \left| a^{j_{qtl}}_{i_B} \right|.$$

- Rescaling: This is done similarly to rescaling as for simulating additive effects. Based on matrix $\mathbf{d}^\star$ and QTL genotypes of a founder population, prior variance $\sigma^2_{D^\star}$ can be calculated. Dominance effects $\mathbf{d}$ are the prior effects $\mathbf{d}^\star$ that are rescaled to achieve the user-defined variance inputs of σ^2_D for the founder population.

Steps for sampling additive × additive effects are.

- Sampling: Each element of matrix $(\mathbf{aa})^\star$ [starting (or prior) values of additive × additive effects] is sampled from a user-defined normal distribution N(0, σ^2_{AA}).
- Rescaling: Prior variance $\sigma^2_{AA^\star}$ can be calculated as we know all individuals' genotype in the population and functional effects of QTL $(\mathbf{aa})^\star$. Calculation of the variance can be done by different methods including variance by pair loci, or by individual as in Chu et al. (2024). Additive × additive effects $(\mathbf{aa})$ are obtained by rescaling prior effects $(\mathbf{aa})^\star$ to achieve the desired variance inputs of σ^2_{AA} for the founder population.

Steps for sampling dominance × dominance effects are similar to those for additive × additive effects. Steps for sampling additive-dominance effects are:

- Sampling: Each element of matrices $(\mathbf{ad})^\star$ and $(\mathbf{da})^\star$ [starting (or prior) values of additive × dominance and dominance × additive effects] is sampled from a user-defined normal distribution N(0, σ^2_{AD}).
- Rescaling: Prior variance $\sigma^2_{AD^\star}$ can be calculated as we know all individuals' genotype in the population and functional effects of QTL $(\mathbf{ad})^\star$ and $(\mathbf{da})^\star$. Note that the additive-dominance effects of an individuals' genotype is the sum of additive × dominance and dominance × additive. Effects $(\mathbf{ad})$ and $(\mathbf{da})$ are obtained by rescaling prior effects $(\mathbf{ad})^\star$ and $(\mathbf{da})^\star$ to achieve the desired variance inputs of σ^2_{AD} for the founder population. For example:

$$(\mathbf{ad}) = (\mathbf{ad})^\star \times \sqrt{\frac{\sigma^2_{AD}}{\sigma^2_{AD^\star}}} \text{ and } (\mathbf{da}) = (\mathbf{da})^\star \times \sqrt{\frac{\sigma^2_{AD}}{\sigma^2_{AD^\star}}}.$$

A centering step is not included for dominance and epistatic effects, but the total genetic value g_i is centered to achieve the population mean of zero based on the founder population. For example, the model for simulating individuals' phenotypes is:

$$y_i = \mu + g_i + e_i \tag{9}$$

TABLE 3 Overview of factors investigated in example 1 and 2.

Factor	Levels
Example 1	
Simulation model for individuals' genetic values	Additive genetic effect only
Multi-allelic assumption with percentage of QTL having bi-allelic state (6 levels)	Bi-allele (100%)
	Bi-allele (80%) + tri-allele (20%)
	Bi-allele (50%) + tri-allele (50%)
	Bi-allele (80%) + quad-allele (20%)
	Bi-allele (50%) + quad-allele (50%)
	Bi-allele (20%) + quad-allele (80%)
Genetic effects included in prediction model for selection (1 level)	Additive genetic effect only
Ploidy (2 levels)	Diploids
	Tetraploids
Example 2	
Simulation model for individuals' genetic values	Additive, dominance, additive × additive, additive-dominance, and dominance × dominance effects
Multi-allelic assumption with percentage of QTL having bi-allelic state (2 levels)	Bi-allele (100%)
	Bi-allele (20%) + quad-allele (80%)
Genetic effects included in prediction model for selection (3 levels)	Additive only
	Additive and dominance
	Additive, dominance, epistatic effects
Ploidy (2 levels)	Diploids
	Tetraploids

where g_i is the genetic values of individual i that is constructed as in (Equation 8); μ is the population mean to re-adjust the mean of the founder population to zero.

In ADAM-Multi, model (Equation 9) can be extended to repeated records, inclusion of non-genetic effects, and multiple traits. The functional genetic effects of additive, dominance, and epistasis are independent of allele frequencies, and in the simulation, they are kept constant across generations. Modeling of genetic recombination during meiosis uses bivalent chromosome pairing (Voorrips and Maliepaard, 2012). A breeding scheme is simulated by combining series of actions: mating, reproduction, phenotyping, genotyping, prediction of breeding values and different selection methods. The use of ADAM-Multi is demonstrated in two examples that study the effects of multi-allelic versus bi-allelic assumptions and the use of different prediction models on accuracy of prediction and genetic gains of breeding programs for potato.

2.3 Example 1

The example uses a simulation model with additive effects only for a single-population breeding scheme. The investigated factors include different multi-allelic assumptions (6 levels) and two levels of ploidy (Table 3). In total, there were 6 × 2 = 12 scenarios simulated in example 1.

The simulation model for individuals' phenotype in example 1 was:

$$y_i = \mu + \sum_{i_B}^{n_B} t_{i_B,i}^{a,1} a_{i_B}^{j_1} + \sum_{i_B}^{n_B} t_{i_B,i}^{a,2} a_{i_B}^{j_2} + \cdots + \sum_{i_B}^{n_B} t_{i_B,i}^{a,n_{qtl}} a_{i_B}^{n_{qtl}} + e_i \qquad (10)$$

Simulation of genetic values in (Equation 10) is an extension of (Equation 1) to the sum effects of n_{qtl}, or equivalent to the additive part of (Equation 8). Notations and symbols are the same as in (Equation 8). The environment term e_i was drawn from a normal distribution N(0, σ_e^2) with $\sigma_e^2 = 2$. This example used a simplified simulation model as compared to Equation 9 because we would like to assess accuracy of predicted breeding values in selection. When non-additive genetics are included in the model, the definition of accuracy of predicted breeding values with multi-allelic assumption is unclear in literature.

Figure 1 shows the simulation pipeline for this example. The genome of founders was simulated to form LD between QTL and markers using a Fisher-Wright inheritance model (Fisher, 1930). The LD genome consisted of 12 chromosomes with genetic distances emulating that of potato (Massa et al., 2018). The total genome length was 888.6 cM. The initial genome that had 50k marker loci and 10k QTL with an equal frequency for each allele. A historical population with effective population size of 200 was simulated for 1,000 generations of random mating, a simulated bottleneck, and an

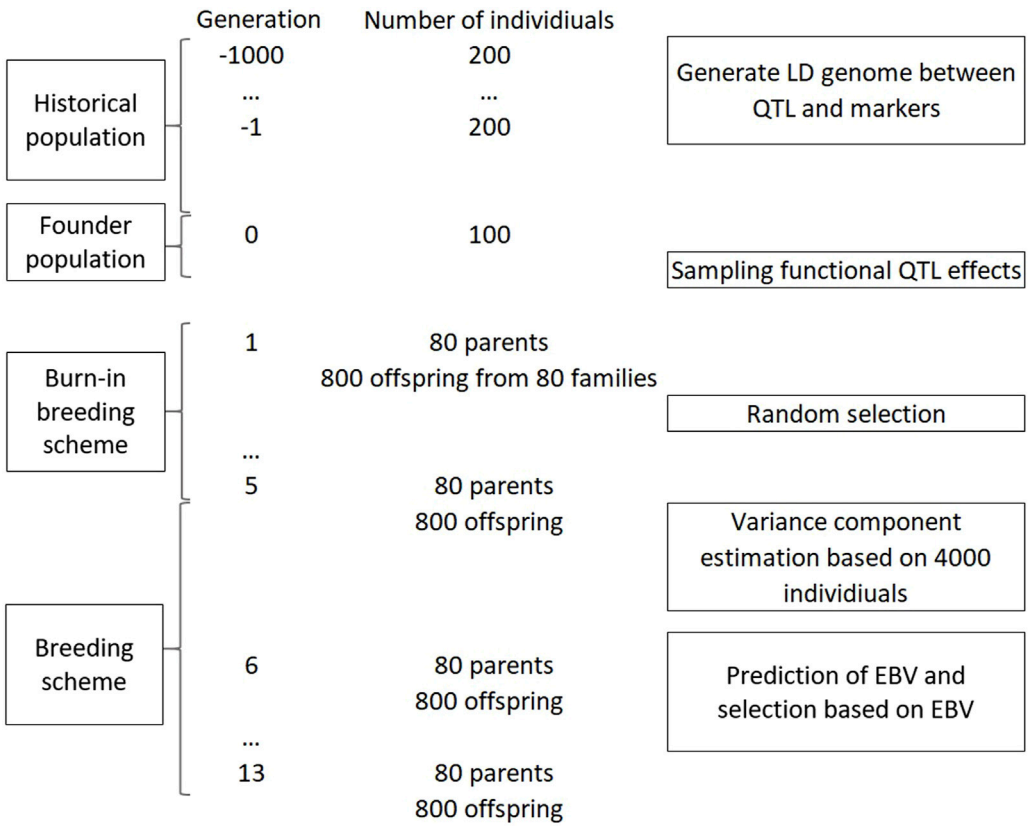

FIGURE 1
An overview of simulation pipeline in example 1 and 2.

inheritance pattern of standard Mendelian principles and bivalent chromosome pairing (Voorrips and Maliepaard, 2012). In example 1, 12 founder populations were created corresponding to 6 levels of multi-allelic assumptions and 2 levels of ploidy (Table 3). The 100-individual founders for each of the populations was created, where the genome of these founders consisted of 2k QTL and 10k bi-allelic markers. The QTL and markers were drawn randomly from loci that were segregating with a minor allele frequency ≥0.05. For markers, bi-allelic state was assumed in all scenarios. The percentage of QTL having segerating bi-, tri-, or quad-alleles was corresponding to the assumption of the scenario. In case of multiple allellisms, for example, quad-allelic assumption for a QTL, each of the four alleles must have a minor allele frequency of at least 0.05. The LD pattern in case of bi-allelic loci for diploid and tetraploid genomes can be found in Supplementary Appendix 2. The functional additive variance of the founder populations was simulated at $\sigma_A^2 = 1.0$.

A simplified breeding scheme was simulated for 13 discrete generations. In a generation, 80 parents were crossed, pseudo-randomly with no self-pollination, to create 80 families. A parent could mate any other parents, but each parent could contribute to only maximum of four crosses. Each family had 10 full-sib offspring, thus in total there were 800 offspring per generation. From generation 1 to 5, the 80 parents were randomly selected from the 800 offspring. In generation 6 to 13, the selection of the offspring to be parents in the following generation was based on predicted breeding values. At generation 6, variance components were estimated when the phenotype data consisted of 4,000 individuals. This estimation of variance components ensures that extra variation due to unknown variance components were taken into account. These variance component estimates were used in the models for prediction of breeding values in the subsequent generations. Selection of 80 new parents from 800 individuals were carried out based on the genetic evaluation after the phenotypes in a generation were obtained.

Model Equation 11 were used in example 1 for variance component estimation and prediction of breeding values is as follows:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z_u u} + \mathbf{e} \tag{11}$$

where $\mathbf{y}$ is the vector of individual phenotypes; $\mathbf{b}$ is a vector of the fixed effects of individuals' generation; $\mathbf{u}$ is the vector of breeding values $\mathbf{u} \sim \mathbf{N}(\mathbf{0}, \mathbf{G}_u\sigma_u^2)$, where the relationship matrix for additive genetics ($\mathbf{G}_u$) was constructed based on bi-allelic markers using method (VanRaden, 2008) for different levels of ploidy assumed in the scenario. The computation of $\mathbf{G}_u$ was carried out using AGHmatrix R package (Amadeu et al., 2023). $\mathbf{X}$ and $\mathbf{Z}_u$ are design matrices relating individuals to fixed effects and additive genetic effects, respectively. Vector $\mathbf{e}$ is an environmental residual term: $\mathbf{e} \sim \mathbf{N}(\mathbf{0}, \mathbf{I}\sigma_e^2)$, where $\mathbf{I}$ is an identity matrix, σ_e^2 is the environmental residual variance.

TABLE 4 Genotypic values of additive (*a*) and dominance (*d*) at one-locus level when assuming bi-allelic QTL.

Genotype	*a*	*d*
Diploid		
B_1B_1	$-(a_2-a_1)$	0
B_1B_2	0	(d_1+d_2)
B_2B_2	(a_2-a_1)	0
Tetraploid		
$B_1B_1B_1B_1$	$-(a_2-a_1)$	0
$B_1B_1B_1B_2$	$-\frac{1}{2}(a_2-a_1)$	$\frac{3}{4}(d_1+d_2)$
$B_1B_1B_2B_2$	0	(d_1+d_2)
$B_1B_2B_2B_2$	$\frac{1}{2}(a_2-a_1)$	$\frac{3}{4}(d_1+d_2)$
$B_2B_2B_2B_2$	(a_2-a_1)	0

Simulation model Equation 10 is QTL-based whereas Equation 11 is marker-based prediction model. Simulation and prediction models are also different in how the covariate of additive effects is calculated. The covariate in Equation 10 is independent of the allele frequency in the population while the covariate as elements of $\mathbf{G}_u$ in Equation 11 uses the frequency in calculation (Chu et al., 2024). The variances estimated from Equation 11 are statistical parameters whereas the simulated variances in Equation 10 are functional variances (Chu et al., 2024). However, as non-additive genetic effects were not simulated in this example, the functional and statistical variances in example 1 would be identical.

Each founder population (at generation 0) were replicated 5 times, i.e., a total of 50 replicates were simulated. The breeding scheme at generations 1–13 was replicated 10 times per founder population replicate. Variance component estimation was carried out using REML in the DMUAI module of the DMU package (Madsen and Jensen, 2013). The prediction of breeding values was performed with the DMU4 module of the DMU package. Population accuracy of the predicted breeding values were assessed for the individuals in generation 6. The accuracy was the correlation between true $\mathbf{u}$ from the simulated values and predicted $\hat{\mathbf{u}}$ in (Equation 11). Rate of genetic gain was assessed as the rate of increase in the genetic mean of population from generation 5 to 13, i.e., rate of genetic gain = $\frac{\overline{u_{13}}-\overline{u_5}}{8}$, where $\overline{u_5}$ and $\overline{u_{13}}$ are the genetic means of population at generation 5 and 13.

2.4 Example 2

The investigated factors in example 2 included different prediction models for selection, multi-allelic levels and two levels of ploidy (Table 3). This example used four haplotype founder populations from example 1 for scenarios that had multi-allelic assumption of bi-allelic (100%), and bi-allelic (20%) + quad-allelic (80%) QTL at two ploidy levels of diploids and tetraploids. There were three prediction models investigated, thus in total $2\times2\times3=12$ scenarios simulated. The simulation pipeline of this example is the same as in example 1 and Figure 1. However, the simulation model in example 2 included additive, dominance, and epistatic genetic effects, as in (Equation 9). For simulating epistasis, we assumed $n_{ep}=1000$, with each QTL present in precisely one pair. The functional variance inputs for the simulation model were set for additive $\sigma^2_A=1.0$, dominance $\sigma^2_D=0.25$, additive × additive $\sigma^2_{AA}=0.25$, additive-dominance $\sigma^2_{AD}=0.25$, dominance × dominance $\sigma^2_{DD}=0.25$, and environmental term $\sigma^2_e=2$. The GBLUP models for predicting breeding values included (Equation 11) and two others as follows:

$$\mathbf{y}=\mathbf{Xb}+\mathbf{Z_u u}+\mathbf{Z_v v}+\mathbf{e} \quad (12)$$

$$\mathbf{y}=\mathrm{Xb}+\mathbf{Z_u u}+\mathbf{Z_v v}+\mathbf{Z}_{uv}(\mathbf{uu})+\mathbf{Z}_{uv}(\mathbf{uv})+\mathbf{Z}_{vv}(\mathbf{vv})+\mathbf{e} \quad (13)$$

where $\mathbf{u}$ is the vector of breeding (additive) values as described in Equation 11; $\mathbf{v}$ is the vector of dominance values $\mathbf{v}\sim\mathbf{N}(\mathbf{0},\mathbf{G}_v\sigma^2_v)$, where σ^2_v is dominance variance, and the relationship matrix for dominance ($\mathbf{G}_v$) is genomic marker-based, which were calculated using AGHmatrix R package (Amadeu et al., 2023). The construction of dominance relationships between individuals used the method by Vitezica et al. (2017) for diploids, and Endelman et al. (2018) for tetraploids. Vector $(\mathbf{uu})$ is additive × additive $(\mathbf{uu})\sim\mathbf{N}(\mathbf{0},\mathbf{G}_{uu}\sigma^2_{uu})$, where σ^2_{uu} is additive × additive variance, and the relationship matrix $\mathbf{G}_{uu}=\mathbf{G}_u\odot\mathbf{G}_u$, where $\mathbf{G}_u$ is genomic-based relationship as in (Equation 11), $\odot$ is the Hadamard product. Similarly, $(\mathbf{uv})$ is the vector of additive-dominance: $(\mathbf{uv})\sim\mathbf{N}(\mathbf{0},\mathbf{G}_{uv}\sigma^2_{uv})$, where σ^2_{uv} is the variance, and $\mathbf{G}_{uv}=\mathbf{G}_u\odot\mathbf{G}_v$. Vector $(\mathbf{vv})$ is dominance × dominance: $(\mathbf{vv})\sim\mathbf{N}(\mathbf{0},\mathbf{G}_{vv}\sigma^2_{vv})$, where σ^2_{vv} is the variance, and $\mathbf{G}_{vv}=\mathbf{G}_v\odot\mathbf{G}_v$. Other notations and symbols are the same as in (Equation 11). Variances σ^2_u, σ^2_v, σ^2_{uu}, σ^2_{uv} and σ^2_{vv} from Equation 13 are statistical parameters, which are different from functional variances σ^2_A, σ^2_D, σ^2_{AA}, σ^2_{AD} and σ^2_{DD} in Equation 9 (Chu et al., 2024).

Similar to example 1, the breeding scheme at generations 1–13 was replicated for 10 times for each of the five founder population replicates. Variance component estimation and prediction of breeding values were carried out by DMUAI and DMU4 module, respectively, of the DMU package (Madsen and Jensen, 2013). Rate of genetic gain was calculated similar as in example 1, except that the total genetic value g_i was used instead of additive values only.

3 Results

Table 4 shows genotypic values of additive (*a*) using (Equation 1) and dominance (*d*) using (Equation 3) for diploids and tetraploids when assuming $n_B=2$ at a locus with allele B_1 and B_2. With bi-allelic QTL locus of diploids, the genotypic effects of B_1B_1, B_1B_2, and B_2B_2 were $-(a_2-a_1)$, 0 and $+(a_2-a_1)$ for *a*, respectively, and 0, (d_1+d_2) and 0 for *d*, respectively. Table 5 shows epistatic genotypic values for diploids due to interactions between loci *k* and *l* when assuming $n_B=2$ at the two loci. The additive-dominance interaction between the pair of loci *k* and *l* is the sum of additive × dominance and dominance × additive interaction effects for the two loci.

Table 6 shows rate of genetic gains and accuracy of predicted breeding values in a simplified breeding scheme for diploids and tetraploids when different levels of multiple allelism were assumed in example 1. Surprisingly, accuracy of predicted breeding values was not statistically different for altered levels of QTL multi-allelic

TABLE 5 Genotypic values of additive × additive (*aa*), additive - dominance (*ad*), and dominance × dominance (*dd*) in two-loci epistatic interactions when assuming bi-allelic QTL.

Genotype	$(aa)^{kl}$	$(ad)^{kl}$	$(dd)^{kl}$
$B_1^k B_1^k B_1^l B_1^l$	$+\left\{(aa)_1^{kl} - (aa)_2^{kl} - (aa)_3^{kl} + (aa)_4^{kl}\right\}$	0	0
$B_1^k B_1^k B_1^l B_2^l$	0	$+\left\{(ad)_1^{kl} - (ad)_2^{kl} + (ad)_3^{kl} - (ad)_4^{kl}\right\}$	0
$B_1^k B_1^k B_2^l B_2^l$	$-\left\{(aa)_1^{kl} - (aa)_2^{kl} - (aa)_3^{kl} + (aa)_4^{kl}\right\}$	0	0
$B_1^k B_2^k B_1^l B_1^l$	0	$+\left\{(da)_1^{kl} - (da)_2^{kl} + (da)_3^{kl} - (da)_4^{kl}\right\}$	0
$B_1^k B_2^k B_1^l B_2^l$	0	0	$(dd)_1^{kl} + (dd)_2^{kl} + (dd)_3^{kl} + (dd)_4^{kl}$
$B_1^k B_2^k B_2^l B_2^l$	0	$-\left\{(da)_1^{kl} - (da)_2^{kl} + (da)_3^{kl} - (da)_4^{kl}\right\}$	0
$B_2^k B_2^k B_1^l B_1^l$	$-\left\{(aa)_1^{kl} - (aa)_2^{kl} - (aa)_3^{kl} + (aa)_4^{kl}\right\}$	0	0
$B_2^k B_2^k B_1^l B_2^l$	0	$-\left\{(ad)_1^{kl} - (ad)_2^{kl} + (ad)_3^{kl} - (ad)_4^{kl}\right\}$	0
$B_2^k B_2^k B_2^l B_2^l$	$+\left\{(aa)_1^{kl} - (aa)_2^{kl} - (aa)_3^{kl} + (aa)_4^{kl}\right\}$	0	0

TABLE 6 Genetic gain and accuracy in simulation model with additive effects only of example 1.

Multi-allelic assumption	Diploids		Tetraploids	
	Genetic gain	Accuracy	Genetic gain	Accuracy
Bi-allele (100%)	1.086	0.842	1.068	0.798
Bi-allele (80%) + tri-allele (20%)	1.097	0.837	1.066	0.796
Bi-allele (50%) + tri-allele (50%)	1.074	0.838	1.104	0.807
Bi-allele (80%) + quad-allele (20%)	1.065	0.842	1.061	0.792
Bi-allele (50%) + quad-allele (50%)	1.077	0.844	1.146	0.794
Bi-allele (20%) + quad-allele (80%)	1.088	0.842	1.118	0.795
Standard deviation over replicates in range (for column)	0.044–0.061	0.03–0.06	0.047–0.068	0.04–0.06

TABLE 7 Genetic gain in simulation model with additive, dominance and epistatic effects of example 2 when different prediction models were used.

Multi-allelic assumption	Prediction model		
	Additive only	Additive and dominance	Additive, dominance and epistasis
Diploids			
Bi-allele (100%)	1.103	1.117	1.114
Bi-allele (20%) + quad-allele (80%)	1.012	0.991	0.994
Tetraploids			
Bi-allele (100%)	1.283	1.275	1.269
Bi-allele (20%) + quad-allele (80%)	1.224	1.238	1.214
Standard deviation over replicates in range (for column)	0.050–0.064	0.048–0.068	0.045–0.065

assumptions, which occurred in the breeding scheme for both diploids and tetraploids. On the contrary, the rate of genetic gains had increasing tendency with increasing levels of multiple allelic QTL in the scheme for tetraploids whereas the genetic gain did not show this tendency in the scheme for diploids. The increasing tendency in genetic gains was most likely due to increased additive variances with higher multiple allelism in tetraploids. The estimated variances for different scenarios in example 1 could be found in Supplementary Table 1, and the true variances at different generations are in Supplementary Table 2.

Table 7 shows the rate of genetic gains in a breeding scheme where different prediction models were used for selection in example 2. In this example, while the simulation model includes additive, dominance, and epistatic interactions between pairs of loci, different prediction models were used for selection. The prediction models with and without non-additive effects did not lead to statistical differences in rate of genetic gains. In about two thirds of replicates, the model could not estimate epistatic effects (Supplementary Table 3). Different multiple allelism did not lead to a significant change in rate of genetic gains. The variance components estimated from different prediction models in example 2 can be found in Supplementary Table 4, and the true variances of total genetic values at different generations are in Supplementary Table 2.

4 Discussion

Our simulation models allow the presence of multi-allelic loci, which is a more realistic assumption for QTL variants. Our simulation model for additive genetic effects is basically the sum of allelic effects. When bi-allelic QTL is assumed, the simulation models applying Equations 1, 3, and 5–7 are identical to the assumption in common genetic textbook, e.g., Falconer and Mackay (1996) and AlphaSimR package (Gaynor et al., 2021). For example, AlphaSimR defines effects of genotype B_1B_1, B_1B_2 and B_2B_2 as $-a$, 0 and $+a$ for additive, respectively. These effects would be corresponding to values in Table 4 if a was defined: $a = (a_2 - a_1)$. Here, $(a_2 - a_1)$ is the substitution effect of allele B_2 for B_1, which is also the definition of a in AlphaSimR and Falconer and Mackay (1996). When bi-allelic QTL is assumed, dominance effect in different ploidy levels is based on a digenic dominance model, which is consistent to Gaynor et al. (2021). Similarly, the epistatic effects in Gaynor et al. (2021) are a special case of our simulation model with bi-allelic assumption.

Interestingly, different levels of multi-allelic assumptions for QTL did not affect accuracy of predicted breeding values based on bi-allelic markers in example 1. This may be due to high density of markers and well-structure population where each clone had many full and half-sibs. Multiple markers can link to a given QTL. Therefore, effects of all alleles at the QTL with multi-allelic state could be estimated using bi-allelic markers. For example, three different bi-allelic marker loci that were closely linked to a QTL could combined to code for up to eight different alleles of the QTL. This could be the reason that regardless of possible multiple alleles in QTL, bi-allelic markers with reasonably high density could predict breeding values in many genomic selection programs (Chu et al., 2019; Samorè and Fontanesi, 2016; Hayes et al., 2013).

In example 1, the differences in rate of genetic gain between different degree of multi-allelisms is primarily due to genetic variances. Although the base population variances are simulated as the same values between two populations, existence of multi-allelism can have higher potential variance, or lower loss of genetic variance under selection. For example, selection led to a removal of a "bad" allele at a QTL in the population. The genetic variance due to QTL would be zero in the bi-allelic case, but might be not in the multi-allelic population. However, maintenance of multi-allelic state might require a bigger effective population size. Otherwise, the multi-allelic state could be lost due to random sampling. This could be the explanation for a higher genetic variance of multi-allelic population in case of tetraploids, but not in diploids in example 1.

While the simulation model including additive, dominance and epistasis was the same for scenarios in example 2, different prediction models (Equations 11–13) were employed for selection. Definition of accuracy of prediction is unclear in literature when different prediction models were used in this case, and particularly when multi-allelic QTL was assumed. Therefore, the rate of genetic gains was used as the main criteria to compare prediction models. Surprisingly, the use of different prediction models did not lead to significant changes in the rate of genetic gains. In other words, the use of correct prediction model for selection, i.e., prediction model and simulation model were more similar, did not improve genetic gains of the breeding scheme. The higher level of multi-allelic assumptions for QTL tended to reduce the genetic gains in example 2, which might be due to lower accuracy of prediction. However, the reduction was not significant.

Nonetheless, examples in this paper are small-scale studies to test ADAM-Multi for multi-allelic features. Many other factors that may affect genetic gain, accuracy of predictions and genetic variances in multi-allelic populations include LD between markers and QTL alleles, population structure, population size, and prediction model. Like other software (Gaynor et al., 2021; Pook et al., 2020; Younis et al., 2023), ADAM-Multi uses functional effects for simulating genotypic values of individuals. Functional effects are independent of allele frequency, thus convenient for studying the consequence of selection in breeding programs (Chu et al., 2024). However, the functional effects and variance parameters cannot be obtained directly by model estimation using real data. Therefore, it is difficult to ensure user-defined parameters for the simulated populations. Just like other software (Gaynor et al., 2021; Pook et al., 2020; Younis et al., 2023), ADAM-Multi is still missing an important feature for a transformation between functional and statistical parameters. More theories are needed to be developed for this transformation, particularly, in case of multi-allelic QTL. Nonetheless, with a more realistic assumption of QTL, ADAM-Multi opens research possibility to study the use of genotyping technology of bi-allelic markers, or the need of new genotyping technology to improve accuracy of selection. Particularly, this assumption of QTL remains very relevant for genomic prediction studies involving multiple breeds and populations. For example, different functional effects of QTL could be assumed in two populations, e.g., González-Diéguez et al. (2021).

This paper presented single-trait models, but the program, ADAM-Multi, can be used for simulating multiple traits with different levels of correlations. For example, the scaling and rescaling procedures in simulation of multiple traits use matrix multiplication, inversion and Cholesky decomposition instead of number multiplication, devision and square root calculations as indicated in this paper. Another note is that the number of alleles n_B could be defined individually for each of QTL (Supplementary Appendix 1). However, this paper assumes a defined n_B for all QTL even when not all alleles in a QTL are segregating. With this assumption, non-segregating alleles' effects can affect the mean, but the mean can be altered with adding constant values to μ just like in Equation 9. On the contrary, non-segregating alleles do not affect the functional or statistical variances as the frequencies of these alleles are fixed. In addition, additive effects are centered in our simulation models, but the dominance effects are not. This assumption of

dominance leads to a positive effect of heterozygous genotypes, as recommended in Wellmann and Bennewitz (2011).

5 Conclusion

This paper presented a simulation model capable of simulating genotypic effects generalized for multiple allelic models and different ploidy levels. This model accommodates genotypic effects of additive, dominance, and epistasis. When assuming bi-allelic QTL, the generalized model becomes identical to the model assumption in common simulation programs, and in genetic textbooks. This model is integrated in our software ADAM-Multi.

In a small-scale study, we have shown that with a reasonable density of bi-allelic markers and a well-structured population, genomic models can effectively predict breeding values despite the presence of multi-allelic QTL. It was also shown that the inclusion of non-additive genetic effects in the prediction model for selection did not lead to a significant improvement in the rate of genetic gains of a breeding program.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Author contributions

TC: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing–original draft, Writing–review and editing. JJ: Funding acquisition, Methodology, Project administration, Resources, Supervision, Validation, Writing–review and editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. The research as a part of project RESPOT was funded by Ministry of Food, Agriculture and Fisheries of Denmark under the Green Development and Demonstrations Program (grant no. 34009-20-1643).

Acknowledgments

We would like to thank Merethe Bagge and the breeders from Danespo for discussions on their breeding program for potatoes, and the presence of multi-allelism in potato population.

Conflict of interest

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

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2025.1513615/full#supplementary-material

References

Álvarez-Castro, J. M., and Carlborg, O. R. (2007). A unified model for functional and statistical epistasis and its application in quantitative trait loci analysis. *Genetics.* 176, 1151–1167. doi:10.1534/genetics.106.067348

Álvarez-Castro, J. M., and Crujeiras, R. M. (2019). Orthogonal decomposition of the genetic variance for epistatic traits under linkage disequilibrium-applications to the analysis of bateson-dobzhansky-müller incompatibilities and sign epistasis. *Front. Genet.* 10, 54. doi:10.3389/fgene.2019.00054

Álvarez-Castro, J. M., and Yang, R.-C. (2011). Multiallelic models of genetic effects and variance decomposition in non-equilibrium populations. *Genetica* 139, 1119–1134. doi:10.1007/s10709-011-9614-9

Amadeu, R. R., Garcia, A. A. F., Munoz, P. R., and Ferrão, L. F. V. (2023). AGHmatrix: genetic relationship matrices in R. *Bioinformatics.* 39, btad445. doi:10.1093/bioinformatics/btad445

Bengtsson, C., Thomasen, J. R., Kargo, M., Bouquet, A., and Slagboom, M. (2022). Emphasis on resilience in dairy cattle breeding: possibilities and consequences. *J. Dairy Sci.* 105, 7588–7599. doi:10.3168/jds.2021-21049

Biová, J., Kaňovská, I., Chan, Y. O., Immadi, M. S., Joshi, T., Bilyeu, K., et al. (2024). Natural and artificial selection of multiple alleles revealed through genomic analyses. *Front. Genet.* 14, 1320652. doi:10.3389/fgene.2023.1320652

Chen, C. J., Garrick, D., Fernando, R., Karaman, E., Stricker, C., Keehan, M., et al. (2022). XSim version 2: simulation of modern breeding programs. *G3 (Bethesda).* 12, jkac032. doi:10.1093/g3journal/jkac032

Christensen, O. F., Madsen, P., Nielsen, B., Ostersen, T., and Su, G. (2012). Single-step methods for genomic evaluation in pigs. *Animal Int. J. Animal Biosci.* 6, 1565–1571. doi:10.1017/s1751731112000742

Chu, T. T., Bastiaansen, J. W. M., Berg, P., Romé, H., Marois, D., Henshall, J., et al. (2019). Use of genomic information to exploit genotype-by-environment interactions for body weight of broiler chicken in bio-secure and production environments. *Genet. Sel. Evol.* 51, 50. doi:10.1186/s12711-019-0493-3

Chu, T. T., Henryon, M., Jensen, J., Ask, B., and Christensen, O. F. (2021). Statistical model and testing designs to increase response to selection with constrained inbreeding

in genomic breeding programs for pigs affected by social genetic effects. *Genet. Sel. Evol.* 53, 1. doi:10.1186/s12711-020-00598-8

Chu, T. T., Kristensen, P., and Jensen, J. (2024). Simulation of functional additive and non-additive genetic effects using statistical estimates from quantitative genetic models. *Hered. (Edinb).* 133, 33–42. doi:10.1038/s41437-024-00690-5

Chu, T. T., Sørensen, A. C., Lund, M. S., Meier, K., Nielsen, T., and Su, G. (2020). Phenotypically selective genotyping realizes more genetic gains in a rainbow trout breeding program in the presence of genotype-by-environment interactions. *Front. Genet.* 866, 866. doi:10.3389/fgene.2020.00866

Da, Y. (2015). Multi-allelic haplotype model based on genetic partition for genomic prediction and variance component estimation using SNP markers. *BMC Genet.* 16, 144. doi:10.1186/s12863-015-0301-1

Endelman, J. B., Carley, C. A. S., Bethke, P. C., Coombs, J. J., Clough, M. E., da Silva, W. L., et al. (2018). Genetic variance partitioning and genome-wide prediction with allele dosage information in autotetraploid potato. *Genetics.* 209, 77–87. doi:10.1534/genetics.118.300685

Falconer, D. S., and Mackay, T. F. (1996). *Introduction to quantitative genetics*. England: Pearson Prentice Hall.

Faux, A. M., Gorjanc, G., Gaynor, R. C., Battagin, M., Edwards, S. M., Wilson, D. L., et al. (2016). AlphaSim: software for breeding program simulation. *Plant Genome* 9. doi:10.3835/plantgenome2016.02.0013

Fisher, R. A. (1930). *The genetical theory of natural selection*. Oxford: Clarendon.

Gaynor, R. C., Gorjanc, G., and Hickey, J. M. (2021). AlphaSimR: an R package for breeding program simulations. *G3 (Bethesda).* 11, jkaa017. doi:10.1093/g3journal/jkaa017

Gebreyesus, G., Sahana, G., Christian Sørensen, A., Lund, M. S., and Su, G. (2020). Novel approach to incorporate information about recessive lethal genes increases the accuracy of genomic prediction for mortality traits. *Hered. (Edinb).* 125, 155–166. doi:10.1038/s41437-020-0329-5

González-Diéguez, D., Legarra, A., Charcosset, A., Moreau, L., Lehermeier, C., Teyssèdre, S., et al. (2021). Genomic prediction of hybrid crops allows disentangling dominance and epistasis. *Genetics* 218, iyab026. doi:10.1093/genetics/iyab026

Hayes, B. J., Lewin, H. A., and Goddard, M. E. (2013). The future of livestock breeding: genomic selection for efficiency, reduced emissions intensity, and adaptation. *Trends Genet.* 29, 206–214. doi:10.1016/j.tig.2012.11.009

Henryon, M., Liu, H., Berg, P., Su, G., Nielsen, H. M., Gebregiwergis, G. T., et al. (2019). Pedigree relationships to control inbreeding in optimum-contribution selection realise more genetic gain than genomic relationships. *Genet. Sel. Evol.* 51, 39. doi:10.1186/s12711-019-0475-5

Henryon, M., Ostersen, T., Ask, B., Sørensen, A. C., and Berg, P. (2015). Most of the long-term genetic gain from optimum-contribution selection can be realised with restrictions imposed during optimisation. *Genet. Sel. Evol.* 47, 21. doi:10.1186/s12711-015-0107-7

Jiang, Y., Chen, S., Wang, X., Liu, M., Iacono, W. G., Hewitt, J. K., et al. (2020). Association analysis and meta-analysis of multi-allelic variants for large-scale sequence data. *Genes* 11, 586. doi:10.3390/genes11050586

Liu, H., Tessema, B. B., Jensen, J., Cericola, F., Andersen, J. R., and Sørensen, A. C. (2019). ADAM-plant: a software for stochastic simulations of plant breeding from molecular to phenotypic level and from simple selection to complex speed breeding programs. *Front. Plant Sci.* 9, 1926. doi:10.3389/fpls.2018.01926

Madsen, P., and Jensen, J. (2013). *DMU: a user's guide. A package for analysing multivariate mixed models*. Available at: https://dmu.ghpc.au.dk/dmu/DMU/(Accessed September 12, 2021).

Massa, A. N., Manrique-Carpintero, N. C., Coombs, J., Haynes, K. G., Bethke, P. C., Brandt, T. L., et al. (2018). Linkage analysis and QTL mapping in a tetraploid russet mapping population of potato. *BMC Genet.* 19, 87. doi:10.1186/s12863-018-0672-1

Pedersen, L., Sørensen, A., Henryon, M., Ansari-Mahyari, S., and Berg, P. (2009). ADAM: a computer program to simulate selective breeding schemes for animals. *Livest. Sci.* 121, 343–344. doi:10.1016/j.livsci.2008.06.028

Pérez-Enciso, M., Ramírez-Ayala, L. C., and Zingaretti, L. M. (2020). SeqBreed: a python tool to evaluate genomic prediction in complex scenarios. *Genet. Sel. Evol. GSE* 52, 7. doi:10.1186/s12711-020-0530-2

Pook, T., Schlather, M., and Simianer, H. (2020). MoBPS - modular breeding program simulator. *G3 Genes. Genomes. Genetics* 10, 1915–1918. doi:10.1534/g3.120.401193

Romé, H., Chu, T. T., Marois, D., Huang, C.-H., Madsen, P., and Jensen, J. (2023). Estimation and consequences of direct-maternal genetic and environmental covariances in models for genetic evaluation in broilers. *Genet. Sel. Evol.* 55, 58. doi:10.1186/s12711-023-00829-8

Samorè, A. B., and Fontanesi, L. (2016). Genomic selection in pigs: state of the art and perspectives. *Ital. J. Anim. Sci.* 15, 211–232. doi:10.1080/1828051X.2016.1172034

Sargolzaei, M., and Schenkel, F. S. (2009). QMSim: a large-scale genome simulator for livestock. *Bioinformatics.* 25, 680–681. doi:10.1093/bioinformatics/btp045

Tessema, B. B., Liu, H., Sørensen, A. C., Andersen, J. R., and Jensen, J. (2020). Strategies using genomic selection to increase genetic gain in breeding programs for wheat. *Front. Genet.* 11, 578123. doi:10.3389/fgene.2020.578123

Thérèse Navarro, A., Tumino, G., Voorrips, R. E., Arens, P., Smulders, M. J. M., van de Weg, E., et al. (2022). Multiallelic models for QTL mapping in diverse polyploid populations. *BMC Bioinforma.* 23, 67. doi:10.1186/s12859-022-04607-z

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91, 4414–4423. doi:10.3168/jds.2007-0980

Vitezica, Z. G., Legarra, A., Toro, M. A., and Varona, L. (2017). Orthogonal estimates of variances for additive, dominance, and epistatic effects in populations. *Dominance, Epistatic Eff. Populations* 206, 1297–1307. doi:10.1534/genetics.116.199406

Voorrips, R. E., and Maliepaard, C. A. (2012). The simulation of meiosis in diploid and tetraploid organisms using various genetic models. *BMC Bioinforma.* 13, 248. doi:10.1186/1471-2105-13-248

Weber, S. E., Frisch, M., Snowdon, R. J., and Voss-Fels, K. P. (2023). Haplotype blocks for genomic prediction: a comparative evaluation in multiple crop datasets. *Front. plant Sci.* 14, 1217589. doi:10.3389/fpls.2023.1217589

Wellmann, R., and Bennewitz, J. (2011). The contribution of dominance to the understanding of quantitative genetic variation. *Genet. Res.* 93, 139–154. doi:10.1017/S0016672310000649

Yang, R.-C., and Álvarez-Castro, J. M. (2008). *Functional and statistical genetic effects with multiple alleles*.

Younis, O. G., Turchetta, M., Ariza Suarez, D., Yates, S., Studer, B., Athanasiadis, I. N., et al. (2023). ChromaX: a fast and scalable breeding program simulator. *Bioinformatics* 39, btad691. doi:10.1093/bioinformatics/btad691

Zaalberg, R. M., Villumsen, T. M., Jensen, J., and Chu, T. T. (2022). Effective selection for lower mortality in organic pigs through selection for total number born and number of dead piglets. *Animals.* 12, 1796. doi:10.3390/ani12141796