Journal of Theoretical Biology 607 (2025) 112110

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/jtbi

A single-locus quantitative genetic model incorporating DNA methylation

L. Ayres *, H. Bovenhuis, M.P.L. Calus

Animal Breeding and Genomics, Wageningen University & Research, 6700 AH, the Netherlands

ARTICLE INFO

Keywords:
Quantitative genetics
Single locus
Breeding value
DNA methylation
Genomic prediction
Epigenetics

ABSTRACT

We describe a single-locus quantitative genetic model that incorporates effects due to DNA methylation. Extending Fisher's decomposition of the genotypic value, we distinguish two quantities to predict an individual's phenotypic or genetic values: the "basic genetic value" and the "expressed genetic value". We show how these quantities relate to the concept of breeding value and derive their corresponding formulas, along with those for phenotypic variance and covariance between relatives. The resulting parameters are influenced by several factors, including the population distribution of DNA methylation levels, the functional relationship between methylation and phenotype, the magnitudes of genetic and methylation effects, and allele frequencies. We show that under the conditions modeled, the presence of DNA methylation does not bias estimated breeding values.

1. Introduction

Epigenetics is concerned with the mechanisms and consequences of chromatin modifications that do not involve changes in DNA sequence. Because the chromatin state is a principal determinant of gene expression, epigenetic mechanisms underlie various biological phenomena. The most well-known forms of chromatin modification are histone modifications, nucleosome remodelling, and DNA methylation (Klug et al., 2019).

DNA methylation is an important cellular regulatory mechanism. It refers to the addition of a methyl group to the nitrogenous base of a nucleoside residue in deoxyribonucleic acid. In vertebrates, the most relevant and well-understood methylation is the reaction that causes cytosine to become 5-methylcytosine (Voet and Voet, 2010; Li and Zhang, 2014). This reaction usually occurs on CpG sequences, and the cytosines on both DNA strands are methylated. The addition of a methyl group to the C5 atom of cytosine's pyrimidine ring changes the local structure of the DNA double helix, affecting its binding affinity with other molecules. As a result, further nuclear processes may follow (e.g., binding of other proteins, histone modifications). More importantly, DNA methylation can regulate gene expression. In mammals, the methylation of cytidine residues in gene promoter regions is associated with repression of mRNA transcription (Li and Zhang, 2014).

Recent advances in the field of genomics have enabled large-scale quantification of DNA methylation. DNA methylation levels are currently measured through assays such as whole-genome bisulfite sequencing (WGBS) (Frommer et al., 1992; Lister and Ecker, 2009), reduced representation bisulfite sequencing (RRBS) (Meissner et al., 2005), bisulfite microarrays (Bibikova et al., 2011), methylated DNA immunoprecipitation (MeDIP) (Weber et al., 2005), and nanopore sequencing (Wallace et al., 2010). In WGBS and RRBS, cytosine methylation is quantified by the proportion of methylated reads at specific loci, thus ranging from zero (when no reads are methylated) to one (all reads are methylated). Intermediate values indicate partial methylation, resulting from a mixture of methylated and unmethylated DNA molecules or strands.

Methods for genomic prediction using multiple loci and large-scale omics data, including methylation, have been proposed by employing mixed model methodology (Guo et al., 2016; Morgante et al., 2020; Christensen et al., 2021), reproducing kernel Hilbert spaces regression (Hu et al., 2015), deep learning (Chai et al., 2021), and neural networks (Zhao et al., 2022). Single-locus models are valuable for developing concepts, discerning parameters that describe DNA methylation and the phenotype, quantifying sources of observed variation, and informing improvements to predictive methods. Indeed, some authors have explored such models in the context of genomic imprinting (Spencer, 2002; de Koning et al., 2002; Santure and Spencer, 2006; Spencer, 2009; Álvarez-Castro, 2014; O'Brien and Wolf, 2019), whose molecular basis is linked to DNA methylation. Unlike in genomic imprinting models, we assume maternal and paternal alleles are interchangeable and disregard parent-of-origin effects.

We aim to build a foundational one-locus model to serve as a framework for incorporating DNA methylation into genomic prediction. This approach provides a critical step towards genome-wide models capable of incorporating both genetic and epigenetic contributions to trait variation.

* Corresponding author.
E-mail address: leviayres@icloud.com (L. Ayres).

https://doi.org/10.1016/j.jtbi.2025.112110
Received 28 January 2025; Accepted 31 March 2025
Available online 4 April 2025

2. The model

2.1. Setting

In a population of diploid, sexually reproducing organisms that mate at random, we consider a quantitative trait and one autosomal locus with two alleles. Each individual has its phenotype measured and its genotype ascertained. In addition, each individual has its methylation level measured at a given age, at a specified site or region of the genome.

We refer to the two alleles as A_1 and A_2. Because the locus is biallelic, there are only three possible genotypes: homozygous for one of the alleles (A_1A_1), heterozygous (A_1A_2) or homozygous for the other allele (A_2A_2). If N denotes the number of copies of A_2, the coding $N \in \{0, 1, 2\}$ captures all genotypes. Let p be the frequency of allele A_2 in the population, so that $\mathbb{E}N = 2p$. We shall refer to the methylation level as the "methylotype" of the individual, in analogy with the term genotype. The methylotype, denoted by M, can be defined as the proportion of methylated reads to facilitate comparisons between individuals, although not a strict requirement of the model we present. We shall assume that the distribution of methylation levels in the population does not depend on the genotype, meaning $\mathrm{Cov}(N, M) = 0$. Further, we assume that the methylation levels of individuals are mutually independent.

2.2. Fisher's decomposition

In Fisher's classic decomposition, the phenotypic measurements Y are treated as a continuous random variable and are regressed on the genotypes N. A simple linear regression model is fitted to the data:

$$Y = \mu + \alpha N + E, \tag{1}$$

where E is a residual effect and μ, α are regression coefficients. The least-squares solution for α is

$$\alpha = \frac{\mathrm{Cov}(N, Y)}{\mathrm{Var}(N)}. \tag{2}$$

In quantitative genetic theory, α is called the *average effect of allelic substitution* (Fisher, 1941).

Under model (1), the conditional expectation minus the mean of the phenotype

$$\mathbb{E}(Y|N = n) - \mathbb{E}Y = \alpha(n - 2p) \tag{3}$$

is called the *additive genetic value* of the genotype that has n copies of the A_2 allele. It can be shown that the additive genetic value of a genotype is twice the expected deviation of its offspring's phenotype from the population mean. For this reason, it is also called the *breeding value* of a genotype (Lynch and Walsh, 1998).

2.3. Incorporating methylation

In Fisher's decomposition, phenotypic values are regressed on the gene content N, that is, the number of copies of an allele (Álvarez-Castro, 2023). We now extend the model to include methylation level—the proportion of DNA methylation affecting the expression of the locus—and an interaction term to account for heterogeneous methylation effects:

$$Y = \mu + \alpha N + \beta M + \gamma NM + E, \tag{4}$$

where Y is the phenotype, N is the number of A_2 alleles in the genotype, M is the methylotype, E is a residual effect, and μ, α, β, γ are regression coefficients.

Similar to expression (3) for the additive genetic value, we may consider a corresponding conditional expectation for the model. For this, we also condition the phenotype Y on the observed methylation level m:

$$\mathbb{E}(Y \mid N = n, M = m) - \mathbb{E}Y. \tag{5}$$

2.4. Expressed genetic value

This quantity, given by expression (5), represents the predicted phenotype of the individual with genotype n and at the observed level of methylation m, conveyed as a deviation from the average phenotype. We will call it the *expressed genetic value* of the genotype-methylotype combination because it reflects the gene expression at a particular time in a certain tissue of the individual. In terms of model (4), we obtain

$$\mathbb{E}(Y \mid N = n, M = m) - \mathbb{E}Y = \alpha(n - 2p) + \beta\left(m - \bar{M}\right) + \gamma\left(nm - 2\bar{M}p\right), \tag{6}$$

where $\bar{M} = \mathbb{E}M$. When mating is random, it can be shown that, unlike the additive genetic value, the expressed genetic value of a genotype is not equal to twice the expected deviation of its offspring's mean phenotype from the population mean (see Table B.1 in Appendix B); thus we call it a genetic value rather than a breeding value. The expectation of the expressed genetic values is equal to zero.

2.5. Basic genetic value

Instead of conditioning the expectation on the observed methylation level m, we may condition it on the methylation value of zero, that is

$$\mathbb{E}(Y \mid N = n, M = 0) - \mathbb{E}Y = \alpha(n - 2p) - \beta\bar{M} - 2\gamma\bar{M}p. \tag{7}$$

This quantity represents the predicted phenotype of the individual if the locus were unmethylated (conveyed as a deviation from the average phenotype). Given that methylation states are dynamic, an unmethylated state provides a baseline for comparison of individuals, as this is when genes can be fully expressed. We refer to this quantity as the *basic genetic value* of the genotype. The expected basic genetic value of the progeny of an individual with a given basic genetic value is not one-half its basic genetic value (see Table B.2 in Appendix B). The mean basic genetic value in a population equals $-\bar{M}(\beta + 2\gamma p)$.

2.6. Breeding value

The traditional breeding value is given by $\mathbb{E}(Y \mid N = n) - \mathbb{E}Y$. To find $\mathbb{E}(Y \mid N = n)$, we apply the law of total probability for expectations, so that the breeding value is equal to

$$\mathbb{E}_M[\mathbb{E}(Y \mid N = n, M)] - \mathbb{E}Y = (n - 2p)\left(\alpha + \gamma\bar{M}\right). \tag{8}$$

Under random mating, the expected phenotypic deviation of the progeny of an individual is one-half of the breeding value of this individual (see Table B.3 in Appendix B). The breeding value in our model has an expression that coincides with the quantity defined below:

$$\mathbb{E}\left(Y \mid N = n, M = \bar{M}\right) - \mathbb{E}Y = (n - 2p)\left(\alpha + \gamma\bar{M}\right).$$

This quantity represents the predicted genotypic value of the individual at the locus if its methylation level was equal to the population average. We also refer to it as the breeding value of the individual. The average breeding value in a population is zero.

Table 1 summarizes the definitions and formulas for the expressed genetic value, basic genetic value, breeding value, and their respective variances.

2.7. Variances

In this section, we provide analytic expressions for the variances of the phenotype, the expressed and basic genetic values, the breeding values, and the phenotypic covariance between relatives.

2.7.1. Phenotypic variance

By definition,

$$V_P = \mathrm{Var}(Y) = \mathbb{E}\left(Y^2\right) - (\mathbb{E}Y)^2 = \mathbb{E}\left[(\mu + \alpha N + \beta M + \gamma NM + E)^2\right] - [\mathbb{E}(\mu + \alpha N + \beta M + \gamma NM + E)]^2.$$

Table 1
Expressed genetic value, basic genetic value, breeding value, and their variances.

Quantity	Concept	Definition	Value	Variance
Expressed genetic value	Predicted phenotypic deviation, given the genotype and the observed methylation state.	$\mathbb{E}(Y \mid N = n, M = m) - \mathbb{E}Y$	$\alpha(n-2p) + \beta(m-\bar{M}) + \gamma(nm - 2\bar{M}p)$	$2p(1-p)\alpha^2 + \bar{M}\left[2p(1-p)\gamma(2\alpha+\gamma\bar{M})\right] + V_M\left[2p(1-p)\gamma^2 + (\beta+2p\gamma)^2\right]$
Basic genetic value	Predicted phenotypic deviation, given the genotype and supposing an unmethylated state.	$\mathbb{E}(Y \mid N = n, M = 0) - \mathbb{E}Y$	$\alpha(n-2p) - \beta\bar{M} - 2\gamma\bar{M}p$	$2p(1-p)\alpha^2$
Breeding value	Predicted phenotypic deviation, given the genotype. Related to the expected phenotype of offspring.	$\mathbb{E}(Y \mid N = n) - \mathbb{E}Y$	$(n-2p)(\alpha+\gamma\bar{M})$	$2p(1-p)(\alpha+\gamma\bar{M})^2$

Under the assumptions of random mating and independence between N and M, the above expression simplifies to (for a derivation, see Appendix A.1)

$$V_P = 2p(1-p)\alpha^2 + \bar{M}\left[2p(1-p)\gamma(2\alpha+\gamma\bar{M})\right] + V_M\left[2p(1-p)\gamma^2 + (\beta+2p\gamma)^2\right] + V_E. \quad (9)$$

2.7.2. Expressed genetic variance

The variance of the expressed genetic values is given by

$$V_{exp} = 2p(1-p)\alpha^2 + \bar{M}\left[2p(1-p)\gamma(2\alpha+\gamma\bar{M})\right] + V_M\left[2p(1-p)\gamma^2 + (\beta+2p\gamma)^2\right]. \quad (10)$$

Clearly, the variance of the expressed genetic values is equal to the phenotypic variance minus the environmental variance. We will refer to the variance of expressed genetic values as the "expressed genetic variance", for it involves genetic and epigenetic (e.g., DNA methylation) factors determining gene expression.

2.7.3. Variance of basic genetic values

The variance of the basic genetic values in a population is equal to (cf. Appendix A.3)

$$V_{bgv} = 2p(1-p)\alpha^2, \quad (11)$$

which is the additive genetic variance in Fisher's classic model (however, see Section 3.2 about the estimation bias of α).

2.7.4. Variance of breeding values

The variance of the breeding values in a population is equal to (cf. Appendix A.4)

$$V_A = 2p(1-p)(\alpha+\gamma\bar{M})^2. \quad (12)$$

2.7.5. Covariance between relatives

The covariance between the genotypes of relatives is given by

$$\mathrm{Cov}(N_i, N_j) = 2r_{ij}p(1-p),$$

where r_{ij} is the expected additive genetic relationship between i and j (Sorensen, 2023). Under the assumptions of $\mathrm{Cov}(M_i, M_j) = \mathrm{Cov}(N_i, M_i) = 0$, the phenotypic covariance between relatives in groups i and j is given by the following expression (see Appendix A.2 for its derivation):

$$\mathrm{Cov}(Y_i, Y_j) = 2r_{ij}p(1-p)(\alpha+\gamma\bar{M})^2. \quad (13)$$

3. Results and discussion

A few hypothetical cases are presented in this section to illustrate how the model captures genotype-to-phenotype relationships influenced by methylation levels. Figs. 1–4 show plots of expected phenotypes against methylation levels, with each line corresponding to a genotype (A_1A_1, A_1A_2 or A_2A_2).

Fig. 1 shows a scenario where higher methylation levels are associated with lower phenotypic values. The expected phenotype is identical across genotypes when the locus is completely methylated ($m = 1$).

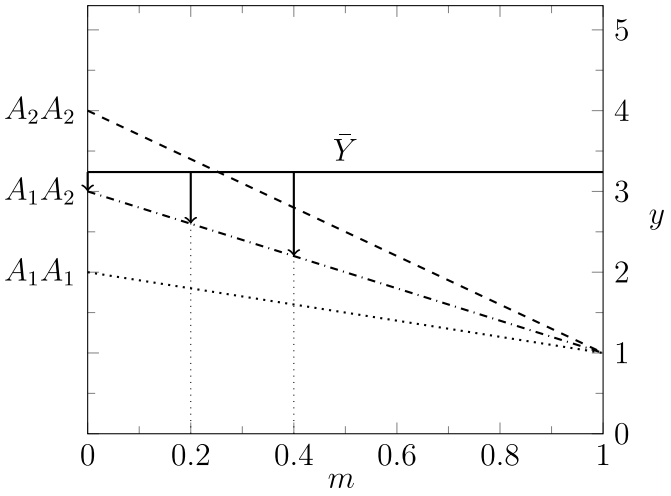

Fig. 1. Predicted phenotype versus methylation level. Each line corresponds to a genotype. The parameters in this example are $p = 0.9$, $\bar{M} = 0.2$, $\mu = 2$, $\alpha = 1$, $\beta = -1$, and $\gamma = -1$. From left to right: the basic genetic value ($M = 0$), the breeding value ($M = \bar{M}$), and the expressed genetic value ($M = m$) of a heterozygote individual with $m = 0.4$ are represented by line segments with arrows. The horizontal line at $y = \bar{Y}$ represents the average phenotype in the population.

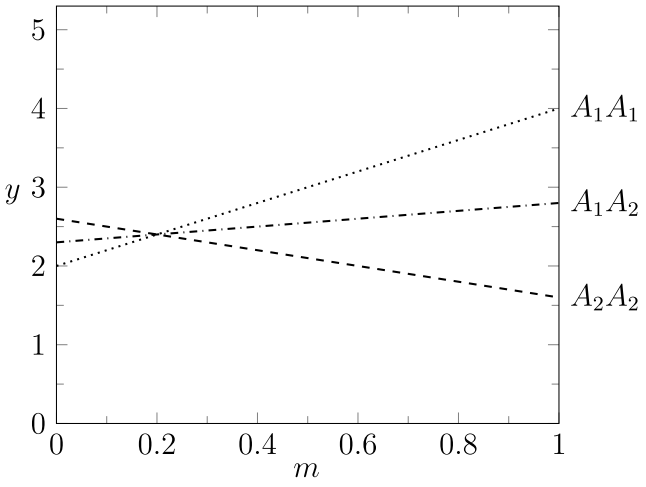

Fig. 2. Predicted phenotype versus methylation level. Each line corresponds to a genotype. The parameters in this example are $\mu = 2$, $\alpha = 0.3$, $\beta = 2$, and $\gamma = -1.5$. The homozygous genotypes are re-ranked at $m = 0.2$.

This pattern is conceptually coherent with methylation sites in the promoter region, where full methylation silences gene expression, resulting in equal average phenotypes due to the lack of a functional gene contribution. The figure also depicts the breeding, basic and expressed genetic values of a hypothetical heterozygous individual with a 40 % methylation rate.

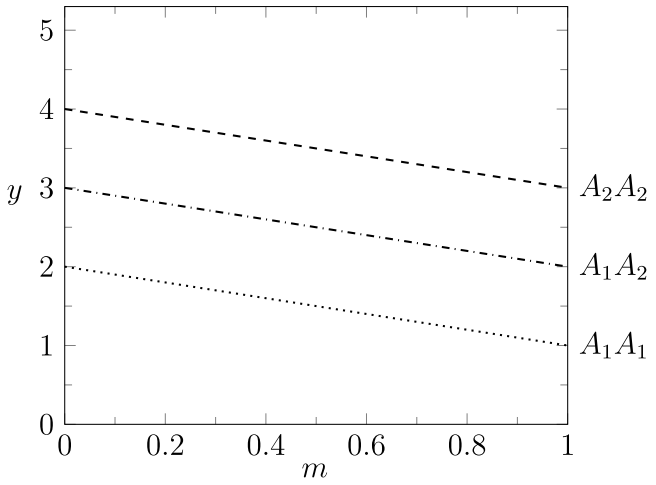

Fig. 3. Predicted phenotype versus methylation level. A case of no interaction ($\gamma = 0$). Each parallel line corresponds to a genotype. The parameters in this example are $\mu = 2$, $\alpha = 1$, $\beta = -1$, and $\gamma = 0$.

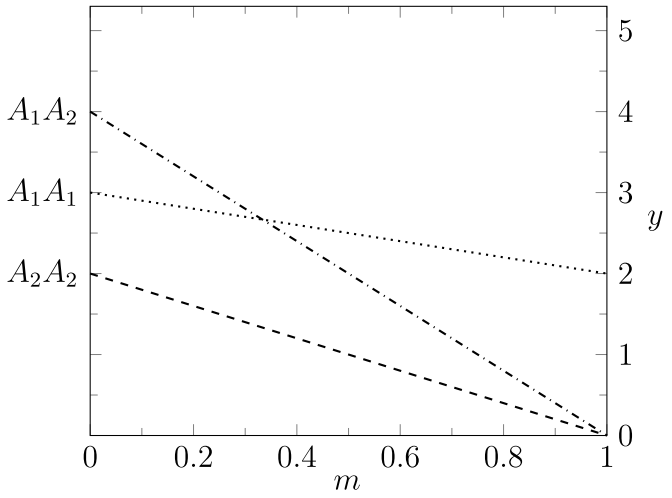

Fig. 4. Predicted phenotype versus methylation level. Parameterization with indicator variables (cf. Appendix C). The parameters in this example are $\mu = 3$, $\alpha_1 = 1$, $\alpha_2 = -1$, $\beta = -1$, $\gamma_1 = -2$, and $\gamma_2 = 1$.

Fig. 2 represents a case where changing methylation levels may either exacerbate or diminish the phenotypic differences among the genotypes in a population; notably, genotypes may be re-ranked depending on the observed methylation level. Slopes differ when $\gamma \neq 0$, indicating variation in the genotype-specific responses to methylation (i.e., there is interaction).

By contrast, Fig. 3 illustrates a scenario with no interaction, where all genotypes have the same slope. The case could be fit with a model without the interaction term. In experimental design, such a model is referred to as an analysis of covariance (Cochran, 1957; Montgomery, 2017), with M serving as the covariate.

In all cases, at any given methylation level, the lines of the homozygotes are equidistant from the line of the heterozygote. This constraint, characteristic of additive models, may be addressed by using a dominance model with indicator variables, which allows for a more flexible adjustment of slopes and intercepts (Fig. 4; cf. Appendix C).

3.1. Equivalence to basic model

When a locus is unmethylated, $M = 0$. Thus

$$\mathbb{E}(Y \mid N = n, M = 0) - \mathbb{E}Y = \alpha(n - 2p).$$

Hence, when there is no methylation, the model reduces to the classic one. Similarly, for the variance formula, in the absence of methylation in the population, $\bar{M} = 0$ and $V_M = 0$, yielding

$$\mathrm{Var}(Y) = V_P = 2\alpha^2 p(1-p) + V_E. \tag{14}$$

3.2. Absence of information on methylation and bias

Consider that the true data generating mechanism is affected by methylation as given by the model (4). Absence of methylation data leads to the classic, reduced model (1), which does not estimate β or γ. Let α^* and μ^* denote the allelic substitution effect and the intercept, respectively, in the reduced model. Under the true model, we have that

$$\mathbb{E}[Y \mid N = n] = (\mu + \beta\bar{M}) + (\alpha + \gamma\bar{M})n,$$

but the reduced model attempts to fit

$$\mathbb{E}[Y \mid N = n] = \mu^* + \alpha^* n.$$

Hence, omitting methylation creates biases in the estimates of μ and α. These biases are equal to $\beta\bar{M}$ and $\gamma\bar{M}$, respectively. There is no bias in α^* if $\gamma = 0$ (all genotypes share the same slope) or if $\bar{M} = 0$ (there is no methylation in the population).

To assess the bias of breeding values under model (1), we compare Eqs. (3) and (8):

$$\left[\alpha^*(n-2p)\right] - \left[(\alpha + \gamma\bar{M})(n-2p)\right] = (n-2p)\left[(\alpha^* - \alpha) - \gamma\bar{M}\right].$$

Since $\mathbb{E}(\alpha^* - \alpha) = \gamma\bar{M}$, the expected value of the previous expression is zero. Therefore, we conclude that there is no bias in the breeding value estimates of the classic model, given the assumptions used so far.

3.3. Phenotypic variance

Eq. (9) reveals that the phenotypic variance is determined by four factors: (i) the allele frequencies, (ii) the environmental variance, (iii) the parameters α, β, γ, and (iv) the distribution of methylation levels in the population (more precisely, its variance and mean). In Figs. 5–7, the phenotypic variance is plotted as a function of the allele frequency, while the other parameters are held fixed within each figure. The variances of the expressed genetic values, basic genetic values, and breeding values are also displayed together with the phenotypic variance. The curve that represents the phenotypic variance as a function of p is symmetric around the line $p = 1/2$ if $\gamma = -\beta$ (e.g., Fig. 6). Fig. 7 shows a case where a dominance model with indicator variables is used (for a

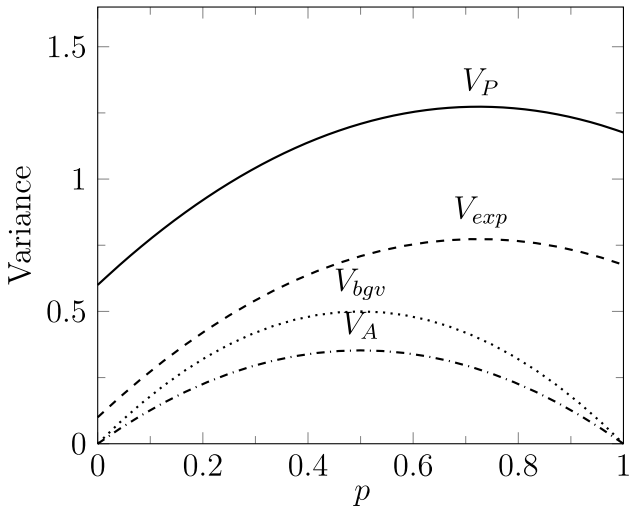

Fig. 5. Phenotypic (V_P), expressed ($V_{\exp}$), basic (V_{bgv}), and additive genetic (V_A) variances as functions of allele frequency (p). Each line corresponds to a type of variance. The parameters in this example are $\alpha = 1$, $\beta = -1$, $\gamma = -0.8$, $\bar{M} = 0.2$, $V_M = 0.1$, and $V_E = 0.5$.

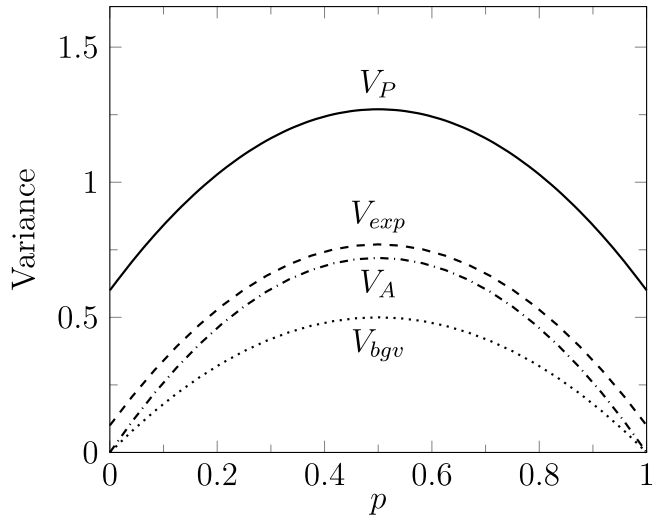

Fig. 6. Phenotypic (V_P), expressed ($V_{\exp}$), basic (V_{bgv}), and additive genetic (V_A) variances as functions of allele frequency (p). The curve is symmetric around $p = 0.5$ when $\gamma = -\beta$. The parameters in this example are $\alpha = 1$, $\beta = -1$, $\gamma = 1$, $\bar{M} = 0.2$, $V_M = 0.1$, and $V_E = 0.5$.

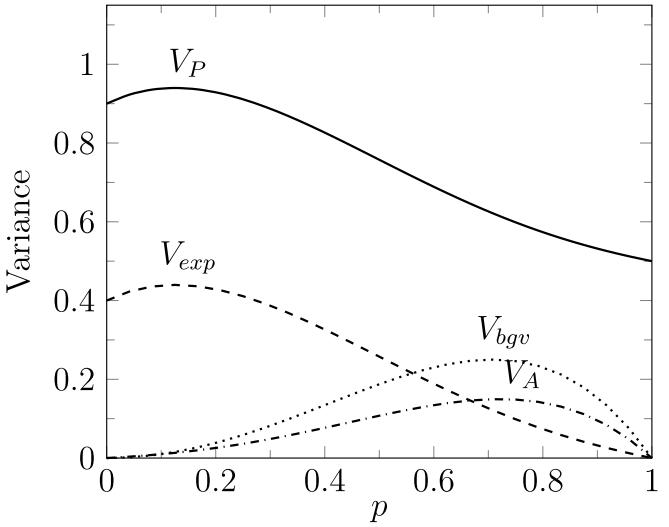

Fig. 7. Phenotypic (V_P), expressed ($V_{\exp}$), basic (V_{bgv}), and additive genetic (V_A) variances as functions of allele frequency (p). Parameterization with indicator variables. The parameters in this example are $\alpha_1 = -1$, $\alpha_2 = -1$, $\beta = -2$, $\gamma_1 = 1$, $\gamma_2 = 2$, $\bar{M} = 0.2$, $V_M = 0.1$, $V_E = 0.5$.

description of this model, see Appendix C). The distribution of methylation levels used for all examples is shown in Fig. 8.

By assuming independent methylation levels among related individuals and no covariance between genotype and methylotype, the models we present share features with genotype-environment interactions, where methylation acts as a continuous environmental gradient (Mather and Jinks, 1982; Falconer and Mackay, 1996). Consequently, Figs. 1–4 resemble reaction norm graphs.

We emphasize that, because DNA methylation patterns are tissue-specific, the expressed genetic value and the basic genetic value pertain to specific tissues or cell types of a multicellular organism. Moreover, expressed genetic values should ideally be compared among individuals of similar ages due to the "epigenetic clock", a process by which methylation levels change as individuals age (Horvath, 2013; Dufek et al., 2024). Therefore, expressed genetic values should be defined in the context of a particular tissue or cell type, at a specific age, and with reference to a specific locus and methylation site.

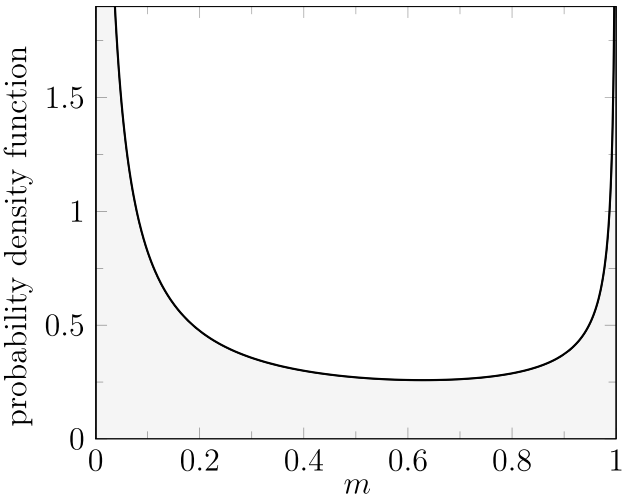

Fig. 8. Distribution of methylation levels used in Figs. 5 and 6. A beta distribution with parameters $a = 0.12$ and $b = 0.48$ has mean $\bar{M} = 0.2$ and variance $V_M = 0.1$.

3.4. Limitations of the model

For simplicity of exposition, we have assumed a linear relationship between methylation and phenotype. This simplification may overlook non-monotonic relationships or other functions that could better represent the underlying biological processes.

In the derivation of the formula of covariance between relatives, the covariance between methylation levels in any two relatives was set to zero; that is, we assumed $\mathrm{Cov}(M_i, M_j) = 0$. This assumption can be violated in two ways: when close relatives share similar environments or genes affecting methylation levels; and in transgenerational epigenetic inheritance. In animal breeding, shared environments are typically addressed by adding a common environment or litter effect in models for full siblings. Transgenerational epigenetic inheritance in vertebrates remains an ongoing topic of investigation. While much evidence suggests it is weak or rare (Goddard and Whitelaw, 2014; Fitz-James and Cavalli, 2022; Bird, 2024), the lack of powerful experimental designs and the problem of confounding make it challenging to fully assess its prevalence and mechanisms. Plants, by contrast, have shown clearer instances of transgenerational epigenetic inheritance across multiple generations, as seen in *Arabidopsis thaliana* (Johannes et al., 2009; Bošković and Rando, 2018; Furci et al., 2019). Similarly, studies in *Caenorhabditis elegans* (Klosin et al., 2017) and *Drosophila melanogaster* (Ciabrelli et al., 2017) have demonstrated instances of transgenerational epigenetic inheritance, albeit via mechanisms other than DNA methylation.

Lastly, to treat N and M as independent random variables may be unrealistic depending on the methylation site and genomic positions being considered. For example, in a study of 15 million CpG methylation sites in a human population, it was found that approximately 10 % of the sites had their methylation rate associated with a *cis*-acting DNA sequence variant (Stefansson et al., 2024).

3.5. Applications to observed data sets

While the model is presented in an expository manner to advance theoretical understanding, it can be readily applied to observed data for exploratory analyses. Phenotypes and genotypes may be grouped, a methylation metric selected, and the multiple linear regression with an interaction term performed for each genotype-methylation locus combination. Based on the regression estimates, one could discover the nature of the model for a particular locus (as illustrated in Figs. 1–4). The model accommodates various methylation metrics, such as the proportion of methylated reads at cytosine sites, global methylation levels, or

average methylation rates in promoter regions. While genotype data can take various forms, the assumption of a biallelic locus makes single-nucleotide polymorphisms (SNPs) a practical choice. However, testing all possible combinations of methylated cytosines and SNPs can result in a combinatorial explosion, posing computational challenges. This issue may be addressed by restricting the analyses to methylation sites and loci that are spatially proximal in the genome.

3.6. Extensions to multiple loci

Due to the polygenic nature of most complex traits, single-locus models offer limited potential for achieving high prediction accuracies. Consequently, extending the models to include multiple loci and methylation sites is a desirable goal. Mixed model methodologies achieve this by predicting SNP effects through covariance structures among related individuals (Meuwissen et al., 2001; VanRaden, 2008; de los Campos et al., 2013).

We have presented a model for the one-locus case and provided several formulas under the assumptions of random mating, independence of methylation levels between relatives, and independence of methylation and genotype. Expanding the model to multiple loci and multiple methylation sites introduces complexities such as epistasis, linkage, and a combinatorial increase in interaction terms, which bring estimation and identifiability issues. Many models cope with these challenges by focusing solely on additive effects. However, genome-wide approaches should require careful management of interaction terms, covariances between methylation and genotype, and methylation covariance among relatives. Overcoming these challenges is needed to develop more robust, genome-wide and methylome-wide prediction models.

4. Conclusion

We have presented a single-locus model that incorporates DNA methylation as an extra variable. Our approach builds on Fisher's classic decomposition of the genotypic value and phenotypic variance. Statistically, the model is formulated as a multiple linear regression with two predictor variables and an interaction term. For simplicity, we assumed that methylation levels were independent across individuals and uncorrelated with the genotype.

The expressed genetic value is the expected phenotypic deviation, conditional on the genotype and on the observed methylation level; the basic genetic value is the expected phenotypic deviation, conditional on the genotype and on the methylation value of zero. The breeding value is the expected phenotypic deviation, conditional on the genotype; conditioning the methylation level on the population average also yields the breeding value. Under the aforementioned assumptions, the presence of methylation introduces no bias in the estimated breeding values.

The phenotypic variance is determined by the functional relationship between phenotype, methylation and genotype, the distribution of methylation values in a population, the allele frequency, and the environmental variance. The phenotypic covariance is determined by the additive genetic relationship, the regression coefficients α and γ, the allele frequency, and the average methylation rate in the population.

Data availability

No data was used for the research described in the article.

Funding

This work was supported by the European Union's Horizon 2020 research and innovation programme under grant 101000236 (GEroNIMO project).

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

L. Ayres reports financial support was provided by H2020 Food Security Sustainable Agriculture and Forestry Marine Maritime and Inland Water Research and the Bioeconomy. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

L. Ayres: Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Visualization; **H. Bovenhuis:** Conceptualization, Methodology, Writing – review & editing, Visualization, Supervision, Funding acquisition, Resources; **M.P.L. Calus:** Conceptualization, Methodology, Writing – review & editing, Visualization, Supervision, Funding acquisition, Resources, Project administration.

Appendix A. Derivation of variances

A.1. Derivation of the phenotypic variance formula

The following steps lead to Eq. (9). The first term expands to

$$\begin{aligned}&\mathbb{E}\big(\mu^2+\alpha^2N^2+\beta^2M^2+\gamma^2M^2N^2+E^2\\&\quad+2\mu\alpha N+2\mu\beta M+2\mu\gamma MN+2\mu E\\&\quad+2\alpha\beta MN+2\alpha\gamma MN^2+2\alpha NE\\&\quad+2\beta\gamma M^2N+2\beta ME+2\gamma MNE\big).\end{aligned}$$

By linearity of expectation, this is equal to

$$\begin{aligned}&\mathbb{E}(\mu^2)+\mathbb{E}(\alpha^2N^2)+\mathbb{E}(\beta^2M^2)+\mathbb{E}(\gamma^2M^2N^2)\\&\quad+\mathbb{E}(E^2)+\mathbb{E}(2\mu\alpha N)+\mathbb{E}(2\mu\beta M)+\mathbb{E}(2\mu\gamma MN)\\&\quad+\mathbb{E}(2\alpha\beta MN)+\mathbb{E}(2\alpha\gamma MN^2)+\mathbb{E}(2\alpha NE)\\&\quad+\mathbb{E}(2\mu E)+\mathbb{E}(2\beta\gamma M^2N)+\mathbb{E}(2\beta ME)\\&\quad+\mathbb{E}(2\gamma MNE)\\&=\mu^2+\alpha^2\mathbb{E}(N^2)+\beta^2\mathbb{E}(M^2)+\gamma^2\mathbb{E}(M^2N^2)\\&\quad+\mathbb{E}(E^2)+2\mu\alpha\mathbb{E}(N)+2\mu\beta\mathbb{E}(M)\\&\quad+2\mu\gamma\mathbb{E}(MN)+2\mu\mathbb{E}(E)+2\alpha\beta\mathbb{E}(MN)\\&\quad+2\alpha\gamma\mathbb{E}(MN^2)+2\alpha\mathbb{E}(NE)+2\beta\gamma\mathbb{E}(M^2N)\\&\quad+2\beta\mathbb{E}(ME)+2\gamma\mathbb{E}(MNE).\end{aligned}$$

We assume independence between M and N. This implies $\mathbb{E}(MN)=\mathbb{E}M\mathbb{E}N$ and $\mathbb{E}\big(M^2N^2\big)=\mathbb{E}\big(M^2\big)\mathbb{E}\big(N^2\big)$. Additionally, $\mathbb{E}E=0$, $\mathbb{E}N=2p$, and under Hardy–Weinberg equilibrium $\mathbb{E}\big(N^2\big)=2p(1+p)$. To shorten the notation, let $\bar{M}=\mathbb{E}M$, $V_M=\mathrm{Var}(M)$, $V_P=\mathrm{Var}(Y)$, and $V_E=\mathrm{Var}(E)$. Then

$$\begin{aligned}\mathbb{E}\big(Y^2\big)&=\mu^2+2\alpha^2p(1+p)+\beta^2\mathbb{E}\big(M^2\big)\\&\quad+2\gamma^2\mathbb{E}\big(M^2\big)p(1+p)+V_E+4\mu\alpha p\\&\quad+4\beta\gamma\mathbb{E}\big(M^2\big)p+2\mu\beta\bar{M}+4\mu\gamma\bar{M}p\\&\quad+4\alpha\beta\bar{M}p+4\alpha\gamma\bar{M}p(1+p)\\&=\mu^2+2\alpha^2p(1+p)+\beta^2\big(V_M+\bar{M}^2\big)\\&\quad+2\gamma^2\big(V_M+\bar{M}^2\big)p(1+p)+V_E+4\mu\alpha p\\&\quad+4\beta\gamma\big(V_M+\bar{M}^2\big)p+2\mu\beta\bar{M}+4\mu\gamma\bar{M}p\\&\quad+4\alpha\beta\bar{M}p+4\alpha\gamma\bar{M}p(1+p).\end{aligned}$$

The second term is:

$$\begin{aligned}(\mathbb{E}Y)^2 &= [\mathbb{E}(\mu+\alpha N+\beta M+\gamma NM+E)]^2\\ &= \left[\mu+2\alpha p+\beta\bar{M}+2\gamma\bar{M}p\right]^2\\ &= \mu^2+4\alpha^2p^2+\beta^2\bar{M}^2+4\gamma^2\bar{M}^2p^2\\ &\quad+4\mu\alpha p+2\mu\beta\bar{M}+4\mu\gamma\bar{M}p\\ &\quad+4\alpha\beta\bar{M}p+4\beta\gamma\bar{M}^2p+8\alpha\gamma\bar{M}p^2.\end{aligned}$$

Their difference can be simplified to

$$\begin{aligned}V_P &= 2\alpha^2p(1-p)+\bar{M}\left[2\gamma p(1-p)\left(\gamma\bar{M}+2\alpha\right)\right]\\ &\quad+V_M\left[2\gamma^2p(1-p)+(\beta+2\gamma p)^2\right]+V_E.\end{aligned}$$

A.2. Derivation of the phenotypic covariance formula

$$\begin{aligned}\mathrm{Cov}(Y_i,Y_j) = \mathrm{Cov}(&\mu+\alpha N_i+\beta M_i+\gamma N_iM_i+E_i,\\ &\mu+\alpha N_j+\beta M_j+\gamma N_jM_j+E_j).\end{aligned}$$

By the property of the covariance of sums of random variables, this expands to:

$$\begin{aligned}\mathrm{Cov}(Y_i,Y_j) &= \alpha^2\mathrm{Cov}(N_i,N_j)+\alpha\beta\mathrm{Cov}(N_i,M_j)\\ &\quad+\alpha\gamma\mathrm{Cov}(N_i,N_jM_j)\\ &\quad+\beta\alpha\mathrm{Cov}(M_i,N_j)\\ &\quad+\beta^2\mathrm{Cov}(M_i,M_j)\\ &\quad+\beta\gamma\mathrm{Cov}(M_i,N_jM_j)\\ &\quad+\gamma\alpha\mathrm{Cov}(N_iM_i,N_j)\\ &\quad+\gamma\beta\mathrm{Cov}(N_iM_i,M_j)\\ &\quad+\gamma^2\mathrm{Cov}(N_iM_i,N_jM_j).\end{aligned}$$

We assume $\mathrm{Cov}(N_i,M_j)=\mathrm{Cov}(M_i,M_j)=0$, from which it follows that $\mathrm{Cov}(N_iM_i,M_j)=0$. Further, we use the identity $\mathrm{Cov}(N_i,N_j)=2r_{ij}p(1-p)$. Thus:

$$\begin{aligned}\mathrm{Cov}(Y_i,Y_j) &= \alpha^2\left[2r_{ij}p(1-p)\right]\\ &\quad+2\alpha\gamma\mathrm{Cov}(N_i,N_jM_j)\\ &\quad+\gamma^2\mathrm{Cov}(N_iM_i,N_jM_j).\end{aligned}$$

We seek $\mathrm{Cov}(N_i,N_jM_j)$, which can be rewritten as

$$\mathbb{E}(N_iN_jM_j)-\mathbb{E}N\mathbb{E}(N_iM_j)=\mathbb{E}(N_iN_j)\mathbb{E}M-\mathbb{E}N\mathbb{E}N\mathbb{E}M.$$

Thus

$$\begin{aligned}\mathrm{Cov}(N_i,N_jM_j) &= \mathbb{E}(N_iN_j)\bar{M}-(\mathbb{E}N)^2\bar{M}\\ &= \mathbb{E}(N_iN_j)\bar{M}-4p^2\bar{M}.\end{aligned}$$

It follows from $\mathrm{Cov}(N_i,N_j)=2r_{ij}p(1-p)$ that $\mathbb{E}(N_iN_j)=2r_{ij}p(1-p)+4p^2$. Therefore

$$\begin{aligned}\mathrm{Cov}(N_i,N_jM_j) &= \left[2r_{ij}p(1-p)+4p^2\right]\bar{M}-4p^2\bar{M}\\ &= 2r_{ij}p(1-p)\bar{M}.\end{aligned}$$

The expression for $\mathrm{Cov}(N_iM_i,N_jM_j)$ is obtained in a similar way:

$$\begin{aligned}\mathrm{Cov}(N_iM_i,N_jM_j) &= \mathbb{E}(N_iN_jM_iM_j)-[\mathbb{E}(NM)]^2\\ &= \mathbb{E}(N_iN_j)\bar{M}^2-4p^2\bar{M}^2\\ &= 2r_{ij}p(1-p)\bar{M}^2.\end{aligned}$$

We can now use them in the expression for the phenotypic covariance:

$$\begin{aligned}\mathrm{Cov}(Y_i,Y_j) &= \alpha^2\left[2r_{ij}p(1-p)\right]+2\alpha\gamma\left[2r_{ij}p(1-p)\bar{M}\right]+\gamma^2\left[2r_{ij}p(1-p)\bar{M}^2\right]\\ &= 2r_{ij}p(1-p)\left(\alpha+\gamma\bar{M}\right)^2.\end{aligned}$$

A.3. Derivation of the variance of basic genetic values

$$\begin{aligned}V_{bgv} &= \mathbb{E}\left[\left(\alpha(N-2p)-\beta\bar{M}-2\gamma\bar{M}p\right)^2\right]\\ &\quad-\left\{\mathbb{E}\left[\alpha(N-2p)-\beta\bar{M}-2\gamma\bar{M}p\right]\right\}^2\\ &= \mathbb{E}\left(\alpha^2N^2-4\alpha^2pN+4\alpha^2p^2-2\alpha\beta\bar{M}N\right.\\ &\quad+\beta^2\bar{M}^2+4\alpha\beta\bar{M}p+4\gamma^2\bar{M}^2p^2\\ &\quad\left.+4\beta\gamma\bar{M}^2p-4\alpha\gamma\bar{M}pN+8\alpha\gamma\bar{M}p^2\right)\\ &\quad-\left[-\bar{M}(\beta+2\gamma p)\right]^2\\ &= 2\alpha^2p(1+p)-8\alpha^2p^2+4\alpha^2p^2-4\alpha\beta\bar{M}p\\ &\quad+\beta^2\bar{M}^2+4\alpha\beta\bar{M}p+4\gamma^2\bar{M}^2p^2\\ &\quad+4\beta\gamma\bar{M}^2p-8\alpha\gamma\bar{M}p^2+8\alpha\gamma\bar{M}p^2\\ &\quad-\left(\beta^2\bar{M}^2+4\gamma^2\bar{M}^2p^2+4\beta\gamma\bar{M}^2p\right)\\ &= 2\alpha^2p-2\alpha^2p^2=2p(1-p)\alpha^2.\end{aligned}$$

A.4. Derivation of the variance of breeding values

$$\begin{aligned}V_A &= \mathbb{E}\left\{\left[(\alpha+\gamma\bar{M})(N-2p)\right]^2\right\}-\left\{\mathbb{E}\left[(\alpha+\gamma\bar{M})(N-2p)\right]\right\}^2\\ &= \mathbb{E}\left(\gamma^2\bar{M}^2N^2-4\gamma^2\bar{M}^2Np+4\gamma^2\bar{M}^2p^2\right.\\ &\quad+2\alpha\gamma\bar{M}N^2-8\alpha\gamma\bar{M}Np+8\alpha\gamma\bar{M}p^2\\ &\quad\left.+\alpha^2N^2-4\alpha^2Np+4\alpha^2p^2\right)-0^2\\ &= -8\gamma^2\bar{M}^2p^2+4\gamma^2\bar{M}^2p^2\\ &\quad+\left(2\gamma^2\bar{M}^2p^2+2\gamma^2\bar{M}^2p\right)-16\alpha\gamma\bar{M}p^2\\ &\quad+\left(4\alpha\gamma\bar{M}p^2+4\alpha\gamma\bar{M}p\right)+8\alpha\gamma\bar{M}p^2\\ &\quad-8\alpha^2p^2+4\alpha^2p^2+\left(2\alpha^2p^2+2\alpha^2p\right)\\ &= -2\gamma^2\bar{M}^2p^2+2\gamma^2\bar{M}^2p-4\alpha\gamma\bar{M}p^2\\ &\quad+4\alpha\gamma\bar{M}p-2\alpha^2p^2+2\alpha^2p\\ &= 2p(1-p)\left(\alpha+\gamma\bar{M}\right)^2.\end{aligned}$$

Appendix B. Genetic values of progeny

Assuming random mating, the expected expressed genetic value of the progeny of an A_1A_1 individual is

$$p\left[(\alpha+\gamma\bar{M})(1-2p)\right]+(1-p)\left[-2\left(\alpha+\gamma\bar{M}\right)p\right]=-\left(\alpha+\gamma\bar{M}\right)p;$$

that of an A_1A_2 individual is

$$\begin{aligned}&\frac{1}{2}\left\{p\left[(\alpha+\gamma\bar{M})(1-2p)\right]+(1-p)\left[-2\left(\alpha+\gamma\bar{M}\right)p\right]\right\}\\ &\quad+\frac{1}{2}\left\{p\left[2\left(\alpha+\gamma\bar{M}\right)(1-p)\right]+(1-p)\left[(\alpha+\gamma\bar{M})(1-2p)\right]\right\}\\ &= \frac{1}{2}\left(\alpha+\gamma\bar{M}\right)(1-2p).\end{aligned}$$

Lastly, that of an A_2A_2 individual is

$$\begin{aligned}&p\left[2\left(\alpha+\gamma\bar{M}\right)(1-p)\right]+(1-p)\left[(\alpha+\gamma\bar{M})(1-2p)\right]\\ &= \left(\alpha+\gamma\bar{M}\right)(1-p).\end{aligned}$$

Table B.1 summarizes these results, showing that that the expected expressed genetic value of the progeny of a parent is not one half the expressed genetic value of the parent.

Similar calculations for the basic genetic value and the breeding value yield Tables B.2 and B.3, respectively.

Table B.1
Expressed genetic value of parent and expected expressed genetic value of progeny, according to genotype of parent.

Genotype	Expressed genetic value	
	Parent	*Progeny*
A_1A_1	$\alpha(1-2p)+\beta(m-\bar{M})+\gamma(m-2\bar{M}p)$	$(\alpha+\gamma\bar{M})(-p)$
A_1A_2	$\alpha(2-2p)+\beta(m-\bar{M})+\gamma(2m-2\bar{M}p)$	$(\alpha+\gamma\bar{M})\frac{1}{2}(1-2p)$
A_2A_2	$\alpha(3-2p)+\beta(m-\bar{M})+\gamma(3m-2\bar{M}p)$	$(\alpha+\gamma\bar{M})(1-p)$

Table B.2
Expected basic genetic value of progeny according to the genotype of the parent.

Genotype	Basic genetic value	
	Parent	*Progeny*
A_1A_1	$-\bar{M}(\beta+2\gamma p)-2\alpha p$	$-\bar{M}(\beta+2\gamma p)-\alpha p$
A_1A_2	$-\bar{M}(\beta+2\gamma p)-2\alpha p+\alpha$	$-\bar{M}(\beta+2\gamma p)-\alpha p+\frac{1}{2}\alpha$
A_2A_2	$-\bar{M}(\beta+2\gamma p)-2\alpha p+2\alpha$	$-\bar{M}(\beta+2\gamma p)-\alpha p+\alpha$

Table B.3
Average breeding value of progeny according to the genotype of a parent.

Genotype	Breeding value	
	Parent	*Progeny*
A_1A_1	$(\alpha+\gamma\bar{M})(-2p)$	$(\alpha+\gamma\bar{M})(-p)$
A_1A_2	$(\alpha+\gamma\bar{M})(1-2p)$	$(\alpha+\gamma\bar{M})\frac{1}{2}(1-2p)$
A_2A_2	$(\alpha+\gamma\bar{M})(2-2p)$	$(\alpha+\gamma\bar{M})(1-p)$

Appendix C. Model with indicator variables

We may represent the genotype as a factor with three levels using two indicator variables, X_1 and X_2, with the coding defined in Table C.1. This parameterization provides more flexibility in modeling the relationships for each genotype, as illustrated in Fig. 4.

$$Y=\mu+\alpha_1X_1+\alpha_2X_2+\beta M+\gamma_1X_1M+\gamma_2X_2M+E \tag{C.1}$$

Each genotype has its own slope and intercept:

$$Y=\begin{cases}\mu+\beta M+E, & \text{if } A_1A_1,\\ (\mu+\alpha_1)+(\beta+\gamma_1)M+E, & \text{if } A_1A_2,\\ (\mu+\alpha_2)+(\beta+\gamma_2)M+E, & \text{if } A_2A_2.\end{cases}$$

When an individual has no DNA methylation for the locus, $M=0$ and

$$\mathbb{E}\big(Y\mid X_1=x_1,X_2=x_2,M=0\big)=\mu+\alpha_1x_1+\alpha_2x_2.$$

If there is Hardy–Weinberg equilibrium, with X_1 and M, X_2 and M independent, then

$$\mathbb{E}Y=\mu+2\alpha_1p(1-p)+\alpha_2p^2+\bar{M}\left[\beta+2\gamma_1p(1-p)+\gamma_2p^2\right].$$

Thus, under model (C.1) the basic genetic value is given by:

$$\begin{aligned}&\mathbb{E}\big(Y\mid X_1=x_1,X_2=x_2,M=0\big)-\mathbb{E}Y\\&\quad=\alpha_1x_1+\alpha_2x_2-\left\{2\alpha_1p(1-p)+\alpha_2p^2+\bar{M}\left[\beta+2\gamma_1p(1-p)+\gamma_2p^2\right]\right\}.\end{aligned}$$

Similarly, the expressed genetic value is

$$\begin{aligned}&\mathbb{E}\big(Y\mid X_1=x_1,X_2=x_2,M=m\big)-\mathbb{E}Y\\&\quad=\alpha_1x_1+\alpha_2x_2+\big(\beta+\gamma_1x_1+\gamma_2x_2\big)m\\&\qquad-\left\{2\alpha_1p(1-p)+\alpha_2p^2+\bar{M}\left[\beta+2\gamma_1p(1-p)+\gamma_2p^2\right]\right\},\end{aligned}$$

while the breeding value is

$$\begin{aligned}&\mathbb{E}\big(Y\mid X_1=x_1,X_2=x_2,M=\bar{M}\big)-\mathbb{E}Y\\&\quad=\alpha_1x_1+\alpha_2x_2\\&\qquad-\left\{2\alpha_1p(1-p)+\alpha_2p^2+\bar{M}\left[\gamma_1x_1+\gamma_2x_2+2\gamma_1p(1-p)+\gamma_2p^2\right]\right\}.\end{aligned}$$

The following variance formulas are given for the model with indicator variables. We omit their proofs, which follow the same path as the previous derivations for the other model.

Table C.1
Indicator variable representation of genotypes.

Genotype	X_1	X_2
A_1A_1	0	0
A_1A_2	1	0
A_2A_2	0	1

The formula for the phenotypic variance is:

$$\begin{aligned}V_P=\;&p(1-p)\left\{\bar{M}^2\left\{\gamma_1^2[2-4p(1-p)]+\gamma_2^2p(1+p)-4\gamma_1\gamma_2p^2\right\}\right.\\&-4\alpha_1\left\{\bar{M}\left[\gamma_2p^2+\gamma_1(2p(1-p)-1)\right]+\alpha_2p^2\right\}\\&+2\alpha_2p\bar{M}\left[\gamma_2(1+p)-2\gamma_1p\right]+\alpha_1^2[2-4p(1-p)]+\alpha_2^2p(1+p)\Big\}\\&+V_M\left\{\beta^2+2\beta\gamma_2p^2+\gamma_2p^2+2\gamma_1\big(2\beta+\gamma_1\big)p(1-p)\right\}+V_E.\end{aligned}$$

By subtracting the environmental variance from the phenotypic variance, we obtain the expressed genetic variance (i.e., the variance of the expressed genetic values):

$$\begin{aligned}V_{exp}=\;&p(1-p)\left\{\bar{M}^2\left\{\gamma_1^2[2-4p(1-p)]+\gamma_2^2p(1+p)-4\gamma_1\gamma_2p^2\right\}\right.\\&-4\alpha_1\left\{\bar{M}\left[\gamma_2p^2+\gamma_1(2p(1-p)-1)\right]+\alpha_2p^2\right\}\\&+2\alpha_2p\bar{M}\left[\gamma_2(1+p)-2\gamma_1p\right]+\alpha_1^2[2-4p(1-p)]+\alpha_2^2p(1+p)\Big\}\\&+V_M\left\{\beta^2+2\beta\gamma_2p^2+\gamma_2p^2+2\gamma_1\big(2\beta+\gamma_1\big)p(1-p)\right\}.\end{aligned}$$

The variance of the basic genetic values is given by:

$$V_{bgv}=\alpha_1^2p(1-p)[1-2p(1-p)]+\alpha_2^2(1-p)^2p(2-p)-4\alpha_1\alpha_2p(1-p)^3.$$

The variance of the breeding values is given by:

$$\begin{aligned}V_A=\;&\big(\alpha_1+\gamma_1\bar{M}\big)^2p(1-p)[1-2p(1-p)]\\&+\big(\alpha_2+\gamma_2\bar{M}\big)^2(1-p)^2p(2-p)\\&-4\big(\alpha_1+\gamma_1\bar{M}\big)\big(\alpha_2+\gamma_2\bar{M}\big)p(1-p)^3.\end{aligned}$$

References

Álvarez-Castro, J.M., 2014. Dissecting genetic effects with imprinting. Front. Ecol. Evol. 2, 1–9.

Álvarez-Castro, J.M., 2023. Genes, Environments and Interactions. Springer.

Bibikova, M., Barnes, B., Tsan, C., Ho, V., Klotzle, B., Le, J.M., Delano, D., Zhang, L., Schroth, G.P., Gunderson, K.L., Fan, J.-B., Shen, R., 2011. High density DNA methylation array with single CpG site resolution. Genomics 98 (4), 288–295.

Bird, A., 2024. Transgenerational epigenetic inheritance: a critical perspective. Front. Epigenet. Epigenom. 2, 1–6.

Bošković, A., Rando, O.J., 2018. Transgenerational epigenetic inheritance. Annu. Rev. Genet. 52, 21–41.

de los Campos, G., Hickey, J.M., Pong-Wong, R., Daetwyler, H.D., Calus, M. P.L., 2013. Whole-genome regression and prediction methods applied to plant and animal breeding. Genetics 193 (2), 327–345.

Chai, H., Zhou, X., Zhang, Z., Rao, J., Zhao, H., Zhao, H., Yang, Y., 2021. Integrating multi-omics data through deep learning for accurate cancer prognosis prediction. Comput. Biol. Med. 134 (104481), 1–8.

Christensen, O.F., Börner, V., Varona, L., Legarra, A., 2021. Genetic evaluation including intermediate omics features. Genetics 219 (2), 1–14.

Ciabrelli, F., Comoglio, F., Fellous, S., Bonev, B., Ninova, M., Szabo, Q., Xuéreb, A., Klopp, C., Aravin, A., Paro, R., Bantignies, F., Cavalli, G., 2017. Stable polycomb-dependent transgenerational inheritance of chromatin states in *Drosophila*. Nat. Genet. 49 (6), 876–886.

Cochran, W.G., 1957. Analysis of covariance: its nature and uses. Biometrics 13 (3), 261–281.

Dufek, G., Katriel, G., Snir, S., Pellegrini, M., 2024. Exponential dynamics of DNA methylation with age. J. Theor. Biol. 579, 1–6.

Falconer, D.S., Mackay, T. F.C., 1996. Introduction to Quantitative Genetics. Longman. Fourth ed.

Fisher, R.A., 1941. Average excess and average effect of a gene substitution. Ann. Eugen. 11, 53–63.

Fitz-James, M.H., Cavalli, G., 2022. Molecular mechanisms of transgenerational epigenetic inheritance. Nat. Rev. Genet. 23 (6), 325–341.

Frommer, M., McDonald, L.E., Millar, D.S., Collis, C.M., Watt, F., Grigg, G.W., Molloy, P.L., Paul, C.L., 1992. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc. Natl. Acad. Sci. USA 89 (5), 1827–1831.

Furci, L., Jain, R., Stassen, J., Berkowitz, O., Whelan, J., Roquis, D., Baillet, V., Colot, V., Johannes, F., Ton, J., 2019. Identification and characterisation of hypomethylated DNA loci controlling quantitative resistance in *Arabidopsis*. eLife 8, 1–23.

Goddard, M.E., Whitelaw, E., 2014. The use of epigenetic phenomena for the improvement of sheep and cattle. Front. Genet. 5 (247), 1–6.

Guo, Z., Magwire, M.M., Basten, C.J., Xu, Z., Wang, D., 2016. Evaluation of the utility of gene expression and metabolic information for genomic prediction in maize. Theor. Appl. Genet. 129 (12), 2413–2427.

Horvath, S., 2013. DNA methylation age of human tissues and cell types. Genome Biol. 14 (10), 1–19.

Hu, Y., Morota, G., Rosa, G. J.M., Gianola, D., 2015. Prediction of plant height in *Arabidopsis thaliana* using DNA methylation data. Genetics 201 (2), 779–793.

Johannes, F., Porcher, E., Teixeira, F.K., Saliba-Colombani, V., Simon, M., Agier, N., Bulski, A., Albuisson, J., Heredia, F., Audigier, P., Bouchez, D., Dillmann, C., Guerche, P.,

Hospital, F., Colot, V., 2009. Assessing the impact of transgenerational epigenetic variation on complex traits. PLoS Genet. 5 (6), 1–11.
Klosin, A., Casas, E., Hidalgo-Carcedo, C., Vavouri, T., Lehner, B., 2017. Transgenerational transmission of environmental information in *C. elegans*. Science 356 (6335), 320–323.
Klug, W.S., Cummings, M.R., Spencer, C.A., Palladino, M.A., Killian, D.J., 2019. Concepts of Genetics. Pearson. Twelfth ed.
de Koning, D.-J., Bovenhuis, H., van Arendonk, J. A.M., 2002. On the detection of imprinted quantitative trait loci in experimental crosses of outbred species. Genetics 161 (2), 931–938.
Li, E., Zhang, Y., 2014. DNA methylation in mammals. Cold Spring Harbor Perspect. Biol. 6 (5), 1–21.
Lister, R., Ecker, J.R., 2009. Finding the fifth base: genome-wide sequencing of cytosine methylation. Genome Res. 19 (6), 959–966.
Lynch, M., Walsh, B., 1998. Genetics and Analysis of Quantitative Traits. Sinauer Associates.
Mather, K., Jinks, J.L., 1982. Biometrical Genetics. Chapman and Hall.
Meissner, A., Gnirke, A., Bell, G.W., Ramsahoye, B., Lander, E.S., Jaenisch, R., 2005. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. Nucleic Acids Res. 33 (18), 5868–5877.
Meuwissen, T. H.E., Hayes, B.J., Goddard, M.E., 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157 (4), 1819–1829.
Montgomery, D.C., 2017. Design and Analysis of Experiments. Wiley. Ninth ed.
Morgante, F., Huang, W., Sørensen, P., Maltecca, C., Mackay, T. F.C., 2020. Leveraging multiple layers of data to predict *Drosophila* complex traits. G3 (Bethesda) 10 (12), 4599–4613.
O'Brien, E.K., Wolf, J.B., 2019. Evolutionary quantitative genetics of genomic imprinting. Genetics 211 (1), 75–88.
Santure, A.W., Spencer, H.G., 2006. Influence of mom and dad: quantitative genetic models for maternal effects and genomic imprinting. Genetics 173 (4), 2297–2316.
Sorensen, D., 2023. Statistical Learning in Genetics. Springer.
Spencer, H.G., 2002. The correlation between relatives on the supposition of genomic imprinting. Genetics 161 (1), 411–417.
Spencer, H.G., 2009. Effects of genomic imprinting on quantitative traits. Genetica 136, 285–293.
Stefansson, O.A., Sigurpalsdottir, B.D., Rognvaldsson, S., Halldorsson, G.H., Juliusson, K., Sveinbjornsson, G., et al., 2024. The correlation between CpG methylation and gene expression is driven by sequence variants. Nat. Genet. 56 (8), 1624–1631.
VanRaden, P.M., 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91 (11), 4414–4423.
Voet, D., Voet, J.G., 2010. Biochemistry. Wiley. Fourth ed.
Wallace, E. V.B., Stoddart, D., Heron, A.J., Mikhailova, E., Maglia, G., Donohoe, T.J., Bayley, H., 2010. Identification of epigenetic DNA modifications with a protein nanopore. Chem. Commun. 46, 8195–8197.
Weber, M., Davies, J.J., Wittig, D., Oakeley, E.J., Haase, M., Lam, W.L., Schübeler, D., 2005. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nat. Genet. 37 (8), 853–862.
Zhao, T., Zeng, J., Cheng, H., 2022. Extend mixed models to multilayer neural networks for genomic prediction including intermediate omics data. Genetics 221 (1), 1–10.