



Genetic architecture and genomic prediction for yield, winter damage, and digestibility traits in timothy (*Phleum pratense* L.) using genotyping-by-sequencing data

N. Vargas Jurado¹ · H. Kärkkäinen¹ · D. Fischer¹ · O. Bitz¹ · O. Manninen² · P. Pärssinen² · M. Isoahti¹ · I. Strandén¹ · E. A. Mäntysaari¹

Received: 9 August 2024 / Accepted: 18 February 2025 / Published online: 18 March 2025
© The Author(s) 2025

Abstract

Key message Accurate prediction of genomic breeding values for Timothy was possible using genomic best linear unbiased prediction.

Abstract Timothy (*Phleum pratense* L.) is a grass species of great importance for Finnish agricultural production systems. Genotyping-by-sequencing along with genomic prediction methods offer the possibility to develop breeding materials efficiently. In addition, knowledge about the relationships among traits may be used to increase rates of genetic gain. Still, the quality of the genotypes and the validation population may affect the accuracy of predictions. The objectives of the study were (i) to estimate variance components for yield, winter damage and digestibility traits, and (ii) to assess the accuracy of genomic predictions. Variance components were estimated using genomic residual maximum likelihood where the genomic relationship matrix was scaled using a novel approach. Genomic breeding values were estimated using genomic best linear unbiased prediction in single- and multiple-trait settings, and for different marker filtering criteria. Estimates of heritability ranged from 0.13 ± 0.03 to 0.86 ± 0.05 for yield at first cut and organic matter digestibility at second cut, respectively. Genetic correlations ranged from -0.72 ± 0.12 to 0.59 ± 0.04 between yield at first cut and winter damage, and between digestibility at first and second cuts, respectively. Accuracy of prediction was not severely affected by the quality of genotyping. Using family cross-validation and single-trait models, predictive ability ranged from 0.18 to 0.62 for winter damage and digestibility at second cut, respectively. In addition, validation using forward prediction showed that estimated genomic breeding values were moderately accurate with little dispersion. Thus, genomic prediction constitutes a valuable tool for improving Timothy in Finland.

Introduction

Timothy (*Phleum pratense* L.) is an important perennial grass species widely used in cool and humid climates across the Nordic countries, central Europe, northeastern and northwestern North America, Russia, and Japan (Berg et al. 1996). Due to its high yield, tolerance to winter conditions in northern latitudes, and high nutritional value, Timothy

is commonly used for feeding livestock, either alone or in combination with other grasses and legumes (Niemeläinen et al. 2004), and is offered fresh or preserved as hay or silage (Bélanger et al. 2001). In addition, in Finland, Timothy has been used as a cover crop to reduce bare soil conditions after the production of cash crops such as oats, barley, and peas (Peltonen-Sainio et al. 2022). As such, Timothy cultivation plays an important economic and environmental role in Finnish production systems.

With the intent of improving economically important traits such as yield, digestibility, winter survival, and disease resistance, selection strategies for Timothy have focused on traditional methods, such as progeny testing in multiple environments (Bélanger et al. 2001). Therefore, development of new varieties has been a lengthy and an expensive process. The use of genomic information combined with performance records in the framework of genomic selection

Communicated by Hai-Chun Jing.

✉ N. Vargas Jurado
napoleon.vargas@luke.fi

¹ Natural Resources Institute Finland, Luke, 31600 Jokioinen, Finland

² Boreal Plant Breeding Ltd., 31600 Jokioinen, Finland

(GS) provides an alternative to increase the rate of genetic change and accuracy of selection (Meuwissen et al. 2001). Timothy is a hexaploid ($2n=6x=42$), outcrossing, wind pollinated species (Cai and Bullen 1994; Tanhuanpää and Manninen 2012), and cultivars are usually open pollinated populations. However, its genomic constitution (autohexaploid vs. allohexaploid) is still unclear (Cai and Bullen 1994; Cai et al. 2003), and its reference genome is not yet available. This creates additional challenges for genomic prediction methods (e.g., genome-wide association studies) that rely on a map of the genome. However, genotype-by-sequencing (GBS) methods provide an alternative to obtain genomic information in a cost-effective manner (Gorjanc et al. 2015) without the explicit need for a reference genome. Still, challenges remain due to low sequencing depths or allelic biases which may affect predictions of genetic merit (Gerard et al. 2018; de Bem Oliveira et al. 2020).

Knowledge about the genetic architecture of traits of economic importance in Timothy is limited. The heritability of several nutritive quality traits of Timothy clones, including organic matter digestibility and crude protein content, has been previously estimated (Ashikaga et al. 2008) for Japanese populations. However, for the local Finnish (or European) populations and for many other traits of economic importance such information is not yet available. Information about the genetic architecture of traits is not only important for selection decisions but estimates of variance components are also necessary for genomic evaluation. Moreover, determining relationships among traits, whether favorable or unfavorable, is crucial for the development of selection indices (Cerón-Rojas and Crossa 2018). Indices that combine estimates of the genetic merit of several traits along with economic weights can be used to increase rates of genetic gain for sets of traits with favorable correlations while simultaneously limiting changes in traits with unfavorable or antagonistic correlations (Batista et al. 2021). Also, indices can be tailored for specific environments thus making efficient use of the genetic resources available (Louw 1990).

Genomic best linear unbiased prediction models (GBLUP) using GBS data have been recently applied to Timothy (Kovi et al. 2019, 2021) and other grass species (Guo et al. 2018; Fiedler et al. 2018). Despite the successful application of genomic prediction, estimates of variance components or genomic breeding values (GEBV) from one population do not necessarily generalize to a different population. In addition, the accuracy of GEBV may depend on how the training and validation sets were selected (Kovi et al. 2021). Furthermore, when using GBS data in polyploid species, factors related to the quality of sequencing such as read depth and call rate, and bias due to overdispersion may affect the accuracy of predictions as the genomic relationship matrix may be impacted (Cericola et al. 2018; Guo et al.

2018). Therefore, validating GEBV resulting from genomic evaluations of a Finnish Timothy population where marker information is obtained from GBS data is of paramount importance. The objectives of the current study were to (i) estimate the narrow-sense heritability and genetic correlations for traits of economic importance in Finnish Timothy, and (ii) determine the accuracy of GEBV prediction by using cross-validation through a forward-prediction strategy.

Materials and methods

Sequencing and genotypes

Variant data was obtained by LGC Genomics, Berlin, Germany, using double digest restriction associated DNA sequencing (ddRAD), a Genotyping-by-sequencing (GBS) method to reduce the complexity of the genome often used for non-model species with no prior or little genomic knowledge (Peterson et al. 2012). However, due to the lack of a published reference genome for Timothy, a “mock” reference genome was constructed from the ddRAD data from six individuals that represented diverse genetic backgrounds. These six samples were also used to determine the most efficient combination of restriction enzymes as measured by the number of variants produced. As such, it was determined that a sequencing depth of 1.5 million read pairs per sample was sufficient. The combination of restriction enzymes PstI-ApeKI resulted in the highest number of identified variants. After demultiplexing of the library groups, sequencing adapter remnants from all reads were clipped and reads with a final length < 20 bases were discarded. Following this step, a quality trimming was applied, where reads containing “N” bases were removed and reads were trimmed at the 3'-end to a minimum Phred quality score of 20 over a window of ten bases. The quality trimmed reads were then aligned against the cluster reference using Bowtie 2 (Langmead and Salzberg 2012; version 2.2.3) and variants were called with FreeBayes (Garrison and Marth 2012; version 1.0.2–16) where further filtering criteria were applied (Supplemental File). The resulting variant information was provided in the variant call format (VCF) text file.

A total of $n = 1764$ mother clones were genotyped using the ddRAD method described above. Out of these, 107 mothers were genotyped multiple times (two to five) such that the repeated samples could be used to provide a measure of the consistency of the sequencing approach. Read depth information extracted from the associated VCF files included the number of reads for the reference (R_A) and alternate (R_B) alleles for each of the genotyped individuals. In addition, a single nucleotide polymorphism (SNP) marker was considered missing if both R_A and R_B were zero. In contrast to SNP chip data, where discrete allele counts are available

(i.e., allele dosage), the current study used allele frequencies instead. Thus, genotypes were defined as (Guo et al. 2018):

$$f_{ij} = \frac{R_{A_{ij}}}{R_{A_{ij}} + R_{B_{ij}}} \quad (1)$$

where f_{ij} was the estimated allele frequency of the reference allele for individual i ($i = 1, \dots, n$) and marker j ($j = 1, \dots, m$), and $R_{A_{ij}}$ and $R_{B_{ij}}$ were the read depths of the reference and alternate alleles, respectively. As defined in (1), f_{ij} were continuous variables (or continuous genotypes) on the interval $[0, 1]$. The number of SNP markers (m) depended on the filtering criteria described below. Also, in cases where an individual was sequenced multiple times, their allele frequencies were averaged into one allele frequency.

Due to the use of the ratio in (1), small or large values of the number of reads may lead to biased estimates of allele frequency and thus impact genomic predictions. As such, the impact of different filtering criteria on predictions was tested. In addition to minor allele frequency (MAF) ≥ 0.05 , markers were filtered (removed) based on their average minimum and maximum number of total reads (R_T), and four levels of call rate (100%, 90%, 80%, and 70%). The minimum and maximum average number of reads were set to 10 or 20 and 100 or 200, respectively. Both criteria affected the number of markers available for later use in genomic prediction (Table 1). Before filtering and across all samples there were a total of 280 808 common markers. For this set of markers, call rate ranged from 100.0 to 2.6% with a mean (SD) of 72.4% (24.4%), and there were 222 063 markers (or 79.1%

of all markers) with a call rate of at least 70%. Missing SNPs for an individual were imputed using the marker average.

After filtering based on MAF, a minimum and maximum read depth of 10 and 200, respectively, and for a call rate of 70% the number of markers was $m = 58,816$. Averaged over markers, the mean total read depth was 37.88 (SD 33.09), ranging from 4.92 to 220.72, while the mean allele frequency was 0.77 (SD 0.20). The distribution of read depth and allele frequencies for the genotyped mother clones is presented in Fig. 1. In addition, the correlation between repeated samples for mother clones genotyped multiple times was, on average, 0.85 (SD 0.10).

Genomic relationships and scaling factors

The genomic relationship matrix ($\mathbf{G}; n \times n$) was constructed as (Van Raden 2009):

$$\mathbf{G} = \frac{(\mathbf{F} - \mathbf{1}_n \bar{\mathbf{f}}')(\mathbf{F} - \mathbf{1}_n \bar{\mathbf{f}}')'}{h} \quad (2)$$

where $\mathbf{F}$ was the n by m matrix of genotypes, $\bar{\mathbf{f}} = [\bar{f}_1, \dots, \bar{f}_m]'$ was an m by 1 vector of mean allele frequencies, $\mathbf{1}_n$ was an n by 1 vector of ones, and h was a scaling factor. In the current work three scaling factors and their impact on $\mathbf{G}$ were studied (i) a naïve approach with a scaling factor derived from allele frequencies, (ii) the de-biasing strategy of Cericola et al. (2018), and (iii) a novel approach based on the Beta-Binomial distribution. The naïve approach used a scaling factor defined as (Guo et al. 2018):

$$h = \frac{1}{k} \sum_{j=1}^m \bar{f}_j (1 - \bar{f}_j) \quad (3)$$

where k was the ploidy of an individual (assumed to be 6 for Timothy) and $\bar{f}_j$ was the mean allele frequency of marker j . Due to the low number of reads or bias in the sequencing reads, using (3) as a scaling factor may inflate the diagonal elements of $\mathbf{G}$ and subsequently the resulting GEBV. The de-biasing strategy by Cericola et al. (2018), on the other hand, addressed this inflation by assuming that the reads for a given allele followed a Binomial distribution:

$$R_{A_{ij}} \sim \text{Binomial}(R_{T_{ij}}, p_j)$$

where i indexed over individuals and j indexed over markers, $R_{A_{ij}}$ was the number of reads for the reference allele, $R_{T_{ij}} = R_{A_{ij}} + R_{B_{ij}}$ was the total number of reads, and p_j was the probability of sampling the reference allele, assumed to be constant. Under these assumptions, the mean and variance of the sequencing reads were $\mathbb{E}[R_{A_{ij}}] = R_{T_{ij}} p_j$ and

Table 1 Number of markers according to different filtering criteria including, minimum and maximum average read depth (RD), and call rate (%)

Minimum RD	Maximum RD	Call rate (%)	Markers
20	100	100	2883
		90	26,452
		80	29,587
		70	30,692
		100	4525
	200	90	29,353
		80	32,488
		70	33,593
		100	4529
		90	41,053
10	100	80	51,703
		70	57,298
	200	100	4529
		90	41,053
	100	80	51,703
		70	57,298
	200	100	4529
		90	41,053

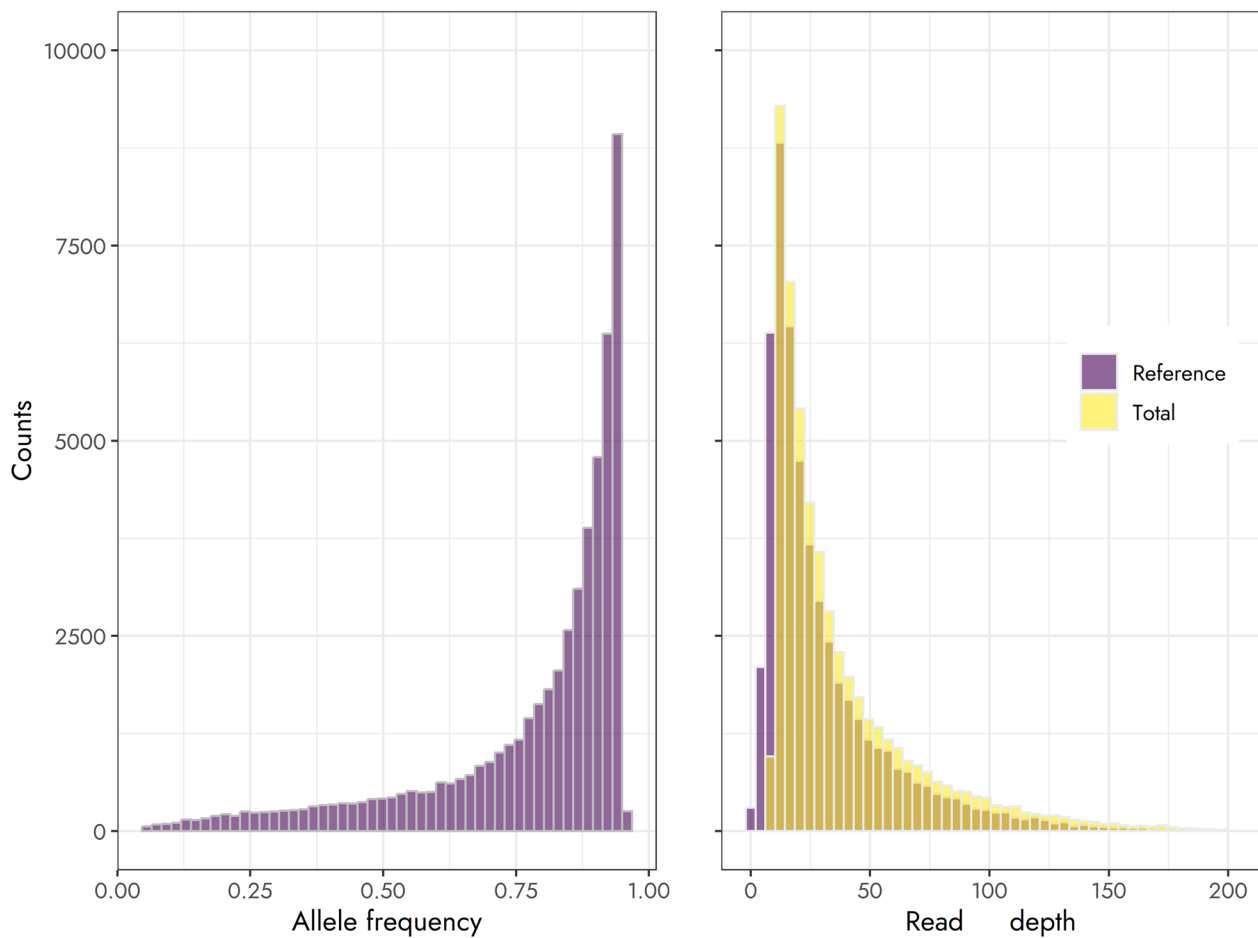


Fig. 1 Distribution of allele frequency (for the reference allele) and read depth for Timothy (*Phleum pratense* L.) clones obtained from genotyping-by-sequencing data

$\mathbb{V}\text{ar}[R_{A_{ij}}] = R_{T_{ij}}p_j(1 - p_j)$ such that the mean and variance of the resulting allele frequencies $f_{ij} = R_{A_{ij}}/R_{T_{ij}}$ were $\mathbb{E}[f_{ij}] = p_j$ and $\mathbb{V}\text{ar}[f_{ij}] = p_j(1 - p_j)/R_{T_{ij}}$, respectively. Using these moments and a Normal distribution approximation, Cericola et al. (2018) proposed to correct the diagonals of $\mathbf{G}$ by $G_{ii} = \tilde{G}_{ii}(1 - w_i)$, where $\tilde{G}_{ii}$ and G_{ii} were the biased and the corrected i^{th} diagonal element of $\mathbf{G}$, respectively, $w_i = (k - 1)/(\bar{R}_{T_i} + k - 1)$ was an adjustment factor, and $\bar{R}_{T_i}$ was the average (total) read depth for an individual (averaged across markers).

In the current study a novel approach, denoted as the Beta-Binomial model, was considered which addressed the bias (overdispersion) in GBS data by assuming that sequencing reads at locus j can instead be modeled using the following hierarchy:

$$\begin{aligned} R_{A_{ij}}|p_{ij} &\sim \text{Binomial}(R_{T_{ij}}, p_{ij}) \\ p_{ij} &\sim \text{Beta}(\mu_j, \tau_j) \end{aligned} \quad (4)$$

where μ_j and τ_j were location and dispersion parameters of the Beta distribution, respectively, such that $\mathbb{E}[p_{ij}] = \mu_j$ and $\mathbb{V}\text{ar}[p_{ij}] = \mu_j(1 - \mu_j)/\tau_j$. Above, μ_j can be considered the underlying “true” allele frequency for locus j . As specified in (4), p_{ij} could vary across individuals but the location and dispersion parameters were common to all individuals at a given locus. Conditional on p_{ij} , the expected value and variance of the sequencing reads were the same as in those for the Binomial model. However, for the hierarchical model (4), the marginal mean and variance of the sequencing reads were instead $\mathbb{E}[R_{A_{ij}}] = R_{T_{ij}}\mu_j$ and $\mathbb{V}\text{ar}[R_{A_{ij}}] = \mu_j(1 - \mu_j)[1 + (R_{T_{ij}} - 1)\tau_j]$ which can be rec-

ognized as the first two moments of a Beta-Binomial distribution. Similarly, the mean and variance of the estimated allele frequencies for the Beta-Binomial model were given by $\mathbb{E}[f_{ij}] = \mu_j$ and $\text{Var}[f_{ij}] = \mu_j(1 - \mu_j) \left[1 + (\bar{R}_{T_j} - 1)\tau_j \right] / \bar{R}_{T_j}$, respectively. Thus, the assumed uncertainty in the underlying frequency of a marker (p_{ij}) was propagated to the estimated allele frequency (f_{ij}) as observed by a larger variance.

The scaling factor for **G** proposed in the current work was then given by:

$$h = \sum_{j=1}^m \bar{f}_j(1 - \bar{f}_j) \frac{\left[1 + (\bar{R}_{T_j} - 1)\hat{\tau}_j \right]}{\bar{R}_{T_j}} \quad (5)$$

where the mean allele frequency for marker j ($\bar{f}_j$) was used as an estimate of the location parameter μ_j , $\hat{\tau}_j$ was an estimate of the overdispersion parameter, and $\bar{R}_{T_j}$ was the mean read depth at locus j , averaged across individuals. The scaling factor (5) can be written more generally as

$h = \sum_{j=1}^m \bar{f}_j(1 - \bar{f}_j)\omega_j$ where $\omega_j = \left[1 + (\bar{R}_{T_j} - 1)\hat{\tau}_j \right] / \bar{R}_{T_j}$ was a heterogeneity factor (Williams 1982). When $\hat{\tau}_j$ was zero, ω_j was equal to $1/\bar{R}_{T_j}$ and its value decreased for increasing $\bar{R}_{T_j}$. On the other hand, if $\hat{\tau}_j > 0$ and $\bar{R}_{T_j}$ was large, then ω_j was almost entirely determined by the overdispersion parameter τ_j (Fig. 2). Unlike the de-biasing approach of Cericola et al. (2018), where the adjustment factor was independent of the allele frequency, the scaling factor (5) depends on the overdispersion factor estimated at each locus. Also, note that in (5) the ploidy number is absent because it is a function of the heterogeneity factor. Estimates of τ_j were obtained by maximizing the logarithm of the Beta-Binomial likelihood using the optim function in R (R Core Team 2024) with the L-BFGS-B algorithm. Other options for estimating the overdispersion parameter include the quasi-likelihood approach (weighted least squares) of Williams (1982) or Bayesian approaches.

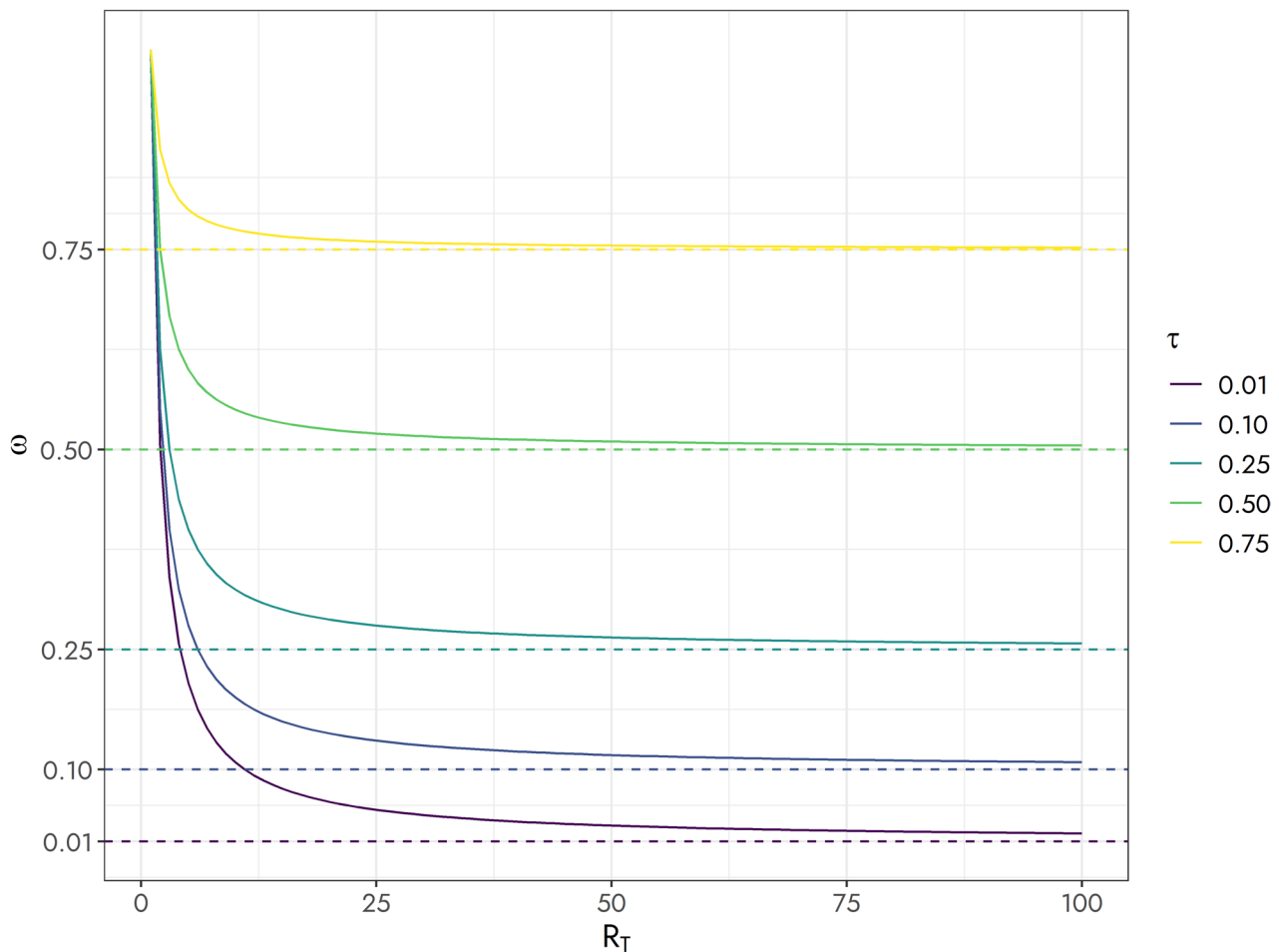


Fig. 2 Heterogeneity factor (ω) for the Beta-Binomial model as a function of total read depth (R_T) and overdispersion parameter (τ)

Field trials and performance records

Timothy (*Phleum pratense* L.) lines evaluated were developed and provided by Boreal Plant Breeding Ltd. The data contained 11,375 performance records corresponding to the progeny of the 1764 genotyped mother clones collected from 2001 to 2023, with an average of approximately 495 records per year. Between 1 and 4 progeny crosses corresponding to approximately 194 (SD 101) mothers were evaluated each year and in an average of 4 locations including southern and northern Finland, Czech Republic, Latvia, and Canada. Progeny lines were the result of top- and poly-crosses. In a top-cross a every mother clone is intermated to the same pollinator variety or other population, while in a poly-cross every mother clone is intermated to all other mother clones in the poly-cross field.

The traits included yield at first, second, and third cut (kg DM/ha), winter damage (%), and digestible organic matter in dry matter (D-value; %; Nousiainen et al. 2003) at first and second cut. Test plots were harvested using an experimental harvester where grass yield was measured, and representative samples were taken for laboratory analysis. Cuttings were made in silage stage, at first cut in growth stages 3.1–3.3 according to Moore (Moore et al. 1991). In Finland the first cut was normally made between 10 and 17th of June, the second cut six weeks after the first cut and the third cut in late August/beginning of September. Winter damage was defined as the proportion of deceased individuals in a test plot in the early spring on the second and third year after sowing. D-value was measured with FOSS DS 2500 near-infrared spectroscopy (NIRS) analyzer. D-value of the calibration samples was calculated as $(100 - \text{ash content}) \times \text{OM digestibility}$, where OM (organic matter) digestibility was assessed by cellulase digestibility, and the ash content of the samples was determined as the proportion of weight remaining after placing the samples in a furnace at 550 °C for 6 h (Rinne and Nykänen 2000).

Prior to this study, performance records were corrected for spatial effects. The trial design in those field trials was a row-column design with two replications. The statistical model used can be described as:

$$y_{rcl} = \mu + \theta_r + \phi_c + \eta_l + \epsilon_{rcl}$$

where μ was the overall mean, θ_r was the effect of row block r , ϕ_c was the effect of column block c , and η_l was the effect of treatment l . Statistical analyses were done with SPSS (IBM SPSS Statistics, version 27, 2019), and the best model was selected based on Akaike's information criterion (AIC). In addition, the performed quality control included removal of outliers. The best linear unbiased estimators obtained after pre-correction were used as the dependent variable (traits) in the subsequent genomic prediction models. Summary

Table 2 Summary statistics including, number of records, mean, standard deviation, minimum, and maximum for Timothy (*Phleum pratense* L.) traits used in the genomic evaluation

Trait	<i>N</i>	Mean	SD	Min	Max
Yield at 1st cut (kg/ha)	11,192	6460.7	2511.7	354.0	18,476.0
Yield at 2nd cut (kg/ha)	11,214	4159.8	1634.2	269.0	11,224.0
Yield at 3rd cut (kg/ha)	7745	2623.8	978.4	404.0	6675.0
Winter damage (%)	3052	12.1	12.1	−3.0	79.0
D-value at 1st cut (%)	7383	68.2	3.3	54.0	77.9
D-value at 2nd cut (%)	7007	67.6	4.4	53.0	78.9

statistics including mean, SD, and number of records for these traits are provided in Table 2.

Estimation of variance components

The following multiple-trait GBLUP model was used to estimate variance components for the traits mentioned above:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}_m\mathbf{m} + \mathbf{Z}_c\mathbf{c} + \mathbf{e}, \quad (6)$$

where $\mathbf{y} = [\mathbf{y}_1, \dots, \mathbf{y}_T]'$ was the vector of phenotypic records with traits stacked, $\mathbf{b} = [\mathbf{b}_1, \dots, \mathbf{b}_T]'$ was the vector of fixed effects including trial by year interaction (210 levels), $\mathbf{m} = [\mathbf{m}_1, \dots, \mathbf{m}_T]'$ was the vector of random genetic effects (GEBV) for mother clones, $\mathbf{c} = [\mathbf{c}_1, \dots, \mathbf{c}_T]'$ was the random vector of permanent environmental effects (progeny ID within trial, 5601 levels), and $\mathbf{e} = [\mathbf{e}_1, \dots, \mathbf{e}_T]'$ was the residual. The block-diagonal matrices $\mathbf{X} = \bigoplus_{i=1}^T \mathbf{X}_i$, $\mathbf{Z}_m = \bigoplus_{i=1}^T \mathbf{Z}_{m_i}$, and $\mathbf{Z}_c = \bigoplus_{i=1}^T \mathbf{Z}_{c_i}$ related records to the fixed, additive, and permanent environmental effects, respectively, and $\bigoplus$ denotes the direct sum of matrices. In addition, $T=6$ was the number of traits. Furthermore, it was assumed that $\mathbf{m} \sim N(\mathbf{0}, \mathbf{G} \otimes \mathbf{V}_M)$, $\mathbf{c} \sim N(\mathbf{0}, \mathbf{I} \otimes \mathbf{V}_C)$, and $\mathbf{e} \sim N(\mathbf{0}, \mathbf{I} \otimes \mathbf{V}_E)$ where $\otimes$ denotes the Kronecker (direct) product, $\mathbf{G}$ was the genomic relationship matrix calculated using the Beta-Binomial scaling factor (defined in (2) and (5)), $\mathbf{V}_M$, $\mathbf{V}_C$, and $\mathbf{V}_E$ were the 6 by 6 mother, permanent, and residual (co)variance matrices among traits, respectively, and $\sigma_{M_k}^2$, $\sigma_{C_k}^2$, and $\sigma_{E_k}^2$ correspond to their k^{th} diagonal element. Variance components for model (6) were estimated using genomic Monte-Carlo (MC) Expectation–Maximization REML (EM-REML) in MiX99 (Vuori et al. 2006; Mätiläinen et al. 2019) with 10 MC samples per round. Convergence was assumed when the round-to-round change in the variance component estimates was less than 1×10^{-11} . Standard errors for (co)variance parameters were calculated with 200 additional MC samples after convergence.

Narrow-sense heritability (h_k^2) and repeatability (c_k^2) for trait $k = 1, \dots, 6$ were calculated as (Legarra 2016; Becker 1992):

$$h_k^2 = \frac{\overline{\text{diag}}(\mathbf{G}) \times \sigma_{A_k}^2}{\sigma_{M_k}^2 \times \overline{\text{diag}}(\mathbf{G}) + \sigma_{C_k}^2 + \sigma_{e_k}^2}, c_k^2 = \frac{\overline{\text{diag}}(\mathbf{G}) \times \sigma_{A_k}^2 + \sigma_{C_k}^2}{\sigma_{M_k}^2 \times \overline{\text{diag}}(\mathbf{G}) + \sigma_{C_k}^2 + \sigma_{e_k}^2}$$

where $\sigma_{A_k}^2 = 4\sigma_{M_k}^2$ corresponded to the additive variance for trait k and $\overline{\text{diag}}(\mathbf{G})$ was the average of the diagonal elements of $\mathbf{G}$ calculated using the Beta-Binomial method. Given that the mother variance ($\sigma_{M_k}^2$) explains only $\frac{1}{4}$ of the genetic variance of the progeny, it was multiplied by 4 (Becker 1992) to obtain the additive variance ($\sigma_{A_k}^2$). In addition, phenotypic and genetic correlations among the traits described above were calculated as:

$$r_{l_{jk}} = \frac{\sigma_{l_{jk}}}{\sqrt{\sigma_{l_{jj}}^2 \sigma_{l_{kk}}^2}}$$

where $r_{l_{jk}}$ was the correlation between traits j and k for the l effect (for $l=A, P$ for additive and phenotypic, respectively), and the additive variance and phenotypic variances were defined as $\mathbf{V}_A = 4\mathbf{V}_M$ and $\mathbf{V}_P = \mathbf{V}_M + \mathbf{V}_C + \mathbf{V}_E$, respectively

Validation of GEBV

Estimation of GEBV was performed using both single-trait and multi-trait GBLUP models (6). The single-trait model was similarly defined as:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}_m\mathbf{m} + \mathbf{Z}_c\mathbf{c} + \mathbf{e}, \quad (7)$$

where $\mathbf{y}$ was the vector of phenotypic records for a given trait, $\mathbf{X}$, $\mathbf{Z}_m$, and $\mathbf{Z}_c$ corresponded to the design matrices for the fixed, mother, and permanent environmental effects, respectively. In addition, $\mathbf{m} \sim N(\mathbf{0}, \sigma_M^2 \mathbf{G})$, $\mathbf{c} \sim N(\mathbf{0}, \mathbf{I}\sigma_C^2)$, and $\mathbf{e} \sim N(\mathbf{0}, \mathbf{I}\sigma_e^2)$ were, as before, the mother, permanent environmental, and residual effects, the variances σ_M^2 , σ_C^2 , and σ_e^2 were the appropriate diagonal element of their corresponding matrix ($\mathbf{V}_M$, $\mathbf{V}_C$, or $\mathbf{V}_E$) estimated from the multiple-trait model (6), and $\mathbf{I}$ was an identity matrix. In the current work, prediction, and validation of GEBV was not performed for specific environments or locations but across all locations. However, to account for potential differences in performance among environments (or locations), both models (6) and (7) included the trial by year and the permanent environmental effects, which incorporate trial location in their definition. Because the $\mathbf{G}$ matrix scaled using the naïve and Beta-Binomial model are the same up to a constant multiplier (e.g., all elements of $\mathbf{G}$ are multiplied by a scalar) no differences are expected in terms of estimates of variance components or the corresponding GEBV. On the other hand, for the debiased approach potential differences may arise. Thus, such impacts are considered in subsequent analyses.

GEBV were validated using two strategies: (i) a family cross-validation, and (ii) a forward-prediction strategy. For both (i) and (ii) two sets of individuals were defined: the estimation set which corresponded to mother clones whose progeny records were used for estimating GEBV using models (6) and (7), and a validation set which corresponded to mother clones whose progeny records were not used in the evaluation but instead were used for calculating predictive ability (and other accuracy measures). However, the genomic information of all mother clones was used in the estimation of GEBV such that the predicted GEBV for individuals in the validation set were derived from the genomic relationships with individuals in the estimation set. For (i), a total of 18 half-sib progeny families were defined such that there were at least 50 mothers, and their progeny, in each family. Thus, there were 18 individual evaluations where the performance records corresponding to the progeny of the mother clones in a given family were not used to estimate their GEBV. The number of mother clones in the validation set ranged from 54 to 262 representing approximately 5% and 23% of all genotyped mothers, and were associated with 48 and 251 individual progenies, respectively. For (ii), on the other hand, performance records from the last four years of recording (2020, 2021, 2022, and 2023) were considered part of the validation set and thus not used in the genomic evaluation (estimation of GEBV). This resulted in 8217 phenotypic records for the estimation set and 3158 for the validation set. That is, about 27% of the records were removed. A total of 338 mother clones were identified that had records in the full data set but did not have records in the estimation set.

For the mothers in the validation set, predictive ability was defined as the correlation between GEBV and mean progeny performance, adjusted for fixed and permanent environmental effects. The impact of the marker filtering criteria previously described on predictive ability was tested by constructing $\mathbf{G}$ with the marker sets shown in Table 1 and fitting the model (6) or (7). In addition, the validation method described in Legarra and Reverter (2018) was used, where the full data set consisted of the progeny records of all genotyped mothers, while the reduced data set was defined as the progeny records from mothers remaining in the evaluation after removing mothers in a validation group. For the full and reduced sets, GEBV were obtained, and the following criteria were calculated: (i) the correlation between GEBV in the full and reduced sets, and (ii) the slope (or dispersion) and (iii) the coefficient of determination from the regression of GEBV from the full set on the GEBV from the reduced set. To avoid confusion with overdispersion in the read depth of SNPs, the terms over- and under-prediction are used to refer to over- and -under-dispersion of GEBV.

Results

Scaling factor

Averaging across the marker sets (Table 1), the mean of the diagonal elements of $\mathbf{G}$ was 2.32 ± 0.09 , 2.06 ± 0.07 , and 1.26 ± 0.01 when scaled using the naïve, debiased, and Beta-Binomial approaches, respectively. In addition, the average of the diagonals of $\mathbf{G}$ increased along with the number of markers for the naïve and debiased approaches. This was less the case when $\mathbf{G}$ was scaled using the Beta-Binomial scaling factor (Fig. 3).

Across all markers, the mean estimate of the overdispersion parameter $\hat{\tau}$ was 0.26 (SD 0.16), with individual values ranging from 0.00 to 0.76. In the same manner, the mean heterogeneity factor ($\hat{\omega}$) was 0.30 (SD 0.15) with values ranging from 0.01 to 0.77. In addition, the correlation between $\hat{\omega}$ and $\hat{\tau}$ was 0.99.

Estimates of variance components

Estimates of additive variance were $60,342.8 \pm 12,184.4$, $102,520.8 \pm 10,432.4$, and $21,229.6 \pm 3912.0$, for yield at first, second, and third cuts, respectively, 6.60 ± 2.28 for winter damage, and 0.72 ± 0.08 and 1.24 ± 0.12 for D-value at first and second cuts, respectively. Similarly, estimates of permanent environmental variances were $79,075.3 \pm 9021.0$, $29,488.8 \pm 3685.9$, and $11,370.6 \pm 1988.1$, for yield at first, second, and third cuts, respectively, 5.88 ± 1.65 for winter damage, and 0.03 ± 0.02 and 0.10 ± 0.03 for D-value at first and second cuts, respectively. Likewise, estimates of residual variances were $482,725.8 \pm 10,220.0$, $179,893.3 \pm 4114.8$, and $82,912.9 \pm 2319.9$ for yield at first, second, and third cuts, respectively, 35.44 ± 1.84 for winter damage, and 0.89 ± 0.02 and 1.32 ± 0.04 for D-value at first and second cuts, respectively. Lastly, favorable genetic covariances were found among yield traits, ranging from 2331.4 ± 1161.7 to 5256.7 ± 1144.9 , between cuts 1 and 3, and between cuts 2 and 3, respectively. Similarly, a smaller covariance of 0.14 ± 0.02 was found among D-value traits. On the other

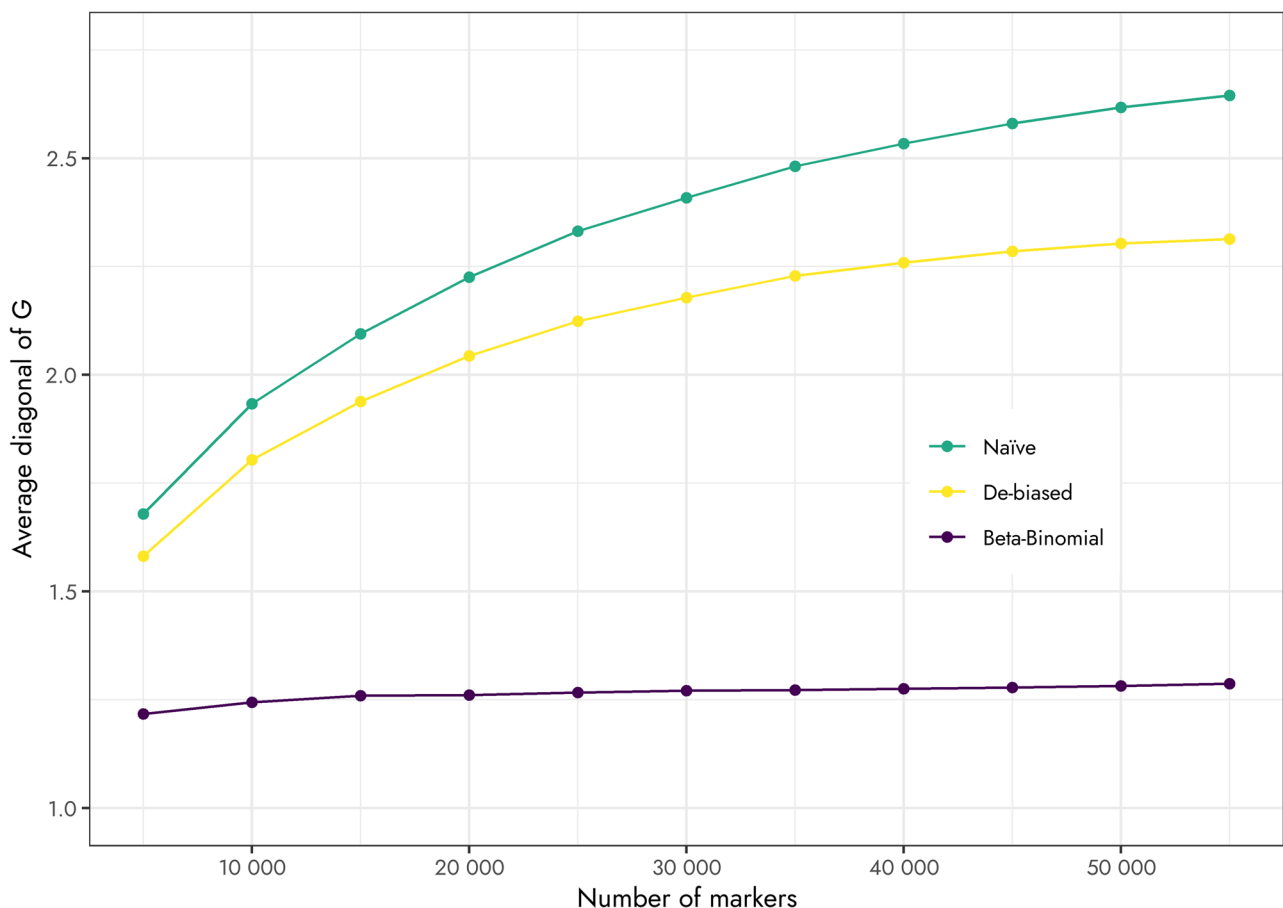


Fig. 3 Average diagonal values of $\mathbf{G}$ scaled under naïve, debiased, and Beta-Binomial model assumption for Timothy clones (*Phleum pratense* L.) sequenced using genotyping-by-sequencing

hand, negative covariances ranging from -114.3 ± 31.0 to -0.04 ± 0.01 were found between winter damage and yield at first cut, and between winter damage and D-value at first cut, respectively.

Estimates of heritability ranged from 0.13 ± 0.02 for yield at first cut to 0.86 ± 0.05 for D-value at second cut (Table 3). Estimates of genetic correlations among traits ranged from -0.72 ± 0.12 between yield at first cut and winter damage to 0.59 ± 0.04 between D-value at first and second cut. Phenotypic correlations, on the other hand were smaller in magnitude, ranging from -0.45 ± 0.02 between yield at first cut and winter damage to 0.23 ± 0.01 between yield at second and third cuts.

Estimates of repeatability were 0.27 ± 0.02 , 0.66 ± 0.04 , 0.38 ± 0.04 , 0.33 ± 0.05 , 0.83 ± 0.05 , and 0.91 ± 0.05 for yield at 1st, 2nd and 3rd cuts, winter damage, and D-value at 1st and 2nd cuts, respectively.

Accuracy of genomic prediction

Filtering markers based on minimum read depth had little effect on predictive ability for both the family cross-validation and forward-prediction strategies ($P=0.31$; Fig. 4; Supplementary Table S1). Averaging across traits, marginal means for predictive ability was 0.412 ± 0.002 and 0.418 ± 0.002 , and 0.324 ± 0.004 and 0.331 ± 0.004 for minimum read depth of 10 and 20, respectively, for the family cross-validation and forward-prediction strategies, respectively. This represented an increase in predictive ability, with respect to a minimum read depth of 10, of 1.4%, and 1.7%, respectively.

On the other hand, a larger effect was observed when filtering markers on maximum read depth ($P=0.014$; Supplementary Table S2). Overall, marginal means for predictive ability were 0.422 ± 0.002 and 0.408 ± 0.002 , and 0.332 ± 0.004 and 0.323 ± 0.004 for maximum read depth of 100 and 200, respectively, for the family cross-validation and forward-prediction strategies, respectively. These changes in predictive ability, with respect to maximum read

depth of 200, represented an increase of 3.4% and 2.8%, respectively. Similarly, call rate had a considerable effect on predictive ability ($P<0.01$; Supplementary Table S3). Marginal means for predictive ability were 0.348 ± 0.006 , 0.437 ± 0.006 , 0.444 ± 0.006 , and 0.432 ± 0.006 for call rates of 100%, 90%, 80%, and 70% for the family cross-validation approach. Similarly, for the forward-prediction strategy, marginal means for predictive ability were 0.280 ± 0.002 , 0.337 ± 0.002 , 0.348 ± 0.002 , and 0.345 ± 0.002 for call rates of 100%, 90%, 80% and 70%, respectively. The maximum change in predictive ability, compared to a call rate of 100%, represented an increase of 27% and 24% for the cross-validation and forward-prediction strategies, respectively.

Predictive ability decreased when using the debiased approach ($P<0.01$). Averaging over traits, marginal means for predictive ability were 0.30 ± 0.02 and 0.39 ± 0.02 for the debiased, and Beta-binomial (and naïve) approaches, respectively. This represented a 23% decrease in predictive ability. The decrease in predictive ability ranged from 14 to 48% for winter damage and yield at second cut, respectively.

Overall, multiple-trait models resulted similar predictive abilities ($P=0.61$). Averaging across traits, marginal means for predictive ability were 0.414 ± 0.004 and 0.417 ± 0.004 , and 0.325 ± 0.002 and 0.330 ± 0.002 for single- and multiple-trait models, respectively, for the family cross-validation and forward-prediction scenarios, respectively. However, the effect of single- and multi-trait models on predictive ability was, in general, not consistent across traits or validation strategies (Fig. 5; Supplementary Table S4). For the family cross-validation strategy, the use of multi-trait models increased the predictive ability for winter damage by 0.05 (or 27%), while it resulted in a slight decrease (less than 3%) for the remaining traits. For the forward-prediction strategy, on the other hand, fitting the multi-trait GBLUP increased the predictive ability for yield at the second cut, and D-value at the first and second cuts by 0.033, 0.030, and 0.023, respectively.

Additional validation criteria, including correlations between GEBV for the full and reduced sets, slope (or

Table 3 Estimates of heritability (diagonal, bold), genetic correlations (lower triangle), and phenotypic correlations (upper triangle) for yield traits (kg/ha), winter damage (%), and D-value (%) in Timothy (*Phleum pratense* L.).

Trait	Yield at 1st cut	Yield at 2nd cut	Yield at 3rd cut	Winter damage	D-value at 1st cut	D-value at 2nd cut
Yield at 1st cut	0.13 (0.03)	0.10 (0.01)	0.11 (0.01)	-0.45 (0.02)	-0.17 (0.01)	0.11 (0.01)
Yield at 2nd cut	0.13 (0.08)	0.53 (0.04)	0.23 (0.01)	-0.01 (0.02)	-0.04 (0.01)	-0.29 (0.01)
Yield at 3rd cut	0.26 (0.10)	0.45 (0.06)	0.26 (0.04)	-0.01 (0.03)	0.01 (0.01)	-0.05 (0.02)
Winter damage	-0.72 (0.12)	0.37 (0.11)	-0.09 (0.14)	0.19 (0.05)	0.19 (0.02)	-0.15 (0.02)
D-value at 1st cut	-0.42 (0.08)	-0.26 (0.04)	0.03 (0.07)	-0.08 (0.10)	0.80 (0.06)	0.12 (0.03)
D-value at 2nd cut	0.04 (0.08)	-0.64 (0.06)	-0.08 (0.08)	-0.52 (0.11)	0.59 (0.04)	0.86 (0.05)

Parameter estimates obtained from a multiple-trait model with the genomic relationship matrix scaled using the Beta-Binomial approach. Standard errors of estimates in parentheses

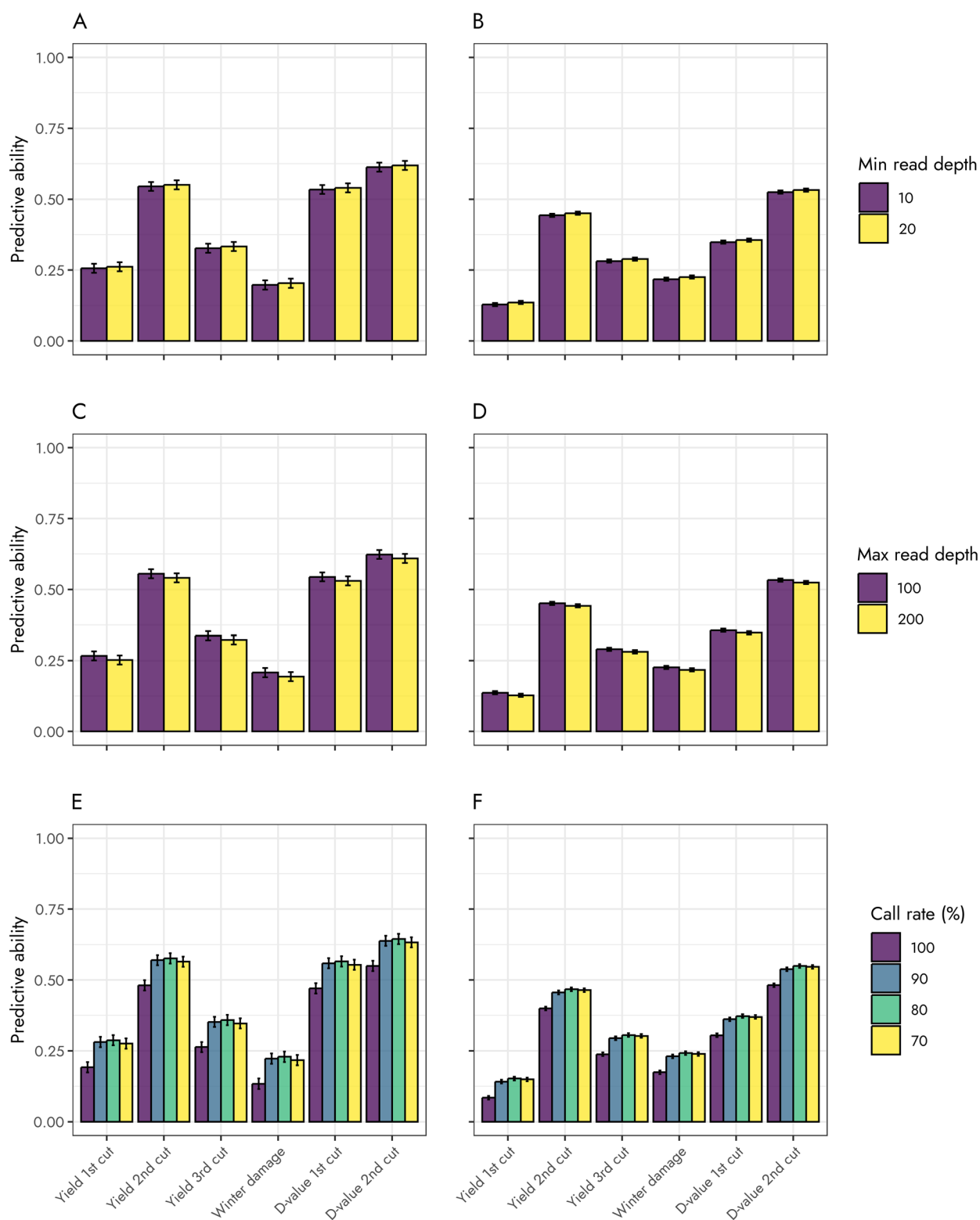


Fig. 4 Marginal means for the effect of minimum (top row) and maximum (middle row) read depth, call rate (bottom row) and validation strategy on predictive ability (correlation between genomic breeding value and mean progeny phenotypes adjusted for fixed and permanent environmental effects), averaged over type of model (single- vs multi-

ple-trait), for yield traits (kg/ha), winter damage (%), and D-value (%) in Timothy (*Phleum pratense* L.). Marginal means for family cross-validation in the left column (A, C, E) and for forward prediction on the right column (B, D, F)

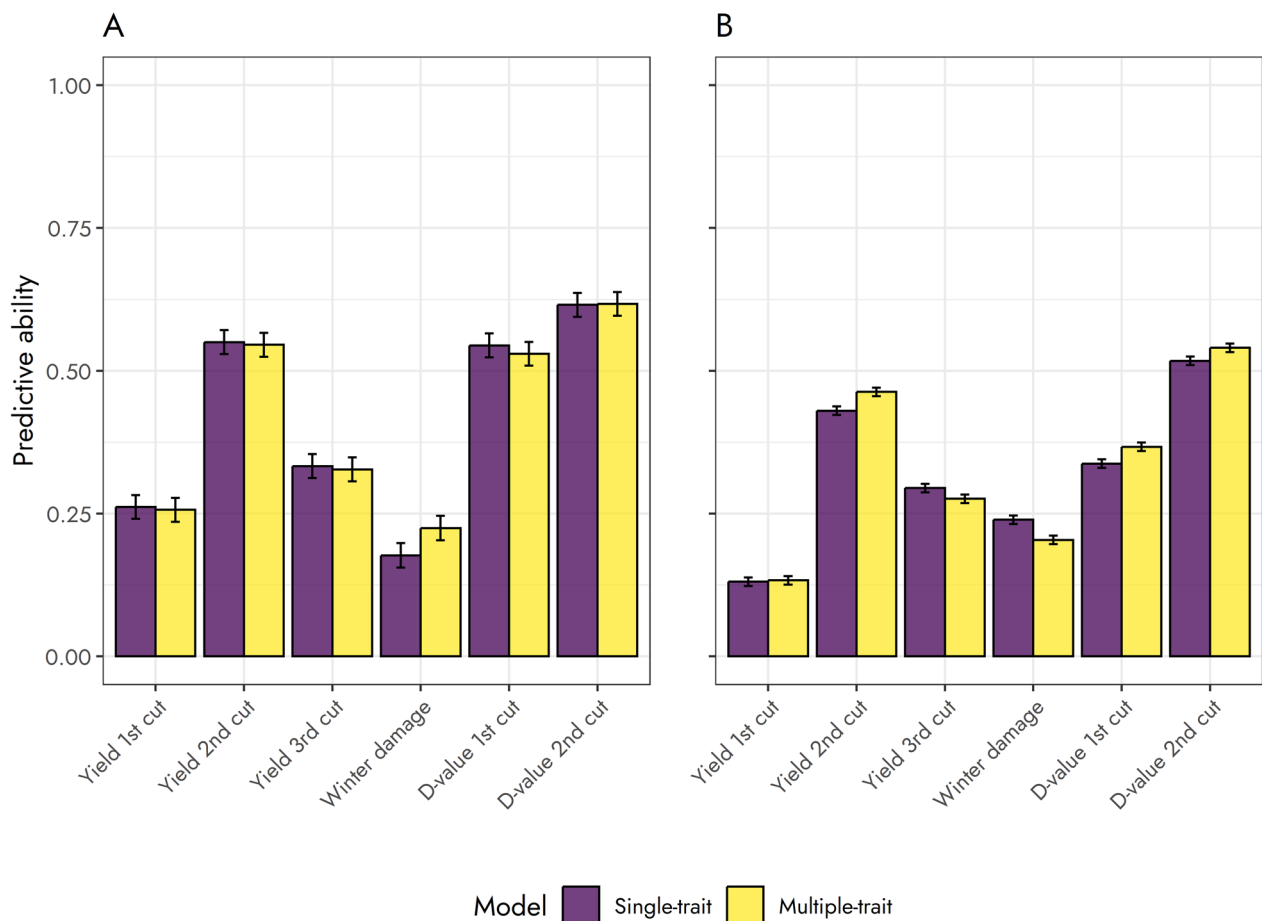


Fig. 5 Marginal means for the effect of type of model (single- and multiple-trait) and validation strategy (family cross-validation (A) and forward prediction (B)) on predictive ability (correlation between genomic breeding value and mean progeny phenotypes adjusted for

fixed and permanent environmental effects), averaged over minimum and maximum read depth, and call rate, for yield traits (kg/ha), winter damage (%), and D-value (%) in Timothy (*Phleum pratense* L.)

dispersion, β_1), and coefficient of determination (R^2) from regression of GEBV from the full on the reduced set, for the family cross-validation and the forward-prediction strategies are presented in Table 4. Overall, the correlations were higher for the forward-prediction strategy and for single-trait GBLUP. Multi-trait GBLUP increased predictive ability only for yield at second cut and D-value at first cut and only in the forward-prediction scenario. On the other hand, GEBV accuracy decreased when fitting multi-trait GBLUP, with the largest change in accuracy for D-value at second cut in the family cross-validation scenario.

Dispersion was small, in general, and not consistent across traits or validation strategies, but was less pervasive for single-trait GBLUP models. For the family cross-validation strategy, the largest over-prediction and under-prediction were observed for yield at second cut and winter damage, respectively. However, winter damage changed from under-prediction when using single-trait GBLUP to over-prediction when fitting multi-trait models. Moreover,

GEBV were over-predicted for all traits when using multi-trait GBLUP. For the forward-prediction strategy, GEBV for winter damage were under-predicted the most when fitting single-trait GBLUP, which was reduced when using multi-trait models. On the other hand, a large increase in over-prediction was observed for yield at third cut when fitting multi-trait GBLUP.

In the family cross-validation strategy, R^2 values from the regression of GEBV from the full set on the reduced set ranged from 0.20 to 0.52 for yield at second cut and winter damage, respectively. However, for the forward-prediction strategy the R^2 values ranged from 0.33 to 0.63 for yield at third cut and winter damage, respectively. Overall, the R^2 values were larger for the forward-prediction strategy and the single-trait GBLUP models. In the family cross-validation scenario, fitting multi-trait GBLUP reduced the R^2 values, while an increase in R^2 was observed only for the forward-prediction strategy for yield at second cut and D-value at first cut.

Table 4 Correlation between genomic breeding values (GEBV), slope (or dispersion; β_1), and coefficient of determination (R^2) for GEBV from full and reduced data sets for family cross-validation, and forward-prediction strategies for yield traits (kg/ha), winter damage (%), and D-value (%) in Timothy (*Phleum pratense* L.)

Model	Trait	Family cross-validation			Forward prediction		
		Correlation ^a	β_1^b	R^{2b}	Correlation ¹	β_1^b	R^{2b}
Single-trait	Yield first cut	0.61	1.01	0.40	0.71	0.95	0.51
	Yield second cut	0.43	0.76	0.20	0.63	0.92	0.40
	Yield third cut	0.61	1.01	0.42	0.57	0.75	0.33
	Winter damage	0.70	1.10	0.52	0.79	1.37	0.63
	D-value first cut	0.56	0.92	0.34	0.55	0.96	0.30
	D-value second cut	0.62	0.96	0.42	0.73	0.99	0.53
Multiple-trait	Yield first cut	0.60	0.98	0.38	0.67	0.94	0.46
	Yield second cut	0.62	0.73	0.20	0.64	0.92	0.41
	Yield third cut	0.59	0.97	0.38	0.57	0.73	0.33
	Winter damage	0.57	0.90	0.35	0.76	1.12	0.58
	D-value first cut	0.52	0.88	0.29	0.59	1.00	0.35
	D-value second cut	0.50	0.82	0.28	0.69	1.03	0.48

^aPearson correlation coefficient between GEBV obtained from full and reduced data sets. ^bSlope and coefficient of determination from regressing GEBV obtained from a full model on GEBV obtained from the reduced model

Discussion

Genotypes

A similar distribution of read depth to the one presented in our study has been reported for ryegrass (Guo et al. 2018). Due to the skewed distribution of read depth, applying a minimum read depth criterion of 10 instead of 20 increased the number of available markers substantially more than choosing a maximum read depth criterion of 200 instead of 100. While markers at the lower end of the distribution may be affected by low coverage and allelic bias, those at the upper end of the distribution may be outliers that are also affected by allelic bias (Gerard et al. 2018). Therefore, restricting markers to a pre-specified range may be necessary to improve the quality of the available genotypes.

Scaling factor

To our knowledge, this is the first time that overdispersion has been estimated for GBS data and used to scale $\mathbf{G}$. The increase in the mean values of the diagonals of $\mathbf{G}$ for the naïve (Binomial) and the debiased approach by Cericola et al. (2018) suggest that the marker variance is considerably larger than $p_j(1 - p_j)$, and that overdispersion plays a considerable role when sequencing Timothy materials. Indeed, this has been reported by Gerard et al. (2018), where other systematic biases in addition to overdispersion, including sequencing error, allelic bias, and outliers, were modeled to increase the accuracy of genotyping for polyploid species. Similarly, Clark et al. (2019) used a Bayesian model to adjust SNPs for biases by accounting

for overdispersion and a contamination rate (i.e., sequencing error). Nevertheless, when calculating genomic relationships within a population, biases may have less of an impact when the number of markers is large (i.e., may be averaged out), such that GEBV and their ranking are less affected by the scaling of $\mathbf{G}$.

Because the scaling factors for the naïve and Beta-Binomial approaches are constants (i.e., a single scalar value for all individuals genotyped), the corresponding $\mathbf{G}$ matrices are also identical (up to a constant). As such, with proper scaling the corresponding estimates of variance and GEBV would also be identical. Such is not the case for the debiased approach as each element of the diagonal of $\mathbf{G}$ is adjusted (or corrected) differently. While the overall adjustment of the diagonal values of $\mathbf{G}$ for the debiased approach was small for this Timothy population, the corresponding GEBV were subsequently impacted.

While the main aim of developing the scaling factor was to address the inflation of the diagonals of $\mathbf{G}$ due to overdispersion in the GBS data, when properly scaled, $\mathbf{G}$ also provides information about inbreeding (i.e., genomic inbreeding; Villanueva et al. 2021). In our current work the average diagonal of $\mathbf{G}$ was 1.26, suggesting an inbreeding coefficient of 0.26 (or 26%). While little information related to inbreeding levels of Timothy is available, estimated inbreeding coefficients for ryegrass ranged from 0.27 to 0.65, depending on the type of individual (F1 vs twice selfing of F1 individuals; Harris et al. 2023). However, it should be noted that estimates of heterozygosity using GBS may be unreliable because a site may be regarded as homozygous simply because it was not sequenced sufficiently (Wang et al. 2020).

Estimates of variance components

To our knowledge, (co)variance estimates for traits of economic importance in Timothy have not been previously reported, further emphasizing the importance of our current work. On the other hand, variance estimates for dry matter yield and disease resistance has been reported for other grass species (Guo et al. 2018). However, due to differences in the definition of traits (e.g., kg/m² vs kg/ha) their reported variances differ substantially to the ones from our study. Thus, parameters from one population may not always (easily) translate from one population to another.

In addition to narrow-sense heritability, information related to broad-sense heritability has been reported for Timothy (Ashikaga et al., 2011) where estimates of genotypic variance were estimated from analysis of variance. Other strategies for estimating broad-sense heritability require estimation of variances due to genotype, year, location, and their interaction, or dominance (and epistatic) effects (Schmidt et al. 2019; Wittenburg et al. 2013). Due to unbalanced trials and the need for correcting for spatial effects, and lack of information on allele dosage in our study, estimates of broad-sense heritability could not be obtained.

Repeatability estimates indicate that the permanent environment of an individual contributes considerably to the variability observed in Timothy. Such contributions were more noticeable for traits with lower heritability (e.g., yield at first cut, winter damage). Because plots may be recorded multiple times in a trial, accounting for permanent environmental effects can help to better separate environmental and additive effects, thereby improving estimates of heritability.

The heritability of yield traits, ranging from 0.11 to 0.44, reflects a considerable environmental effect. For the first cut, performance is determined by how well individuals are established in a field, as well as by the effect of the previous winter after the first year of recording. For the second cut, on the other hand, environmental conditions would be more stable and homogeneous, such that performance depends more on the genetic potential of the individual. Finally, because not every location is suitable for a third cut, and because of changing weather conditions from late summer to fall, performance at this stage is again heavily influenced by the environment. Our estimates differ considerably from those reported by Ashikaga et al. (2016), who reported a broad-sense heritability of 0.79 and 0.91 for dry matter yield at first and second cuts, respectively. These discrepancies, however, likely reflect not only the different populations considered but also the methods used to obtain the estimates as Ashikaga et al. (2016) used partial least-squares regression for predictions and functions of the mean squares derived from analysis of variance to estimate the variance components, and the clones under evaluation had not been selected for nutritive value traits.

The estimate of heritability for winter damage was moderately low suggesting a strong environmental component, and the variability in winter conditions at the locations in which the Timothy materials were tested (e.g., Czech Republic and northern Finland). On the other hand, Ashikaga et al. (2016) reported a broad-sense heritability of 0.66 for winter survival, which was scored on a scale from 1 to 9, with 1 and 9 representing poor and good survival, respectively. Thus, in addition to population differences, winter damage and winter survival may represent different traits, and the differences in heritability may be expected.

Estimates of heritability for digestibility traits (D-value) were high for both cuts, suggesting a stronger genetic component, and the possibility of faster genetic progress for these traits. While not the same traits, Ashikaga et al. (2011) and Ashikaga et al. (2016) reported high heritability estimates for nutritional quality traits, including low-digestible fiber, organic cell wall, and water-soluble carbohydrate fractions, which are strongly associated with digestibility of plant material (Huhtanen and Krizsan 2023). Furthermore, their heritability estimates at the second cut were higher than those at the first cut, which was also the case in our study. Thus, despite potential environmental effects due to possible hot and dry summers, a considerable additive variance still influences digestibility traits in the Finnish Timothy population.

Estimates of additive variance (or heritability) coupled with selection intensity and cycle length (or generation interval), can be used to obtain knowledge about the response to selection (rate of gain) for different breeding strategies (Lush 1937; Cobb et al. 2019). While larger additive variances (and heritability) are associated, in general, with a larger response to selection, faster genetic gains can be achieved by decreasing cycle length (Cobb et al. 2019). Thus, given the estimates of additive variance for yield and digestibility in Timothy, using genomic selection to both reduce the cycle length and increase accuracy of selection can potentially increase the rates of genetic gain.

Positive correlations among yield traits (e.g., yield at second and third cuts), and among D-value traits (e.g., first and second cuts) indicate that selection for one of the traits is likely to have a positive impact on the remaining traits, potentially increasing the rate of genetic progress. Moreover, individual D-value traits were negatively correlated with yield traits (e.g., yield and D-value for the second cut), which is consistent with the results in Claessens et al. (2004) who reported a phenotypic correlation of -0.44 between dry matter yield and true in-vitro digestibility. These negative relationships between digestibility and dry matter yield have been previously reported for Timothy under northern climate conditions (Thorvaldsson et al. 2007; Hyrkäs et al. 2018). On the other hand, a positive genetic correlation between dry matter yield and winter hardiness, defined as the combined

effect of tillers per unit after winter damage, was found for ryegrass (Fè et al. 2015). These favorable genetic correlations among yield traits coupled with unfavorable genetic correlations between yield traits and digestibility, and between some yield traits and winter damage reflect the challenging relationships between these traits and environmental conditions. For instance, individuals affected by harsh winter conditions are likely to have poor performance at their first cut of the following season, and individuals who are still in the growing phases when winter conditions arrive are also likely to be impacted such that their subsequent performance next season will also be affected. In the same manner, high yielding individuals may have low digestibility. Thus, indices should be implemented for the selection of suitable materials that combine high yield, digestibility, and tolerance to winter conditions.

Genomic prediction and validation

Although substantially increasing (or decreasing) the number of available markers, filtering criteria related to minimum and maximum read depth had only a small effect on predictive ability. On the other hand, the larger effect observed due to call rate showed that genomic relationships were more accurately calculated with more markers, despite the potential effect bias and overdispersion. However, with a larger proportion of markers requiring imputation, a slight decrease in predictive ability was observed at the lowest level of call rate. In addition, the lack of a reference genome, lack of allele dosage information (i.e., allele counts), and our use of continuous genotypes (i.e., ratio of number of reads) limits the types of methods that can be used for increasing the accuracy of imputation.

While multi-trait models have shown improvement of predictive ability in ryegrass (Arojju et al. 2020), a clear advantage of multi-trait models was not observed in the current study. In our validations, no phenotypic records were available for genotyped individuals in the validation set, such that predictions of GEBV for these individuals were based entirely on the genomic relationships with other genotyped individuals in the estimation set, and on the genetic relationships among traits. Thus, uncertainty in the estimates of genetic correlations, along with collinearity may lead to poor predictive ability for multiple-trait models.

The discrepancy in predictive ability across traits for the two validation strategies is likely due to differences in the size and structure of the training and validation sets. For the family cross-validation strategy, records are removed (to a certain extent) uniformly across years. On the other hand, a considerable portion of the records were collected in the last three years (2020 to 2022) such that the predictive ability is more affected for the forward-prediction strategy. Overall, the predictive ability for yield traits was similar

to that reported by Kovi et al. (2021) for a second generation of full-sib families. These authors reported prediction accuracies ranging from 0.18 for yield at first cut to 0.54 for yield at second cut which are comparable to those found in the current work. Similar predictive abilities have also been reported for dry matter yield in switchgrass and ryegrass (Fiedler et al. 2018; Guo et al. 2018). On the other hand, the accuracy reported by Kovi et al. (2021) for yield at the third cut was much lower (−0.03 to 0.01) than the one obtained in our study. However, the differences in accuracy may be due to a larger validation population (more individuals genotyped) and larger number of phenotypic records in our study. Also, Kovi et al. (2021) used hierarchical clustering to define validation groups as compared to the family cross-validation and forward-prediction strategies in our study.

Predictive ability for winter damage was the lowest for single-trait models in the family cross-validation scenario. An increase in predictive ability for multiple-trait GBLUP was perhaps due to some half-sibs having records for winter damage and other traits, thus increasing the available information for the estimation of GEBV. On the other hand, for the forward-prediction strategy such may have not been the case, leading to a decrease in predictive ability. In addition, in the forward-prediction validation scenario, winter damage was found to have the most under-prediction when using single-trait GBLUP, probably due to the low number of records overall and outside of Finland, and the lower heritability of the trait. Such under-prediction was ameliorated when multi-trait models were used. Thus, multiple-trait models may be more useful when there is interest in improving winter survival.

Overall, the predictive ability for digestibility traits was moderately high, likely due to high heritability and a relatively large number of records in the training population. In switchgrass, in-vitro dry matter digestibility had a predictive ability of 0.69 when cross-validation was performed across the entire dataset (Fiedler et al. 2018). This predictive ability is consistent with that obtained from the family cross-validation in our study. On the other hand, the predictive ability for organic matter digestibility in ryegrass was approximately 0.20 (Arojju et al. 2020) which more closely corresponds to the predictive ability of D-value at first cut in our study for the forward-prediction strategy. For both validation strategies, D-value GEBV were not affected by over- or under-prediction. In addition, single- and multi-trait GBLUP were similarly accurate.

Using the Legarra and Reverter (2018) methodology, moderately high correlations between GEBV from full and reduced sets, along with minimal dispersion (under- or over-prediction), and moderate R^2 values were found for all traits and for both the family cross-validation and forward-prediction methodologies. This suggest that the predictions of GEBV were accurate, likely due to the high number of

records and moderate heritability for the traits. Bornhofen et al. (2022) reported similar dispersion values for GEBV of dry matter yield and forage quality traits in ryegrass, ranging from 0.55 to 1.51. Similarly, the correlations between GEBV obtained from full and reduced models reported by these authors ranged from about 0.45 to 0.95. As in our study, their reported values depended on the model used (e.g., single- vs multiple-trait) and validation strategy (e.g., tenfold cross-validation vs family validation).

The impact of environment is evident in many grass populations and can be included in the model through G×E interactions (Bornhofen et al. 2022). For models including many traits and environments as it would be the case in our study, estimates of variance components derived from multiple-trait and multiple-environment models may be unreliable, especially with limited number of records and unbalanced data across locations. An alternative is to use so-called reaction norm models (Kolmodin et al. 2002), which can produce a parsimonious model for estimating G×E interactions that allows for many environments. Reaction norm models will be tested as more data become available. Therefore, with the intention of providing reliable estimates of variance components and accurate genomic predictions such models were not considered in our current work. Still, to minimize the impact of environment on estimates of variance components and prediction of GEBV, trial location was included in the definition of contemporary group and permanent environment in the model.

Conclusion

In the present study, genomic relationships derived from GBS data along with phenotypic records were used to estimate variance components of several traits of economic importance for a Finnish Timothy population. The moderate to large genetic variances, and the corresponding heritability, of some yield and digestibility traits showed the potential for making genetic gains in these traits. However, for the low heritable traits (e.g., winter damage), favorable correlations with other traits can be used to increase genetic gain. On the other hand, negative correlations between some yield traits and winter damage (and digestibility traits) highlight the need to use selection indices to reduce the effects of these unfavorable relationships.

The current work constituted a first step toward a routine and comprehensive genomic evaluation of Timothy in Finland. Markers obtained from GBS data were used in single- and multi-trait GBLUP to predict GEBV. While read depth and call rate directly affected the number of markers available for downstream analyses, only call rate had a considerable effect on predictive ability. Moreover, the potential benefits of using multi-trait GBLUP were not consistent

and need to be further investigated. Nevertheless, the high accuracy of some yield and digestibility traits under the different validation strategies used emphasize the possibility for developing materials with improved yield and nutritional qualities.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00122-025-04860-9>.

Author contributions All authors have contributed to the study design and conception. Sequencing and genotype data collecting, and preparation was performed by DF, OB, OM, EAM, and NVJ. Phenotypic data collection and processing was performed by PP and MI. Statistical analysis, modeling, and genomic prediction by NVJ, HK, IM, and EAM. First draft of the manuscript is by NVJ. All authors have commented on previous versions of the manuscript, and all authors read and approved the final manuscript.

Funding Open access funding provided by Natural Resources Institute Finland. This project was supported by Ministry of Agriculture and Forestry, Makera fund VN/7519/2020, seed industry (Hankkija, Peltoniemi, Tilasiemi) and meat industry (Atria, HKScan, Snellman).

Data availability The data used for this study are property of Boreal Plant Breeding Ltd., and as such not publicly available.

Declarations

Conflict of interests The authors have no relevant financial or non-financial interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aroju SK, Cao M, Trollove M, Barrett BA, Inch C, Eady C, Stewart A, Faville MJ (2020) Multi-trait genomic prediction improves predictive ability for dry matter yield and water-soluble carbohydrates in perennial ryegrass. *Front Plant Sci* 11:1197. <https://doi.org/10.3389/fpls.2020.01197>
- Ashikaga K, Tanaka T, Fujii H, Tamaki H, Sato K, Deguchi K, Iida K (2011) Relationship between the first and second crops and estimation of genetic parameters of the second crop on the nutritive value of timothy (*Phleum pratense* L.). *Euphytica* 182:325–334. <https://doi.org/10.1007/s10681-011-0420-3>
- Ashikaga K, Tanaka T, Fujii H, Tamaki H, Sato K (2016) Simultaneous selection for nutritive value and agronomic traits in timothy (*Phleum pratense* L.). *Euphytica*. <https://doi.org/10.1007/s10681-015-1583-0>

- Ashikaga K, Tamaki H, Deguchi K, Sato K (2008) Heritability of nutritive value in first crop of Timothy *Phleum pratense* L. *Jpn J Grassl Sci*. <https://doi.org/10.14941/grass.54.19>
- Batista LG, Gaynor RC, Margarido GRA, Byrne T, Amer P, Gorjanc G, Hickey JM (2021) Long-term comparison between index selection and optimal independent culling in plant breeding programs with genomic prediction. *PLoS ONE* 16(5):e0235554. <https://doi.org/10.1371/journal.pone.0235554>
- Becker WA (1992) Manual of quantitative genetics, 5th edn. Academic Enterprises, USA
- Bélanger G, Michaud R, Jefferson PG, Tremblay GF, Bréard A (2001) Improving the nutritive value of timothy through management and breeding. *Can J Plant Sci* 81:577–585. <https://doi.org/10.4141/P00-143>
- Berg CC, McElroy AR, Kunelius HT (1996) Timothy. In: Moser LE, Buxton DR, Casler MD (eds) Cool-season forage grasses, ASA, CSSA, SSSA, Madison, pp 643–664. <https://doi.org/10.2134/agronmonogr34.c20>
- Bornhofen E, Fè D, Lenk I, Greve M, Didion T, Jensen CS, Asp T, Janss L (2022) Leveraging spatiotemporal genomic breeding value estimates of dry matter yield and herbage quality in ryegrass via random regression models. *Plant Genome* 15:e20255. <https://doi.org/10.1002/tpg2.20255>
- Cai Q, Bullen MR (1994) Analysis of genome-specific sequences in phleum species: identification and use for study of genomic relationships. *Theoret Appl Genet* 88:831–837. <https://doi.org/10.1007/BF01253993>
- Cai H-W, Yuyama N, Tamaki H, Yoshizawa A (2003) Isolation and characterization of simple sequence repeat markers in the hexaploid forage grass timothy (*Phleum pratense* L.). *Theoret Appl Genet* 107:1337–1349. <https://doi.org/10.1007/s00122-003-1386-x>
- Cericola F, Lenk I, Fè D, Byrne S, Jensen CS, Pedersen MG, Asp T, Jensen J, Janss L (2018) Optimized use of low-depth genotyping-by-sequencing for genomic prediction among multiparental family pools and single plants in perennial ryegrass (*Lolium perenne* L.). *Front Plant Sci*. <https://doi.org/10.3389/fpls.2018.0036>
- Cerón-Rojas JJ, Crossa J (2018) Linear selection indices in modern plant breeding. Springer, Cham. <https://doi.org/10.1007/978-3-319-91223-3>
- Claessens A, Michaud R, Bélanger G, Mather DE (2004) Characteristics of timothy genotypes divergently selected for fiber traits. *Crop Sci* 44(1):81–88. <https://doi.org/10.2135/cropsci2004.8100>
- Clark LV, Lipka AE, Sacks EJ (2019) polyRAD: genotype calling with uncertainty from sequencing data in polyploids and diploids. *G3 Genes Genom Genet*. <https://doi.org/10.1534/g3.118.200913>
- Cobb JN, Juma RU, Biswas PS, Arbelaez JD, Rutkoski J, Atlin G, Hagen T, Quinn M, Ng EH (2019) Enhancing the rate of genetic gain in public-sector plant breeding programs: lessons from the breeder's equation. *Theor Appl Genet* 132:627–645. <https://doi.org/10.1007/s00122-019-03317-0>
- de Bem OI, Amadeu RR, Ferrão LFV, Muñoz PR (2020) Optimizing whole-genomic prediction for autotetraploid blueberry breeding. *Heredity* 25(6):437–448. <https://doi.org/10.1038/s41437-020-00357-x>
- Fè D, Pedersen MG, Jensen CS, Jensen J (2015) Genetic and environmental variation in a commercial breeding program of perennial ryegrass. *Crop Sci* 55:631–640. <https://doi.org/10.2135/cropsci2014.06.0441>
- Fiedler JD, Lanzatella C, Edmé SJ, Palmer NA, Sarath G, Mitchell R, Tobias CM (2018) Genomic prediction accuracy for switchgrass traits related to bioenergy within differentiated populations. *BMC Plant Biol*. <https://doi.org/10.1186/s12870-018-1360-z>
- Garrison E, and Marth G (2012) Haplotype-based variant detection from short-read sequencing. arXiv preprint. <https://doi.org/10.48550/arXiv.1207.3907>. Accessed 1 July 2024.
- Gerard D, Ventrone Ferrão LF, Franco Garcia AA, Stephens M (2018) Genotyping polyploids from messy sequencing data. *Genetics* 210:789–807. <https://doi.org/10.1534/genetics.118.301468>
- Gorjanc G, Cleveland MA, Houston RD, Hickey JM (2015) Potential of genotyping-by-sequencing for genomic selection in livestock populations. *Genet Sel Evol* 47:1–12. <https://doi.org/10.1186/s12711-015-0102-z>
- Guo X, Cericola F, Fè D, Pedersen MG, Lenk I, Jensen CS, Jensen J, Janss L (2018) Genomic prediction in tetraploid ryegrass using allele frequencies based on genotyping by sequencing. *Front Plant Sci* 9:1–14. <https://doi.org/10.3389/fpls.2018.01165>
- Harris C, Hall M, Arrowfield R, Herridge R, Eady C, Macknight R, Brownfield L (2023) Assessing inbreeding in perennial ryegrass (*Lolium perenne*) as a step towards F1 hybrid breeding. *Plant Breeding* 142(4):518–526. <https://doi.org/10.1111/pbr.13099>
- Huhtanen P, Krizsan SJ (2023) Nutritional uniformity of cell wall and lignin fractions in grass and red clover silage evaluated by the Lucas test with application to forage feed evaluation. *Anim Feed Sci Tech* 306:1–12. <https://doi.org/10.1016/j.anifeedsci.2023.115819>
- Hyrkäs M, Korhonen P, Pitkänen T, Rinne M, Kaseva J (2018) Grass growth models for estimating digestibility and dry matter yield of forage grasses in Finland. In: Proceedings of the 27th General Meeting of the European Grassland Federation, Cork, Ireland, pp 252–254.
- Kolmodin R, Strandberg E, Madsen P, Jensen J, Jorjani H (2002) Genotype by environment interaction in nordic dairy cattle studied using reaction norms. *Acta Agric Scand A Anim Sci* 52(1):11–24. <https://doi.org/10.1080/09064700252806380>
- Kovi MR, Marum P, Alsheikh M, Amdahl H, Asp T, Rognli OA (2019) Implementation of genomic predictions in timothy (*Phleum pratense* L.). In: Proceedings of the joint 20th symposium of the European Grassland Federation and the 33rd meeting of the EUCARPIA section 'Fodder crops and amenity grasses'. Zürich, Switzerland, pp 262–274. <https://doi.org/10.3929/ethz-b-000369264>
- Kovi MR, Pashapu AR, Amdahl H, Alsheikh M, Marum P, Rognli OA (2021) Impact of genetic relatedness on the genomic prediction accuracies in Timothy (*Phleum pratense* L.). In: Proceedings of the 34th meeting of the EUCARPIA 'Fodder crops and amenity grasses' section in cooperation with the EUCARPIA Festulolium Working Group. Freising, Germany, pp 109–112. <https://doi.org/10.3929/ethz-b-000511542>
- Langmead R, Salzberg BL (2012) Bowtie 2: fast and sensitive alignment of short DNA sequences to the human genome. *Nat Methods* 9(4):357–359. <https://doi.org/10.1038/nmeth.1923>
- Legarra A (2016) Comparing estimates of genetic variance across different relationship models. *Theoret Pop Biol* 107:26–30. <https://doi.org/10.1016/j.tpb.2015.08.005>
- Legarra A, Reverter A (2018) Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. *Genet Sel Evol* 50(53):1–18. <https://doi.org/10.1186/s12711-018-0426-6>
- Louw JH (1990) A selection index to cope with genotype-environment interaction with an application to wheat breeding. *Plant Breed* 104(4):346–352. <https://doi.org/10.1111/j.1439-0523.1990.tb00446.x>
- Lush JL (1937) Animal breeding plans. Iowa State University Press, Ames
- Matilainen K, Strandén I, Mäntysaari EA (2019) Efficient monte carlo algorithm for restricted maximum likelihood estimation of genetic parameters. *J Anim Breed Genet* 136:252–261. <https://doi.org/10.1111/jbg.12375>

- Meuwissen TH, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genomewide dense marker maps. *Genetics* 157:1819–1829. <https://doi.org/10.1093/genetics/157.4.1819>
- Moore KJ, Moser LE, Vogel KP, Waller SS, Johnson BE, Pedersen JF (1991) Describing and quantifying growth stages of perennial forage grasses. *Agron J* 83:1073–1077. <https://doi.org/10.2134/agronj1991.00021962008300060027x>
- Niemeläinen O, Nissinen O, Jauhiainen L, Isolahti M. Timothy and tall fescue mixtures increase forage yield and improve yield quality in comparison with pure timothy in Finland (2004) In: Land use systems in grassland dominated regions. Proceedings of the 20th General Meeting of the European Grassland Federation. Luzern. Switzerland, pp 516–518
- Nousiainen J, Rinne M, Hellämäki M, Huhtanen P (2003) Prediction of the digestibility of the primary growth of grass silages harvested at different stages of maturity from chemical composition and pepsin-cellulase solubility. *Anim Feed Sci Technol*. [https://doi.org/10.1016/S0377-8401\(02\)00283-3](https://doi.org/10.1016/S0377-8401(02)00283-3)
- Peltonen-Sainio P, Jauhiainen L, Känkänen H, Joonas J, Hydén T, Mattila TJ (2022) Farmers' experiences of how under-sown clovers, ryegrasses, and timothy perform in northern european crop production systems. *Agronomy* 12:1–18. <https://doi.org/10.3390/agronomy12061401>
- Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest radseq: an inexpensive method for de novo snp discovery and genotyping in model and non-model species. *PlosOne* 7(5):1–11. <https://doi.org/10.1371/journal.pone.0037135>
- R Core Team (2024). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>. Accessed 1 July 2024.
- Rinne M, Nykänen A (2000) Timing of primary growth harvest affects the yield and nutritive value of timothy-red clover mixtures. *Agric Food Sci Finland*. 9(2):121–134
- Schmidt P, Hartung J, Rath J, Piepho H (2019) Estimating broad-sense heritability with unbalanced data from agricultural cultivar trials. *Crop Sci* 59:525–536. <https://doi.org/10.2135/cropsci2018.06.0376>
- Tanhuanpää P, Manninen O (2012) High SSR diversity but little differentiation between accessions of nordic timothy (*Phleum pratense* L.). *Hereditas*. <https://doi.org/10.1111/j.1601-5223.2012.02244.x>
- Thorvaldsson G, Tremblay FG, Kunelius HT (2007) The effects of growth temperature on digestibility and fibre concentration of seven temperate grass species. *Acta Agric Scand B Soil Plant Sci* 57(4):322–328. <https://doi.org/10.1080/09064710600984221>
- Van Raden PM (2009) Efficient Methods to compute genomic predictions. *J Dairy Sci* 91:4414–4423. <https://doi.org/10.3168/jds.2007-0980>
- Villanueva B, Fernández A, Saura M, Caballero A, Fernández J, Morales-González E, Toro MA, Pong-Wong R (2021) The value of genomic relationship matrices to estimate levels of inbreeding. *Genet Sel Evol*. <https://doi.org/10.1186/s12711-021-00635-0>
- Vuori K, Strandén I, Lidauer M, Mäntysaari EA (2006) MiX99-effective solver for large and complex linear mixed models. In: Proceedings of the 8th World congress on genetics applied to livestock production, Belo Horizonte, Brazil, pp 27–33
- Wang N, Yuan Y, Wang H, Yu D, Liu Y, Zhang A, Gowda M, Nair SK, Hao Z, Lu Y, San Vicente F, Prasanna BM, Li X, Zhang X (2020) Applications of genotyping-by-sequencing (GBS) in maize genetics and breeding. *Sci Rep* 10:16308. <https://doi.org/10.1038/s41598-020-73321-8>
- Williams DA (1982) Extra-Binomial variation in logistic linear models. *J R Stat Soc C Appl Stat* 31(2):144–148. <https://doi.org/10.2307/2347977>
- Wittenburg D, Melzer N, Willmitzer L, Lisec J, Kesting U, Reinsch N, Reipsilber D (2013) Milk metabolites and their genetic variability. *J Dairy Sci* 96(4):2557–2569. <https://doi.org/10.3168/jds.2012-5635>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.