Genomic Prediction in Tetraploid Ryegrass Using Allele Frequencies Based on Genotyping by Sequencing

Xiangyu Guo[1]*, Fabio Cericola[2], Dario Fè[3], Morten G. Pedersen[3], Ingo Lenk[3], Christian S. Jensen[3], Just Jensen[1] and Lucas L. Janss[1]

[1] Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Aarhus University, Aarhus, Denmark, [2] Rijk Zwaan B.V., De Lier, Netherlands, [3] Research Division, DLF Seeds A/S, Store Heddinge, Denmark

Perennial ryegrass is an outbreeding forage species and is one of the most widely used forage grasses in temperate regions. The aim of this study was to investigate the possibility of implementing genomic prediction in tetraploid perennial ryegrass, to study the effects of different sequencing depth when using genotyping-by-sequencing (GBS), and to determine optimal number of single-nucleotide polymorphism (SNP) markers and sequencing depth for GBS data when applied in tetraploids. A total of 1,515 F_2 tetraploid ryegrass families were included in the study and phenotypes and genotypes were scored on family-pools. The traits considered were dry matter yield (DM), rust resistance (RUST), and heading date (HD). The genomic information was obtained in the form of allele frequencies of pooled family samples using GBS. Different SNP filtering strategies were designed. The strategies included filtering out SNPs having low average depth (FILTLOW), having high average depth (FILTHIGH), and having both low average and high average depth (FILTBOTH). In addition, SNPs were kept randomly with different data sizes (RAN). The accuracy of genomic prediction was evaluated by using a "leave single F_2 family out" cross validation scheme, and the predictive ability and bias were assessed by correlating phenotypes corrected for fixed effects with predicted additive breeding values. Among all the filtering scenarios, the highest estimates for genomic heritability of family means were 0.45, 0.74, and 0.73 for DM, HD and RUST, respectively. The predictive ability generally increased as the number of SNPs included in the analysis increased. The highest predictive ability for DM was 0.34 (137,191 SNPs having average depth higher than 10), for HD was 0.77 (185,297 SNPs having average depth lower than 60), and for RUST was 0.55 (188,832 SNPs having average depth higher than 1). Genomic prediction can help to optimize the breeding of tetraploid ryegrass. GBS data including about 80–100 K SNPs are needed for accurate prediction of additive breeding values in tetraploid ryegrass. Using only SNPs with sequencing depth between 10 and 20 gave highest predictive ability, and showed the potential to obtain accurate prediction from medium-low coverage GBS in tetraploids.

Keywords: ryegrass, tetraploid, genomic prediction, genotyping-by-sequencing, sequencing depth

INTRODUCTION

Perennial ryegrass (*Lolium perenne* L.) is one of the most widely sown forage grasses in temperate regions (Humphreys, 2005). Low production costs and the perennial character provide high agronomic value, and it is widely used for feeding ruminants (Jensen et al., 2001). The popularity of cultivating perennial ryegrass is mainly due to its re-growth capacity after defoliation and its palatability, digestibility, and nutrient content as feed for ruminants compared with other forage species (Wilkins, 1991).

Compared to diploid ryegrass, the tillers and seed heads of tetraploid ryegrass are longer and the leaves are wider. Tetraploid ryegrass is more open and more prone to wear, but is less susceptible to snow mold and has a better drought tolerance, leading to better performance under continental conditions with frequent dry periods. Palatability and digestibility are better in tetraploid varieties than in diploid varieties, and tetraploids perform better than diploids during grazing (Wilkins, 1991) and lead to a higher animal production (Lantinga and Groot, 1996; O'Donovan and Delaby, 2005).

Perennial ryegrass is an allogamous species (Cornish et al., 1979) due to a gametophytic self-incompatibility system (Cornish et al., 1979). For this reason, it is generally bred, maintained and commercialized as heterogeneous families. Evaluation of F_2 families is frequently used in breeding programs for outcrossing species such as perennial ryegrass. An F_2 family consists of the offspring from random interbreeding a full-sib F_1 family, which are the offspring from an initial bi-parental cross. F_2 families are evaluated in plot experiments over several locations and years to obtain measurements on yield, agronomic traits, and disease resistance.

Perennial ryegrass breeding has mainly relied on prediction of genetic merit using phenotypic information only (Conaghan and Casler, 2011; Hayes et al., 2013). Using this system, relevant improvements for yield and quality-related traits have been achieved (Wilkins and Humphreys, 2003; McDonagh et al., 2016). However, compared with traits such as rust resistance and spring growth, gains for yield traits like dry matter and seed yield were not as high as expected (Sampoux et al., 2011). In addition, phenotypic selection is costly and time consuming, needing up to 10 years to complete a selection cycle (Wilkins and Humphreys, 2003; Lin et al., 2016). In recent decades, the development of marker technology allowed adoption of genomic prediction (GP) strategies, which have been highly beneficial and led to a reduction of cost in practical animal and plant breeding programs (Hickey et al., 2017). In GP, dense markers distributed across the whole genome can be used simultaneously to predict

breeding values (Meuwissen et al., 2001). The quantitative trait loci (QTLs) affecting the traits of interest are assumed in linkage disequilibrium (LD) with one or more single-nucleotide polymorphism (SNP) markers. Thus, a sufficiently dense and well-distributed set of markers allows all QTLs to be in LD with at least one marker, and this LD can be exploited in GP to ensure accurate prediction of breeding values as a basis for selection decisions.

The prospects for implementing GP in forage grass breeding were recently reviewed by Hayes et al. (2013). Several GP studies have been reported for crops such as maize and wheat (Crossa et al., 2010, 2014), and the first investigations in diploid perennial ryegrass also demonstrated great potential for using GP (Fè et al., 2015a, 2016). However, GP studies for tetraploid ryegrass, to our knowledge, have not yet been carried out. The implementation of GP in tetraploid ryegrass may be more challenging than in diploid ryegrass, because families of tetraploids will show a higher heterozygosity than families of diploids. This may hamper accurate estimation of genomic relationships and genomic breeding values.

Genotyping-by-sequencing (GBS) was developed by Elshire et al. (2011) as a robust genotyping approach. GBS uses methylation sensitive restriction enzymes to reduce genome complexity. GBS is a good approach to estimate genome-wide allele frequency profiles in pooled samples for outbred heterogeneous varieties (Byrne et al., 2013). Moreover, for association studies and GP studies, calling of genotypes can be avoided by directly using allele frequencies from GBS, which facilitates measurements on family pools (Ashraf et al., 2014). Use of GBS data also poses some challenges; in particular, sequencing depth needs to be optimized carefully. At low depth, genotyping errors and missing values are an issue (Poland and Rife, 2012), and result in biased estimates of allele-effect and heritability (Ashraf et al., 2014, 2016). At higher sequencing depth the accuracy of genotype estimates is improved (Sims et al., 2014), but under a fixed budget, the number of samples that can be sequenced would be reduced, which reduces power of the entire experiment (Ashraf et al., 2014). Several investigations on how sequencing depth affects association studies and estimation of genomic heritability have been conducted (Garner, 2011; Sims et al., 2014; Ashraf et al., 2016). As reviewed by Poland and Rife (2012), GBS has become a flexible and low cost tool for plant genetics and breeding. It has been demonstrated that GBS can effectively generate high-density genome wide markers in a range of species (Elshire et al., 2011; Poland and Rife, 2012; Poland et al., 2012; Beissinger et al., 2013; Crossa et al., 2013; Zhang et al., 2015; Fè et al., 2016; Cericola et al., 2018). With GBS, an accurate GP model was derived for the large, complex, and polyploid wheat genome (Poland et al., 2012). In addition, GBS also has been applied on diploid ryegrass for genomic prediction (Fè et al., 2015a, 2016). However, to our knowledge the optimization of sequencing depth for GP in tetraploid ryegrass has not been reported yet.

The aims of this study were: (1) to investigate the possibility of implementing genomic prediction in tetraploid perennial ryegrass, (2) to study the effects of different sequencing depth when using GBS, and (3) to determine the optimal number of

---

Abbreviations: CV, cross-validation; DM, dry matter yield; FILTBOTH, the strategy filtering out SNPs having both low average and high average depth; FILTHIGH, the strategy filtering out SNPs having high average depth; FILTLOW, the strategy filtering out SNPs having low average depth; GBS, genotyping-by-sequencing; GEBVs, genomic breeding values; GP, genomic prediction; GSLM, family × sowing year × location × management effects; GSLMF, family × sowing year × location × management × farming year effects; G×E, genotype by environment; HD, heading date; LD, linkage disequilibrium; QTLs, quantitative trait loci; RAN, the strategy keeping SNPs randomly with different data sizes; RUST, rust resistance; SNP, single-nucleotide polymorphism.

SNPs to include in genomic prediction when GBS are applied in tetraploid ryegrass.

MATERIALS AND METHODS

Plant Material

Both phenotype and genotype data were derived from 1,515 F_2 families from a commercial breeding program from DLF A/S, Denmark. F_2 families originated from a pair-cross between two parents; F_1 seeds from both parent plants were pooled; F_1 families were sown in small protected plots to cross-fertilize; and finally F_2 seeds were harvested and used for field-testing of F_2 families. A detailed description of testing procedures was provided by Fè et al. (2015b).

Phenotypic records, defined below, consisted of historical data from F_2 families, which were sown between 2004 and 2016 at 8 locations in Denmark, the Netherlands, France, and United Kingdom, and cultivated according to the local management schemes. In all locations, F_2 families were tested in trials including 12 families in a randomized experiment with two replicates for each family. Details of testing and recording procedures were the same as for diploid ryegrass as described previously (Fè et al., 2015b). The dataset analyzed included records of three traits:

(1) Dry matter yield (DM), expressed in kg/m^2 and obtained from multiple cuts over 2 years. For analyses, the total yield during the first year and the total yield during the second year were used so that each family had yield measurements from two years; to validate genomic predictions, the average yield of the two years was predicted.
(2) Heading date (HD), defined as the day on which spikes are visible over the general plots, and expressed in days since January 1st. HD was scored in plots for seed multiplication, which were farmed for one cropping season only.
(3) Rust resistance (RUST), measured during the period of maximum infection, both in regular yield plots, and in mini plots, which were cultivated only for 1 year. The level of infection was determined by visual scoring from 1 (plants completely covered by rust) to 9 (no sign of rust infection). Plots were cut between the different scoring time points to make the scores independent.

Descriptive statistics including mean value, standard deviation, minimum, maximum, number of families, number of records, number of plots, and number of sowing year × location × management levels are listed in **Table 1**.

Filtering of GBS Data and Calculation of Allele Frequencies for Each Family

Genotypic data was produced as described previously (Fè et al., 2015a). In total, 1,515 F_2 families were sequenced. A total of 51 libraries were prepared, with up to 96 families per library. Each library was sequenced on multiple lanes of an Illumina HiSeq2000 (single-end). On average, 12.9 million 100 bp single-end reads were produced per sample. A draft sequence assembly (Byrne et al., 2015) was used for the alignment of data for each family, and initially 18.6 million SNPs were identified. A first, quite liberal, filtering of the raw SNP data was performed by removing: (1) SNPs with missing rate higher than 50%; (2) SNPs with allele frequencies lower than 0.01 or higher than 0.99; (3) SNPs with average read depth smaller than 1. This left 188,832 SNPs available for our analysis, which included further, more stringent, filtering steps for the SNPs. The average read depth for the 188,832 SNPs ranged from 1 to 278, with mean of 19. The distribution of average read depth for each SNP is shown in **Figure 1**.

Differently from SNP chip data, where genotypes are explicitly called, the genotype of a SNP is obtained here in the form of an allele frequency ($\widehat{g_{ij}}$), which is estimated as the ratio between
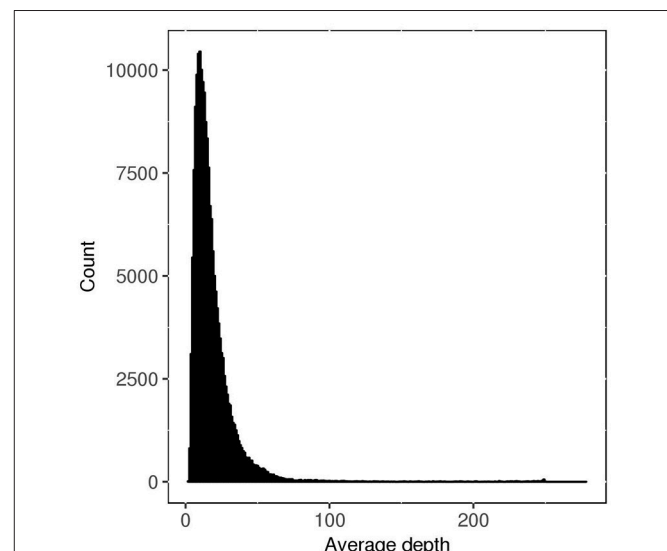

FIGURE 1 | Distribution of average read depth for each SNP before filtering by different strategies.

TABLE 1 | Descriptive statistics[a] for three traits.

Trait[b]	No. Fam	No. Rec	No. Plot	No. YLM	Mean	SD	Min	Max
DM	1,188	5,312	3,414	27	1.33	0.37	0.41	2.5
HD	979	1,810	1,810	7	155.64	7.51	136	178
RUST	1,506	13,545	5,368	22	5.64	1.99	1	9

[a]No. Fam, number of families; No. Rec, number of records; No. Plot, number of plots; No. YLM, number of sowing year × location × management levels; Mean, mean value; SD, standard deviation; Min, minimum value; Max, maximum value.
[b]DM, dry matter yield; HD, heading date; RUST, rust resistant.

number of reads for alternative allele ($S_{1_{ij}}$) and the total number of reads ($S_{T_{ij}}$), which is the sum of number of reads for the reference allele ($S_{2_{ij}}$) and $S_{1_{ij}}$, for each sample:

$$\widehat{g_{ij}} = \frac{S_{1_{ij}}}{S_{T_{ij}}} = \frac{S_{1_{ij}}}{S_{1_{ij}} + S_{2_{ij}}}.$$

SNP Filtering Strategies

In order to study the effect of sequencing depth of GBS data, additional SNP filtering was performed. First, SNPs having average depth lower than a certain value were filtered out in 11 levels, with minimum depth from 1 to 90 (FILTLOW1 to FILTLOW11); second, SNPs having average depth higher than a certain value were filtered out in 11 levels, with maximum depth from 100 to 5 (FILTHIGH1 to FILTHIGH11); third, SNPs having average depth outside a certain range were filtered out (equivalent to keeping SNPs with average depth within that range), using 12 different ranges (FILTBOTH1 to FILTBOTH12); finally, SNPs were kept randomly with 11 different data sizes from 5 to 180 k (RAN5 to RAN180), and repeated for 10 times. In summary, there were four filtering strategies, FILTLOW, FILTHIGH, FILTBOTH, and RAN, and the number of scenarios was, respectively, 11, 11, 12, and 11, where the latter (RAN) was repeated 10 times. A summary of SNPs used in each filtering scenario is shown in **Figure 2** and the details are shown in **Supplementary Table 1**.

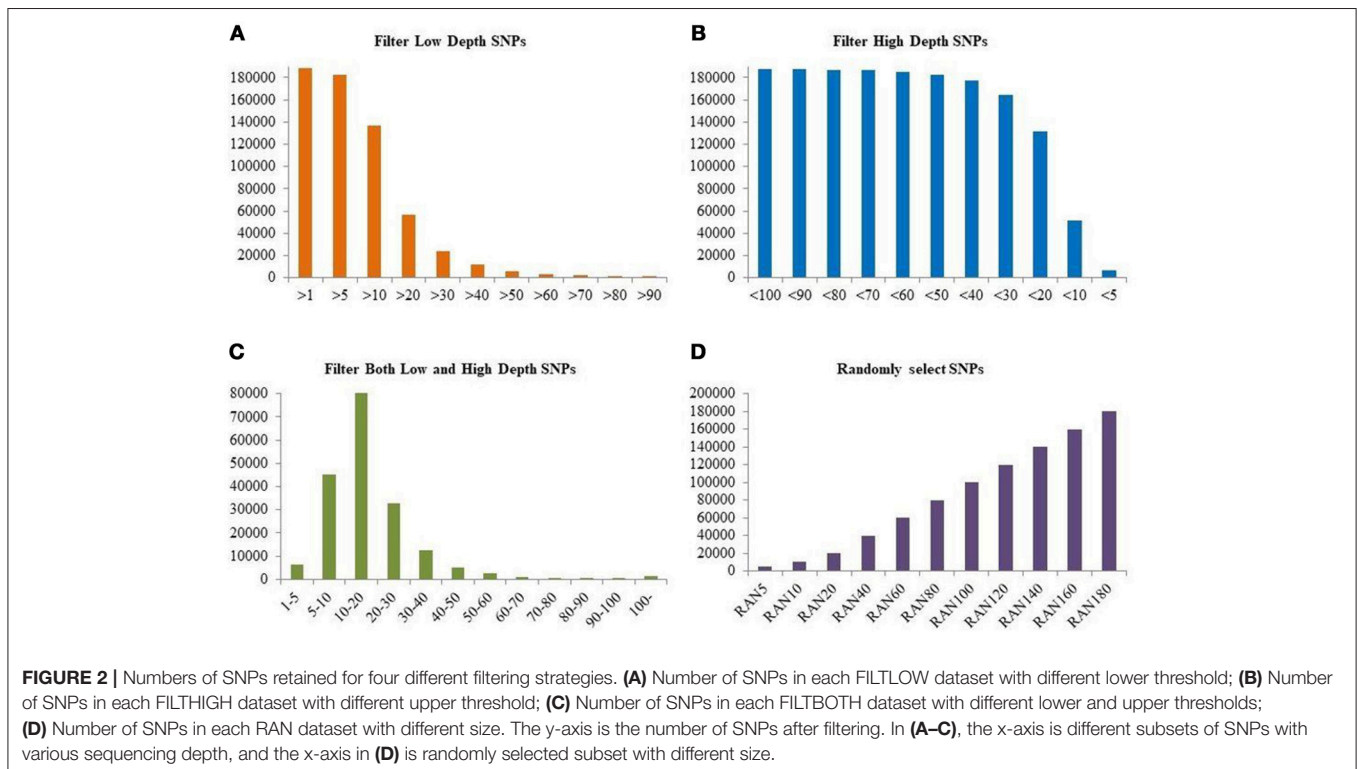## Statistical Model and Methods

A single trait model was used to estimate variance components and fixed effects, and to predict breeding values as well as other random effects in the model:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z_g g} + \mathbf{Z_a a} + \mathbf{Z_p p} + \mathbf{Z_{i_1} i_1} + \mathbf{Z_{i_2} i_2} + \mathbf{e},$$

in which **y** was the vector of phenotypic values of the trait DM, HD or RUST; **b** was the vector of fixed effects (sowing year × location × management × trial × farming year); **g** was the vector of additive genomic family effects; **a** was the vector of residual genetic family effects which were not explained by the genomic information; **p** was the vector of random plot effects; $\mathbf{i_1}$ and $\mathbf{i_2}$ were vectors of genotype by environment (G×E) effects [$\mathbf{i_1}$: family × sowing year × location × management (GSLM), $\mathbf{i_2}$: family × sowing year × location × management × farming year (GSLMF)]; and **e** was the vector of random residual effects. **X**, $\mathbf{Z_g}$, $\mathbf{Z_a}$, $\mathbf{Z_p}$, $\mathbf{Z_{i_1}}$, and $\mathbf{Z_{i_2}}$ were incidence matrices associating **b**, **g**, **a**, **p**, $\mathbf{i_1}$, and $\mathbf{i_2}$ with **y**. The random effects were assumed to be independent of each other and normally distributed, that is, $\mathbf{g} \sim N(0, \mathbf{G^*}\sigma_g^2)$, $\mathbf{a} \sim N(0, \mathbf{I}\sigma_a^2)$, $\mathbf{p} \sim N(0, \mathbf{I}\sigma_p^2)$, $\mathbf{i_1} \sim N(0, \mathbf{I}\sigma_{i_1}^2)$, $\mathbf{i_2} \sim N(0, \mathbf{I}\sigma_{i_2}^2)$, $\mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)$, in which $\mathbf{G^*}$ was the corrected **G** matrix of additive genomic relationships constructed based on the genomic information, **I** was the identity matrix, and σ_g^2, σ_a^2, σ_p^2, $\sigma_{i_1}^2$, $\sigma_{i_2}^2$, and σ_e^2 were the variances of additive genomic effects, residual genetic effects, random plot effects, first genetic by environment effects, second genetic by environment effects, and residuals, respectively. For DM and RUST, the general model was applied in the analysis, while for HD, the effects of **p** and $\mathbf{i_2}$ were excluded since there was only one score and only one environment in each family for HD.

The method to compute the **G** matrix was based on a modification of VanRaden (2008) to use allele frequencies

FIGURE 2 | Numbers of SNPs retained for four different filtering strategies. **(A)** Number of SNPs in each FILTLOW dataset with different lower threshold; **(B)** Number of SNPs in each FILTHIGH dataset with different upper threshold; **(C)** Number of SNPs in each FILTBOTH dataset with different lower and upper thresholds; **(D)** Number of SNPs in each RAN dataset with different size. The y-axis is the number of SNPs after filtering. In **(A–C)**, the x-axis is different subsets of SNPs with various sequencing depth, and the x-axis in **(D)** is randomly selected subset with different size.

(ranging between 0 and 1) instead of SNP genotype calls. A matrix ($\mathbf{F}$) with allele frequencies for each sample was centered by the mean SNP frequencies to create matrix $\mathbf{M}$ ($\mathbf{M_j} = \mathbf{F_j} - \overline{\mathbf{F_j}}$). Then, the $\mathbf{G}$ matrix was obtained by computing $\mathbf{M}$ multiplied by its own transpose and scaled by the sum of expected SNP variances across genotypes ($\mathbf{G} = \mathbf{MM'}/\mathbf{K}$). The scale parameter used for tetraploid F_2 families is half that used for diploid F_2 families as computed by Ashraf et al. (2014) and as applied in the study by Fè et al. (2015a), because the number of alleles in F_2 family pools is eight for tetraploid families, which is double that of diploid families:

$$\mathbf{K} = 0.125 \sum \overline{F_j}(1 - \overline{F_j}).$$

Finally, the $\mathbf{G}$ matrix was corrected for the extra binomial variance due to limited sequencing depth. The correction was derived by Cericola et al. (2018) and simply can be done according to ploidy number and the average depth of the sample. Corrected $\mathbf{G}$ matrix ($\mathbf{G^*}$) was calculated by scaling down the diagonal elements of each individual as follows:

$$D_{c_i} = D_{b_i}\left(1 - \frac{n-1}{\overline{S_{T_i}} + n - 1}\right),$$

where D_{b_i} is the ith element of the biased diagonal element in $\mathbf{G}$ and D_{c_i} is the corrected diagonal element in $\mathbf{G^*}$, $\overline{S_{T_i}}$ is the average $S_{T_{ij}}$ for each individual across all SNPs, and n is the ploidy number, which is eight as mentioned before.

For each of the four filtering scenarios, single trait analyses were run on the subsets of SNPs, which were previously created according to different filtering strategies (**Figure 2** and **Supplementary Table 1**). Variance components and their standard errors (SE) were estimated by restricted maximum likelihood (REML) using the DMU software package (Madsen and Jensen, 2013).

The phenotypic variance of family means was calculated as the sum of weighted variance components:

$$\sigma_{P_f}^2 = \overline{G^*}\sigma_g^2 + \sigma_a^2 + \sigma_p^2/n_p + \sigma_{i_1}^2/n_{i_1} + \sigma_{i_2}^2/n_{i_2} + \sigma_e^2/n_e,$$

where $\overline{G^*}$ is the average diagonal of the corrected genomic relationship matrix ($\mathbf{G^*}$ matrix), n_p is the average number of plots for each family, n_{i_1} and n_{i_2} are the average numbers of environments for each family, and n_e is average number of replicates across all fields for each family. Accordingly, genomic family heritability based on multiple plots was calculated as $h_f^2 = \frac{\overline{G^*}\sigma_g^2}{\sigma_{P_f}^2}$. To evaluate importance of each random effect in the model, phenotypic variance of a single plot was also calculated:

$$\sigma_{P_p}^2 = \overline{G^*}\sigma_g^2 + \sigma_a^2 + \sigma_p^2 + \sigma_{i_1}^2 + \sigma_{i_2}^2 + \sigma_e^2.$$

In the calculation of $\sigma_{P_f}^2$, $\sigma_{P_p}^2$ and h_f^2, σ_p^2, and $\sigma_{i_2}^2$ were not considered for HD due to the reduced recording strategy for this trait. This was used to compute the relative contribution of each random effect to the total phenotypic plot variance.

Cross-Validation

To estimate the accuracy of genomic breeding values (GEBVs), a leave-one-family-out cross-validation (CV) strategy was applied. In each CV round, the phenotypes of one family were masked and then all other families were used to train the prediction model and to predict the family with phenotypes masked.

Before CV, the whole dataset was used to estimate variance components and fixed effects. Corrected phenotypes (y_c) were computed by subtracting the estimates of the fixed effects. Predictive ability was measured as $cor(\overline{y_c}, \hat{g})$, which is theoretically not larger than the square root of h_f^2 (Legarra et al., 2008) because breeding values predict genetic effects and not environment. $\overline{y_c}$ is the average y_c for each family. Furthermore, to assess bias of predictions, regression coefficient of $\overline{y_c}$ on $\hat{g}$ was calculated. The deviation of this regression coefficient from 1 represents the level of bias.

RESULTS

In order to interpret the results from each scenario, **Figures 2–8** were created. **Figure 2** and **Supplementary Table 1** show the numbers of SNPs retained for four different filtering strategies. **Figure 2** shows four bar charts according to the data filtering levels. **Figures 3–5** show the estimated heritability, predictive ability and bias in different SNP filtering scenarios for three traits, DM, HD and RUST, respectively. In these figures, line charts were plotted as a function of number of SNPs included in each model. **Figures 6–8** show the percentages of explained variance, i.e., each variance components over the total phenotypic variance, for three traits. In these three figures, bar charts were plotted for all scenarios.

SNP Filtering Strategies

The first filtering strategy FILTLOW used an increasing lower threshold for average SNP read depth, and the number of SNPs included decreased from 188,832 for FILTLOW1 to 1,587 for FILTLOW11. The second filtering strategy FILTHIGH used a decreasing upper threshold for average SNP read depth, and the number of SNPs included decreased from 187,516 for FILTHIGH1 to 6,389 for FILTHIGH11. In this data, a large proportion of SNPs had read depth between 10 and 20, which caused large reductions in the numbers of SNP when either the lower threshold for read depth increased to 20, or when the upper threshold for read depth decreased to 10. For instance, between FILTLOW >10 and FILTLOW >20, the number of SNPs kept dropped from 73 to 30%, and between FILTHIGH <20 and FILTHIGH <10, the number of SNPs kept dropped from 70 to 27%. The SNPs with depth lower than 10 and 5 accounted for 27 and 3% of the full data, respectively. In the third filtering strategy, SNPs were kept in a certain interval of average read depth. The percentage of SNPs kept varied from 43% for FILTBOTH3 having average read depth from 10 to 20, to 0.1% for FILTBOTH11 having average read depth from 90 to 100. In addition to three filtering strategies for average read depth, the RAN filtering strategy kept random subsets from the full dataset, ranging from 5 K (RAN5) to 180 K (RAN180) SNPs;
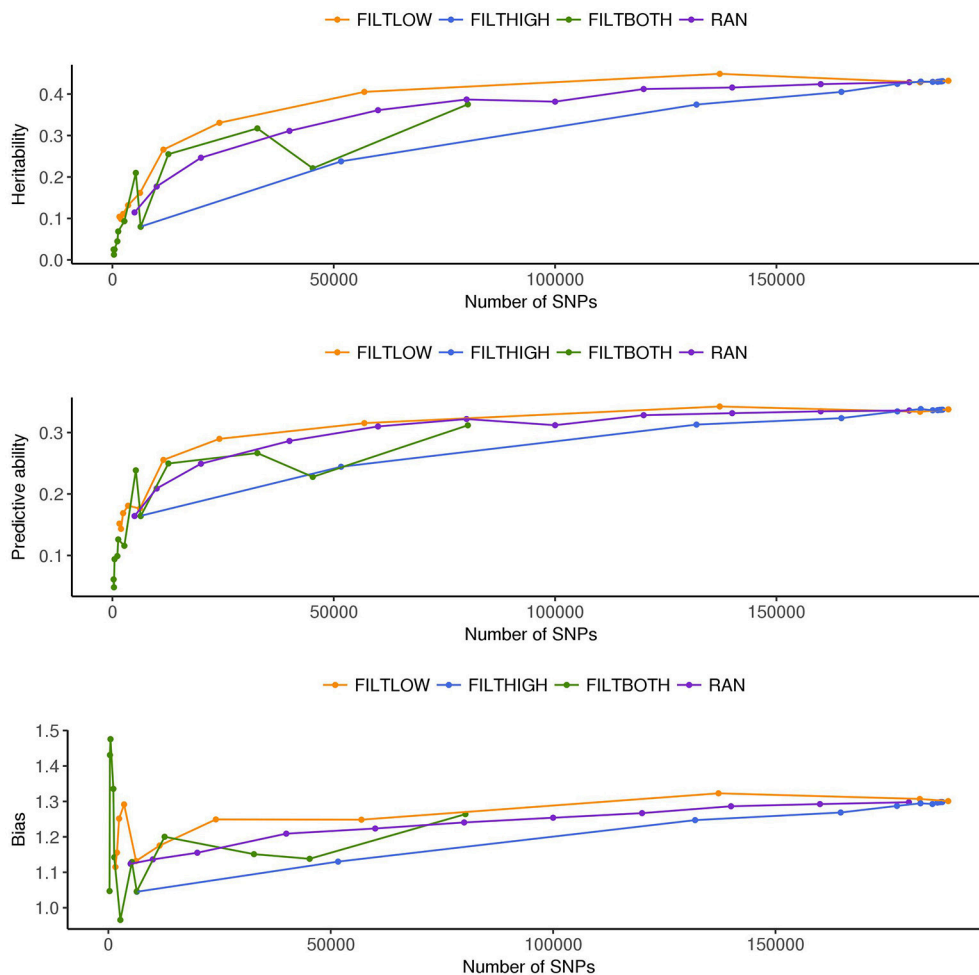
FIGURE 3 | Estimated heritability, predictive ability and bias in different SNP filtering scenarios[1] for dry matter yield. [1] ●FILTLOW, strategy filtering out SNPs having low average depth; ●FILTHIGH, strategy filtering out SNPs having high average depth; ●FILTBOTH, strategy filtering out SNPs having both low average and high average depth; ●RAN, strategy keeping SNPs randomly with different data size.

in percentages this corresponds to 3% of SNPs in RAN5 to 95% of SNPs in RAN180.

Variance Components and Heritabilities

Figures 3–5 show the effects of different filtering strategies on estimates of h_f^2 for the traits DM, DH, and RUST, respectively. For DM, the highest estimate of h_f^2 was 0.45 (FILTLOW3). In the FILTLOW scenarios, the estimated h_f^2 increased slightly from FILTLOW1 to FILTLOW3, and decreased fast afterwards. In the FILTHIGH scenarios, the estimated h_f^2 generally decreased from FILTHIGH1 to FILTHIGH11, and the decrease was more obvious from FILTHIGH8. In the FILTBOTH scenarios, the estimated h_f^2 was highest in FILTBOTH3, showed a small decrease in FILTBOTH4, and larger reductions in other FILTBOTH scenarios. In the RAN scenarios, the estimated h_f^2 increased along with the number of SNPs included in the model, i.e., increased from RAN5 to RAN180, with rate of increase gradually

reducing. For HD, the trends of heritability estimates within the four filtering strategies were the same as for DM, and the highest estimate of h_f^2 was also for FILTLOW3 at 0.74. For RUST, the highest estimate of h_f^2 was 0.73 for FILTHIGH1. In the FILTLOW scenarios, the estimated h_f^2 were similar for FILTLOW1 to FILTLOW3, and also decreased fast afterwards. The trends for the other three filtering strategies (FILTHIGH, FILTBOTH and RAN) were the same as for DM and HD.

In our analysis model, we also include a variance component for residual genetic effects (σ_a^2), i.e., the part of genetic effects that cannot be explained by genomic markers. **Figures 6–8** show that the percentage of σ_a^2 over $\sigma_{P_p}^2$ changed in the different scenarios. When the number of SNPs increased, the percentage of additive genetic variance explained by markers generally increased while the percentage of residual genetic variance decreased, and the percentage of total genetic variance (sum of $\overline{\mathbf{G}^*}\sigma_g^2$ and σ_a^2) over $\sigma_{P_p}^2$ remained relatively similar for all scenarios.
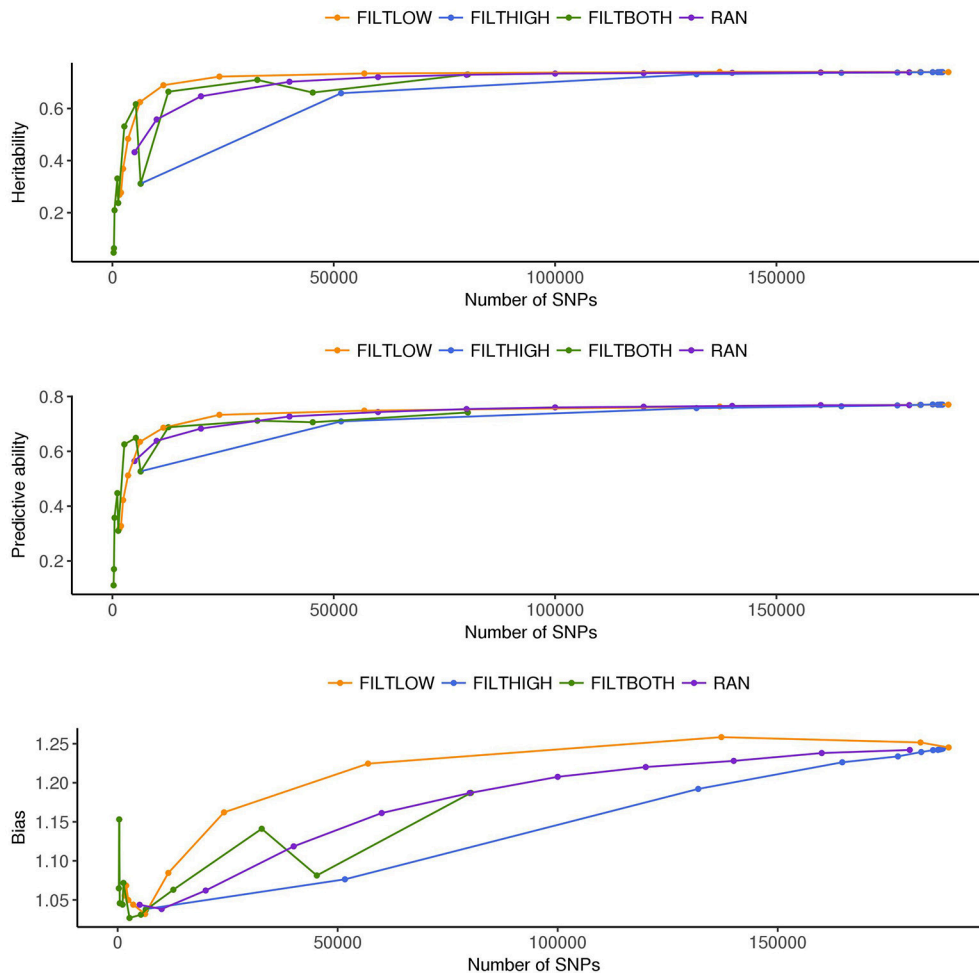
FIGURE 4 | Estimated heritability, predictive ability and bias in different SNP filtering scenarios[1] for heading date. [1] ●FILTLOW, strategy filtering out SNPs having low average depth; ●FILTHIGH, strategy filtering out SNPs having high average depth; ●FILTBOTH, strategy filtering out SNPs having both low average and high average depth; ●RAN, strategy keeping SNPs randomly with different data size.

For the variance of plot effects estimated for DM and RUST, the percentages of σ_p^2 over $\sigma_{P_p}^2$ were consistent across all the scenarios but different between DM and RUST. The percentage of variance due to plot effects in DM was about twice as large as the plot variance in RUST (**Figure 6**, **8**). For DM, σ_p^2 and total genetic variance had similar magnitude, but for RUST, σ_p^2 only accounted for 29% of total genetic variance.

As shown in **Figures 6**, **8**, the estimates of variance for G×E interactions GSLM ($\sigma_{i_1}^2$) and GSLMF ($\sigma_{i_2}^2$) were similar among all the scenarios, but different in DM and RUST. For DM, estimates of $\sigma_{i_1}^2$ were not significantly different from 0. However, for RUST, estimates for both $\sigma_{i_1}^2$ and $\sigma_{i_2}^2$ were significantly different from 0, with the average percentages of $\sigma_{i_2}^2$ being slightly larger than $\sigma_{i_1}^2$. For HD, the estimates of $\sigma_{i_1}^2$ varied more between scenarios than for DM and RUST, and the percentage of $\sigma_{i_1}^2$ over $\sigma_{P_p}^2$ ranged from 11 to 28%. When the number of SNPs increased, this percentage generally decreased. For HD, less phenotypic records

were available and therefore variance components were estimated with lower accuracy.

The estimation of residual variance (σ_e^2) was generally consistent among all scenarios. The average percentage of σ_e^2 was 59, 22, and 25%, for DM, HD and RUST, respectively. The largest difference between residual variance estimates for the different scenarios was 5% for HD (24% in FILTHIGH5 vs. 19% in FILTLOW11), 2% for DM (60% in FILTHIGH11 vs. 58% in FILTLOW3), and 1% in RUST (26% in FILTHIGH11 vs. 25% in FILTLOW11).

Details on estimated variance components and heritabilities, together with their standard errors (SE), for three traits in all F_2 families are available in **Supplementary Table 2** (DM), **Supplementary Table 3** (HD) and **Supplementary Table 4** (RUST).

Cross-Validation

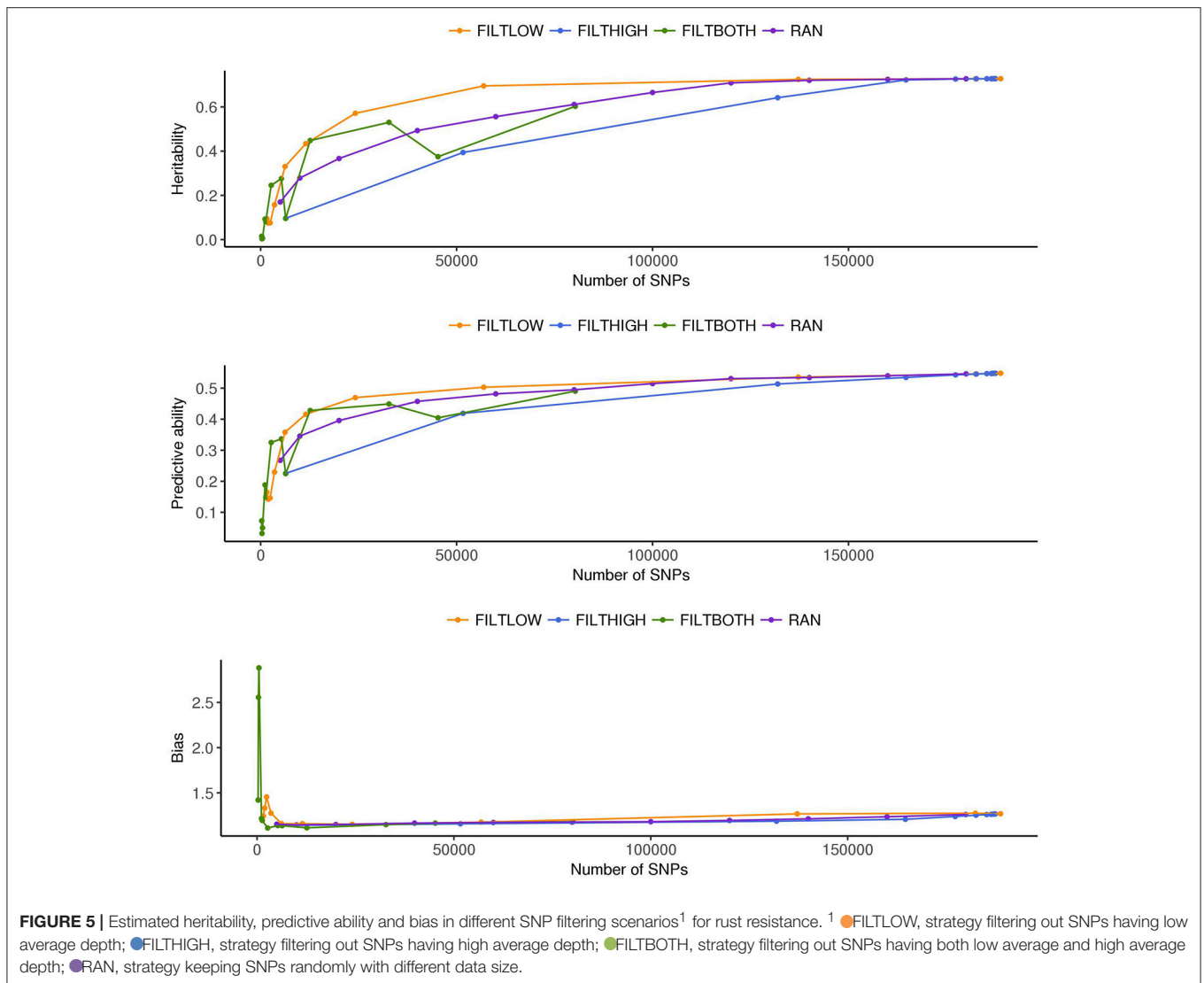Detailed results from CV for three traits are available in **Supplementary Table 5** (HD).

FIGURE 5 | Estimated heritability, predictive ability and bias in different SNP filtering scenarios[1] for rust resistance. [1] ●FILTLOW, strategy filtering out SNPs having low average depth; ●FILTHIGH, strategy filtering out SNPs having high average depth; ●FILTBOTH, strategy filtering out SNPs having both low average and high average depth; ●RAN, strategy keeping SNPs randomly with different data size.

Figure 3 (DM), Figure 4 (HD), and Figure 5 (RUST), show that the predictive ability generally increased when the number of SNPs included in the analysis increased. The highest predictive ability for DM was provided by dataset FILTLOW3 (0.34) with 137,191 SNPs having average depth higher than 10, the highest predictive ability for HD was provided by dataset FILTHIGH5 (0.77) with 185,297 SNPs having average depth lower than 60, and the highest predictive ability for RUST was provided by dataset FILTLOW1 (0.55) with 188,832 SNPs having average depth higher than 1, which was equivalent to including all markers.

Randomly filtering out SNPs and varying the number of SNPs showed that predictive ability generally increased with increasing number of SNPs included in the analysis. Above 80–100 K SNPs, effects of further increases were limited.

Overall, with an increase in the number of SNPs included in the analysis, the bias, which is the deviation from 1 for the regression of predictions on observed phenotypes, also increased. For all three traits, the FILTLOW strategy showed more biased predictions than RAN, whereas the FILTHIGH strategy showed

less biased prediction than RAN. In addition, larger bias was always observed together with better predictive ability. For DM, strategy FILTLOW3 provided best predictive ability, but the bias when using this subset of SNPs was also high (regression coefficient was 1.32). For HD, the most biased prediction was also provided by dataset FILTLOW3 (regression coefficient was 1.26), though the best predictive ability was provided by the dataset FILTHIGH5, the predictive ability when using dataset FILTLOW3 was only 0.01 lower than FILTHIGH5. For RUST, FILTLOW1 provided highest predictive ability, and the bias provided by this dataset was slightly larger (regression coefficient was 1.27) than other scenarios, except for models with few SNPs included.

DISCUSSION

To investigate the potential for genomic prediction in tetraploid ryegrass we analyzed data from 1,515 F_2 families. All families were genotyped using GBS with an average sequencing depth
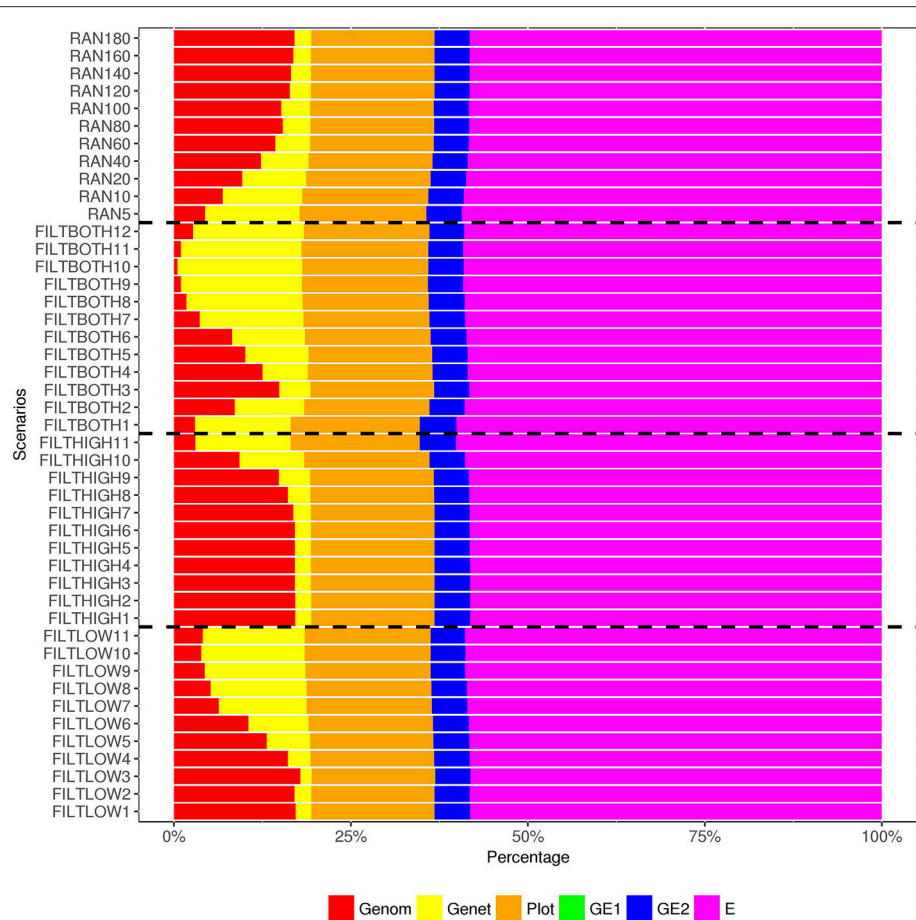
FIGURE 6 | Percentage of variance components[1,2] over the total phenotypic variance for dry matter yield. [1] ■Genom, Additive genomic variance; ■Gene, Residual genetic variance; ■PLOT, random plot variance; ■GE1, family × sowing year × location × management variance; ■GE2, family × sowing year × location × management × farming year variance; ■E, residual environment variance. [2] GE1 is too small to be visible, so that there are only five variances can be observed in this figure.

of 19. GBS data, with various strategies for filtering SNPs, were used in GP models, and we compared heritabilities and predictive abilities to determine optimal SNP numbers and sequencing depth for genomic prediction in tetraploid ryegrass. Among all the filtering scenarios, the highest estimates for genomic heritability of family means were 0.45, 0.74 and 0.73 for DM, HD and RUST, respectively. The predictive ability generally increased as the number of SNPs included in the analysis increased. The highest predictive ability for DM was 0.34 (137,191 SNPs having average depth higher than 10), for HD was 0.77 (185,297 SNPs having average depth lower than 60), and for RUST was 0.55 (188,832 SNPs having average depth higher than 1).

Heritabilities and Variance Components

Several studies have reported heritabilities in diploid ryegrass for the same traits studied here. Fè et al. (2015b) reported analysis of total DM in two years, and heritability ranged from 0.20 to 0.25, and the estimates of heritability of total DM over two years were slightly higher than in first and second year separately. In the current study, DM was defined as the total dry matter yield in

each of two farming years, and modeled as a trait with repeated records while the overall year effect was included in the fixed effects. The estimate of h_f^2 was higher than heritabilities reported by Fè et al. (2015b). Compared with the heritability for HD in diploids, where estimates ranged from 0.49 to 0.68 (Fè et al., 2015a,b) and from 0.07 to 0.22 (Ashraf et al., 2016), the estimates in the current study are higher. RUST was investigated in diploid varieties by Ravel and Charmet (1996), Waldron et al. (1998), Fè et al. (2015b), and Fè et al. (2016), and the estimates of h_f^2 in the current study were in the range reported for diploid ryegrass. In the previous study on diploid ryegrass by Fè et al. (2015b), estimated heritability for DM was similar to that for RUST, but in the current study we find a larger difference between estimated heritabilities for DM and RUST.

G×E effects accounted for about 10% of total variance for HD in diploid ryegrass (Fè et al., 2015a), which is similar to results in the current study. Although the proportion of total phenotypic variance explained by genetic marker information was much less in DM than in HD and RUST, the variances of G×E effects were also small in DM.
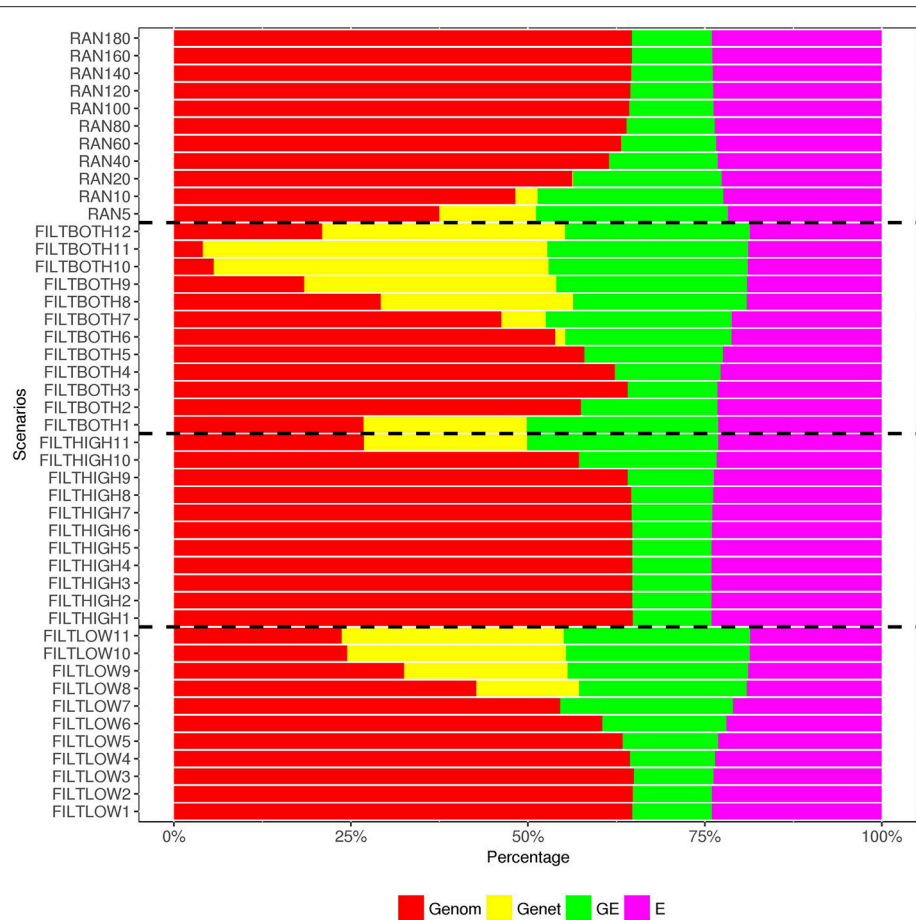
FIGURE 7 | Percentage of variance components[1] over the total phenotypic variance for heading date. [1] ■Genom, Additive genomic variance; ■Gene, Residual genetic variance; ■GE, family × sowing year × location × management variance; ■E, residual environment variance.

The proportions of variances due to the two G×E effects were different for DM and RUST. For DM, the second G×E effect (GSLMF) was important, but the first G×E effect (GSLM) explained only a small part, which indicated that growth season had a large effect on DM and can modify the ranking of different families. In diploid ryegrass, the genetic and phenotypic correlation between DM in the two years were 0.62 and 0.39 respectively (Fè et al., 2015b), which also implies large G×E effects. Variance of G×E effects in RUST was different from G×E in DM. For RUST, both the first and the second G×E effects accounted for similar amount of variances (around 12% of total phenotypic variance), which is comparable to results from the previous study on diploids (Fè et al., 2016).

The proportion of residual variance at the level of single plots was different among the three traits. The residual variance accounted for more than 50% of phenotypic variance in DM but only around 25% in both HD and RUST. The large amount of residual variance in DM indicates larger measurement errors in DM, and necessarily leaves only relatively small proportions of variance that can be attributed to the other effects.

SNP Filtering Strategies

Sequencing depth is an important factor when utilizing GBS data. An increase of sequencing depth means that the average number of times a locus been sequenced is increased, so that the accuracy of measuring the frequency of the reference allele is also increased. However, increasing sequencing depth also increases the cost of sequencing. Therefore, it is crucial to investigate the optimal sequencing depth when using GBS data. In the current study, different SNP filtering scenarios were compared with regard to parameter estimation and genomic prediction results. Four SNP filtering strategies were applied on the full GBS dataset, creating subsets of SNPs with different sequencing depth and/or different numbers of SNPs.

A previous study on diploid ryegrass (Ashraf et al., 2016) used GBS data with sequencing depth varying from 0 to 60, and divided the SNPs in 5 groups with depth interval of 10. Ashraf et al. (2016) did not correct for low accuracy of allele frequency estimates at low sequencing depth, and showed this creates a general trend of increasing genomic heritability with increasing sequencing depth. In the current study, we corrected for the effects of low accuracy at low sequencing
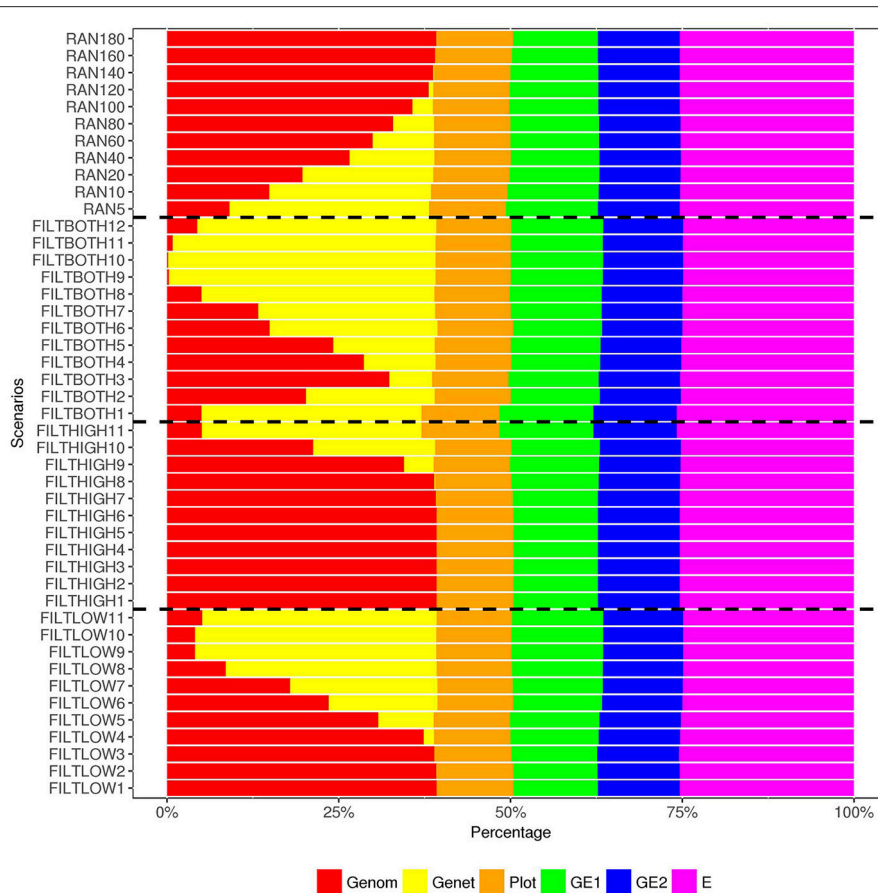
FIGURE 8 | Percentage of variance components[1] over the total phenotypic variance for rust resistance. [1] ■Genom, Additive genomic variance; ■Gene, Residual genetic variance; ■PLOT, random plot variance; ■GE1, family × sowing year × location × management variance; ■GE2, family × sowing year × location × management × farming year variance; ■E, residual environment variance.

depth, based on Cericola et al. (2018), and generally see no more clear linear correlation between sequencing depth and heritability. For instance, the FILTBOTH strategy also grouped SNPs into different depth intervals, and highest estimates of genomic heritability were found for the middle to lower levels FILTBOTH2 (depth 5–10) and FILTBOTH3 (depth 10–20). Comparable heritabilities were found between FILTBOTH 3 and RAN80, where both scenarios included similar amount of SNPs while the later one covering a larger range of sequencing depths (1–278). Hence, the corrections for bias from Cericola et al. (2018) are removing obvious trends related to sequencing depth, and seem to effectively remedy the problem of biased heritabilities at low sequencing depth reported by Ashraf et al. (2016).

For prediction accuracy, the impact from GBS sequencing depth was investigated in simulated biparental segregating populations (Gorjanc et al., 2017) as well as in outbred livestock populations (Gorjanc et al., 2015). The results from these two simulation studies showed that GBS data with low coverage (∼1X) could provide prediction accuracy comparable to SNP array data. When using field data, most of the studies were focused on settings with inbred individuals, e.g., in wheat

(Poland et al., 2012) and maize (Crossa et al., 2013). The accuracy of genomic prediction using low-coverage GBS data were comparable with SNP array or diversity array technology data in inbred populations (Poland et al., 2012; Crossa et al., 2013). Different from these simulation studies or studies on inbred populations, the current study is based on the commercial tetraploid data using family-pools. In our data, we cannot confirm that GBS data with low sequencing depth of about 1X already gives accurate predictions. As expected, high heterozygosity in tetraploid ryegrass, combined with use of family-pools, requires higher sequencing depth for accurate genomic prediction. In the current study, SNPs with sequencing depth between 10 and 20 (FILTBOTH3) delivered desirable predictive ability.

In the current study, by filtering out SNPs with either low sequencing depth (FILTLOW) or with high sequencing depth (FILTHIGH), the optimal sequencing depth for practical genomic prediction in tetraploid ryegrass was investigated. In FILTLOW groups, FILTLOW1 to FILTLOW3 gave most accurate predictions. The number of SNPs included in the models with highest predictive ability was about 140–180 k. The lowest sequencing depth for SNPs in FILTLOW3 was 10.

The similar high predictive ability provided by FILTLOW1 to FILTLOW3 indicated that excluding low sequencing depth (1–10) SNPs did not affect the predictive ability significantly. In FILTHIGH groups, FILTHIGH1 to FILTHIGH9 gave similar predictive abilities, which indicated that accurate predictions can be reached even by including only SNPs with sequencing depth lower than 20. This can simply be an effect of still having sufficiently large numbers of SNPs with depth lower than 20, and removing SNPs with high depth may reduce some noise caused by repetitive sequences. Hence, filtering out SNPs with high depth can increase the proportion of useful information without reducing the prediction accuracy. Compared with the RAN strategy, filtering out SNPs with low depth provided higher predictive ability than using a similar number of randomly chosen SNPs, and when comparing the FILTHIGH strategy with the RAN strategy, filtering out SNPs with high depth provided similar predictive ability as using a similar number of randomly chosen SNPs. For the three traits investigated in the current study, the best predictive abilities were not achieved with exactly the same filtering strategy, however, differences between the best filtering strategies were small. In practical breeding, single trait evaluation can be carried out by using **G** matrices built from different sets of SNPs. It is also feasible to apply index selection on a combination of traits with different weights by using a same set of SNPs, which can provide globally accurate predictions. For example, in the current study, even though the highest predictive ability was provided by FILTLOW3, FILTHIGH5, and FILTLOW1 for DM, HD, and RUST, respectively, FILTLOW3 can be chosen as a scenario that provided accurate predictions for all the three traits analyzed. In addition, applying different sets of SNPs at the same time is also achievable by using random regression models disregarding the higher demand of computing resources.

In addition to genomic prediction accuracy, bias was also investigated in this study. In general, it was observed that predictions were biased, and with increasing number of SNPs included in the model, more biased predictions were observed. This can be due to many factors. The definition of the **G** matrix could be one of the reasons. When using GBS data, the allele frequencies can suffer from inaccuracy due to low sequencing depth, which can induce bias into the prediction. However, in the current study, biases due to low sequencing depth was corrected for using the method reported by Cericola et al. (2018). In addition, G×E interactions were modeled in a rather simple way, and bias of prediction may be reduced by better modeling of G×E effects (Fè et al., 2015b).

For diploid heterozygotic species, a minimum sequencing depth of around 10X is needed to obtain accurate calling (Chenuil, 2012). However, for tetraploid species, the requirement of sequencing depth for accurate calling of tetraploid genotypes was reported to be 60–80X (Uitdewilligen et al., 2013). For genomic prediction, however, it is not necessary to obtain accurate calling for each individual sample. The results in the current study indicate that high predictive ability can be obtained using much lower sequencing depth because only the frequency and not the individual genotypes needs to be called.

CONCLUSIONS

In the current study, phenotypic records for three traits dry matter yield (DM), rust resistance (RUST), and heading date (HD), as well as GBS data were used to obtain genomic predictions for 1,515 tetraploid F_2 ryegrass families. Different SNP filtering strategies by filtering out SNPs according to average depth and number of SNPs were compared. The estimates of genomic heritability of family means were 0.45, 0.74, and 0.73 for DM, HD and RUST, respectively. The predictive ability for DM was 0.34, for HD was 0.77, and for RUST was 0.55. The estimation of genomic heritability and the predictive ability for DM, HD and RUST clearly showed that genomic prediction can be implemented in tetraploid perennial ryegrass. Comparison of different filtering strategies showed that using only SNPs with sequencing depth between 10 and 20 would not reduce predictive ability, and showed the potential to obtain accurate prediction from medium-low coverage GBS in tetraploids. Adding SNPs with sequencing depth lower than 10 in the model also lead to accurate predictions. The predictive ability generally increased as the number of SNPs included in the analysis increased. GBS data including 80–100 K SNPs were needed for accurate prediction of additive breeding values in tetraploid ryegrass. The results clearly illustrate that genomic prediction using GBS data can help to optimize the breeding program for tetraploid ryegrass.

AUTHOR CONTRIBUTIONS

XG implemented and carried out the statistical analysis, interpreted the results and had a major role in drafting the manuscript. FC and DF contributed to the statistical analysis and the result interpretation, reviewed the manuscript. MP managed the production of the plant material and the acquisition of the phenotypic data. IL carried out part of the GBS sequencing, and subsequent bioinformatics analysis. CJ conceived the experiment and reviewed the manuscript. JJ conceived the experiment, contributed in developing the statistical models and interpreting the results, reviewed the manuscript. LJ contributed in developing the statistical models and interpreting the results, reviewed the manuscript. All authors read and approved the final manuscript.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2018.01165/full#supplementary-material

REFERENCES

Ashraf, B. H., Byrne, S., Fe, D., Czaban, A., Asp, T., Pedersen, M. G., et al. (2016). Estimating genomic heritabilities at the level of family-pool samples of perennial ryegrass using genotyping-by-sequencing. *Theor. Appl. Genet.* 129, 45–52. doi: 10.1007/s00122-015-2607-9

Ashraf, B. H., Jensen, J., Asp, T., and Janss, L. L. (2014). Association studies using family pools of outcrossing crops based on allele-frequency estimates from DNA sequencing. *Theor. Appl. Genet. Theoretische Und Angewandte Genetik* 127, 1331–1341. doi: 10.1007/s00122-014-2300-4

Beissinger, T. M., Hirsch, C. N., Sekhon, R. S., Foerster, J. M., Johnson, J. M., Muttoni, G., et al. (2013). Marker density and read depth for genotyping populations using genotyping-by-sequencing. *Genetics* 193, 1073–1081. doi: 10.1534/genetics.112.147710

Byrne, S., Czaban, A., Studer, B., Panitz, F., Bendixen, C., and Asp, T. (2013). Genome wide allele frequency fingerprints (GWAFFs) of populations via genotyping by sequencing. *PLoS ONE* 8:e57438. doi: 10.1371/journal.pone.0057438

Byrne, S. L., Nagy, I., Pfeifer, M., Armstead, I., Swain, S., Studer, B., et al. (2015). A synteny-based draft genome sequence of the forage grass *Lolium perenne*. *Plant J.* 84, 816–826. doi: 10.1111/tpj.13037

Cericola, F., Lenk, I., Fè, D., Byrne, S., Jensen, C. S., Pedersen, M. G., et al. (2018). Optimized use of low-depth genotyping-by-sequencing for genomic prediction among multi-parental family pools and single plants in perennial ryegrass (*Lolium perenne* L.). *Front. Plant Sci.* 9:369. doi: 10.3389/fpls.2018. 00369

Chenuil, A. (2012). How to infer reliable diploid genotypes from NGS or traditional sequence data: from basic probability to experimental optimization. *J. Evol. Biol.* 25, 949–960. doi: 10.1111/j.1420-9101.2012. 02488.x

Conaghan, P., and Casler, M. D. (2011). A theoretical and practical analysis of the optimum breeding system for perennial ryegrass. *Irish J. Agric. Food Res.* 50, 47–63. Available online at: http://www.jstor.org/stable/41348155

Cornish, M. A., Hayward, M. D., and Lawrence, M. J. (1979). Self-incompatibility in ryegrass I. Genetic control in diploid *Lolium perenne L. Heridity* 43, 95–106. doi: 10.1038/hdy.1979.63

Crossa, J., Beyene, Y., Kassa, S., Perez, P., Hickey, J. M., Chen, C., et al. (2013). Genomic prediction in maize breeding populations with genotyping-by-sequencing. *G3 (Bethesda)* 3, 1903–1926. doi: 10.1534/g3.113.0 08227

Crossa, J., Campos Gde, L., Perez, P., Gianola, D., Burgueno, J., Araus, J. L., et al. (2010). Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics* 186, 713–724. doi: 10.1534/genetics.110.118521

Crossa, J., Perez, P., Hickey, J., Burgueno, J., Ornella, L., Ceron-Rojas, J., et al. (2014). Genomic prediction in CIMMYT maize and wheat breeding programs. *Heredity (Edinb)* 112, 48–60. doi: 10.1038/hdy. 2013.16

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6:e19379. doi: 10.1371/journal.pone.00 19379

Fè, D., Ashraf, B. H., Pedersen, M. G., Janss, L., Byrne, S., Roulund, N., et al. (2016). Accuracy of genomic prediction in a commercial perennial ryegrass breeding program. *Plant Genome* 9. doi: 10.3835/plantgenome2015. 11.0110

Fè, D., Cericola, F., Byrne, S., Lenk, I., Ashraf, B. H., Pedersen, M. G., et al. (2015a). Genomic dissection and prediction of heading date in perennial ryegrass. *BMC Genomics* 16:921. doi: 10.1186/s12864-015-2163-3

Fè, D., Pedersen, M. G., Jensen, C. S., and Jensen, J. (2015b). Genetic and environmental variation in a commercial breeding program of perennial ryegrass. *Crop Sci.* 55, 631–640. doi: 10.2135/cropsci2014.06.0441

Garner, C. (2011). Confounded by sequencing depth in association studies of rare alleles. *Genet. Epidemiol.* 35, 261–268. doi: 10.1002/gepi.20574

Gorjanc, G., Cleveland, M. A., Houston, R. D., and Hickey, J. M. (2015). Potential of genotyping-by-sequencing for genomic selection in livestock populations. *Genet. Sel. Evol.* 47:12. doi: 10.1186/s12711-015-0102-z

Gorjanc, G., Dumasy, J.-F., Gonen, S., Gaynor, R. C., Antolin, R., and Hickey, J. M. (2017). Potential of low-coverage genotyping-by-sequencing and imputation for cost-effective genomic selection in biparental segregating populations. *Crop Sci.* 57, 1404–1420. doi: 10.2135/cropsci2016. 08.0675

Hayes, B. J., Cogan, N. O. I., Pembleton, L. W., Goddard, M. E., Wang, J., Spangenberg, G. C., et al. (2013). Prospects for genomic selection in forage plant species. *Plant Breed.* 132, 133–143. doi: 10.1111/pbr. 12037

Hickey, J. M., Chiurugwi, T., Mackay, I., Powell, W., and Implementing Genomic Selection in CGIAR Breeding Programs Workshop Participants (2017). Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. *Nat. Genet.* 49, 1297–1303. doi: 10.1038/ng.3920

Humphreys, M. O. (2005). Genetic improvement of forage crops–past, present and future. *J. Agric. Sci.* 143, 441–448. doi: 10.1017/s0021859605 005599

Jensen, C. S., Salchert, K., and Nielsen, K. K. (2001). A Terminal flower1-like gene from perennial ryegrass involved in floral transition and axillary meristem identity. *Plant Physiol.* 125, 1517–1528. doi: 10.1104/pp.125.3.1517

Lantinga, E. A., and Groot, J. C. J. (1996). Optimization of grassland production and herbage feed quality in an ecological context. *EAAP Publication* 84, 58–67.

Legarra, A., Robert-Granié, C., Manfredi, E., and Elsen, J.-M. (2008). Performance of genomic selection in mice. *Genetics* 180, 611–618. doi: 10.1534/genetics.108.088575

Lin, Z., Cogan, N. O. I., Pembleton, L. W., Spangenberg, G. C., Forster, J. W., Hayes, B. J., et al. (2016). Genetic gain and inbreeding from genomic selection in a simulated commercial breeding program for perennial ryegrass. *Plant Genome* 9. doi: 10.3835/plantgenome2015.06.0046

Madsen, P., and Jensen, J. (2013). *A User's Guide to DMU.* Tjele: University of Aarhus, Faculty Agricultural Sciences (DJF), Department of Genetics and Biotechnology Research Centre Foulum. Available online at: http://dmu.agrsci. dk/DMU/Doc/Current/

McDonagh, J., O'Donovan, M., McEvoy, M., and Gilliland, T. J. (2016). Genetic gain in perennial ryegrass (*Lolium perenne*) varieties 1973 to 2013. *Euphytica* 212, 187–199. doi: 10.1007/s10681-016-1754-7

Meuwissen, T. H., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829.

O'Donovan, M., and Delaby, L. (2005). A comparison of perennial ryegrass cultivars differing in heading date and grass ploidy with spring calving dairy cows grazed at two different stocking rates. *Anim. Res.* 54, 337–350. doi: 10.1051/animres:2005027

Poland, J., Endelman, J., Dawson, J., Rutkoski, J., Wu, S., Manes, Y., et al. (2012). Genomic selection in wheat breeding using genotyping-by-sequencing. *Plant Genome J.* 5, 103–113. doi: 10.3835/plantgenome2012.06.0006

Poland, J. A., and Rife, T. W. (2012). Genotyping-by-Sequencing for Plant Breeding and Genetics. *Plant Genome J.* 5, 92–102. doi: 10.3835/plantgenome2012.05.0005

Ravel, C., and Charmet, G. (1996). A comprehensive multisite recurrent selection strategy in perennial ryegrass. *Euphytica* 88, 215–226. doi: 10.1007/BF00023893

Sampoux, J.-P., Baudouin, P., Bayle, B., Béguier, V., Bourdon, P., Chosson, J.-F., et al. (2011). Breeding perennial grasses for forage usage: an experimental assessment of trait changes in diploid perennial ryegrass (*Lolium perenne* L.) cultivars released in the last four decades. *Field Crops Res.* 123, 117–129. doi: 10.1016/j.fcr.2011.05.007

Sims, D., Sudbery, I., Ilott, N. E., Heger, A., and Ponting, C. P. (2014). Sequencing depth and coverage: key considerations in genomic analyses. *Nat. Rev. Genet.* 15, 121–132. doi: 10.1038/nrg3642

Uitdewilligen, J. G., Wolters, A. M., D'Hoop, B. B., Borm, T. J., Visser, R. G., and van Eck, H. J. (2013). A next-generation sequencing method for genotyping-by-sequencing of highly heterozygous autotetraploid potato. *PLoS ONE* 8:e62355. doi: 10.1371/journal.pone.0062355

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

Waldron, B. L., Ehlke, N. J., Wyse, D. L., and Vellekson, D. J. (1998). Genetic variation and predicted gain from selection for winterhardiness

and turf quality in a perennial ryegrass topcross population. *Crop Sci.* 38, 817–822.

Wilkins, P. W. (1991). Breeding perennial ryegrass for agriculture. *Euphytica* 52, 201–214. doi: 10.1007/BF00029397

Wilkins, P. W., and Humphreys, M. O. (2003). Progress in breeding perennial forage grasses for temperate agriculture. *J. Agric. Sci.* 140, 129–150. doi: 10.1017/s0021859603003058

Zhang, X., Perez-Rodriguez, P., Semagn, K., Beyene, Y., Babu, R., Lopez-Cruz, M. A., et al. (2015). Genomic prediction in biparental tropical maize populations in water-stressed and well-watered environments using low-density and GBS SNPs. *Heredity (Edinb)* 114, 291–299. doi: 10.1038/hdy.2014.99