Genomic prediction of agronomic traits in perennial ryegrass (*Lolium perenne* L.) and genotype x environment interactions at the limit of the species distribution

Natasha H. Johansen[1], Andrea Bellucci[1,2], Pernille B. Hansen[1,3], Petter Marum[4], Helga Amdahl[4], Kristin Håland Gylstrøm[4], Odd Arne Rognli[5], Vilma Kemešytė[6], Gintaras Brazauskas[6], Morten Greve[7], Christer Persson[8], Mika Isolahti[9], Áslaug Helgadóttir[10], Rene Aavola[11], Torben Asp[1], Guillaume P. Ramstein[1].

1. Center for Quantitative Genetics and Genomics, Aarhus University, Denmark
2. Carlsberg Research Laboratory, Denmark
3. Nordic Seed A/S, Denmark
4. Graminor, Norway
5. Norwegian University of Life Sciences, Norway
6. Institute of Agriculture, Lithuanian Research Centre for Agriculture and Forestry, Lithuania
7. DLF Seeds A/S, Denmark
8. Lantmännen Seed, Sweden
9. Boreal Plant Breeding, Finland
10. The Agricultural University of Iceland, Iceland
11. Centre of Estonian Rural Research and Knowledge, Estonia

Abstract:

Background: In breeding the aim is to identify and accumulate beneficial variants. However, detection of these variants may be challenging in the presence of extensive genotype x environment interactions (GxE), as variant effects will be conditional on environment. The study assesses the performance of **264** diploid perennial ryegrass accessions in a multi-environment field trial.

Methods: We investigate the extent of GxE, for yield (total dry matter) and persistence traits, i.e. winter kill, and spring cover, under environmental conditions experienced in Nordic and Baltic regions at the limit of the species distribution. Two different approaches to modelling GxE were tested: reaction norm and envirotyping. Models were validated under different breeding scenarios.

Results: Our analysis documented the presence of significant GxE interaction for all traits investigated in the study. Validation showed improvements in prediction accuracy when accounting for GxE: up to 4% for grain yield when predicting in unobserved environments, and up to 21% for spring cover when predicting in unobserved germplasm. Genome-wide-association analyses (GWAS) were utilized to detect genetic variants with marginal effects (environment-independent effect) and conditional effects (environment-dependent effects). Results showed the presence of large-effect genetic variants with marginal effects, in addition to few QTL whose effects were adaptive under specific environmental conditions while neutral or deleterious under different environmental conditions.

Conclusion: This study demonstrates the usefulness and limitations of genomic prediction models for predicting GxE in highly diverse samples, and describes the extent of GxE interactions at the limit of species distribution for perennial ryegrass. Finally, our study points toward adaptive variation, which may enhance persistence of perennial ryegrass populations in Nordic and Baltic growing conditions.

1. Introduction

Climate changes will introduce further challenges for plant breeders and crop growers in the coming years, as they will lead to increased fluctuations in temperature and rainfall, and therefore, more unpredictable weather patterns (Raza et al. 2019). Consequently, to ensure that future food demand can be supported, it is necessary to adapt current food production practices and develop high yielding crop varieties that are adapted to unpredictable weather patterns (Kahiluoto et al. 2019**).** Different selection strategies have been developed to facilitate effective determination of superior breeding material. These selection strategies include, among other, Genomic Prediction (GP) (Meuwissen, Hayes, and Goddard 2001) which has become an important tool to predict the values of accessions from genetic markers.

Plant breeders evaluate candidate lines/populations in multiple environmental trials (MET) to ensure that these lines/populations have stable performance, i.e., broad adaptation, across a range of different environments. These environments are referred to as the target population of environments (TPE), which encompasses the intended production environments of the lines. However, broad adaptation does not necessarily imply superior performance across environments, as some genetic material may express superior performance in a narrow range of environments while performing poorly across the majority of the TPE (Ågren et al. 2013).

Hence, differences in phenotypic performance between lines/populations are not exclusively caused by genetic effects but are also affected by genotype by environment interactions (GxE), i.e., differential responses of genotypes over an environmental gradient. GxE can be expressed as (i) non-cross-over interaction, i.e., genotypes maintain their rankings but the variation in phenotypic performances is changed which is also referred to as scale changes, or alternatively (ii) cross-over interaction where the rankings of the genotypes are changed. Cross-over interactions occur when the rankings of genotypes are environment-dependent (El-Soda et al. 2014). Traditionally, plant breeders have ignored or minimised GxE interactions to avoid the challenge of modelling this component, however studies show that the prediction accuracy of GP models can be enhanced in MET, by accounting for GxE (Jarquin et al. 2020; Millet et al. 2019). Furthermore, the integration of GxE in a statistical model allows for prediction of phenotypic performance in unobserved environments, which can be of great advantage to breeders.

Different approaches have been developed to integrate GxE into genomic prediction models. These approaches include, amongst others extensions of GBLUP models which account for covariance of environmental values and GxE interactions: reaction norm (Jarquín et al. 2014) and envirotyping approaches (Costa-Neto, Fritsche-Neto, and Crossa 2021; Costa-Neto et al. 2021; Costa-Neto, Crossa, and Fritsche-Neto 2021; Xu 2016). These approaches differ regarding how the environmental covariates are transformed. The reaction-norm approach models GxE by genotypes' linear response to environmental covariates. In reaction-norm models, more complex interactions, i.e., non-linear interactions, are not captured (Jarquín et al. 2014).

Envirotyping is an approach by which environments are characterised by determining the environmental covariates that affect a given plant species' growth (Costa-Neto, Fritsche-Neto, and Crossa 2021; Costa-Neto et al. 2021; Costa-Neto, Crossa, and Fritsche-Neto 2021; Xu 2016). The envirotyping approach is comparable to the reaction-norm approach but differs in how environmental relatedness is calculated. For the reaction-norm approach environmental relatedness is based on quantitative descriptors (e.g., mean temperature) whereas for the envirotyping approach environmental gradients are partitioned into environmental zones, or typologies, according to physiological expectations about survival, tolerance, or optimal growth for the given crop species, as inspired by Shelford's Law of Tolerance (Shelford 1931). The number of occurrences of these environmental zones are determined for each environment, which allows for environmental relatedness to be based on the frequency of each environmental-zone that an environment experiences and therefore also informs about the value of the environment for the given crop (Costa-Neto, Fritsche-Neto, and Crossa 2021; Costa-Neto, Crossa, and Fritsche-Neto 2021; Costa-Neto et al. 2021).

The aim of GP is to identify superior lines based on genome-wide variants, but GP models usually do not detect large-effect variants explicitly. Genome-Wide-Association-studies (GWAS) are commonly used to detect statistical associations between variants, or quantitative trait loci (QTL), and phenotypes. Linkage disequilibrium (LD), i.e., non-random association of alleles at different loci, is integral for GWAS as the power and ability to detect genotype-phenotype associations is lower when neutral alleles are not associated with causal variants (Flint-Garcia, Thornsberry, and Buckler 2003). Association analysis in populations with rapid LD decay therefore requires high marker densities, as genetic variants will be linked together in smaller blocks, compared to populations with more extensive LD. GWAS are commonly used to detect QTL based on their average association with traits, also termed marginal effects (Atwell et al. 2010). However, GWAS are also a useful tool for the detection of variants with conditional effects, which can occur in the presence of local adaptation. Conditional neutrality of variant effect and genetic trades-offs (maladaptive pleiotropy) may both produce conditional variant effects. Conditional neutrality implies that a variant may be beneficial or alternatively deleterious under certain environmental settings, while effectively neutral in others, whereas alleles showing maladaptive pleiotropy will confer a fitness advantage under certain environmental conditions, whereas the same allele may under different environmental conditions incur a fitness cost (Anderson et al. 2013; Lasky, Forester, and Reimherr 2018).

Perennial ryegrass (*Lolium perenne* L.*)* is a highly valued temperate forage grass in dairy production, because of high yields and nutritional value. Additionally, this crop expresses a high level of genetic diversity, due to its obligate outbreeding nature (Brazauskas et al. 2011). Recently, there has been an interest, amongst commercial breeders of perennial ryegrass, in extending the production of this crop to more Northern regions of Europe (Solberg, Rognli, and Østrem 1994; Helgadóttir et al. 2018).

Nordic regions, which have previously been unsuitable for production of perennial ryegrass, can become fit for production of certain robust and cold-tolerant varieties in the near future, as temperatures increase in these regions in response to climate changes (Ergon et al. 2018; Helgadóttir et al. 2018). However, there are challenges to this endeavour, including the presence of high levels of GxE interactions, which have previously been documented in perennial ryegrass (Fè et al. 2015; Grogan and Gilliland 2011). These GxE interactions include rank changes, i.e., environment-specific ranking of accessions in respect to productivity (Grogan and Gilliland 2011) and large GxE interactions at year and field level (Fè et al. 2016).

Improved insight into the extent of GxE, in perennial ryegrass, in Nordic and Baltic regions may benefit future breeding efforts toward more robust perennial ryegrass germplasm. Few studies have previously investigated GxE in perennial ryegrass and general questions remain about the effective estimation of GxE models in diverse samples. Hence, this study aims to: (i) investigate and compare the performance of perennial ryegrass accessions in a multi-environment field trial, to improve biomass yield and winter survival; (ii) determine the extent of GxE interactions in perennial ryegrass to assess their importance for adaptation to extreme environments at the margin of the species' distribution, and (iii) detect adaptive variants which confer significant advantages, in all or specific environments.

2. Material And Methods

2.1 Plant material and phenotyping

A total of 264 diploid perennial ryegrass accessions were included in this study. These accessions consisted of advanced cultivars, wild/semi-wild accessions, and landraces obtained from gene banks worldwide, including the Nordic Genetic Resource Center (Rognli et al. 2018). These accessions were multiplied to obtain enough seeds for regular field trials. For the study, a diverse set of accessions was chosen, in respect to geographic origin, which spanned Central Europe, Northern Europe, Russia, US and Japan (Supplementary Table 1). The experiment was conducted across 2014-2016 in Norway (60°45' N, 11°12' E), Estonia (58°45' N, 26°24' E), Sweden (55°56' N, 13°6' E), Lithuania (55°23' N, 23°52' E), Iceland (64°9' N, 21°45' W) and Denmark (55°20' N, 12°23' E).

Replicated field trials were established in late-spring or summer in 2014, in Norway (20th of June), Estonia (2nd of August), Sweden (28th of May), Lithuania (10th of June), Iceland (9th of July), and Denmark (25th of June). The elevation, in metres above sea level, or m.a.sl, at the trial sites were as follows: Norway (187 m.a.sl), Estonia (68 m.a.sl), Sweden (65 m.a.sl), Lithuania (33 m.a.sl), Iceland (35 m.a.sl) and Denmark (35 m.a.sl). The following traits were measured in the study: total dry matter yield from three harvests during the growing season (TDM3), spring cover (SpringCover), winter kill (WinterKill). The traits SpringCover and WinterKill are transformed to a scale ranging from 1 to 9, where SpringCover, i.e., percentage of plot covered by the crop, is a measure of the crop's establishment ability, while WinterKill is measured as the decrease in plot cover from autumn to spring, with 1 for highest survival, 9 for lowest. The data was unbalanced with respect to the number of accessions scored in each country, with the number of accessions varying between 108 and 262 depending on the environment, while across all environments 264 accessions were included. Hence, of the 264 accessions included in the study, a 'core' set of 108 accessions were scored in all countries.

Table 1: Number of diploid accessions scored in each country for the traits

	Norway		**Estonia**		**Sweden**		**Lithuania**		**Iceland**		**Denmark**	
	2015	**2016**	**2015**	**2016**	**2015**	**2016**	**2015**	**2016**	**2015**	**2016**	**2015**	**2016**
TDM3	213	213	144	144	108	108	118	118	-	-	262	262
SpringCover	213	213	144	144	-	-	118	118	142	142	-	-
WinterKill	-	-	-	144	-	108	-	118	-	-	-	-

2.2 Genotyping-by-sequencing

Sequence data was produced using genotyping-by-sequencing (Elshire et al. 2011), with the methylation-sensitive restriction enzyme ApeKI to target the low copy fraction of the genome. Sampling and library preparation followed the protocol described by (Byrne et al. 2013). The accessions were genotyped based on a pooled sample.

After alignment of sequencing reads to the reference genome (Nagy et al. 2022) and SNP calling by GATK (McKenna et al. 2010), a total of 380,252 SNPs were obtained. VCF-tools (Danecek et al. 2011) were used to filter the SNP markers that did not meet the following requirements: minDP $\geq$ 10, min-mean DP $\geq$ 20, max-meanDP $\geq$ 100, minQ $\geq$ 30, minor allele frequency across accessions (MAF) $\geq$ 0.05, missing rate lower than 20%. After filtering, 151,074 SNPs remained. We then calculated the allele frequency, in each accession *i* where pseudo-counts were used to reduce bias introduced by low read depths: $p_{ik} = \frac{AD_{ALT,ik}+1}{AD_{ALT,i}k+AD_{REF,ik}+2}$, where $AD_{REF,ik}$ and $AD_{ALT,ik}$ are the depths of the alternate and reference alleles, respectively, in accession *i* at marker *k*. Missing information at a given SNP was imputed as the average allele frequency across accessions.

2.3 Genomic relationship matrices

To construct the covariance matrix for additive effects the genomic relationship matrix (GRM) was constructed as $G_A = \frac{\mathbf{XX^T}}{tr(\mathbf{XX^T})/n}$ (Vitezica et al. 2017), where **X** is the centered matrix of alternate allele frequency matrix for each accession and marker, *tr* is the sum of the matrix diagonals, and *n* the number of accessions. To construct the covariance matrix for dominance effects, we assumed Hardy-Weinberg equilibrium in the populations, and estimated marker-specific heterozygosity as $H_{ik} = 2p_{ik}\, q_{ik}$, where p_{ik} is the frequency of the alternative allele, and q_{ik} is the frequency of the reference allele in accession *i* at marker *k*. Dominance codes were then computed as the residuals of estimated heterozygosity regressed on alternative allele frequency, by extending the approach of (Vitezica et al. 2017). The genomic relationship matrix for the dominance effects was then constructed similarly to the additive GRM.

2.4 Environmental data and environmental covariates

Daily weather and climatic variables were extracted for the six trial locations in the period 2014-2016. The information was collected with the R package EnvRtype (Costa-Neto et al. 2021). Different environmental covariance matrices are constructed for two GxE modelling approaches: (i) reaction norm (RN) and (ii) envirotyping (ET). Environmental similarity was determined based on 18 environmental covariates

(Supplementary table 2). For the RN approach we determined the environment means for each environmental covariate x time-window combination.

For the ET approach, daily measurements for each of the environmental covariates were partitioned into quantile intervals (0.01, 0.25, 0.50, 0.99) specific to each time window, intended to characterize environmental zones or envirotypes. The frequency of each envirotype was then determined for each country x year x time-interval combination. For TDM3, the environmental covariates were extracted from the period of the 1st of September, in the year prior to harvest, to the 31st of August the following year (harvest year). For SpringCover and WinterKill, environmental covariates were extracted from the 1st of September in the year prior to when the trait was scored until the 31st of May (year trait is scored). The R package EnvRtype requires the specification of lower and upper temperature bounds for crop development for the model crop, to which we used experimentally determined temperature bounds as estimated in (Monks, Sadatasilan, and Moot 2009). In both approaches the time interval was specified to span a 30-day window. The covariance matrices, denoted by **E**, were computed with the same approach as used for the genomic relationship matrices. For the RN and ET models we defined environments as country-year combinations, hence 12 environments were included in the study (Table 1).

2.5 Phenotypic analysis

Linear mixed models were fitted separately in each location to estimate the effects of accession and their interaction with year. Phenotypic records were regressed on accession (fixed), year (random), the accession-by-year interactions (random), spatial effects nested in year (random), and model residuals. The effects of year, the effects of accession-by-year interactions, and model residuals were modelled as independent and identically distributed (*i.i.d.*), following a normal distribution with the same variance for each type of random effect. Random spatial effects differed by countries, due to differences in experimental design: in Denmark, effects of trial and effects of row and column across the field; in Iceland and Lithuania, effects of replicate and effects of row and column in each replicate; in Estonia, Norway and Sweden, effects of trial and effects of row and column in each trial. Effects of trials and replicates were modelled as *i.i.d.*, whereas effects of rows and columns were modelled as normally distributed following a first-order autoregressive correlation structure, by row and column. In each year, accession means were calculated as the sum of the estimated effects of accession and accession-by-year interactions. Due to differences in managements across countries, accession means were centered and scaled in each country x year combination, based on the mean and standard deviation estimated in the core collection of 108 accessions (to account for the unbalanced number of accessions by country).

2.6 Genomic models

2.6.1 Main effects of genotype and environment (M1)

Model M1 is a genomic model (GBLUP) which accounts for the main effect of genotype and environment.

$$y_{ij} = \mu + a_i + e_j + \varepsilon_{ij} \text{ (M1)}$$

Where y_{ij} is the estimated mean phenotypic performance for accession i in environment j (country-year combination); e_j the random effect of environment j $e \sim N(0, \sigma_E^2\mathbf{E})$; μ is the intercept, or grand mean; a_i is the random additive genetic effect of accession i, $a \sim N\left(0, \sigma_{G_A}^2\mathbf{G_A}\right)$; and ε_{ijk} is the error term, $\varepsilon \sim N(0, \sigma_\varepsilon^2\mathbf{I})$, where **I** is the identity matrix.

2.6.2 Additive genetic effect and environment interaction (M2)

The M1 model is extended by adding the GxE interaction term, which is calculated by the Hadamard product of the covariance for the accessions and the covariance for the environments, given as $\mathbf{G_A} \otimes \mathbf{E}$, where $ae \sim N\left(0, \mathbf{G_A} \otimes \mathbf{E}\, \sigma_{G_{AE}}^2\right)$;

$$y_{ij} = \mu + a_i + e_j + ae_{ij} + \varepsilon_{ij} \qquad \text{(M2)}$$

2.6.3 Additive and dominance genetic effects and their interaction with the environment (M3)

In model M3 model we added the random genetic effect of dominance d_i and its interaction with the environment, where $d \sim N(0, \mathbf{G_D}\sigma_{G_D}^2)$ and $de \sim N\left(0, \mathbf{G_D} \otimes \mathbf{E}\, \sigma_{G_{DE}}^2\right)$.

$$y_{ij} = \mu + a_i + e_j + ae_{ij} + d_i + de_{ij} + \varepsilon_{ij} \qquad \text{(M3)}$$

2.6.4 Heritability

Narrow-sense heritability, h^2, a measure of the phenotypic variance which can be explained by the additive genetic variance was calculated based on the variance components from the M3 model as:

$$h^2 = \frac{V_A}{V_P} = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_D^2 + \sigma_{AxE}^2 + \sigma_{DxE}^2 + \sigma_\varepsilon^2}$$

2.7 Model comparison and prediction accuracy

Likelihood ratio tests based on restricted maximum likelihoods (REML) were used to compare models, and to investigate the statistical significance of additional variance components. The likelihood ratios were calculated as: $LR = -2\,[l_0(\theta) - l_1(\theta)]$, where $l_1(\theta)$ is the log-likelihood of the tested model, and $l_0(\theta)$

the log-likelihood of the null model. The likelihood ratio test statistic follows a chi-square distribution, where the degrees of freedom is the difference in the number of parameters between the models.

Model performance was compared under two different validation schemes. For this purpose, the data was separated into a training set and a validation set. Prediction accuracy was calculated for each fold as the Pearson correlation between observed accession means and predicted accession means $\hat{y}_{ij}$: $Cor(y_{ij}, \hat{y}_{ij})$. Finally, model performance was calculated as the mean correlation across all folds within the given validation scheme. The validation schemes under which the models were evaluated were: (i) Leave-one-country-out, where the models' ability to extrapolate to unobserved environments was tested, and (ii) Leave-one-genetic-cluster-out, where we divided accessions into 10 genetic clusters based on genetic similarity. Genetic clusters were inferred by hierarchical clustering based on Ward's criterion on the matrix of alternate allele frequencies **X**, with Euclidean distance as distance metric. Hence, this validation scheme tested the models' ability to predict the performance of unobserved accession clusters in observed environments. An additional validation-scheme, Leave-one-random-cluster-out, was included, where accessions are randomly assigned to ten clusters, the validation-scheme is used to evaluate model ability to predict unobserved accessions when related genetic material is included in the training-set.

2.8 Genome-Wide-Association study (GWAS)

A linear mixed model approach was applied to identify associations between accession means and genetic variants. Two different GWAS models were fitted: (i) a 'marginal' GWAS model to detect QTL with consistent effects in the suite of environments included in the study; and (ii) a 'conditional' GWAS to discover environment-specific QTL effects.

To account for population structure, we included the fixed effects of the first three principal components from a principal component analysis (PCA) of the centered allele frequency matrix **X**.

In marginal GWAS, the fixed effect of the tested SNP was $s_i\alpha$, where s_i is the alternate allele frequency at the tested SNP for accession *i* and α is the effect of the alternate allele on the phenotype:

$$y_{ijk} = q'\beta + c_j + s_i\alpha + a_i + d_i + ac_{ij} + dc_{ij} + e_{jk} + \varepsilon_{ijk} \qquad \text{(Marginal GWAS)}$$

where **q** is the vector of the principal components for accession *i* and **β** is the vector of associated effects; c_j is the fixed effect of country *j*; a_i and d_i are the additive and dominance genetic effects as described above; ac_{ij} and dc_{ij} are interactions between additive genetic effects and country, $ac \sim N\left(0,\ \mathbf{G_A} \otimes \mathbf{I}\, \sigma^2_{G_{AC}}\right)$ and $dc \sim N\left(0,\ \mathbf{G_D} \otimes \mathbf{I}\, \sigma^2_{G_{DC}}\right)$; e_{jk} is the random effect of environment *jk* (combination of country *j* and year *k*); ε_{ijk} is the model residual.

To detect genetic variants in linkage with QTL which have environment-specific effects, we extended the marginal GWAS model by including the interaction between SNP and country:

$$y_{ijk} = q'\beta + c_j + s_i(\alpha + \gamma_j) + a_i + d_i + ac_{ij} + dc_{ij} + e_{jk} + \varepsilon_{ijk} \quad \text{(Conditional GWAS)}$$

where δ_j is effect of the SNP in country j. Significance was assessed by a Wald test. To correct for multiple testing, statistical significance was assessed by the false discovery rate (FDR) (Benjamini and Hochberg 1995).

2.9 Software

All models were constructed in R 4.2.0 **(**Development Core Team**)** within high-performance computing cluster GenomeDK. The R package MM4LMM version 3.0.2 (Laporte, Charcosset, and Mary-Huard 2022) was used to extract model likelihoods, model validation and to run the GWAS analysis. The R package EnvRtype version 1.1.1 (Costa-Neto et al. 2021) was used to extract environmental covariates and construct the environmental covariance matrices for the two modelling approaches. For the phenotypic analysis we used the software ASREML-R version 4.2 (Butler et al., n.d.). The R package Dendextend was used to infer genetic clusters (Galili 2015).

3. Results

3.1 Perennial ryegrass accessions are highly diverse and exhibit minimal population structure

Generally, the germplasm included in this study was quite diverse, with the first PC accounting for less than 2 % of genomic variation. Additionally, we observed rapid LD-decay, with the genome-wide average LD decreasing to below $r^2 = 0.20$ after 200 bp (Figure 1). This is indicative of a germplasm being highly genetically dissimilar, and that alleles at loci in close physical proximity are largely randomly associated. There was a moderate correlation (0.62) between the additive and dominance genomic relationship matrices for the off-diagonal elements, which indicates that there is some collinearity between the additive and the dominance effects. We identified 10 genetic clusters based on hierarchical clustering, which were used for the leave-one-genetic-cluster-out validation scheme (Figure 1).

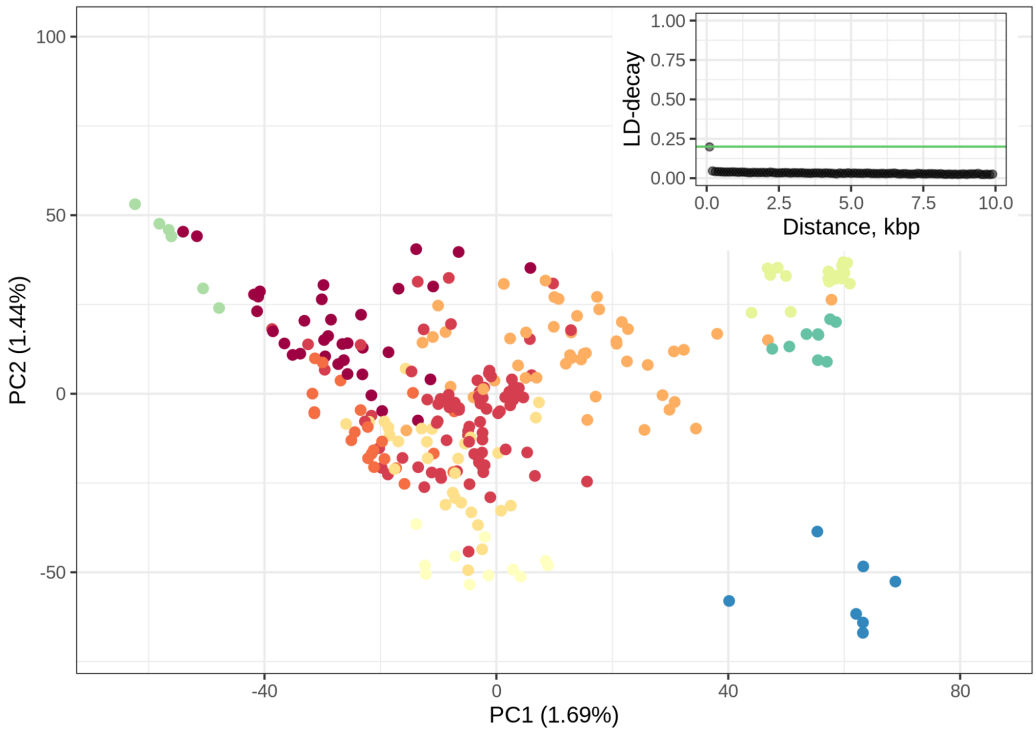


Figure 1: PCA of the 264 perennial ryegrass accessions based on their additive allele dosage. Colours indicate which genetic cluster the accessions have been assigned, based on hierarchical clustering. Linkage disequilibrium (LD) decay is shown in the upper-right corner, as a function of distance in kilobase pairs (kbp). Threshold shown at LD, $r^2 = 0.20$.

3.2 Observed differences among accessions depend on additive and dominance genetic effects and their interactions with climate

Estimated variance components indicate that the additive genetic components contributed most to the phenotypic variance, regardless of trait or modelling approach, RN or ET (Figure 2). The additive genetic component amounted to 37-41 % of the variance for TDM3, 41-43 % for SpringCover and 28-33 % for WinterKill (Figure 2). For TDM3, the interaction between the dominance genetic effect and environment (DxEnv) constituted about 17% of phenotypic variance, while the interaction between the additive genetic effect and environment (AxEnv) constituted about 15-18%. Therefore, the two forms of GxE had approximately equal contributions to the phenotypic variance in TDM3. This indicates that heterozygosity, and in turn genetic variation within accession, may be important for environment-specific interactions and local adaptation.

For SpringCover and WinterKill environment-specific interactions are better captured by the interaction between the additive genetic component and the environment (AxEnv), with this component accounting for 25-31% of the variance in SpringCover and 22-26% of the variance in WinterKill, whereas the interaction between the dominance component and the environment (DxEnv) contributed with 0.4-0.9% of the variance in SpringCover and 11-14% in WinterKill. The dominance genetic component contributed minimally to the phenotypic performance in SpringCover and WinterKill; less than 0.1%, while compared to 14% for TDM3. The environmental component contributed minimally to the phenotypic variance for all traits, as accession means were centered and scaled in each environment (based on means and standard deviations in the core collection of 108 accessions), to avoid confounding by management effects. Lastly, heritability estimates did not differ markedly between the RN or ET modelling approach for any of the traits, with all traits showing moderate narrow-sense heritabilities (Table 2).

Table 2: Narrow-sense heritability (h^2) for all traits, as estimated by model M3 (A + Env + D + AxEnv + DxEnv) for the Reaction-norm (RN) and Envirotyping (ET) modelling approaches.

	TDM3	SpringCover	WinterKill
RN	0.37	0.43	0.28
ET	0.41	0.41	0.32

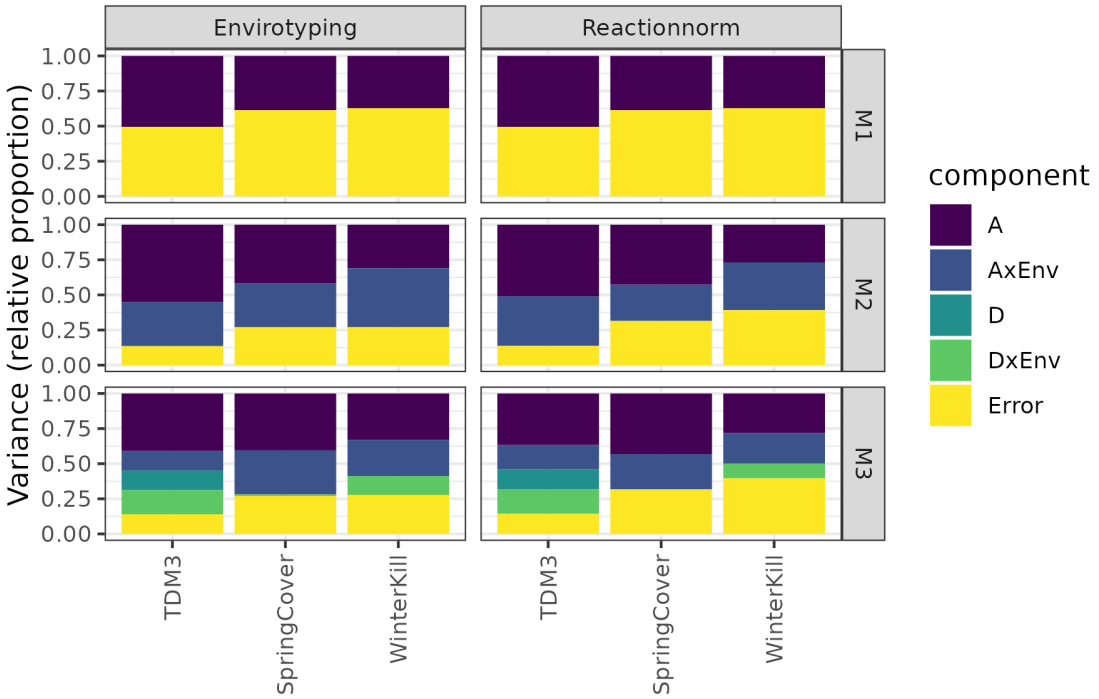


Figure 2: Relative proportion of variance explained by the different model components, excluding the environmental component. M1: main effects of genotype and environment; M2: main effects of genotype and their interactions with environment; M3: main effects of additive and dominance genetic effects and their interactions with the environment. A: additive genetic variance component, AxEnv: variance component for the interaction between additive genetic effect and environment, D: dominance genetic variance component, DxEnv: variance component for the interaction between the dominance genetic effect and the environment, where error is the variance component for the residuals.

We compared the fits of the nested models M1, M2 and M3 (Table 3). Results show that the addition of an interaction term between the additive genetic component and the environment significantly (from M1 to M2) improves model fit for all traits. The addition of a dominance term, which accounts for the assumed level of heterozygosity on phenotypic performance, and an interaction term between the dominance effect and the environment, led to significant model improvement for TDM3, but not for SpringCover and WinterKill. These results suggest the presence of QTL with environment-specific effects which may confer adaptation to specific climatic conditions. Additionally, results indicate that environment-specific marker effects on TDM3 may be mediated by dominance or genetic diversity. In contrast, for SpringCover and WinterKill, results indicate that environment-specific marker effects are mainly mediated by additive genetic effects. Independent of the trait, we find that the RN and ET models do not differ in their ability to capture GxE interactions.

Table 3: Log-likelihood ratio (LLR) statistics and P-values (Pval) for model comparisons, from likelihood-ratio tests. M1: main effects of genotype and environment; M2: main effects of genotype and their interactions with environment; M3: main effects of additive and dominance genetic effects and their interactions with the environment.

		TDM3	SpringCover	WinterKill
RN	**M1-M2**	$LLR = 436.4, Pval = 9.89 \cdot 10^{-95}$	$LLR = 181.9, Pval = 1.88 \cdot 10^{-41}$	$LLR = 8.6, Pval = 3.34 \cdot 10^{-3}$
	M2-M3	$LLR = 22.9, Pval = 1.02 \cdot 10^{-5}$	$LLR = -0.06, \quad Pval = 1$	$LLR = 1.0, Pval = 0.599$
ET	**M1-M2**	$LLR = 474.5, Pval = 3.30 \cdot 10^{-105}$	$LLR = 198.7, Pval = 3.68 \cdot 10^{-45}$	$LLR = 16.8, Pval = 3.34 \cdot 10^{-3}$
	M2-M3	$LLR = 24.3, Pval = 5.24 \cdot 10^{-6}$	$LLR = -0.07, \quad Pval = 1$	$LLR = 1.6, Pval = 0.445$

3.3. Accounting for interactions between additive effects and climate variables improves genomic prediction accuracy

Variance component estimates and model comparison with likelihood ratio tests indicate that including GxE interactions might improve model performance, i.e., prediction accuracy. To determine if GxE interactions improves evaluation of accessions in diverse samples and environments, the prediction accuracy of our models is evaluated under three validation schemes: i) unobserved environments (leave-one-country-out validation), ii) unobserved germplasm, grouped based on relatedness (leave-one-genetic-cluster-out validation), and iii) unobserved germplasm, randomly grouped (leave-one-random-cluster-out validation). Prediction accuracy (PA) in unobserved environments (leave-one-country-out validation scheme), increased with the inclusion of GxE interactions for TDM3, but not for SpringCover and WinterKill. PA for TDM3 increased by up to 4% when accounting for additive GxE, whereas for SpringCover PA decreased by 2-4% depending on the modelling approach, while for WinterKill inclusion of additional model terms led to an 8% decrease in PA with the ET approach and a 24% decrease with RN (Table 4). The reduction in PA with the inclusion of GxE terms, as observed for SpringCover and WinterKill, may suggest the presence of QTL with highly environment-specific effects which would limit extrapolation of marker effects to novel environments.

Results of leave-one-genetic-cluster-out validation show that for TDM3, model M1 cannot accurately extrapolate to novel germplasm, with PA effectively around zero (Table 4). However, addition of an interaction term between the additive genetic effect and environment (AxEnv) improves PA for yield by up to 11%. Further extension of the model by inclusion of additional terms (D and DxEnv) further increases PA by up to 6%. For SpringCover and WinterKill, we observe similar improvements in PA, which increased by up to 21-22% and 7-9%, respectively, when accounting for AxEnv. These results support the presence of QTL with environment-specific effects which covary with the environmental covariates used to characterize the environments. These moderate improvements in PA further indicate that despite the accessions being

genetically diverse, the models are capable of extrapolating GxE interactions to genetically distinct clusters of accessions. For the prediction of unobserved germplasm, randomly assigned into clusters (leave-one-random-cluster-out validation scheme), the PA of the M1 model is much improved for all traits compared to the PAs observed in leave-one-genetic-cluster-out validation. However, limit improvement in PA is observed when extending the M1 with GxE terms under this validation scheme compared to the leave-one-genetic-cluster out scheme, with PA increasing by up to 2% (TDM3), 5-6% (SpringCover) and 2-3% (WinterKill). The large improvement in PA for the leave-one-genetic-cluster-out validation compared to the leave-one-random-cluster-out validation may suggest that the additive genetic effects alone may be less useful when predicting in novel and genetic dissimilar germplasm, and environment-specific marker effects may provide useful information in these challenging situations.

Our results show that inclusion of GxE may lead to improved PA in one validation scheme but confer no improvement or potentially reduce PA in another (Figure 3), e.g., for SpringCover and WinterKill where we observed improved PA in novel germplasm but lower PA in novel environments. Neither of the two GxE modelling approaches (RN and ET) consistently outperformed the other under the three validation schemes. For lack of clear evidence of either modelling approach being superior, we will focus the next analyses on the more widely used RN modelling approach.

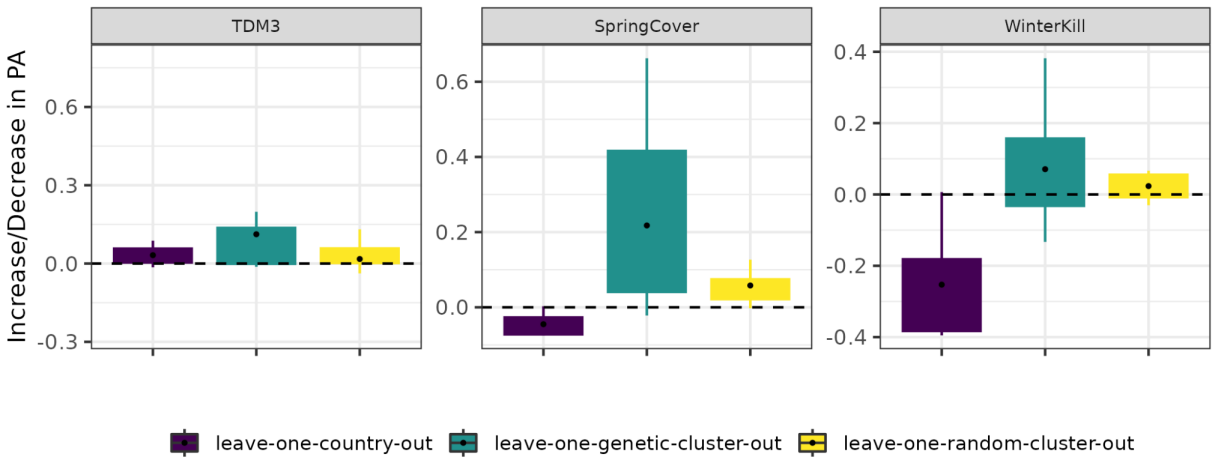


Figure 3: Differences in prediction accuracy (PA) from M1 to M3 for each validation scheme, specifically for the reaction-norm approach. The average correlation across validation sets is shown with black dot.

Table 4: Mean prediction accuracy and standard deviation across validation sets, in the three validation schemes (Leave-one-country-out, Leave-one-cluster-out, and Leave-one-random-cluster-out) for the two modelling approaches: Reaction-norm (RN) and Envirotyping (ET). M1: main effects of genotype and environment; M2: main effects of genotype and their interactions with environment; M3: main effects of additive and dominance genetic effects and their interactions with the environment.

		TDM3	SpringCover	WinterKill
Leave-one-country-out				
	M1	0.57±0.12	0.52±0.08	0.53±0.11
RN	**M2**	0.61±0.12	0.48±0.11	0.27±0.13
	M3	0.61±0.12	0.48±0.11	0.29±0.13
	M1	0.57±0.12	0.52±0.08	0.53±0.11
ET	**M2**	0.59±0.11	0.50±0.10	0.42±0.01
	M3	0.59±0.12	0.50±0.10	0.45±0.05
Leave-one-genetic-cluster-out				
	M1	0.03±0.27	0.15±0.24	0.16±0.20
RN	**M2**	0.14±0.27	0.37±0.17	0.23±0.19
	M3	0.20±0.19	0.36±0.17	0.25±0.14
	M1	0.03±0.27	0.16±0.23	0.16±0.20
ET	**M2**	0.14±0.27	0.37±0.17	0.25±0.21
	M3	0.20±0.16	0.36±0.17	0.29±0.15
Leave-one-random-cluster-out				
	M1	0.45±0.16	0.43±0.10	0.44±0.22
RN	**M2**	0.47±0.15	0.48±0.07	0.46±0.20
	M3	0.46±0.15	0.48±0.07	0.46±0.20
	M1	0.45±0.16	0.43±0.10	0.44±0.22
ET	**M2**	0.47±0.15	0.49±0.07	0.46±0.19
	M3	0.46±0.15	0.48±0.07	0.47±0.19

1

3.4 Genomic prediction accuracy depends on countries

Environments included in this study are diverse and each uniquely challenging for the crop, which can increase the difficulty of capturing any GxE interaction. We compare prediction accuracy for each country to determine whether certain environments constitute a larger challenge for our models (Figure 4). We note that PA for TDM3 and SpringCover is relatively stable across environments, while PA for WinterKill varies greatly across countries. These results indicate that WinterKill may be a trait which is particularly challenging to predict in novel environments. However, it is uncertain whether this is due to sparse sampling of environments (i.e., too few environments in a large area), highly environment-specific effect marker effects, and/or selected environmental variables not correlating well with phenotypic performance in specific environments.

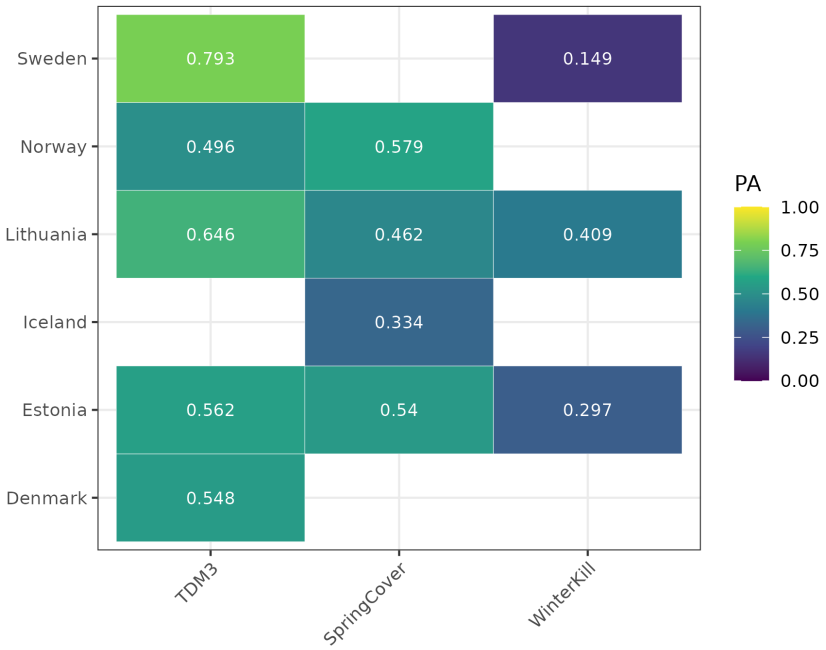


Figure 4: Prediction accuracy (PA) for each country in the leave-one-country-out validation scheme, where the GxE effects are modelled with the reaction-norm (RN) approach. PA is shown for the M3 model which includes the main effect of additive and dominance genetic effects and their interaction with the environment.

3.5 Genomic prediction models predict different components of phenotypic variability with varying accuracy

Validations show improved PA when accounting for GxE for some traits and under certain validation schemes, while for others there is a cost to including GxE (Figure 3). Here we investigate which model components constitute a greater challenge for our GP models. We compare the correlation between accession means y_{ij} (for any accession *i* in environment *j*), and the estimated effects for each model component i.e., additive genetic effects ($\hat{a}_i$), dominance genetic effects ($\hat{d}_i$), interactions between additive genetic effects and environment ($\widehat{ae}_{ij}$), and dominance genetic effects and environment ($\widehat{de}_{ij}$). Correlations between component effects and accession means are calculated in validation sets. This analysis highlights the components which are more challenging for our models to extrapolate and interpolate (Figure 5).

The genetic effects ($\hat{a}_i$ and $\hat{d}_i$) are generally well predicted for the leave-one-country-out validation scheme, with $cor\left(y_{ij}, \widehat{a_{ij}}\right)$ (averaged over validation sets) ranging from 0.49 to 0.59, while $cor\left(y_{ij}, \widehat{d_{ij}}\right)$ ranged from 0.39 to 0.56. When interpolating to novel germplasm, where random accessions are assigned to the validation set, the genetic effects are still reasonably well predicted, with $cor\left(y_{ij}, \widehat{a_{ij}}\right)$ ranging from 0.41 to 0.43, while $cor\left(y_{ij}, \widehat{d_{ij}}\right)$ ranged from 0.19 to 0.37. However, in the leave-one-genetic-cluster validation scheme extrapolation of genetic effects fails, as $cor\left(y_{ij}, \widehat{a_{ij}}\right)$ ranged from 0.07 to 0.16, and $cor\left(y_{ij}, \widehat{d_{ij}}\right)$ ranged from -0.05 to 0.14. The leave-one-genetic-cluster validation scheme is furthermore characterized by large standard deviations indicating that for certain genetic clusters the genetic effects may be well predicted, whereas for other clusters $cor\left(y_{ij}, \widehat{a_{ij}}\right)$ and $cor\left(y_{ij}, \widehat{d_{ij}}\right)$ are either effectively zero or negatively correlated with phenotypic performance (Figure 5). Correlations between accession means and estimated GxE interactions show that coefficients $\widehat{ae}_{ij}$ and $\widehat{de}_{ij}$ are generally more challenging to predict under the leave-one-country-out validation scheme, as suggested by Figure 4, which show a consistently larger cost to accounting for GxE terms in novel environments, as shown by a decrease in PA from the M1 to M3 model. Together these results underlines that the lower PA for the leave-one-genetic-cluster-out validation scheme, specifically with M1 model, compared to the other validation schemes is due to the models not being able to extrapolate genetic effects ($\hat{a}_i$ and $\hat{d}_i$) to novel, and genetically distinct, germplasm. Our results further indicate that under the circumstances where the main genetic effects are not well predicted then there is a greater benefit to accounting for GxE, as shown by larger increases in PA under the leave-one-genetic-cluster-out scheme compared to the leave-one-random-cluster-out where main genetic effects are more useful to predict accession means.

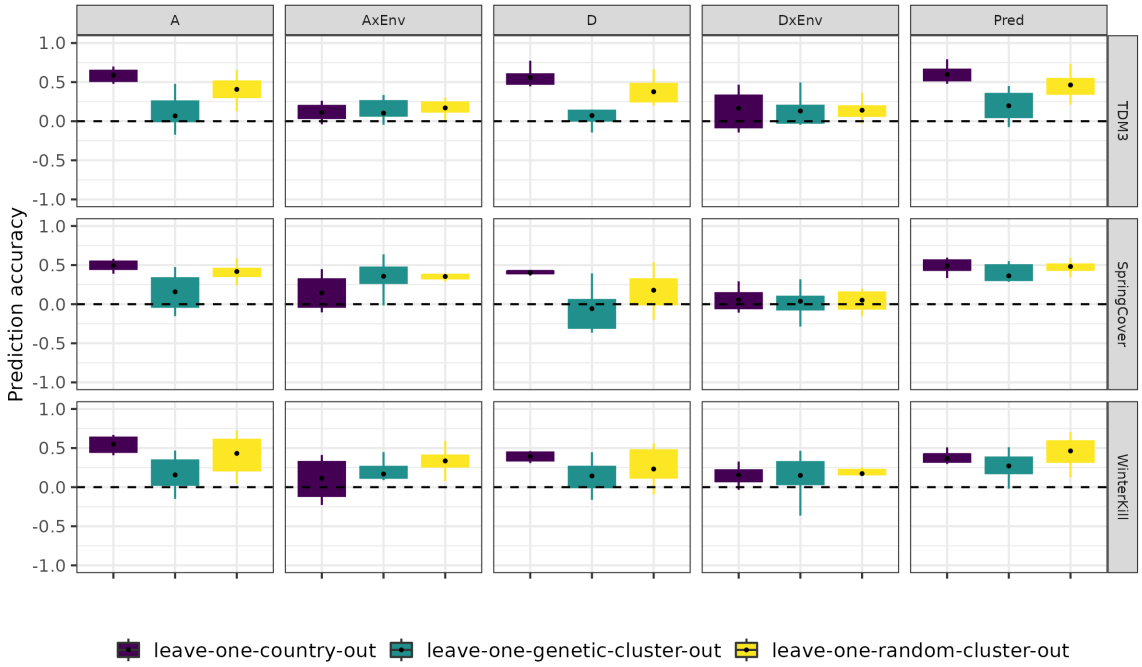


Figure 5: Correlations in left-out folds for the three validation schemes (Leave-one-country-out scheme, Leave-one-genetic-cluster-out and Leave-one-random-cluster-out) from model M3 fitted with the Reaction-norm approach. The following correlations are shown: Pred., Observed accession means vs. predicted accession means; A, Observed accession means vs. Additive breeding values coefficients; D, Observed accession means vs. Dominance breeding values coefficients; Env, Observed accession means vs. Environment coefficients; AxEnv, Observed accession means vs. AxEnv coefficients; DxEnv, Observed accession means vs. DxEnv coefficients.

3.6 Genome-wide association studies detected marginal and country-dependent marker effects

We conducted GWAS to detect genetic variants with marginal effects (average associations across environments) and conditional effects (country-specific association) (Figure 6).

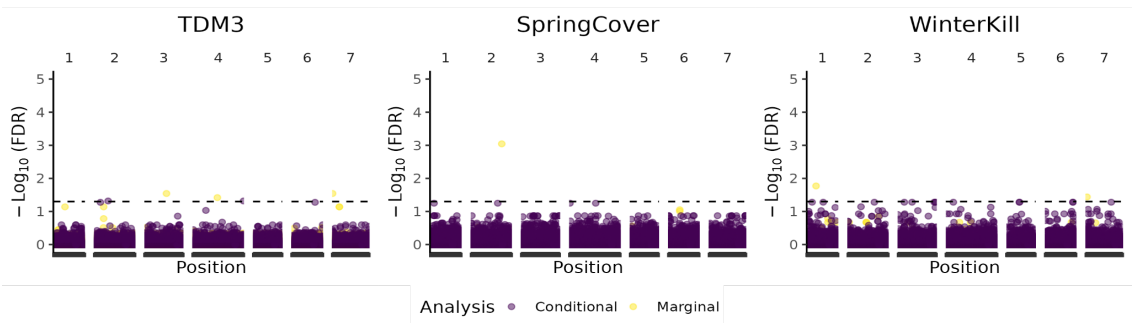


Figure 6: GWAS analysis showing results for a marginal GWAS analysis (environment-independent QTL) and conditional GWAS analysis (environment-specific QTL effect). Significance threshold adjusted with FDR (False-discovery rate).

Previous analyses indicated extensive additive genetic effects (Figure 2 and Table 3) and moderate narrow-sense heritabilities (Table 2) for all traits, suggesting that a large proportion of the phenotypic variation may be described by additive genetic variation. However, few significant QTL were detected by the marginal GWAS. These results suggest that the investigated traits are polygenic, and are characterized by many small-effect QTL. Genome-wide analyses of GxE, i.e., likelihood-ratio tests (Table 3) and GP validation (Table 4), indicated the presence of environment-specific marker effects. Few conditional effects were highly significant (FDR ≤ 0.05) for TDM3. For SpringCover, there was some evidence for potential QTL (FDR ≤ 0.10) conferring a disadvantage in certain environments (e.g., in Lithuania), but not in others (e.g., Denmark and Norway) (Figure 7). Similarly, for WinterKill we detected potential QTL which may confer a disadvantage in Lithuania but not in Sweden (Figure 7).

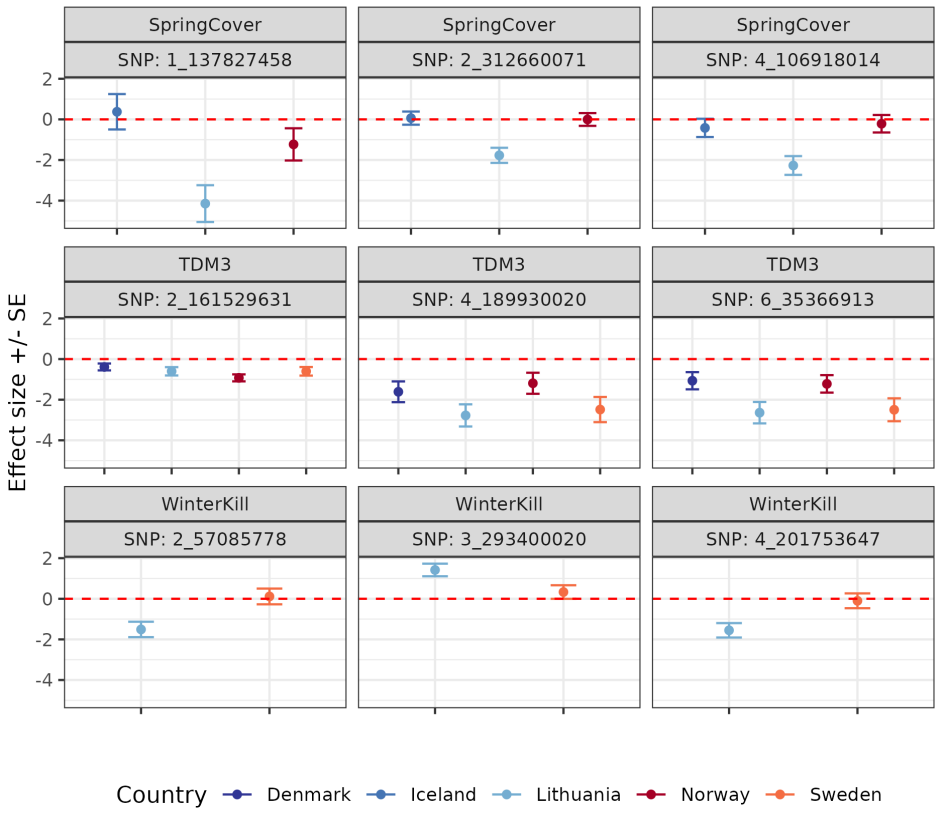


Figure 7: Effect estimates ± standard errors (SE) for the SNPs with the lowest P-value from the conditional GWAS.

For both the marginal and conditional GWAS we found few significant QTL, which may be due in part to the germplasm being genetically dissimilar as demonstrated by limited LD (Figure 1). Hence, only common and large-effect QTL would be detected by our models. Alternatively, it is possible that the underlying genetic architecture is highly polygenic, both in regard to the additive genetic effects and the additive GxE effects, and implying that most QTL have small effects.

4. Discussion

4.1 Diverse accessions contain unique genetic variation important for broadening the genetic base of perennial ryegrass

This study included a geographically and genetically diverse germplasm, to enhance the probability of including and identifying genetic variation that may confer an advantage under the environmental conditions experienced in the Nordic and Baltic countries. Utilising genetically diverse accessions is favourable in prebreeding, as less advanced breeding material, or potentially unselected material, may be enriched for higher genetic diversity and may contain novel adaptive variants not present in current breeding programs, which may be introgressed into more advanced breeding material for enhanced phenotypic performance under biotic and abiotic stressors as demonstrated in previous studies (Singh et al. 2021; Meseka et al. 2013). The present study sought to quantify and estimate the extent of GxE interactions in environments that are at the limit of the species' distribution. GxE interactions are environment specific, hence while previous studies have investigated GxE interactions in perennial ryegrass, most of these studies have investigated GxE in temperate regions, which are currently more suitable for its production e.g., in central and southern Europe (Fois et al. 2021), Denmark (Fè et al. 2015, 2016) and Ireland (Grogan and Gilliland 2011). Consequently, the findings of these studies are presumably not indicative of the level of GxE interactions that are present, or occur, in the more extreme environments experienced in the northern regions, investigated in the present study. GxE interactions in environments comparable to those investigated in this study have previously been documented in perennial ryegrass, by (Helgadóttir et al. 2018) who found that commercial cultivars are not widely adapted to the environmental conditions across the Nordic and Baltic regions, which underlines the importance of including unselected material in future breeding efforts.

Additionally, the accessions included in the study have been collected from vastly different geographic origins, which is a benefit as local adaptation to some of these environments may confer an advantage for the environments tested in the current study. Hence, if instead we had included genetically more similar accessions, we would have had to constrain us to only include highly related individuals (from similar geographic backgrounds), which would have been a limitation.

4.2 Genotype x environment interactions pose challenges in diverse samples of environments and accessions

In the current study we compared a standard reaction-norm approach, as demonstrated by (Jarquín et al. 2014) to a more recent envirotyping approach as demonstrated by (Xu 2016) and later extended by (Costa-Neto, Fritsche-Neto, and Crossa 2021; Costa-Neto, Crossa, and Fritsche-Neto 2021). In accordance with the findings of previous studies we found phenotypic variation to be highly affected by GxE interactions for both ordinal traits, i.e., SpringCover and WinterKill, as well as yield traits. The inclusion of interactions between additive genetic effects and environments (AxEnv) led to moderate improvements in model fit and prediction

accuracy for all traits. Inclusion of dominance and dominance GxE terms further improved prediction and model fit. Previous studies, in hybrid maize, have similarly documented the importance of dominance genetic effects in regard to yield production, with (Rogers et al. 2021) demonstrating that the dominance genetic effect was a more effective predictor of yield performance than the additive genetic effect. However, when model performance was evaluated based on their ability to predict the phenotypic performance of novel germplasm, where training and validation sets were genetically more similar (leave-one-random-cluster-out) or when predicting the performance observed germplasm but in novel environments (leave-one-country-out), there were more limited improvements in prediction accuracy with the inclusion of GxE terms. Previous GxE studies (e.g., in wheat) experienced more marked improvements in prediction accuracy when accounting for GxE, with some studies citing increases in prediction accuracy by up to 82% when predicting yield performance of unobserved lines, modelled with a reaction-norm approach (Jarquín et al. 2017), while (Heslot et al. 2014) reported an increase in mean prediction accuracy of approximately 25-28% for yield production in winter wheat. It should be noted that the study by (Heslot et al. 2014) utilized crop growth models that incorporate developmental stage specific environmental covariates (stress covariates), and hence the environmental covariates included in that study was presumably more physiologically relevant than the environmental covariates included in this study. Generally, GxE studies cover less heterogeneous areas than those investigated in this study, e.g., Northern France (Jarquín et al. 2014) and France (Heslot et al. 2014). Additionally, these studies also investigated a significantly denser spatial and temporal grid, with the study by (Heslot et al. 2014) including 44 environments (location-year combinations) from 2006 to 2011, while the study by Jarquin et al. (2014) investigated 134 locations across eight years. The importance of employing a dense sampling approach is further underlined by (Rogers and Holland 2022), where they noted that more sparsely sampled areas suffered higher reductions in prediction accuracy as these environments were less represented by the environments included in the training set. This is potentially also why we observed reduced prediction accuracy for yield in Iceland, which is environmentally most dissimilar to the other environments (Figure 4). Therefore, more effective extrapolation of GxE might be achieved with denser environmental grids, even with large differences among environments which limit the transferability of marker effects to novel environments.

4.3 Genetic diversity and rapid LD decay in perennial ryegrass limit the accuracy of genomic prediction

The improvements in prediction accuracy, observed for certain validation schemes, are more modest than those demonstrated in previously mentioned studies. This is to be expected as the accessions are genetically diverse, and include unselected material where genomic relationships between accessions are limited. This is further compounded, as we inferred genetic clusters based on genetic similarity, resulting in increased genetic dissimilarity between the accessions in the training and validation sets. Consequently, the lack of

population structure and low levels of genetic similarity between accessions in the training set and the validation set might have limited our ability to efficiently extrapolate both additive genetic effects and GxE interactions to unobserved accessions. This is supported by the findings of (Pembleton et al. 2018), which demonstrated that prediction accuracies were negatively affected by the inclusion of genetically distinct subgroups of perennial ryegrass, and that prediction accuracies could be enhanced by removing genetic material that is genetically distinct from the majority of the germplasm contained in the training sets. Consequently, while low genetic similarity, in general, is a challenge across species, it is possible that there is a particularly large cost to including genetically dissimilar material when performing genomic prediction in perennial ryegrass, a species characterized by rapid LD decay of less than 1 kb (Hayes et al. 2013) or less than 3 kb (Ponting et al. 2007). In our study we found rapid declines in linkage disequilibrium (LD), with LD decay within 200 bp (Figure 1), hence we underline the importance of including genetically similar germplasm for genomic selection in perennial ryegrass. However, focusing on advanced and highly related individuals would have limited the probability of identifying novel adaptive genetic variation which may be introgressed into advanced breeding material and in turn may enrich the genetic diversity in perennial ryegrass breeding populations.

4.4 Applications for Breeders: perennial ryegrass and its future in the northern regions

For breeders to be able to produce perennial ryegrass in more northern regions it is of importance to identify genetic material that can survive the cold and variable climates experienced in these regions. In the present study, we were able to identify genetic variants in linkage with QTL that might be associated with increased winter survival. These variants appear to have environment-independent effects across the environments where WinterKill was scored. The presence of extensive GxE (Figure 2, Table 3 and Figure 6) suggests that it may be advantageous to breed for local adaptation to specific environmental conditions or countries, as introgression of QTL with environment-specific effects may be advantageous in certain conditions while deleterious in others. Furthermore, depending on the stochasticity of the environmental stressors, at the intended production area, considerations regarding whether to select for increased GxE interactions should be made. As environments with reliable environmental cues, e.g. seasonal changes in day-length or temperature reliably signalling future selective environments, there is an increased chance of GxE interactions improving population persistence under stressing conditions, whereas if the environment is highly stochastic and environmental cues for acclimatisation are unreliable such interactions may be deleterious for survival (Reed et al. 2010). Hence, if the environments at the intended production area are characterised by variable and increasingly unpredictable weather patterns it may be more beneficial to select for accessions with smaller GxE interactions.

4.5 Conclusions

Our study highlighted extensive GxE in perennial ryegrass for the environments experienced in the Nordic and Baltic regions. We found no clear evidence for different modelling approaches surpassing the other (either reaction norm based on linear response to environmental variables, or envirotyping, based on non-linear responses). The environments included in this study were highly diverse, and we observed that the statistical models were challenged in modelling the environmental effect on phenotypic performance for the majority of the traits, presumably because of sparsity in the spatial grid. While there was some benefit in accounting for GxE when predicting the performance of unobserved and genetically dissimilar germplasm, these moderate improvements in prediction were not observed for validation schemes where the additive genetic effects were most useful to predict phenotypic performance (leave-one-random-cluster out validation). Despite rapid LD decay in our diverse panel of accessions, we were able to detect several candidate QTL associated with multiple traits, including QTL with country-specific effects. Together, our results suggest that there is adaptive genetic variation in the perennial ryegrass diversity panel that could be advantageous to introduce into current breeding programs. Further work investigating the extent of conditional neutrality relative to maladaptive pleiotropy in respect to local adaptation may be useful for future breeding efforts in perennial ryegrass, as this may inform breeders about the costs and benefits in selecting for increased local adaptation compared to broad adaptation.

Acknowledgements

This study was carried out under the auspices of the 'Public Private Partnership for pre-breeding in perennial ryegrass' project (urn:nbn:se:norden:org:diva-12149), funded by the Nordic Council of Ministers, and the 'Development of Molecular Markers for Genomic Selection of Adaptation in Perennial Ryegrass' project funded by the Research Council of Lithuania (Grant No. MIP-64/2015). NordGen is acknowledged for their help obtaining the perennial ryegrass accessions for the project. NHJ was supported by a grant from the Aarhus University Research Foundation (Grant No. AUFF-F-2021-7-6).

References

Ågren, Jon, Christopher G. Oakley, John K. McKay, John T. Lovell, and Douglas W. Schemske. 2013. "Genetic Mapping of Adaptation Reveals Fitness Tradeoffs in Arabidopsis Thaliana." *Proceedings of the National Academy of Sciences* 110 (52): 21077–82.

Anderson, Jill T., Cheng-ruei Lee, Catherine A. Rushworth, Robert I. Colautti, and Thomas Mitchell-olds. 2013. "Genetic Trade-offs and Conditional Neutrality Contribute to Local Adaptation." *Molecular Ecology* 22 (3): 699–708.

Atwell, Susanna, Yu S. Huang, Bjarni J. Vilhjálmsson, Glenda Willems, Matthew Horton, Yan Li, Dazhe Meng, et al. 2010. "Genome-Wide Association Study of 107 Phenotypes in Arabidopsis Thaliana Inbred Lines." *Nature* 465 (7298): 627–31.

Benjamini, Yoav, and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society Series B: Statistical Methodology* 57 (1): 289–300.

Brazauskas, Gintaras, Ingo Lenk, Morten Greve Pedersen, Bruno Studer, and Thomas Lübberstedt. 2011. "Genetic Variation, Population Structure, and Linkage Disequilibrium in European Elite Germplasm of Perennial Ryegrass." *Plant Science* 181 (4): 412–20.

Butler, D. G., B. R. Cullis, A. R. Gilmour, B. J. Gogel, and R. Thompson. n.d. "ASReml Estimates Variance Components under a General Linear Mixed Model by Residual Maximum Likelihood (REML)."

Byrne, Stephen, Adrian Czaban, Bruno Studer, Frank Panitz, Christian Bendixen, and Torben Asp. 2013. "Genome Wide Allele Frequency Fingerprints (GWAFFs) of Populations via Genotyping by Sequencing." *PloS One* 8 (3): e57438.

Costa-Neto, Germano, Jose Crossa, and Roberto Fritsche-Neto. 2021. "Enviromic Assembly Increases Accuracy and Reduces Costs of the Genomic Prediction for Yield Plasticity in Maize." *Frontiers in Plant Science* 12 (October):717552.

Costa-Neto, Germano, Roberto Fritsche-Neto, and José Crossa. 2021. "Nonlinear Kernels, Dominance, and Envirotyping Data Increase the Accuracy of Genome-Based Prediction in Multi-Environment Trials." *Heredity* 126 (1): 92–106.

Costa-Neto, Germano, Giovanni Galli, Humberto Fanelli Carvalho, José Crossa, and Roberto Fritsche-Neto. 2021. "EnvRtype: A Software to Interplay Enviromics and Quantitative Genomics in Agriculture." *G3* 11 (4). https://doi.org/10.1093/g3journal/jkab040.

Danecek, Petr, Adam Auton, Goncalo Abecasis, Cornelis A. Albers, Eric Banks, Mark A. DePristo, Robert E. Handsaker, et al. 2011. "The Variant Call Format and VCFtools." *Bioinformatics* 27 (15): 2156–58.

Elshire, Robert J., Jeffrey C. Glaubitz, Qi Sun, Jesse A. Poland, Ken Kawamoto, Edward S. Buckler, and

Sharon E. Mitchell. 2011. “A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species.” *PloS One* 6 (5): e19379.

El-Soda, Mohamed, Marcos Malosetti, Bas J. Zwaan, Maarten Koornneef, and Mark G. M. Aarts. 2014. “Genotype×environment Interaction QTL Mapping in Plants: Lessons from Arabidopsis.” *Trends in Plant Science* 19 (6): 390–98.

Ergon, Å., G. Seddaiu, P. Korhonen, P. Virkajärvi, G. Bellocchi, M. Jørgensen, L. Østrem, D. Reheul, and F. Volaire. 2018. “How Can Forage Production in Nordic and Mediterranean Europe Adapt to the Challenges and Opportunities Arising from Climate Change?” *European Journal of Agronomy: The Journal of the European Society for Agronomy* 92 (January):97–106.

Fè, Dario, Bilal H. Ashraf, Morten G. Pedersen, Luc Janss, Stephen Byrne, Niels Roulund, Ingo Lenk, et al. 2016. “Accuracy of Genomic Prediction in a Commercial Perennial Ryegrass Breeding Program.” *The Plant Genome* 9 (3). https://doi.org/10.3835/plantgenome2015.11.0110.

Fè, Dario, Morten Greve Pedersen, Christian S. Jensen, and Just Jensen. 2015. “Genetic and Environmental Variation in a Commercial Breeding Program of Perennial Ryegrass.” *Crop Science* 55 (2): 631–40.

Flint-Garcia, Sherry A., Jeffry M. Thornsberry, and Edward S. Buckler 4th. 2003. “Structure of Linkage Disequilibrium in Plants.” *Annual Review of Plant Biology* 54 (1): 357–74.

Fois, Mattia, Marta Malinowska, Franz Xaver Schubiger, and Torben Asp. 2021. “Genomic Prediction and Genotype-by-Environment Interaction Analysis of Crown and Stem Rust in Ryegrasses in European Multi-Site Trials.” *Agronomy* 11 (1119). https://doi.org/10.3390/agronomy11061119.

Galili, Tal. 2015. “Dendextend: An R Package for Visualizing, Adjusting and Comparing Trees of Hierarchical Clustering.” *Bioinformatics (Oxford, England)* 31 (22): 3718–20.

Grogan, D., and T. J. Gilliland. 2011. “A Review of Perennial Ryegrass Variety Evaluation in Ireland.” *Irish Journal of Agricultural and Food Research* 50 (1): 65–81.

Hayes, Benjamin J., Noel O. I. Cogan, Luke W. Pembleton, Michael E. Goddard, Junping Wang, German C. Spangenberg, and John W. Forster. 2013. “Prospects for Genomic Selection in Forage Plant Species.” *Plant Breeding* 132 (2): 133–43.

Helgadóttir, Á., R. Aavola, M. Isolahti, P. Marum, C. Persson, A. Aleliūnas, G. Brazauskas, T. A. Krisjánsdóttir, T. Asp, and O. A. Rognli. 2018. “Adaptability and Phenotypic Stability of Lolium Perenne L. Cultivars of Diverse Origin Grown at the Margin of the Species Distribution.” *Journal of Agronomy and Crop Science* 204 (5): 493–504.

Heslot, Nicolas, Deniz Akdemir, Mark E. Sorrells, and Jean-Luc Jannink. 2014. “Integrating Environmental Covariates and Crop Modeling into the Genomic Selection Framework to Predict Genotype by Environment Interactions.” *TAG. Theoretical and Applied Genetics* 127 (2): 463–80.

Jarquín, Diego, José Crossa, Xavier Lacaze, Philippe Du Cheyron, Joëlle Daucourt, Josiane Lorgeou, François Piraux, et al. 2014. "A Reaction Norm Model for Genomic Selection Using High-Dimensional Genomic and Environmental Data." *TAG. Theoretical and Applied Genetics* 127 (3): 595–607.

Jarquín, Diego, Cristiano Lemes da Silva, R. Chris Gaynor, Jesse Poland, Allan Fritz, Reka Howard, Sarah Battenfield, and Jose Crossa. 2017. "Increasing Genomic-enabled Prediction Accuracy by Modeling Genotype × Environment Interactions in Kansas Wheat." *The Plant Genome* 10 (2): lantgenome2016.12.0130.

Jarquin, Diego, Natalia de Leon, Cinta Romay, Martin Bohn, Edward S. Buckler, Ignacio Ciampitti, Jode Edwards, et al. 2020. "Utility of Climatic Information via Combining Ability Models to Improve Genomic Prediction for Yield Within the Genomes to Fields Maize Project." *Frontiers in Genetics* 11:592769.

Kahiluoto, Helena, Janne Kaseva, Jan Balek, Jørgen E. Olesen, Margarita Ruiz-Ramos, Anne Gobin, Kurt Christian Kersebaum, et al. 2019. "Decline in Climate Resilience of European Wheat." *Proceedings of the National Academy of Sciences of the United States of America* 116 (1): 123–28.

Laporte, Fabien, Alain Charcosset, and Tristan Mary-Huard. 2022. "Efficient ReML Inference in Variance Component Mixed Models Using a Min-Max Algorithm." *PLoS Computational Biology* 18 (1): e1009659.

Lasky, Jesse R., Brenna R. Forester, and Matthew Reimherr. 2018. "Coherent Synthesis of Genomic Associations with Phenotypes and Home Environments." *Molecular Ecology Resources* 18 (1): 91–106.

McKenna, Aaron, Matthew Hanna, Eric Banks, Andrey Sivachenko, Kristian Cibulskis, Andrew Kernytsky, Kiran Garimella, et al. 2010. "The Genome Analysis Toolkit: A MapReduce Framework for Analyzing next-Generation DNA Sequencing Data." *Genome Research* 20 (9): 1297–1303.

Meseka, Silvestro, Morakinyo Fakorede, Sam Ajala, Baffour Badu-Apraku, and Abebe Menkir. 2013. "Introgression of Alleles from Maize Landraces to Improve Drought Tolerance in an Adapted Germplasm." *Journal of Crop Improvement* 27 (1): 96–112.

Meuwissen, T. H., B. J. Hayes, and M. E. Goddard. 2001. "Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps." *Genetics* 157 (4): 1819–29.

Millet, Emilie J., Willem Kruijer, Aude Coupel-Ledru, Santiago Alvarez Prado, Llorenç Cabrera-Bosquet, Sébastien Lacube, Alain Charcosset, Claude Welcker, Fred van Eeuwijk, and François Tardieu. 2019. "Genomic Prediction of Maize Yield across European Environmental Conditions." *Nature Genetics* 51 (6): 952–56.

Monks, D., K. Sadatasilan, and D. Moot. 2009. "Cardinal Temperatures and Thermal Time Requirements for Germination of Annual and Perennial Temperate Pasture Species" 39:95–110.

Nagy, Istvan, Elisabeth Veeckman, Chang Liu, Michiel Van Bel, Klaas Vandepoele, Christian Sig Jensen, Tom Ruttink, and Torben Asp. 2022. "Chromosome-Scale Assembly and Annotation of the Perennial Ryegrass Genome." *BMC Genomics* 23 (1): 505.

Pembleton, Luke W., Courtney Inch, Rebecca C. Baillie, Michelle C. Drayton, Preeti Thakur, Yvonne O. Ogaji, German C. Spangenberg, John W. Forster, Hans D. Daetwyler, and Noel O. I. Cogan. 2018. "Exploitation of Data from Breeding Programs Supports Rapid Implementation of Genomic Selection for Key Agronomic Traits in Perennial Ryegrass." *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik* 131 (9): 1891–1902.

Ponting, Rebecca C., Michelle C. Drayton, Noel O. I. Cogan, Mark P. Dobrowolski, Germán C. Spangenberg, Kevin F. Smith, and John W. Forster. 2007. "SNP Discovery, Validation, Haplotype Structure and Linkage Disequilibrium in Full-Length Herbage Nutritive Quality Genes of Perennial Ryegrass (Lolium Perenne L.)." *Molecular Genetics and Genomics: MGG* 278 (5): 585–97.

Raza, Ali, Ali Razzaq, Sundas Saher Mehmood, Xiling Zou, Xuekun Zhang, Yan Lv, and Jinsong Xu. 2019. "Impact of Climate Change on Crops Adaptation and Strategies to Tackle Its Outcome: A Review." *Plants* 8 (2). https://doi.org/10.3390/plants8020034.

Reed, Thomas E., Robin S. Waples, Daniel E. Schindler, Jeffrey J. Hard, and Michael T. Kinnison. 2010. "Phenotypic Plasticity and Population Viability: The Importance of Environmental Predictability." *Proceedings. Biological Sciences* 277 (1699): 3391–3400.

Rogers, Anna R., Jeffrey C. Dunne, Cinta Romay, Martin Bohn, Edward S. Buckler, Ignacio A. Ciampitti, Jode Edwards, et al. 2021. "The Importance of Dominance and Genotype-by-Environment Interactions on Grain Yield Variation in a Large-Scale Public Cooperative Maize Experiment." *G3* 11 (2). https://doi.org/10.1093/g3journal/jkaa050.

Rogers, Anna R., and James B. Holland. 2022. "Environment-Specific Genomic Prediction Ability in Maize Using Environmental Covariates Depends on Environmental Similarity to Training Data." *G3* 12 (2). https://doi.org/10.1093/g3journal/jkab440.

Rognli, O. A., R. Aavola, A. Aleliūnas, T. Asp, G. Brazauskas, K. H. Gylstrøm, A. Helgadottir, et al. 2018. "Utilization of Genebank Accessions to Improve Northern Adaptation of Perennial Ryegrass (Lolium Perenne L.)." In *Breeding Grasses and Protein Crops in the Era of Genomics*, 3–8. Springer International Publishing.

Shelford, V. E. 1931. "Some Concepts of Bioecology." *Ecology* 12 (3): 455–67.

Singh, Sukhwinder, A. Jighly, D. Sehgal, J. Burgueño, R. Joukhadar, S. K. Singh, A. Sharma, et al. 2021. "Direct Introgression of Untapped Diversity into Elite Wheat Lines." *Nature Food* 2 (10): 819–27.

Solberg, E. T., A. Rognli, and L. Østrem. 1994. "Potential for Improving Adaptation of Lolium Perenne L. to Continental Climates in Norway." *Euphytica* 77:183–91.

Vitezica, Zulma G., Andrés Legarra, Miguel A. Toro, and Luis Varona. 2017. "Orthogonal Estimates of Variances for Additive, Dominance, and Epistatic Effects in Populations." *Genetics* 206 (3): 1297–1307.

Xu, Yunbi. 2016. "Envirotyping for Deciphering Environmental Impacts on Crop Plants." *TAG. Theoretical and Applied Genetics.* 129 (4): 653–73.

1

2 **Supplementary**

Supplementary Table 1: List of accessions names, project IDs, geographic origin and accession type (breeding line, landrace, cultivar, or ecotype) for the 265 diploid accessions included in the study.

	Accession name	Accession number	Project id	Country of Origin	Accession type
1	1159	LIA368	40053	Lithuania	Breeding line
2	1159	LIA66	40057	Lithuania	Breeding line
3	1248	LIA365	40058	Lithuania	Breeding line
4	114-3	144-3	40128	Estonia	Breeding line
5	14G2000331	14G2000331	40104	Czech Republic	Ecotype
6	14G2000332	14G2000332	40105	Czech Republic	Ecotype
7	14G2000333	14G2000333	40106	Czech Republic	Ecotype
8	14G2000503	14G2000503	40109	Czech Republic	Ecotype
9	14G2000504	14G2000504	40107	Czech Republic	Ecotype
10	14G2000523	14G2000523	40108	Czech Republic	Ecotype
11	14G2000586	14G2000586	40103	Czech Republic	Ecotype
12	14G2000587	14G2000587	40102	Czech Republic	Ecotype
13	16-57-1	NGB4262	40227	Norway	Ecotype
14	16-57-2	NGB4263	40228	Norway	Ecotype
15	16-59-2	NGB4264	40229	Norway	Ecotype
16	16-62-3	NGB4267	40230	Norway	Ecotype
17	16-62-4	NGB4268	40231	Norway	Ecotype
18	177/06	LVA02520	40090	Latvia	Breeding line
19	264/06	LVA02529	40099	Latvia	Ecotype
20	2903	LIA1414	40071	Lithuania	Ecotype
21	2910	LIA1416	40078	Lithuania	Ecotype
22	342	PI598414	40188	Turkey	Ecotype
23	346	PI598515	40189	Turkey	Ecotype
24	363	PI598516	40190	Turkey	Ecotype
25	45-4	45-4	40127	Estonia	Breeding line
26	481	W6-11256	40200	Turkey	Ecotype
27	512	W6-11264	40201	Turkey	Ecotype
28	513	PI598519	40191	Turkey	Ecotype
29	537	LIA65	40080	Lithuania	Breeding line
30	693	W6-11322	40202	Turkey	Ecotype
31	ABY-BA10106.82	PI619003	40195	Norway	Ecotype
32	ABY-BA10109.82	PI610803	40193	Norway	Ecotype
33	ABY-BA10111.82	PI610802	40192	Norway	Ecotype
34	ABY-BA8602.00	PI598429	40176	Italy	Ecotype
35	ABY-BA8603.00	PI598430	40177	Italy	Ecotype
36	ABY-BA8604.00	PI598431	40178	Italy	Ecotype
37	ABY-BA8614.68	PI598432	40179	Italy	Ecotype
38	ABY-BA8616.00	PI598433	40180	Italy	Ecotype
39	ABY-BA8621.82	PI598434	40181	Italy	Ecotype
40	ABY-BA9080.A81	PI598453	40186	Romania	Ecotype
41	ABY-BA9088.82	W6-9290	40198	Switzerland	Ecotype

42	ABY-BA9091.72	PI598440	40182	Switzerland	Ecotype
43	ABY-BA9092.72	PI598441	40183	Switzerland	Ecotype
44	ABY-BA9097.84	PI598443	40185	Switzerland	Ecotype
45	ABY-BA9101.72	W6-9297	40199	Switzerland	Ecotype
46	ABY-BA9108.79	PI610825	40194	Switzerland	Ecotype
47	ABY-BA9792.81	PI595043	40173	UK	Ecotype
48	ABY-BA9798.82	PI595044	40174	UK	Ecotype
49	ABY-BA9803.80	PI595047	40175	UK	Ecotype
50	ABY-BA9983.81	P1598454	40187	Romania	Ecotype
51	ABY-BA9819	BA9819	40356	UK	Ecotype
52	ABY-Ba9832	Ba9832	40357	UK	Ecotype
53	ALLEGRO	NGB13338	40219	Denmark	Cultivar
54	AMADO	NGB1623	40203	Denmark	Cultivar
55	ANNA	POL155165	40397	Poland	Cultivar
56	APUS	NGB7508	40238	Sweden	Cultivar
57	ARGONA	POL155193	40400	Poland	Cultivar
58	ARKE	POL155194	40401	Poland	Cultivar
59	ASKELAND-16-60-1	NGB4265	40232	Norway	Ecotype
60	Abermagic	Abermagic	40277	UK	Cultivar
61	Algutsrum-HAJ0203	NGB16597	40253	Sweden	Ecotype
62	Aria	Vir50422	40146	Sweden	Cultivar
63	Arka	14G2000008	40115	Poland	Cultivar
64	Arsenal	Arsenal	40285	Germany	Cultivar
65	Banat	DE161	40003	Romania	Cultivar
66	Belida	NGB1633	40204	Denmark	Cultivar
67	Benestad-UE1504	NGB333	40254	Sweden	Ecotype
68	BJörkeröd-PW2702	NGB43449	40255	Sweden	Ecotype
69	Ba11434	Ba11434	40313	Poland	Ecotype
70	Ba11445	Ba11445	40314	Poland	Ecotype
71	Ba11448	Ba11448	40315	Poland	Ecotype
72	Ba11451	Ba11451	40317	Poland	Ecotype
73	Ba11461	Ba11461	40316	Poland	Ecotype
74	Ba12275	Ba12275	40358	Czech Republic	Ecotype
75	Ba12276	Ba12276	40359	Czech Republic	Ecotype
76	Ba12277	Ba12277	40360	Czech Republic	-
77	Ba12947	Ba12947	40319	Hungary	-
78	Ba12948	Ba12948	40320	Hungary	-
79	Ba12949	Ba12949	40321	Hungary	-
80	Ba12950	Ba12950	40322	Hungary	-
81	Ba12951	Ba12951	40323	Hungary	-
82	Ba12952	Ba12952	40324	Hungary	-
83	Ba12953	Ba12953	40325	Hungary	-
84	Ba12954	Ba12954	40326	Hungary	-
85	Ba12955	Ba12955	40327	Hungary	-
86	Ba12956	Ba12956	40328	Hungary	-
87	Ba12957	Ba12957	40329	Hungary	-
88	Ba12958	Ba12958	40330	France	-
89	Ba12965	Ba12965	40331	France	Ecotype

90	Ba12967	Ba12967	40333	France	Ecotype
91	Ba12969	Ba12969	40334	France	Ecotype
92	Ba12970	Ba12970	40335	France	Ecotype
93	Ba12971	Ba12971	40336	France	Ecotype
94	Ba12972	Ba12972	40337	France	Ecotype
95	Ba12973	Ba12973	40388	France	Ecotype
96	Ba12974	Ba12974	40339	France	Ecotype
97	Ba12975	Ba12975	40340	France	Ecotype
98	Ba12976	Ba12976	40342	France	Ecotype
99	Ba12977	Ba12977	40343	France	Ecotype
100	Ba12978	Ba12978	40344	France	Ecotype
101	Ba12979	Ba12979	40345	France	Ecotype
102	Ba12980	Ba12980	40346	France	Ecotype
103	Ba12981	Ba12981	40347	France	Ecotype
104	Ba12982	Ba12982	40348	France	Ecotype
105	Ba12983	Ba12983	40349	France	Ecotype
106	Ba12986	Ba12986	40352	France	Ecotype
107	Ba12987	Ba12987	40353	France	Ecotype
108	Ba12988	Ba12988	40354	France	Ecotype
109	Ba12989	Ba12989	4035	France	Ecotype
110	Ba13004	Ba13004	40366	Germany	Ecotype
111	Ba13005	Ba13005	40367	Germany	Ecotype
112	Ba13006	Ba13006	40368	Germany	Ecotype
113	Ba13007	Ba13007	40369	Germany	Ecotype
114	Ba13011	Ba13011	40373	Germany	-
115	Babolnai	PI298092	40172	Hungary	Landrace
116	Baca	Vir42146	40135	Czechoslovakia	Cultivar
117	Barnhem	Barnherm	40294	Holland	Cultivar
118	Bronsyn	Bronsyn	40288	Holland	Cultivar
119	Burlina1	Burlina1	40282	Denmark	Cultivar
120	CHANTAL	NGB15401	40205	Denmark	Cultivar
121	Calvano1	Calvano1	40284	Denmark	Cultivar
122	DASAS-TRIFOLIUM	NGB4117	40206	Denmark	Cultivar
123	DE28845	De28845	40031	Germany	Ecotype
124	DE28886	DE28846	40029	Germany	Ecotype
125	DE28848	DE28848	40011	Germany	Ecotype
126	DE28849	DE28849	40030	Germany	Ecotype
127	DE28851	DE28851	40018	Germany	Ecotype
128	DE28856	DE28856	40012	Germany	Ecotype
129	DE39913	DE39913	40032	Germany	Ecotype
130	DE39932	DE39932	40026	Germany	Ecotype
131	DE33943	DE33943	40017	Germany	Ecotype
132	DE50653	DE50653	40015	Germany	Ecotype
133	DE50654	DE50654	40019	Germany	Ecotype
134	DE50667	DE50667	40023	Germany	Ecotype
135	DE50668	DE50668	40027	Germany	Ecotype
136	DE51971	DE51971	40024	Germany	Ecotype
137	DE51987	DE51987	40028	Germany	Ecotype

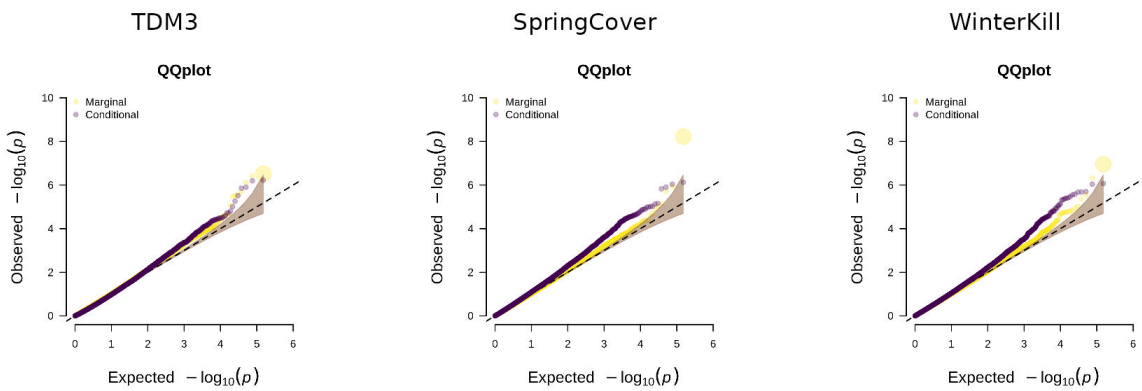
138	DE51990	DE51990	40022	Germany	Ecotype
139	DE54063	DE54063	40040	Romania	Ecotype
140	DE54064	DE54064	40041	Romania	Ecotype
141	DE54065	DE54065	40042	Romania	Ecotype
142	DE54084	DE54084	40044	Romania	Ecotype
143	DE54085	DE54085	40045	Romania	Ecotype
144	DE54088	DE54088	40046	Romania	Ecotype
145	DE54089	DE54089	40047	Romania	Ecotype
146	DE54090	DE54090	40048	Romania	Ecotype
147	DE54091	DE54091	40049	Romania	Ecotype
148	DE54092	DE54092	40050	Romania	Ecotype
149	DE54093	DE54093	40051	Romania	Ecotype
150	DE54456	DE54456	40020	Germany	Ecotype
151	DE54461	DE54461	40016	Germany	Ecotype
152	DE59277	DE59277	40025	Germany	Ecotype
153	DE59284	DE59284	40014	Germany	Ecotype
154	DE62301	DE62301	40038	France	Ecotype
155	DE62308	DE62308	40035	France	Ecotype
156	DE62309	DE62309	40033	France	Ecotype
157	DE62310	DE62310	40034	France	Ecotype
158	DE62312	DE62312	40052	France	Ecotype
159	DE62313	DE62313	40039	France	Ecotype
160	DE62315	DE62315	40036	France	Ecotype
161	DELTA	NGB2731	40239	Sweden	Cultivar
162	E12	NGB2594	40234	Sweden	Breeding line
163	EST279	EST279	40130	Estonia	Breeding line
164	E-4-Kockoko	Vir40298	40134	Poland	Cultivar
165	Fjälkinge-SB2501	NGB4341	40256	Sweden	Ecotype
166	Fure	NGB2209	40226	Norway	Cultivar
167	Gothem-TL0305	NGB1533	40257	Sweden	Ecotype
168	Gunne	NHB2732	40240	Sweden	Cultivar
169	Georgikon	PI632542	40197	Hungary	Cultivar
170	Hagestad-PW1301	NGB4342	40258	Sweden	Ecotype
171	Häljaröd-JK3103	NGB4345	40259	Sweden	Ecotype
172	Hörsene-TL0102	NGB1523	40260	Sweden	Ecotype
173	Inka	POL155176	40398	Poland	Cultivar
174	JO-0110	DE7063	40001	Finland	Breeding line
175	Jo-231	Vir38563	40133	Finland	Cultivar
176	Karcagi	PI298091	40171	Hungary	Landrace
177	Karcagi	PI632510	40196	Hungary	Cultivar
178	Kihelkonna-RA02066	EST963	40125	Estonia	Ecotype
179	Leningradskij-809	DE4573	40010	Soviet Union	Breeding line
180	Leninggradskii	Vir51004	40151	Russia	Cultivar
181	Leningradskii-809	Vir20258	40131	Russia	Cultivar
182	Lorina	Vir49885	40154	France	Cultivar
183	Markinskij-24	DE4569	40007	Soviet Union	Breeding line
184	Moskovwkij-84	DE4570	40008	Soviet Union	Breeding line
185	Martlett	Vir40889	40144	Austria	Cultivar

186	NIGA	POL155230	40404	Poland	Cultivar
187	NIRA	POL155192	40399	Poland	Cultivar
188	NK-200	14G2000204	40119	USA	Cultivar
189	Norlea	PI278773	40170	Canda	Cultivar
190	Pasavy	DE4571	40006	Soviet Union	Cultivar
191	Patora	NGB1637	40207	Denmark	Cultivar
192	Pavo	NGB13331	40242	Sweden	Cultivar
193	PI197270	PI197270	40163	Finland	Landrace
194	PI198070	PI198070	40164	Sweden	Landrace
195	PI204710	PI204710	40165	Turkey	Ecotype
196	PI205278	PI205278	40166	Turkey	Ecotype
197	PI272120	PI272120	40167	Poland	Ecotype
198	PI272121	PI272121	40168	Poland	-
199	PI274637	PI274637	40169	Poland	Ecotype
200	POL133323	POL133323	40374	Poland	Ecotype
201	POL133323	POL133323	40374	Ukraine	Ecotype
202	POL133324	POL133324	40375	Ukraine	Ecotype
203	POL133325	POL133325	40376	Slovakia	Ecotype
204	POL133326	POL133326	40377	Slovakia	Ecotype
205	POL133327	POL133327	40378	Ukraine	Ecotype
206	POL133328	POL133328	40379	Poland	-
207	POL133398	POL133398	40380	Poland	-
208	POL133399	POL133399	40381	Poland	-
209	POL133400	POL133400	40382	Poland	-
210	POL133401	POL133401	40383	Poland	Ecotype
211	POL155145	POL155145	40384	Poland	Ecotype
212	POL155147	POL155147	40385	Poland	Ecotype
213	POL155148	POL155148	40386	Poland	Ecotype
214	POL155150	POL155150	40387	Poland	Ecotype
215	POL155151	POL155151	40388	Poland	Ecotype
216	POL155152	POL155152	40389	Poland	Ecotype
217	POL155153	POL155153	40390	Poland	Ecotype
218	POL155154	POL155154	40391	Poland	Ecotype
219	POL155155	POL155155	40392	Poland	Ecotype
220	POL155158	POL155158	40393	Poland	Ecotype
221	POL155159	POL155159	40394	Poland	Ecotype
222	POL155160	POL155160	40395	Poland	Ecotype
223	Presto-Pajbjerg	NGB8378	40396	Denmark	Cultivar
224	Priekulskij-59	DE4572	40209	Soviet Union	Breeding line
226	Perma	Vir36944	40116	Holland	Cultivar
227	Portrush	Portrush	40153	UK	Cultivar
228	Portstewart	Portstewart	40297	UK	Cultivar
229	Prosperowo/BY	Vir47191	40296	Poland	Cultivar
230	Pürksi-Karja-RA0271	EST968	40139	Estonia	Ecotype
231	RIIKKA	NGB8417	20221	Finland	Cultivar
232	RONJA	BGB13330	40243	Sweden	Cultivar
233	RONJA	NGB14164	40244	Sweden	Cultivar
234	Raidi	EST49	40123	Estonia	Cultivar

235	SERVO	NGB2444	40245	Sweden	Cultivar
236	SILDIG-DAENO III	NGB11675	40218	Denmark	Cultivar
237	SILDIG HUNSBALLE	NGB1641	40211	Denmark	Cultivar
238	SIVERSKIJ-809	DE22479	40004	Soviet Union	Cultivar
239	SVEA	NGB2730	40246	Sweden	Cultivar
240	SVENSK-01408-NO	NGB11673	40235	Sweden	Breeding line
241	SW-E39	NGB13887	40236	Sweden	Breeding line
242	SW-E50	NGB13888	40237	Sweden	Breeding line
243	TACA-Trifolium	NGB1642	40212	Denmark	Cultivar
244	TAYA	NGB13328	40213	Denmark	Cultivar
245	TOFTA-TL0102	NGB1568	40261	Sweden	Ecotype
246	TRANI	NHB4092	40215	Denmark	Cultivar
247	Toronto	Toronto	40287	Germany	Cultivar
248	VALINGE	NGB2602	40220	Finland	Cultivar
249	VALINGE	NGB2590	40251	Sweden	Landrace
250	VALINGE	NGB4090	50252	Sweden	Landrace
251	VEJA	DE22481	40005	Soviet Union	Cultivar
252	VERNA-PAJBJERG	NGB1647	40217	Denmark	Cultivar
253	VIKTORIA	NGB6267	40248	Sweden	Landrace
254	VIRIS	NGB2360	40249	Sweden	Landrace
255	VIVA	NGB2729	40250	Sweden	Cultivar
256	VOLL-16-62-2	NGB4266	40233	Norway	Cultivar
257	Vir37860	Vir37860	40132	Romania	Cultivar
258	Vir47200	Vir47200	40155	Poland	Ecotype
259	Vir47226	Vir47226	40142	Poland	Ecotype
260	Vir49962	Vir49962	40145	Moldova	Ecotype
261	Vir51517	Vir51517	40159	USA	-
262	Vir51519	Vir51519	40161	Russia	Ecotype
263	Vir51520	Vir51520	40162	Ukraine	Ecotype
264	WIR20258	NGB4261	40223	Norway	Breeding line
265	Yomulin/PT	Vir47196	40140	Poland	Cultivar

Supplementary table 2: List of environmental covariates used to construct the environmental similarity matrices used for the genomic prediction models.

ID	Environmental covariate
ALLSKY_SFS_SW_DWN	All sky insolation incident on a horizontal surface
ALLSKY_SFC_LW_DWN	Thermal infrared longwave radiative flux
ETP	Evatransportation
FRUE	Effect of temperature and radiation use efficiency
n	Duration of sunshine hours
N	Daylight hours
PETP	Atmospheric water deficit
PRECTOT	Rainfall precipitation
RH2M	Relative air humidity at 2 Meters
RTA	Global Solar Radiation based on latitude and Julian day
SPV	Slope of saturation vapour pressure curve
T2MDEW	Dew-point temperature at 2 Meters
T2M	Temperature at 2 Meters
T2M_MIN	Minimum air temperature at 2 Meters
T2M_MAX	Maximum air temperature at 2 Meters
T2M_RANGE	Temperature range at 2 Meters
VPD	Vapour pressure deficiency
WS2M	Wind speed at 2 Meters



Supplementary Figure 1: QQ-plots for the p-values from the GWAS analyses. P-values largely follow the diagonal, which suggests that our models effectively control for confounding factors like population stratification.

Supplementary table 3: P-values for SNPs, below the genomewide threshold (FDR).
Signif. codes: '*' 0.001, '**' 0.01, '*' 0.05, '' 0.10, 'NS' Non-significant**

Trait	Analysis	Chr	Position	P-value	FDR	Threshold
SpringCover	Conditional	1	137827458	$1.487 \cdot 10^{-6}$	0.056	NS
SpringCover	Conditional	2	312660071	$7.544 \cdot 10^{-7}$	0.056	NS
SpringCover	Conditional	4	106918014	$1.251 \cdot 10^{-6}$	0.056	NS
SpringCover	Conditional	4	291504000	$9.135 \cdot 10^{-7}$	0.056	NS
SpringCover	Marginal	2	340594974	$5.985 \cdot 10^{-9}$	0.0009	**
SpringCover	Marginal	6	198771042	$1.952 \cdot 10^{-6}$	0.098	NS
SpringCover	Marginal	6	198771051	$1.171 \cdot 10^{-6}$	0.088	NS
TDM3	Conditional	2	161529631	$1.255 \cdot 10^{-6}$	0.052	NS
TDM3	Conditional	2	221602036	$6.015 \cdot 10^{-7}$	0.048	*
TDM3	Conditional	4	189930020	$3.072 \cdot 10^{-6}$	0.092	NS
TDM3	Conditional	4	96059970	$6.376 \cdot 10^{-7}$	0.048	*
TDM3	Conditional	6	35366913	$1.399 \cdot 10^{-6}$	0.052	NS
TDM3	Marginal	1	193676613	$3.362 \cdot 10^{-6}$	0.072	NS
TDM3	Marginal	2	188828089	$2.678 \cdot 10^{-6}$	0.072	NS
TDM3	Marginal	3	306736484	$2.962 \cdot 10^{-7}$	0.028	*
TDM3	Marginal	4	277056746	$7.569 \cdot 10^{-7}$	0.038	*
TDM3	Marginal	7	103894370	$3.788 \cdot 10^{-7}$	0.028	*
TDM3	Marginal	7	159935238	$3.165 \cdot 10^{-6}$	0.072	NS
TDM3	Marginal	7	159935246	$2.054 \cdot 10^{-6}$	0.072	NS
WinterKill	Conditional	1	12416317	$1.871 \cdot 10^{-5}$	0.098	NS
WinterKill	Conditional	1	149556644	$3.328 \cdot 10^{-6}$	0.052	NS
WinterKill	Conditional	1	212715131	$1.553 \cdot 10^{-5}$	0.094	NS
WinterKill	Conditional	1	227248048	$2.109 \cdot 10^{-6}$	0.052	NS
WinterKill	Conditional	1	74714761	$8.925 \cdot 10^{-6}$	0.075	NS
WinterKill	Conditional	2	218558177	$7.143 \cdot 10^{-6}$	0.071	NS
WinterKill	Conditional	2	312660071	$8.553 \cdot 10^{-7}$	0.052	NS
WinterKill	Conditional	2	340594974	$1.560 \cdot 10^{-5}$	0.094	NS
WinterKill	Conditional	2	57085778	$1.551 \cdot 10^{-5}$	0.094	NS
WinterKill	Conditional	3	168639278	$9.139 \cdot 10^{-7}$	0.052	NS
WinterKill	Conditional	3	249015291	$4.021 \cdot 10^{-6}$	0.052	NS
WinterKill	Conditional	3	293400020	$1.549 \cdot 10^{-5}$	0.094	NS
WinterKill	Conditional	3	74496447	$4.860 \cdot 10^{-6}$	0.052	NS
WinterKill	Conditional	3	83644758	$4.382 \cdot 10^{-6}$	0.052	NS
WinterKill	Conditional	3	83644766	$4.728 \cdot 10^{-6}$	0.052	NS
WinterKill	Conditional	3	83644804	$9.002 \cdot 10^{-6}$	0.075	NS
WinterKill	Conditional	4	152399748	$2.794 \cdot 10^{-6}$	0.052	NS
WinterKill	Conditional	4	201753647	$1.152 \cdot 10^{-5}$	0.085	NS
WinterKill	Conditional	4	53655409	$7.862 \cdot 10^{-6}$	0.074	NS
WinterKill	Conditional	5	203021725	$2.087 \cdot 10^{-6}$	0.052	NS
WinterKill	Conditional	5	203021751	$2.621 \cdot 10^{-6}$	0.052	NS
WinterKill	Conditional	6	107590067	$1.696 \cdot 10^{-5}$	0.098	NS
WinterKill	Conditional	6	64101962	$1.897 \cdot 10^{-5}$	0.098	NS
WinterKill	Conditional	6	75411051	$3.729 \cdot 10^{-6}$	0.052	NS
WinterKill	Conditional	6	75411052	$4.011 \cdot 10^{-6}$	0.052	NS
WinterKill	Conditional	7	124140775	$1.176 \cdot 10^{-5}$	0.085	NS
WinterKill	Conditional	7	203027370	$1.871 \cdot 10^{-5}$	0.098	NS
WinterKill	Conditional	7	203277106	$1.194 \cdot 10^{-5}$	0.085	NS
WinterKill	Conditional	7	325317951	$1.998 \cdot 10^{-6}$	0.052	NS
WinterKill	Marginal	1	176887443	$1.121 \cdot 10^{-7}$	0.016	*
WinterKill	Marginal	7	114853062	$4.824 \cdot 10^{-7}$	0.036	*

1