

Genomic prediction of yield and deep root growth in wheat using a semi-field phenotyping system with a water availability gradient

Xiangyu Guo (✉ xiangyu.guo@mbg.au.dk)

Center for Quantitative Genetics and Genomics, Aarhus University <https://orcid.org/0000-0001-7354-1519>

Simon F. Svane

Department of Plant and Environmental Science, University of Copenhagen, 1871 Frederiksberg

Winnie S. Füchtbauer

Sejet Plant Breeding I/S, 8700 Horsens

Jeppe R. Andersen

Nordic Seed A/S, 8300 Odder

Just Jensen

Center for Quantitative Genetics and Genomics, Aarhus University, 8830 Tjele

Kristian Thorup-Kristensen

Department of Plant and Environmental Science, University of Copenhagen, 1871 Frederiksberg

Research article

Keywords: genomic prediction, wheat, semi-field, grain-related, deep root

Posted Date: January 29th, 2020

DOI: <https://doi.org/10.21203/rs.2.22201/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background Deeper roots can help plants to take up available resources in deep soil ensuring better growth and higher yields under conditions of limited water supply. The objective was to explore the possibility of genomic prediction for grain-related and root traits of winter wheat measured in a semi-field root phenotyping facility allowing for a water availability gradient and interaction of genomic by water availability gradient. Genotyped winter wheat lines were grown as rows within two subunits of the semi-field facility. Along each row, a water availability gradient was applied by varying the depth distance to a subsurface-irrigation system. Root data were collected based on images taken from a minirhizotron tube below each row in one side of each bed. Results The analysis comprised four grain-related traits: grain yield (GY), protein concentration (PC), total nitrogen content (NC), and thousand kernel weight (TKW) measured on each half row along the water gradient. In addition, two root traits, total root length between 1.2 m and 2 m depth (TRL) and root length in four intervals (IRL) on each tube were analyzed. Two sets of models with or without the effect of neighbors from both sides of each row were applied. Variance components of all effects in the models were estimated. Predictive ability (PA) was measured as the correlation between mean phenotypes corrected for fixed effects and predicted genomic breeding values for each line. Estimated heritabilities (h^2) ranged from 0.263 to 0.680 for grain-related traits and from 0.030 to 0.055 for root traits. No interaction of genomic by water availability gradient was found. The relative amount of genetic variance was similar for grain-related and root traits. The PA ranged from 0.246 to 0.477 for grain-related traits and from 0.057 to 0.062 for root traits. Including neighbor effects in the model increased the h^2 and the PA for GY and NC. Conclusions In conclusion, it is possible to do genomic prediction for grain-related and root traits in the semi-field facility. Accuracy of root records is low using the current evaluation system due to large environmental variance, and more accurate root records are needed.

Background

Wheat, one of the most important food crops, is grown across widely different regions in the world [1]. To meet the demand of a growing human population, an intensification of agriculture, which allows greater yield to be obtained from existing farmland with sustainable inputs of water and fertilizer, is needed [2, 3]. Many of the constraints on yield in agricultural systems were found in the root system of crops [4]. However, the investigation of the root system is difficult because of the complexity of the soil environment compared with laboratory conditions [5]. To investigate the deep root system in a non-laboratory condition as well as to obtain direct and stable measurements, a semi-field root phenotyping facility has been constructed recently [6]. This facility had a capacity of 300 rows in each of four experimental units (beds) so that relatively large plant populations could be sown and studied. The facility is designed for both direct phenotyping of roots traits through minirhizotron and indirect screening of genotypes based on a water stress gradient.

Plant breeding has been successful in increasing the yield of most major crops [7]. This has contributed to a steady increase in cereal production without large increase in the acreage devoted for the production

[7]. The significant progress has been driven by the extensive research in developing and establishing new breeding technologies and rapid uptake of such new technology by breeding programs [8]. A suggested strategy to improve breeding has been the utilize of molecular markers in breeding program through marker-assisted selection (MAS) [9], and the use of genome-wide association studies (GWAS) combined with subsequent MAS has been shown as an effective tool in breeding program for qualitative traits under simple genetic control [10]. For the complex polygenic quantitative traits, e.g. grain yield, the application of MAS did not provide considerable genetic improvement. Instead, the implementation of genomic selection (GS) has been studied and shown as a promising method to be applied in the breeding program for polygenic traits [11].

In GS, a large set of markers are simultaneously incorporated into a model in order to obtain genomic prediction (GP) of genomic estimated breeding values (GEBVs) [11]. The efficiency of a breeding program can be improved and the cost of resources can be reduced by applying GP to predict GEBVs of individuals for selection as early as possible [12]. The accuracy of GP in wheat has been investigated in several recent studies, in which different population sizes and various models/methods were compared [8, 13-16]. Many factors can affect the prediction accuracy in GP, for example, size of training data, relatedness between training and test individuals, cross-validation strategies, marker density, as well as the model/method used for prediction of GEBVs [17, 18].

One of the most popular statistical models used for prediction of GEBVs is genomic best linear unbiased prediction (GBLUP) model, which integrates the genomic relationship matrix (**G**) constructed from all available markers [19]. Based on the GBLUP model, various extensions of model have been proposed and investigated. For example, in addition to additive genetic effects, non-additive genetic effects such as dominance and epistatic effects can also be included in the model [20]. Another example is modeling of genotype-by-environment interactions [21].

In practice, micro-environments differ across a field, and neighboring individuals can have similarity in their phenotypic performance because of the sharing of micro-environment [22]. Therefore, some studies investigated the spatial effects in linear mixed model by fitting the spatial effects using different strategies [22-24]. Another issue which should be considered when plants are grown in an individual row field trial is the influence from the neighboring individuals. There can be through competition for resources, such as water and nutrients from the soil, and light capture above ground. It was shown that the inclusion of such effects from neighbors could decrease error in estimation of genetic effects [25].

The objectives of this study were to 1) investigate genetic variation in grain-related and deep root traits in winter wheat; 2) study genomic by water availability interaction; 3) quantifying neighbor effects when lines were grown in adjacent rows; 4) analyze the possibilities of predicting breeding value of new wheat lines based on the genomic prediction from the semi-field root phenotyping facility.

Results

In this study, four grain-related traits were grain yield (GY), protein concentration (PC), total nitrogen content (NC) and thousand kernel weight (TKW), and two root traits were tube root length (TRL) and interval root length (IRL). Based on the grain-related and root data from a semi-field phenotyping system, variance components (VCs) for each trait were estimated and genomic cross-validation were applied and evaluated.

Descriptive statistics

Table 1 gives the descriptive statistics of all the traits analyzed in this study. As shown in **Table 1**, the number of observations for grain-related traits were 1,043 to 1,045. Some records were missing compared with the potential of 1,200 (2 beds, 2 position, 2 areas and 150 rows) records in principle, due to failure in growth or seeding errors in these rows. The average of GY, PC, NC and TKW, respectively, were 7.44 t ha⁻¹, 11.49%, 135.82 Kg N ha⁻¹ and 39.99 g. For root traits, there were 255 records for TRL and 986 records for IRL. The average of TRL was 99.79 cm, and the average for IRL was 25.81 cm.

Estimation of variance components

Estimated relative variance components (RVCs) for grain-related traits are shown in **Fig. 1**. For each grain-related trait, two different models with neighbor effects (GM2) or without neighbor effects (GM1) were applied. The proportion of each VC was calculated for the wet and dry areas separately due to the different spatial variance within each area according to the preliminary results.

Averaged for wet and dry areas, the RVC of genomic effect (***g***) when using GM1 which is equivalent to estimated narrow sense heritability (h^2), were 0.414 for GY, 0.263 for TKW, 0.350 for PC and 0.447 for NC. For PC, in the wet area was around 3.8% larger than in the dry area. For other grain-related traits, were more similar between wet and dry areas. In GM2, including of neighbor effects improved the h^2 for GY and NC. Averaged across two areas, the h^2 increased to 0.633 for GY and 0.680 for NC.

The RVC for line effect (***l***) were 0.391 for GY, 0.697 for TKW, 0.537 for PC, and 0.335 for NC, averaged across wet and dry areas by using GM1. For PC, the RVC for ***l*** in the wet area was 4.00% larger than in the dry area. The RVC for ***l*** were similar between two areas in other grain-related traits. Significant change when using GM2 was also observed in the RVC of ***l*** for GY and NC. When including neighbor effects in the model, RVC for ***l*** decreased to 0.086 for GY and to 0.026 for NC.

The RVC for row effect (***r***) were relatively small compared with ***g*** and ***l***, the average of two areas using GM1 were 0.060 for GY, 0.006 for TKW, 0.013 for PC, and 0.056 for NC. There was also a trend of decreasing for the RVC of ***r*** in GY and NC when including neighbor effects in the model (GM2)

When using GM1, the RVC due to spatial effect (***s***) were 0.027 for GY, 0.008 for TKW and 0.034 for PC in the wet area, while in the dry area, they were 0.032 (GY), 0.018 (TKW) and 0.071 (PC). The RVC due to ***s*** for NC was same for both wet and dry areas. A tendency of decreasing was also observed for the RVC of ***s*** in GY and NC when including neighbor effects in the model (GM2).

The RVC due to residual effect (e) were quite similar and stable between wet and dry areas as well as different models. When using GM1, the RVC for e were 0.106 for GY, 0.023 for TKW, 0.050 for PC and 0.128 for NC. The inclusion of neighbor effects did not change the RVC for e for TKW and PC, while for GY and NC the RVC of e decreased.

Consistent with the significant change of RVC in GY and NC when using GM2, a higher proportion of genomic effects for neighbor (g_n) and genomic effects for line (l_n) was observed in these two traits compared with TKW and PC. The RVC of g_n were 0.034 for GY and 0.013 for NC, and the RVC of l_n were 0.084 for GY and 0.110 for NC. The neighbor effects were small in TKW and PC. And there was no difference for neighbor effects between wet and dry areas.

Estimated RVC for each line in root traits are shown in **Fig. 2**. Due to model complexity, only the models without neighbor effects (RM1 and RM2) were applied for root traits. Compared with grain-related traits, the were small in root traits. The were 0.030 for TRL and 0.055 for IRL, the RVC of l were 0.171 for TRL and 0.252 for IRL. However, the RVC for s and e were large in both root traits. The RVC for s were 0.464 for TRL and 0.177 for IRL, and the RVC for e were 0.334 for TLR and 0.481 for IRL. For IRL, the effects of l accounted for 0.036 of the total phenotypic variances.

In addition to the estimation of VCs, coefficient of genetic variation, calculated as the ratio between square root of variance explained by g and the mean value of observations, were also computed and the results are presented in **Fig. 3**. In order to compare grain-related and root traits, only the results from models without neighbor effects are reported here. For grain-related traits, the coefficient of genomic variation (c_v) were 8.24% for GY, 3.81% for TKW, 2.49% for PC and 7.80% for NC. The c_v for root traits were 3.88% for TRL and 4.20% for IRL. Even though the of root traits was low compared with grain-related traits, the c_v of root traits were higher than for TKW and PC.

Genomic prediction and cross-validation

Leave-one-line-out cross-validation was conducted by using two models with or without neighbor effects. The predictive ability (PA) for each line in grain-related and root traits are presented in **Fig. 4**. This is the accuracy of predicting a genotyped line that does not have their own phenotypic records. The horizontal line on top of each bar in **Fig. 4** shows square root of h^2 (h) of line means for each trait and model, which shows the maximum potential correlation for each trait. The h averaged across wet and dry areas using GM1 were 0.643 (GY), 0.512 (TKW), 0.591 (PC) and 0.669 (NC). When using GM1, the PA was 0.344 for GY, 0.246 for TKW, 0.260 for PC, and 0.338 for NC. Considering neighbor effects in the model significantly increased the PA of direct breeding values for GY and NC, which were 0.475 (GY, h was 0.795) and 0.477 (NC, h was 0.825) when using GM2. Compared with GY and NC, the PA for PC increased a bit to 0.282 (h was 0.582) when using GM2, while the PA of TKW, which was 0.237 (h was 0.507), did not increase when using GM2. When using models without neighbor effects, the h was 0.172 (IRL) and 0.234 (TRL), and the PA for TRL was 0.057, and for IRL was 0.062.

The inflation of prediction for each line in grain-related and root traits are presented in **Fig. 5**. When using GM1, the regression coefficients were 1.045 for GY, 1.122 for TKW, 0.788 for PC and 1.054 for NC. When using GM2 by considering neighbor effects, the regression coefficients were 1.072 for GY, 1.094 for TKW, 0.855 for PC and 1.078 for NC. The inflation of prediction from GM1 and GM2 were similar. For root traits, the regression coefficients were 1.218 for TRL and 1.082 for IRL. The regression coefficients for all the grain-related and root traits were not significantly different from 1.

Discussion

In this study, grain-related and root traits of winter wheat measured in a semi-field root phenotyping facility [6] were analyzed. Variance components were estimated and genomic predictions were explored by using models with or without neighbor effects.

The models used in the current study did not include genomic by water availability area (wet and dry) effects since preliminary investigation showed that the proportion of variance due to genomic by area interactions were close to zero. The previous study also showed that there was no line by water availability area interaction [26, 27]. One of the reasons could be the design of the water availability gradient. The observations of grain-related traits were obtained for wet and dry areas by dividing the row into two equal sized areas from the either north or south part of the row. However, the water supply was not a strict low or high level of case-control experimental design. This gradient of water availability could reduce the variation causing difficulty in detecting genomic by area effects. Other factors might influence the lack of subsurface irrigation response, and it has been speculated that a lack of nutrients will lower growth potential when water is supplied to a nutrient poor subsoil, and do not provide water to the nutrient rich topsoil being dry at all locations across the water availability gradient [26]. Therefore, for the current study, genomic by area effects were excluded in the models.

Heritabilities and variance components

There were four grain-related and two root traits analyzed in this study. The proportion of total phenotypic variance accounted for by genomic effects (g), equivalent to the narrow sense h^2 , was estimated. Using different models, the estimates of h^2 of GY ranged from 0.41 to 0.67. The estimates were higher than the h^2 reported from some previous studies for phenotypic prediction [28, 29]. One of the studies on winter wheat GY reported h^2 , which ranged from 0.26 to 0.75, for 12 F_3 wheat populations [30]. The h^2 of TKW was estimated as 0.26 from both models. The h^2 of TKW reported from previous studies were similar or higher than in the current study, e.g. 0.32 [31] and 0.57 [32]. One study using genomic data obtained the h^2 estimated as 0.52 [33], which is also higher than in the current study. The reason could be the inclusion of line effects in the current models while in [33] line effects were not included due to lack of replication. The estimates of h^2 were around 0.35 for PC using both models. Compared with 0.51 reported in a previous study [33], the lower estimates in the current study can also be caused by fitting of line effects as discussed earlier. Estimated h^2 for NC range from 0.45 to 0.72 using different models in the current

study. One previous study showed that the broad sense heritability of NC as 0.31 in which 31% of the genetic variance was contributed by genotype and 66% was contributed by genotype by year [34].

The estimates of RVC for line effects (I) varied from 0.34 to 0.70 for grain-related traits when using GM1, and from 0.03 to 0.70 when using GM2. The variance of I includes part of the genetic variance which cannot be explained by genomic markers as well as non-additive genetic effects. The effects due to common area and origin of the seeds used when establishing the experiment could also be included into the variance of I . Decreasing RVC for I when including neighbor effects in the model, for GY and NC, also suggested that the variance of I included the variances from other sources e.g. neighbors. In the ideal randomization of the experiment, each line will not have the same neighbor twice, to prevent the variance of I including variances from neighbors, but in the current study, the randomization could be one of the reasons for this issue. Some preliminary studies were done by comparing models including I or not, from which the results suggested that the exclusion of I can create more inflation in genomic predictions (results not shown). Therefore, the models reported in the current study always considered the effects of I .

Spatial effects were modeled in this study and the estimates of RVC for spatial effects ranged from 0.01 to 0.07 for grain-related traits with both wet and dry areas and in both GM1 and GM2. Modeling neighbor effects did not affect the proportion of variances caused by spatial effects. The modeling of spatial effects was for the two water availability areas separately, since preliminary results showed that the pattern and the degree of spatial effects were different when analyzing the data within each water availability area. Spatial effects were investigated by some other studies in different ways. For example, considerable spatial variation was found in wheat data [23]. In their study, spatial variation was corrected by introducing a covariable, which was calculated as the value of phenotypic in each plot minus mean phenotypic value of neighboring plots. This introduces a regression of a function of the data that might lead to underestimation of total phenotypic variance. In this study, this was avoided by modeling the spatial effect as a running sum of random effects due to rows surrounding each plot. Another discovery from this study was that a more pronounced field variation was observed under dry conditions, and led to a lower h^2 than the other two fields either under full irrigation or under mild water stress [23].

Neighbor effects were considered by applying GM2 and both genomic effects for neighbors and line effects for neighbors were modeled. The estimates of RVC for g_n ranged from 0.003 to 0.042, and ranged from 0 to 0.110 for I_n . Modeling of neighbor effects led to significant change of variance structures for both GY and NC, whereas the changes in TKW and PC were small. For both GY and NC, the RVC accounted by g increased considerably but the I effect decreased in the same amount. This indicates that if neighbor effects are excluded from the model a large part of this effects is picked up by the I effect. In a study for cassava data [35], neighbor effects were investigated as a competitive ability. Through their simulation study, accuracy in estimating trait genotypic effect was found increased when including competitive effects in the model. Through their analysis for nine trials and four traits, significant competition effects were observed for 12 of the 36 combinations.

For the two root traits analyzed in this study, the estimates of h^2 ranged from 0.03 to 0.05, whereas the proportion of other effects was large. Though the narrow sense h^2 for root traits were low compared with grain-related traits analyzed in this study, there were relative same amount of genetic variation in the root traits. The c_v of root traits were even higher than TKW and PC, which suggested that there were comparable amounts of genetic variation for root traits. However, the root measurements were dominated by other effects e.g. spatial effects and measurement errors, which caused the limited h^2 obtained for root traits using the current phenotyping methods for root length. In addition, the current number of records also limited model complexity for these two traits.

Genomic prediction and validation

Both predictive ability and inflation of prediction were investigated in this study for both grain-related and root traits, and the results from models with or without neighbor effects were compared.

The PA measured as the correlation between corrected phenotypic values (y_c) and GEBVs range from 0.25 to 0.34 for the four grain-related traits when using GM1, and from 0.24 to 0.48 when using GM2 including neighbor effects. The PA obtained for grain-related traits ranged from 44.9% for PC using GM1 to 66.2% for GY using GM2 of the potential maximum correlation, which is the square root of h^2 for each trait. These ratios, which show the accuracy of predicting the underlying true additive genetic value, clearly show the power of GP utilizing information from related individuals to predict the additive genetic value of specific lines. Inclusion of neighbor effects significantly improved the PA for GY and NC. The increase in PA for these two traits was consistent with the marked change in the VC pattern when including neighbor effects. Improvement of PA when modeling neighbor effects were also found in the previously mentioned study for cassava where up to 25% increase in accuracy was observed [35]. The PA obtained in this experiment is expected to be relatively low due to the small size of the population used to train the models.

The earlier reported predictive ability (PA) for TKW ranged from around 0.26 to 0.50 by using two different validation schemes [33]. In this previous study, the 10-fold cross-validations, where the whole population was randomly divided into 10 groups, gave higher PA than the leave-set-out cross-validations, where each set was from one of the breeding cycles included in the study. Relatively larger PA compared with the current study could be explained by the materials used for the previous study was 1,152 F6 winter wheat lines from four different breeding cycles, which was much larger than in the current study, i.e. the increased accuracy can be explained by the larger training population. Effect from size of training dataset on the genomic prediction accuracy was investigated in the previous studies [15, 16]. For example, in the study for 309 spring barley lines, it was concluded that using less than 200 lines in the training set could result in low prediction accuracy [15]. Another study explored a total of 988 advanced wheat breeding lines for yield, lodging, and starch content, and the results showed that training population of around 700 lines were enough to yield the highest observed predictive abilities [16]. In the current study, only 84 lines were involved in the cross-validation procedure, considering this, the degree of PA was reasonable, but higher PA could be expected by increasing the number of lines included in the experiment.

Compared with TKW, the PA obtained for PC were a bit higher and ranged from 0.26 to 0.28. The PA for PC was also investigated in previous study [33]. The authors reported that in the 10-fold cross-validation scheme, the PA for PC were also bit higher than for TKW, however, the PA for PC obtained from the leave-set-out cross-validation scheme were lower than TKW. The PA of PC reported from another two studies were above 0.6 by using genomic-assisted models, which merged the merits of genomic selection with phenotypic selection in preliminary GY trials [36, 37].

The PA of NC were very close to the results for GY, and were 0.34 using GM1 and 0.48 using GM2. One of the reasons could be that the NC was estimated from grain GY and grain PC concentration. NC is required to maintain a photosynthetically active canopy to support GY and PC [38]. Grain NC is a reflection of nitrogen use efficiency, which could increase GY at a given level of nitrogen fertilizer [34]. The moderate PA obtained for NC can ensure a reasonable genetic progress when breeding for grain NC and facilitate the breeding for nitrogen use efficiency.

For root traits in the present study, the PA were both below 0.1, which was consistent with the low estimates of VC due to additive genetic effects. The PA of TRL was 33.4% of the potential maximum correlation and the PA of IRL was 26.4% of the potential maximum correlation. These ratios actually show the accuracy of predicting the underlying true additive genetic value. Though the PA were low, it was clearly shown that reasonable accuracies can also be obtained for the studied root traits in the current phenotyping facility. However, estimation of genetic variation, under the circumstance the current level of spatial and measurement errors, the chance to obtain reasonable predictions seem small.

The accuracy obtained for genomic predictions of additive genetic effects were expected for both grain-related and root traits since a relative limited training population were used in this study. Previous studies clearly showed that increasing the training population significantly improve the prediction accuracy [15, 16].

Conclusions

In this study, genomic prediction was carried out for grain-related and root traits of winter wheat measured in a semi-field root phenotyping facility. Phenotypic and genotypic data were analyzed for 84 winter wheat lines. Two sets of models were used in order to study consequences of including neighbor effects in the model. We report estimated variance components and performance of genomic prediction using leave-one-line-out strategy.

Results showed that the grain-related traits, GY, TKW, PC, and NC had high to moderate narrow sense heritabilities. For root traits, heritabilities were low. However, the relative amount of genetic variation for root traits were of similar magnitude as for the grain-related traits but the low heritabilities were caused by very large environmental variance due to spatial effects and measurement error.

In no case did we find any genomic by water availability interaction, which could be due to the water availability design in this study was a gradual gradient and it might weaken the power to detect the

genomic by water availability interaction.

For the above ground grain-related traits, we found effects of neighbors. The effects were generally small but including neighbor effects increased the estimated additive genetic variance of the traits and increased accuracy of predicting breeding values for lines without own phenotypic records.

For both grain-related and root traits it was possible to obtain genomic predictions of additive genetic effects with an accuracy as expected based on the limited training population available.

Especially the root traits had large environmental variance due to large spatial effects and measurement error. More accuracy in the methods for phenotyping deep root length are needed to obtain stronger genomic predictions.

Methods

(See Methods in the Supplementary Files)

List Of Abbreviations

CV: leave-one-line-out cross-validation

GBLUP: genomic best linear unbiased prediction

GEBVs: genomic estimated breeding values

GP: genomic prediction

GS: genomic selection

GWAS: genome-wide association studies

GY: grain yield

IRL: root length in four intervals

MAS: marker-assisted selection

MR: multispectral minirhizotron

NC: total nitrogen content

PA: Predictive ability

PC: protein concentration

REML: restricted maximum likelihood method

RVC: Relative variance components

TKW: thousand kernel weight

TRL: total root length between 1.2 m and 2 m depth

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Competing interests

The authors declare that they have no competing interests.

Funding

This project was funded by Innovation Fund Denmark (Grant no 5184-000358), Crop Innovation Denmark, Promilleafgiftsfonden, Plan Danmark and Godfred Birkedal Hartmanns Familiefond.

Authors' contributions

XG implemented and carried out the statistical analysis, interpreted the results and drafted the manuscript. SFS, WSF, JRA, and JJ contributed in developing the statistical models and interpreting the results. The experimental design was conceived by KTK and SFS. All authors read and approved the final manuscript.

Acknowledgements

The authors thank to the preliminary work for model building did by Hsin Yuan Jason Tsai and valuable suggestions given by other colleagues at Aarhus University. The authors also thank to Olga Popovic, Jesper Svendsgaard and many other researchers and research technicians at Copenhagen University for providing data. In addition, Lene Krusell, Finn Borum, Jens Due Jensen and many other plant breeders from Sejet and Nordic Seed are gratefully acknowledged for the contributions of plant material and technical assistance.

References

1. Tadesse W, Sanchez-Garcia M, Assefa SG, Amri A, Bishaw Z, Ogbonnaya FC, Baum M: **Genetic Gains in Wheat Breeding and Its Role in Feeding the World** %J *Crop Breeding, Genetics and Genomics*. 2019, **1**(1):e190005.
2. Gregory PJ, Atkinson CJ, Bengough AG, Else MA, Fernández-Fernández F, Harrison RJ, Schmidt S: **Contributions of roots and rootstocks to sustainable, intensified crop production**. *Journal of Experimental Botany* 2013, **64**(5):1209-1222.
3. Dai A: **Increasing drought under global warming in observations and models**. *Nature Climate Change* 2013, **3**(1):52-58.
4. Wasson AP, Chiu GS, Zwart AB, Binns TR: **Differentiating Wheat Genotypes by Bayesian Hierarchical Nonlinear Mixed Modeling of Wheat Root Density**. 2017, **8**(282).
5. Rich SM, Watt M: **Soil conditions and cereal root system architecture: review and considerations for linking Darwin and Weaver**. *Journal of Experimental Botany* 2013, **64**(5):1193-1208.
6. Svane SF, Jensen CS, Thorup-Kristensen K: **Construction of a large-scale semi-field facility to study genotypic differences in deep root growth and resources acquisition**. *Plant Methods* 2019, **15**(1):26.
7. Prohens J: **Plant Breeding: A Success Story to be Continued Thanks to the Advances in Genomics**. *Front Plant Sci* 2011, **2**:51.
8. Norman A, Taylor J, Tanaka E, Telfer P, Edwards J, Martinant J-P, Kuchel HJT, Genetics A: **Increased genomic prediction accuracy in wheat breeding using a large Australian panel**. 2017, **130**(12):2543-2555.
9. Koebner RMD, Summers RW: **21st century wheat breeding: plot selection or plate detection?** *Trends in Biotechnology* 2003, **21**(2):59-63.
10. Xu Y, Crouch JH: **Marker-Assisted Selection in Plant Breeding: From Publications to Practice**. 2008, **48**(2):391-407.
11. Meuwissen T, Hayes B, Goddard M: **Prediction of total genetic value using genome-wide dense marker maps**. *Genetics* 2001, **157**:1819-1829.
12. Mackay I, Ober E, Hickey J: **GplusE: beyond genomic selection**. *Food Energy Secur* 2015, **4**(1):25-35.
13. He S, Schulthess AW, Mirdita V, Zhao Y, Korzun V, Bothe R, Ebmeyer E, Reif JC, Jiang Y: **Genomic selection in a commercial winter wheat population**. *Theoretical and Applied Genetics* 2016, **129**(3):641-651.
14. Isidro J, Jannink J-L, Akdemir D, Poland J, Heslot N, Sorrells ME: **Training set optimization under population structure in genomic selection**. *Theoretical and Applied Genetics* 2015, **128**(1):145-158.
15. Nielsen NH, Jahoor A, Jensen JD, Orabi J, Cericola F, Edriss V, Jensen J: **Genomic Prediction of Seed Quality Traits Using Advanced Barley Breeding Lines**. *PLoS one* 2016, **11**(10):e0164494-e0164494.
16. Cericola F, Jahoor A, Orabi J, Andersen JR, Janss LL, Jensen J: **Optimizing Training Population Size and Genotyping Strategy for Genomic Prediction Using Association Study Results and Pedigree**

- Information. A Case of Study in Advanced Wheat Breeding Lines.** *PloS one* 2017, **12**(1):e0169606-e0169606.
17. Robertsen C, Hjortshøj R, Janss L: **Genomic Selection in Cereal Breeding.** *Agronomy* 2019, **9**(2).
 18. Norman A, Taylor J, Edwards J, Kuchel H: **Optimising Genomic Selection in Wheat: Effect of Marker Density, Population Size and Population Structure on Prediction Accuracy.** *G3 (Bethesda, Md)* 2018, **8**(9):2889-2899.
 19. VanRaden PM: **Efficient methods to compute genomic predictions.** *Journal of dairy science* 2008, **91**(11):4414-4423.
 20. Jiang Y, Schmidt RH, Zhao Y, Reif JC: **A quantitative genetic framework highlights the role of epistatic effects for grain-yield heterosis in bread wheat.** *Nature Genetics* 2017, **49**:1741.
 21. Pérez-Rodríguez P, Crossa J, Rutkoski J, Poland J, Singh R, Legarra A, Autrique E, Campos Gdl, Burgueño J, Dreisigacker S: **Single-Step Genomic and Pedigree Genotype × Environment Interaction Models for Predicting Wheat Lines in International Environments.** *The Plant Genome* 2017, **10**.
 22. Mao X, Dutta S, Wong R, Nettleton D: **Adjusting for Spatial Effects in Genomic Prediction;** 2019.
 23. Lado B, Matus I, Rodríguez A, Inostroza L, Poland J, Belzile F, del Pozo A, Quincke M, Castro M, von Zitzewitz J: **Increased genomic prediction accuracy in wheat breeding through spatial adjustment of field trial data.** *G3 (Bethesda, Md)* 2013, **3**(12):2105-2114.
 24. Bernal-Vasquez A-M, Möhring J, Schmidt M, Schönleben M, Schön C-C, Piepho H-P: **The importance of phenotypic data analysis for genomic prediction - a case study comparing different spatial models in rye.** *BMC Genomics* 2014, **15**(1):646.
 25. Rebetzke GJ, Fischer RA, van Herwaarden AF, Bonnett DG, Chenu K, Rattey AR, Fettell NA: **Plot size matters: interference from intergenotypic competition in plant phenotyping studies.** *Functional Plant Biology* 2014, **41**(2).
 26. Svane SF: **Deep root phenotyping.** *Ph.d. Thesis.* Copenhagen: University of Copenhagen, Faculty of Science, Department of Plant and Environmental Sciences, Section for Crop Sciences; 2019.
 27. Svane SF, Dam EB, Carstensen JM, Thorup-Kristensen K: **A multispectral camera system for automated minirhizotron image analysis.** *Plant and Soil* 2019, **441**(1):657-672.
 28. Teich AH: **Heritability of grain yield, plant height and test weight of a population of winter wheat adapted to Southwestern Ontario.** *Theoretical and Applied Genetics* 1984, **68**(1):21-23.
 29. Rife TW, Graybosch RA, Poland JA: **A Field-Based Analysis of Genetic Improvement for Grain Yield in Winter Wheat Cultivars Developed in the US Central Plains from 1992 to 2014.** *Crop Science* 2019, **59**(3):905-910.
 30. Aziz T, Khalil I, Hussain Q, Shah T, Ahmad N, Sohail A: **Heritability and Selection Response for Grain Yield and Associated Traits in F 3 Wheat Populations.** 2018, **34**:767-774.
 31. Kaya Y, Akcura M: **Effects of genotype and environment on grain yield and quality traits in bread wheat (*T. aestivum* L.).** *Food Science and Technology* 2014, **34**:386-393.

32. Rustgi S, Shafqat MN, Kumar N, Baenziger PS, Ali ML, Dweikat I, Campbell BT, Gill KS: **Genetic dissection of yield and its component traits using high-density composite map of wheat chromosome 3A: bridging gaps between QTLs and underlying genes.** *PLoS one* 2013, **8**(7):e70526-e70526.
33. Kristensen PS, Jahoor A, Andersen JR, Orabi J, Janss L, Jensen J: **Multi-Trait and Trait-Assisted Genomic Prediction of Winter Wheat Quality Traits Using Advanced Lines from Four Breeding Cycles.** *Crop Breeding, Genetics and Genomics* 2019, **1**(2):e190010.
34. Guttieri MJ, Frels K, Regassa T, Waters BM, Baenziger PS: **Variation for nitrogen use efficiency traits in current and historical great plains hard winter wheat.** *Euphytica* 2017, **213**(4):87.
35. Elias AA, Rabbi I, Kulakow P, Jannink J-L: **Improving Genomic Prediction in Cassava Field Experiments by Accounting for Interplot Competition.** *G3 (Bethesda, Md)* 2018, **8**(3):933-944.
36. Michel S, Ametz C, Gungor H, Akgöl B, Epure D, Grausgruber H, Löschenberger F, Buerstmayr H: **Genomic assisted selection for enhancing line breeding: merging genomic and phenotypic selection in winter wheat breeding programs with preliminary yield trials.** *TAG Theoretical and applied genetics Theoretische und angewandte Genetik* 2017, **130**(2):363-376.
37. Michel S, Löschenberger F, Ametz C, Pachler B, Sparry E, Buerstmayr H: **Simultaneous selection for grain yield and protein content in genomics-assisted wheat breeding.** *Theoretical and Applied Genetics* 2019, **132**(6):1745-1760.
38. Cormier F, Foulkes M, Hirel B, Gouache D, Moënne-Loccoz Y, Gouis J: **Breeding for increased nitrogen-use efficiency: A review for wheat (*T. aestivum* L.).** *Plant Breeding* 2016, **135**.
39. Madsen P, Jensen J: **A users guide to DMU. A package for analysing multivariate mixed models;** 2013.
40. Legarra A, Robert-Granié C, Manfredi E, Elsen J-M: **Performance of Genomic Selection in Mice.** *Genetics* 2008, **180**(1):611-618.

Tables

Table 1 Descriptive statistics for all the traits

Type	Trait	Unit	No. of records	Average	S.D.	Min	Max
Grain-related	GY	t ha ⁻¹	1,043	7.44	1.60	1.52	12.71
	TKW	g	1,044	39.99	3.82	25.47	50.35
	PC	%	1,045	11.49	0.93	9.41	15.50
	NC	kg N ha ⁻¹	1,043	135.82	26.72	28.70	218.07
Root	TRL	cm	255	99.79	37.13	16.21	198.59
	IRL	cm	986	25.81	20.01	0.01	80.46

GY is grain yield, TKW is thousand kernel weight, PC is protein concentration, NC is total nitrogen content, TRL is tube root length, IRL is interval root length.

Table 2 Editing rules for root data and number of records kept in each step

Step	Rule	No. of records
0	full data	21,057
1	remove records with no observation	20,479
2	divide depth into 8 intervals and remove records with soil depth larger than 2 m (4 m along tube) or smaller than 1.2 m (2 m along tube)	16,211
3	for each interval, remove records out of mean \pm 3sd	15,960
4	remove records from lines without genomic information	14,270

$$\text{Soil depth} = (\text{Tube depth} - 420\text{mm}) \cdot \sin(23.5^\circ) + 570\text{mm}$$

Table 3 Depth interval, range of tube (mm) and soil (cm) depth together with number of records

Interval	Min tube depth	Max tube depth	Min soil depth	Max soil depth	No. records
1	0	1,500	0	100.06	0
2	1,500	2,000	100.06	120.00	0
3	2,000	2,500	120.00	139.94	3,529
4	2,500	3,000	139.94	159.88	3,753
5	3,000	3,500	159.88	179.81	3,508
6	3,500	4,000	179.8	199.75	3,480
7	4,000	4,500	199.75	219.69	0
8	4,500	5,006	219.69	239.87	0

Soil depth = (Tube depth-420mm) \times sin(23.5°) + 570mm

Supplementary File Legends

Additional file 1:

Table S1 Genotypic data

Table S2 Grain-related phenotypic data

Table S3 Tube root length phenotypic data

Table S4 Interval root length phenotypic data

Figures

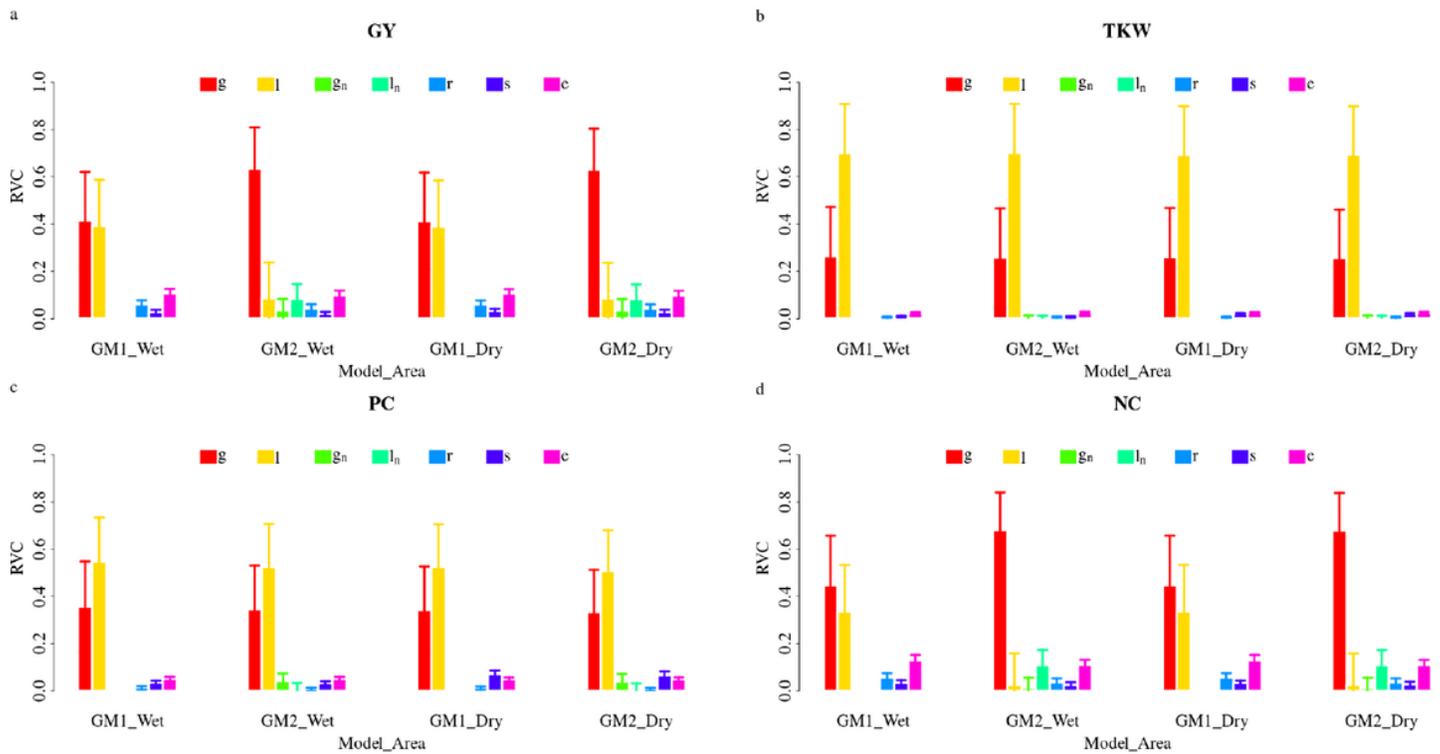


Figure 1

Estimated relative variance components for grain-related traits in different models and wet or dry parts a is grain yield (GY); b is thousand kernel weight (TKW); c is protein concentration (PC); and d is total nitrogen content (NC). g is additive genomic effect; l is line effect; gn is additive genomic effect for neighbors; ln is line effect for neighbors; r is row effect; s is spatial effect; e is residual effect. y-axis is relative variance components (RVC); x-axis is the model used and the area records obtained from.

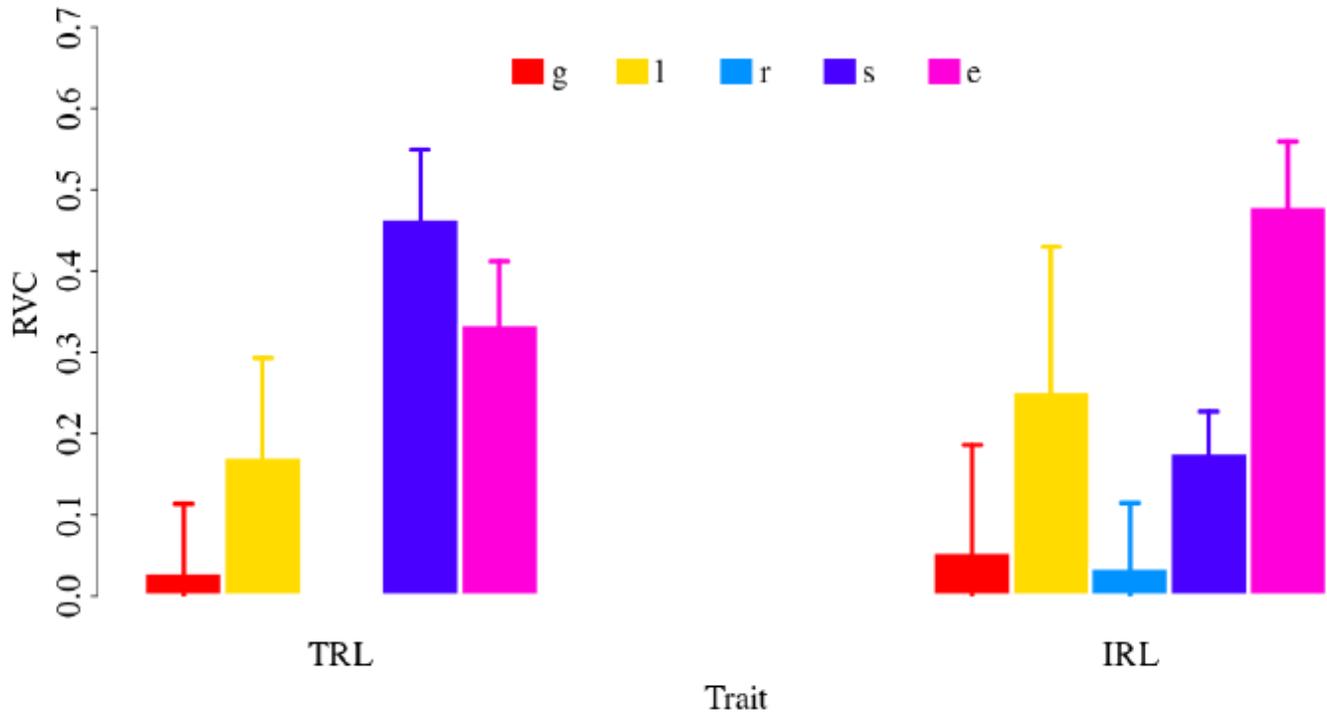


Figure 2

Estimated relative variance components for root traits g is additive genomic effect; l is line effect; r is row effect; s is spatial effect; e is residual effect. y-axis is relative variance components (RVC); x-axis is trait; TRL is tube root length, IRL is interval root length.

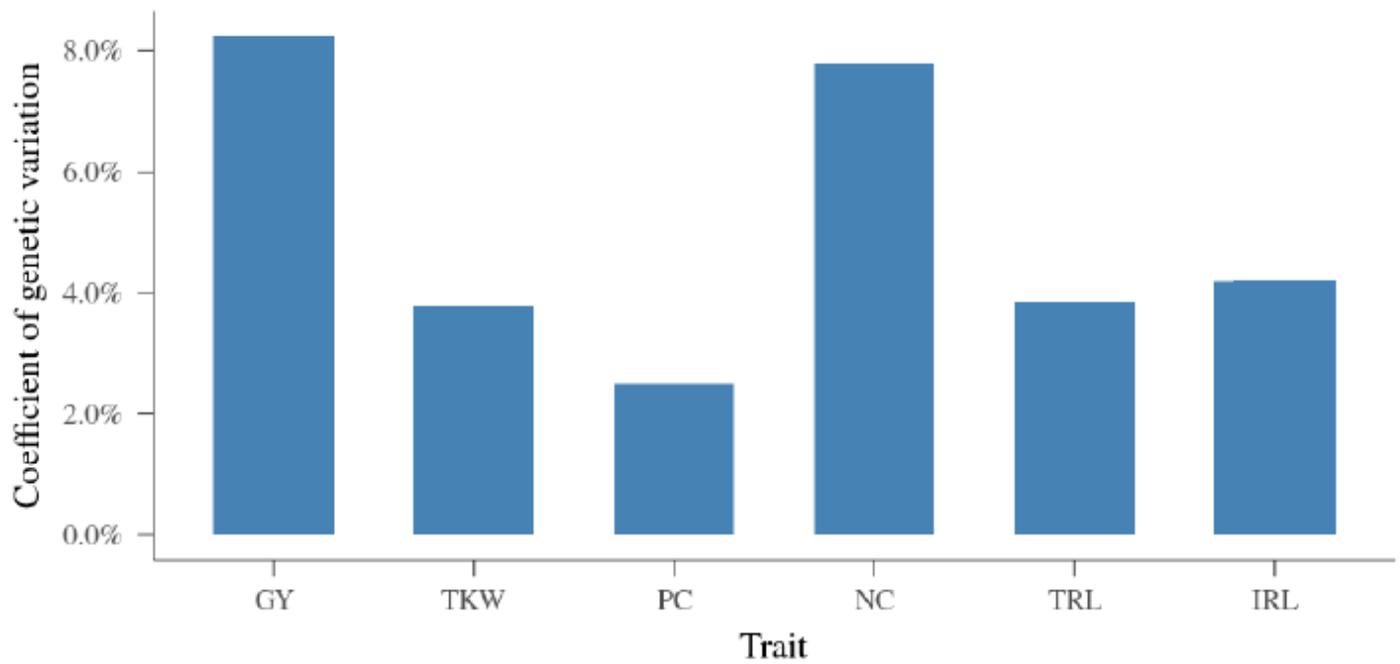


Figure 3

Coefficient of genetic variation for grain-related and root traits y-axis is coefficient of genetic variation; x-axis is trait; GY is grain yield, TKW is thousand kernel weight, PC is protein concentration, NC is total nitrogen content, TRL is tube root length, IRL is interval root length.

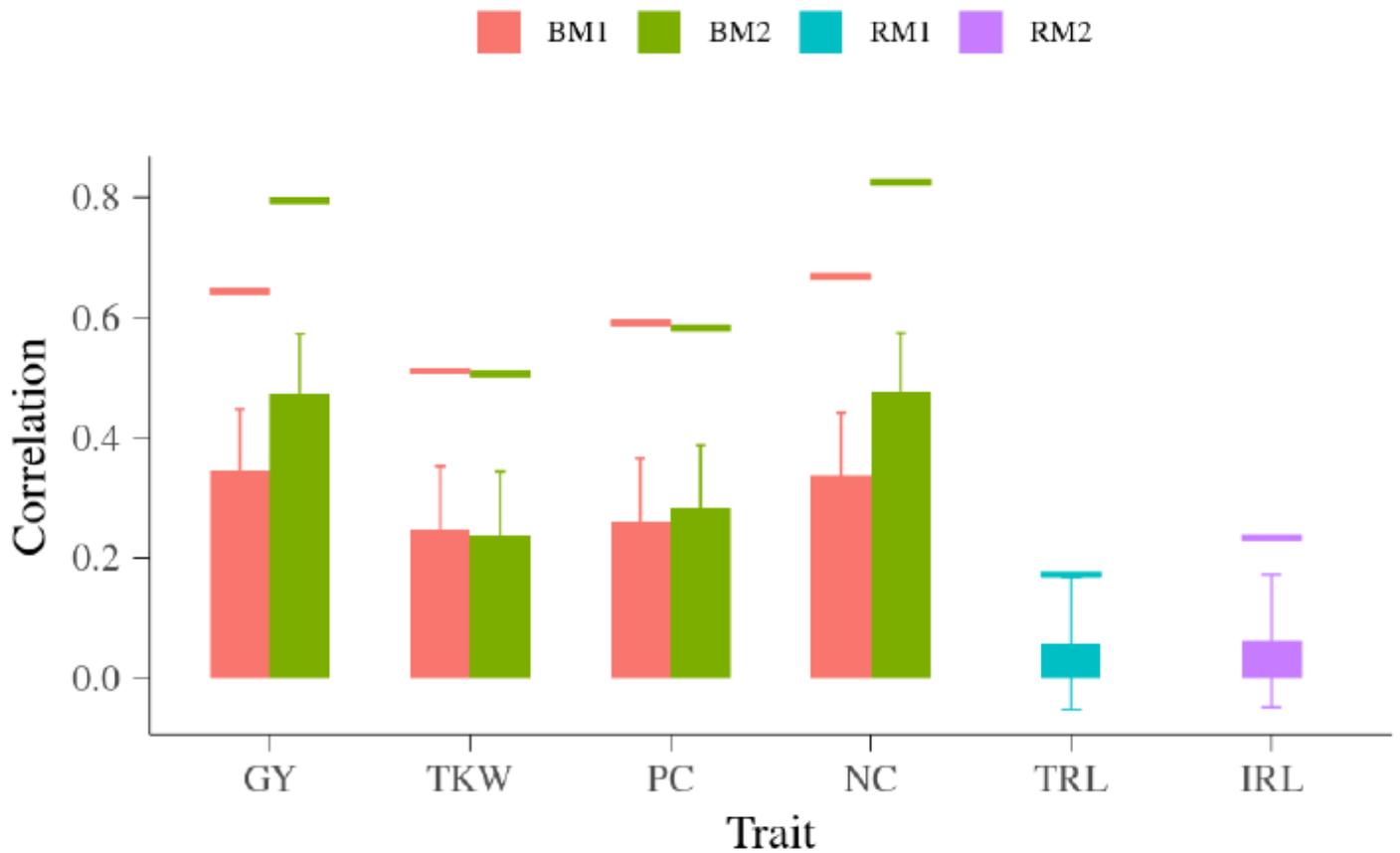


Figure 4

Predictive ability for each line in grain-related and root traits y-axis is correlation between $(y_c)^{\sim}$ and \hat{g} ; x-axis is trait; GY is grain yield, TKW is thousand kernel weight, PC is protein concentration, NC is total nitrogen content, TRL is tube root length, IRL is interval root length; horizontal line on top of each bar is square root of heritability of line mean for each trait and model, which shows the potential maximum correlation for each trait.

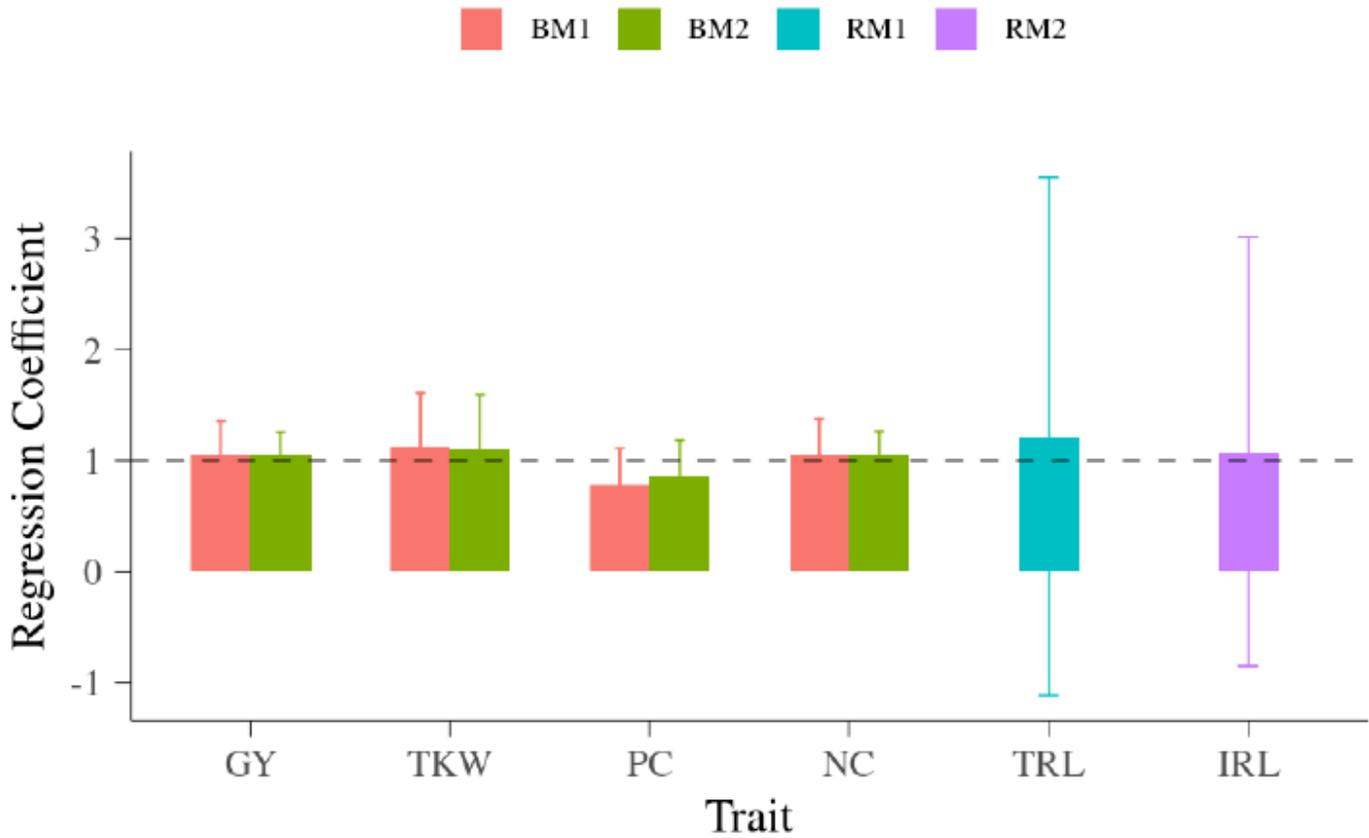


Figure 5

Inflation of prediction for each line in grain-related and root traits y-axis is regression coefficient of (y_c) on \hat{g} ; x-axis is trait; GY is grain yield, TKW is thousand kernel weight, PC is protein concentration, NC is total nitrogen content, TRL is tube root length, IRL is interval root length; horizontal dashed line is regression coefficient of 1, which is the situation of no inflation.

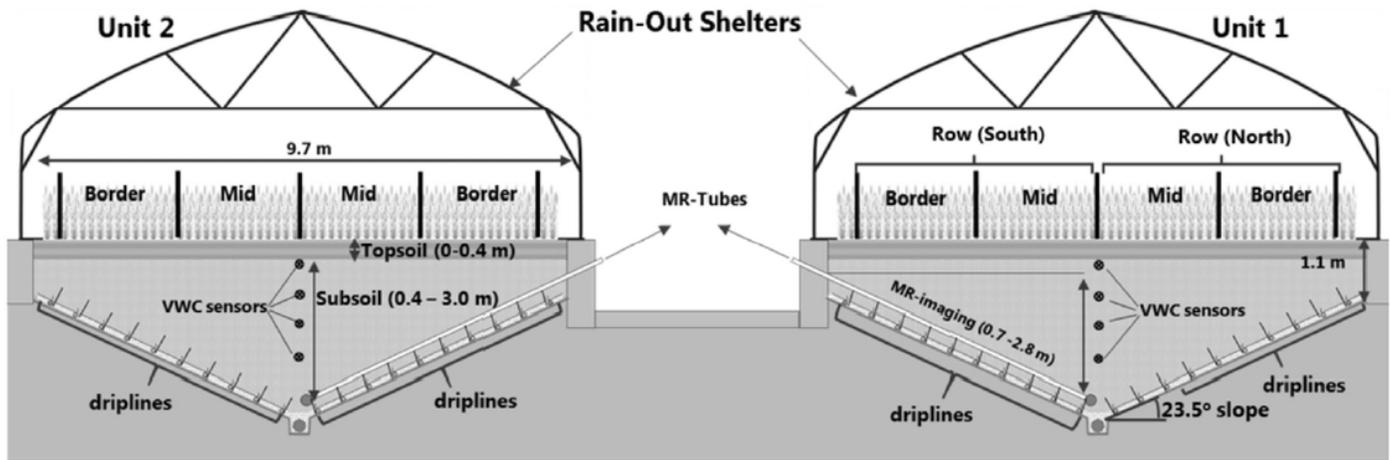


Figure 6

The semi-field facility displayed as a cross-section of two identical units [6, 26]

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1.xlsx](#)
- [Methods.pdf](#)