УДК 633.11:577.21

IMPLEMENTATION OF GENOME-WIDE SELECTION IN WHEAT

© 2012 г. **G. Charmet, E. Storlie**

INRA-UMP, UMR Genetics, Diversity, Ecophysiology of Cereals,
F63100 Clermont-Ferrand, France, e-mail: gilles.charmet@clermont.inra.fr

Received November 19, 2011. Accepted for publication January 10, 2012.

With the expected development of thousands of molecular markers in most crops, the marker-assisted selection theory has recently shifted from the use of a few markers targeted in QTL regions (or derived from candidate or validated genes) to the use of many more markers covering the whole genome. These genome-wide markers are already used for association analysis between polymorphisms for anonymous markers and qualitative or quantitative traits. The condition for success is that a sufficient level of linkage disequilibrium (LD) exists between the adjacent markers used for genotyping and the true genes or QTLs. This LD is known to vary among species and type of genetic material. In selfing species, particularly among breeding lines, LD has been reported to range up to 1 cM or more. In such conditions, uncharacterized markers can be used to predict the breeding value of a trait without referring to actual QTLs. We present an example applying DArT markers to the INRA wheat breeding material in an attempt to implement whole genome selection as an alternative to phenotypic selection. This study assesses different models (GBLUP, Ridge Regression BLUP, Bayesian Ridge Regression and Lasso) and their ability to predict the yields of genotypes evaluated in a multi-site network from 2000 to 2009 in a highly unbalanced design. The prediction coefficients obtained by cross-validation techniques are encouraging, given the small size of the training population.

Keywords: genomic selection, breeding value, GBLUP, Ridge regression, LASSO.

Introduction

To satisfy the demand of the growing world population, agriculture faces the challenge of delivering safe, high-quality, and health-promoting food and feed in an economical, environmentally sensitive, and sustainable manner while maintaining yield and stability across different environments affected by climate change. Grain cereals – mainly wheat, rice and maize- represent a major renewable resource and are among the most widely grown crop worldwide. Wheat is the most widely grown crop worldwide with an average global annual harvest of 621 million tons of grains. Wheat demand is expected to increase from 621 mt to 760 mt in 2020, to 813 mt in 2030 and more than 900 mt in 2050 (FAO, 2002). This implies annual production growth rate of about 2 %, while it was limited to 0,9 % from 1985 to 1995. Moreover, the rate of yield increase has slowed down from 1995 to 2005 in nearly every country (Complementary strategies ..., 2009; Wheat facts ..., 2009), and it is close to 0 in EU, particularly in the major producing countries like France, Germany and UK.

This yield increase should be achieved by «sustainable intensification». Thus, accelerating genetic progress is recognized as a priority in most countries. Genetic progress per year is given by the general formula:

$\Delta G = i \cdot h^2 \cdot \sigma_P / L$ (where i is selection intensity, h^2 trait heritability, σ_P phenotypic variability and L the duration of selection cycle).

The utilization of markers has been proposed as a means to improve trait heritability and increase selection intensity (by reducing need/cost of phenotyping). Marker assisted selection (MAS) can be used to accelerate and improve the transfer of traits under mono or oligogenic control. For example, MAS can facilitate the transfer from an unadapted source into an elite genetic background through recurrent backcrosses. It has been demonstrated that the use of markers tightly linked to the causal gene can avoid costly and time consuming phenotyping, while genome wide markers enable

to recover most of the elite background in two or three backcrosses instead of 6–7 while limiting linkage drag. In the case of complex traits such as yield, there are likely many genes with quantitative effects (QTLs), which have not all been identified. Therefore, the former strategy of marker assisted transfer cannot be applied, and is generally replaced by marker assisted recurrent selection, whose objective is to increase the frequency of favourable alleles at most QTL. By this way, the probability of identifying lines which cumulate favourable alleles is also increased.

In their pioneering work, Lande and Thompson (1990) proposed an extension of the index selection theory by adding a molecular score to the classical phenotypic score. They introduced the theory for optimizing weights given to each component and demonstrated that this index is in any case at least as efficient as the phenotypic score alone. Note that this approach of marker assisted recurrent selection used only markers which have been identified as being significantly associated (linked) to QTL. The efficiency of MAS/phenotype selection is higher when the trait has a low heritability, the population size is large and the detected QTLs explain a large proportion of the trait variation. Thus further studies have shown that efficiency is improved when including QTLs with small effects, even if they are false positives, rather than being too stringent during the QTL detection step (Moreau *et al*., 1998; Bernardo, 2006). This combined index theory has been adapted, particularly by removing the phenotypic component. Hospital *et al*. (1997) showed that the use of marker index only allows early selection, without trait evaluation, thereby shortening selection cycles and accelerating genetic gain per cycle. However, after several cycles of selection, some favorable alleles may become fixed, and recombination will decrease linkage disequilibrium between QTLs and markers. It is then necessary to regularly re-estimate the associations between QTLs and markers and their effects on the trait (Gimelfarb, Lande, 1994). The interest of marker assisted selection for quantitative traits has been experimentally demonstrated (e.g. Eathington *et al*., 2007; Blanc *et al*., 2008) and they are currently used in routine by most large plant breeding companies.

However, the efficiency of these marker assisted selection methods can still be limited by the first step of QTL detection, whose power can be low for QTL with small effects in breeding populations of limited sizes. For complex traits, like grain yield, the most likely hypothesis is that they are controlled by a very high number of genes, most with small effects below the detection threshold. Therefore, a large number of QTL are not accounted for by markers included in the selection index.

A further step was proposed by Whittacker (2000), who suggested including all markers in the selection index, thus skipping the QTL detection step. As the number of markers is generally higher than the number of genotypes, classical multiple regression with markers as fixed effects cannot be used. Therefore Whittacker (2000) suggested using ridge regression models to overcome this overparemetrization problem. This method is based on introducing a penalization parameter, λ, which reduces the space dimensionality. Meuwissen *et al*. (2001) applied ridge regression and several Bayesian approaches to animal populations for predicting breeding values. They proposed the use of genome-wide markers to predict the genetic value of individuals. Therefore, it is appropriate to name these methods «genomic prediction». However, as genomic predictions are intended for selection purposes, the expression «genomic selection» has become common (e.g. Goddard, Hayes, 2007).

The most efficient use of genomic selection is to replace costly and time consuming phenotyping by a prediction of the genetic value of the trait under selection (or any multitrait index). Thus, the main expected advantage is to shorten selection cycles. However, to benefit from shorter cycles, the genetic gain per selection cycle should be close to that expected from phenotypic or combined MAS + phenotypic selection.

The relative efficiency relies on the accuracy of prediction of the true genetic value by the marker score. Abundant theoretical quantitative genetics literature often report the correlation between genomic marker predictions and «true» breeding value or phenotype. The true breeding value is known only in simulated data, in which QTL effects are given to simulated or real markers, and then these effects are summed to obtain the «real» genetic value. In real datasets, the true genetic value is unknown, and it should be remembered that the phenotype is only a predictor of this breeding value, but usually the only available to compare perfor-

mance of marker-based predictors. The quality of a prediction, as measured by this correlation, relies itself on the level of linkage disequilibrium between a QTL and the linked marker. The relevant parameter is the rI, as it was demonstrated that the sample size required to detect a QTL by a nearby marker is $1/r^2$ times the size required if we has tested the QTL itself (Balding *et al*., 2007). The quality of the global prediction of breeding value will depend on the effectiveness of the markers to capture most of the information brought by QTLs. Thus marker density should be high enough, in order that every QTL be in sufficient LD with an adjacent marker. The extent of LD has been extensively studied in animal and plant species, and we should keep in mind that it is a property of each particular genepool, and no generalization is straightforward among germplasm or breeding programmes. For example, this LD range is expected to be large in biparental populations, and Lorenzana *et al*. (2009) obtained reasonably good prediction with as few as 96 markers in simulated maize progenies. But in progenies from more complex mating schemes, the required marker density will be higher (Bernardo, Yu, 2007; Blanc *et al*., 2008; Heffner *et al.*, 2009; Jannink *et al*., 2010). Moreover, the LD pattern changes from one generation to the next, since recombination reduces the range of LD.

For practical applications in breeding programs, one has to estimate marker effects and add them to obtain the genomic estimate of breeding value (GEBV). This estimation requires both genotypic and phenotypic information in a so-called «reference» or «training» population. Then, marker effects can be used to estimate GEBV in a «target» population with only the genotypic information, and, subsequently, selection can be made on the GEBV instead of the phenotypes. Genomic Selection (GS) can be repeated on the progeny of crosses between GEBV-selected individuals and so on. However, as the LD between markers and QTL decreases from one generation to the next, GEBV predictions are less and less accurate. Therefore, new phenotypic measurements are needed to re-estimate marker effects (see Heffner *et al*., 2010).

In this manuscript, we report on some preliminary results about the implementation of genomic prediction of yield in the INRA wheat breeding programme. We used both simulated and real data and discus some of the issues related to genomic prediction for wheat in France. The presented results only deal with the initial prediction of target populations using marker effects estimated from training populations sampled by cross-validation.

Material and Methods

The INRA wheat breeding program is a carried out in three main research units: Clermont-Ferrand, Estrées-Mons and Rennes. Each breeder makes 100–150 crosses every year, using registered varieties (most recently in western Europe) for 50 % of parents and breeding lines from previous cycles of the program for the remaining 50 %. F_2 to F_4 plants are conducted as bulked families with around 2000 plants per cross, then F_5 grains from selected spikes are sown in single rows in a classical pedigree design. Bulked grains of F_6 lines are sown is two replicate trials with randomized 6–10 mI plots in a single location, then the best F_7 in 3–4 replicates, and the most advanced F_8–F_9 lines are evaluated in a network with 4 replicates in 8–10 locations, according to their precocity group. To have a more balanced design, we kept data from 6 locations with the higher number of common genotypes. Therefore, 30–50 most fixed «new» lines enter the most advanced evaluation network each year. Some of them are evaluated only one year, some two or three consecutive years before being presented to official registration for the best ones. As breeding lines are used as genitors only once sufficient phenotypic data are available, i.e. in F_8, the duration of the selection cycle can be estimated to at least 8 years, and more likely 9–10 to take into account the use of registered varieties in crossing schemes. In this study, we used those lines which have been evaluated in the complete multisite network between 2000 and 2009. After discarding some lines with too few data or to many missing markers, this gave a dataset of 318 breeding lines.

DArT markers were provided by Triticarte company (www.triticarte.com.au). After cleaning markers with more than 5 % missing data and minor allele frequency >5 %, we obtained a dataset with 2121 polymorphic markers.

As often reported in the literature, we used Monte-Carlo methods to simulate «true» breeding values to be estimated by GS prediction. For this, a subset of 50, 100 or 250 markers were sampled and given an additive effect drawn from a $\boldsymbol{N}(0,1)$

distribution. Then the sum of the 100 QTL effects was summed for each individual to estimate its «true» breeding value (TBV), and its «realized» genetic variance σ_g^2. Finally a random was generated using a $\boldsymbol{N}$ $(0, \sigma_e^2)$ and added to the TBV to generate a «simulated phenotype» (simP), where σ_e^2 being set as $\sigma_e^2 = \sigma_g^2 \cdot (1 + h^2) / h^2$ to achieve the desired heritability of simP.

For estimating the accuracy of prediction on real data, we focused on yield, whose broad sense heritability in our design was estimated to 0,37. Because of the highly unbalanced design, we first had to correct for other factor and estimate a corrected genetic main effect. This was achieved through the use of mixed models, with environments and blocks within environments as fixed effects and genotypes as random effects, whose variance being modeled by an identity matrix to avoid confusion with further BLUP prediction using marker estimates of additive relationship matrix. Then the BLUP for each of the 318 lines were used as observed phenotypes (obsP).

Several statistical models are being compared for their prediction accuracy as measured by the correlations between GEBV and either TBV, simP or obsP.

Four statistical methods have been used to predict GEBV

- The ridge regression, as described by Whittacker *et al*. (2000) using a home written R programme (R development core team, 2011). Basically, this methods uses a mixed linear model to estimate best linear unbiased predictor (BLUP), assuming that markers have random effects with common variance. RRBLUP uses a penalty parameter, λ^2 in the estimator to shrink marker effects and to avoid over-fitting (Piepho, 2009). In this study, $\lambda^2 = \sigma_e^2 / \sigma_g^2$, where σ_e^2 is the residual variance and σ_g^2 is the marker effect variance – estimated from the additive genetic variance divided by the number of markers.
- The GBLUP (Coster, 2010), using the pedigree library of R. The XX function solve the classical BLUP equation (Henderson, 1975), using a marker-based estimate of the additive relationship matrix.
- Bayesian Ridge Regression and LASSO (De los Campos, Pérez, 2010; Pérez *et al.*, 2010) as implemented in the BRR library of R.

Results

Figure 1 shows the correlations between GEBV and either TBV or simP for the 4 prediction methods on simulated data with 100 QTL at 3 trait heritabilities.

Similar to other publications, the prediction accuracy increases with simulated trait heritabilities and the correlations with TBV are all higher than that with simP at a given heritability. It should be remembered that the correlation with phenotype cannot exceed h, the square-root of trait heritability, which is verified in Fig. 1. Whatever the trait heritability and the measure of accuracy, the 4 methods rank in similar order, the G-BLUP being the least efficient and the Bayesian approaches the most, particularly LASSO.

The accuracies of the four methods on the BLUP prediction of yield in each of the 6 locations and on the overall BLUP prediction are given in Table and illustrated in Fig. 2.

On this real trait averaged over environments, i.e. the best estimate of the additive main genotype effects, the ranking of the 4 methods is quite similar to that obtained on simulated data. The two Bayesian approaches (RRB and LASSO) clearly outperform the mixed model approaches. However the ridge regression appears to be less accurate than G-BLUP, which was the least efficient on simulated data.

Prediction accuracies of the 4 methods, i.e. correlations between GEBV and obsP obtained in six different locations vary from one location to another, likely according to the within location broad sense heritability. Moreover, in some locations all 4 methods gave similar correlations, while in others there are significant differences among them. More remarkably, the ranking of the 4 methods according to their accuracy differs from the ranking observed on simulated data or even the obsP on all environments. This is particularly true for the G-BLUP method, which is never worst, and it even outperforms the Bayesian methods in two locations. It clearly appears that these BLUP estimates of yield, using single locations, differ from the overall estimate, likely due to GxE interactions.

Discussion

In this preliminary attempt to predict the breeding values of elite wheat lines using genomic

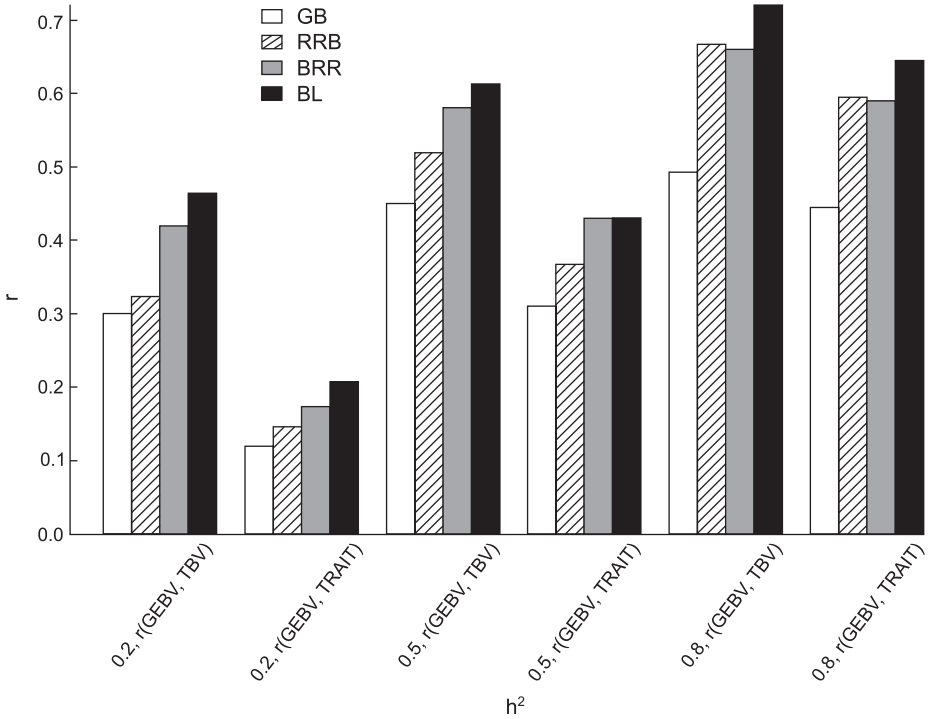

Fig. 1. Correlation between GEBV and simulated true breeding value (TBV) or simulated phenotypes (TRAIT) for 3 heritability values.

RRB: Ridge regresion BLUP, GB: G-BLUP, BRR: Bayesian Ridge regression, BL: Bayesian LASSO. Mean of 100 simulations with 100 QTL with normally distributed effects.

Table

Mean (and standard deviation from 100 cross-validations) of correlations between GEBV estimated from four statistical models and yield predicted in each of the 6 locations and using all locations

Site / Model	cf	di	em	lm	lu	re	All sites
RRB	.289(.14)	.471(.12)	.276(.15)	.332(.12)	.278(.08)	.297(.10)	.488(.12)
GB	.395(.10)	.447(.15)	.330(.12)	.494(.08)	.294(.11)	.344(.11)	.522(.08)
BRR	.329(.12)	.479(.15)	.350(.11)	.492(.08)	.276(.16)	.348(.13)	.506(.11)
BL	.312(.14)	.479(.13)	.333(.12)	.456(.08)	.316(.14)	.324(.12)	.504(.11)

Notes. RRB: Ridge regression BLUP, GB: G-BLUP, BRR: Bayesian Ridge regression, BL: Bayesian LASSO.

markers, results obtained on real data are in accordance with those obtained on simulated traits of similar heritabilities. Indeed the correlation between GEBV and either simulated or observed phenotype is around 0,5. We may assume that the correlation with TBV of real data will also be similar to that obtained on simulated data, i.e. in the range 0,6–0,7. This value is encouraging, and compared to those reported by Crossa *et al*. (2010) who reported accuracy values ranging from 0,355 to 0,608 according to the method and the environment. Heffner *et al*. (2010) recently reported somewhat lower correlation, but they used a more conservative approach, as the used yield in one year as training data and correlate GEBV with yield in another year. If true, an accuracy of 0,6 for TBV is encouraging, since phenotype itself cannot be viewed as a perfect predictor of TBV. Therefore

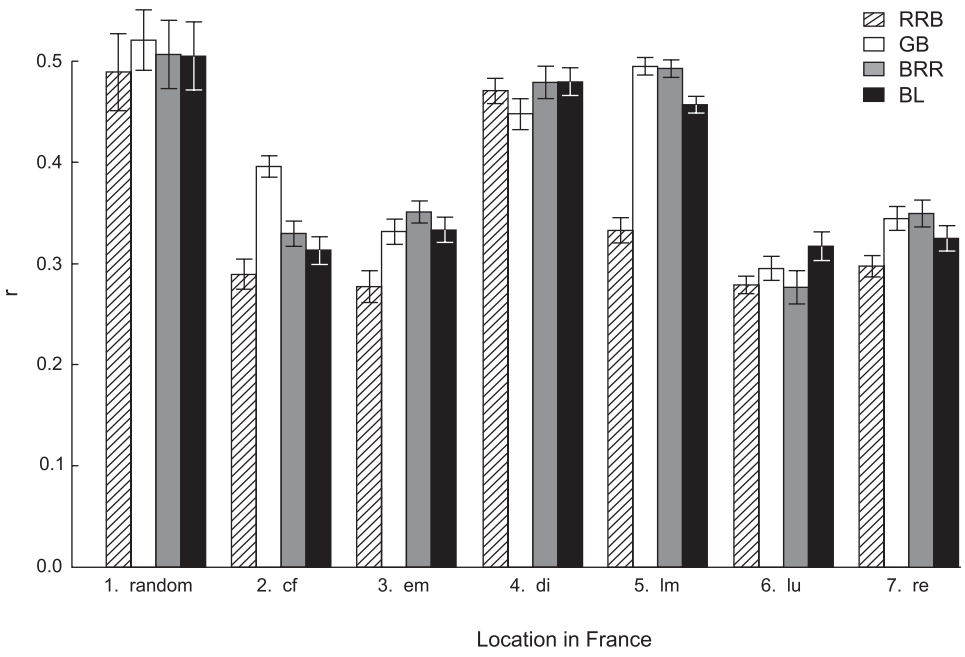

Fig. 2. Mean (and standard deviation from 100 cross-validations) of correlations between GEBV estimated from four statistical models and yield predicted in each of the 6 locations and using all locations.

RRB: Ridge regression BLUP, GB: G-BLUP, BRR: Bayesian Ridge regression, BL: Bayesian LASSO.

selection based on GEBV may not be worse than that based on phenotypes.

However, all studies published so far have failed to obtain very high prediction accuracies. This may be due to the small size of the training population, which is most often lower than 1000. Hayes *et al*. (2009) gave an estimate of the training population required to achieve an accuracy of 0,8, according to trait heritability. For a trait with $h^2 = 0,5$, the theoretical population size is about 5000, nearly twentyfold more than in the present study. Another limitation could come from sparse marker coverage. However, the average marker density achieved with the Dart markers, although unevenly distributed on the genome, seems to be sufficient. This is related to the minimal extent of LD range in the studied material, which itself depends on the number of founder lines and number of generations or the effective population size, as discussed by Heffner *et al*. (2010). As we do not have reliable map positions for every marker, we do not present the pattern of LD in the studied material. However at first glance there are some high values of LD between markers at a few cM apart. Other parameters which affect prediction accuracy have been recently discussed (Zhong *et al*., 2009; Iwata, Jannink, 2011).

The correlation values obtained in this study appear high enough to provide prediction accuracies of TBV of the same magnitude as that provided by replicated phenotypic trials. However, compared to dairy cow, the economic advantage of replacing phenotype prediction by genomic prediction is much less obvious in wheat. Indeed, reliable phenotypic prediction of breeding value of a bull for milk production requires measuring milk production of some or hundreds of its daughters (progeny tests). This requires at least 5–6 years, and the cost is estimated to be around 40 000 € per bull (D. Boichard, pers. comm.). In wheat, 4-time replicated plots in 8–10 locations are usually considered enough to get reliable estimates of mean breeding values of a breeding line, which costs a few hundred euros. Thus the main interest of GS in wheat is shortening selection cycles to accelerate genetic gain. This should only be achieved if fast pure line fixation methods are implemented. This could be accelerated using single seed descent with off season generation in different environments (such as the shuttle breeding used in CIMMYT's programmes), or under controlled conditions using doubled haploid methods, which allow the production of and intermating of GS-selected pure

lines in only 2–3 years instead of 7–10 in classical pedigree selection.

In the framework of the French National Breedwheat programme, a fair comparison of one cycle of phenotypic selection vs two cycles of GEBV-based selection will be carried out on about 1000 DH lines from 34 breeders' crosses over a 6-year period.

References

Balding D.J., Bishop M., Cannings C. Handbook of Statistical Genetics. Chichester, UK: John Wiley and Sons Eds, 2007. V. 2. P. 919–921.

Bernardo R., Moreau L., Charcosset A. Number and fitness of selected individuals in marker-assisted and phenotypic recurrent selection // Crop Sci. 2006. V. 46. P. 1972–1980.

Bernardo R., Yu J.M. Prospects for genomewide selection for quantitative traits in maize // Crop Sci. 2007. V. 47. P. 1082–1090.

Blanc G., Charcosset A., Veyrieras J.B. *et al*. Marker-assisted selection efficiency in multiple connected populations: a simulation study based on the results of a QTL detection experiment in maize // Euphytica. 2008. V. 161. P. 71–84.

Complementary strategies to raise wheat yield potential // Proc. Symp. Complementary strategies to raise wheat yield potential, CIMMYT headquarters / Eds M. Reynolds, D. Eaton. 10–13 November 2009. Mexico: CIMMYT, 2009. 33 p.

Coster A. Package 'pedigree'. 2010. Available at http://cran.r-project.org/web/packages/pedigree/pedigree.pdf

Crossa J., de los Campos G., Pérez P. *et al*. Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers // Genetics. 2010. V. 186. P. 713–724.

De los Campos G., Pérez P. The Bayesian Linear Regression Package V.1.2. 2010. Available at http://cran.r-project.org/web/packages/BLR/index.html.

Eathington S.R., Crosbie T.M., Edwards M.D. *et al*. Molecular markers in a commercial breeding program // Crop Sci. 2007. V. 47. P. 154–163.

FAO. World Agriculture: Towards 2015/2030. Roma: FAO, 2002. 77 p.

Gimelfarb A., Lande R. Simulation of marker assisted selection in hybrid populations // Genet. Res. 1994. V. 63. P. 39–47.

Goddard M.E., Hayes B.J. Genomic selection // J. Anim. Breed. Genet. 2007. V. 124. P. 323–330.

Hayes B.J., Bowman P.J., Chamberlain A.J., Goddard M.E. Genomic selection in dairy cattle: Progress and challenges: Invited review // J. Dairy Sci. 2009. V. 92. V. 433–443.

Heffner E.L., Sorrells M.E., Jannink J.L. Genomic selection for crop improvement // Crop Sci. 2009. V. 49. P. 1–12.

Heffner E.L., Lorenz A.J., Jannink J.-L. Sorrells M.E. Plant breeding with genomic selection: gain per unit time and cost // Crop Sci. 2010. V. 50. P. 1681–1690.

Henderson C.R. Best linear unbiased estimation and prediction under a selection model // Biometrics. 1975. V. 31. P. 423–444.

Hospital F., Moreau L., Lacoudre F. *et al*. More on the efficiency of marker-assisted selection // Theor. Appl. Genet. 1997. V. 95. P. 1181–1189.

Iwata H., Jannink J.L. Accuracy of genomic selection prediction in barley breeding programs: a simulation study based on the real single nucleotide polymorphism data of barley breeding lines // Crop Sci. 2011. V. 51. P. 1915–1927.

Jannink J.L., Lorenz A.J., Iwata H. Genomic selection in plant breeding: from theory to practice // Brief. Funct. Genom. Proteom. 2010. V. 9. P. 166–177.

Lande R., Thompson R. Efficiency of marker-assisted selection in the improvement of quantitative traits // Genetics. 1990. V. 124. P. 743–756.

Lorenzana R.E., Bernardo R. Accuracy of genotypic value predictions for marker-based selection in biparental plant populations // Theor. Appl. Genet. 2009. V. 120. P. 151–161.

Meuwissen T.H.E., Hayes B., Goddard M.E. Prediction of total genetic value using genome-wide dense marker maps // Genetics. 2001. V. 157. P. 1819–1829.

Moreau L., Charcosset A., Hospital F., Gallais A. Marker-assisted selection efficiency in populations of finite size // Genetics. 1998. V. 148. P. 1353–1365.

Pérez P., de los Campos G., Crossa J., Gianola D. Genomic-enabled prediction based on molecular markers and pedigree using the BLR package in R // Plant Genome. 2010. V. 3. P. 106–116.

Piepho H.P. Ridge regression and extensions for genomewide selection in maize // Crop Sci. 2009. V. 49. P. 1165–1175.

R Development Core Team. R: a language and environment for statistical computing. Vienna: R foundation for statistical computing, 2011. Available at http://www.R-project.org.

Wheat Facts and Futures / Eds J. Dixon, H.J. Braun, P. Kosina, J. Crouch. Mexico, D.F.: CIMMYT, 2009. 95 p.

Whittaker J.C., Thompson R., Denham M.C. Marker-assisted selection using ridge regression // Genet. Res. 2000. V. 75. P. 249–252.

Zhong S.Q., Dekkers J.C.M., Fernando R.L., Janink J.L. Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: A barley case study // Genetics. 2009. V. 182. P. 355–364.

ПРИМЕНЕНИЕ МЕТОДА ГЕНОМНОЙ СЕЛЕКЦИИ НА ПШЕНИЦЕ

Ж. Шармэ, Э. Сторли

Национальный институт сельскохозяйственных исследований (INRA) в Клермон-Ферране, Клермон-Ферран, Франция, e-mail: gilles.charmet@clermont.inra.fr

Ввиду ожидаемой разработки тысяч молекулярных маркеров для большинства культур сместились акценты в теории MAS-селекции (маркер-опосредованной селекции) от маркирования определенных QTL (локусов количественных признаков) несколькими маркерами в сторону так называемой геномной селекции с помощью большого числа маркеров, покрывающих весь геном. Наборы маркеров, покрывающие геном, уже используются для анализа ассоциаций между полиморфизмами по маркерам и признаками (качественными или количественными). При этом обязательным является условие, чтобы ген (или QTL) находился в достаточном неравновесии по сцеплению (LD) с прилегающими к нему маркерами, используемыми для генотипирования. Величина LD варьирует от вида к виду и зависит от типа генетического материала. Так, сообщалось, что при анализе самоопыляющихся видов (особенно селекционных линий таких видов) величина LD составляет до 1 cM и более. При таких условиях для предсказания селекционной ценности признака можно использовать маркеры, не прибегая к анализу локусов количественных признаков. Используя DArT-маркеры на селекционном материале INRA, мы демонстрируем пример применения метода геномной селекции в качестве альтернативы традиционному подходу, основанному на фенотипической оценке. В исследовании проводится оценка возможности использования различных моделей («GBLUP», «Ridge Regression BLUP», «Bayesian Ridge Regression» и «Lasso») для предсказания урожайности генотипов, оцененных в широкой сети испытательных участков с 2000 по 2009 гг. С учетом небольшого размера обучающей популяции в ходе перекрестной проверки получены удовлетворительные предсказательные коэффициенты.

Ключевые слова: геномная селекция, селекционная ценность, метод «GBLUP», метод «Ridge Regression», метод «LASSO».