Asian Jr. of Microbiol. Biotech. Env. Sc. Vol. 24, No. (1) : 2022 : 90-97
© Global Science Publications
ISSN-0972-3005

http://doi.org/10.53550/AJMBES.2022.v24i01.017

EVALUATION OF THE EFFICIENCY OF GENOMIC SELECTION APPROACH FOR PREDICTING SHEATH BLIGHT RESISTANCE IN RICE (*ORYZA SATIVA* L.) USING BAYESIAN MODELS

MAHANTESH[1]*, K. GANESAMURTHY[2], SAYAN DAS[3], R. SARASWATHI[4], C. GOPALAKRISHNAN[5] AND R. GNANAM[6]

[1]Centre for Plant Breeding and Genetics, Tamil Nadu Agricultural University,
Coimbatore 641 003, Tamil Nadu, India 641003
[2] Department of Rice, Centre for Plant Breeding and Genetics, Tamil Nadu Agricultural University,
Coimbatore, Tamil Nadu, India
[3]Molecular Breeding Lead, Pioneer Hi-Bred Private Limited, Tunkikalsa Village, Medak 502 336,
Telangana, India
[4] Department of Rice, Centre for Plant Breeding and Genetics, Tamil Nadu Agricultural University,
Coimbatore 641 003, Tamil Nadu, India
[5]Department of Plant Pathology, Centre for Plant Protection studies, Tamil Nadu Agricultural University,
Coimbatore 641 003, Tamil Nadu, India
[6]Department of Plant Molecular Biology & Bioinformatics, CPMB, Tamil Nadu Agricultural University,
Coimbatore 641 003, Tamil Nadu, India

(Received 10 November, 2021; Accepted 23 December, 2021)

Key words *: Rice, Sheath blight, SNPs, Genomic selection, Bayesian*

Abstract–Rice sheath blight (ShB) is one of the most serious fungal diseases caused by *Rhizoctonia solani*, instigating significant yield losses in many rice-growing regions of the world. Intensive studies indicated that resistance for sheath blight is controlled possibly by polygenes. Because of complex inheritance, it's very difficult to exploit and tap all the genomic regions conferring resistance using classical approaches of quantitative trait loci (QTL) mapping, it's very important to have a planned strategy to harness such resistance mechanism. One of the most promising approach is genomic selection (GS). The research was undertaken with an objective to validate genomic selection approach for predicting sheath blight resistance involving 1545 Recombinant inbred lines (RILs) derived from eleven bi-parental populations from crosses between resistant and susceptible parents. Where, Jasmine 85, Tetep and MTU 9992 were resistant parents and TN1, Swarna Sub1, II32B, IR54 and IRBB4 were susceptible parents. During rainy season (2020) the F7 recombinant inbred lines (RIL) were screened for their reaction to sheath blight in two hot spot locations. The genotyping was done with Illumina platform having 6564 SNP markers. Three Bayesian models were used, Bayesian A, Bayesian B and Bayesian CPi to train the statistical model for calculation of marker effects and genomic estimated breeding values (GEBV). The prediction accuracy of training set across models ranged from 0.69 to 0.72, lowest and highest prediction accuracies were observed with Bayesian A and Bayesian B models respectively and the average prediction accuracy of tenfold cross validation with different models was 0.60. Bayesian B model exhibited higher prediction accuracies compared to other models studied. The results are lucrative, all in all, higher prediction accuracies across the models studied suggest genomic selection as a promising breeding strategy for predicting sheath blight resistance in Rice.

INTRODUCTION

Rice (*Oryza sativa* L.) feeds more than half of the world's population and genetic improvement of this food crop can serve as a major component of sustainable food production.

Rice sheath blight (ShB) is one of the most devastating fungal diseases of rice, causing significant yield losses in many rice-growing regions of the world. This disease has become popular recently because of intensification of rice-cropping systems with the development of new short stature, high tillering, high yielding cultivars, high plant densities, and an increase in nitrogen fertilization, these morphological and microenvironment situations are very much congenial for the growth and multiplication of the sheath blight fungus, in India it's prevalence is mainly confined to coastal places of India where farmers grow very high yielding varieties and hot humid climate adds to that. These factors promote disease spread by providing a favorable microclimate for the disease agent due to a dense leaf canopy with an increased leaf-to-leaf and leaf-to-sheath contact (Banniza *et al*., 2007).

The necrotrophic Sheath Blight pathogen possess a broad range of hosts,there are few germplasm lines which are known to show resistant reaction against this pathogen, most of the breeders are focused on harnessing these resistant sources to breed cultivars which are resistant to tolerant for this disease.

Because of lack of authentic and reliable sources of resistance, breeding for sheath blight has been challenging in Rice (Jia *et al*. 2009; Zuo *et al*. 2010; Srinivasachary, *et al*., 2011). Upon intensive study it's believed to be controlled by many genomic regions dispersed across the genome (Zuo *et al*. 2013). It is widely believed that quantitative nature of resistance could be the expedient for evolving varieties with durable/horizontal resistance (Poland *et al*. 2013).

Breeding for disease resistance involving many genes is a challenge. One of popular approach which is very popular now a days which can help in breeding for complextraits is genomic selection, because of reduction in cost involved in genotyping and development of robust statistical models this approach is becoming very popular while breeding for quantitative resistance. Genomic selection uses large number of markers scattered across the genome which are in LD with many genomic regions which are associated with phenotype of interest to generate marker effectsfor all the loci dispersed across the genome (Meuwissen *et al*., 2001). It has been shown to be especially effective for improving quantitative traits, both in simulations (Bernardo and Yu 2007) and in empirical studies (Heslot *et al*., 2013; Lorenz *et al*., 2012; Rutkoski *et al*., 2011, 2012, 2014).

In many studies bi-parental populations with good genetic relationship between the training and test populations have exhibited better prediction accuracies in comparison with three way and complex cross populations, the reason could be controlled population structure and greater linkage disequilibrium between markers and QTLs (Bernardo and Yu, 2007).

For better exploitation of genomic selection approach for predicting sheath blight resistance, it is very important to study the predictability of statistical models used for calculation of marker effects.The current investigation was done with eleven bi-parental populations created by design to validate the efficiency of genomic selection approach to predict sheath blight resistance with three Bayesian models.

MATERIALS AND METHODS

Parent materials and phenotyping of F_7 RILs for ShB

A total of 250 germplasm lines were screened for identification of lines which were resistant and susceptible for Sheath blight. Crosses were made involving Jasmine 85, Tetep & MTU 9992 as resistant to moderately resistant parents and TN1, Swarna Sub1, II32B, IR54 & IRBB4 as susceptible parents. The total of 1545 RILs across eleven populations were used for the study to tap all the genomic regions governing sheath blight resistance dispersed across the genome. The RILs were generated by following single seed descent method (SSD)at Rapid Generation Advancement/ Speed breeding facility of Pioneer Hi-Bred Pvt. Ltd. Research Centre at Tunkikalsa village, Medak district, Telangana. The eleven crosses used for the study were, Jasmine 85XTN1, Jasmine 85XSwarna Sub1, Jasmine 85XII32B, Jasmine 85XIR54, TetepXTN1, TetepXSwarna Sub1, TetepXII32B, TetepXIR54, MTU 9992XTN1, MTU 9992XII32B andMTU 9992XIRBB4. The 1545 RILs derived from these eleven crosses were phenotyped for sheath bight reaction in two hot spot locations (Seethanagaram and Draksharam) of East Godavari District of Andhra Pradesh state, India (lat.16^0 08′ N and long. 81^0 08 E).

The experiments consisting of F_7progenies along with parental lines were planted in Randomized complete design with two replications. Row length of 1.2 meter with row-to-row distance 15 cmand plant to plant distance 10 cm was considered to ensure dense population which is congenial for the development of disease. TN1 was used as susceptible check and was sown after every two rows as well as all along the border to increase the disease pressure asto serve as spreader rows. In the present study, thevirulent local East Godavari isolate of rice sheath blight pathogen was utilized for disease screening. Before the inoculation, the fungus was cultivated in potato dextrose agar medium at optimal temperature for 3–4 days, followed by transferring of disc of medium with mycelia for multiplication. To ensure stringent screening for better disease development, artificial inoculation was done by spraying the mycelia uniformly at the base of plant at maximum tillering stage. The data was recorded at peak milking stage to dough stage by visualizing the relative lesion length to height (%) using 1-9 scale based on development of lesionfrom the lower to upper part of planton a scale from 1 (Resistant) to 9 (Susceptible) thereby getting total of five phenotypic categories, where score 1: 0-20%, score 3: 21-30%, score 5: 31-45%, score 7: 46-65%, score 9: 66-100%.

SNP genotyping

All the RILs used forthe studywere genotyped using Infinium marker platform which is a fixed plexcomprising of 6564 markers, the genotyping was done at marker technology lab of Pioneer Hi-Bred International Limited at Johnston, Iowa State, United States of America.

GS modeling

Genomic selection follows a three-step process (Figure 1). First, all the individuals which are part of training set are genotyped and phenotyped and effects are estimated for all molecular markers, GEBVs (predicted values) were calculated for all the individuals which are part of same training set using the marker effects generated and were correlated with phenotypic values to get prediction accuracy, this is referred as data fit analysis of training set. Second, the training set is validated by considering independent data set, different approaches of cross validation are used to understand predictive ability of training set. Third, members of untested populations are solely genotyped and then selected based on their predicted phenotypes (GEBVs) according to the marker effects estimated in the training set.

For the current investigation, BayesA, BayesB and BayesC Pimodels were used for training the GS model to generate marker effects to calculate GEBV's of the breeding lines under investigation.The statistical analysis was done in "R" program with BGLR package with 50,000 iterations.

The Bayesian models assume a prior marker effects distribution and are of the form:

$$y = 1n\mu + X\hat{a} + \mathring{a}$$

Where X is the incidence matrix for the markers and $\hat{a}$ is the vector of k marker effects.

The prediction accuracy of training set not only depends on statistical model used for calculation of marker effects but also on many factors, viz., phenotypic precision, size of the training set, number of markers used, LD between markers and traits of interest, marker type, heritability of the trait, genetic relationship between training and test set, type of populations present in training set and test set etc.

Tenfold cross validation analysis

To assess the accuracy of the model's ability to predict the untested lines, tenfold cross-validation (CV) simulations were done. The training set comprising of 1545 lines from eleven population with phenotypic and genotypic data was used to

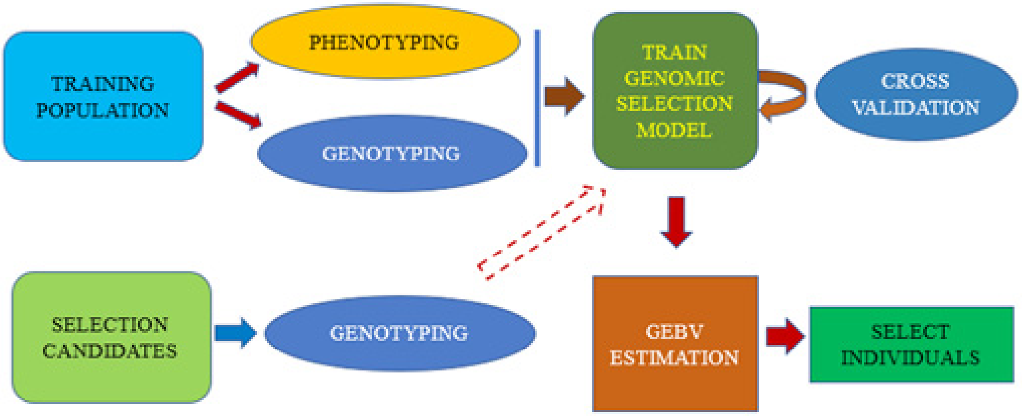

Fig. 1. Showing the different steps of genomic selection (GS) used for crop improvement program

perform repeated tenfold cross validation. The training set was randomly divided into ten portions, the statistical model was trained on nine portions with 1390 lines (training set), the remaining 155 lines (validation set) GEBV was predicted using the marker effects of nine portion's training set, these GEBV of 155 lines were correlated with phenotypic values to know the prediction accuracy. All these steps were repeated ten times to ensure that each portion was once used for prediction of GEBV, finally the obtained accuracies across tenfold were averaged to understand the predictive ability of training set with all the models used for the study.

RESULTS AND DISCUSSION

The frequency distribution of 1545 F7 RILs evaluated showed continuous variation across all population studied (Figure 2). The genotypic analysis was done with large number of markers which were uniformly distributed throughout the genome (Table 1), polymorphic markers between parents across populations studied ranged from 1407 to 2849, MTU 9992XTN1 and MTU 9992XIRBB4 possessed lowest and highest number of informative markers (Table 2).

The principal components analysis results (Figure 3) with marker data clearly indicated that there was ample diversity within as well as across populations present in the training set. Marker effects generated after genotypic analysis (Figure 4, 5 and 6) clearly illustrated that several loci scattered across the genome are contributing to sheath blight resistance which demonstrated that the resistance to sheath blight was governed by many genes with additive effect, this was reported by earlier researchers (Zuo *et al*., 2013). As of now, more than 50 ShB resistance quantitative trait loci (ShBR QTLs) were mapped to all the 12 rice chromosomes (Jia *et al*., 2009; Zuo *et al*., 2010; Xu *et al*., 2011 and Wang *et al*., 2012) because of advancement in genotyping technology and availability of genotypic information at cheaper price but handling several loci through marker assisted selection is near to impossible.

Table 1. Summary of marker data used for analysis and SNPs distribution on each chromosome

Chromosome	SNPs	Length (cM)
Ch1	639	181.8
Ch2	846	162.84
Ch3	598	164.04
Ch4	594	129.6
Ch5	583	128.58
Ch6	577	124.4
Ch7	457	118.6
Ch8	495	121.2
Ch9	427	93
Ch10	324	84.01
Ch11	541	117.9
Ch12	483	109.5
Total	6564	1535.47

The prediction accuracy (data fit) of training set across three models studied (BayesA, BayesB and Bayes CPi) ranged from 0.69 to 0.72, lowest and highest prediction accuracies were observed with BayesA and BayesB models respectively (Figure 7, 8 and 9).

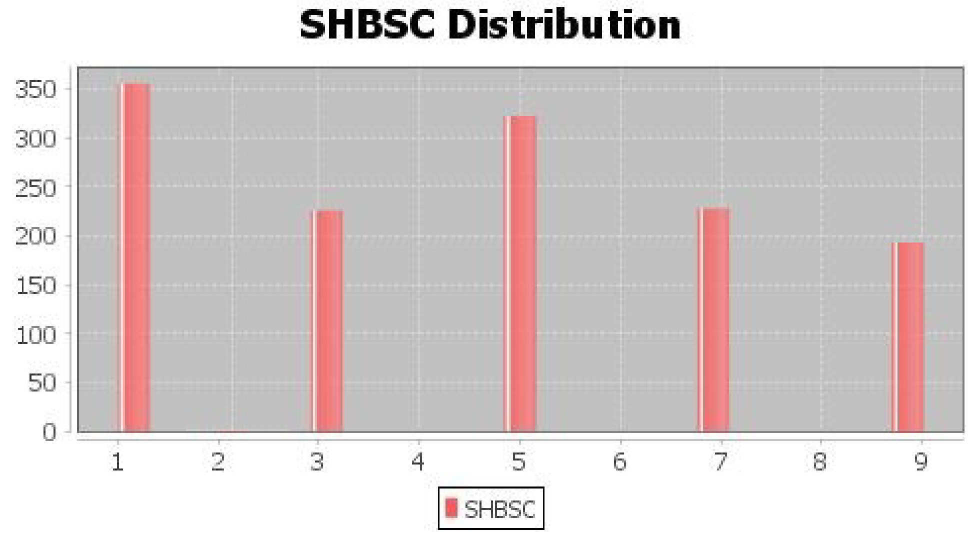

Fig. 2. Distribution of sheath blight phenotypic scores into five classes or categories

Table 2. The informative markers available across the genome for each population used for analysis

Populations	No of RILs	Total Markers	Polymorphic Markers
Jasmine 85/TN1	121	6564	2522
Jasmine 85/Swarna Sub1	139	6564	2627
Jasmine 85/II32B	144	6564	2586
Jasmine 85/IR54	161	6564	2663
Tetep/TN1	221	6564	2806
Tetep/Swarna Sub1	158	6564	2278
Tetep/II32B	241	6564	2702
Tetep/IR54	94	6564	2796
MTU 9992/TN1	50	6564	1407
MTU 9992/II32B	122	6564	2314
MTU 9992/IRBB4	94	6564	2849
Total	**1545**		

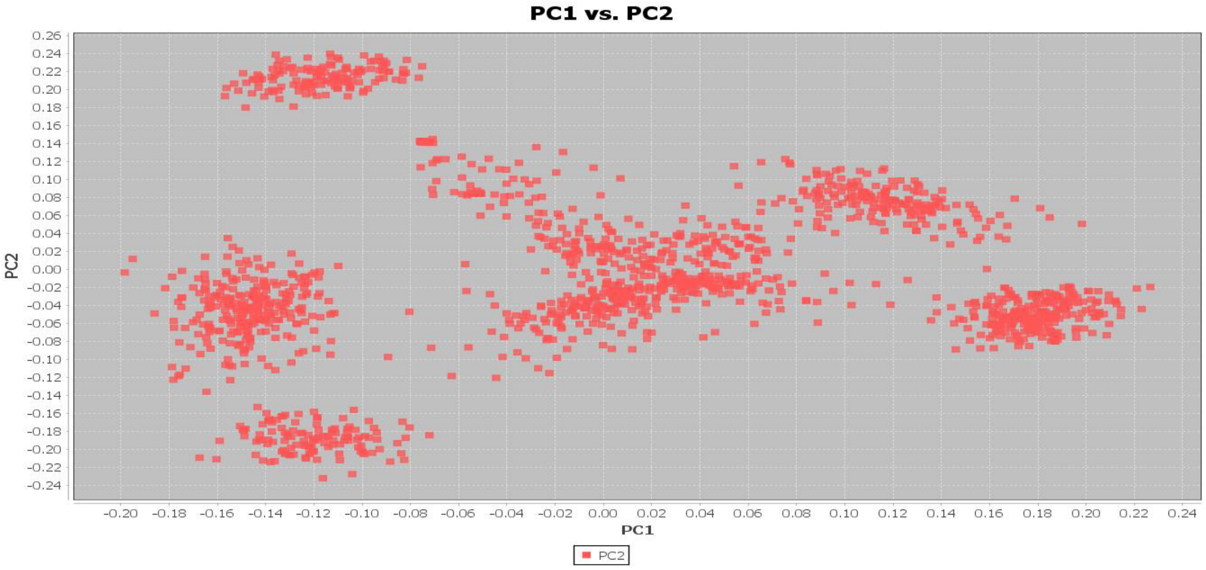

Fig. 3. Depiction of principal components results generated through principal co-ordinate analysis with all eleven populations used for the study

Higher prediction accuracy of the training set in the current study could be attributed to larger size of the training set, higher number of markers, high LD between markers and QTLs, robust statistical models used for calculation of marker effects, greater genetic relationship between populations which were part of training and validation set etc.

The tenfold cross validation prediction accuracy results of BayesA ranged from 0.46 to 0.69, BayesB results ranged from 0.58 to 0.64 and Bayes CPi results ranged from 0.54 to 0.68 (Table 3). The average prediction accuracy across tenfold analysis was 0.60 (BayesA, BayesB and Bayes CPi). The box plot visualization clearly indicates that cross validation prediction accuracy results were consistent with BayesB model across tenfold, other two models exhibited huge range of values across tenfold (Figure 10)and hence BayesB could be

Table 3. Tenfold cross validation analysis results of all five genomic selection models studied

Tenfold	BayesA	BayesB	BayesCPi
1	0.541081	0.581831	0.580503
2	0.667377	0.644163	0.545643
3	0.614166	0.593093	0.581049
4	0.569762	0.606939	0.623068
5	0.612357	0.598611	0.682747
6	0.626874	0.609494	0.543562
7	0.466312	0.604107	0.613744
8	0.697367	0.614845	0.616392
9	0.615836	0.632851	0.623488
10	0.647432	0.590061	0.624802
Average	0.605856	0.6076	0.6035

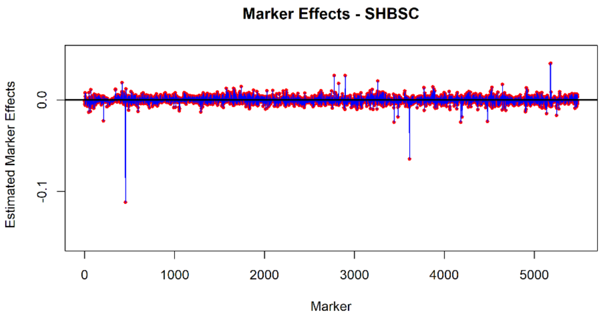

Fig. 4. Marker effects of loci spread across the genome associated with sheath blight generated by marker trait association using BayesA model

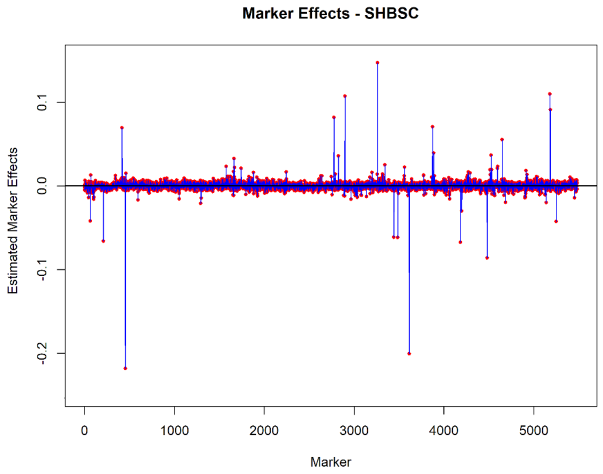

Fig. 5. Marker effects of loci spread across the genome associated with sheath blight generated by marker trait association using BayesB model

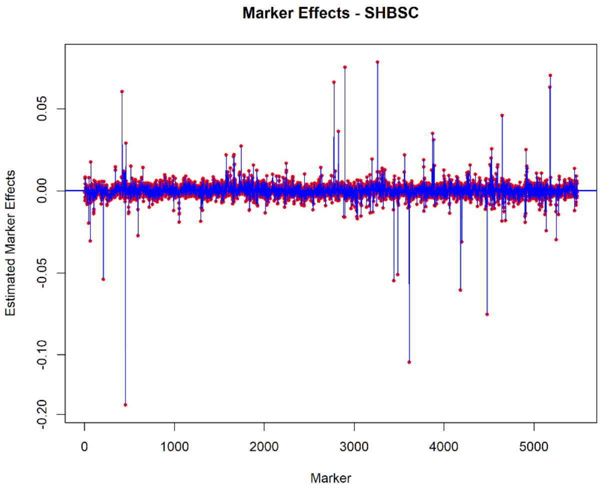

Fig. 6. Marker effects of loci spread across the genome associated with sheath blight generated by marker trait association using BayesCPi model

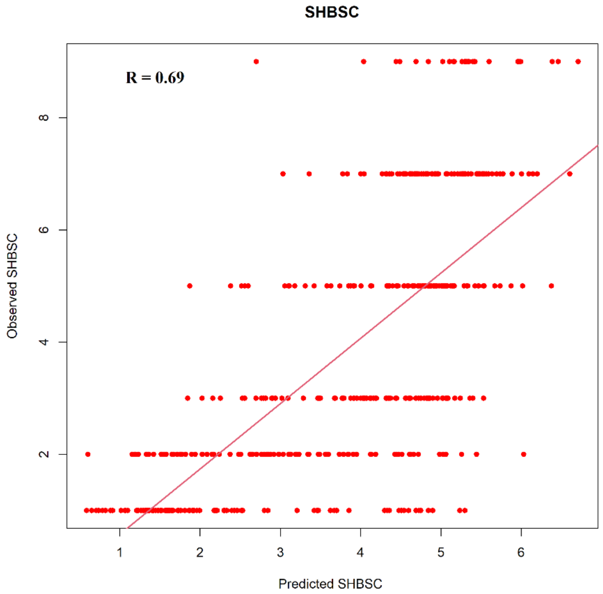

Fig. 7. Prediction accuracy (Data fit) results of BayesA model across eleven populations studied

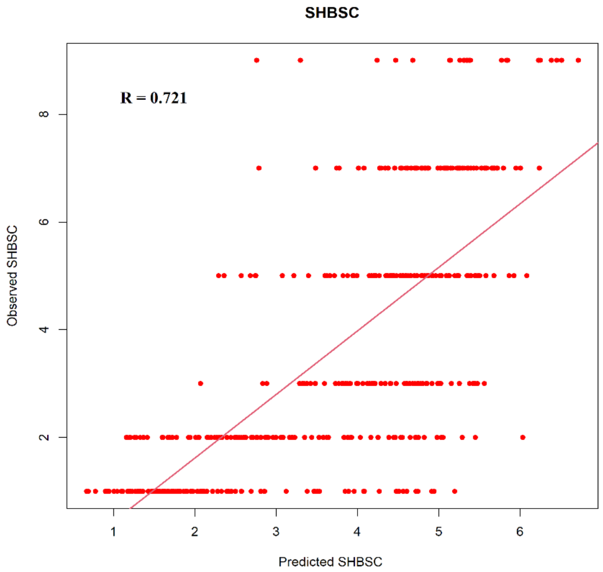

Fig. 8. Prediction accuracy (Data fit) results of BayesB model across eleven populations studied

considered superior over BayesA and BayesCPi models. When large number of markers usedwith good genetic relationship between training and test set, Bayesian models appears to be robust but one of the challenges could be computational power that can be further improved by using advanced statistical models (Heffner *et al.* 2011). As cost involved in genotyping has drastically reduced due to path breaking technologies in biotech industry, genomic selection can be successfully and efficiently

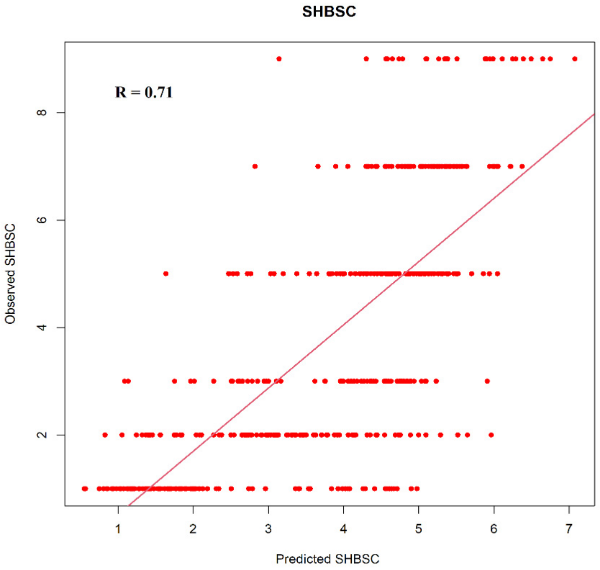

Fig. 9. Prediction accuracy (Data fit) results of BayesCPi model across eleven populations studied

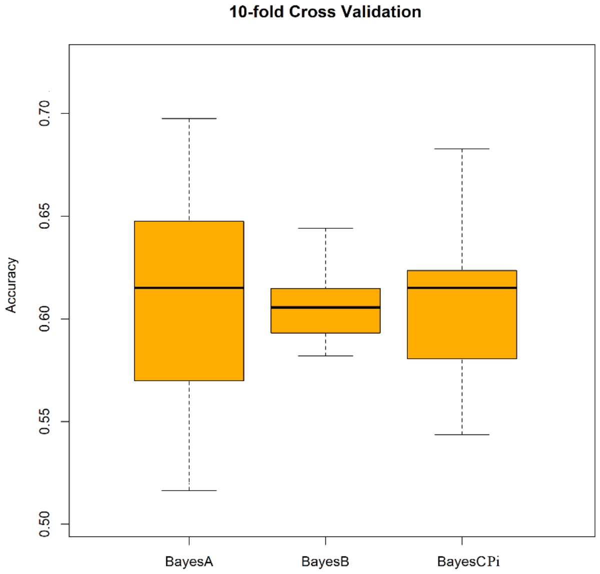

Fig. 10. Box plot of tenfold cross validation analysis results of all three genomic selection models studied

implemented to tackle breeding for sheath blight resistance.

CONCLUSION

From the data fit and tenfold cross validation results it is evident that Bayesian models exhibit high prediction accuracy, genomic selection method can be successfully used for predicting for sheath blight resistance in Rice. Pyramiding all the QTL identified so far into a susceptible varieties is challenging task as resistance is governed by not only several large effect QTLs but also medium to small effect QTLs as well. The inheritance of sheath blight resistance is complex, hence genomic selection could be rewarding for breeding for sheath blight resistance as this approach considers marker effects of all loci dispersed across the genome to provide genomic estimated breeding values which can be used for selection or rejection of breeding lines with resistance to sheath blight.

ACKNOWLEDGEMENTS

I would like to thank Corteva Agriscience (Pioneer Hi-Bred Private Limited, Tunkikalsa Village, Medak, Telangana, India) for providing all the facilities to carry out my research work. I greatly acknowledge my advisory committee members for their continued suggestion, support and guidance.

REFERENCES

Banniza, S., Sy, A. A., Bridge, P.D., Simons, S.A. and Holderness, M. 2007. Characterization of Populations of Rhizoctoniasolani in Paddy Rice Fields in Côte d'Ivoire. Published Online: 22 Feb 2007. *https://doi.org/10.1094/PHYTO.1999.89.5.414.*

Bernardo, R. and Yu, J. 2007. Prospects for Genome wide Selection for Quantitative Traits in Maize. *Crop Science*. 47: 1082-1090.

Jia, Y., Liu, G.J., Costanzo, S., Lee, S.H. and Dai, Y.T. 2009. Current progress on genetic interactions of rice with rice blast and sheath blight fungi. *Frontiers of Agriculture in China.* 3 : 231–239.

Habier, D., Fernando, R.L., Kizilkaya, K. and Garrick, D.J. 2011. Extension of the bayesian alphabet for genomic selection. *BMC Bioinformatics*. 12 : 186.

Heffner, E.L., Jannink, and Sorrells, M.E. 2011a. Genomic selection accuracy using multifamily prediction models in a wheat breeding program. *The Plant Genome.* 4 : 65-75.

Heffner, E.L., J. Jannink, J., Iwata, H., Souza, E. and Sorrells M.E. 2011b. Genomic selection accuracy for grain quality traits in biparental wheat populations. *Crop Science*. 51 : 2597-2606.

Heslot, N., Jannink, J.L. and Sorrells, M.E. 2013b. Using genomic prediction to characterize environments and optimize prediction accuracy in applied breeding data. *Crop Science*. 53(June): 921–933.

Heslot, N. 2012. Genomic selection in plant breeding: a comparison of models. *Crop Science.* 52: 146-160.

Heslot, N., Yang, H.P., Sorrells, M.E. and Jannink, J.L. 2012. Genomic Selection in Plant Breeding: A Comparison of Models. *Crop Science*. 52 : 146–160.

Heslot, N., Rutkoski, J., Poland, J., Jannink, J.L. and Sorrells, M.E. 2013. Impact of Marker Ascertainment Bias on Genomic Selection Accuracy and Estimates of Genetic Diversity. *PLoS One*. 8(9): 74612.

Jia, Y., Liu, G.J., Costanzo, S., Lee, S.H. and Dai, YT. 2009. Current progress on genetic interactions of rice with rice blast and sheath blight fungi. *Frontiers of Agriculture in China*. 3:231–239.

Lorenz, A.J., Smith, K.P. and Jannink, J.L. 2012. Potential and optimization of genomic selection for Fusarium head blight resistance in six-row barley. *Crop Science.* 52 : 1609–1621.

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

Rutkoski, J., Benson, J., Jia, Y., Brown-Guedira, G., Jannink, J.L. and Sorrells, M. 2012. Evaluation of Genomic Prediction Methods for Fusarium Head Blight Resistance in Wheat. *The Plant Genome*. 5 (2): 51.

Rutkoski, J.E., Heffner, E.L. and Sorrells, M.E. 2011. Genomic selection for durable stem rust resistance in wheat. *Euphytica*. 179: 161–173.

Rutkoski, J.E., Poland, J.A., Singh, R.P., Huerta-Espino, J., Bhavani, S. and Barbier, H. 2014. Genomic Selection for Quantitative Adult Plant Stem Rust Resistance in Wheat. *The Plant Genome*. 7(3).

Srinivasachary, L., Willocquet, L. and Savary, S. 2011. Resistance to rice sheath blight (*Rhizoctonia solani* Kuhn) [teleomorph: *Thanatephorus cucumeris* (A.B. Frank) Donk.] disease: Current status and perspectives. *Euphytica.* 178 : 1-22.

Wang, Y., Pinson, S.R.M., Fjellstrom, R.G. and Tabien, R.E. 2012. Phenotypic gain from introgression of two QTL, qSB9-2 and qSB12-1 for rice sheath blight resistance. *Molecular Breeding*. 30 : 293–303.

Xu, Q., Yuan, X.P., Yu, H.Y., Wang, Y., Tang, S.X. and Wei, X.H. 2011. Mapping quantitative trait loci for sheath blight resistance in rice using double haploid population. *Plant Breeding*. 130 : 404–406.

Zuo, S.M., Zhang, Y.F., Chen, Z.X., Chen, X.J. and Pan, X.B. 2010. Current progress on genetics and breeding in resistance to rice sheath blight. *Scientia Sin Vitae*. 40:1014–1023.

Zuo, S.M., Yin, Y.J., Zhang, L., Zhang, Y.F., Chen, Z.X. and Pan, X.B. 2013. Fine mapping of qSB-11LE, the QTL that confers partial resistance on rice sheath blight. *Theoretical and Applied Genetics*. 126 : 1257–1272.