Genomic Selection for Ordinal Stripe Rust Resistance Traits

Lance F. Merrick[1], Xianming Chen[2], Brian P. Ward[3], Arron H. Carter[1]
[1]Department of Crop and Soil Sciences, Washington State University, Pullman, WA 99164, USA
[2]USDA-ARS Wheat Health, Genetics and Quality Research Unit and
Department of Plant Pathology, Washington State University, Pullman, WA 99164, USA
[3]USDA-ARS Plant Science Research Unit, Raleigh, NC 27695, USA

WASHINGTON STATE UNIVERSITY

WSU WINTER WHEAT BREEDING
EST. 1894

ARCS ADVANCING SCIENCE IN AMERICA™

INTRODUCTION

Most genomic selection regression models use linear models that assume continuous and normally distributed phenotypes. Disease resistance, such as stripe rust resistance (caused by *Puccinia striiformis* Westend. f. sp. *tritici* Erikss.), is commonly expressed in ordinal scales and percentages. Disease severity (SEV) and infection type (IT) generally do not follow the assumptions of linear models and have skewed distributions due to high levels of resistance in breeding programs. Stripe rust is a major breeding objective because it can cause more than 90% yield losses in fields planted with susceptible cultivars (Liu et al. 2020).

When faced with data that does not follow the assumptions of linear models, researchers have four options (Montesinos-López et al. 2015). They may either ignore the lack of normality, transform the phenotypes, use generalized linear mixed models (GLMM), or use supervised learning algorithms and classification models with no restriction on the distribution of response variables that are less sensitive when modeling ordinal scores.

OBJECTIVES

Our objectives for this study were: 1) to compare classification and regression genomic selection models for stripe rust disease severity (SEV) and infection type (IT); 2) Inform breeding decisions for the best genomic selection model for stripe rust.

MATERIAL and METHODS

Phenotypic Data

- PNW winter wheat diversity panel lines consisting of 452 lines evaluated for IT and SEV in Pullman, WA in 2013-2015.
- An initial scale for IT was 0-9 and SEV was 0-100% of leaf coverage (Figure 1 and 2).

Genotype Data

- Lines were genotyped using genotyping by sequencing (GBS) through the USDA-ARS Genotyping Laboratory in Raleigh, NC.
- After filtering and imputation, 41,856 SNPs were used.

Genomic Selection Models

- Regression: phenotypic data compared were the unadjusted values, square-root transformed (TF), and adjusted values calculated via generalized linear mixed models (GLMMs) using a poisson distribution.
- rrBLUP was used for all regression models and performed in R.
- Classification: models compared three class types for IT 0-9 classes, 0-2, classes, and a binary 0-1 classes, and classification for SEV.
- Support vector machine with a radial kernel and performed in R.

Validation and Significant Tests

- 5-Fold Cross-Validation.
- Pearson's correlation squared (r^2) was used as accuracy between training and testing data.
- Tukey's HSD test was used to compare models.

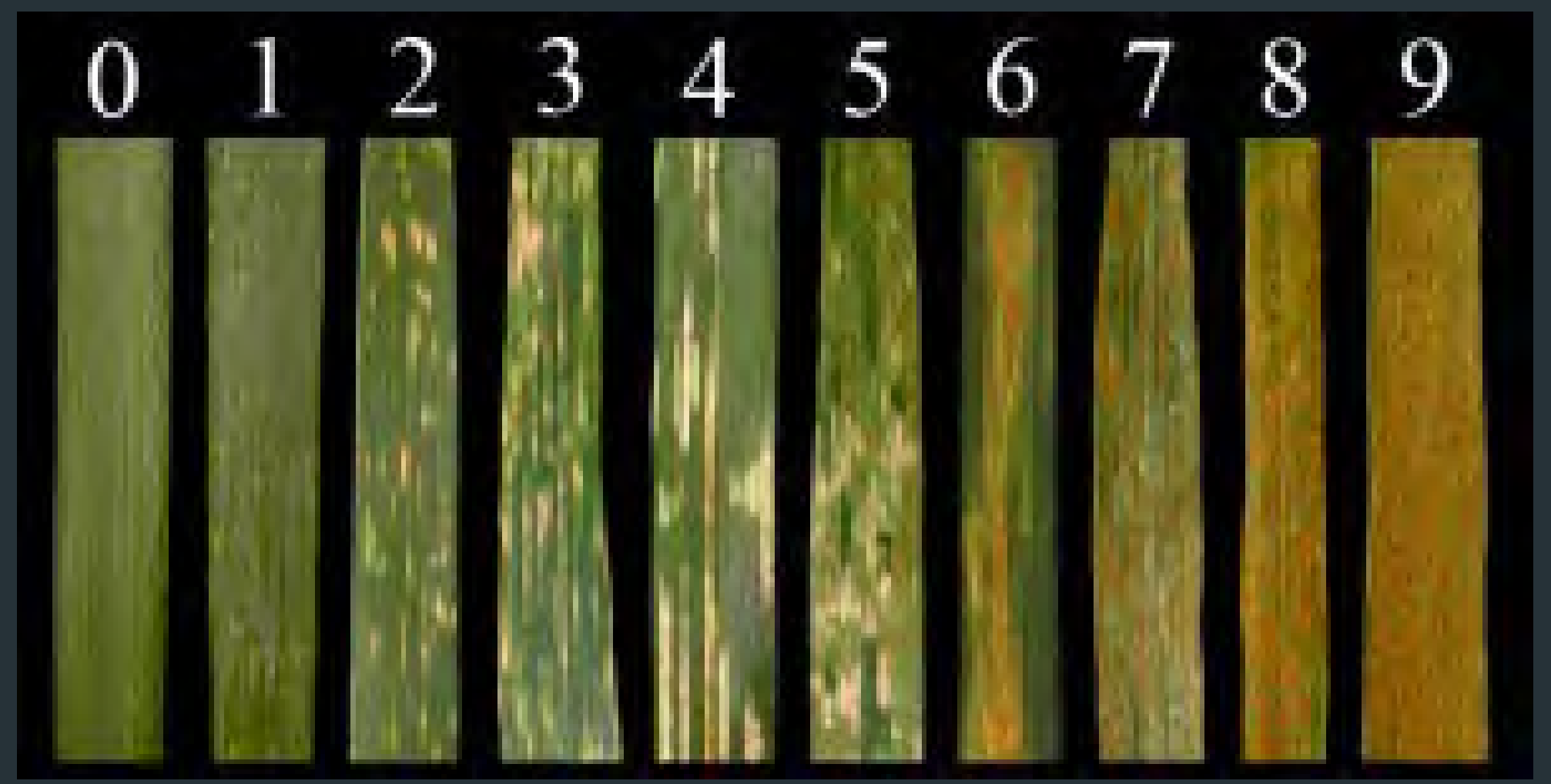


Figure 1. Infection Type: 0-9 Scale.

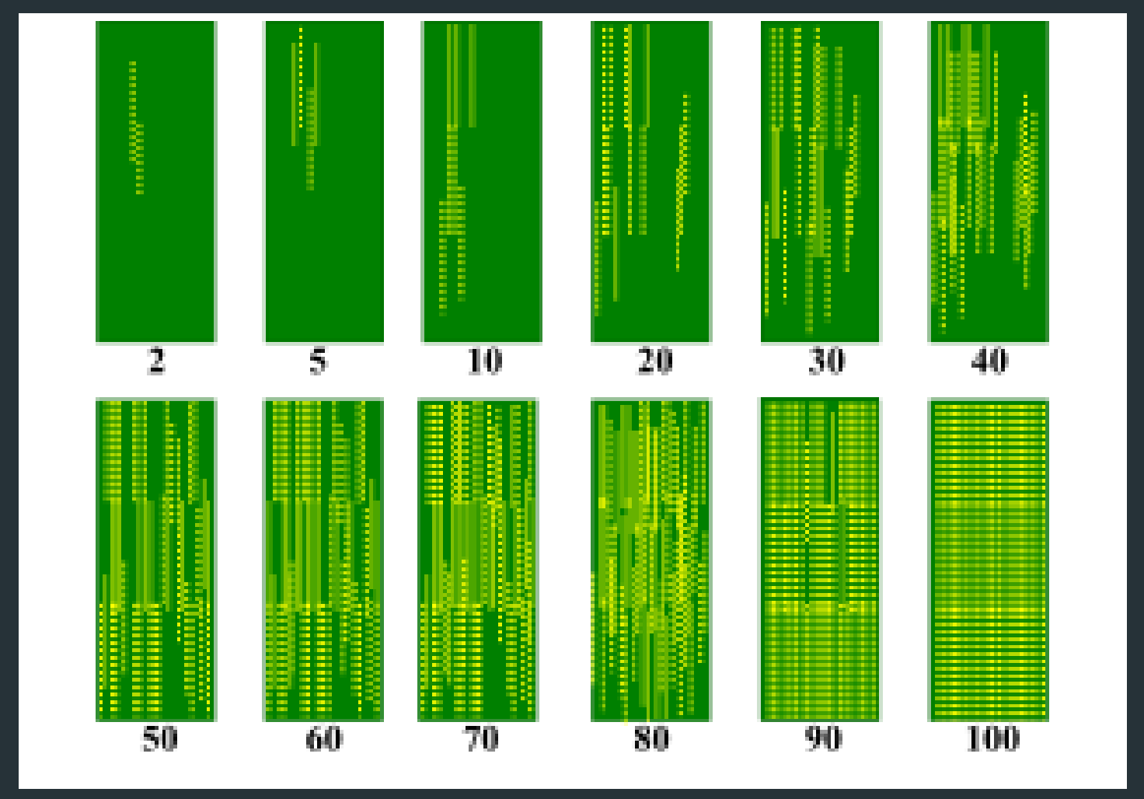


Figure 2. Disease Severity: 0-100%.

Table 1. Comparison of Trait effect on accuracy using Tukey's HSD Test.

Comparison	Variable	Accuracy (r^2)	HSD
Trait	Infection Type	0.44	a
	Disease Severity	0.41	a

- No significant differences between IT and SEV when averaged over all models according to Tukey's HSD pairwise comparisons.

Table 2. Comparison of Model Type on accuracy using Tukey's HSD Test.

Comparison	Variable	Accuracy (r^2)	HSD
Model Type	Regression	0.43	a
	Classification	0.42	a

- No significant differences between regression and classification when averaged over all models according to Tukey's HSD pairwise comparisons.

Table 3. Model comparison using Tukey's HSD Test.

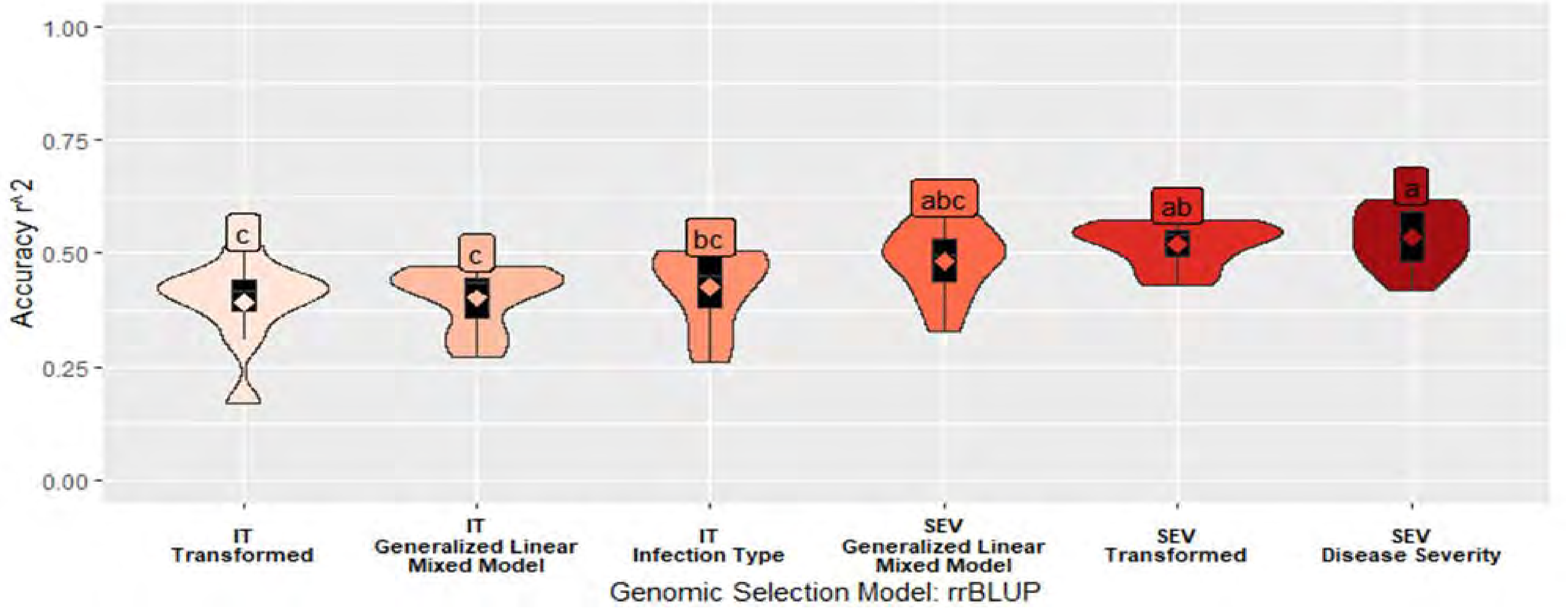
Comparison	Variable	Accuracy (r^2)	HSD
Model	IT 0-1 Classes	0.72	a
	IT 0-2 Classes	0.62	ab
	Disease Severity (SEV)	0.56	bc
	SEV Transformed	0.52	bcd
	SEV Generalized Linear Mixed Model	0.45	cde
	Infection Type (IT)	0.37	def
	IT Generalized Linear Mixed Model	0.35	ef
	IT Transformed	0.33	ef
	IT 0-9 Classes	0.23	fg
	SEV Classification	0.11	g

Breeders can use a binary classification system with very high accuracy to decide whether to keep or discard lines for disease resistance.

The study was funded by the O.A. Vogel Research Foundation and the National Institute of Food and Agriculture (NIFA) of the U.S. Department of Agriculture (Award number 2016-68004-24770) and Hatch project 1014919.

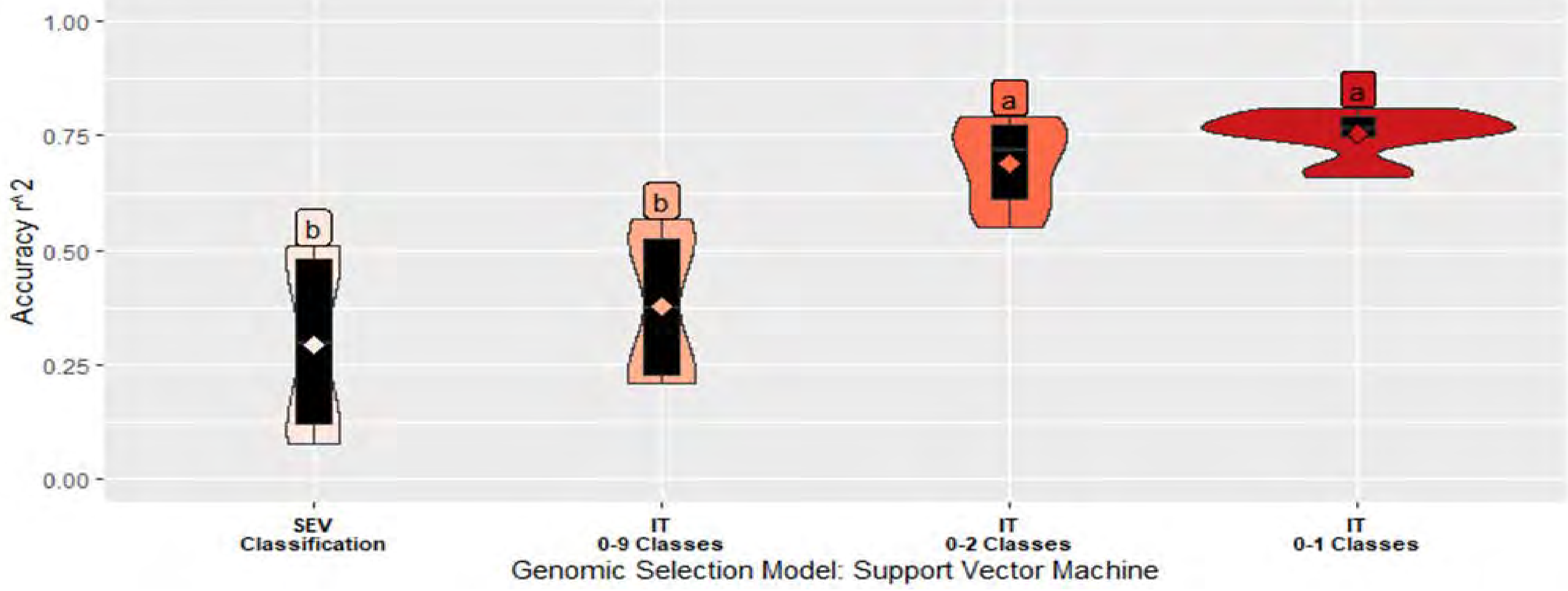
RESULTS

Figure 3. Comparison of regression models dealing with skewed phenotypes by using unadjusted, square-root transformations, and generalized linear mixed models.



- Unadjusted data had the highest mean accuracy for both SEV and IT regression.

Figure 4. Comparison of classification models for dealing with different classification scenarios by using 0-9 classes, 0-2 classes, and 0-1 classes.



- The reduced and binary class resulted in significantly higher accuracy.

DISCUSSION

- There is no significant difference between the average mean for traits, or model type.
- By using a reduced scale or binary classification system, breeders can accurately decide whether to keep or discard lines for disease resistance.
- This study showed the validity of using genomic selection for selecting lines with high stripe rust resistance.

REFERENCES

- Liu, Qie, Li, et al (2020) Genome-Wide Mapping of Quantitative Trait Loci Conferring All-Stage and High-Temperature Adult-Plant Resistance to Stripe Rust in Spring Wheat Landrace PI 181410. IJMS 21:478. https://doi.org/10.3390/ijms21020478
- Montesinos-López, O.A.; Montesinos-López, A.; Pérez-Rodríguez, P.; Eskridge, K.; He, X.; Juliana, P.; Singh, P.; Crossa, J. Genomic Prediction Models for Count Data. JABES 2015, 20, 533–554, doi:10.1007/s13253-015-0223-4.