Evaluation of Genomic Prediction Algorithms for Reducing Selection and Breeding Cycles in Shea Tree (*Vitellaria Paradoxa*)

J. B. Odoi[1,2,4], **H. Prasad**[3], **B. Arfang**[1], **R. Kitiyo**[1], **A. Ozimati**[5], **P. Gibson**[4], **R. Edema**[1,4], **S. Gwali**[2], **T. L. Odong**[1]

[1] Makerere University [2] National Forestry Resources Research Institute (NARO) [3] World Agroforestry (ICRAF) [4] Makerere Regional Centre for Crop Improvement (MaRCCI) [5] National Crop Resources Research Institute (NARO)

*Corresponding author. @ juventineboaz@gmail.com ☎ +256782568822

Abstract. The focus of this study was to determine the genomic prediction (GP) algorithms with the highest prediction accuracies for reducing the breeding and selection cycles in *Vitellaria paradoxa*. The efficiency of the GP algorithms were compared to evaluate five Shea tree growth traits in 708 genotypes with 30734 Single Nucleotide Polymorphic (SNPs) markers, which were reduced to 27063 after removing duplicates. Five hundred forty-nine (77.54%) Shea tree training population and 159 (22.46%) training population were genotyped for 30734 single nucleotide polymorphisms (SNPs) and phenotyped for five Shea tree growth traits. We built a model using phenotype and marker data from a training population by optimizing its genomic prediction accuracy for effectiveness of GS. The phenotype and marker data were used for cross validation of the prediction accuracies of the different models. Prediction accuracies varied among the genomic prediction algorithms based on the five phenotypic traits. We determined the best genomic algorithm that is more suitable for reduction of selection and breeding cycles in *Vitellaria paradoxa*. The GP algorithms were evaluated and we conclude that rrBLUP is the best for improving the prediction accuracy for reducing the breeding cycle in Shea tree.

Keywords: Imputation, genomic selection, phenotypic traits, genotypes, validation.

Introduction

Shea tree (*Vitellaria paradoxa* C.F. Gaertn.subsp. *nilotica* (Kotschy) belongs to the Order Ebenales, family Sapotaceae, genus Vitellaria (Byakagaba *et. al.*, 2011, Makerere University Institute of Environment and Natural Resources. 1998). The specie plays economic and ecological roles to the communities in the savannah parklands of Africa where it has been growing for over 1000 years (Jasaw *et al.,* 2015 and Issaka, 2017). The tree's importance is derived from its products which include: edible fruits, oils, income, cosmetics and chocolate production which are well traded worldwide (Gwali *et. al.* 2012). Since its domestication over 1000 years ago, it has not yet fully been improved for important agronomic traits and this limits

its cultivation as a commercial crop despite of its commercial importance. Conventional Shea breeding efforts have been tried in West Africa with no any such effort in Uganda (Asante *et al.* 2012, Gwali *et al.* 2015, Mawa *et al.* 2016 and Issaka 2017). Application of genomic selection as a new approach of breeding system would go a long way in cutting down the long breeding period that would be taken using conventional means (Koebner and Summers, 2003, Reece & Ejnavarzala, 2006 and Grattapaglia 2014). When clearly identified, molecular markers can be used in Shea breeding programmes to predict the presence of preferred traits important for trees improvement (Chataway, 2005 and Hayes *et al.* 2009). In fact, genomic selection efficiency can be improved with increased genomic prediction accuracies. Studies conducted by Cleverland *et al.* (2013) revealed that the Genomic Estimated Breeding Value (GEBV) accuracies in pigs were comparable to those of dairy cattle with increased size of the training set. Several algorithms/models are now in place for prediction and validation of their prediction accuracies. Choice of the prediction models that yield the highest prediction accuracies of breeding values for traits of interest is important in any tree improvement programme (Calus 2010).

Although (Bernando *et al.* 2007) recommended Bayesian methods in evaluating genomic selection, likewise, quantitative tools such as Best Linear Unbiased Prediction (BLUP) have also successfully been used elsewhere (Silva *et al.* 2000). It is important to note here that different genomic prediction models will have varied prediction accuracies based on their trait genetic architecture. A number of tree species such as Pines, poplars, eucalyptus and apple fruit tree (Resende *et al.* 2012a; Zapata-Valenzuela *et al.* 2012 and Kumar *et al.* 2012b) have been advanced from such methods. The main focus of this study was to reduce the selection and breeding cycles in Shea tree (*V. paradoxa*) for its improvement in Uganda. Furthermore, the study aimed at testing the effectiveness of genomic prediction for improvement of Shea tree for diameter, total height and leaf index traits and identifying the tree's traits significantly important for reducing selection and breeding cycles.

Materials and Methods

Genotyping

Shea Tree DNA extraction was conducted from Biosciences Eastern and Central Africa-International Livestock Research Institute (BeCA-ILRI) Integrated Genotyping Service & Support (IGSS) Unit. Genotyping by sequencing DArTseq (1.0) technology was performed in accordance to Barilli *et al.* (2018), Raman *et al*, (2014) and Kilian *et al*, 2012).

The Shea tree reference genome was accessed from BREED project to align the reads and detect the single nucleotide polymorphism (SNPs) using SNP callers in accordance to Glaubitz *et al.*, (2014) and Wang (2010). SNP calling was performed using TASSEL 5.0. Missing data of 10% were sought and imputed using snpReady software (Granato and Fritsche-Neto, 2018). Search for SNP markers for polymorphic information content, expected heterozygosity, compliance with Hardy Weinberg Equilibrium and number of polymorphic loci done.

A total of 205,781 reads of sequenced data were obtained on analysis and 78% were good barcoded reads. The total alignment rate of the reads to the Shea tree reference genome (Diversity Arrays Technology 2017) was 90%, with 78.5% exactly aligned once to an existing model genome. Initially, the FASTQ files were demultiplexed according to the assigned barcode. The reads from the samples were trimmed and tags identified as follows: a kmer length of 60 bp, minimum quality of 50ng/ul and quantity of 30%. The files were then cleaned using

kdc2_standalone_twoLetter2DArTSnp by converting two letter genotypic data into DArT SNP genotypic data. After imputation, 30734 Single Nucleotide Polymorphic (SNPs) markers were available from 708 genotyped Shea tree seedlings which were reduced to 27063 SNP markers after removing duplicates using the R package DartR. These were later used in the genomic analyses.

Phenotyping

A total of 708 Shea tree lines from five districts in Uganda (Amuru, Arua, Katakwi, Moyo and Otuke) were used in this study. Out of the total number of lines, 549 (77.54% of the whole population) lines with both genotypic and phenotypic data composed the training set and the remaining 159 (22.46%) lines were considered as the breeding set.

All these lines were phenotyped for five Shea tree growth traits which included root collar diameter- measured right above the soil, total height- measured from above the ground to the apex of the highest leaf, leaf length measured from the petiole and the leaf blade, leaf width and petiole length.

Imputation

Imputation was performed and Nipals; PPCA; MissForest; SVD; EM and Knn imputation methods were ran on the dataset with an additional 10% introduced missing values. Missing values were automatically estimated by the software as a linear combo of the k most significant eigengenes. They were iterated until the change in the estimated solution fell below threshold. A collection of imputation methods in the algorithms was scored by a calculated Simple Matching Coefficient (SMC) and the optimal candidate (highest SMC) used to impute the original genotypic dataset. Here, genotypic data were imputed using a dynamically selected best-performing method by testing a number of different imputation algorithms (Nipals, PPCA, MissForest, SVD, EM), calculating a SMC which reflected their performance. The **NIPALS** algorithm (Wold 1973) which had the highest SMC was then used to impute the dataset. The algorithm with Prediction models were developed for each trait and cross validated by means of a 'validation' population, a randomly sampled subset of individuals of the same training population that did not participate in the estimation of marker effects.

Genomic Prediction Models

We used sixteen Genomic Prediction algorithms to evaluate prediction accuracies of five Shea tree growth traits. Prediction models were developed for each trait and cross validated using 'validation' (test) population (Bai and Li 2019). Genomic predictions were performed using the KD Compute platform with the 27063 unique SNPs and the raw phenotypic data.

Data Analysis

Plant Material and Phenotype Data

The Shea tree plants used in this study were derived from the open pollinated seeds in the natural populations of the Shea parklands of eastern, Northern and West Nile regions of Uganda. The training population, used in this study consisted of 180 Half-sib (HS) families composed of 708 individuals. The HS families were raised in a tree nursery at Ngetta Zonal

Agricultural Research and Development Institute (NgeZARDI), Lira district. Their parents were selected from five populations located in the districts of Katakwi, Otuke, Amuru, Moyo and Arua in Uganda. Out of 752 genotyped trees, 708_were selected after filtration for high quality SNPs. The training set was comprised of 549 (77.54% of the whole population) lines with both genotypic and phenotypic data and the remaining 159 (22.46%) lines were considered as the breeding set.

Genotype Data

The DArTseq and SNP platforms images were analysed using DArTsoft v.7.4.7 (DArT P/L, Canberra, Australia) and scored using DArTsoft as binary data (1/0). The DArT software automatically computed the SNP markers (call rate, polymorphic information content (PIC). A Mixed Linear Model (MLM) based method, genomic best linear unbiased prediction (GBLUP), was performed by using the "kinship.BLUP" function of the R package rrBLUP ver. 4.337 with linear ridge kernel regression (RR) and non-linear Gaussian kernel regression (GAUSS) and the Bayesian linear regression models Bayesian Ridge Regression (BRR), Bayesian Lasso (BL), BayesA, BayesB, and BayesC implemented in the R package BGLR ver. 1.0.363. The additive relationship matrix (kinship matrix) was computed by the "A.mat" function of the R package rrBLUP.

We verified the algorithms by assessing their prediction accuracies. A 10-fold cross-validation scheme with 5 replications totalling 50 repetitions and the prediction accuracies were a result of the pooled correlations between the GEBV and the true phenotypic values from each to compare their prediction accuracies. Cross Validation was performed by randomly subletting the marker data by filtering samples that had a known phenotype. This in turn was further separated into 10 parts; one part as the test set and the remaining 9 as the training set for each iteration. Correlation scores were calculated by collapsing results across all folds within a repetition. The prediction accuracies were a result of the pooled correlations between the GEBV and the true phenotypic values from each trait. The genomic prediction accuracies were evaluated using their correlations between their GEBV and phenotypic means (estimated breeding values) as in (Crossa *et al.* 2010). Four prediction models out of the sixteen validated ones provided satisfactory selection accuracies (correlation between the observed and predicted breeding values obtained by cross-validation) useful for calculating the genomic estimated breeding values (GEBV) of the candidates.

The GP Model performed in this current study was: $y = 1_n\mu + Zg + e$.
Where:
y represents vector of observation; 1_n represents $n \times 1$ vectors which had one; μ represents the population mean; Z stands for the design matrix which relates the random variables to genetic values; $g \sim N(0, K\sigma^2 g)$ stands for the genetic values; K represents relatedness matrices; $\sigma^2 g$ represents the genetic variance; $e \sim N(0, I\sigma^2 e)$ stands for the residuals; I stands for identity matrix and $\sigma^2 e$ represents the residual variance. BLUPs was used to perform Cross Validation by randomly subletting the marker data by filtering samples that had a known phenotype.

Results

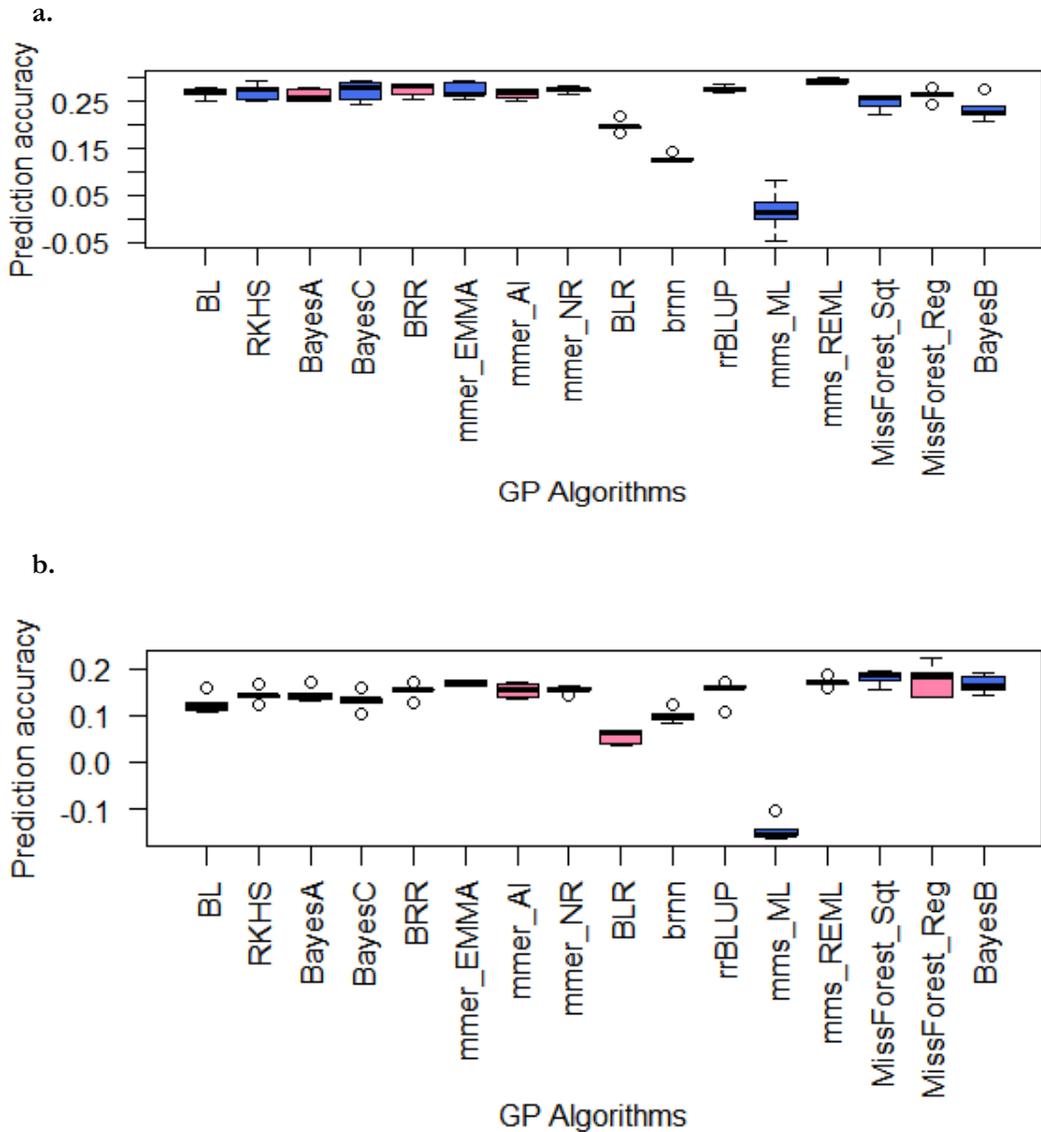### Genomic Selection Algorithms and Prediction Accuracies

The prediction accuracy was high for leaf length and combination of tree diameter and leaf size characteristics and lower in leaf width and petiole length. GBLUP (RR and GAUSS), Ridge

Regression, and Bayesian regressions had higher accuracies for all the traits. The prediction models included the SNP information and the four phenotypic growth traits. MissForest_Sqt (r = 0.251) models proved more reliable for the prediction of height growth trait. On the other hand, the Bayesian Ridge Regression (BRR) model yielded the highest prediction accuracy (r = 0.27), followed by Bayesian Lasso (BL), (r = 0.26) and rrBLUP (r = 0.25), implemented in the R package BGLR ver. 1.0.363. Assessment of the leaf biomass trait resulted in rrBLUP (r = 0.26) and MissForest_Sqt (r = 0.252) with the highest prediction accuracy. It was also observed that rrBLUP (r = 0.297) and mms+REML (r = 0.296) among others were most suitable for leaf size index prediction (Table 1).

Over all, rrBLUP (r = 0.2767), BLR (r = 0.2744), mmer_EMMA (r = 0.2726) and bglr_Bayes (0.2722) were the top four algorithms with highest accuracies (Figure 7). The prediction accuracies were evaluated by 10 folds of the covariance of the parental population partitioned by each fold to verify the models. Five repetitions and 10 folds were applied maintaining the prediction accuracy as Pearson's correlation coefficients (r = 0 at r < 0) between observed and predicted genotypic values. The missing data were imputed using Nipals which had the highest SMC (0.804).

Table 1. Prediction accuracies of genomic prediction algorithms for the Shea tree growth traits.

| Algorithms	Trait prediction accuracies			
	Tree height	*Tree diameter*	*Leaves per tree*	*Leaf size index*
MissForest_Sqt	**0.25**	0.2	**0.252**	**0.27**
BL	0.20	**0.26**	0.2	**0.253**
RKHS	0.179	0.2	0.2	**0.26**
BRR	0.14	**0.27**	0.168	**0.272**
mmer_AI	0.132	0.2	0.2	0.2
mmer_NR	0.131	0.16	0.19	0.075
mmer_Emma	0.13	0.162	0.23	**0.273**
BLR	0.098	0.122	0.179	0.15
MissForest_Reg	0.095	0.214	0.2	0.22
BayesB	0.08	0.216	0.21	0.21
BayesA	0.066	0.2	0.179	**0.25**
BayesC	0.058	0.213	0.25	**0.27**
rrBLUP	-0.88	**0.25**	**0.26**	**0.297**
mms+ML	-0.153	0.004	0.1	0.06
mms+REML	-0.15	0.24	0.23	**0.296**
brnn	-0.01	0.09	0.08	0.1

a.

b.

Figure 1. Comparison of prediction accuracies for the four Shea tree growth traits in a) PC1 and b) PC2 using 16 genomic prediction algorithms. Prediction accuracy was measured as the Pearson's correlation coefficient (r) between predicted genotypic values and phenotypic values.

Choice of Genomic Prediction Algorithm

Choice of an algorithm with high genomic prediction accuracy and precision gives an upper hand to tree improvement over other algorithms with relatively lower accuracy (Xie and Yanchu 2000). Each of the evaluated prediction algorithms indicated varying levels of prediction accuracies with each of the traits in studied (Table 1).

The ANOVA for the differences in the prediction accuracies of the different algorithms was significant (P <0.001); the s.e.d. was 0.01089; Least significant differences (l.s.d.) of means (5% level) 0.02178 confirming their significance.

Figure 2 below clearly show how close the predicted parameters are too close to the middle line for the rrBLUP plot (blue colour) as compared to the rest. It indicates that the predicted values are closer to the observed values in rrBLUP than in other algorithms.
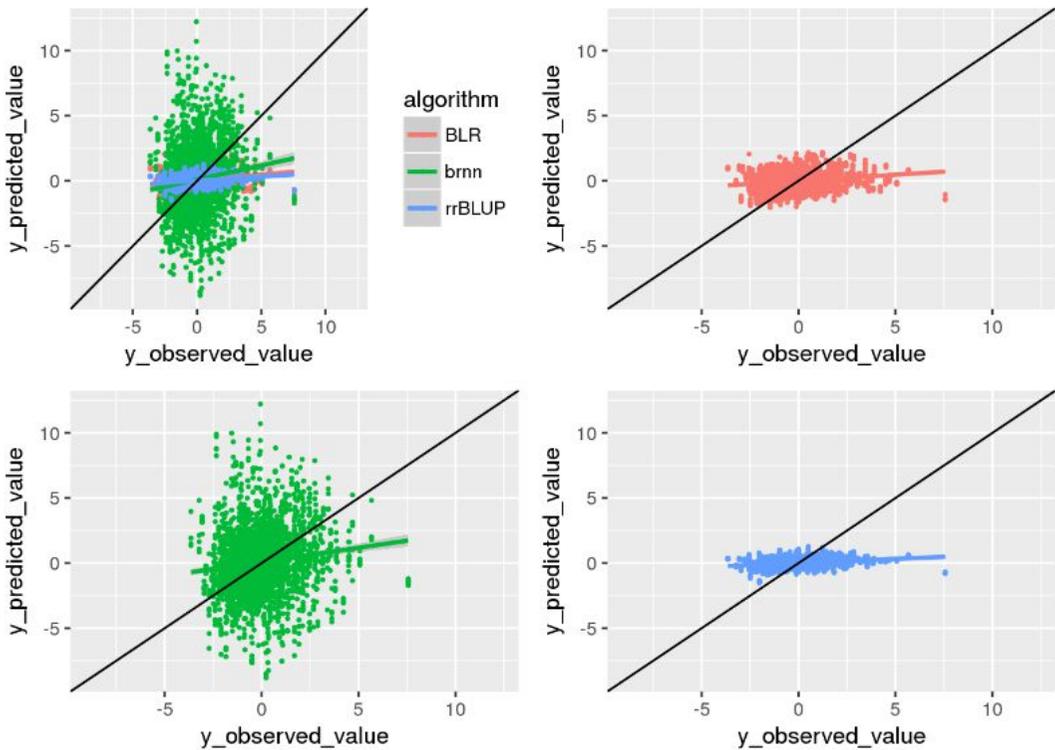
Figure 2. Scatter plot showing overall variability and dispersion of predicted values and observed values of Shea genotypes for fast growth. The colours represent the different algorithms while the dots show distribution of genotype scores.

The heat map reveals how the Shea tree genotypes have clustered according to their faster growth traits. Two clusters were formed revealing two groups/population as those with first growth traits and those with low growth traits (Figure 3).

Tree Traits Significantly Affecting Selection and Breeding Cycles

Three growth traits except "number of leaves" and "leaf petiole length" were significantly correlated to one another (Table 2). Root collar diameter and leaf size were most correlated and formed 54.8% of the total variation in Shea trees growth and maturity in Uganda.

Table 2. Pearson product-moment correlation coefficient of the five Shea tree growth traits with their *p*- values (italicized).

	Root collar	Height	No. leaves	Leaf length	Leaf width	Petiole length
Root collar	1.000	*0.001*	*0.009*	*0.001*	*0.001*	*0.001*
Height	0.305	1.000	*0.010*	*0.001*	*0.001*	*0.001*
No leaves	0.110	0.109	1.000	*0.030*	*0.33**	*0.688**
Leaf length	0.400	0.206	0.092	1.000	*0.001*	*0.001*
Leaf width	0.249	0.156	-0.041*	1.000	1.000	*0.001*
Petiole length	0.305	0.198	-0.017*	0.454	0.278	1.000

Stem diameter (in this case, root collar diameter) and leaf size index indicated moderate correlation (Pearson's r = 0.4, p-value ≤ 0.001) to one another and these are good estimators of tree growth and thus maturity. This is in line with Okao *et al.,* (2019) who found out that broad leaved and larger stemmed Shea tree accessions from Palabek were growing faster (height growth rate registered at annual rate of 125.5 cm) than other accessions planted within the same period and location.
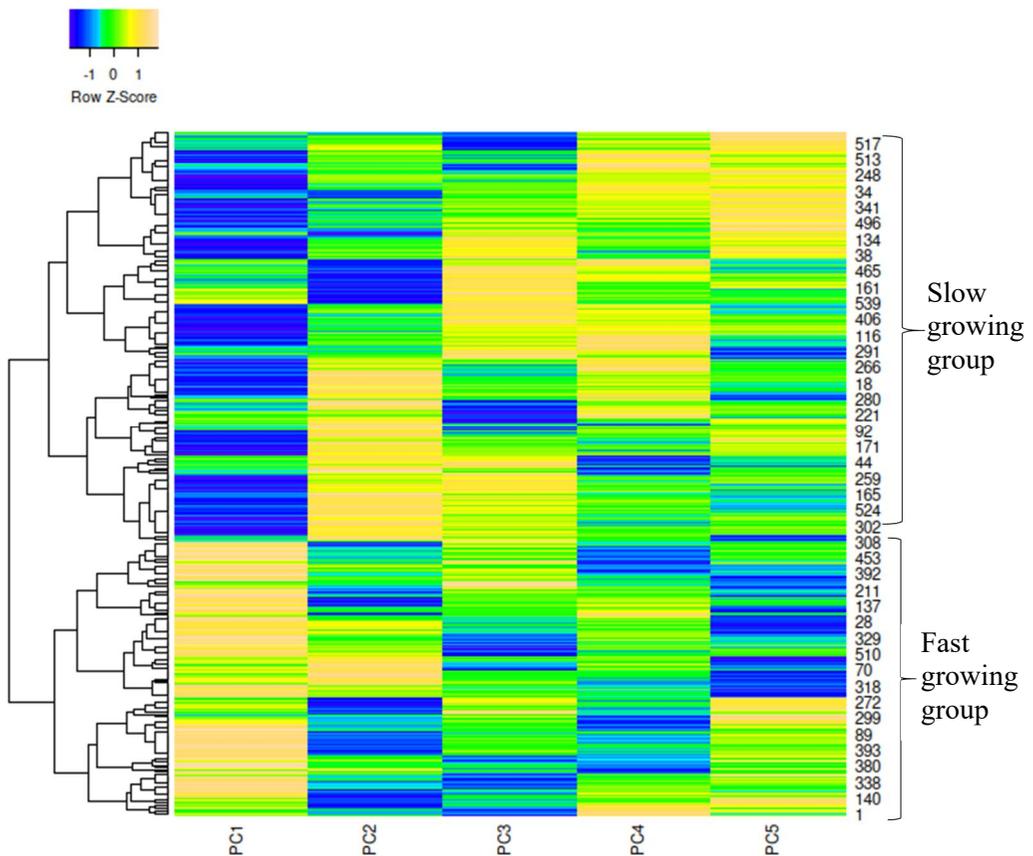
Figure 3. Genetic and phenotypic population structure of the parental population showing heat map of Shea tree growth traits and ward's hierarchical clustering based on Euclidean distance between genotypes. Up to 75% of the total variations in the populations is explained by the first two PC's.

Discussion and Conclusion

Sixteen genomic prediction algorithms were evaluated on the five traits and their prediction accuracies compared. Each prediction algorithms exhibited quite a different accuracy from trait to trait (figure 7). Ridge regression–best linear unbiased prediction (rr-BLUP) proved most accurate for total height with prediction accuracy of 0.277 at a very high precision. The algorithm assumes that traits are controlled by many loci of small effects so that all marker effects are random, normally distributed and with a common variance. Although Grattapaglia (2014) retaliated that rr-BLUP is best used for genomic prediction at start before other models, this study has proved it more accurate and with higher precision for genomic prediction in trees than all the other fifteen algorithms (table 2). rrBLUP treats SNPs as fixed effects and other factors as random effect. In fact, Lorenz *et al.* 2011 recommended rr-BLUP of being very effective and faster in computation time and prediction efficiency. The model works better in animal and plant breeding by including marker interactions which increase its prediction accuracy (Souza *et al.* 2019, Tang *et al.* 2016, Zapata-Valenzuela *et al.* 2013). Actually Hani *et al.* (2018) reported selection methods of genetic algorithm that are more suitable for scheduling problems. On the other hand, models such as rrBLUP that treats all SNPs as random effects have been reported elsewhere as having the highest prediction accuracy in genomic selection in Citrus (Mai *et al.* 2017). The algorithm was reported to capture additive genetic effects which is hereditary. Actually, additive genetic effects are the major contributors of genetic variation in any given population. Therefore, prediction algorithms that takes in account additive genetic effects is the best for genomic selection in perennial species like Shea tree to enable selected traits to be passed on to the offspring even after a long period. This indicates that rrBLUP algorithm accelerates breeding process by reducing the selection and breeding cycles in Shea trees (Bai and Li 2019).

The rrBLUP higher performance in the prediction of tree girth and leaf size index traits clearly confirm the high accuracy and precision in the genomic prediction for early maturity in Shea tree growth. This is true because these traits are important in rapid increase of any plant growth. Actually, bigger tree leaves mean higher rate of manufacturing photosynthates useful for metabolic activities required for faster growth. This has been confirmed in the heat map (Figure 9) which partitioned the population clusters into fast growing and slow growing genotypes.

This study reveals variation in the accuracy of genomic prediction algorithms among the four traits of study. Tree diameter and leaf size index had higher prediction accuracy throughout all the sixteen evaluated algorithms. These are the traits that determine growth and maturity in trees, important for reducing selection and breeding cycles.

References

Asante, W.J., Banidiyia, M.A., Tom-Dery D. (2012). Effect of planting depth on the germination and initial growth and development of Shea (Vitellaria paradoxa C.F. Gaertn.). Int. J. Biosci. (IJB) **2**, 146–152

Bai, J. and Li S. (2019). Genetic Breeding and Molecular Marker-Assisted Selective Breeding of Largemouth Bass, Advances in Agronomy

Barilli, E., Cobos, M.J., Carrillo, E., Kilian, A., Carling, J., and Rubiales D. (2018). A High-Density Integrated DArTseq SNP-Based Genetic Map of Pisum fulvum and Identification of QTLs Controlling Rust Resistance. Front Plant Sci. 9: 167

Bernando, R., Yu, J.M., (2007). Prospects for genomewide selection for quantitative traits in maize. Crop Sci; 47: 1082–1090.

Byakagaba, P., Eilu G., Okullo, J.B.L., Tumwebaze, S.B., Mwavu, E.N. (2011). Population structure and regeneration status of Vitellaria paradoxa (C.F.Gaertn.) under different land management regimes in Uganda. Agric. J. 6 (1):14-22.

Calus, M.P.L. (2010) Genomic breeding value prediction: methods and procedure. Animal4:2, pp. 157–164

Chataway, J. (2005). Introduction: is it possible to create pro-poor agriculture-related biotechnology? J. Int. Dev., 17, 597-610.

Cleveland, M.A., and Hickey J.M. (2013). Practical implementation of cost-effective genomic selection in commercial pig breeding using imputation. J Anim Sci; 91: 3583–3592

Crossa, J., de los C.G., Perez, P., Gianola D. et al (2010). Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. Genetics 186:713–406

Glaubitz, J.C., Casstevens, T.M., Lu, F., Harriman, J., Elshire, R.J., Sun, Q. et al (2014). TASSEL-GBS: a high-capacity genotyping by sequencing analysis pipeline. PLoS One 9, e90346.

Granato, I., and Fritsche-Neto R. (2018). snpReady: Preparing genotypic datasets in order to run 501 genomic; analysis. R package version 0.9.6. Available: https://CRAN.R-502 project.org/package=snpReady.

Grattapaglia, D. (2014). Breeding Forest Trees by Genomic Selection: Current Progress and the Way Forward Researchgate.net pp. 652-678 DOI: 10.1007/978-94-007-7572-5_26

Gwali, S., Vaillant, A., Nakabonge, G., Okullo J.B.L., Eilu, E., Muchugi, A. (2015). Genetic diversity in Shea tree (Vitellaria paradoxa subspecies nilotica) ethno-varieties in Uganda assessed with microsatellite markers. Forests, Trees and Livelihoods. Volume 24, Issue 3 pp 163-175

Gwali, S., Nakabonge, G., Okullo, J.B.L., Eilu, E., Nyeko, P., Vuzi, P. (2012). Morphological variation among Shea tree (Vitellaria paradoxa subsp. nilotica) 'ethno-varieties' in Uganda Genetic. Resources and Crop Evolution **59**, Issue 8, pp 1883–1898

Hani, A., AL-Rawashdeh, Riad A., Ahmad and Ahmad O. Hasan (2018). The Efficient Selection Methods of Genetic Algorithm used in Scheduling Problems, International Journal of Development Research, 8, (10), 23213-23217.

Hayes, B.J., Bowman, P.J., Chamberlain, A.J., Goddard M.E. (2009). Invited review: genomic selection in dairy cattle: progress and challenges. J Dairy Sci 92:433–443

Issaka, A., Konstantin, V.K. and Reiner F. (2017). Morphological and genetic diversity of Shea tree (*Vitellaria paradoxa*) in the savannah regions of Ghana. Genetic Resources and Crop Evolution Volume 64, Issue 6, pp 1253–1268

Jasaw, G.S., Osamu, S., and Takeuchi, K. (2015). Shea (Vitellaria paradoxa) Butter Production and Resource Use by Urban and Rural Processors in Northern Ghana. **Sustainability** ISSN 2071-1050 www.mdpi.com/journal/sustainability

Kilian, A., Wenzl, P., Huttner, E., Carling, J., Xia, L., Blois, H. et al (2012). Diversity arrays technology: a generic genome profiling technology on open platforms. Methods Mol. Biol. 888, 67–89. 10.1007/978-1-61779-870-2_5 [PubMed] [CrossRef] [Google Scholar]

Koebner, R.M.D., and Summers R.W. (2003). 21st century wheat breeding: plot selection or plate detection? Trends in Biotechnology, 21, 59-63.

Kumar, S., Chagne, D., Bink MCAM, Volz R.K. et al (2012b). Genomic Selection for fruit quality traits in apple (Malus domestica Borkh.). PLoS One 7:e36674

Lorenz, A.J., Chao S.M., Asoro, F.G., Heffner, E.L. et al (2011). Genomic Selection in Plant Breeding: Knowledge and Prospects. Adv Agron 110(110):77–123

Mai, F.M., Keisuke, N., Eli, K., Hiromi, K-K., Akio, O., Shingo, G., Terutaka, Y., Atsushi, I., Hiroko, H., Takeshi, H., Satomi, M., Yuichi, K., Atsushi, T., Asao, F., Yasukazu, N., Tokurou, S. and Hiroyoshi I. (2017). Genome-wide association study and genomic prediction in citrus: Potential of genomics-assisted breeding for fruit quality traits. Scientific Reports 7: 4721 DOI: 10.1038/s41598-017-05100-x.

Makerere University Institute of Environment and Natural Resources (1998). Vitellaria paradoxa. The IUCN Red List of Threatened Species 1998: e.T37083A10029534. http://dx.doi.org/10.2305/IUCN.UK.1998.RLTS.T37083A10029534.en.

Raman, H., Raman, R., Kilian, A., Deterring, F., Carling, J., Coombes, N. et al (2014). Genome-wide delineation of natural variation for pod shatter resistance in Brassica napus. PLoS ONE 9:e101673. 10.1371/journal.pone.0101673 [PMC free article] [PubMed] [CrossRef] [Google Scholar]

Reece, J.D. and Ejnavarzala H. (2006). Genes to feed the world: the weakest link? Food Policy, In Press.

Resende, M.D.V., Resende, M.F.R., Sansaloni, C.P., Petroli, C.D. et al (2012a). Genomic selection for growth and wood quality in Eucalyptus: capturing the missing heritability and accelerating breeding for complex traits in forest trees. New Phytol 194:116–128

Silva J.C.E., Wellendorf, H., Borralho, N.M.G. (2000). Prediction of breeding values and expected genetic gains in diameter growth, wood density and spiral grain from parental selection in Picea abies (L.) KARST. Silvae Genetica 49:102–109

Souza, L.M., Francisco, F.R., Gonçalves, P.S., Scaloppi, Jr, Le Guen, E.J., Fritsche-Neto V, Souza A.P. (2019). Genomic Selection in Rubber Tree Breeding: A Comparison of Models and Methods for dealing with G x E bioRxiv

Tang, C., Yang, M., Fang, Y., Luo Gao Y., Xiao S.X et al. (2016). The rubber tree genome 566 reveals new insights into rubber production and species adaptation. Nat. Plants 2, 16073

Wang, L., Li, P., Brutnell, T.P. (2010). Exploring plant transcriptomes using ultra high-throughput sequencing. *Brief. Funct. Genomics, 9*, 118-128.

Wold, H. (1973). Nonlinear iterative partial least squares (NIPALS) modelling: Some current developments. In P. R. Krishnaiah (Ed.), Multivariate Analysis (pp. 383–487). New York: Academic Press.

Xie, C-Y., and Yanchuk A.D. (2000). Breeding values of parental trees, genetic worth of seed orchard seedlots and yields of improved stocks in British Columbia Research Branch B.C. Ministry of Forests

Yang, L., Xu, M., Koo, Y., He, J. and Poethig, R.S. (2013). Sugar promotes vegetative phase change in Arabidopsis thaliana by repressing the expression of MIR156A and MIR156C. eLife 2: e00260.

Zapata-Valenzuela, J., Isik, F., Maltecca, C., Wegrzyn, J. et al. (2012). SNP markers trace familial linkages in a cloned population of Pinus taeda – prospects for genomic selection. Tree Genet Genomes 8:1307–1318

Zapata-Valenzuela, J., Whetten, R.W., Neale, D., Mckeand, S., and Isik, F. (2013). Genomic estimated breeding values using genomic relationship matrices in a cloned population of loblolly pine. G3 (Bethesda) 3, 909-916.