Check for updates

Received: 1 August 2023 | Accepted: 27 November 2023

DOI: 10.1002/csc2.21170

Crop Science

ORIGINAL ARTICLE

Crop Breeding & Genetics

Evaluating breeding for broad versus narrow adaptation for cassava in Nigeria using stochastic simulation

Moshood A. Bakare[1,2] | **Siraj Ismail Kayondo**[2] | **Peter Kulakow**[2] | **Ismail Yusuf Rabbi**[2] | **Jean-Luc Jannink**[1,3]

[1]Plant Breeding and Genetics Section, School of Integrative Plant Science, College of Agriculture and Life Sciences, Cornell University, Ithaca, New York, USA

[2]International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria

[3]USDA-ARS, Robert W. Holley Center for Agriculture and Health, Ithaca, New York, USA

Correspondence
Jean-Luc Jannink, USDA-ARS, Robert W. Holley Center for Agriculture and Health, Ithaca, NY, USA. Email: jeanluc.work@gmail.com

Assigned to Associate Editor Alexander Lipka.

Funding information
Foreign, Commonwealth and Development Office, Grant/Award Number: INV-007637; Bill and Melinda Gates Foundation, Grant/Award Number: INV-007637

Abstract

The cassava (*Manihot esculenta* Crantz) breeding program at the International Institute of Tropical Agriculture (IITA) has adopted genomic selection to accelerate genetic gain. The program continues to develop varieties broadly adapted across Nigeria's diverse agroclimatic zones. However, for this purpose, genotype-by-environment interaction (GEI) presents a challenge. To decide whether broad adaptation breeding is a good strategy, we evaluated broad versus narrow adaptation strategies using stochastic simulation, assessing genetic gain, genetic variance, heritability, and selection accuracy at 0 versus realistic levels of GEI variance. To parameterize the models, we analyzed historical data from four phenotypic evaluation stages of the IITA breeding program to estimate genetic and error variances, and genetic correlations across environments. Based on these observed parameters, the genomic-enabled breeding programs exhibited higher genetic gain than the conventional program for both GEI variances. At realistic GEI variance, the narrow adaptation program showed higher genetic gain than the broad adaptation program. Across all programs, the genetic variance declined over time, though the genomic-enabled programs showed higher initial variance due to the selection of parents at earlier stages. At realistic GEI variance, an increase in genetic variance was observed in the narrow adaptation program due to its conversion of GEI between mega-environments into main genetic variance within mega-environments. This higher genetic variance led to higher heritabilities and selection accuracies. This study highlights the potential of genomic selection in accelerating genetic gain and suggests that dividing the IITA cassava breeding program to target more than one mega-environment should be considered.

Abbreviations: AYT, advanced yield trial; CET, clonal evaluation trial; GBLUP, genomic best linear unbiased predictor; GEBV, genomic estimated breeding value; GEI, genotype-by-environment interaction; GS, genomic selection; ME, mega-environment; PYT, preliminary yield trial; QTL, quantitative trait locus; SDN, seedling nursery; TP, training population; UYT, uniform yield trial.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Crop Science* published by Wiley Periodicals LLC on behalf of *Crop Science* Society of America . This article has been contributed to by U.S. Government employees and their work is in the public domain in the USA.

1 | INTRODUCTION

The breeding strategy used by International Institute of Tropical Agriculture (IITA) cassava (*Manihot esculenta* Crantz) breeders has evolved over time. Phenotypic recurrent selection was mainly used from the inception of the IITA cassava breeding program until 2012, when genomic selection (GS) was integrated to accelerate the rate of genetic improvement for complex traits like fresh root yield (de Oliveira et al., 2012; Wolfe et al., 2017; Yonis et al., 2020). GS is a marker-based selection approach using genome-wide and densely distributed molecular markers to increase the efficiency of selection on polygenic traits (T. H. E. Meuwissen et al., 2001). This approach involves using a training population (TP) and selection candidates. The TP needs to be phenotyped for desired traits in a target population of environments (TPEs) and genotyped to train the statistical model to predict genomic estimated breeding values (GEBVs) on the selection candidates (Jannink et al., 2010).

The implementation of GS-enabled population improvement is a complex task involving a cyclic crossing of promising parental clones, evaluation in several stages, and selection procedures over long periods of time (Ali et al., 2020). The assessment of the long-term effects of the chosen breeding strategy, particularly in the early phase of a breeding program, is demanding (Hoyos-Villegas et al., 2019). It is also impractical and not cost-effective to experiment with different breeding strategies due to limited resources budgeted for breeding.

A broad adaptation breeding program is a scenario where genotype-by-environment interaction (GEI) is ignored, and available resources are used to target all environments the breeding institute serves in a single breeding program (Braun et al., 1997; Ewing et al., 2019). In contrast, a narrow adaptation breeding program would partition all environments into two or more mega-environments (MEs) and create separate breeding programs where the available resources are partitioned between MEs (Gauch & Zobel, 1997). IITA's cassava breeding strategy is an example of the former, as it aims to develop broadly adapted cultivars with high and stable yield performance across all of Nigeria and other west-African countries. However, in the presence of high GEI, the broad-adaptation breeding strategy will be challenging due to the difficulty of finding individuals that perform well across all environments. In that case, a narrow adaptation breeding strategy may be appropriate, as it breeds cultivars that respond relatively better under specific set of environments only.

With the recent advances in computing power, simulation can be a valuable tool for plant breeders to cost-effectively evaluate different breeding scenarios and determine the best strategy toward making informed decisions in accordance with simulation results (Sun et al., 2011). Stochastic simulation, as used in this study, mimics the mechanisms of crossing, recombination, and evaluation, including their random components. Hoyos-Villegas et al. (2019) pointed out that stochastic simulation is appropriate to simulate entire breeding programs that are usually too involved to be deterministically modeled. It can also be used to study the rate of genetic improvement over time, selection and predictive accuracy, and cost-effectiveness of GS under different schemes (Gaynor et al., 2021).

There are many open-source simulation applications that are currently accessible to plant breeders, including QU-GENE (Podlich & Cooper, 1998), AlphaSim (Faux et al., 2016), AlphaSimR (Gaynor et al., 2021), Modular Breeding Program Simulator (Pook et al., 2020), ADAM-Plant (Liu et al., 2019), and BreedingSchemeLanguage (Yabe et al., 2017), to support crucial decision-making for cultivar development, particularly in modern breeding programs involving marker- and genomics-assisted selection methods. These applications have been beneficial for assessing existing breeding strategies in actual field-based breeding programs (Wang et al., 2003) and have also been used to develop novel breeding strategies.

The simulation of any novel breeding strategy involves two major phases: (i) burn-in phase, simulating the current strategy, and (ii) future breeding phase, simulating the new strategy that might be adopted. The burn-in phase provides a realistic starting point for all simulated breeding strategies. Gaynor et al. (2017) pointed out three levels of burn-in phase as follows: (i) the simulation of the species' initial genome sequences, (ii) the assignment to the initial sequences to founder haplotypes for the first generation of parental individuals, and (iii) simulation of years of conventional breeding without GS. After the burn-in, a future breeding phase is a testing phase for several simulated breeding programs branched off from the burn-in phase that may involve

Core Ideas

1. Simulation is a cost-effective approach to assess different breeding strategies toward making informed decisions.
2. Genomic selection outperformed a conventional breeding program in terms of rates of genetic gain, irrespective of genotype-by-environment interaction variance.
3. Breeding for narrow adaptation by splitting Nigerian locations into two mega-environments outperformed breeding for broad adaptation.
4. Not all biological complexities of genetic effects and real-world breeding programs can be captured in simulation.

14350653, 0, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/csc2.21170 by Nigeria Hinari NPL, Wiley Online Library on [13/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

either phenotypic or GS or both are tested for a certain number of years.

Stochastic simulations have been implemented in many crops, including wheat (*Triticum*) (Gaynor et al., 2017), maize (*Zea mays* L.) (Powell et al., 2020), clonally propagated crops (Covarrubias-Pazaran et al., 2022; Werner et al., 2020), tea (*Camellia sinensis*) (Lubanga et al., 2022), and sorghum (*Sorghum bicolor* spp.) (Muleta et al., 2019). However, to our knowledge, no simulation studies have been published on cassava (or other clonally propagated crop) to compare breeding for broad versus narrow adaptation in a GS context.

The principal goal of this study is to identify a breeding strategy that will maximize gains for cassava in Nigeria by assessing the relative value of broad adaptation breeding with one breeding program versus narrow adaptation breeding partitioning breeding resources between two sets of testing locations. To achieve this goal, we simulated a conventional breeding scheme based on phenotypic selection as a baseline where the number of crosses, number of progenies per cross, number of replicates per evaluation stage, and other simulation parameters mimicked the actual IITA cassava breeding program. Then we assessed simulated gain in breeding schemes targeting broad versus narrow adaptation in the GS context.

2 | MATERIALS AND METHODS

2.1 | Historical trial data

We chose parameters for the simulation study by analyzing historical data from 198 field trials sourced across four evaluation stages: clonal evaluation trial (CET, 13 trials), preliminary yield trial (PYT, 49), advanced yield trial (AYT, 76), and uniform yield trial (UYT, 60) of the IITA cassava breeding program located in Nigeria. The trials were evaluated for diseases, yield-related traits, and other agronomic traits of interest in nine crop growing seasons (2013–2022) across 11 locations in different agro-ecological zones. These data were used to estimate genetic, GEI, and error variances for each evaluation stage, and genetic correlations across trials (Bakare et al., 2022) to be used to parameterize stochastic simulations of the IITA cassava breeding program. For the CET stage, the trials were established as an augmented randomized complete block design in which the checks were replicated in every block to account for field spatial variability and to estimate the error variance. However, for other evaluation stages (PYT, AYT, and UYT) where the planting materials were sufficient, the trials were laid as randomized complete block design.

2.2 | Statistical analysis of historical data

Statistical analyses of historical data were carried out for each evaluation stage. We modeled fresh root yield trait using the *lmer()* function from the lme4 library (Bates et al., 2015) within the R statistical environment (R Core Team, 2022). The statistical model of fresh root yield for CET trials was fitted as follows:

$$\mathbf{y} = \mu + \mathbf{X}_1\mathbf{b} + \mathbf{X}_2\mathbf{c} + p\beta + \mathbf{Z}_1\mathbf{g} + \epsilon \quad (1)$$

where $\mathbf{y}$ is the $(n \times 1)$ vector of observed fresh root yields, in which n is the number of observations in the trial; μ is the intercept (global mean); $\mathbf{b}$ is the $(b \times 1)$ vector of fixed effect of blocks with its associated incidence matrix $\mathbf{X_1}$ of dimension $n \times b$; $\mathbf{c}$ is the $(c \times 1)$ vector of fixed effect of checks with its associated incidence matrix $\mathbf{X_2}$ of dimension $n \times c$; p denotes the proportion of plant stands harvested as a covariate (e.g., if 28 stands were planted but only 21 were harvested, $p = 0.75$); β is a regression coefficient relating p and y; $\mathbf{g}$ is the $(g \times 1)$ vector of random effect of new genotypes with its associated design matrix $\mathbf{Z_1}$ of dimension $n \times g$, where g is assumed to follow a Gaussiam distribution $g \sim N(0, I_n\sigma_g^2)$, and ϵ is a residual term that is assumed to follow a Gaussian distribution, $\epsilon \sim N(0, I_n\sigma_\epsilon^2)$. For other evaluation stages, the univariate linear mixed model was fitted as follows:

$$\mathbf{y} = \mu + \mathbf{X}_1\mathbf{b} + p\beta + \mathbf{Z}_1\mathbf{g} + \epsilon \quad (2)$$

where all terms were as defined in the previous equation.

In each fitted model, we derived a standardized residual value for each observation as a deviation of the observed data point from the predicted value scaled by the residual standard deviation. We refitted the model after excluding data points whose absolute value of standardized residuals was above 3, considered to be outliers.

2.2.1 | Use of variance estimates from historical data

After data curation within trials, data from all trials were assembled and analyzed using a standard GEI model that included main effects for environment and a random effect for GEI to obtain estimates of genetic, GEI, and error variances. The default genetic variance among founders used by AlphaSimR is 1. To simplify comparisons to other studies and species, we left that default. We adjusted error variances to obtain heritabilities of 0.05 and 0.2 in the seedling nursery (SDN) and CET, respectively. Error variances for the PYT,

14350653, 0, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/csc2.21170 by Nigeria Hinari NPL, Wiley Online Library on [13/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

AYT, and UYT were adjusted to maintain the historical error variance ratios between those trials and the CET. Finally, to choose an appropriate GEI variance, we used the historical GEI variance observed in the UYT, which covered all agro-ecological zones of Nigeria. We set the simulated GEI variance so that the ratio of simulated genetic to GEI variance was equal to the observed ratio (see "Trait simulation" below).

2.3 | Simulated founder genotypes

Stochastic simulation was implemented using the AlphaSimR package (Gaynor et al., 2021) in R (R Core Team, 2022) programming software environment. A diploid founder population size of 25 outbred cassava clones was used to form the genome and genotypes of initial parents in the burn-in phase based on simulated founder sequences. The biallelic genome sequences for 18 chromosome pairs of the founders were simulated using the Markovian Coalescent Simulator (Chen et al., 2009) in AlphaSimR (Gaynor et al., 2021). For each genotype, we considered a genetic map of 18 chromosome pairs with an average genetic length of 1.43 Morgans (M), resulting in a total genetic length of 25.7 M. The effective population size (N_e) was 100 to mimic a history of natural and artificial selection. We assumed each chromosome had 150 randomly sampled segregating quantitative trait loci (QTLs) (2700 genome-wide) and 500 segregating loci as single nucleotide polymorphisms (SNPs) per chromosome (9000 genome-wide), corresponding to the number of observed SNPs. The positions of QTLs and SNP markers were randomly distributed, and sites for SNP markers and QTL across the entire genome did not overlap.

2.4 | Trait simulation

We simulated a single complex trait (fresh root yield) by modeling three genetic effects: additive, dominance, and GEI, referred to as an ADG trait in AlphaSimR (Gaynor et al., 2021). The genetic value of this trait was the sum of additive, dominance, and GEI effects over the 2700 QTLs, modeled as follows:

$$\mathrm{GV}(\mathbf{x}, w) = \mu + a(\mathbf{x}) + d(\mathbf{x}) + g(\mathbf{x}, w) \quad (3)$$

where GV($\mathbf{x}$, w) represented an individual's genetic value, with $\mathbf{x}$ denoting a vector of QTL genotype dosages and w representing an environmental covariate inducing GEI. The intercept (μ) is the trait mean in the founder population. The overall additive effect across the n-vector sites of QTL is derived as $a(\mathbf{x}) = \sum_{i=i}^{n_{\mathrm{QTL}}} a_i \mathbf{x}_i$, where a_i were randomly sampled from a normal distribution set so that the additive genetic variance of the founder population has an expected value of 0 and variance of 1. Dominance effects for all loci were derived as

$$d(\mathbf{x}) = \sum_{i=1}^{n_{\mathrm{QTL}}} \begin{cases} \delta_i \times |a_i|, & if \ \mathbf{x}_i = 1 \\ 0, & otherwise \end{cases}$$

the product of locus-specific dominance degree (δ_i) and absolute value of its additive effect a_i. Dominance degrees were sampled from normal distribution with $\delta_i \sim N(\mu_\delta, \sigma^2_\delta)$, where μ_δ is the mean dominance degree set to 0.20 to simulate positive directional dominance and σ^2_δ is the dominance variance equal to 0.1 (de Andrade et al., 2022; Wolfe et al., 2016). A dominance degree of 0 denotes no dominance, and a dominance degree of 1 signifies complete dominance. Dominance degrees between 0 and 1 correspond to partial dominance, and values above 1 correspond to overdominance.

The GEI effects were modeled as $g(\mathbf{x}, w) = wb(\mathbf{x})$, where w was an environmental covariate $w \sim N(0, 1)$ and $b(\mathbf{x})$ was a genotype-specific slope defined as follows:

$$b(\mathbf{x}) = \mu_G + \sum_{i=1}^{n_{\mathrm{QTL}}} g_i x_{ia}$$

where μ_G represents an intercept value and $\sum g x_a$ denotes sum over all QTL of products of locus-specific genotype-by-environment effects (g) and scaled additive dosages (x_a). For a given environment, the user specifies the environmental covariate w. Given the distribution of w, a covariate of 0 represents the average of all environments (Bajgain et al., 2020; Bakare et al., 2022). We simulated a testing network of nine locations. For each location in each year, the covariate for that location was sampled from a distribution with ranges for each location as in Table 1.

The ranges shown in Table 1 were obtained from the location means of the environment loadings for Factor 2 from Table 3 of Bakare et al. (2022). We chose Factor 2 because (unlike Factor 1) it induced crossover GEI. With this evaluation network, the environmental covariates range from −1.9 to 2.1. Across locations, the distribution of the environmental covariate was therefore not a standard normal but a truncated normal with a variance of 0.44. To preserve the relative ratio of genetic to GEI variances, we took observed variances from historical UYT trials of 6.5 (GEI) and 7.7 (genetic). Thus, we set the simulated GEI variance to (6.5/7.7) × (1/0.44), where the 1 in the formula is default variance of the environmental covariate and the 0.44 comes from the reduced variance due to truncation. We rounded this value to 2.0. Consequently, we simulated two levels of genotype-by-environment variance: variance of GEI = 0 as baseline and GEI variance = 2.0, which is estimated from historical data.

14350653, 0, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/csc2.21170 by Nigeria Hinari NPL, Wiley Online Library on [13/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Ranges of environmental covariates inducing the genotype-by-environment interaction for each of the nine locations in the breeding evaluation network. For each phenotyping year, a covariate was randomly sampled from these ranges for each location.

Environmental covariates	Loc1	Loc2	Loc3	Loc4	Loc5	Loc6	Loc7	Loc8	Loc9
Minimum	−1.9	−0.9	−0.9	−0.8	−0.6	−0.3	0.0	0.1	0.9
Maximum	−0.7	0.3	0.3	0.4	0.6	0.9	1.2	1.3	2.1

Abbreviation: Loc, location.

TABLE 2 Global parameters used for the simulation of evaluation stages in cassava breeding scheme optimization.

	Simulation parameters used across stages for breeding scenarios						
			Broad adaptation		Narrow adaptation		
						ME1	ME2
Stage	nReps	Scaled error variance	nClones	Location(s)	nClones	Location	
F1/SDN	1	19.0	10,000	L5	5000	L5	L5
CET	1	4.0	1000	L5	500	L5	L5
PYT	2	3.7	200	L3, L7	200	L7	L3
AYT	2	2.4	60	L3, L4, L6, L7	60	L6, L7	L3, L4
UYT	3	2.3	30	L1, L2, L3, L4, L6, L7, L8, L9	30	L6, L7, L8, L9	L1, L2, L3, L4

Abbreviations: AYT, advanced yield trial; CET, clonal evaluation trial; ME1, mega-environment 1; ME2, mega-environment 2; nClones, number of clones evaluated; nReps, number of replicate(s) of each clone; PYT, preliminary yield trial; SDN, seedling nursery; UYT, uniform yield trial.

2.5 | Simulation design

The simulation was run for 30 1-year breeding cycles in two phases: 20 years of burn-in and 10 years of future breeding. We ran 50 independent replicates with each of the two levels of genotype-by-environment variance.

2.5.1 | Burn-in phase

The burn-in phase used conventional breeding with phenotypic selection to serve as a common starting point for comparing the breeding scenarios. Before the burn-in phase, we filled the cassava breeding pipeline with outbred clones through seven overlapping cycles of random crossing, evaluation, and selection. For every cycle or simulation replicate, we started with 25 new parents to generate 200 biparental crosses with 50 seedlings per cross, resulting in a base population (cycle 0) size of 10,000 seedlings. The population of 10,000 F1 seedlings resulting from the crossing block was evaluated as an unreplicated seedling nursery trial in one location. Visual selection in the seedling nursery is modeled as selection on yield, where we assumed a low heritability of 0.05, which corresponds to a high error variance (Table 2) relative to the additive variance of the founder population. In the second year, the top 1000 clones from the seedling nursery were advanced to the CET and clonally propagated as an unreplicated trial in a single location. We scaled the observed error variance of 67.6–4.0, which corresponds to a heritability of 0.2 relative to the founder population. In the third year, 200 clones were evaluated in PYT, the first stage of replicated yield trial in two locations with an error variance of 62.4 scaled to 3.7 ($4/67.6 \times 62.4$). In the fourth year, 60 clones were advanced to AYT and evaluated in a replicated field design in four locations with an error variance of 41.1 scaled to 2.4 ($4/67.6 \times 41.1$) per location. In the fifth year, 30 clones were evaluated in the UYT in eight locations with an error variance of 39.5 scaled to 2.3 ($4/67.6 \times 39.5$) per location. In the sixth year, the same 30 clones from previous year were reevaluated in UYT. These clones were also considered as potential candidates for next year's crossing block. In the seventh year, the four clones with the best mean performance over the previous 2 years of UYT evaluation were considered for variety release.

2.6 | Future phase breeding strategies

We compared conventional phenotypic selection with broad- and narrow-adaptation GS programs. Conventional phenotypic selection was simply the continuation of the burn-in strategy described above. Broad adaptation GS used the same field assessment as the conventional program, but clones constituting the TP were also genotyped. Then, the candidate clones were advanced to subsequent stages based on phenotypic value and genomic estimated total genetic values (GETGV). The breeding cycle time was shortened by selecting new parental clones based on GEBV (Crossa et al., 2017; T. Meuwissen et al., 2016) from the combined population

14350653, 0, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/csc2.21170 by Nigeria Hinari NPL, Wiley Online Library on [13/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

of CET, PYT, AYT, and UYT stages. Like the conventional breeding scheme, the SDN and CET were evaluated in one location, and the PYT, AYT, and UYT were evaluated in two, four, and eight locations, respectively. We simulated 200 crosses of 50 progeny planted in the SDN. The five top individuals per family were selected to advance 1000 individuals to CET.

In the narrow adaptation program, the testing sites were split into two MEs (ME1 and ME2). The TPEs were split such that environmental covariates ranging from −0.3 to 2.1 were assigned to ME1 and environmental covariates ranging from −1.9 to 0.4 were assigned to ME2, and both had common location 5, where the CET and SDN were established, with an average environmental covariate of 0.0. The mean environmental covariates of ME1 and ME2 were 0.62 and −0.42, respectively. For each ME, the cohorts at PYT, AYT, and UYT were evaluated at one, two, and four locations, respectively (i.e., half as many locations as for the corresponding stages in the broad adaptation program). In contrast to the broad adaptation program, 100 crosses with 50 progeny each were simulated in each ME, again advancing the five best individuals per family. Thus, 500 individuals were advanced per ME for a total of 1000 individuals as part of the broad adaptation program.

2.7 | Training population and genomic prediction model

The training datasets for GS models were initiated with genotyped and phenotyped individuals starting in the first year of the future breeding phase from the CET, PYT, AYT, and UYT. The initial TP had 1290 clones and 1880 phenotypic records for broad adaptation, representing the total number of clones and clone-by-location combinations across the evaluation stages. In contrast to the narrow adaptation program, each ME had a TP of 940 phenotypic records on 790 clones. In each cycle of selection, the training set was updated by adding all the new individuals from PYT, AYT, and UYT. For computational reasons, we did not retain all CET individuals in the TP. Rather, in each year, only that year's CET, representing the current selection candidates, was included in the TP. Prior to fitting the genomic prediction model, we filtered the simulated SNPs marker dataset using qc.filtering() function of ASRgenomics library (Gezan et al., 2021) to remove markers with minor allele frequency less than 0.05 and heterozygosity greater than 0.95.

The genomic best linear unbiased predictor (GBLUP) model was implemented in a linear mixed model framework using ASREML-R version 4.0 (Butler et al., 2017) as in:

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\mathbf{G}^{-1}\lambda \end{bmatrix} \begin{bmatrix} b \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix} \quad (3)$$

where is $\mathbf{y}$ is the vector of simulated phenotypic value from the TP; b is fixed effect estimates of environment; $\mathbf{u}$ is the vector of random breeding values of clones distributed as $\mathbf{u} \sim N(0, G\sigma^2_{\mathbf{u}})$; in which $\sigma^2_{\mathbf{u}}$ is the additive genetic variance and G is the additive genomic relationship derived from marker information following VanRaden method 1 (VanRaden, 2008). $\mathbf{X}$ and $\mathbf{Z}$ are design matrices for fixed and random effects, respectively. R is the variance structure of residual, $e \sim N(0, R\sigma^2_e)$, which assigned inverse variance weight to each observation according to the empirically observed error variance from each stage. Note that λ is a shrinkage factor expressed as a ratio of error variance to genetic variance of breeding values.

2.8 | Comparison of breeding programs

The progress of each breeding program was tracked across the 50 independent simulation runs based on three metrics: genetic gain, genetic variance, broad- and narrow-sense heritabilities on entry-mean basis, and selection accuracy over the 10 cycles of future breeding. The mean and variance of genetic value were calculated and saved for each cycle at the CET. We visualized the trend of change in genetic mean and variance for CET entries over cycles of selection (years 0–10) where year 0 was defined as the last year of the burn-in phase.

The breeding strategies were also compared based on heritability, which is one of the key parameters driving the rate of genetic gain for a quantitative trait in a breeding program. We calculated the broad-sense heritability for the conventional breeding program from clones pooled from the PYT, AYT, and UYT evaluation stages by fitting a BLUP model. For the GS programs, we fitted a GBLUP model that included both additive and nonadditive effects to estimate broad- and narrow-sense heritabilities from clones constituted from the TP. The performance of each breeding strategy depended on its prediction accuracy, measured as Pearson's correlation between true genetic value and the GEBV selection criterion (Desta & Ortiz, 2014). This accuracy was calculated for each evaluation stage.

3 | RESULTS

3.1 | Parameter estimates from historical data and parental candidate for crossing

The parameter estimates resulting from analyzing historical data provided prior information to simulate phenotypic values, which mimic empirical data from each evaluation stage of the IITA breeding scheme. We observed a decline in both genetic and error variances from early stage (CET) to late stage (UYT) of the breeding program (Figure 1).

14350653, 0, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/csc2.21170 by Nigeria Hinari NPL, Wiley Online Library on [13/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

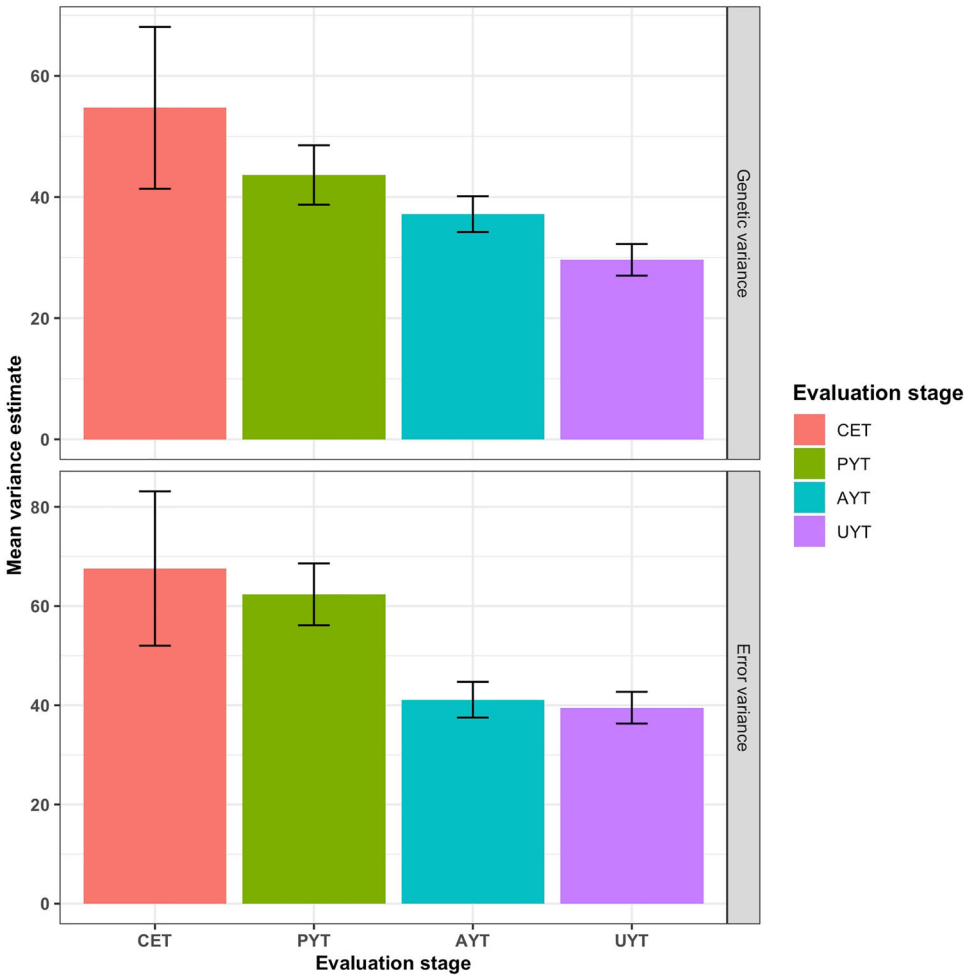

FIGURE 1 A comparison of mean genetic and error variances derived from statistical analysis of 198 historical trial data over four evaluation stages: clonal evaluation trial (CET, 13), preliminary yield trial (PYT, 49), advanced yield trial (AYT, 76), and uniform yield trial (UYT, 60) trials, in the International Institute of Tropical Agriculture (IITA) cassava breeding program. Error bars are ± one standard error of the mean estimated over trials.

The distribution of the evaluation stage from which the 25 selected parents came from across the CET, PYT, AYT, and UYT in each breeding cycle was displayed for each simulated breeding program (Figure S1). In the conventional program, parents could only be selected from the AYT and UYT. From the GS programs, they could already be selected in the CET since they were genotyped. Given those constraints, parents were chosen by truncation selection, irrespective of the evaluation stage they came from. Thus, the percentage of lines coming from each stage was not predetermined but depended on the relative genotypic or breeding values of clones from each stage. Across all programs, most parents were selected from the AYT. The broad and narrow programs differed in that in the latter, more clones were selected from the CET than in the former.

3.2 | Realized parameters for assessing breeding programs

3.2.1 | Genetic gain

The genetic gain was measured as a change in mean genetic values over 10 cycles of selection during the future breeding phase. The plot of the trend of genetic gain showed that genomic-enabled breeding programs outperformed the conventional breeding programs at both levels of GEI variance (Figure 2). In the absence of GEI variance, we observed no significant difference among the three breeding scenarios until after six cycles when genomic-based scenarios were significantly better than the conventional program. The slopes of linear regression of mean genetic values on years of breeding

14350653, 0, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/csc2.21170 by Nigeria Hinari NPL, Wiley Online Library on [13/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

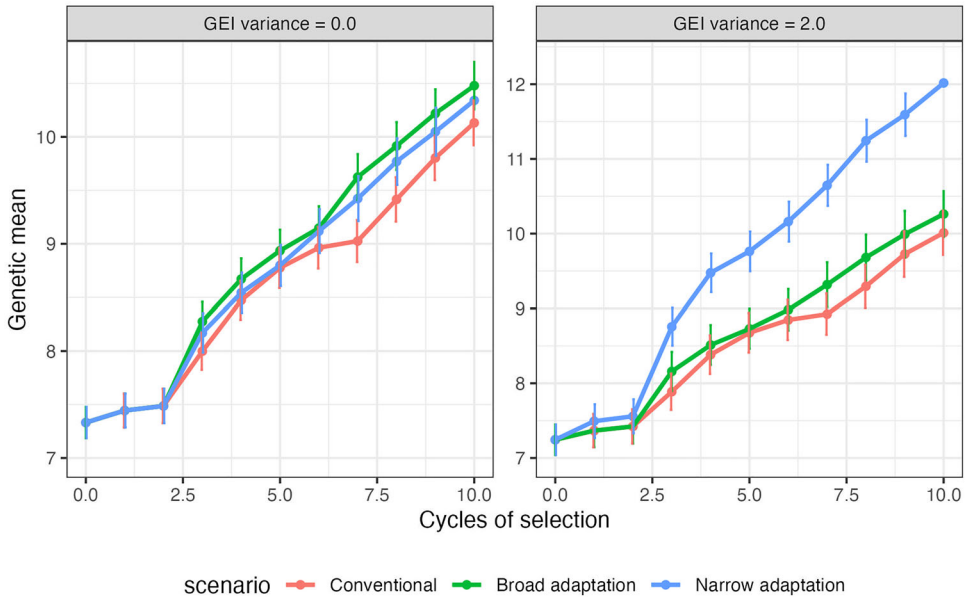

FIGURE 2 The trend of genetic gain of three breeding scenarios at two levels of genotype-by-environment interaction (GEI) variance over 10 cycles of selection averaged across 50 simulation runs. Genetic gain is a change in the mean genetic value of entries at clonal evaluation trial (CET) stage over time. The trend over time is displayed with error bars. The error bars are ± one standard error of the mean of the reported simulated parameter.

were 0.30, 0.35, and 0.33 for conventional, broad, and narrow adaptation, respectively, at GEI variance of 0 (Figure S2). At high GEI variance (GEI variance = 2.0), the narrow adaptation breeding program showed the highest gain (Figure 2), and we observed slopes of 0.30, 0.33, and 0.52 for conventional, broad, and narrow adaptation breeding, respectively (Figure S2).

3.2.2 | Genetic variance

Genetic variance observed at the CET declined for all the breeding strategies over breeding cycles at GEI variance = 0 (Figure 3). The pattern of loss of genetic variance differed across breeding strategies. Genetic variance declined continuously in the conventional program for both levels of GEI variance. In the absence of GEI variance, a bump in CET genetic variance occurred in the third year of the GS programs, after which their genetic variances declined more rapidly than the conventional program (Figure 3). This bump in CET genetic variance coincided with the first CET composed of individuals coming from crosses of parents selected by GS. At high GEI variance, CET genetic variance increased sharply for narrow adaptation at the onset of GS, which coincided with partitioning the program between the two MEs. Afterward, we observed a more rapid decline in the narrow program genetic variance relative to both broad adaptation and conventional programs (Figure 3).

3.2.3 | Heritability

In the conventional program, broad-sense heritability estimates were 0.29 and 0.13 for GEI variances of 0.0 and 2.0, respectively, averaged across 10 cycles of selection (Figure 4). We did not estimate narrow-sense heritability in the conventional program because the program did not use a relationship matrix. For the genomic-based breeding strategies, the presence of GEI variance affected heritabilities in opposite directions for the broad and narrow adaptation programs. For the broad program, heritabilities declined from 0.28 and 0.36 (narrow- and broad-sense) to 0.16 and 0.20 in the presence of GEI (Figure 4). In contrast, for the narrow program, heritabilities increased from 0.24 and 0.29 to 0.35 and 0.41 (Figure 4).

3.2.4 | Selection accuracy

We observed a consistent increase in selection accuracy from early to late evaluation stages for all the breeding strategies at both levels of GEI variance (Figure 5). The mean selection accuracy for the conventional breeding program ranged from 0.18 (SDN) to 0.83 (UYT) and 0.19 (SDN) to 0.79 (UYT) for GEIs of 0.0 and 2.0, respectively. For the broad adaptation program, it ranged from 0.19 (SDN) to 0.86 (UYT) and 0.19 (SDN) to 0.82 (UYT) for GEI of 0.0 and 2.0, respectively. Narrow adaptation showed a similar trend with mean

14350653, 0, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/csc2.21170 by Nigeria Hinari NPL, Wiley Online Library on [13/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

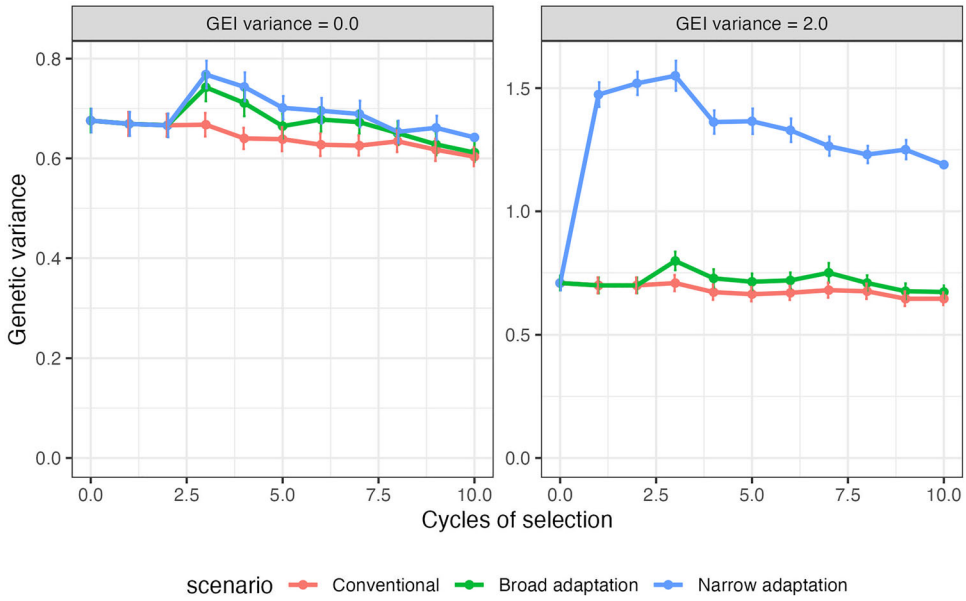

FIGURE 3 Trend of genetic variance of three breeding strategies at two levels of genotype-by-environment interaction (GEI) variance over 10 cycles averaged across 50 simulation runs. Genetic variance is the variance of genetic value of entries at clonal evaluation trial (CET). The error bars are ± one standard error of the mean of reported simulated parameter.

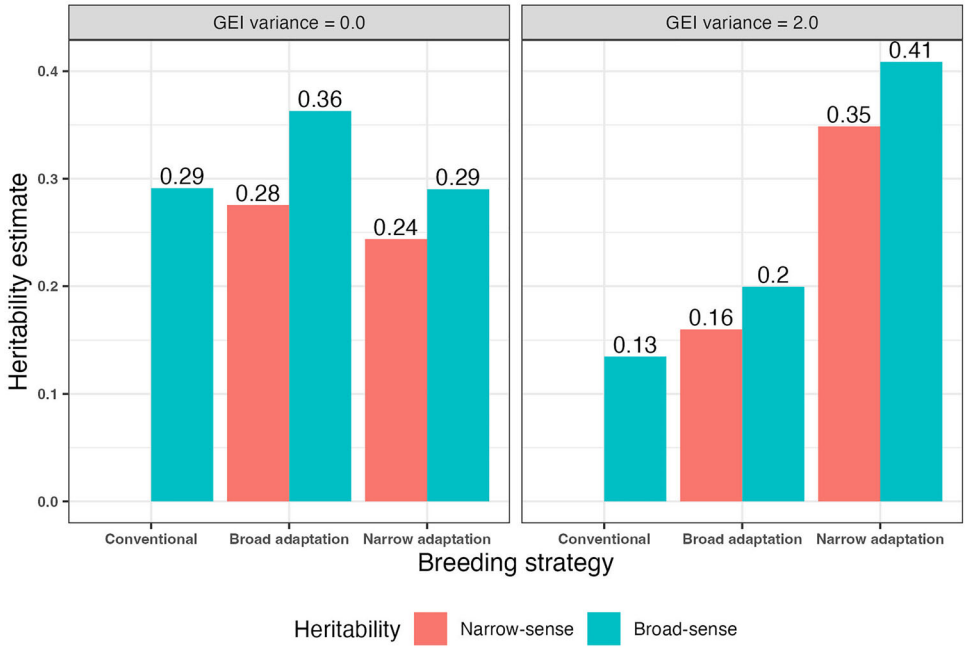

FIGURE 4 A comparison of broad and narrow-sense heritabilities over the three simulated breeding strategies. GEI, genotype-by-environment interaction.

selection accuracy varying from 0.19 (SDN) to 0.80 (UYT) and from 0.13 (SDN) to 0.87 (UYT) for GEIs of 0.0 and 2.0, respectively.

At GEI variance = 0, conventional and broad adaptation accuracies were similar. Narrow adaptation had similar accuracy to those for the SDN and CET but had lower accuracy at PYT, AYT, and UYT. As for GEI variance = 0, the selection accuracy for conventional and broad adaptation was at GEI variance = 2 and had similar medians. However, at GEI variance = 2, accuracies in the narrow adaptation program were lower at SDN and CET but higher at PYT, AYT, and UYT compared to other programs.

14350653, 0, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/csc2.21170 by Nigeria Hinari NPL, Wiley Online Library on [13/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

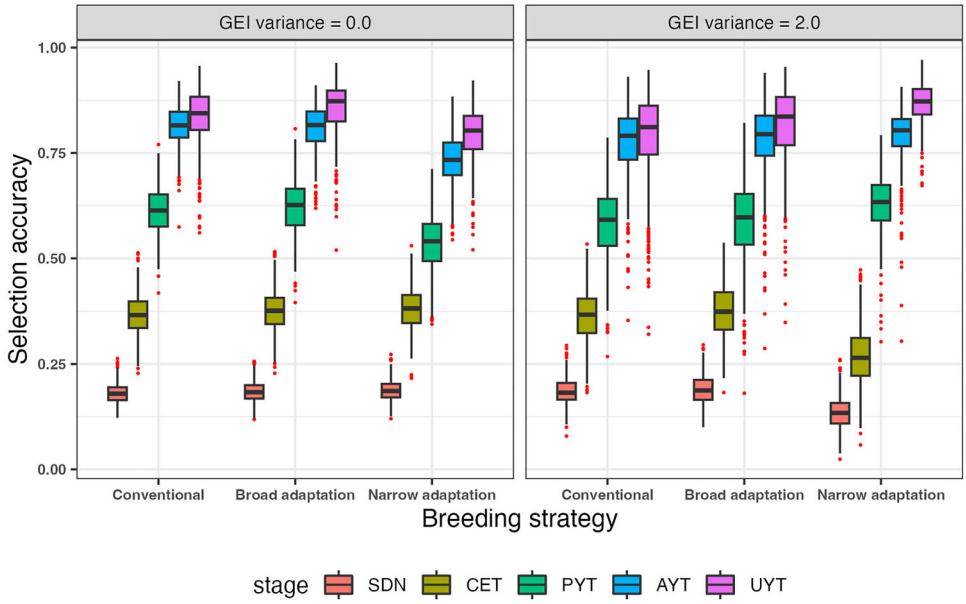

FIGURE 5 Distribution of selection accuracy across evaluation stages of three breeding strategies at two levels of genotype-by-environment interaction (GEI) variance over 10 cycles of selection averaged across 50 simulation runs. Selection accuracy is the Pearson's correlation between true genetic value and phenotypic selection criterion for conventional and estimated breeding value for genomic-enabled breeding strategies. AYT, advanced yield trial; CET, clonal evaluation trial; PYT, preliminary yield trial; SDN, seedling nursery; UYT, uniform yield trial.

4 | DISCUSSION

This study used stochastic simulation to assess the performance of broad and narrow adaptation GS breeding programs for IITA cassava breeding in Nigeria. The primary purpose of the study was to support the IITA program in deciding whether to transition to breeding for narrow adaptation from its current broad adaptation approach. It would be straightforward to come up with simulation parameters and conditions that would favor either broad or narrow adaptation breeding. We tailored these simulations to the IITA program by extensively analyzing their trials and using the resulting variance components to parameterize the simulations. We note, however, that we set the simulated GEI variance based on empirical estimates of *total* GEI variance for cassava in Nigeria. GEI variance simulated by AlphaSimR is always of the *crossover* type, which is more problematic for breeding and more strongly favors narrow adaptation. Thus, our parameterization may have favored the narrow adaptation breeding program. A full assessment of this issue should be the subject of further research.

The way AlphaSimR induces GEI is by supplying an environmental covariate to each testing environment (Gaynor et al., 2021). In a previous study (Bakare et al., 2022), a factor analytic covariance structure applied to IITA's UYT trials showed that testing environments could be characterized by three latent factors. Empirically, the values of environmental loadings on these factors also allowed the calculation of genetic correlations between performance across environments (Figure S3). Though we found the three-factor model to be most parsimonious (and the four-factor model was close, Bakare et al., 2022), AlphaSimR only uses an environmental covariate in *one* dimension. We used the second factor, which induced crossover interaction, to parameterize distributions for the environmental covariates sampled for each phenotypic evaluation in the simulated locations. The limitation of one covariate in AlphaSimR prevented us from simulating environments that closely mimicked the observed distribution of genetic correlations (Figure S4). The empirical and simulated distributions had similar ranges (the former varying from −0.4 to 1.0 and the latter from −0.7 to 1.0), but the shapes of the distributions differed. It is unclear what impact this difference might have on the relative performance of broad versus narrow adaptation programs.

The realized genetic gains from GS-enabled breeding programs were greater than those of a conventional breeding program. This finding was consistent with previous simulation studies in other crop species, including tea, rice, wheat, and sorghum (Lubanga et al., 2022; Muleta et al., 2019; Sabadin et al., 2022; Tessema et al., 2020), which affirmed that breeding programs implementing GS can achieve higher genetic gain than phenotypic selection breeding programs. These findings revealed that the use of the GS technique can enhance the efficiency of the breeding program even in the presence of G × E due to information sharing among genotyped clones over years and locations (Heffner et al., 2009). It

14350653, 0, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/csc2.21170 by Nigeria Hinari NPL, Wiley Online Library on [13/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

is worth noting that the conventional program still achieved reasonable genetic gain, albeit slightly lower, even in the presence of GEI.

The genetic variance over breeding cycles declined for all three breeding strategies, regardless of the level of GEI variance. At a GEI variance of 0, the bump in genetic variance observed in the GS programs (Figure 3) was attributed to the fact that year 3 was the first year in which progeny from parents selected by GS were evaluated in the CET. Progeny from GS had higher variance than progeny from the burn-in because in the latter parents only derived from the AYT and UYT stages, whereas under GS, parents could come from the CET through UYT stages (Figure S1). Variance among individuals in the AYT and UYT is lower than that of earlier stages (Figure 1), both because of selection bottlenecks and because of the Bulmer effect (Bulmer, 1971). Consequently, progenies resulting from crossing in the GS program had higher variances than from crossing in the conventional program.

When GEI was present, the initial increase in genetic variance of narrow adaptation was due to individuals being partitioned into two MEs to train GS model. The model AlphaSimR uses to induce GEI shows how this partitioning affects main genetic and GEI variances. As described, each environment is characterized by a single covariate, w. The GEI deviation for individual x is $wb(x)$, where $b(x)$ is a genotype-specific slope determined by the genotype of x and parameterized to have 0 mean and a variance equal to the overall GEI variance. The variance of the product is $\text{var}(wb) = E(w)^2\ \text{var}(b) + \text{var}(w)E(b)^2 + \text{var}(w)\text{var}(b)$. Over *all* environments, $E(w) = 0$, $\text{var}(w) = 1$, $E(b) = 0$, and $\text{var}(b) = \text{varGEI}$, causing the desired GEI variance (Gaynor, 2023). But selecting for narrow adaptation necessarily means ***not*** evaluating over all environments, but over a subset. Assume that this subset is characterized by covariates w_n where the n subscript indicates values sampled within the narrow TPE. The distribution of w_n is different from that of w. In particular, $E(w_n) \neq 0$ and $\text{var}(w_n) < \text{var}(w)$. The distribution of b is unchanged. Therefore, when selecting for narrow adaptation, we have:

$$\text{var}\left(w_n b\right) = E\left(w_n\right)^2\ \text{var}\left(b\right) + \text{var}\left(w_n\right)\text{var}\left(b\right)$$

The two components of $\text{var}(w_n b)$ correspond to the two impacts of selecting for narrow adaptation: the $E(w_n)^2\text{var}(b)$ component corresponds to the increase in the genetic variance within the narrow TPE relative to the broad all-environment TPE. It results from the conversion of a fraction of the GEI variance present across all environments into main genetic variance within narrow MEs. The $\text{var}(w_n)\text{var}(b)$ component corresponds to the fact that within a narrow TPE, the G × E variance will be lower than the all-environment G × E variance because $\text{var}(w_n) < \text{var}(w)$. To see that $E(w_n)^2\text{var}(b)$ does correspond to an increment of the genotypic variance within the narrow TPE, calculate that increment as follows:

$$\underset{x}{\text{var}}\left[E\left(w_n b|x\right)\right] = \underset{x}{\text{var}}\left[E\left(w_n\right)E(b|x)\right] = E\left(w_n\right)^2\text{var}\left(b\right)$$

Here, the first equality is correct because the covariance between w_n and b is 0, and w_n does not depend on x. The second equality is correct because $E(w_n)$ is a constant. Note that this increment is nonzero precisely because $E(w_n) \neq 0$.

The broad-sense heritabilities estimated from GS-enabled breeding programs were higher than those of the conventional breeding program. This reflected the earlier selection of parents, such that parents were more variable under GS than conventional breeding. In the absence of GEI, we had higher heritability estimates for broad adaptation breeding programs relative to narrow adaptation programs. This higher heritability reflected the increased replication over locations possible in the larger broad adaptation program. At a GEI variance of 2.0, however, it was evident that narrow adaptation breeding programs showed higher heritability estimates compared to broad adaptation breeding programs. This higher heritability was a direct consequence of the higher within ME genetic variance discussed above. Targeting narrow adaptation increased the heritability because the within ME genotypic variance was greater than the genotypic variance across all environments.

The selection accuracy, measured as the correlation between true genetic value and the selection criterion, increased consistently from the early stage (SDN) to the late stage (UYT) for all breeding strategies. This increase could be attributed to an increase in the number of replications and locations, which increased the heritability of the trait. However, the observed variation in selection accuracy could be due to at least two factors. First, we selected the best individuals between the evaluation stages based on the highest phenotypic value or estimated total genetic value, resulting in a reduction in genetic variation. Second, in a stochastic simulation, there will be variation in accuracy from stage to stage, year to year, and replicate to replicate. The accuracy in the narrow breeding program was similar to the broad program at SDN and CET because they both had one observation at one location, but it was lower at PYT, AYT, and UYT because for those stages, the narrow program only had half as many locations. At GEI = 2, the higher accuracy observed in narrow breeding program came from the fact that GEI variance has been converted to main effect genetic variance, resulting in higher heritability and consequently higher accuracy.

14350653, 0, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/csc2.21170 by Nigeria Hinari NPL, Wiley Online Library on [13/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

5 | CONCLUSIONS

In conclusion, the results suggested that use of GS-enabled breeding strategies can increase genetic gain of cassava breeding programs. The clearest practical result from the study was that under GEI conditions, as we believe they are experienced in the IITA program, more rapid gain was observed when breeding for narrow than broad adaptation. The recommendation in favor of breeding for narrow adaptation is strengthened by the fact that even when GEI = 0, the condition that should most favor breeding for broad adaptation, gains by breeding for narrow adaptation were a close second. While the results provide valuable insights, there are some limitations to consider. First, the study focuses on a specific breeding program and crop (cassava), which might limit the generalizability of the findings to other breeding programs or crops. Additionally, the simulation-based approach used in this study did not account for the costs of labor, phenotyping, and genotyping. The underlying assumptions and parameter values used may not fully capture the complexities of real-world breeding programs. Future research could involve validation of these findings using real-world data to strengthen the findings and expand their applicability to other breeding programs and crops.

AUTHOR CONTRIBUTIONS

Moshood A. Bakare: Conceptualization; data curation; formal analysis; methodology; visualization; writing—original draft; writing—review and editing. **Siraj Ismail Kayondo**: Conceptualization; visualization; writing—review and editing. **Peter Kulakow**: Funding acquisition; project administration; supervision. **Ismail Yusuf Rabbi**: Funding acquisition; project administration; supervision. **Jean-Luc Jannink**: Conceptualization; funding acquisition; methodology; supervision; visualization; writing—review and editing

ACKNOWLEDGMENTS

The authors express their appreciation to the UK's Foreign, Commonwealth & Development Office (FCDO) and the Bill & Melinda Gates Foundation (Grant INV-007637, http://www.gatesfoundation.org) for their financial support. They also extend their deep appreciation to the dedicated staff of the International Institute of Tropical Agriculture (IITA) of the cassava breeding program that assisted with the fieldwork and data collection for statistical analyses.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The simulated datasets and R script used to support findings in this study can be found on the author's github repository: https://github.com/mab658/cassavaSimulationModels.

ORCID

Moshood A. Bakare https://orcid.org/0000-0003-1910-8233

Siraj Ismail Kayondo https://orcid.org/0000-0002-3212-5727

Jean-Luc Jannink https://orcid.org/0000-0003-4849-628X

REFERENCES

Ali, M., Zhang, L., Delacy, I., Arief, V., Dieters, M., Pfeiffer, W. H., Wang, J., & Li, H. (2020). Modeling and simulation of recurrent phenotypic and genomic selections in plant breeding under the presence of epistasis. *Crop Journal*, *8*(5), 866–877. https://doi.org/10.1016/j.cj.2020.04.002

Bajgain, P., Zhang, X., & Anderson, J. A. (2020). Dominance and G×E interaction effects improve genomic prediction and genetic gain in intermediate wheatgrass (*Thinopyrum intermedium*). *The Plant Genome*, *13*(1), e20012. https://doi.org/10.1002/tpg2.20012

Bakare, M. A., Kayondo, S. I., Aghogho, C. I., Wolfe, M. D., Parkes, E. Y., Kulakow, P., Egesi, C., Jannink, J.-L., & Rabbi, I. Y. (2022). Parsimonious genotype by environment interaction covariance models for cassava (*Manihot esculenta*). *Frontiers in Plant Science*, *13*, 978248. https://doi.org/10.3389/fpls.2022.978248

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Braun, H.-J., Rajaram, S., & Ginkel, M. (1997). CIMMYT's approach to breeding for wide adaptation. In P. M. A. Tigerstedt (Ed.), *Adaptation in plant breeding* (pp. 197–205). Springer. https://doi.org/10.1007/978-94-015-8806-5_25

Bulmer, M. G. (1971). The effect of selection on genetic variability. *American Naturalist*, *105*, 201–221. https://doi.org/10.1086/282718h

Butler, D. G., Cullis, B. R., Gilmour, A. R., Gogel, B. J., & Thompson, R. (2017). *ASReml-R reference manual version 4*. VSN International Ltd.

Chen, G. K., Marjoram, P., & Wall, J. D. (2009). Fast and flexible simulation of DNA sequence data. *Genome Research*, *19*(1), 136–142. https://doi.org/10.1101/gr.083634.108

Covarrubias-Pazaran, G., Gebeyehu, Z., Gemenet, D., Werner, C., Labroo, M., Sirak, S., Coaldrake, P., Rabbi, I., Kayondo, S. I., Parkes, E., Kanju, E., Mbanjo, E. G. N., Agbona, A., Kulakow, P., Quinn, M., & Debaene, J. (2022). Breeding schemes: What are they, how to formalize them, and how to improve them? *Frontiers in Plant Science*, *12*, 791859.

Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., De Los Campos, G., Burgueño, J., González-Camacho, J. M., Pérez-Elizalde, S., Beyene, Y., Dreisigacker, S., Singh, R., Zhang, X., Gowda, M., Roorkiwal, M., Rutkoski, J., & Varshney, R. K. (2017). Genomic selection in plant breeding: Methods, models, and perspectives. *Trends in Plant Science*, *22*(11), 961–975. https://doi.org/10.1016/j.tplants.2017.08.011

14350653, 0, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/csc2.21170 by Nigeria Hinari NPL, Wiley Online Library on [13/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

de Andrade, L. R. B., Sousa, M. B. E., Wolfe, M., Jannink, J.-L., De Resende, M. D. V., Azevedo, C. F., & de Oliveira, E. J. (2022). Increasing cassava root yield: Additive-dominant genetic models for selection of parents and clones. *Frontiers in Plant Science*, *13*, 1071156. https://doi.org/10.3389/fpls.2022.1071156

de Oliveira, E. J., De Resende, M. D. V., Da Silva Santos, V., Ferreira, C. F., Oliveira, G. A. F., Da Silva, M. S., de Oliveira, L. A., & Aguilar-Vildoso, C. I. (2012). Genome-wide selection in cassava. *Euphytica*, *187*(2), 263–276. https://doi.org/10.1007/s10681-012-0722-0

Desta, Z. A., & Ortiz, R. (2014). Genomic selection: Genome-wide prediction in plant improvement. *Trends in Plant Science*, *19*(9), 592–601. https://doi.org/10.1016/j.tplants.2014.05.006

Ewing, P. M., Runck, B. C., Kono, T. Y. J., & Kantar, M. B. (2019). The home field advantage of modern plant breeding. *PloS One*, *14*(12), e0227079. https://doi.org/10.1371/journal.pone.0227079

Faux, A.-M., Gorjanc, G., Gaynor, R. C., Battagin, M., Edwards, S. M., Wilson, D. L., Hearne, S. J., Gonen, S., & Hickey, J. M. (2016). AlphaSim: Software for breeding program simulation. *The Plant Genome*, *9*(3), plantgenome2016.02.0013. https://doi.org/10.3835/plantgenome2016.02.0013

Gauch, H. G., & Zobel, R. W. (1997). Identifying mega-environments and targeting genotypes. *Crop Science*, *37*(2), 311–326. https://doi.org/10.2135/cropsci1997.0011183X003700020002x

Gaynor, C. (2023). *Breeding program simulations* [R package AlphaSimR version 1.5.3]. https://CRAN.R-project.org/package=AlphaSimR

Gaynor, R. C., Gorjanc, G., Bentley, A. R., Ober, E. S., Howell, P., Jackson, R., Mackay, I. J., & Hickey, J. M. (2017). A two-part strategy for using genomic selection to develop inbred lines. *Crop Science*, *57*(5), 2372–2386. https://doi.org/10.2135/cropsci2016.09.0742

Gaynor, R. C., Gorjanc, G., & Hickey, J. M. (2021). AlphaSimR: An R package for breeding program simulations. *G3: Genes, Genomes, Genetics*, *11*(2), jkaa017. https://doi.org/10.1093/g3journal/jkaa017

Gezan, S., de Oliveira, A., & Murray, D. (2021). *ASRgenomics: An R package with complementary genomic functions*. VSN International.

Heffner, E. L., Sorrells, M. E., & Jannink, J.-L. (2009). Genomic selection for crop improvement. *Crop Science*, *49*(1), 1–12. https://doi.org/10.2135/cropsci2008.08.0512

Hoyos-Villegas, V., Arief, V. N., Yang, W.-H., Sun, M., Delacy, I. H., Barrett, B. A., Jahufer, Z., & Basford, K. E. (2019). QuLine-Plus: Extending plant breeding strategy and genetic model simulation to cross-pollinated populations—Case studies in forage breeding. *Heredity*, *122*(5), 684–695. https://doi.org/10.1038/s41437-018-0156-0

Jannink, J.-L., Lorenz, A. J., & Iwata, H. (2010). Genomic selection in plant breeding: From theory to practice. *Briefings in Functional Genomics and Proteomics*, *9*(2), 166–177. https://doi.org/10.1093/bfgp/elq001

Liu, H., Tessema, B. B., Jensen, J., Cericola, F., Andersen, J. R., & Sørensen, A. C. (2019). ADAM-plant: A software for stochastic simulations of plant breeding from molecular to phenotypic level and from simple selection to complex speed breeding programs. *Frontiers in Plant Science*, *9*(January), 1926. https://doi.org/10.3389/fpls.2018.01926

Lubanga, N., Gorjanc, G., Massawe, F., Mayes, S., & Bancic, J. (2022). Genomic selection strategies to increase genetic gain in tea breeding programs. *The Plant Genome*, *16*, e20282.

Meuwissen, T., Hayes, B., & Goddard, M. (2016). Genomic selection: A paradigm shift in animal breeding. *Animal Frontiers*, *6*(1), 6–14. https://doi.org/10.2527/af.2016-0002

Meuwissen, T. H. E., Hayes, B. J., & Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, *157*(4), 1819–1829. https://doi.org/10.1093/genetics/157.4.1819

Muleta, K. T., Pressoir, G., & Morris, G. P. (2019). Optimizing genomic selection for a sorghum breeding program in Haiti: A simulation study. *G3: Genes, Genomes, Genetics*, *9*(2), 391–401. https://doi.org/10.1534/g3.118.200932

Podlich, D. W., & Cooper, M. (1998). QU-GENE: A simulation platform for quantitative analysis of genetic models. *Bioinformatics*, *14*(7), 632–653.

Pook, T., Schlather, M., & Simianer, H. (2020). MoBPS-modular breeding program simulator. *G3: Genes, Genomes, Genetics*, *10*(6), 1915–1918. https://doi.org/10.1534/g3.120.401193

Powell, O., Gaynor, R. C., Gorjanc, G., Werner, C. R., & Hickey, J. M. (2020). A two-part strategy using genomic selection in hybrid crop breeding programs. *BioRxiv*. https://www.biorxiv.org/content/10.1101/2020.05.24.113258v1.article-info

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Sabadin, F., Dovale, J. C., Platten, J. D., & Fritsche-Neto, R. (2022). Optimizing self-pollinated crop breeding employing genomic selection: From schemes to updating training sets. *Frontiers in Plant Science*, *13*, 3770. https://doi.org/10.3389/fpls.2022.935885

Sun, X., Peng, T., & Mumm, R. H. (2011). The role and basics of computer simulation in support of critical decisions in plant breeding. *Molecular Breeding*, *28*(4), 421–436. https://doi.org/10.1007/s11032-011-9630-6

Tessema, B. B., Liu, H., Sørensen, A. C., Andersen, J. R., & Jensen, J. (2020). Strategies using genomic selection to increase genetic gain in breeding programs for wheat. *Frontiers in Genetics*, *11*, 578123. https://doi.org/10.3389/fgene.2020.578123

Vanraden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science*, *91*(11), 4414–4423. https://doi.org/10.3168/jds.2007-0980

Wang, J., van Ginkel, M., Podlich, D., Ye, G., Trethowan, R., Pfeiffer, W., Delacy, I. H., Cooper, M., & Rajaram, S. (2003). Comparison of two breeding strategies by computer simulation. *Crop Science*, *43*(5), 1764–1773. https://doi.org/10.2135/cropsci2003.1764

Werner, C. R., Gaynor, R. C., Sargent, D. J., Lillo, A., Gorjanc, G., & Hickey, J. M. (2020). Genomic selection strategies for clonally propagated crops. *BioRxiv*. https://www.biorxiv.org/content/10.1101/2020.06.15.152017v1

Wolfe, M. D., Del Carpio, D. P., Alabi, O., Ezenwaka, L. C., Ikeogu, U. N., Kayondo, I. S., Lozano, R., Okeke, U. G., Ozimati, A. A., & Williams, E. (2017). Prospects for genomic selection in cassava breeding. *The Plant Genome*, *10*(3), plantgenome2017.03.0015.

Wolfe, M. D., Kulakow, P., Rabbi, I. Y., & Jannink, J.-L. (2016). Marker-based estimates reveal significant nonadditive effects in clonally propagated cassava (*Manihot esculenta*): Implications for the prediction of total genetic value and the selection of varieties. *G3: Genes, Genomes, Genetics*, *6*(11), 3497–3506. https://doi.org/10.1534/g3.116.033332

Yabe, S., Iwata, H., & Jannink, J.-L. (2017). A simple package to script and simulate breeding schemes: The breeding scheme language. *Crop*

14350653, 0, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/csc2.21170 by Nigeria Hinari NPL, Wiley Online Library on [13/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Science, *57*(3), 1347–1354. https://doi.org/10.2135/cropsci2016.06.0538

Yonis, B. O., Del Carpio, D. P., Wolfe, M., Jannink, J.-L., Kulakow, P., & Rabbi, I. (2020). Improving root characterisation for genomic prediction in cassava. *Scientific Reports*, *10*(1), 8003. https://doi.org/10.1038/s41598-020-64963-9

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Bakare, M. A., Kayondo, S. I., Kulakow, P., Rabbi, I. Y., & Jannink, J.-L. (2024). Evaluating breeding for broad versus narrow adaptation for cassava in Nigeria using stochastic simulation. *Crop Science*, 1–14. https://doi.org/10.1002/csc2.21170

14350653, 0, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/csc2.21170 by Nigeria Hinari NPL, Wiley Online Library on [13/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License