



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

## Modelling illustrates that genomic selection provides new opportunities for intercrop breeding

### Citation for published version:

Bancic, J, Werner, C, Gaynor, C, Gorjanc, G, Odeny, DA, Ojulong, H, Dawson, I, Hoad, SP & Hickey, J 2021, 'Modelling illustrates that genomic selection provides new opportunities for intercrop breeding', *Frontiers in plant science*, vol. 12, 605172. <https://doi.org/10.3389/fpls.2021.605172>

### Digital Object Identifier (DOI):

[10.3389/fpls.2021.605172](https://doi.org/10.3389/fpls.2021.605172)

### Link:

[Link to publication record in Edinburgh Research Explorer](#)

### Document Version:

Publisher's PDF, also known as Version of record

### Published In:

Frontiers in plant science

### Publisher Rights Statement:

Copyright © 2021 Bancic, Werner, Gaynor, Gorjanc, Odeny, Ojulong, Dawson, Hoad and Hickey. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

### General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

### Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.





# Modeling Illustrates That Genomic Selection Provides New Opportunities for Intercrop Breeding

Jon Bančić<sup>1,2\*</sup>, Christian R. Werner<sup>1\*</sup>, R. Chris Gaynor<sup>1</sup>, Gregor Gorjanc<sup>1</sup>, Damaris A. Odeny<sup>3</sup>, Henry F. Ojulong<sup>3</sup>, Ian K. Dawson<sup>2</sup>, Stephen P. Hoad<sup>2</sup> and John M. Hickey<sup>1</sup>

<sup>1</sup> Royal (Dick) School of Veterinary Studies, The Roslin Institute, University of Edinburgh, Easter Bush Research Centre, Midlothian, United Kingdom, <sup>2</sup> Scotland's Rural College (SRUC), Edinburgh, United Kingdom, <sup>3</sup> International Crops Research Institute for the Semi-Arid Tropics, ICRAF House, Nairobi, Kenya

## OPEN ACCESS

### Edited by:

Jérôme Enjalbert,  
Institut National de Recherche pour  
l'Agriculture, l'Alimentation et  
l'Environnement (INRAE), France

### Reviewed by:

Lizhen Zhang,  
China Agricultural University, China  
Andrés Ricardo Schwember,  
Pontificia Universidad Católica  
de Chile, Chile

### \*Correspondence:

Jon Bančić  
jon.bancic@roslin.ed.ac.uk;  
jon.bancic@gmail.com  
Christian R. Werner  
c.werner@cgiar.org

### Specialty section:

This article was submitted to  
Plant Breeding,  
a section of the journal  
Frontiers in Plant Science

**Received:** 11 September 2020

**Accepted:** 05 January 2021

**Published:** 09 February 2021

### Citation:

Bančić J, Werner CR, Gaynor RC,  
Gorjanc G, Odeny DA, Ojulong HF,  
Dawson IK, Hoad SP and Hickey JM  
(2021) Modeling Illustrates That  
Genomic Selection Provides  
New Opportunities for Intercrop  
Breeding.  
*Front. Plant Sci.* 12:605172.  
doi: 10.3389/fpls.2021.605172

Intercrop breeding programs using genomic selection can produce faster genetic gain than intercrop breeding programs using phenotypic selection. Intercropping is an agricultural practice in which two or more component crops are grown together. It can lead to enhanced soil structure and fertility, improved weed suppression, and better control of pests and diseases. Especially in subsistence agriculture, intercropping has great potential to optimize farming and increase profitability. However, breeding for intercrop varieties is complex as it requires simultaneous improvement of two or more component crops that combine well in the field. We hypothesize that genomic selection can significantly simplify and accelerate the process of breeding crops for intercropping. Therefore, we used stochastic simulation to compare four different intercrop breeding programs implementing genomic selection and an intercrop breeding program entirely based on phenotypic selection. We assumed three different levels of genetic correlation between monocrop grain yield and intercrop grain yield to investigate how the different breeding strategies are impacted by this factor. We found that all four simulated breeding programs using genomic selection produced significantly more intercrop genetic gain than the phenotypic selection program regardless of the genetic correlation with monocrop yield. We suggest a genomic selection strategy which combines monocrop and intercrop trait information to predict general intercropping ability to increase selection accuracy in the early stages of a breeding program and to minimize the generation interval.

**Keywords:** sustainable agriculture, intercrop breeding program designs, genomic selection, intercropping ability, stochastic simulation

## INTRODUCTION

Intercropping is an agricultural practice in which two or more component crops are grown together (Vandermeer, 1989). A common combination is a cereal with a legume, such as maize with beans in Latin America (Zimmermann, 1996), and millet/maize/sorghum with pigeon pea in India and East Africa (Dass and Sudhishri, 2010; Kiwia et al., 2019). Intercropping can lead to enhanced soil structure and fertility, the conservation of soil moisture, improved weed suppression, and better

control of pests and diseases, enabling greater yields and higher profitability (Brooker et al., 2015; Litrico and Violle, 2015). Recent meta-analysis showed low- and high- input intercropping systems on average delivered 16 to 29% more grain per hectare while using 19 to 36% less fertilizer per unit output than in conventional monocrop production, respectively (Li et al., 2020). Intercropping also allows for simultaneous cultivation of crops with different nutritional profiles, which can contribute to improving diets (Dawson et al., 2019a) and to increasing the stability and resilience of food systems (Himmelstein et al., 2017; Raseduzzaman and Jensen, 2017). Due to these characteristics, intercropping has great potential to optimize farming, especially in subsistence agricultural systems, which has recently led to an increased interest in the development and evaluation of efficient intercrop production (Dawson et al., 2019b).

Despite the potential benefits of intercropping, intercrop breeding has received only very little attention to date, with varieties specifically bred for intercrop production being unavailable (Brooker et al., 2015; Litrico and Violle, 2015). This lack of attention is due to in large part two reasons:

- (i) In advanced economies, major global crop species are predominantly grown as monocrops (Leff et al., 2004) and the majority of breeding programs are focused on generating varieties adapted to monocrop production (Acquaah, 2012).
- (ii) Intercrop breeding is more complex than monocrop breeding. Breeding for intercrop production requires the optimization of two or more component crops simultaneously (Francis, 1981; Wright, 1985); intercrop varieties ideally exhibit both a high *per se* performance and combine well with the other component crop(s) (Davis and Woolley, 1993).

As a result, the literature on intercrop breeding methodology is rare (Hamblin et al., 1976; Wright, 1985; Hill, 1996) and almost no progress in approaches has been made over the last few decades, with one recent exception (Sampoux et al., 2020). The crop varieties currently used for intercropping have typically been bred for monocrop production, and most often their performance in intercropping has not even been evaluated in advance (Brooker et al., 2015), strongly restricting the potential benefits of this practice.

Genomic selection offers many opportunities to address the complexity of intercrop breeding programs and aid the simultaneous improvement of two or more component crops that combine well in the field. Genomic selection uses associations between genome-wide markers and phenotypic performance to predict the value of genotypes based on their genomic markers (Meuwissen et al., 2001; Lorenz et al., 2011; Hickey et al., 2017). In the context of an intercrop breeding program, genomic selection could be used in several ways to increase the rate of genetic gain:

- (i) Selection accuracy can be increased for individual performance and combined performance of the component crops in an intercrop.

- (ii) The generation interval can be reduced, since new crossing parents can be selected based on their genomic predicted values as soon as they are genotyped.
- (iii) Selection intensity can be increased, since thousands of potential intercrop combinations could be evaluated without testing all of them in the field.

This study aimed to test whether genomic selection can speed up and better facilitate the process of breeding for intercropping. To test this, we compared four different intercrop breeding programs which implemented genomic selection to an intercrop breeding program using only phenotypic selection, using a stochastic simulations approach. We assumed three different levels of genetic correlation between monocrop grain yield and intercrop grain yield to investigate how the different breeding strategies are affected. We found all four breeding programs using genomic selection produced significantly more intercrop genetic gain than the phenotypic selection program, regardless of the genetic correlation. The combination of monocrop and intercrop trait information in genomic models to predict general intercropping ability seems to be the best strategy to both increase selection accuracy in the early stages of a breeding program and to minimize the generation interval.

## MATERIALS AND METHODS

Stochastic simulations were used to evaluate the potential of genomic selection for intercrop breeding. We compared four different intercrop breeding programs implementing genomic selection and an intercrop breeding program using phenotypic selection for long-term efficacy for maximizing intercrop grain yield. Below, we have subdivided Material and Methods into three sections that describe: first, the simulation of the founder genotype population; second, the simulation of the recent (burn-in) breeding phase using a phenotypic selection breeding program; and third, the simulation of the future breeding phase to compare four different genomic selection breeding programs to the phenotypic selection breeding program. These topics are briefly reviewed below before detail is provided.

Simulation of the founder genotype population:

- (i) Genome simulation: a genome sequence was simulated for two hypothetical component crops in intercrop production.
- (ii) Simulation of founder genotypes: the simulated genome sequences were used to generate a base population of 100 founder genotypes for each of the two component crops.
- (iii) Simulation of genetic values: for each of the two component crops, two traits were simulated, representing monocrop grain yield and intercrop grain yield. Genetic values for the two traits were calculated by summing the additive genetic effects for both traits at 10,000 quantitative trait nucleotides and three different genetic correlations (0.4, 0.7, and 0.9) were simulated.
- (iv) Simulation of phenotypes: phenotypes were simulated for monocrop grain yield and intercrop grain yield.

Phenotypes representing monocrop grain yield were generated by adding random error to the genetic values for monocrop grain yield. Phenotypes representing intercrop grain yield were generated by adding random error to the mean genetic values for intercrop grain yield of two genotypes from both component crops.

Simulation of the recent (burn-in) breeding phase: A phenotypic selection breeding program was simulated for 20 years (burn-in) to provide a common starting point for the comparison of the different intercrop breeding programs during the future breeding phase.

Simulation of the future breeding phase: Four different genomic selection breeding programs were simulated and compared to the phenotypic selection breeding program for an additional 20 years of future breeding. The four genomic selection breeding programs included three variations of a Conventional genomic selection breeding program and a Grid genomic selection breeding program.

## Simulation of the Founder Genotype Population

### Genome Simulation

Genome sequences were simulated for two hypothetical component crops used in intercropping. For modeling purposes, the two crops' genomes were assumed to have the same characteristics. Each genome sequence consisted of 10 chromosome pairs. Each chromosome had a genetic length of 1.43 Morgans and a physical length of  $8 \times 10^8$  base pairs. The chromosome sequences were generated using the Markovian coalescent simulator (MaCS, Chen et al., 2009) implemented in AlphaSimR (R Core Team, 2019; Gaynor et al., 2020). Recombination rate was derived as the ratio between genetic length and physical length (i.e., 1.43 Morgans/ $8 \times 10^8$  base pairs =  $1.8 \times 10^{-9}$  per base pair). The per-site mutation rate was set to  $2 \times 10^{-9}$  per base pair. Effective population size was set to 50, with linear piecewise increases up to 32,000 at 100,000 generations ago, as described by Gaynor et al. (2017).

### Simulation of Founder Genotypes

The simulated genome sequences were used to generate a base population of 100 founder genotypes in Hardy-Weinberg equilibrium, for each of the two component crop species. These genotypes were formed by randomly sampling 10 chromosome pairs per genotype. A set of 1,000 bi-allelic quantitative trait nucleotides (QTNs) and 2,000 single nucleotide polymorphisms (SNPs) were randomly selected along each chromosome. This was done to simulate the structure of a quantitative trait that was controlled by 10,000 QTN and a SNP marker array with 20,000 genome-wide SNP markers. The founder genotypes were converted to doubled haploids (DH) and served as initial parents in the burn-in phase.

### Simulation of Genetic Values

For each of the two component crops, two traits were simulated:

- (i) Monocrop grain yield, representing the yield of a genotype under monocrop production.
- (ii) Intercrop grain yield, representing the total yield of two genotypes, each from one of the two component crops, under intercrop production.

Genetic values for the two traits were calculated by summing the additive genetic effects for both traits across all 10,000 QTN. Additive effects were sampled from a standard normal distribution and scaled to obtain an additive variance of  $\sigma_A^2 = 1$  in the founder population, as described in detail in the vignette of the AlphaSimR package (Gaynor et al., 2020).

Three different genetic correlations (0.4, 0.7, and 0.9) were simulated to represent different degrees of genotype-by-cropping interaction (Davis and Woolley, 1993).

### Simulation of Phenotypes

Phenotypes were simulated for monocrop grain yield and for intercrop grain yield. Phenotypes for monocrop grain yield were generated by adding random error to the genetic values for monocrop grain yield. The random error was sampled from a normal distribution with mean zero and error variance  $\sigma_e^2$ , defined by the target level of heritability at each stage of a breeding program. Entry-mean values for narrow-sense heritability ( $h^2$ ) in the founder population were set to 0.1 in the doubled haploid stage and 0.33 in the preliminary yield trial stage. Narrow-sense heritabilities in later breeding stages increased as a result of an increased number of replicates ( $r$ ) per genotype (Tables 1, 2).

Phenotypes for intercrop grain yield were generated by adding random error to the mean genetic values for intercrop grain yield of two genotypes from the two component crops. The following equation was used to calculate intercrop grain yield:

$$y_{i,j}^{ic} = \frac{a_{A,i}^{ic} + a_{B,j}^{ic}}{2} + e_{i,j},$$

where  $y_{i,j}^{ic}$  is the intercrop grain yield,  $a_{A,i}^{ic}$  is the genetic value for intercrop grain yield of genotype  $i$  from component crop A,  $a_{B,j}^{ic}$  is the genetic value for intercrop grain yield of genotype  $j$  from component crop B, and  $e_{i,j}$  is the random error. The random error for intercrop grain yield was also sampled from a normal distribution with mean zero and error variance defined by the target level of narrow-sense heritability at each stage of a breeding program.

Narrow-sense heritabilities for monocrop grain yield and intercrop grain yield were calculated as (Bernardo, 2010):

$$h^2 = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_e^2/r}.$$

## Simulation of the Recent (Burn-in) Breeding Phase

An intercrop breeding program using phenotypic selection was simulated to provide a common starting point (burn-in) for the comparison of the five intercrop breeding programs during the future breeding phase. The simulation modeled 20 years of

**TABLE 1** | Summary of *per stage* parameters and annual operational costs for four 'medium' breeding programs.

Year	Stage	Reps	$h^2$	Pheno	Baseline-GS	PYT-GS	DH-GS
1	Cross			100 × 50 DH	100 × 47 DH	100 × 47 DH	40 × 80 DH
2	DH	1	0.10	5,000	4,700	4,700	3,200
3	PYT	4	0.33	500	500	500	500
4	GIA 1	4	0.33	50 <sup>1</sup>	50 <sup>1</sup>	50 <sup>1</sup>	50 <sup>1</sup>
5	GIA 2	24	0.50	13 <sup>2</sup>	13 <sup>2</sup>	13 <sup>2</sup>	13 <sup>2</sup>
6	SIA 1	16	0.67	9*	9*	9*	9*
7	SIA 2	32	0.80	3*	3*	3*	3*
Cost (United States \$)				493,200	492,200	492,200	495,200

Cost is set at approximately United States \$500K. The cost breakdown is shown for the Phenotypic selection breeding program (Pheno) and the three Conventional genomic selection breeding programs.

Baseline-GS, the Baseline genomic selection breeding program; PYT-GS, the Preliminary yield trial genomic selection breeding program; DH-GS, the Doubled haploid genomic selection breeding program; Reps, the effective number of replications (i.e., locations);  $h^2$ , narrow-sense heritability; DH, the doubled haploid stage; PYT, the preliminary yield trial stage; GIA 1 and 2, the general intercropping ability stages 1 and 2; SIA 1 and 2, the specific intercropping ability stages 1 and 2. <sup>1</sup> denotes testing with one probe variety; <sup>2</sup> denotes testing with three probe varieties; \*number of specific intercrop combinations.

phenotypic selection to reflect prior breeding that has taken place in the two component crops.

In brief, the four key features of the phenotypic selection breeding program (**Figure 1**) were:

- (i) A crossing block of 80 DH lines was used to develop 100 bi-parental populations each year for each of the two component crops.
- (ii) New DH lines were developed from each bi-parental cross.
- (iii) A 2-year monocrop testing phase, in which monocrop grain yield was evaluated for each of the two component crop species.
- (iv) A 4-year intercrop testing phase, in which intercrop grain yield was evaluated for each intercrop combination of two genotypes from the two component crops. New parents were selected in the second year of the intercrop testing phase, giving a generation interval of 5 years.

The time from crossing to the release of a pair of component crop varieties for intercrop production was 7 years.

In what follows, the four key features of the phenotypic selection breeding program are explained in more detail. Simulation parameters, including heritability, the number of trial replications, and the number of tested genotypes at each stage

of the breeding program, are shown in **Table 1**. In the context of the phenotypic selection breeding program for intercropping varieties, we introduce the following terms:

- (i) Probe variety: represents a genotype of one of the two component crops that has good intercropping ability with genotypes from the other component crop. It is comparable to a tester in a hybrid breeding program, which is used to evaluate the general combining ability of genotypes from one heterotic pool with another heterotic pool.
- (ii) General intercropping ability (GIA): the average intercrop grain yield of a genotype from one component crop grown with genotypes from the other component crop. It is evaluated using one or several probe genotypes from the other component crop.
- (iii) Specific intercropping ability (SIA): the intercrop grain yield of a specific intercrop combination of two genotypes from the two component crops.

### Crossing Block (Year 1)

Each year, for each crop, a crossing block of 80 DH lines was used to produce 100 bi-parental crosses (**Table 1**). Parental combinations were chosen at random from all 3,160 possible pairwise combinations.

### Development of Doubled Haploids (Year 2)

From each bi-parental cross, 50 DH lines were produced for each of the two component crops. The resulting 5,000 DH lines per crop were advanced to the monocrop testing phase and tested in the same year.

### Monocrop Testing Phase (Years 2 and 3)

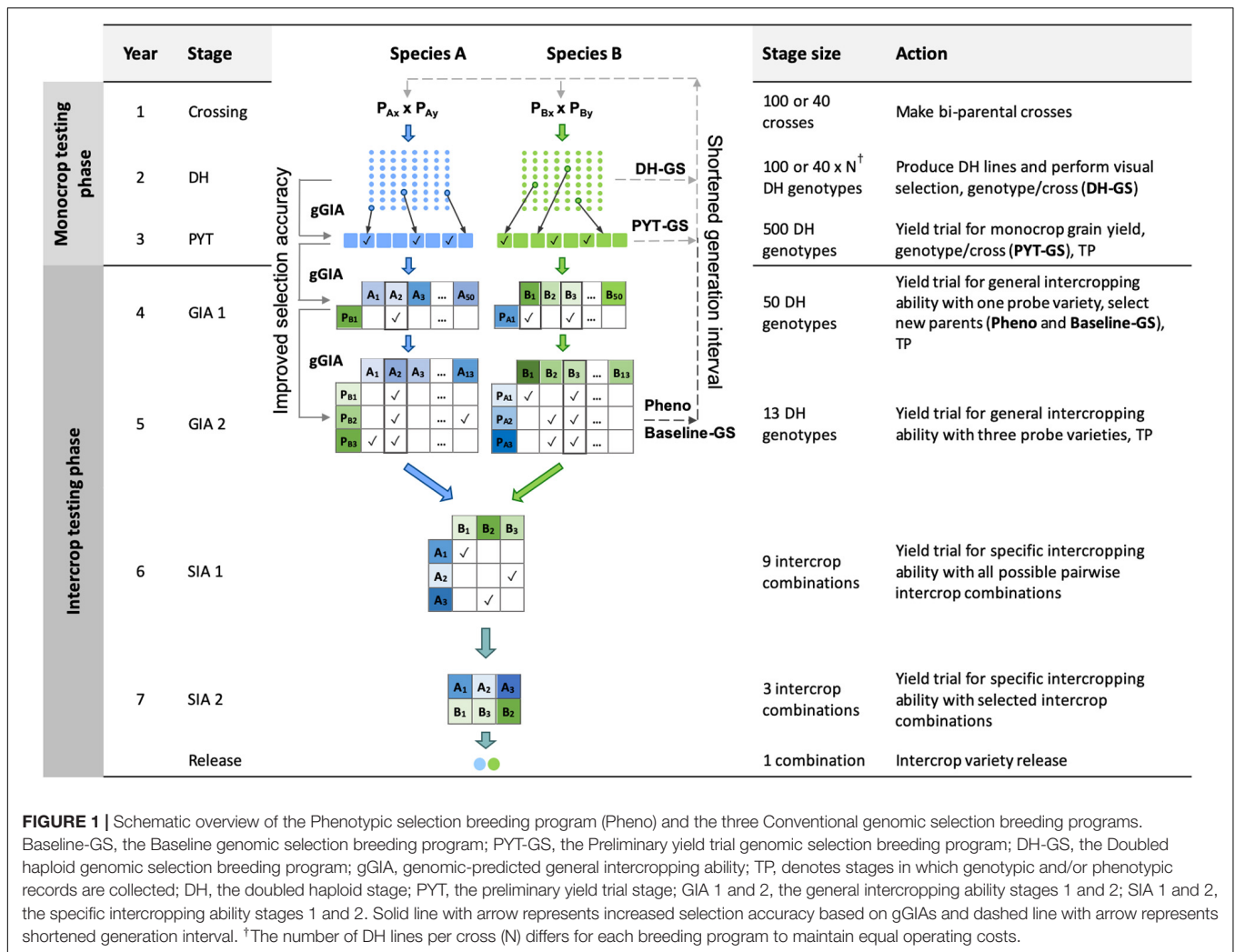
The monocrop testing phase spanned 2 years. Performance was evaluated as monocrop grain yield. Monocrop testing included the doubled haploid stage (DH stage, year 2) and the preliminary yield trial stage (PYT stage, year 3). In the DH stage, seed was increased and phenotypic selection was based on single plants within each bi-parental cross, to ensure there was variation in the later stages. In the PYT stage, phenotypic selection was based on multi-location trial plots (**Table 1**).

**TABLE 2** | Summary of *per stage* parameters and annual operational costs for the 'medium' Grid genomic selection breeding program (Grid-GS).

Year	Stage	Reps	$h^2$	Grid-GS
1	Cross			40 × 90
2	DH	1	0.10	3,600
3	Grid	1	0.10	900/250,000*
4	SIA1	16	0.67	50*
5	SIA2	32	0.80	8*
Cost (United States \$)				493,800

Cost is set at approximately United States \$500K.

Reps, the effective number of replications (i.e., locations);  $h^2$ , narrow-sense heritability; DH, the doubled haploid stage; Grid, the Grid stage; SIA 1 and 2, the specific intercropping ability stages 1 and 2; \*number of specific intercrop combinations.



### Intercrop Testing Phase (Years 4 to 7)

The intercrop testing phase spanned 4 years. Performance was evaluated as intercrop grain yield, i.e., the total yield from both simultaneously grown genotypes of the two component crops. The intercrop testing phase included two general intercropping ability testing stages (GIA1, year 4; and GIA2, year 5) and two specific intercropping ability testing stages (SIA1, year 6; and SIA2, year 7).

In the GIA1 stage, phenotypic selection was based on general intercropping ability in a yield trial with one probe variety. In the GIA2 stage, phenotypic selection was based on general intercropping ability using three probe varieties. Each year, the four best performing genotypes from the GIA2 stage replaced the probe varieties from the previous year. New parents were selected in the GIA1 stage. Each year, the 20 best performing genotypes from the GIA1 stage were used to replace the 20 oldest parents in the crossing block. Hence, every genotype stayed in the crossing block for four crossing cycles. The generation interval was 5 years.

In the SIA1 stage, all possible pairwise combinations of the selected lines were tested in a yield trial. In the SIA2 stage, the best combinations from the SIA1 stage were tested in a multi-location

trial (Table 1). The highest yielding intercrop combination was then released as an intercrop variety combination.

### Simulation of the Future Breeding Phase

The future breeding phase was used to evaluate the phenotypic selection breeding program and the four genomic selection breeding programs for an additional 20 years of breeding. The genomic selection breeding programs included three variations of a Conventional genomic selection breeding program and a Grid-GS breeding program. The three Conventional genomic selection breeding programs replaced phenotypic selection by genomic selection at different stages of the phenotypic selection breeding program (Figure 1). They comprised a Baseline-GS, a PYT-GS and a DH-GS breeding program. The Grid-GS breeding program reorganized the phenotypic selection breeding program to enable the evaluation of a greater number of specific intercrop combinations using genomic selection. Details on all programs are given below.

In order to obtain approximately equal annual operating costs, the number of doubled haploids per bi-parental cross was reduced in the genomic selection breeding programs to

compensate for additional costs due to genotyping. **Table 1** shows the resources used for the phenotypic selection breeding program and the three Conventional genomic selection breeding programs. **Table 2** shows the resources used for the Grid-GS breeding program. Estimated applied costs in calculations were \$20 per monocrop test plot, \$50 per intercrop test plot, \$35 for producing a doubled haploid line and \$20 for producing a single genotype by array genotyping. The former two values were based on consultations with breeders at ICRISAT and the latter two values were previously used by Gaynor et al. (2017).

### The Baseline Genomic Selection Breeding Program (Baseline-GS)

In the Baseline-GS breeding program, genomic selection was used to replace phenotypic selection in the PYT stage and in the GIA1 stage. Each year, the best 20 genotypes from the GIA1 stage were selected as new parents using genomic selection to replace the oldest 20 parents in the crossing block. As for the phenotypic selection breeding program, the generation interval was 5 years.

### The Preliminary Yield Trial Genomic Selection Breeding Program (PYT-GS)

In the PYT-GS breeding program, genomic selection was used to replace phenotypic selection in the PYT stage and in the GIA1 stage. Each year, the best 80 genotypes from the PYT stage and last year's crossing block were selected as new parents using genomic selection.

### The Doubled Haploid Genomic Selection Breeding Program (DH-GS)

In the DH-GS breeding program, genomic selection was used to replace phenotypic selection in the DH stage, the PYT stage and the GIA1 stage. Each year, 80 genotypes from the DH stage were selected as new parents to replace the entire last year's crossing block. As preliminary results showed a rapid decrease in genetic variance when the best 80 genotypes were selected, we implemented a maximum avoidance crossing scheme using genomic selection to reduce the rate of decrease (Kimura and Crow, 1963).

### Genomic Selection Model for the Three Conventional Genomic Selection Breeding Programs

The three Conventional genomic selection breeding programs (above) used a multivariate ridge regression genomic selection model (RR-BLUP) to obtain genomic predictions of general intercropping abilities (gGIA) for each component crop separately.

In this model, monocrop grain yield from the PYT stage and mean intercrop grain yield with one or three probes, respectively, from the GIA1 and the GIA2 stage were fitted simultaneously. Genomic predictions of general intercropping ability can be directly calculated using intercrop grain yield from the GIA1 and the GIA2 stage as phenotypic information. In addition, the multivariate model uses information on monocrop grain yield, which was included as a correlated trait.

The following model was used:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{a} + \mathbf{e},$$

expanded in matrix form as:

$$\begin{bmatrix} \mathbf{y}_m \\ \mathbf{y}_{ic1} \\ \mathbf{y}_{ic2} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_m & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_{ic1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{X}_{ic2} \end{bmatrix} \begin{bmatrix} \mathbf{b}_m \\ \mathbf{b}_{ic1} \\ \mathbf{b}_{ic2} \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_m & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_{ic1} \\ \mathbf{0} & \mathbf{Z}_{ic2} \end{bmatrix} \begin{bmatrix} \mathbf{a}_m \\ \mathbf{a}_{ic} \end{bmatrix} + \begin{bmatrix} \mathbf{e}_m \\ \mathbf{e}_{ic1} \\ \mathbf{e}_{ic2} \end{bmatrix},$$

where  $\mathbf{y}_m$ ,  $\mathbf{y}_{ic1}$  and  $\mathbf{y}_{ic2}$  respectively denote the vectors of monocrop grain yield from the PYT stage, and mean intercrop grain yield with one or three probes from the GIA1 and the GIA2 stage;  $\mathbf{b}_m$ ,  $\mathbf{b}_{ic1}$  and  $\mathbf{b}_{ic2}$  respectively denote the vectors for the fixed effects of year and stage for PYT, GIA1, and GIA2;  $\mathbf{a}_m$  and  $\mathbf{a}_{ic}$  respectively denote the vectors of the marker effects for monocrop grain yield and intercrop grain yield;  $\mathbf{X}_m$ ,  $\mathbf{X}_{ic1}$ ,  $\mathbf{X}_{ic2}$ ,  $\mathbf{Z}_m$ ,  $\mathbf{Z}_{ic1}$  and  $\mathbf{Z}_{ic2}$  denote the corresponding incidence matrices; and  $\mathbf{e}_m$ ,  $\mathbf{e}_{ic1}$  and  $\mathbf{e}_{ic2}$  denote the corresponding vectors of residuals.

Additive genetic (G) and residual (R) variance-covariance matrices were:

$$\mathbf{G} = \text{var} \begin{bmatrix} \mathbf{a}_m \\ \mathbf{a}_{ic} \end{bmatrix} = \begin{bmatrix} \sigma_{A_m}^2 & \sigma_{A_m,ic} \\ \sigma_{A_m,ic} & \sigma_{A_{ic}}^2 \end{bmatrix},$$

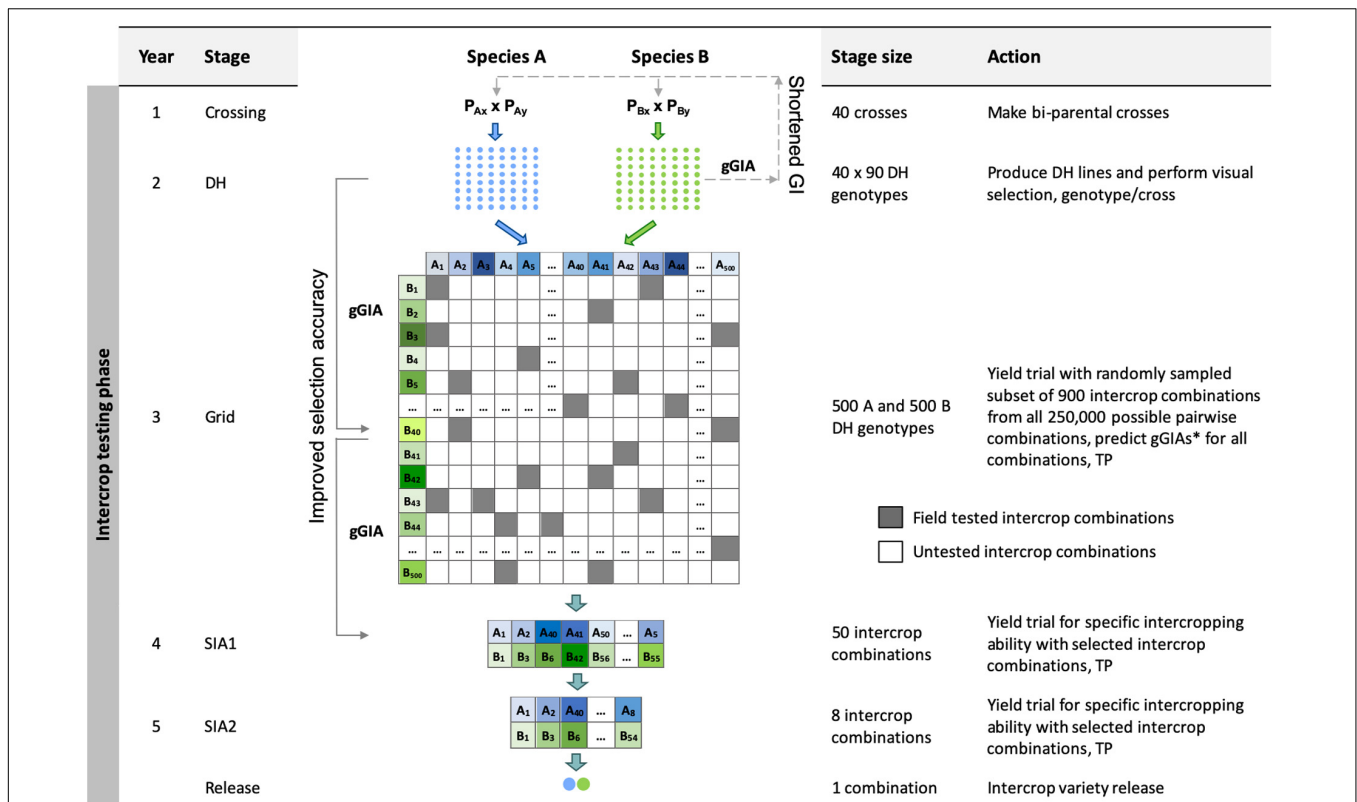
$$\mathbf{R} = \text{var} \begin{bmatrix} \mathbf{e}_m \\ \mathbf{e}_{ic1} \\ \mathbf{e}_{ic2} \end{bmatrix} = \begin{bmatrix} \sigma_e^2/r_m & 0 & 0 \\ 0 & \sigma_e^2/r_{ic1} & 0 \\ 0 & 0 & \sigma_e^2/r_{ic2} \end{bmatrix},$$

where  $\sigma_{A_m}^2$  and  $\sigma_{A_{ic}}^2$  respectively denote the additive genetic variances for monocrop grain yield and intercrop grain yield, and  $\sigma_{A_m,ic}$  denotes the additive genetic covariance between the two traits; and  $\sigma_e^2$  denotes the residual variance.  $\mathbf{R}$  modeled heterogeneous residual variances by weighting  $\sigma_e^2$  for the effective number of replications ( $r$ ) in a particular stage (**Table 2**). To reduce computation time, additive genetic variances were assumed known and calculated each year using the true additive genetic effects.

The initial training population at the start of the future breeding phase consisted of all genotypes from the PYT stage of the last 5 years of the burn-in phase. This training population consisted of 2,500 genotypes and 2,739 phenotypic records from the PYT, the GIA1 and the GIA2 stages. In every year of the future breeding phase, 500 new genotypes from the PYT stage were added to the training population, as well as 500, 50 and 13 new phenotypic records from the PYT, the GIA1 and the GIA2 stages, respectively. The training population was updated using a 5-year sliding window approach, in which it always contained the most recent 5 years of training data.

### Grid Genomic Selection Breeding Program (Grid-GS)

The Grid genomic selection breeding program reorganized the phenotypic selection breeding program to enable the evaluation of a greater number of specific intercrop combinations using genomic selection (**Figure 2**). To achieve this, the PYT, the GIA1 and GIA2 stages were replaced by a single 'grid' stage.



**FIGURE 2 |** Schematic overview of the Grid genomic selection breeding program (Grid-GS). gGIA, genomic-predicted general intercropping ability; TP, denotes stages in which genotypic and/or phenotypic records are collected; DH, the doubled haploid stage; Grid, the grid stage; SIA 1 and 2, the specific intercropping ability stages 1 and 2. Solid line with arrow represents increased selection accuracy based on gGIAs and dashed line with arrow represents shortened generation interval.

The reorganized program design also tested an increased number of specific intercrop combinations at the SIA1 and the SIA2 stages (Table 2).

The grid stage involved field testing of 900 intercrop combinations. At first, genomic prediction of general intercropping ability (gGIA) was used at the DH stage to select the best 500 DH lines from each component crop. From all 250,000 possible intercrop combinations between the 500 DH lines from each component crop, 900 were randomly sampled for field testing at a single location (Table 2).

Genomic predictions of general intercropping ability were calculated for all 250,000 intercrop combinations using the following equation:

$$gGIA_{i,j}^{ic} = \frac{gGIA_{A,i} + gGIA_{B,j}}{2},$$

where  $gGIA_{i,j}^{ic}$  is the mean genomic-predicted general intercropping ability; and  $gGIA_{A,i}$  and  $gGIA_{B,j}$  are respectively the genomic-predicted general intercropping abilities of the  $i$ -th and  $j$ -th genotypes of component crops A and B. The best 50 predicted intercrop combinations were then advanced to the SIA1 stage (compared to nine intercrop combinations in our four other breeding programs).

Each year, 80 genotypes from the DH stage were selected as new parents to replace the entire last year’s crossing block. As

preliminary results showed a rapid decrease in genetic variance when the best 80 genotypes were selected, we implemented a maximum avoidance crossing scheme with genomic selection to reduce the rate of decrease (Kimura and Crow, 1963). The generation interval was 2 years and the total length of the breeding program from initial crosses to the release of the intercrop variety pair was 5 years.

### Genomic Selection Model for the Grid-GS Breeding Program

Genomic predictions of general intercropping ability were calculated using a ridge regression model (RR-BLUP) which predicted marker effects for both component crops simultaneously based on intercrop grain yield.

The following model was used:

$$y = Xb + Z_A a_A + Z_B a_B + e,$$

expanded in matrix form as:

$$y_{ic} = Xb + \begin{bmatrix} Z_A & 0 \\ 0 & Z_B \end{bmatrix} \begin{bmatrix} a_A \\ a_B \end{bmatrix} + e_{ic},$$

where  $y_{ic}$  denotes the vector of intercrop grain yield from the grid, SIA1 and SIA2 stages;  $b$  denotes the vector for fixed effects of year and stage for grid, SIA1, and SIA2;  $a_A$  and  $a_B$  respectively



denote the vectors of the marker effects for intercrop grain yield in component crops A and B;  $\mathbf{X}$ ,  $\mathbf{Z}_A$  and  $\mathbf{Z}_B$  denote the incidence matrices; and  $\mathbf{e}_{ic}$  denotes the vector of residual effects.

The residual ( $\mathbf{R}$ ) variance-covariance matrix modeled heterogeneous residual variances by weighting  $\sigma_e^2$  for the effective number of replications ( $r$ ) in a particular stage (**Supplementary Table 4**), as described for the genomic selection model used in the Conventional genomic selection breeding programs. To reduce computation time, additive genetic variances were assumed known and calculated each year using the true additive genetic effects.

To initialize the training population, the grid stage was already simulated during the last 5 years of the burn-in breeding phase. For each of the 5 years, the best 23 genotypes at the PYT stage for each component crop were selected based on their genomic-predicted general intercropping abilities. These selected genotypes were then used to generate all 529 possible intercrop combinations, which were then tested in the field. At the beginning of the future breeding phase, the initial training population thus consisted of 115 genotypes from each component crop and 2,645 intercrop grain yield phenotypes ( $5 \times 529$  different intercrop combinations). In every year of the future breeding phase, 500 new genotypes from each component crop were added to the training population, as well as 900, 50, and 8 intercrop grain yield records, respectively, from the grid, SIA1 and SIA2 stages. The training population was updated using a 5-year sliding window approach, in which the training population always contained the most recent 5 years of data. This training population contained a total of 2,500 genotypes and 4,790 phenotypic records.

## Comparison of the Intercrop Breeding Program Designs

The performance of the five intercrop breeding programs (the four using genomic selection and the phenotype-alone comparison) was evaluated by measuring the mean intercrop genetic value over time in the DH stage of both component crops as follows:

$$\bar{a}_{A,B}^{ic} = \frac{\bar{a}_A^{ic} + \bar{a}_B^{ic}}{2},$$

with  $\bar{a}_A^{ic}$  and  $\bar{a}_B^{ic}$  being the mean intercrop genetic values of the genotypes in the DH stage from component crops A and B, respectively. Mean intercrop genetic values of the two component crops were centered at 0 for the last year of the burn-in breeding phase. Intercrop genetic variance was measured as variance of the mean intercrop genetic values. Direct comparisons between breeding program designs for intercrop genetic gain and intercrop genetic variance were reported as ratios. These were calculated by performing a paired  $t$ -test (Welch) on log-transformed values from the 30 simulation replicates; the log-transformed differences from the  $t$ -test were then back-transformed to obtain ratios (**Supplementary Table 5**; Gaynor et al., 2017).

Prediction accuracy was evaluated as the correlation coefficient between the true and predicted performance at the DH stage. In the phenotypic selection breeding program,

the phenotype served as a predictor of intercropping ability. Prediction accuracy was evaluated as the correlation between the phenotypic value (i.e., monocrop grain yield) and true intercrop genetic value. In all four genomic selection breeding programs, prediction accuracy was measured as the correlation between the genomic-predicted general intercropping ability and the true intercrop genetic value of the doubled haploids.

Comparisons of the five breeding programs were done under three different levels of annual operating budget (see the start of Simulation of the Future Breeding Phase): (i) a 'large' budget (United States \$1M); (ii) a 'medium' budget (United States \$500K); and (iii) a 'small' budget (United States \$250K). Since the results for all our breeding programs showed similar rankings across these budgets, the methods presented above and the results presented below are described only for the medium budget scenario (**Tables 1, 2**). The parameters applied for breeding program designs and the results of the simulations at other budget levels are presented in the **Supplementary Materials (Supplementary Tables 1–4 and Supplementary Figures 1, 2)**.

## RESULTS

Our results show that intercrop breeding programs using genomic selection can produce faster genetic gain than an intercrop breeding program using only phenotypic selection. All four breeding programs using genomic selection produced more intercrop genetic gain than the phenotypic selection breeding program ( $\sim 1.3$ – $2.5$  times), regardless of the genetic correlation between monocrop grain yield and intercrop grain yield. However, the three Conventional genomic selection breeding programs produced increasingly more genetic gain when the genetic correlation between monocrop grain yield and intercrop grain yield increased, while the Grid-GS breeding program produced slightly less genetic gain when the genetic correlation increased. The DH-GS breeding program always produced the most genetic gain among the three Conventional genomic selection breeding programs (2.1 and 2.5 times the gain of the phenotypic selection breeding program at correlations of 0.4 and 0.9, respectively). Intercrop breeding using genomic selection also gave a faster reduction in genetic variance than intercrop breeding with phenotypic selection, regardless of the genetic correlation between monocrop yield and intercrop yield. Selection accuracy for intercropping ability was higher when genomic selection was compared to phenotypic selection. Selection accuracy in the three Conventional genomic selection breeding programs and the phenotypic selection breeding program increased when the genetic correlation between monocrop grain yield and intercrop grain yield increased, while selection accuracy in the Grid-GS breeding program was similar under different levels of genetic correlation. The general trends and rankings observed under the medium annual budget were representative of the trends observed under low and high annual budgets. Our findings are discussed in more detail below in terms of gain, genetic variance and selection accuracy.

## Intercrop Genetic Gain

Intercrop breeding using genomic selection produced faster genetic gain than intercrop breeding with phenotypic selection. This is shown in **Figure 3**, which plots intercrop genetic gain as mean intercrop genetic value in the DH stage for the entire future breeding phase. The three panels show intercrop genetic gain for the five simulated breeding programs under different levels of genetic correlation between monocrop grain yield and intercrop grain yield. All four breeding programs using genomic selection produced significantly more intercrop genetic gain than the phenotypic selection program under all three levels of genetic correlation between monocrop yield and intercrop yield.

**Figure 3** also shows that the three Conventional genomic selection breeding programs produced increasingly more genetic gain when the genetic correlation between monocrop grain yield and intercrop grain yield increased, while the Grid-GS breeding program produced slightly less genetic gain when the genetic correlation increased. As a result, the ranking of the four genomic selection breeding programs for genetic gain was dependent on the level of genetic correlation. When the genetic correlation was low (0.4), the Grid-GS breeding program produced the most genetic gain over time, closely followed by the DH-GS breeding program. Both breeding programs produced more than twice the genetic gain of the phenotypic selection breeding program. However, when the genetic correlation was high (0.9), the Grid-GS breeding program produced less genetic gain than all three Conventional genomic selection breeding programs. It generated 1.3 times the gain of the phenotypic selection breeding program,

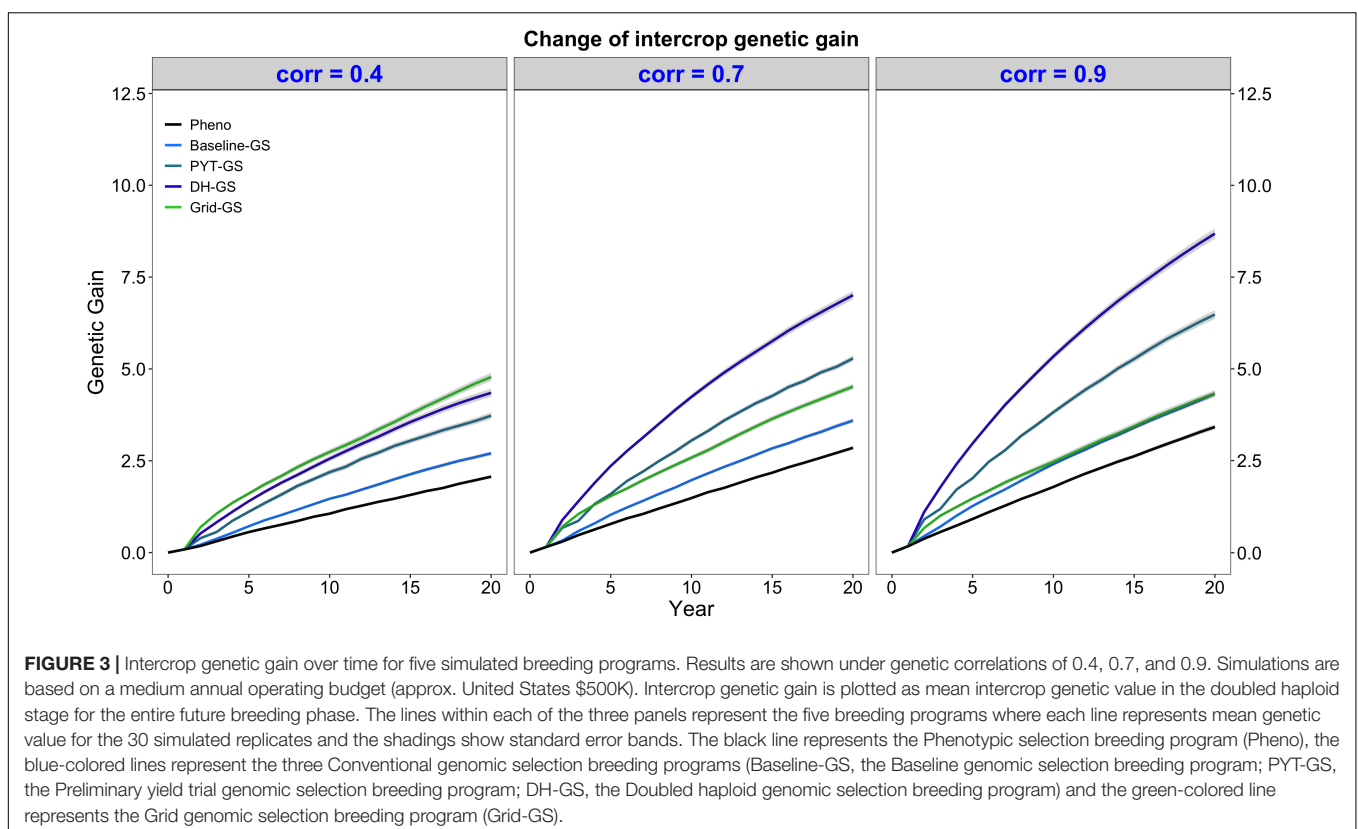
while the DH-GS breeding program produced 2.5 times the genetic gain of the phenotypic selection breeding program.

**Figure 3** also shows that the DH-GS breeding program always produced the most genetic gain of the three Conventional genomic selection breeding programs, followed by PYT-GS and Baseline-GS breeding programs. The relative performance of the DH-GS breeding program compared to the other two Conventional genomic selection breeding programs increased when the genetic correlation between monocrop grain yield and intercrop grain yield increased. When the genetic correlation was low (0.4), the DH-GS breeding program generated 1.2 times the genetic gain of the PYT-GS breeding program and 1.6 times the gain of the Baseline-GS breeding program. When the genetic correlation was high (0.9), it generated 1.3 times the genetic gain of the PYT-GS breeding program and twice the gain of the Baseline-GS breeding program.

All breeding programs produced more genetic gain when the annual operating budget was high (**Supplementary Table 5** and **Supplementary Figure 2a**) and less genetic gain when the annual operating costs were low (**Supplementary Table 5** and **Supplementary Figure 1a**). The general trends and rankings observed under the medium annual budget were however representative of the trends observed under low and high annual budgets.

## Intercrop Genetic Variance

Intercrop breeding using genomic selection gave a faster reduction in genetic variance than intercrop breeding with



phenotypic selection. This is shown in **Figure 4**, which plots the genetic variance of the intercrop genetic values in the DH stage for the entire future breeding phase. All four breeding programs using genomic selection gave a faster reduction in genetic variance than the phenotypic selection breeding program under all three levels of genetic correlation between monocrop yield and intercrop yield.

**Figure 4** also shows that the Grid-GS breeding program gave the fastest reduction in genetic variance at the end of the future breeding phase under all three levels of genetic correlation. The Baseline-GS breeding program gave the slowest reduction in genetic variance among the four breeding programs using genomic selection. The DH-GS and the PYT-GS breeding programs always showed a similar reduction in genetic variance and ranked between the other two breeding programs using genomic selection. However, these two breeding programs became more similar to the Grid-GS breeding program as the genetic correlation increased. When the genetic correlation was high (0.9), the Grid-GS, the PYT-GS and the DH-GS breeding programs all showed a similar reduction in genetic variance at the end of the future breeding phase.

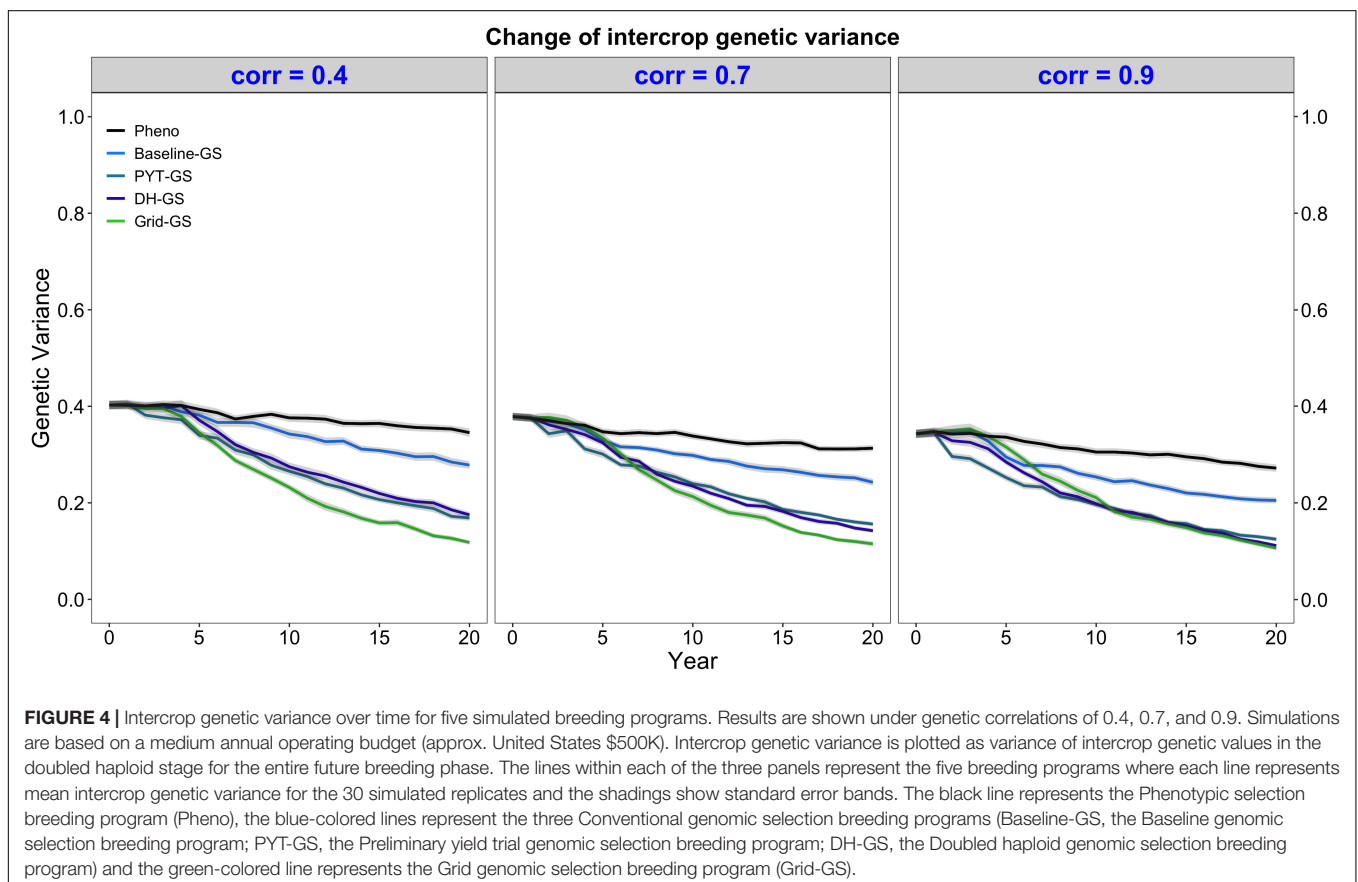
All breeding programs gave a faster reduction in genetic variance when the annual operating budget was high (**Supplementary Figure 1b**) and a slower reduction in genetic variance when annual operating costs were low (**Supplementary Figure 2b**). The general trends observed under the medium

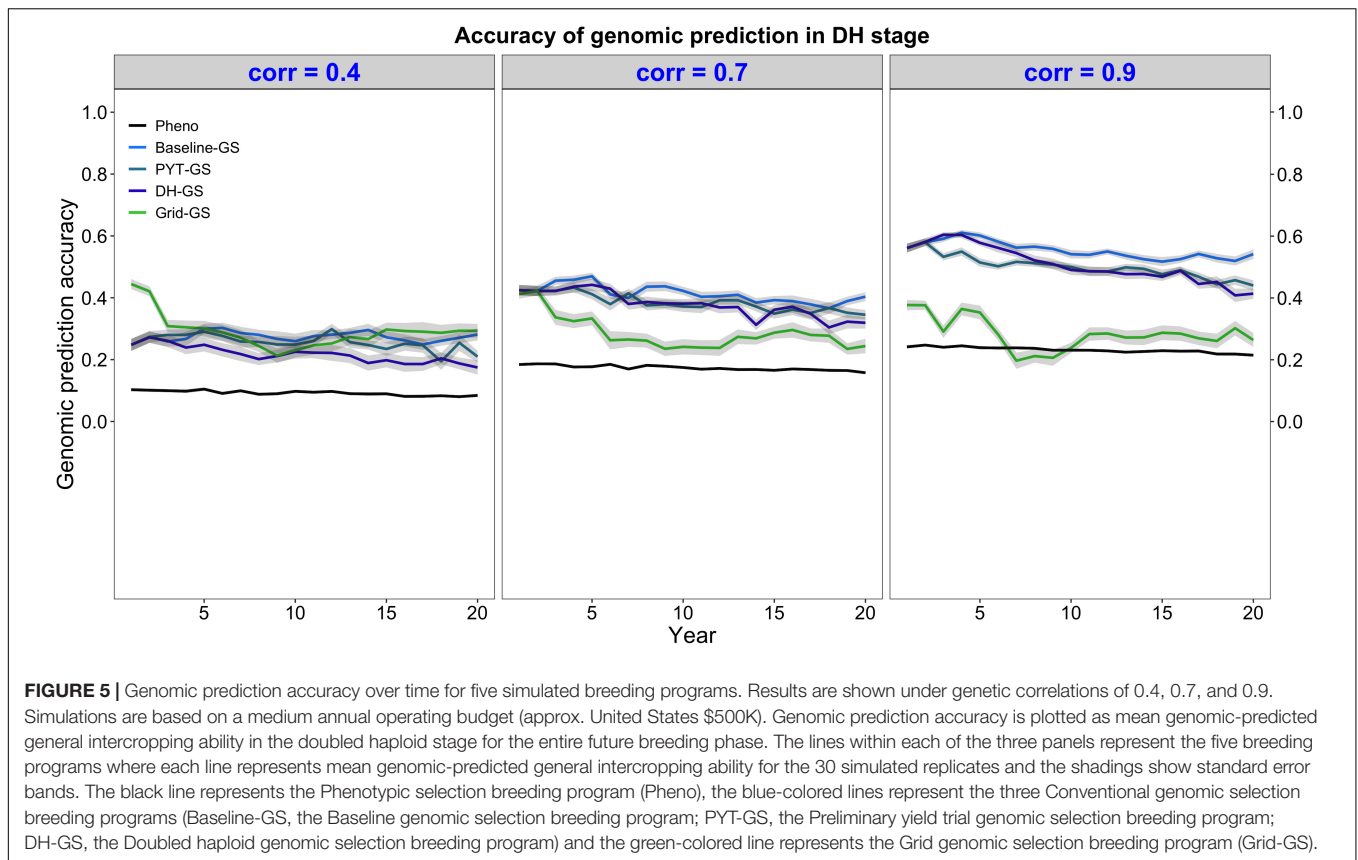
annual budget were however representative of the trends observed under low and high annual budgets.

## Genomic Selection Accuracy

Genomic selection for intercropping ability was more accurate than phenotypic selection for intercropping ability. This is shown in **Figure 5**, which plots the mean selection accuracy for general intercropping ability in the DH stage for the entire future breeding phase. All four breeding programs using genomic selection showed on average higher accuracy than the phenotypic selection breeding program under all three levels of correlation between monocrop yield and intercrop yield.

**Figure 5** also shows that selection accuracy for intercropping ability in the three Conventional genomic selection breeding programs and the phenotypic selection breeding program increased when the genetic correlation between monocrop grain yield and intercrop grain yield increased. Selection accuracy in the Grid-GS breeding program, on the other hand, was on average similar under all three levels of genetic correlation. When the genetic correlation was low (0.4), all four breeding programs using genomic selection showed on average a relatively similar selection accuracy. However, when the genetic correlation was high (0.9), the three Conventional genomic selection breeding programs showed a significantly higher selection accuracy than the Grid-GS breeding program. During most of the future breeding phase, the selection accuracy of the Grid-GS breeding





program was even lower than the selection accuracy of the phenotypic selection breeding program.

The four breeding programs using genomic selection showed a higher selection accuracy when the annual operating budget was high (**Supplementary Table 5** and **Supplementary Figure 1a**) and a lower selection accuracy when the annual operating costs were low (**Supplementary Table 5** and **Supplementary Figure 2a**). The general trends observed under the medium annual budget were however representative of the trends observed under low and high annual budgets.

## DISCUSSION

High-performance intercrop production systems require more efficient intercrop breeding approaches that make use of advances in breeding (Dawson et al., 2019b). While it offers potential advantages, genomic selection also incurs additional costs, so it is necessary to understand the balance between benefits and costs. Stochastic simulations are becoming widely used to explore the efficiency of genomic selection in monoculture breeding (e.g., Gaynor et al., 2017; Gorjanc et al., 2018; Muleta et al., 2019), but to our knowledge our use of simulations to explore genomic selection's value for intercrop breeding is unique. Through simulations, we have shown that intercrop breeding programs using genomic selection can produce faster genetic gain than intercrop breeding programs which only use phenotypic

selection, working to a common cost basis that reflects the resources available for a mediumly invested breeding initiative.

To discuss our results, we first examine the value of genomic selection to increase selection accuracy and reduce the generation interval in breeding crops for intercrop production. We also explain that maximizing the rate of genetic gain using genomic selection can significantly increase genetic gain in the short term, but may impair long-term genetic gain due to rapid depletion of genetic variance. We then describe the value of strategies which reduce the loss of genetic variance to solve this problem, such as maximum avoidance crossing schemes or optimal contribution selection. We explain why the performance of the different genomic selection breeding programs was dependent on the genetic correlation between monocrop grain yield and intercrop grain yield, and we conclude that the DH-GS breeding program should be used in intercrop breeding unless the genetic correlation between the two traits is known to be low. We finish by discussing the most important limitations of our simulations and explain why we believe that our results are still valid in the context of real-world intercrop breeding programs.

## Genomic Selection Increases Intercrop Genetic Gain

In phenotypic selection breeding programs, new crossing parents are usually selected after several years of intensive testing in multiple environments. This enables high selection accuracies but also results in long generations intervals, substantially restricting

the rate of genetic gain. Replacing phenotypic selection by genomic selection increases selection accuracy in early testing stages and thereby allows for selection of new parents based on their genomic predicted performance as soon as they can be genotyped. Our observations showed that all the intercrop breeding programs using genomic selection that we tested produced faster genetic gain than the phenotypic selection breeding program. This was observed regardless of the genetic correlation between monocrop grain yield and intercrop grain yield, and under three operating budgets. We observed that the major drivers of increased genetic gain were both an increased selection accuracy in early selection stages and a reduction of the generation interval. Our results were consistent with those of Gaynor et al. (2017) who used stochastic simulations to evaluate genomic selection strategies in plant breeding programs for developing inbred lines. We refer the reader to this study for a more detailed analysis of the relationship between genetic gain, the generation interval and prediction accuracy. As a consequence of increased selection accuracy and the reduced generation interval, all four genomic selection breeding programs also showed a faster reduction in genetic variance over time compared to the phenotypic selection breeding program. We discuss particular features of our findings in more detail below.

### Genomic Selection Accelerates the Reduction of Intercrop Genetic Variance Over Time

We found that all intercrop breeding programs using genomic selection showed a faster reduction of genetic variance than the phenotypic selection breeding program. As genomic selection improved the conversion of genetic variance into genetic gain, the accelerated reduction of intercrop genetic variance was a direct outcome of the increased selection accuracy and the reduced generation interval. While maximizing this conversion will significantly increase the rate of genetic gain in the short term, the long-term genetic gain may be impaired due to a rapid depletion of genetic variance. To solve this problem, maximum avoidance crossing schemes can be used, which maintain genetic variation by selecting the best genotypes within families while ensuring that each family equally contributes to the next generation (Kimura and Crow, 1963). In this way, an over-representation of the top families in future generations is prevented.

We experimented with this approach by applying a maximum avoidance crossing scheme in the DH-GS and the Grid-GS breeding programs, as initial simulations using truncation selection to select new parents resulted in rapid exhaustion of genetic variance. The approach was successful, but a downside of maximum avoidance crossing schemes is that they require a closed population with a constant number of families and a minimum number of progeny per family to ensure the least related crosses are made each generation. While these requirements can be easily met within a simulation framework, practical application of a maximum avoidance crossing scheme may be more challenging, as breeders might introduce new genetic material to their breeding population, and not every crossing event might produce seed. Other, more complex, strategies might be more suitable to reduce the loss of genetic variation in real-world breeding programs, such as optimal

contribution selection and crossing (Meuwissen, 1997; Sonesson et al., 2012; Akdemir and Sánchez, 2016; Gorjanc et al., 2018), and exploring these could be a feature of future work.

### The DH-GS Breeding Program Produces the Most Genetic Gain When the Genetic Correlation Between Monocrop Yield and Intercrop Yield Is Medium to High

In our simulations, the DH-GS breeding program produced approximately two times the genetic gain of the Grid-GS breeding program and approximately 2.5 times the genetic gain of the phenotypic selection breeding program when the genetic correlation between monocrop yield and intercrop yield was medium to high (0.7 and 0.9). The DH-GS scheme benefited from a short generation interval and an increased selection accuracy in the DH, PYT and GIA1 stages.

To obtain genomic predictions of general intercropping abilities for each component crop, the DH-GS breeding program used a multivariate genomic selection model which fitted monocrop grain yield and intercrop grain yield simultaneously. While phenotypic information on intercrop yield came from the GIA1 and the GIA2 stages, the multivariate model enabled us to extract additional information from monocrop yield phenotypes due to the genetic correlation between monocrop yield and intercrop yield. This additional information resulted in increased selection accuracy when the correlation was medium to high. While novel in the context of intercrop breeding, the use of correlated traits in multivariate genomic models is a well-known approach to improve prediction accuracy with wide application in plant and animal breeding (Calus and Veerkamp, 2011; Jia and Jannink, 2012).

The same multivariate genomic selection model was also used in the Baseline-GS and the PYT-GS breeding programs. These two breeding programs also outperformed the phenotypic selection breeding program regardless of the genetic correlation between monocrop yield and intercrop yield, but produced less genetic gain than the DH-GS breeding program. The PYT-GS breeding program benefited from an increased selection accuracy and a reduced generation interval compared to the phenotypic selection breeding program. The Baseline-GS breeding program did not reduce the generation interval. It was used to demonstrate the increase in selection accuracy when genomic selection is used compared to phenotypic selection.

### The Grid-GS Breeding Program Is Advantageous When Genetic Correlation Is Low

In our modeling, the Grid-GS breeding program produced approximately 1.2 times the genetic gain of the DH-GS breeding program and approximately 2.3 times the genetic gain of the phenotypic selection breeding program when the genetic correlation between monocrop yield and intercrop yield was low (0.4). Our findings can be explained by the fact that the Grid-GS genomic selection model did not consider monocrop yield records, so it is unaffected by the genetic correlations between monocrop yield and intercrop yield, and prediction accuracies are similar under all correlations. When the genetic correlation was low, it therefore outperformed the DH-GS

breeding program. However, under a high genetic correlation, the training population size of the DH-GS breeding program was effectively increased by including phenotypic information from monocrop stages, while the training population in the grid stage of the Grid-GS program was not affected by the level of genetic correlation. We hypothesize that a larger training population and an optimized sampling strategy (rather than random sampling) of intercrop combinations in the grid stage would further increase the predictive ability of the genomic model and performance of the Grid-GS breeding program. Sampling strategies such as data-mining tools that exploit both genomic relationships and phenotypic variation to obtain the most representative subset as the training population from factorial design have already been discussed in the context of hybrid breeding and the existing theory could be extended to our Grid-GS design (Guo et al., 2019).

### Unless the Genetic Correlations Between Monocrop and Intercrop Yield Are Known to Be Low, the DH-GS Breeding Program Should Be Used

Unless the genetic correlation between monocrop yield and intercrop yield was low, the DH-GS breeding program produced the most genetic gain. Even when the genetic correlation was 0.4, it was only slightly outperformed by the Grid-GS breeding program. These results indicate that the DH-GS breeding program has great potential to improve breeding for intercrop production.

In practical intercrop breeding programs, the genetic correlation between monocrop traits and intercrop traits will most likely be unknown and can change over time. The estimation of these parameters is difficult and requires large and costly experimental designs (Hamblin et al., 1976; Wright, 1985; Hill, 1996). When data for a precise decision-making process is not available, a strategy is required that delivers consistent performance across a wide parameter space. The DH-GS breeding program achieved substantially higher genetic gains than the phenotypic selection breeding program under all simulated correlations. Hence, we recommend that it is suitable for prompt implementation without prior knowledge about the level of genetic correlation.

A further advantage of the DH-GS breeding program is that it is a relatively simple way to implement genomic selection on top of a phenotypic selection intercrop breeding program, as it only requires minor resource re-allocations to compensate for the extra cost of genotyping. The Grid-GS breeding program, on the other hand, requires extensive restructuring of the breeding program, which might be harder to realize, particularly in low- and middle-income countries with limited resources.

The genomic selection models employed in this study should not be considered the ideal models to use in practice. These models were chosen to provide a reasonable estimate of the performance of genomic selection in an intercrop breeding program. In practice, the choice of a model should be guided by data and models that assess each of the actual component crops that are being considered. These models were not considered in our simulations, because the assumptions of the simulations made them unnecessary. Specifically, our simulations assumed

no interaction between component crops, and that the genetic variance and heritability was the same for each component crop. These assumptions are unlikely to be met in reality.

### Limitations of Applying Stochastic Simulations for Intercrop Breeding Program Design

Our simulations have revealed the value of applying genomic selection in intercrop breeding. However, they are based on various simplified assumptions and do not model the full complexity of an actual intercrop breeding program. In this section, we discuss the most important limitations of our simulations and explain why we believe that our results still remain valid for real-world intercrop breeding. In the below we will discuss in turn assumptions which impact genomic selection accuracy, assumptions about making crosses and seed production, assumptions about the complexity of the breeding goal, and assumptions about the absence of genotype-by-genotype interaction between the two component crops.

#### Assumptions Which Impact Genomic Selection Accuracy

The intercrop prediction accuracies obtained in our simulations are likely to be higher than those realized under real-world conditions. Our simulations may be inflated because: the variance components provided to genomic models were estimated directly from the true simulation parameters; there were no genotyping or phenotyping errors in our data set; and we assumed additive gene action only without non-additive effects (i.e., dominance and epistasis) and genotype-by-environment interaction (Vitezica et al., 2013; Forneris et al., 2017; Jarquín et al., 2017). Also, the population structure of both component crops was much simpler than that found in real breeding programs which may further affect the prediction accuracy (Guo et al., 2014; Isidro et al., 2015).

These factors should affect all genomic selection breeding programs and we do not expect their relative performance to change much under real-world conditions. Inclusion of non-additive effects in more complex genomic selection models, however, may also only provide very low (or negligible) improvements in genetic gain or the prediction accuracy (Hill et al., 2008; Varona et al., 2018); while dominance does not even apply in our simulations since we were dealing with inbred lines. Gaynor et al. (2017) observed similar rankings of breeding programs even when using simulated genotype-by-environment interaction. This suggests that the DH-GS breeding program would still show overall best performance compared to other genomic selection breeding programs. However, the performance of the phenotypic selection breeding program relative to the genomic selection breeding programs could change. If this was to occur, the magnitude of the difference between the best genomic selection breeding program and the phenotypic selection breeding program we have observed leads us to believe that under real-world conditions the genomic selection program would still outperform the phenotypic selection breeding program. Lastly, the population structure might be an issue at the beginning of an intercrop breeding

program when extensive phenotypic and genetic variation will be required for various interaction traits (e.g., days to flowering and plant height) (Litrico and Violle, 2015). However, in our simulations, genomic selection was only implemented after 20 years of phenotypic selection (burn-in), when populations were already more uniform and a sufficient number of training records was available. Such gradual transition is likely to happen also in reality.

### Assumptions About Making Crosses and Seed Production

To minimize complexity, in our simulated breeding programs we assumed no differences in flowering time between crossing parents and that all crosses produce sufficient amounts of seed for immediate next step implementation. In real-world breeding, differences in maturity between potential crossing parents might reduce the number of possible crosses, while some crosses may not immediately produce enough seed, with additional seed multiplication steps required that prolong the breeding process. If dealing with a self-pollinating and an outcrossing component crop simultaneously, these issues might present the most significant challenge (Hamblin and Zimmermann, 1986). As Hamblin et al. (1976) indicate, two self-pollinating crops with large seed production may be the simplest case for intercrop breeding. Breeding programs that use either phenotypic or genomic selection would be similarly affected by these seed production issues. We thus assume that the relative performance of the different breeding programs would be similar under more realistic crossing scenarios.

### Assumptions About the Complexity of the Breeding Goal

In our analysis, comparisons between breeding programs were based on a single quantitative trait representing intercrop grain yield. We also assumed that both component crops equally contributed to intercrop grain yield and its economic value. Real-world breeding programs, however, have to consider multiple quantitative and qualitative traits simultaneously to maximize agronomic performance. Furthermore, it is unlikely that both component crops produce comparable amounts of yield and that both component traits have a similar market value.

In fact, the contribution of each component crop to the total economic value of the combined product will depend on various factors. These include: the cultivation environment (i.e., biological, economic and cultural) and management practices (Francis, 1981; Mead and Riley, 1981); the *per se* yield potential and economic value of each component crop (Hamblin et al., 1976; Francis, 1981; Wright, 1985); and the intended use of the products, especially whether for subsistence use or market (Mead and Riley, 1981).

In theory, a selection index could be developed to enable selection of the best intercrop combinations by combining several key traits and through considering the above factors. Selection indices can allow the assignment of customized economic weights to the component crops, thereby optimizing

their individual yield gains to maximize the market value of the combined crop product. In real-world breeding programs, estimation of the relative (economic) weights for traits of interest is not a trivial exercise, and weights may also need to be changed over time (Mead and Riley, 1981). Moreover, the selection index is likely to differ between different intercrop combinations and should be determined by both the data and breeding objectives. However, in the context of a simulation, the simulated trait also can be considered as the total economic value resulting from a linear selection index (e.g., Smith-Hazel index). We assume that we would observe similar trends for our simulated breeding programs if we were to include multiple traits in an index.

### Assumptions About the Absence of Genotype-by-Genotype Interaction Between the Two Component Crops

In our simulations, we ignored the possible effects of genotype-by-genotype interactions between component crops. In practical intercrop production, these interactions play an important role in determining their productivity, ecosystem service provision and resilience (Dawson et al., 2019a). Although strategies have been outlined through which genetic variants underlying mutualisms between pairs of plant species in natural ecosystems can be characterized, studies reporting genotype-by-genotype interactions are currently relatively scarce (Subrahmaniam et al., 2018). We expect the effect of genotype-by-genotype interactions to be most significant at the start of breeding activities, when material is unadapted to a particular growing system and when they could potentially result in re-ranking of our breeding programs. We expect that through continuous recurrent selection these interactions may become minimal, as the competition component is minimized through continuously improved coexistence between two component crops (Hill, 1996).

## CONCLUSION

Our results show that genomic selection shows great promise in breeding crops for intercrop production. We have demonstrated that genomic selection can significantly increase the rate of genetic gain in intercrop breeding. In particular, the DH-GS breeding strategy provides a simple solution to implement genomic selection on top of an existing phenotypic selection breeding program, without major rearrangements and regardless of the genetic correlation between monocrop yield and intercrop yield. Clearly, the practical challenges of the implementation of genomic selection strategies differ between breeding programs, but we believe that our results will aid breeders in optimizing the implementation process. Overall, the current study can be considered as an initial piece that future modeling work can build on. In our further work we are exploring the utility of different design approaches for crop combinations such as finger millet and groundnut that could be optimized as an important intercrop

for reaching multiple human and environmental health benefits in East Africa (Dawson et al., 2019b).

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## AUTHOR CONTRIBUTIONS

JB, JH, and RG conceived and designed the study. JB developed the plant breeding program simulations. DO and HO provided information on existing breeding programs. JB, CW, and ID wrote the manuscript, with input from all authors. All authors read and approved the final manuscript.

## REFERENCES

- Acquaah, G. (2012). *Principles of Plant Genetics and Breeding*, 2nd Edn. John Wiley & Sons, doi: 10.1017/CBO9781107415324.004
- Akdemir, D., and Sánchez, J. I. (2016). Efficient Breeding by Genomic Mating. *Front. Genet.* 7:1–12. doi: 10.3389/fgene.2016.00210
- Bernardo, R. N. (2010). *Breeding for Quantitative Traits in Plants*. Woodbury, MN: Stemma Press.
- Brooker, R. W., Jones, H. G., Paterson, E., Watson, C., Brooker, R. W., Bennett, A. E., et al. (2015). Improving Intercropping: A Synthesis of Research in Agronomy, Plant Physiology and Ecology. *N. Phytol.* 206, 107–117. doi: 10.1111/nph.13132
- Calus, M. P. L., and Veerkamp, R. F. (2011). Accuracy of Multi-Trait Genomic Selection Using Different Methods. *Genet. Select. Evolut.* 43:26. doi: 10.1186/1297-9686-43-26
- Chen, G. K., Marjoram, P., and Wall, J. D. (2009). Fast and Flexible Simulation of DNA Sequence Data. *Genome Res.* 19, 136–142. doi: 10.1101/gr.083634.108
- Dass, A., and Sudhishri, S. (2010). Intercropping in Finger millet (Eleusine Coracana) with Pulses for Enhanced Productivity, Resource Conservation and Soil Fertility in Uplands of Southern Orissa. *Ind. J. Agronomy* 55, 89–94.
- Davis, J. H. C., and Woolley, J. N. (1993). Genotypic Requirement for Intercropping. *Field Crops Res.* 34, 407–430. doi: 10.1016/0378-4290(93)90124-6
- Dawson, I. K., Park, S. E., Attwood, S. J., Jamnadass, R., Powell, W., Sunderland, T., et al. (2019a). Contributions of Biodiversity to the Sustainable Intensification of Food Production. *Glob. Food Security* 21, 23–37. doi: 10.1016/j.gfs.2019.07.002
- Dawson, I. K., Powell, W., Hendre, P., Bančić, J., Hickey, J. M., Kindt, R., et al. (2019b). The Role of Genetics in Mainstreaming the Production of New and Orphan Crops to Diversify Food Systems and Support Human Nutrition. *N. Phytol.* 224:15895. doi: 10.1111/nph.15895
- Fornieris, N. S., Vitezica, Z. G., Legarra, A., and Pérez-Enciso, M. (2017). Influence of Epistasis on Response to Genomic Selection Using Complete Sequence Data. *Genet. Select. Evol.* 49, 1–14. doi: 10.1186/s12711-017-0340-3
- Francis, C. A. (1981). "Development of Plant Genotypes for Multiple Cropping Systems," in *Plant Breeding II*, ed. K. J. Frey (Iowa: Iowa State University Press).
- Gaynor, R. C., Gorjanc, G., and Hickey, J. M. (2020). AlphaSimR: An R-Package for Breeding Program Simulations. *BioRxiv* 2020:245167. doi: 10.1101/2020.08.10.245167
- Gaynor, R. C., Gorjanc, G., Bentley, A. R., Ober, E. S., Howell, P., Jackson, R., et al. (2017). A Two-Part Strategy for Using Genomic Selection to Develop Inbred Lines. *Crop Sci.* 57, 2372–2386. doi: 10.2135/cropsci2016.09.0742
- Gorjanc, G., Gaynor, R. C., and Hickey, J. M. (2018). Optimal Cross Selection for Long-Term Genetic Gain in Two-Part Programs with Rapid Recurrent Genomic Selection. *Theoret. Appl. Genet.* 131, 1953–1966. doi: 10.1007/s00122-018-3125-3
- Guo, T., Yu, X., Li, X., Zhang, H., Zhu, C., Flint-Garcia, S., et al. (2019). Optimal Designs for Genomic Selection in Hybrid Crops. *Mol. Plant* 12, 390–401. doi: 10.1016/j.molp.2018.12.022
- Guo, Z., Tucker, D. M., Basten, C. J., Gandhi, H., Ersoz, E., Guo, B., et al. (2014). The Impact of Population Structure on Genomic Prediction in Stratified Populations. *TAG Theoret. Appl. Genet.* 127, 749–762. doi: 10.1007/s00122-013-2255-x
- Hamblin, J., and Zimmermann, M. J. O. (1986). Breeding Common Bean for Yield in Mixtures. *Plant Breeding Rev.* 4, 245–272. doi: 10.1002/9781118061015.ch8
- Hamblin, J., Rowell, J. G., and Redden, R. (1976). Selection for Mixed Cropping. *Euphytica* 25, 97–106. doi: 10.1007/BF00041533
- Hickey, J. M., Chiurugwi, T., Mackay, I., and Powell, W. (2017). Genomic Prediction Unifies Animal and Plant Breeding Programs to Form Platforms for Biological Discovery. *Nat. Genet.* 49, 1297–1303. doi: 10.1038/ng.3920
- Hill, J. (1996). Breeding Components for Mixture Performance. *Euphytica* 92, 135–138. doi: 10.1007/BF00022838
- Hill, W. G., Goddard, M. E., and Visscher, P. M. (2008). Data and Theory Point to Mainly Additive Genetic Variance for Complex Traits. *PLoS Genet.* 4:1000008. doi: 10.1371/journal.pgen.1000008
- Himmelstein, J., Ares, A., Gallagher, D., and Myers, J. (2017). A Meta-Analysis of Intercropping in Africa: Impacts on Crop Yield, Farmer Income, and Integrated Pest Management Effects. *Int. J. Agricult. Sustainabil.* 15, 1–10. doi: 10.1080/14735903.2016.1242332
- Isidro, J., Jannink, J. L., Akdemir, D., Poland, J., Heslot, N., and Sorrells, M. E. (2015). Training Set Optimization under Population Structure in Genomic Selection. *TAG. Theoret. Appl. Genet.* 128, 145–158. doi: 10.1007/s00122-014-2418-4
- Jarquín, D., da Silva, C. L., Gaynor, R. C., Poland, J., Fritz, A., Howard, R., et al. (2017). Increasing Genomic-Enabled Prediction Accuracy by Modeling Genotype × Environment Interactions in Kansas Wheat. *Plant Genome* 10:130. doi: 10.3835/plantgenome2016.12.0130
- Jia, Y., and Jannink, J. L. (2012). Multiple-Trait Genomic Selection Methods Increase Genetic Value Prediction Accuracy. *Genetics* 192, 1513–1522. doi: 10.1534/genetics.112.144246
- Kimura, M., and Crow, J. F. (1963). On the Maximum Avoidance of Inbreeding. *Genet. Res.* 4, 399–415. doi: 10.1017/S0016672300003797
- Kiwia, A., Kimani, D., Harawa, R., Jama, B., and Sileshi, G. W. (2019). Sustainable Intensification with Cereal-Legume Intercropping in Eastern and Southern Africa. *Sustainabil.* 11, 1–18. doi: 10.3390/su11102891
- Leff, B., Ramankutty, N., and Foley, J. A. (2004). Geographic Distribution of Major Crops across the World. *Glob. Biogeochem. Cycles* 18:GB002108. doi: 10.1029/2003GB002108
- Li, C., Hoffland, E., Kuyper, T. W., Yu, Y., Zhang, C., Li, H., et al. (2020). Syndromes of Production in Intercropping Impact Yield Gains. *Nat. Plants* 6, 653–660. doi: 10.1038/s41477-020-0680-9

## FUNDING

SRUC authors are grateful for Global Challenge Research Funding on orphan crops (project BB/P022537/1: Formulating Value Chains for Orphan Crops in Africa, 2017–2019, Foundation Award for Global Agriculture and Food Systems). JB was funded through an SRUC studentship Research Excellence Grant.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article, including scripts to run simulations, can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2021.605172/full#supplementary-material>



- Litrico, I., and Violle, C. (2015). Diversity in Plant Breeding: A New Conceptual Framework. *Trends Plant Sci.* 20, 604–613. doi: 10.1016/j.tplants.2015.07.007
- Lorenz, A. J., Chao, S., Asoro, F. G., Heffner, E. L., Hayashi, T., Iwata, H., et al. (2011). Genomic Selection in Plant Breeding. Knowledge and Prospects. *Adv. Agronomy* 110, 77–123. doi: 10.1016/B978-0-12-385531-2.0002-5
- Mead, R., and Riley, J. (1981). A Review of Statistical Ideas Relevant to Intercropping Research. *J. R. Statist. Soc. A* 144, 462–509. doi: 10.2307/2981827
- Meuwissen, T. H. E. (1997). Maximizing the Response of Selection with a Predefined Rate of Inbreeding: Overlapping Generations. *J. Anim. Sci.* 75, 934–940. doi: 10.2527/1998.76102575x
- Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics* 157, 1819–1829. doi: 11290733
- Muleta, K. T., Pressoir, G., and Morris, G. P. (2019). Optimizing Genomic Selection for a Sorghum Breeding Program in Haiti: A Simulation Study. *G3* 9, 391–401. doi: 10.1534/g3.118.200932
- R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Vienna: R Core Team.
- Raseduzzaman, M. D., and Jensen, E. S. (2017). Does Intercropping Enhance Yield Stability in Arable Crop Production? A Meta-Analysis. *Eur. J. Agronomy* 91, 25–33. doi: 10.1016/j.eja.2017.09.009
- Sampoux, J.-P., Giraud, H., and Litrico, I. (2020). Which recurrent selection scheme to improve mixtures of crop species? Theoretical expectations. *G3: Genes Genom. Genet.* 10, 89–107 doi: 10.1534/g3.119.400809
- Sonesson, A. K., Woolliams, J. A., and Meuwissen, T. H. E. (2012). Genomic Selection Requires Genomic Control of Inbreeding. *Genet. Select. Evol.* 44, 1–10. doi: 10.1186/1297-9686-44-27
- Subrahmaniam, H. J., Libourel, C., Journet, E. P., Morel, J. B., Muñoz, S., Niebel, A., et al. (2018). The Genetics Underlying Natural Variation of Plant-Plant Interactions, a Beloved but Forgotten Member of the Family of Biotic Interactions. *Plant J.* 93, 747–770. doi: 10.1111/tpj.13799
- Vandermeer, J. (1989). *The Ecology of Intercropping*. Cambridge: Cambridge University Press.
- Varona, L., Legarra, A., Toro, M. A., and Vitezica, Z. G. (2018). Non-Additive Effects in Genomic Selection. *Front. Genet.* 9:1–12. doi: 10.3389/fgene.2018.00078
- Vitezica, Z. G., Varona, L., and Legarra, A. (2013). On the Additive and Dominant Variance and Covariance of Individuals within the Genomic Selection Scope. *Genetics* 195, 1223–1230. doi: 10.1534/genetics.113.155176
- Wright, A. J. (1985). Selection for Improved Yield in Inter-Specific Mixtures or Intercrops. *Theoret. Appl. Genet.* 69, 399–407. doi: 10.1007/BF00570909
- Zimmermann, M. J. O. (1996). Breeding for Yield in Mixtures of Common Beans (*Phaseolus Vulgaris* L.) and Maize (*Zea Mays* L.). *Euphytica* 92, 129–134. doi: 10.1007/BF00022837

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Bančić, Werner, Gaynor, Gorjanc, Odeny, Ojulung, Dawson, Hoad and Hickey. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.