

15 **Abstract**

16 Hybrid crop breeding programs using a two-part strategy produced the most genetic gain,
17 but a maximum avoidance of inbreeding crossing scheme was required to increase long-term
18 genetic gain. The two-part strategy uses outbred parents to complete multiple generations per
19 year to reduce the generation interval of hybrid crop breeding programs. The maximum
20 avoidance of inbreeding crossing scheme manages genetic variance by maintaining uniform
21 contributions and inbreeding coefficients across all crosses. This study performed stochastic
22 simulations to quantify the potential of a two-part strategy in combination with two crossing
23 schemes to increase the rate of genetic gain in hybrid crop breeding programs. The two crossing
24 schemes were: (i) a circular crossing scheme, and (ii) a maximum avoidance of inbreeding
25 crossing scheme. The results from this study show that the implementation of genomic
26 selection increased the rate of genetic gain, and that the two-part hybrid crop breeding program
27 generated the highest genetic gain. This study also shows that the maximum avoidance of
28 inbreeding crossing scheme increased long-term genetic gain in two-part hybrid crop breeding
29 programs completing multiple selection cycles per year, as a result of maintaining higher levels
30 of genetic variance over time. The flexibility of the two-part strategy offers further
31 opportunities to integrate new technologies to further increase genetic gain in hybrid crop
32 breeding programs, such as the use of outbred training populations. However, the practical
33 implementation of the two-part strategy will require the development of bespoke transition
34 strategies to fundamentally change the data, logistics, and infrastructure that underpin hybrid
35 crop breeding programs.

36

37

38 **Introduction**

39 The two-part strategy produced the most genetic gain in hybrid crop breeding programs,
40 but the maximum avoidance of inbreeding crossing scheme was required for it to increase long-
41 term genetic gain. The two-part strategy uses outbred parents to complete multiple generations
42 per year in hybrid crop breeding programs. In contrast, conventional plus genomic selection
43 strategies are limited in this regard by the time they take to develop inbred lines. The maximum
44 avoidance of inbreeding crossing scheme manages genetic variance by maintaining uniform
45 contributions and inbreeding coefficients across all crosses. This study performed stochastic
46 simulations to quantify the potential of a two-part strategy to increase the rate of genetic gain
47 in hybrid crop breeding programs. A large increase in food production is required to meet the
48 demand for a global population of 9 billion people in 2050. Increasing the rate of genetic gain
49 of breeding programs is one route to achieve sustainable, permanent and cumulative increases
50 in food production. Hybrid crops, and their genetic improvement, have made major
51 contributions to historical increases in food production. For example, genetic merit for maize
52 yield has approximately doubled from 1930 to 2001 (Duvick et al. 2010; Fig. 4.1). However,
53 the current rates of genetic gain in hybrid crop breeding programs are insufficient to meet the
54 estimated 70% increase in overall food production (Alexandratos and Bruinsma 2012) required
55 within the next 30 years.

56 Genomic selection could increase the rate of genetic gain in hybrid crop breeding
57 programs by directly addressing three of the parameters of the breeder's equation. The
58 breeder's equation provides a framework to understand how the rates of genetic gain in
59 breeding programs can be increased (Lush 1943). The breeder's equation shows that genetic
60 gain is a function of (i) the accuracy of ranking selection candidates based on genetic merit,
61 (ii) the intensity of selection, (iii) the genetic variance in the population, and (iv) the generation
62 interval. Genomic selection could increase the rate of genetic gain in hybrid crop breeding

63 programs by reducing the generation interval, increasing the selection intensity and increasing
64 the selection accuracy.

65 Hybrid crop breeding program designs involve multiple stages with selection
66 candidates evaluated against an increasing number of testers in an increasing number of
67 environments. As the accuracy of evaluation increases through the stages of evaluation,
68 individuals are recycled by reciprocal recurrent testcross selection (Bernardo, 2014; Hull,
69 1945) and crossed to create the next set of selection candidates. Reciprocal recurrent testcross
70 selection aims to improve the general combining ability of selection candidates. However, the
71 development of inbred individuals in reciprocal recurrent testcross selection requires time
72 which results in longer breeding cycle times and slows the rate of population improvement.
73 For example, hybrid crop breeding program designs typically have a cycle time of 3 to 4 years
74 and are not radically different from a typical breeding program design for inbred crops.

75 Genomic selection can increase the rate of population improvement in plant breeding
76 programs. Recently, Gaynor et al. (2017) proposed a two-part breeding strategy for inbred
77 crops that explicitly separates a conventional plant breeding program into two distinct
78 components. These components are:

- 79 (i) a population improvement component to develop improved germplasm via
80 recurrent genomic selection, and;
81 (ii) a product development component to identify new inbred varieties within
82 conventional plant breeding program designs.

83 Gaynor et al. (2017) used simulation to compare conventional and two-part breeding program
84 designs in the context of inbred crops. Compared to the conventional design, the two-part
85 strategy generated 2.4 times more genetic gain per unit cost and unit time.

86 Conceptually the two-part strategy is equally suited to hybrid crop breeding programs.
87 In a hybrid crop breeding program, the two-part strategy could enable large increases in genetic

88 gain by shortening the generation interval considerably. However, in the context of hybrid
89 crops, population improvement would need to be driven by reciprocal recurrent genomic
90 selection rather than recurrent genomic selection. Reciprocal recurrent selection aims to
91 improve the general combining ability of individuals in different heterotic pools. Reciprocal
92 recurrent genomic selection (Kinghorn et al. 2010) uses the phenotypes and parental genotypes
93 of hybrids to more accurately estimate the general combining ability of individuals (Rembe et
94 al. 2019).

95 However, the previous implementations of the two-part strategy in inbred crops showed
96 that large reductions of the generation interval came at the expense of genetic variation. Gaynor
97 et al. (2017) showed that two-part breeding programs that used rapid cycling to reduce the
98 generation interval below 0.5 years reduced long-term genetic gain. Using simulation, Gorjanc
99 et al. (2018) showed that long-term genetic gain can be optimised with crossing schemes that
100 balance increases in genetic gain with reductions in genetic variance. Maximum avoidance of
101 inbreeding is a crossing scheme that maintains uniform contributions across generations and
102 uniform inbreeding coefficients across all crosses (Wright 1921; Kimura and Crow 1963). Due
103 to the large number of generations per year that can be completed with the two-part strategy,
104 the use of maximum avoidance of inbreeding could increase long-term genetic gain in hybrid
105 crop breeding programs.

106 The objective of this study was to develop and test the two-part strategy in the context
107 of hybrid crops. Stochastic simulations were used, with a maize breeding program as a model,
108 to compare conventional, conventional plus genomic selection and two-part hybrid crop
109 breeding programs under an assumption of approximately equal operating costs and time. To
110 manage long-term genetic variance, both a circular crossing scheme and a maximum avoidance
111 of inbreeding crossing scheme were used in conjunction with genomic selection. The results
112 show that: (i) the implementation of genomic selection in hybrid crop breeding programs

113 increases the rate of genetic gain, (ii) the two-part strategy was the most cost-effective strategy
114 for implementing genomic selection in hybrid crop breeding programs, and (iii) two-part
115 hybrid crop breeding programs completing multiple selection cycles per year should use
116 methods to manage genetic variance.
117

118 **Methods**

119 Stochastic simulations of entire hybrid crop breeding programs were used to compare:

- 120 • a conventional breeding program not using genomic selection;
- 121 • three conventional plus genomic selection breeding programs, and;
- 122 • two breeding programs implementing the two-part strategy.

123 These breeding programs were compared on an equal time across 40 years of breeding. Each
124 breeding program was constrained to have approximately equal operating costs so that direct
125 comparisons between the different breeding programs would represent their relative
126 effectiveness. The six different breeding programs were compared using 10 independent
127 replicates of a stochastic simulation for three levels of genotype-by-year interaction variance.
128 Each replicate consisted of:

- 129 (i) a burn-in phase shared by all strategies so that each strategy had an identical,
130 realistic starting point, and;
- 131 (ii) a future breeding phase that simulated 20 years of future breeding with each of
132 the different breeding strategies.

133

134 **Burn-In Phase**

135

136 Specifically, the burn-in phase was subdivided into three stages. The first stage
137 simulated the species' genome sequence. The second stage simulated trait architecture and
138 founder genotypes for the initial parents. The third stage simulated 20 years of breeding using
139 the conventional breeding strategy without genomic selection.

140

141 *Generation of whole genome sequence data*

142

143 For each replicate, a genome consisting of 10 chromosome pairs was simulated to
144 resemble the maize genome. These chromosomes were assigned a genetic length of 2.0
145 Morgans and a physical length of 2×10^8 base pairs. Sequences for each chromosome were
146 generated using the Markovian Coalescent Simulator (Chen et al. 2009) within AlphaSimR
147 (Gaynor et al. 2019). Recombination and mutation rates were respectively set to 1.25×10^{-8} per
148 base pair and 1×10^{-8} per base pair. Historical effective population size was simulated,
149 beginning with a single population, as follows; 100,000 at 6,000 generations ago, 10,000 at
150 2,000 generations ago, 5,000 at 1,000 generations ago, 1,000 at 100 generations ago. To mimic
151 the genetic separation of two heterotic groups the population was split 200 generations ago.
152 The final effective population size at the end of the coalescent simulation was set to 100 for
153 each heterotic group. These values were chosen to roughly follow the evolution of effective
154 population size in North American hybrid maize.

155

156 *Founder Genotypes*

157

158 The founders served as the initial parents in the burn-in breeding phase. This was
159 accomplished by randomly sampling gametes from the simulated genome to assign as
160 sequences for the founders. 80 founders were created for each of the two heterotic groups. Sites
161 segregating in the founders' sequences were randomly selected to serve as 2000 single
162 nucleotide polymorphism (SNP) markers per chromosome (20000 total) and 300 quantitative
163 trait nucleotides (QTL) per chromosome (3000 total). The randomly selected sites for SNP

164 markers and QTL were not allowed to overlap. The founders were converted to inbred lines by
165 simulating the formation of doubled haploids (DH).

166

167 *Trait Architecture*

168

169 Three types of biological effects were modelled at each QTL to simulate genetic values:
170 additive effects, dominance effects and genotype-by-year effects. Under the AlphaSimR
171 framework, this is referred to as an ADG trait. We will give only a brief summary of the
172 modelling procedure, while a detailed description can be found in the vignette of the
173 AlphaSimR package (Gaynor et al. 2019).

174 A single trait representing grain yield, controlled by 3,000 QTL, was simulated for all
175 individuals. Each QTL was assigned an additive genetic effect, composed of additive and
176 genotype-by-year effects, and a dominance effect resulting from the interaction between alleles
177 at a heterozygous locus. Epistatic gene action was not considered. Three levels of genotype-
178 by-year variance were examined: 0, 2 and 4 times the genetic variance. Dominance effects
179 were calculated by multiplying the absolute value of each QTL allele effect by a locus-specific
180 dominance deviation (δ). Dominance deviations were sampled from a normal distribution with
181 mean dominance deviation of 0.92 and variance of 0.2, to approximate historical levels of
182 heterosis displayed in commercial maize (Troyer and Wellin 2009). The genetic value of each
183 QTL was then defined as the sum of additive QTL effects and the dominance effect of
184 interacting alleles. Finally, the genetic value of each individual was obtained as the sum of all
185 of the QTL genetic values, accounting for the individuals' genotype at these QTL.

186 The genetic value of each individual was used to produce phenotypic values by adding
187 random error. The random error was sampled from a normal distribution with mean zero. The
188 variance of the random error varied according to the stage of evaluation in the breeding

189 program. This was done to account for increasing accuracy in the evaluation as the number of
190 replications per entry increased. The values for these error variances were set to achieve target
191 levels of heritability. The levels of heritability represented heritability on an entry-mean basis
192 for the 80 founder genotypes when genotype-by-year variance was absent. The levels of
193 heritability that were used are presented in the description of the conventional breeding
194 program below.

195

196 *Conventional Breeding Program*

197

198 Burn-in breeding for yield was simulated using 20 years of breeding in a conventional
199 program without genomic selection. The design of the burn-in program approximated existing
200 maize breeding programs taken from Bernardo (2014; Table 8.3). The key features of the two
201 heterotic groups in this breeding program were:

- 202 (i) a crossing block consisting of 80 DH lines used to develop 80 biparental
203 populations each year;
- 204 (ii) the development of 25 new DH lines from each biparental cross;
- 205 (iii) a 3-year cycle time from crossing to selection of new parents; and
- 206 (iv) a 6-year production interval from crossing to release of a new commercial
207 hybrid.

208 All selection in the burn-in program was performed using phenotypes. These
209 phenotypes represented direct selection on yield using a yield trial. The levels of heritability at
210 a particular selection stage were adapted from the number of DH lines and locations reported
211 in Bernardo (2014; Table 8.3). A schematic for the overall design of the burn-in program is
212 given in Fig. 1 and a detailed description follows below. Each of the stages, described below,

213 were conducted independently in the two heterotic groups. The progression of germplasm
214 through the breeding program was simulated using AlphaSimR (Gaynor et al. 2019).

215

216 *Year 1*

217

218 80 bi-parental populations were created with intra-heterotic group crosses. It was
219 ensured that each of the 80 parental DH lines was used as a male or female only once. Each
220 cross produced 25 F₁ derived doubled haploid lines (Geiger and Gordillo 2009). The 2,000 DH
221 lines were planted in separate plots. No selection was performed at this stage. Each DH line
222 was crossed to a single inbred tester.

223

224 *Year 2*

225

226 The 2,000 DH testcrosses (1 tester x 2,000 DH lines) were evaluated in the testcross 1
227 (TC1) stage. The TC1 stage represented yield measured in unreplicated, two-row plots across
228 6 locations. Selection in the TC1 stage was modelled as the selection on a yield phenotype with
229 a heritability of 0.54. The best performing 400 DH lines were advanced to the next trial based
230 on general combining ability. Each of the 400 DH lines was crossed to 3 inbred testers.

231

232 *Year 3*

233

234 The 1,200 DH testcrosses (3 testers x 400 DH lines) were evaluated in the testcross 2
235 (TC2) stage. The TC2 stage represented yield measured in unreplicated, two-row plots across
236 12 locations. Selection in the TC2 stage was modelled as the selection on a yield phenotype
237 with a heritability of 0.71. The best performing 40 DH lines were advanced to the next trial

238 based on general combining ability. These 40 DH lines were then crossed to 5 ‘elite’ DH lines
239 from the other heterotic group. This produced 200 single cross experimental hybrids.

240

241 *Year 4*

242

243 The 200 experimental hybrids were evaluated in the elite yield trial (EYT) stage. The
244 EYT stage represented yield measured in unreplicated, two-row plots across 24 locations.
245 Selection in the EYT stage was modelled as the selection on a yield phenotype with a
246 heritability of 0.82. The best performing 20 experimental hybrids were advanced to the next
247 trial.

248

249 *Year 5*

250

251 The 20 experimental hybrids were evaluated in the hybrid yield trial 1 (HYT1) stage.
252 The HYT1 stage represented yield measured in unreplicated, two-row plots across 48 locations.
253 Selection in the HYT1 stage was modelled as the selection on a yield phenotype with a
254 heritability of 0.98. The best performing 4 experimental hybrids were advanced to the next
255 trial.

256

257 *Year 6*

258

259 The 4 pre-commercial hybrids were evaluated in the hybrid yield trial 2 (HYT2) stage.
260 The HYT2 stage represented yield measured in on-farm strip tests of pre-commercial hybrids
261 across 600 locations. Selection in the HYT2 stage was modelled as the selection on a yield

262 phenotype with a heritability of 0.99. Commercial hybrids were selected for release from this
263 set of pre-commercial hybrids.

264

265 **Future Breeding**

266

267 The future breeding phase of the simulation modelled alternative breeding programs
268 and compared them against the conventional breeding program. Each breeding program was
269 simulated for an additional 20 years following a common burn-in breeding phase so that each
270 strategy could be evaluated with an equivalent starting point. Two crossing schemes were used
271 in breeding programs with a generation interval of 1 year or less. All breeding programs were
272 constrained to equivalent operating costs (Table 1).

273 **Table 1 Summary of hybrid crop breeding program sizes (number of individuals across both heterotic**
274 **pools) and costs. DH, doubled haploid; EYT, elite yield trial; HYT I, hybrid yield trial 1; HYT II, hybrid**
275 **yield trial 2**

<u>Breeding Program</u>	<u>Parents</u>	<u>DH</u>	<u>TC1</u>	<u>TC2</u>	<u>EYT</u>	<u>HYT1</u>	<u>HYT2</u>	<u>Cost (\$)</u>
Conventional	80	4000	4000	2400	400	40	8	1,276,800
Conventional Plus Genomic Selection Strategies	80	3840	3840	2196	400	40	8	1,273,680
Two-Part Strategies	80	3200	3200	1728	400	40	8	1,271,040

276

277 **Cost Equalizing Strategy**

278 To ensure approximately equal operating costs across different breeding programs, the
279 number of DH lines tested across the product development component was reduced in breeding
280 programs using genomic selection. Table 1 details the exact numbers in each of the breeding
281 programs using genomic selection. Since the two-part breeding programs used multiple cycles
282 per year, they resulted in up to three times as many crosses per year. Therefore, the number of

283 candidate DH lines was reduced further compared to the conventional plus genomic selection
284 breeding programs. To equalize genotyping costs within the population improvement
285 component of the two-part strategy, the number of genotyped offspring per cross was reduced
286 with each additional crossing cycle per year: (i) 1 cycle/year – 80 seeds genotyped per cross;
287 and (ii) 3 cycles/year – 26 seeds genotyped per cross. The remaining components of the
288 breeding programs were kept constant.

289 Estimated costs for genotyping, phenotyping and producing DH lines were used in the
290 equalisation of operating costs. The cost for producing doubled haploid lines was estimated at
291 \$45 based on the lowest publicly advertised price
292 (<http://www.plantbreeding.iastate.edu/DHF/DHF.htm>). Genotype costs and phenotypic
293 evaluation of a yield trial plot, were assumed to be equivalent at a cost of \$15
294 (<http://techservicespro.com/test-locations/>). The cost for crossing, which in any case would be
295 small, was not considered even though it varied between breeding programs.

296

297 **Genomic Selection Training Population & Method**

298

299 Genomic selection in each hybrid crop breeding program used initial training
300 populations comprising the last 3 years of testcross I and II yield trial data from the recent
301 breeding burn-in phase. Separate training populations were developed for each of the two
302 heterotic pools. Thus, the initial training populations comprised phenotypic records on 7,200
303 testcross genotypes. The training populations were updated in subsequent years via a 3-year
304 sliding window approach, in which the oldest year of data was replaced with the data from the
305 newest year. As a consequence of the cost equalisation process the training population sizes
306 were reduced to phenotypic records on 6,858 and 5,664 testcross genotypes for conventional
307 genomic selection and two-part breeding strategies, respectively.

308 The genotypes and testcross phenotypic means of the DH selection candidates were
309 fitted using the genomic selection model used by Bernardo and Yu (2007), and the genetic
310 background of the tester was accounted for by fitting a tester-by-stage fixed effect. A separate
311 genomic selection model was fitted for each heterotic pool. Genomic predictions were
312 calculated using the AlphaSimR function “RRBLUP”. This function fits a ridge regression best
313 linear unbiased prediction model (Whittaker et al. 2000). It models the heterogeneous error
314 variance due to different levels of error in each yield trial by weighting for the effective number
315 of field measurements.

316

317 **Conventional plus Genomic Selection Breeding Programs**

318

319 Three conventional plus genomic selection breeding programs were used to quantify
320 the increase in genetic gain due to the implementation of genomic selection within the
321 traditional structure of a conventional breeding program. The design of these programs used
322 the conventional program as a template. Minimal modifications were made to this template to
323 produce the designs for each strategy.

324

325 *Conventional Genomic Selection Breeding Program*

326

327 The conventional plus genomic selection (ConvGS) breeding program used genomic
328 selection to advance candidate DH lines in the testcross 1 and testcross 2 stages and to select
329 parental lines for the subsequent breeding cycle. The parental lines were selected by choosing
330 the 80 DH lines with the highest genomic estimated breeding values (GEBVs) from a set of
331 candidates that comprised all DH lines from the testcross 2 stage and later yield trials. The

332 minimum cycle time from bi-parental cross to the selection of new parental lines was 3 years,
333 the same as in the conventional breeding program.

334

335 *Genomic Selection Testcross 1 Breeding Program*

336

337 The genomic selection testcross 1 (GS-TC1) breeding program used genomic selection
338 to advance candidate DH lines in the testcross 1 and testcross 2 stages and to select parental
339 lines for the subsequent breeding cycle. The parental lines were selected by choosing the 80
340 DH lines with the highest GEBVs from a set of candidates that comprised all DH lines from
341 the testcross 1 stage and later yield trials. This reduced the minimum cycle time from bi-
342 parental cross to the selection of new parental lines from 3 years in the conventional program
343 to 2 years.

344

345 *Genomic Selection Doubled Haploids Breeding Program*

346

347 The genomic selection doubled haploids (GS-DH) breeding program used genomic
348 selection to advance candidate DH lines in the testcross 1 and testcross 2 stages and to select
349 parental lines for the subsequent breeding cycle. The parental lines were selected by choosing
350 the 80 DH lines with the highest GEBVs from a set of candidates that comprised all DH lines
351 from the DH stage. This reduced the minimum cycle time from bi-parental cross to the selection
352 of new parental lines from 3 years in the conventional program to 1 year.

353

354 **Two-Part Breeding Programs**

355

356 Crossing and selection of new parents in the two-part breeding programs was handled
357 in the population improvement component (Fig. 3), which consisted of one or three crossing
358 cycles per year (TP GS (1 cycle/year) and TP GS (3 cycles/year)). Parents were then grown in
359 greenhouses and at the appropriate stage, crossings were undertaken following a circular
360 scheme or a maximum avoidance of inbreeding scheme as described in the next section.

361 The product development component of the two-part program screened the germplasm
362 to identify new commercial hybrids (Fig. 3). This process began with the production of the new
363 DH lines. The DH lines were screened for testcross and single cross hybrid performance in the
364 same manner as in the conventional plus genomic selection strategies. In two-part breeding
365 programs none of the DH lines were selected for the crossing block, but their genomic and
366 testcross phenotypic data were added to the genomic selection training population. This
367 allowed the genomic selection model used in the population improvement component to be
368 updated over time as new material was evaluated in the field.

369

370 **Crossing of Parents**

371

372 Two crossing schemes were used in hybrid crop breeding programs:

- 373 (i) a circular crossing scheme; and
- 374 (ii) a maximum avoidance of inbreeding crossing scheme.

375 The circular crossing scheme used both between-family and within-family selection to select
376 the 80 selection candidates with the highest genomic estimated breeding values (GEBVs). In
377 each generation 80 crosses were made, ensuring that each parent was used only once as a male
378 and a female. Consequently, this circular crossing scheme is different from the ‘circular design’
379 described by Kimura & Crow (1963), which only conducts within-family selection.

380 Hybrid crop breeding programs using genomic selection, with a generation interval of
381 1 year or less, also used the maximum avoidance of inbreeding crossing scheme. Maximum
382 avoidance of inbreeding is a crossing scheme that maintains uniform contributions and
383 inbreeding coefficients across all crosses (Wright 1921; Kimura and Crow 1963). The
384 maximum avoidance of inbreeding crossing scheme mates the least related crosses in the first
385 generation and maintains this crossing structure over generations. In the present study, the
386 maximum avoidance of inbreeding scheme used within-family genomic selection to select the
387 2 selection candidates with the highest GEBVs per cross as new parents, giving 160 parents in
388 total. In each generation 80 crosses were made, ensuring that each parent was used only once
389 as a male or a female.

390

391 **Comparison of Breeding Programs**

392

393 The performance of each breeding program was measured by comparing genetic gain
394 and genetic variance of hybrids from the EYT stage. These hybrids were the crosses between
395 all DH lines at the EYT stage from the two heterotic pools. The EYT stage was examined
396 because it is the earliest stage in which all breeding programs evaluate DH lines for single cross
397 hybrid performance. Genetic gain and genetic variance in the breeding programs were assessed
398 by plotting mean and variance of true genetic values for hybrids at the EYT stage over time.

399 Accuracy of genomic prediction, defined in the next section, was assessed at the DH
400 stage. The DH stage was examined because it allowed an assessment of the ability to rank all
401 possible DH lines as parents of hybrids in the two heterotic pools. Accuracy of genomic
402 predictions was also assessed in the population improvement components to estimate the
403 accuracy of parent selection in the two-part breeding programs.

404 To aid in visualization, the mean values were centred at the mean value for lines in Year
405 0 for each replicate. Year 0 was defined as the last year of the burn-in phase. Direct comparisons
406 between breeding programs for genetic gain, genetic variance and accuracy of genomic
407 predictions were reported as ratios with 95% confidence intervals (95% CI). These ratios and
408 95% CI were calculated by performing paired Welch's t-tests on log-transformed values from
409 the 10 simulation replicates. The log-transformed differences and 95% CI from the t-test were
410 then back-transformed to obtain ratios (Ramsey and Schafer 2002). All calculations were
411 performed using R (R Development Core Team 2014).

412

413 *Measurement of genomic selection accuracy*

414

415 Accuracy of genomic prediction was defined as the correlation between the general
416 combining ability (GCA) of DH lines and their GEBV. The GCA of an individual was
417 calculated as the sum of all the average effects at the QTL weighted by the individuals genotype
418 at these QTL. The average effects of alleles were calculated using allele frequencies from the
419 corresponding population of DH lines in the other heterotic pool and the true simulated additive
420 and dominance effects for each QTL. Therefore, the GCA of a DH line reflected its average
421 performance as a parent in single cross hybrids when crossed to all DH lines from the other
422 heterotic group.

423

424 **Results**

425 Genomic selection increased the rate of genetic gain compared to phenotypic selection
426 in hybrid crop breeding programs, mainly by reducing the generation interval. The two-part
427 hybrid crop breeding program, with a generation interval of 0.33 years, produced the most
428 genetic gain regardless of genotype-by-year variance. Genomic selection **increased** the
429 selection accuracy compared to phenotypic selection in the early stages of hybrid crop breeding
430 programs. There was a perfect rank correlation between the reduction in the generation interval
431 and the reduction of genetic variance in hybrid crop breeding programs. Genomic selection
432 reduced the efficiency of conversion of genetic variance into genetic gain compared to
433 phenotypic selection. However, the use of the maximum avoidance of inbreeding crossing
434 scheme **slowed** the reduction of genetic variance and increased the efficiency of conversion of
435 genetic variance to genetic gain compared to the circular crossing scheme.

436

437 **Genetic Gain**

438

439 Genomic selection increased the rate of genetic gain compared to phenotypic selection
440 in hybrid crop breeding programs, mainly by reducing the generation interval. This is shown
441 in Fig. 4, which presents the mean genetic value of hybrids at the elite yield trial stage. The
442 first graph shows the trends for the mean for each of the breeding programs evaluated in the
443 future breeding component when genotype-by-year variance is 0. The second graph shows the
444 same trends for genotype-by-year variance of 4. Both graphs show that the two-part breeding
445 program, with the shortest generation interval of 0.33 years, produced the most genetic gain.
446 When genotype-by-year was 0, the TP GS (3 cycles/year) breeding program, which had the
447 shortest generation interval of 0.33 years, generated 2.01 times the genetic gain of the Conv
448 breeding program, which had the longest generation interval of 3 years.

449 Fig. 4 also shows that the ranking of hybrid crop breeding programs for genetic gain
450 was consistent across different genotype-by-year variances. This is shown by the average
451 genetic values of hybrids in the final year (Year 20) of each graph. There was a perfect rank
452 correlation between the generation interval and genetic gain of the breeding programs. Both
453 graphs show the ranking from highest to lowest average genetic value was: TP GS (3
454 cycles/year), GS-DH, TP GS (1 cycle/year), ConvGS, Conv.

455 However, the relative differences between hybrid crop breeding programs using
456 genomic selection and those using phenotypic selection were smaller when genotype-by-year
457 variance was 4. At this level of genotype-by-year variance, the best performing two-part hybrid
458 crop breeding program, TP GS (3 cycles/year), generated 1.96 times the genetic gain of the
459 conventional breeding program. When genotype-by-year variance was 0, this value was 2.01.
460 The relative differences in genetic gain between hybrid crop breeding programs using genomic
461 selection remained constant across different levels of genotype-by-year variance.

462 Fig. 4 also shows that adding genomic selection to the conventional program without
463 reducing the generation interval did not show a significant increase in genetic gain. This is
464 shown by comparing genetic gain in the Conv program with genetic gain in the ConvGS
465 program. The ConvGS program produced 1.08 (95% CI [1.00, 1.16]) and 1.08 (95% CI [0.98,
466 1.18]) times the genetic gain of the Conv program, when genotype-by-year variance was 0 and
467 4, respectively.

468 All breeding programs using genomic selection displayed a similar genetic gain prior
469 to Year 5 (Fig. 4). Year 5 was the first year that hybrids at the elite yield trial stage were derived
470 from parents selected by genomic selection. Therefore, the differences in genetic gain between
471 Year 1 and Year 5 reflect the difference between using genomic selection or phenotypic
472 selection on existing germplasm from the burn-in.

473 The TP GS (1 cycle/year) breeding program did not generate genetic gain in Year 5 and
474 Year 6. This is because no selection was undertaken in the first two generations of future
475 breeding of two-part hybrid crop breeding programs. These first two generations were required
476 to convert the doubled haploid inbred parents from burn-in breeding into outbred parents. The
477 TP GS (3 cycles/year) breeding program did not show this lag as it was able to complete this
478 process and one cycle of selection within the first year of future breeding.

479

480 **Selection Accuracy**

481

482 Genomic selection increased the selection accuracy compared to phenotypic selection
483 in the early stages of hybrid crop breeding programs. This is show in Fig. 5, which plots the
484 correlations between the simulated, true general combining abilities (GCA) for DH lines at the
485 DH stage and their GEBV. The first graph shows the mean selection accuracy for all breeding
486 programs when genotype-by-year variance was 0. The second graph shows the same trends for
487 genotype-by-year variance of 4. The selection accuracies in the hybrid crop breeding programs
488 using genomic selection were higher than those using phenotypic selection. In Year 1 when
489 genotype-by-year variance was 0, all hybrid crop breeding programs using genomic selection
490 had a selection accuracy of 0.73 while the Conv breeding program had a selection accuracy of
491 0.24.

492 Fig. 5 also shows that selection accuracies in all hybrid crop breeding programs
493 decreased over the years of the simulation, and this decrease had a perfect rank correlation with
494 the generation interval of hybrid crop breeding programs. This is shown by the average
495 selection accuracy in the final year (Year 20) of each figure. Both figures show the ranking
496 from highest to lowest average selection accuracy was: ConvGS, TP GS (1 cycle/year), GS-

497 DH, TP GS (3 cycles/year). The ranking of hybrid crop breeding programs for selection
498 accuracy was consistent across different genotype-by-year variances.

499 Selection accuracy in the two-part hybrid crop breeding programs was also measured
500 in the population improvement stage. This is shown in Fig. 6, which plots the correlation
501 between the simulated, true GCA for parental candidates and their GEBV for each cycle of
502 crossing. There were one and three cycles of crossing per year. Therefore, each cycle is plotted
503 at one third of a year increments in Fig. 6. The first figure shows the selection accuracy when
504 genotype-by-year variance was 0. The second figure shows the selection accuracy when
505 genotype-by-year variance was 4. Fig. 6 shows the change in selection accuracy over time
506 differed between the two-part breeding programs. The first figure shows that the TP GS (1
507 cycle/year) breeding program displayed a gradual decrease in selection accuracy over time.
508 The TP GS (3 cycles/year) breeding program displayed a faster decrease in selection accuracy
509 and selection accuracy becomes 0 in Year 16. The second figure shows that selection accuracy
510 remained constant over time when genotype-by-year was 4. Both graphs show yearly
511 oscillations which correspond to the yearly updating of the training population.

512

513 **Genetic Variance**

514

515 All hybrid crop breeding programs displayed a reduction of genetic variance over
516 simulation years. However, hybrid crop breeding programs using genomic selection caused a
517 faster reduction of the genetic variance than the hybrid crop breeding program using
518 phenotypic selection. This is shown in Fig. 7, which plots the change in genetic variance of
519 hybrids at the elite yield trial stage over simulation years. The first graph shows the change in
520 genetic variance for each future breeding program when genotype-by-year variance equals 0.
521 The second graph shows the same breeding programs when genotype-by-year variance equals

522 4. The conventional breeding program displayed a gradual and consistent reduction of genetic
523 variance across simulation years. All hybrid crop breeding programs using genomic selection
524 displayed a large initial reduction of genetic variance. At Year 3 when genotype-by-year was
525 0, the Conv breeding program had 1.25 times the genetic variance of the ConvGS breeding
526 program. The reduction of genetic variance in subsequent years differed between the hybrid
527 crop breeding programs using genomic selection.

528 Fig. 7 shows the change in genetic variance had a perfect rank correlation with the
529 generation interval of hybrid crop breeding programs using genomic selection. This is shown
530 by the genetic variance in the final year (Year 20) of each plot. Both graphs show the ranking
531 from highest to lowest genetic variance was: ConvGS, TP GS (1 cycle/year), GS-DH, TP GS
532 (3 cycles/year). When genotype-by-year variance was 0, the TP GS (3 cycles/year) breeding
533 program had 0.44 times the genetic variance of the GS-DH breeding program and 0.07 times
534 that of the Conv breeding program in Year 20. When genotype-by-year variance was 4, these
535 values were 0.53 and 0.11, respectively. The ranking of hybrid crop breeding programs for
536 genetic variance was consistent across different genotype-by-year variances.

537

538 **Crossing Schemes**

539

540 Genomic selection with the maximum avoidance of inbreeding crossing scheme
541 produced similar genetic gain over the circular crossing scheme, but maintained higher levels
542 of genetic variance. Therefore, the maximum avoidance of inbreeding crossing scheme had a
543 higher conversion efficiency compared to the circular scheme. This is shown in Fig. 8, which
544 plots the genetic gain against the reduction of genetic variance for each hybrid crop breeding
545 program-crossing scheme combination. All hybrid crop breeding programs using genomic
546 selection with the circular crossing scheme had lower efficiency than the conventional breeding

547 program using phenotypic selection. The first graph shows the trends for the mean values when
548 genotype-by-year variance is 0. The second graph shows the same trends for genotype-by-year
549 variance of 4. Fig. 8 also shows that hybrid crop breeding programs using genomic selection
550 with the maximum avoidance crossing scheme had comparable conversion efficiency to the
551 conventional breeding program using phenotypic selection. This ranking was consistent across
552 different levels of genotype-by-year variance. Because the conversion efficiencies followed a
553 very clear non-linear path (Fig. 8) it was not possible to formally test for statistically significant
554 differences.
555

556 **Discussion**

557 The results of this study highlight four points for discussion:

- 558 (i) the impact of reciprocal recurrent genomic selection on the drivers of genetic
559 gain in hybrid crop breeding programs;
- 560 (ii) the impact of crossing schemes on genetic gain in hybrid crop breeding
561 programs;
- 562 (iii) the limitations of the simulation undertaken; and
- 563 (iv) the practical implementation of the two-part strategy in real hybrid crop
564 breeding programs.

565

566 **The impact of reciprocal recurrent genomic selection on the drivers of genetic gain in** 567 **hybrid crop breeding programs**

568

569 *Generation Interval*

570 Reciprocal recurrent genomic selection increased the rate of genetic gain in hybrid crop
571 breeding programs, mainly by a reduction in generation interval. In an animal breeding context
572 reciprocal recurrent genomic selection is a method that uses crossbred data to predict parent
573 specific breeding values (Kinghorn et al. 2010). In a hybrid crop breeding context, parent
574 specific breeding values can be predicted using hybrid phenotypes and inbred parental
575 genotypes, and previous studies have predicted that reciprocal recurrent genomic selection
576 could improve rates of genetic gain (Longin et al. 2013; Rembe et al. 2019). The results in the
577 present study support this, showing that using reciprocal recurrent genomic selection increased
578 genetic gain within a conventional hybrid crop breeding program design. However, the
579 development of inbred parents takes time, which increases the generation interval (Griffing

580 1975). The recently proposed two-part strategy (Gaynor et al., 2017) provides a framework to
581 remove this time delay by using outbred parents. In the context of hybrid crop breeding
582 programs, the two-part strategy enables reciprocal recurrent genomic selection and the
583 completion of multiple generations per year. The two-part strategy for hybrid crop breeding
584 programs outlined in this study used this principle to achieve a generation interval of 0.33
585 years. Over the first ten years of breeding this drove a 1.33-fold increase in genetic gain
586 compared to the best performing conventional plus genomic selection breeding program. While
587 this is an important increase in rate of genetic gain, it is lower than the 3-fold expectation based
588 on the breeder's equation (Lush 1943), which can be explained by decreases in the selection
589 accuracy and genetic variance over time.

590

591 *Selection Accuracy*

592 Rapid decreases in genomic selection accuracy were observed in two-part hybrid crop
593 breeding programs and can be explained by:

- 594 (i) a larger number of generations separating individuals in the 'training
595 population' and selection candidates; and
596 (ii) the reduction of the genetic variance of the selection candidates due to genomic
597 selection.

598 In the present study, the two-part breeding program, with the shortest generation
599 interval of 0.33 years, displayed the fastest decrease in genomic selection accuracy over time.
600 The decreasing trend in selection accuracy is consistent with previous simulations of two-part
601 breeding strategy for inbred crops (Gaynor et al. 2017; Gorjanc et al. 2018), although, the trends
602 in the present study were much larger. There are a number of possible reasons for this.

603 Fluctuations in the average effect (of an allele substitution) over time driven by different
604 trait genetic architectures can partly explain the lower selection accuracy in the present study

605 compared to previous studies. The present study simulated a trait including both additive and
606 dominance variation. When dominance contributes to trait variation, the average effects of
607 alleles can change due to changes in allele frequencies (Falconer and Mackay 1996). The
608 genomic estimated average effects of alleles are confounded with the ‘training population’
609 allele frequency, so they may not provide good estimates for average effects of alleles in the
610 selection candidate whose allele frequency can be different. The previous studies of Gaynor et
611 al. (2017) and Gorjanc et al. (2018) simulated strictly additive traits, so the average effects of
612 alleles remained constant.

613 The more rapid decrease in genomic selection accuracy in this study compared to
614 previous implementations of two-part strategies may also be partially attributable to different
615 strategies for updating the ‘training population’. Previous studies accumulated records in the
616 ‘training population’ across all years of future breeding, while in this present study, the size of
617 the ‘training population’ was kept constant over time by updating the training population using
618 a 3-year sliding window approach. This resulted in a training population that trended to half
619 the size of that used in Gaynor et al. (2017) and a quarter of the size used in Gorjanc et al.
620 (2018). Thirdly, the higher number of generations per year used in population improvement
621 caused a greater divergence in relatedness between the training population and selection
622 candidates compared to Gaynor et al. (2017).

623

624 *Genetic Variance*

625 The two-part strategy caused a rapid reduction of genetic variance in hybrid crop
626 breeding programs. The reduction of genetic variance when using the two-part strategy may
627 largely depend on how the additional genotyping costs in the population improvement are
628 offset. Under a fixed budget, the additional genotyping costs could be offset in two simple
629 ways. In the first strategy, the number of genotyped individuals per cross could be held constant

630 while the number of parents and crosses could be reduced for each additional cycle of
631 population improvement per unit time. However, such a strategy would cause a reduction of
632 the effective population size and this could expose the breeding program to accelerated
633 reduction of genetic variance due to genetic drift (Charlesworth 2009). Therefore, to overcome
634 this risk, this study implemented an alternative strategy that maintained the number of parents
635 and crosses but reduced the number of genotyped individuals per cross. This alternative
636 strategy aims to maintain the effective population size and mitigate the reduction of genetic
637 variance due to genetic drift. The results from the present study show that this alternative
638 strategy was not sufficient to mitigate the accelerated reduction of genetic variance in two-part
639 hybrid crop breeding programs using the circular crossing scheme. The two-part breeding
640 program using the circular crossing scheme, with a generation interval of 0.33 years, displayed
641 the fastest decrease in genetic variance over time. The large decreases in genetic variance with
642 such an aggressive approach limited long-term genetic gain. Therefore, the maximum
643 avoidance of inbreeding crossing scheme was explored.

644

645 **The impact of crossing schemes on the conversion efficiency in hybrid crop breeding** 646 **programs**

647 Reciprocal recurrent genomic selection with the maximum avoidance crossing scheme
648 had the highest conversion efficiency, comparable to that of phenotypic selection. Genomic
649 selection with the circular design crossing scheme had the lowest conversion efficiency. The
650 higher conversion efficiency of the maximum avoidance crossing scheme was driven by a
651 slower reduction of genetic variance over time compared to the circular scheme.

652 Crossing schemes are designed to manage genetic variance in breeding populations by
653 avoiding the mating of closely related individuals. The circular scheme used in the present
654 study does this by equalising the contributions of each cross. However, it can create higher

655 levels of inbreeding compared to other crossing schemes because multiple parents could be
656 selected from the same family if sufficient differences in the estimates of family means exist.
657 Over generations this could result in large decreases in genetic variance due to genetic drift.
658 Maximum avoidance of inbreeding is a crossing scheme that maintains uniform contributions
659 and inbreeding coefficients across all crosses (Wright 1921; Kimura and Crow 1963). The
660 maximum avoidance of inbreeding crossing scheme mates the least related crosses in the first
661 generation and uses within-family truncation selection to choose new parents. In the present
662 study, maximum avoidance worked well. Two-part hybrid crop breeding programs using
663 maximum avoidance showed much smaller reductions in genetic variance than those using
664 circular scheme.

665 The higher conversion efficiency of the maximum avoidance crossing scheme was
666 important for the long-term genetic gain of hybrid crop breeding programs. Long-term genetic
667 gain is dependent on the ability to exploit the within-family component of a breeding value,
668 which is called the Mendelian sampling term (Wray and Thompson 1990; Meuwissen 1997;
669 Pong-Wong and Woolliams 1998; Woolliams et al. 2015). Reciprocal recurrent genomic
670 selection enabled a high within-family selection accuracy (Fig. 8), which both the maximum
671 avoidance and circular scheme crossing schemes exploit. However, the maximum avoidance
672 scheme preserved more genetic variance which also maintained a higher within-family
673 selection accuracy. Therefore, the two-part hybrid crop breeding programs using the maximum
674 avoidance crossing scheme displayed the highest long-term genetic gain.

675

676 **The limitations of the study**

677

678 The simulations conducted in the present study did not model the full complexity of
679 actual hybrid crop breeding programs. In this section the limitations and impact of key
680 assumptions are discussed:

- 681 (i) assumptions about the genetic architecture;
- 682 (ii) assumptions that impact genomic selection accuracy;
- 683 (iii) assumptions about the reciprocal recurrent genomic selection model;
- 684 (iv) assumptions about the making of crosses;
- 685 (v) assumptions about the ratio between genotyping and phenotyping costs;
- 686 (vi) assumptions about the complexity of the breeding goal;
- 687 (vii) assumptions that impact the maintenance of genetic variance.

688 *Assumptions about the genetic architecture*

689 The simulated trait was controlled by 3,000 QTL, a dominance degree of 0.9, a
690 dominance variance of 0.3, and a heterotic pool split that occurred 100 generations ago. These
691 values were chosen as they produced long-term trends for inbred and hybrid performance that
692 reflected those observed in real data (Troyer and Wellin 2009). The main focus of the present
693 study was on genetic gain, which relates to general combining ability (Sprague and Tatum
694 1942). We hypothesise that tuning the parameters to match long term genetic gain results in
695 greater uncertainty in variance due to specific combining ability rather than variance due to
696 general combining ability. Therefore, the assumptions about the genetic architecture of the
697 simulated trait are likely to have limited impact on the conclusions of this study.

698

699 *Assumptions that impact genomic selection accuracy*

700 The genomic selection accuracies observed in these simulations are likely higher than
701 those in real-world conditions. As previously described by Gaynor et al. (2017), this was
702 because of conditions in the simulation that favoured high genomic selection accuracy such as

- 703 (i) molecular markers with no genotyping errors;
704 (ii) genetic control of the trait that did not involve epistasis; and
705 (iii) a closed breeding program.

706 The accuracy of genomic selection affects the genetic gain of the simulated breeding
707 programs. These effects should affect all hybrid crop breeding programs using genomic
708 selection similarly, which suggests that using the two-part strategy should still outperform the
709 other genomic selection breeding strategies. However, the relative performance of breeding
710 programs using phenotypic selection to the breeding programs using genomic selection could
711 change. If this were to occur, the hybrid crop breeding programs using the two-part strategy
712 should still outperform the conventional breeding program because of the magnitude of
713 difference observed in the simulation.

714

715 *Assumptions about the making of crosses*

716 An important assumption was how the present study performed crossing. As in Gaynor
717 et al. (2017), the present study did not consider maturity differences between male and female
718 parents. Maturity differences between prospective parents could result in particular crosses
719 being missed, which could bring additional costs. Maturity differences could have specific
720 impacts on the two crossing schemes used in the present study. Maturity differences could
721 prevent each parent from being used twice, once as a male and once as a female, which was
722 assumed in the circular design crossing scheme. Maturity differences could also result in
723 missed crosses in the maximum avoidance crossing scheme. However, the flexibility of the
724 maximum avoidance crossing scheme could account for maturity differences between
725 prospective parents by replacing them by their next best-ranking siblings. Such maturity
726 differences could have a more substantial impact on two-part breeding programs compared to
727 the conventional plus genomic selection breeding programs, due to a higher number of crosses

728 per year and lower seed availability in the two-part. However, the implications of maturity
729 differences are likely to be relatively small and therefore have a relatively small effect on the
730 performance of hybrid crop breeding programs.

731

732 *Assumptions about the ratio between genotyping and phenotyping costs*

733 The present study considered the costs of genotyping and phenotyping to be equal, and
734 may not reflect current or future cost ratios for different breeding operations. The ratio between
735 genotyping and phenotyping costs is more than likely to reduce in the future. Improvements in
736 technology and the benefits from economy of scale could reduce genotyping costs in the future.
737 Phenotyping costs could also reduce in the future, but likely at a slower rate. Any reduction in
738 the ratio between genotyping and phenotyping costs would reduce the reallocation of resources
739 required for the deployment of genomic selection, which is likely to favour hybrid crop
740 breeding programs with high genotyping requirements, such as those using the two-part
741 strategy.

742

743 *Assumptions about the complexity of the breeding goal*

744 The commercial products of hybrid crop breeding programs are a small number of
745 hybrid varieties that are each planted on thousands to millions of acres. Therefore, these
746 commercial products need to meet a wide range of requirements across a wide range of
747 potential target production environments. Hybrid crop breeding programs must consider
748 multiple traits relating to agronomic performance, disease resistance, and end-use quality. The
749 hybrid crop breeding programs examined in this simulation only considered a single
750 quantitative trait with 3,000 QTL. We assumed that this trait represented grain yield. However,
751 it could equally represent a selection index with a few additional assumptions: all traits are
752 measured on all individuals, all traits are pleiotropic, and economic merit is linear.

753

754 *Assumptions that impact the maintenance of genetic variance*

755 The present study showed there might be a potential risk for the rapid reduction of
756 genetic variance in two-part hybrid crop breeding programs. The present study may
757 overemphasise the issue of reduction of genetic variance in hybrid crop breeding programs,
758 due to a combination of simulating a closed system, non-epistatic trait architecture, high
759 genomic selection accuracy and a simplified breeding goal. Two crossing schemes to mitigate
760 the reduction of genetic variance, the circular design and maximum avoidance, were used in
761 the present study. Still, the two-part hybrid crop breeding programs displayed rapid reductions
762 of genetic variance using both crossing schemes. Optimal contribution selection (Meuwissen
763 1997; Kinghorn et al. 1999; Woolliams et al. 2015; Gorjanc et al. 2018) and optimal cross
764 selection (Allier et al. 2019) enable more complex crossing designs to avert the reduction of
765 genetic variance in breeding programs. Such methods have the added benefit of balancing the
766 choice between the maintenance of genetic variance, versus genetic gain and can, therefore, be
767 tailored to prioritise short- and long-term genetic gain in two-part hybrid crop breeding
768 programs. However, due to the large number of simulated scenarios optimal contribution
769 methods were not used because they have high computational costs compared to the crossing
770 schemes used in the present study.

771

772 **Further opportunities for the two-part strategy in real hybrid crop breeding programs**

773

774 The present study demonstrated that the deployment of the two-part strategy could
775 significantly increase genetic gain compared to current hybrid crop breeding program designs.
776 However, the implementation of the two-part strategy in hybrid crop breeding programs
777 requires some additional developments.

778 Alternative strategies for managing genetic variation need to be explored. The
779 maximum avoidance of inbreeding crossing scheme used in this simulation only works in a
780 closed breeding pipeline, so it cannot accommodate germplasm exchange. The maximum
781 avoidance of inbreeding crossing scheme also offers little freedom to alter the balance between
782 genetic gain and maintenance of diversity. More advanced strategies based on optimal
783 contribution selection offer the potential to address both these limitations and should be
784 explored (Meuwissen 1997; Woolliams et al. 2015). Further, a pre-breeding process could be
785 integrated into the population improvement component of each heterotic pool to introgress
786 external germplasm from gene banks or other sources (Gorjanc et al. 2016; Yang et al. 2019).

787 The creation of an outbred training population could partially mitigate the rapid
788 reduction of genomic selection accuracy in two-part hybrid crop breeding programs. The
789 entries in this outbred training population would consist of progeny from testcrosses to
790 genotyped plants from the population improvement component. The progeny would be
791 evaluated in plots to test the merit of their parents. This strategy is similar to older strategies
792 for early testing of inbred lines (Sprague 1946). This data could be used to increase selection
793 accuracy because it would reduce the genetic distance between the training and prediction
794 individuals, which is known to be a significant determinant of accuracy (Habier et al. 2007;
795 Clark et al. 2012). For example, in the current simulation the maximum number of generations
796 between the selection candidates and the most recent training population lines could be cut in
797 half by bypassing the creation of DH lines. Finally, a more speculative additional use of such
798 an outbred training population would be to directly derive inbred lines from this population via
799 apomixis.

800 The further opportunities for the two-part strategy in hybrid crop breeding programs,
801 outlined here, incur further costs such as additional genotyping. Therefore, the implementation
802 of the two-part strategy requires resource reallocation to account for these additional costs. The

803 present study reallocated resources by reducing the number of selection candidates evaluated
804 in the product development pipeline and maximised the number of crosses per generation in
805 the population improvement component. However, the optimal resource reallocation strategy
806 may differ between breeding programs. For example, breeding programs serving a smaller
807 geographical region could reduce the number of trial locations in the product development
808 pipeline. Therefore, optimal resource reallocation strategies require further research.

809 The deployment of the two-part strategy is a fundamental change to current practices
810 in breeding programs and is currently untested empirically. Therefore, the development of
811 bespoke transition strategies is required which would build the data sets the two-part strategy
812 requires and build empirical confidence in its performance.

813

814 **Conclusions**

815 Hybrid crop breeding programs using a two-part strategy produced the most genetic gain,
816 but a maximum avoidance of inbreeding crossing scheme was required for it to increase long-
817 term genetic gain. The two-part strategy uses outbred parents to complete multiple generations
818 per year in hybrid crop breeding programs. In contrast conventional plus genomic selection
819 strategies are limited in this regard by the time they take to develop inbred lines. The maximum
820 avoidance of inbreeding crossing scheme manages genetic variance by maintaining uniform
821 contributions and inbreeding coefficients across all crosses. This study performed stochastic
822 simulations to quantify the potential of a two-part strategy in combination with two crossing
823 schemes to increase the rate of genetic gain in hybrid crop breeding programs. Three main
824 conclusions can be drawn from the results:

825 (i) the implementation of genomic selection in hybrid crop breeding programs
826 increases the rate of genetic gain;

827 (ii) the two-part strategy was the most cost-effective strategy for implementing
828 genomic selection in hybrid crop breeding programs.

829 (iii) two-part hybrid crop breeding programs completing multiple selection cycles
830 per year should use crossing schemes to manage genetic variance.

831 As well as the benefits outlined in this study, the flexibility of the two-part strategy offers
832 further opportunities to integrate new technologies to further increase genetic gain in hybrid
833 crop breeding programs, such as the use of outbred training populations. However, the
834 practical implementation of the two-part strategy will require the development of bespoke
835 transition strategies to fundamentally change the data, logistics, and infrastructure that
836 underpin hybrid crop breeding programs.

837

838

839

840

841

842

843

844

845

846

847

848

849

850

851

852 **Author contributions statement**

853 OP and JMH conceived the study. OP, JMH and RCG designed the study. OP developed the
854 plant breeding program simulation. OP wrote the manuscript with input from all authors. All
855 authors read and approved the final manuscript.

856

857 **Acknowledgments**

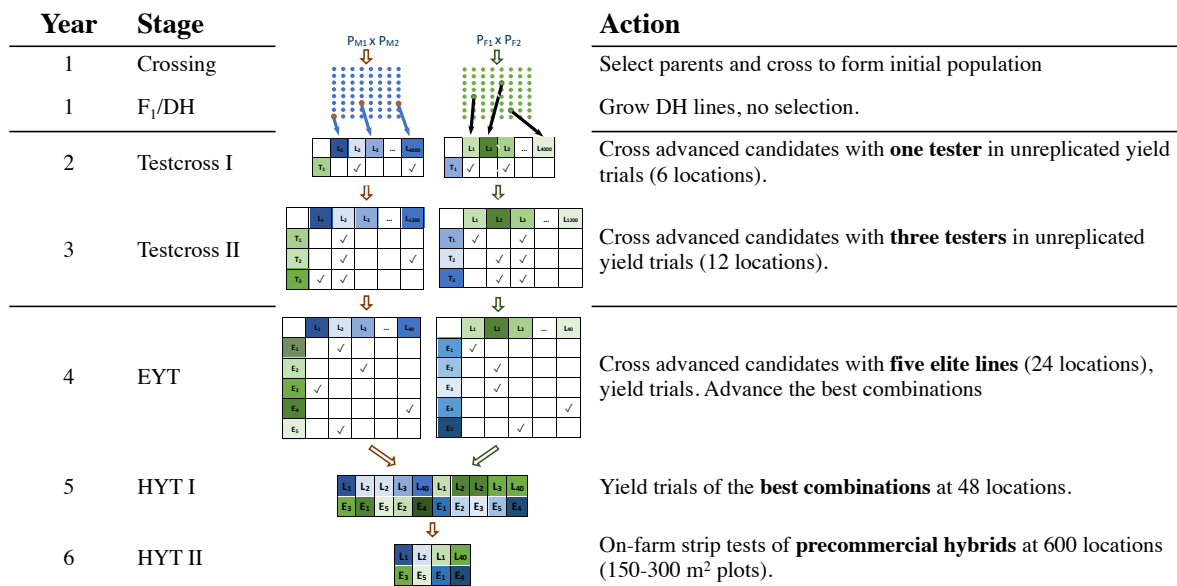
858 The authors acknowledge the financial support from the BBSRC ISPG to The Roslin Institute
859 (BBS/E/D/30002275). This work has made use of the resources provided by the Edinburgh
860 Compute and Data Facility (ECDF) (<http://www.ecdf.ed.ac.uk>).

861

862 **Conflict of interest**

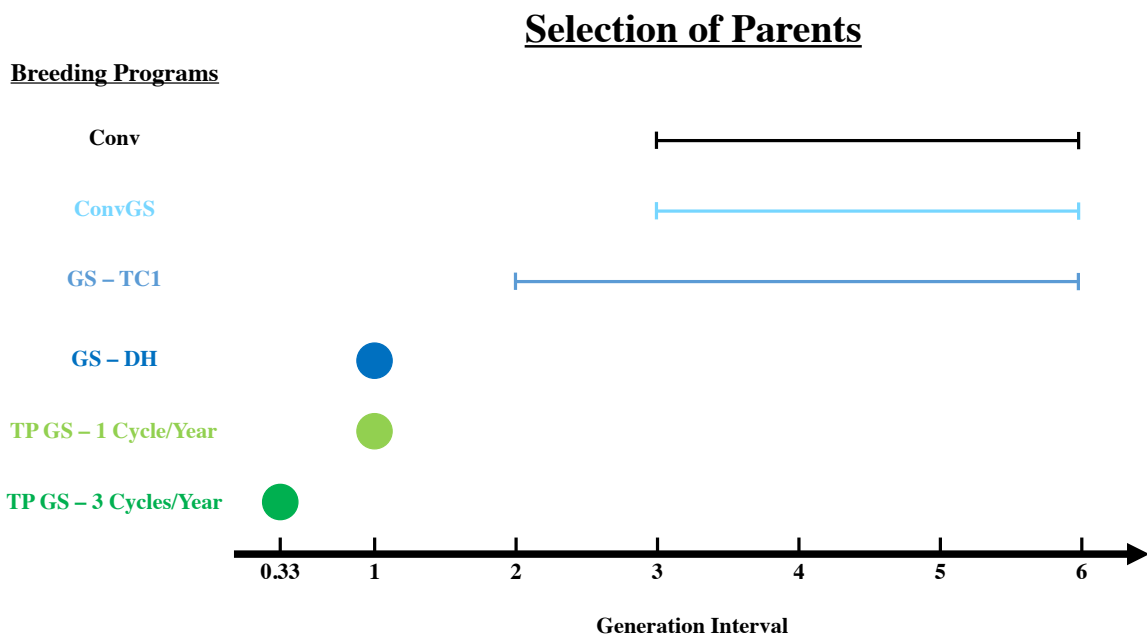
863 The authors declare that they have no conflict of interest.

864 Figures



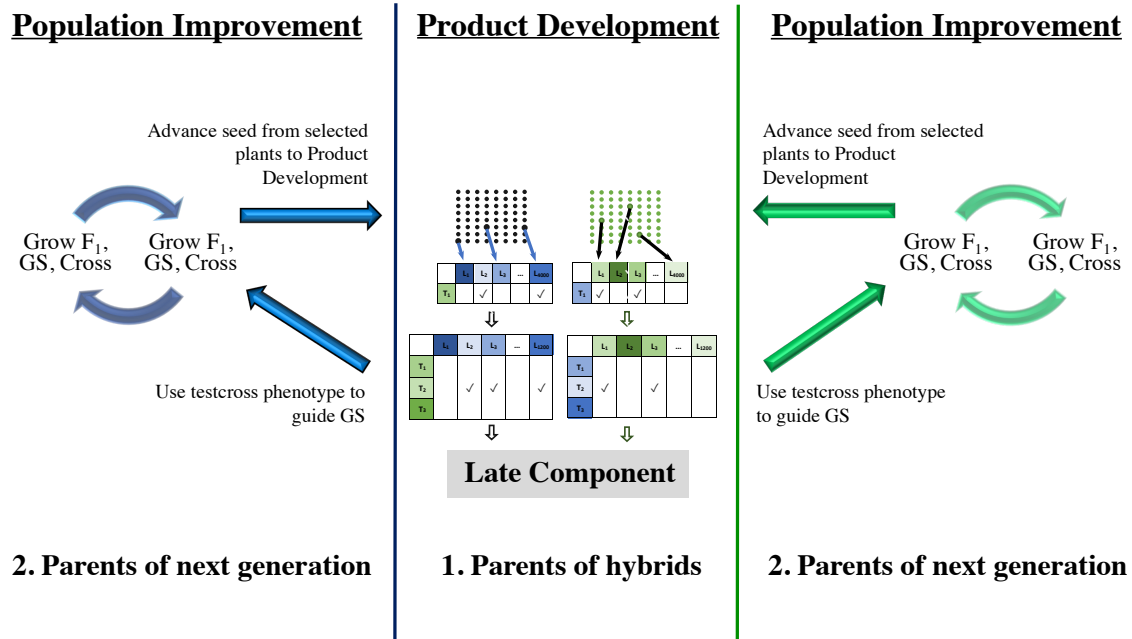
865
866
867
868
869

Fig. 1 Overview of breeding schemes for the conventional hybrid crop breeding program (used in burn-in breeding) and the breeding programs using standard genomic selection strategies. DH, doubled haploid; EYT, elite yield trial; HYT I, hybrid yield trial 1; HYT II, hybrid yield trial 2



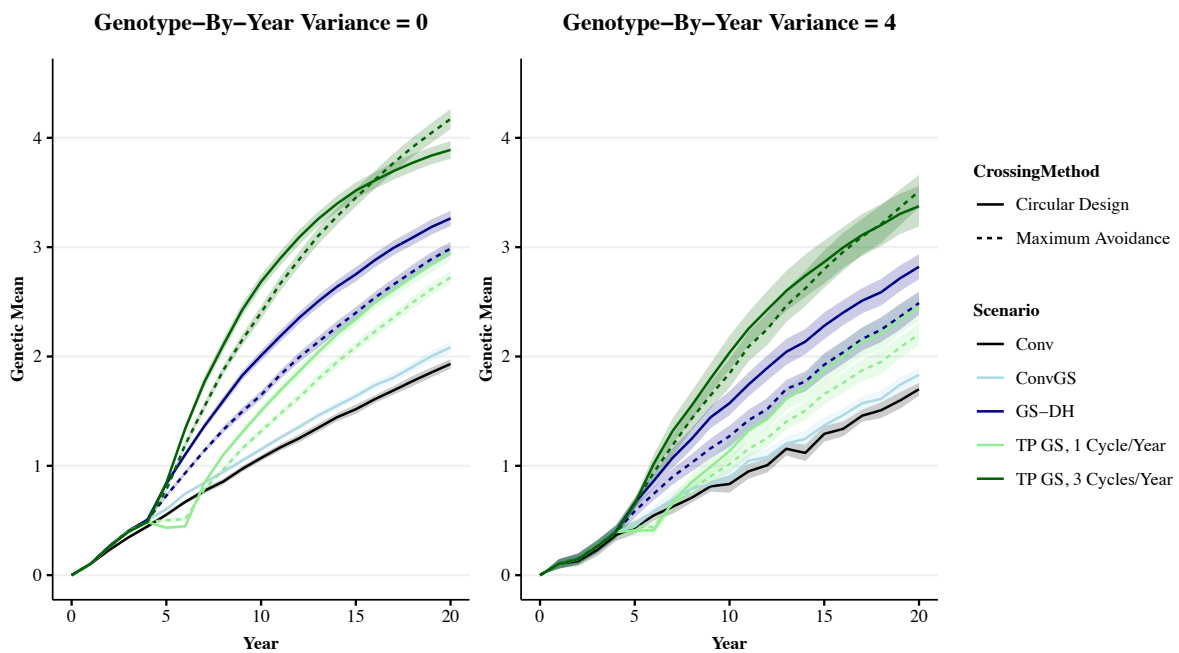
870
871
872
873
874
875
876

Fig. 2 The generation intervals of hybrid crop breeding programs. Conv, conventional breeding program; ConvGS, conventional program with genomic selection; GS-TC1, genomic selection program with parents selected in the testcross 1 stage; GS-DH, genomic selection program with parents selected in the doubled haploid stage; TP GS, 1 Cycle/Year, two-part program with genomic selection; TP GS, 3 Cycles/Year, two-part program with genomic selection



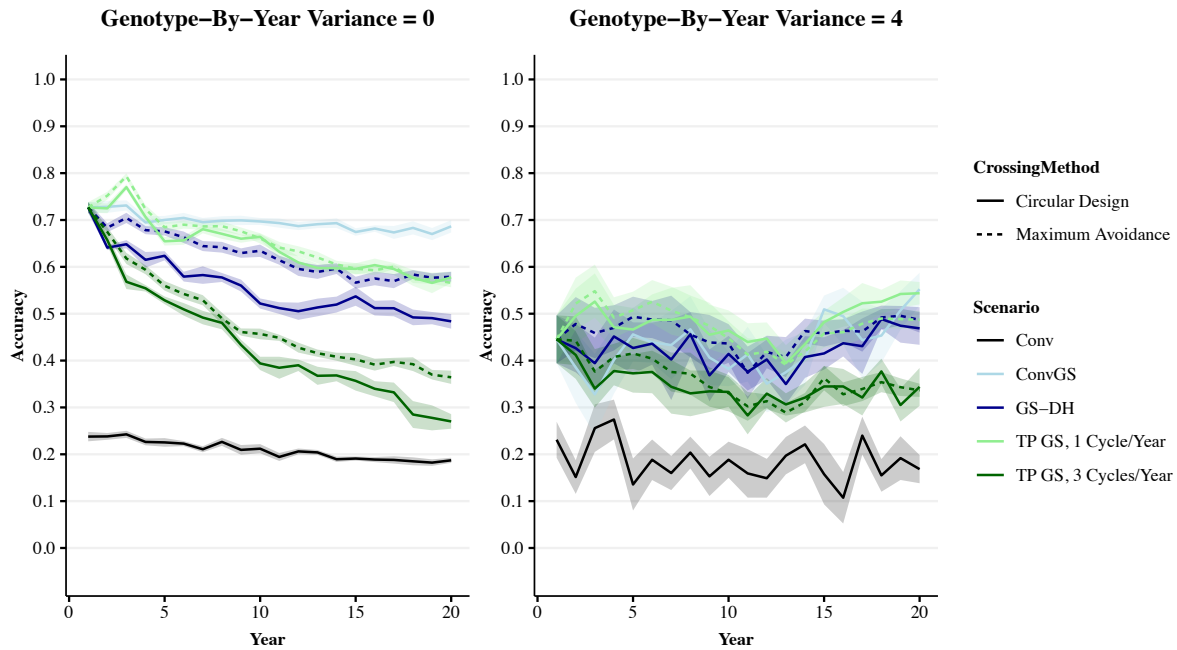
877
878
879
880
881

Fig. 3 Overview of the two-part strategies with testcross genomic selection for hybrid crop breeding programs (TP GS, 1 Cycle/Year, TP GS, 3 Cycles/Year). The number of crosses differed for each two-part breeding program to maintain equal operating costs. See Table 1

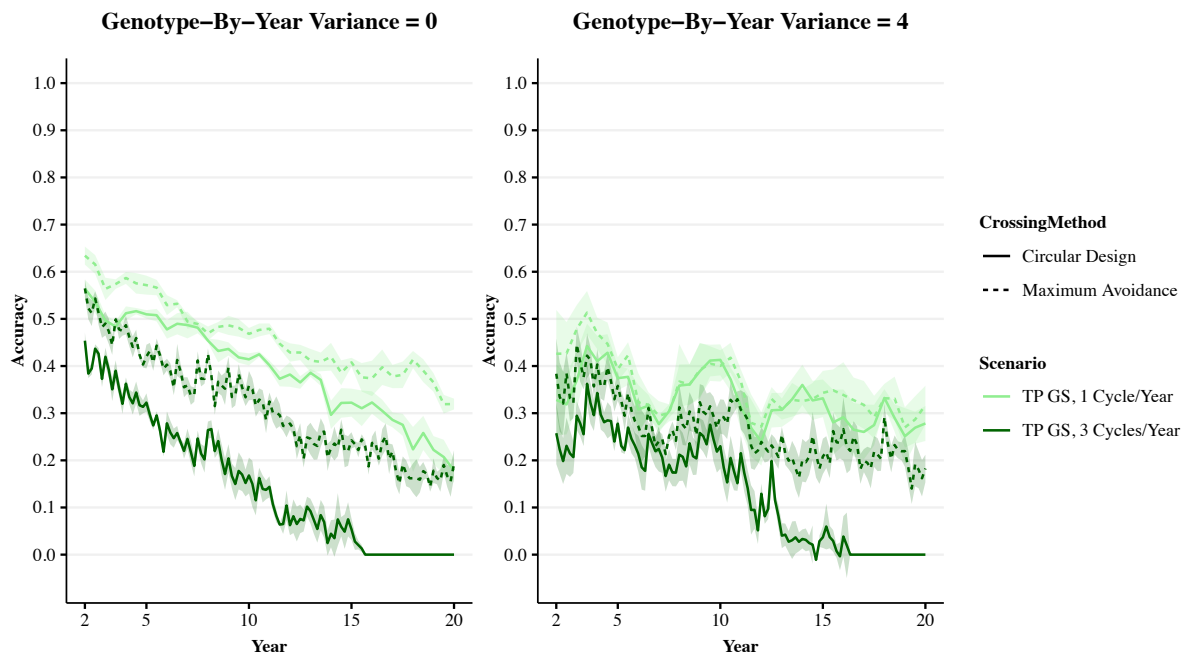


882
883
884
885
886
887
888
889
890

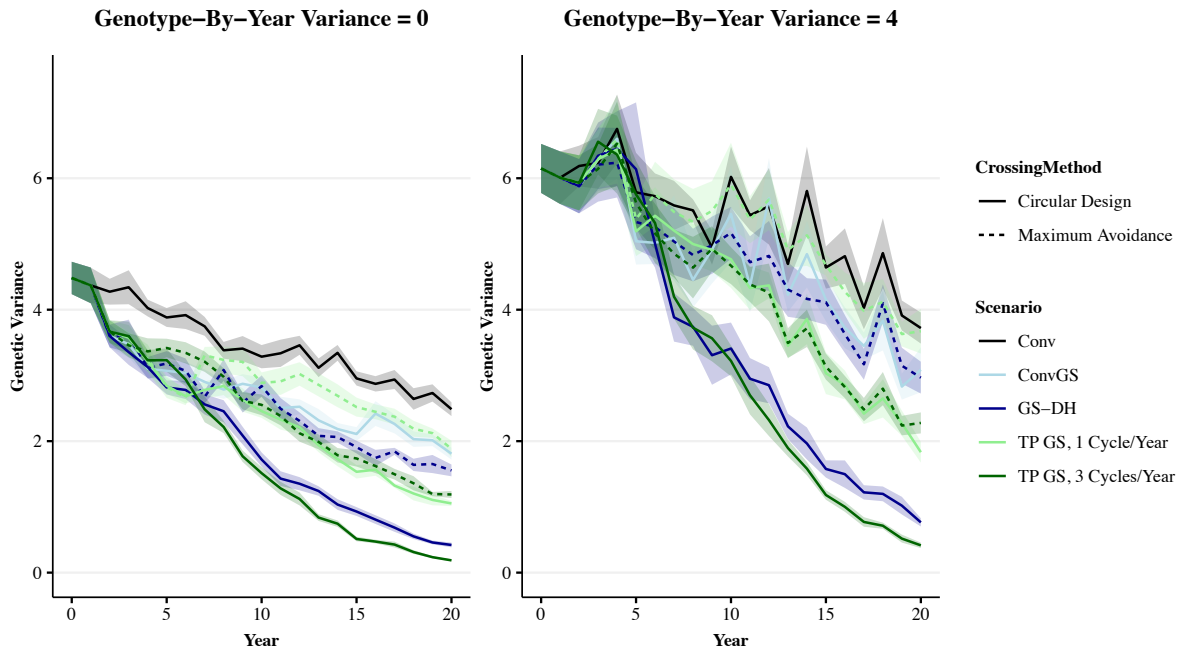
Fig. 4 Genetic gain for all breeding programs over simulation years. Genetic gain when genotype-by-year variance was 0 and 4. Genetic gain is expressed as mean genetic value of hybrids at the elite yield trial stage over time. The mean genetic value for all replicates were centered on 0 in Year 4. Means for all 10 replicates are shown with dark lines, with the shaded area representing the 95% confidence intervals of the mean. Conv, conventional breeding program; Conv GS, conventional program with genomic selection; GS-DH, genomic selection program with parents selected in the doubled haploid stage; TP GS, 1 Cycle/Year, two-part program with genomic selection; TP GS, 3 Cycles/Year, two-part program with genomic selection



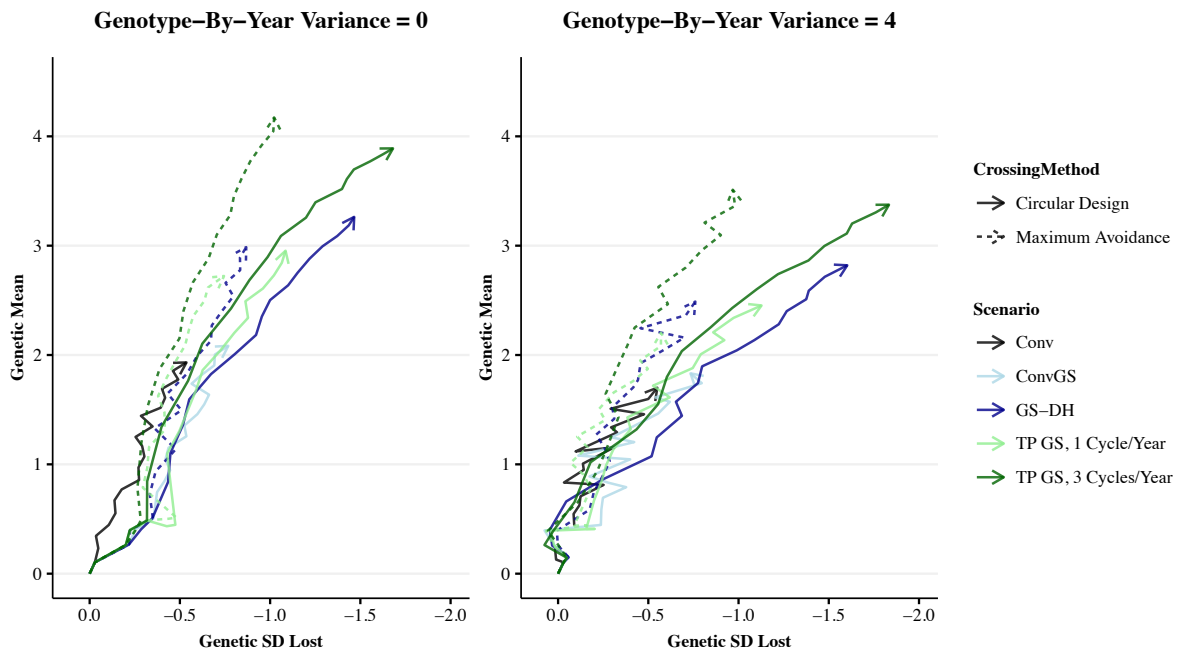
891
 892 **Fig. 5 Selection accuracy for all breeding programs over selection cycles.** Selection accuracy when genotype-by-year variance was 0 and 4. Selection accuracy is expressed as the correlation between true and predicted general combining abilities of doubled haploid (DH) lines at the DH stage over selection cycles. Means for all 10 replicates are shown with dark lines, with the shaded area representing the 95% confidence intervals of the mean. Conv, conventional breeding program; Conv GS, conventional program with genomic selection; GS-DH, genomic selection program with parents selected in the doubled haploid stage; TP GS, 1 Cycle/Year, two-part program with genomic selection; TP GS, 3 Cycles/Year, two-part program with genomic selection
 893
 894
 895
 896
 897
 898
 899



900
 901 **Fig. 6 Within-family selection accuracy in the population improvement components of the two-part**
 902 **breeding programs.** Within-family selection accuracy when genotype-by-year variance was 0 and 4. Within-
 903 family selection accuracy is expressed as the correlation between the simulated, true general combining abilities
 904 (GCA) for parental candidates and their predicted GCA. Means for all 10 replicates are shown with dark lines,
 905 with the shaded area representing the 95% confidence intervals of the mean. TP GS, 1 Cycle/Year, two-part
 906 program with genomic selection; TP GS, 3 Cycles/Year, two-part program with genomic selection
 907



908
 909 **Fig. 7 Genetic variance for all breeding programs over simulation years.** Genetic variance when genotype-by-year variance was 0 and 4. Genetic variance is expressed as the genetic variance among hybrids at the elite yield trial (EYT) stage over simulation years. Means for all 10 replicates are shown with dark lines, with the shaded area representing the 95% confidence intervals of the mean. Conv, conventional breeding program; Conv GS, conventional program with genomic selection; GS-DH, genomic selection program with parents selected in the doubled haploid stage; TP GS, 1 Cycle/Year, two-part program with genomic selection; TP GS, 3 Cycles/Year, two-part program with genomic selection
 911
 912
 913
 914
 915
 916



917
 918 **Fig. 8 Conversion efficiency for all breeding programs over simulation years.** Conversion efficiency when genotype-by-year variance was 0 and 4. Conversion efficiency is presented as the genetic gain against the genetic variance among hybrids at the elite yield trial between Year 0 and Year 20 of the simulation. Means for all 10 replicates are shown. Conv, conventional breeding program; Conv GS, conventional program with genomic selection; GS-DH, genomic selection program with parents selected in the doubled haploid stage; TP GS, 1 Cycle/Year, two-part program with genomic selection; TP GS, 3 Cycles/Year, two-part program with genomic selection
 921
 922
 923
 924

925 **References**

926

927 Alexandratos N, Bruinsma J (2012) World agriculture towards 2030/2050. *Land Use Policy*
928 20:375. [https://doi.org/10.1016/S0264-8377\(03\)00047-4](https://doi.org/10.1016/S0264-8377(03)00047-4)

929 Allier A, Lehermeier C, Charcosset A, et al (2019) Improving Short- and Long-Term Genetic
930 Gain by Accounting for Within-Family Variance in Optimal Cross-Selection. *Front*
931 *Genet* 10:1006. <https://doi.org/10.3389/fgene.2019.01006>

932 Bernardo R (2014) *Essentials of Plant Breeding*. Stemma Press

933 Bernardo R, Yu J (2007) Prospects for Genomewide Selection for Quantitative Traits in
934 Maize. *Crop Science* 47:1082. <https://doi.org/10.2135/cropsci2006.11.0690>

935 Charlesworth B (2009) Fundamental concepts in genetics: Effective population size and
936 patterns of molecular evolution and variation. *Nature Reviews Genetics* 10:195–205.
937 <https://doi.org/10.1038/nrg2526>

938 Chen GK, Marjoram P, Wall JD (2009) Fast and flexible simulation of DNA sequence data.
939 *Genome Res* 19:136–142. <https://doi.org/10.1101/gr.083634.108>

940 Clark S a, Hickey JM, Daetwyler HD, van der Werf JH (2012) The importance of
941 information on relatives for the prediction of genomic breeding values and the
942 implications for the makeup of reference data sets in livestock breeding schemes.
943 *Genetics Selection Evolution* 44:4. <https://doi.org/10.1186/1297-9686-44-4>

944 Duvick DN, Smith JSC, Cooper M (2010) Long-Term Selection in a Commercial Hybrid
945 Maize Breeding Program. In: Janick J (ed) *Plant Breeding Reviews*. John Wiley &
946 Sons, Inc., Oxford, UK, pp 109–151

- 947 Falconer DS, Mackay TFC (1996) Introduction to Quantitative Genetics. Longman, Harlow,
948 UK
- 949 Gaynor RC, Gorjanc G, Bentley AR, et al (2017) A two-part strategy for using genomic
950 selection to develop inbred lines. *Crop Science* 57:2372–2386.
951 <https://doi.org/10.2135/cropsci2016.09.0742>
- 952 Gaynor RC, Gorjanc G, Wilson D, Hickey JM (2019) AlphaSimR. Version 0.11.0URL
953 <http://CRAN.R-project.org/package=AlphaSimR>
- 954 Geiger HH, Gordillo GA (2009) DOUBLED HAPLOIDS IN HYBRID MAIZE BREEDING.
955 *Maydica* 485–499
- 956 Gorjanc G, Gaynor RC, Hickey JM (2018) Optimal cross selection for long-term genetic gain
957 in two-part programs with rapid recurrent genomic selection. *Theor Appl Genet*
958 131:1953–1966. <https://doi.org/10.1007/s00122-018-3125-3>
- 959 Gorjanc G, Jenko J, Hearne SJ, Hickey JM (2016) Initiating maize pre-breeding programs
960 using genomic selection to harness polygenic variation from landrace populations.
961 *BMC Genomics* 17:30. <https://doi.org/10.1186/s12864-015-2345-z>
- 962 Griffing B (1975) Efficiency changes due to use of doubled-haploids in recurrent selection
963 methods. *Theoret Appl Genetics* 46:367–385. <https://doi.org/10.1007/BF00281141>
- 964 Habier D, Fernando RL, Dekkers JCM (2007) The impact of genetic relationship information
965 on genome-assisted breeding values. *Genetics* 177:2389–2397.
966 <https://doi.org/10.1534/genetics.107.081190>
- 967 Hull FH (1945) Recurrent selection and specific combining ability in corn. *J Amer Soc*
968 *Agron*

- 969 Kimura M, Crow JF (1963) On the maximum avoidance of inbreeding. *Genet Res* 4:399–
970 415. <https://doi.org/10.1017/S0016672300003797>
- 971 Kinghorn BP, Hickey JM, Werf JHJVD (2010) Reciprocal Recurrent Genomic Selection for
972 Total Genetic Merit in Crossbred Individuals. *Proc 9th WCGALP* 0036
- 973 Kinghorn BP, Shepherd RK, Woolliams JL (1999). In: *Proceedings of the Association for the*
974 *Advancement of Animal Breeding and Genetics*. p 412
- 975 Longin CFH, Gowda M, Mühleisen J, et al (2013) Hybrid wheat: quantitative genetic
976 parameters and consequences for the design of breeding programs. *Theor Appl Genet*
977 126:2791–2801. <https://doi.org/10.1007/s00122-013-2172-z>
- 978 Lush JL (1943) *Animal Breeding Plans*. Iowa State Press
- 979 Meuwissen THE (1997) Maximizing the response of selection with a predefined rate of
980 inbreeding. *Journal of Animal Science* 75:934–940. <https://doi.org/doi:/1997.754934x>
- 981 Pong-Wong R, A. Woolliams J (1998) Response to mass selection when an identified major
982 gene is segregating. *Genetics Selection Evolution* 30:313–337
- 983 R Development Core Team (2014) *R: A language and environment for statistical computing*.
984 R Foundation for Statistical Computing, Vienna, Austria
- 985 Ramsey FL, Schafer DW (2002) *The Statistical Sleuth: A Course in Methods of Data*
986 *Analysis*, 2nd edn. Duxbury/Thomson Learning, Australia
- 987 Rembe M, Zhao Y, Jiang Y, Reif JC (2019) Reciprocal recurrent genomic selection: an
988 attractive tool to leverage hybrid wheat breeding. *Theor Appl Genet* 132:687–698.
989 <https://doi.org/10.1007/s00122-018-3244-x>

- 990 Sprague GF (1946) THE EXPERIMENTAL BASIS FOR HYBRID MAIZE. Biological
991 Reviews 21:101–120. <https://doi.org/10.1111/j.1469-185X.1946.tb00317.x>
- 992 Sprague GF, Tatum LA (1942) General vs. Specific Combining Ability in Single Crosses of
993 Corn ¹. Agron J 34:923–932.
994 <https://doi.org/10.2134/agronj1942.00021962003400100008x>
- 995 Troyer AF, Wellin EJ (2009) Heterosis decreasing in hybrids: Yield test inbreds. Crop
996 Science 49:1969–1976. <https://doi.org/10.2135/cropsci2009.04.0170>
- 997 Whittaker JC, Thompson R, Denham MC (2000) Marker-assisted selection using ridge
998 regression. Genet Res 75:249–252
- 999 Woolliams JA, Berg P, Dagnachew BS, Meuwissen THE (2015) Genetic contributions and
1000 their optimization. Journal of Animal Breeding and Genetics 132:89–99.
1001 <https://doi.org/10.1111/jbg.12148>
- 1002 Wray NR, Thompson R (1990) Prediction of rates of inbreeding in selected populations.
1003 Genet Res 55:41–54. <https://doi.org/10.1017/S0016672300025180>
- 1004 Wright S (1921) Systems of Mating. II. the Effects of Inbreeding on the Genetic Composition
1005 of a Population. Genetics 6:124–143
- 1006 Yang CJ, Sharma R, Gorjanc G, et al (2019) Origin Specific Genomic Selection: a simple
1007 process to optimize the favourable contribution of parents to progeny. Genetics
1008
1009