bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

1 # A strategy to exploit surrogate sire technology in
2 # livestock breeding programs

3

4 Paolo Gottardo[1], Gregor Gorjanc[1], Mara Battagin[1], R Chris Gaynor[1], Janez Jenko[1],
5 Roger Ros-Freixedes[1], C Bruce A Whitelaw[1], Alan J Mileham[2], William O Herring[3],
6 John M Hickey[1§]

7

8 [1]The Roslin Institute and Royal (Dick) School of Veterinary Studies, The University
9 of Edinburgh, Easter Bush, Midlothian, Scotland, UK

10 [2]Genus PLC, 1525 River Rd., DeForest, Wisconsin 53532, USA

11 [3]The Pig Improvement Company, Genus PLC, 100 Bluegrass Commons Blvd., Ste
12 2200, Hendersonville, Tennessee 37075, USA

13 [§]Corresponding author

14

15 Email addresses:

16 PG: paolo.gottardo@roslin.ed.ac.uk

17 GG: gregor.gorjanc@roslin.ed.ac.uk

18 MB: mara.battagin@roslin.ed.ac.uk

19 RCG: chris.gaynor@roslin.ed.ac.uk

20 JJ: janez.jenko@roslin.ed.ac.uk

21 RRF: roger.ros@roslin.ed.ac.uk

22 CBAW: bruce.whitelaw@roslin.ed.ac.uk

23 AJM: Alan.Mileham@genusplc.com

24 WOH: William.Herring@genusplc.com

25 JMH: john.hickey@roslin.ed.ac.ukJ

26

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

27 # Abstract

28 In this work, we performed simulations to develop and test a strategy for exploiting
29 surrogate sire technology in animal breeding programs. Surrogate sire technology
30 allows the creation of males that lack their own germline cells, but have transplanted
31 spermatogonial stem cells from donor males. With this technology, a single elite male
32 donor could give rise to huge numbers of progeny, potentially as much as all the
33 production animals in a particular time period.

34 One hundred replicates of various scenarios were performed. Scenarios followed a
35 common overall structure but differed in the strategy used to identify elite donors and
36 how these donors were used in the product development part.

37 The results of this study showed that using surrogate sire technology would
38 significantly increase the genetic merit of commercial sires, by as much as 6.5 to 9.2
39 years' worth of genetic gain compared to a conventional breeding program. The
40 simulations suggested that a strategy involving three stages (an initial genomic test
41 followed by two subsequent progeny tests) was the most effective of all the strategies
42 tested.

43 The use of one or a handful of elite donors to generate the production animals would
44 be very different to current practice. While the results demonstrate the great potential
45 of surrogate sire technology there are considerable risks but also other opportunities.
46 Practical implementation of surrogate sire technology would need to account for
47 these.

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

48 # Introduction

49 In this study, we performed simulations to develop a strategy for exploiting surrogate
50 sire technology [1-2] in animal breeding programs (Fig. 1). Surrogate sire technology
51 allows the creation of males that lack their own germline cells, but have transplanted
52 spermatogonial stem cells from other donor males. The concept requires the
53 production of recipient males with an ablated germ line. Rodent males can have their
54 germline ablated using chemotoxic drugs or localised irradiation of the testes, but,
55 importantly for use in livestock breeding, this ablation is incomplete and recipient
56 sperm output is mixture of donor and recipient cells [3]. The mammalian NANOS2
57 gene seems to be absolutely required for the maintenance of germ line cells in males
58 only [4]. In mice, Nanos 2 knock out males the testes completely lack germ-line cells,
59 but there is no effect in females [4]. NANOS 2 knock out pigs have been produced
60 using CRISP/Cas9 gene editing [1] and boars homozygous for the knockout likely
61 provide ideal recipients for the surrogate sire concept."

62 With this technology, a single elite male donor could give rise to huge
63 numbers of progeny, potentially as much as all the production animals in a particular
64 time period. This potential offers many advantages. Firstly, it would reduce the
65 genetic lag between the elite nucleus animals and the production animals. Secondly, it
66 could enable better matching of specific management plans to the genetics. Thirdly, as
67 we outline in the discussion it could enable exploitation of combining ability. The
68 latter could increase production on farm and increase investment and innovation in
69 breeding by enabling a greater ability to protect intellectual property.

70 Typically, animal breeding programs are implicitly or explicitly organized in
71 pyramid structures with layers (Fig. 2). The top layer is the nucleus, which is

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

72 improved using recurrent selection. Nowadays most selection decisions are made
73 using genomic based testing rather than traditional phenotype based testing [4–8]. The
74 middle layer is the multiplication, where the nucleus genetics is multiplied and
75 sometimes crosses between purebred lines are produced. The base layer is the
76 commercial sector, where the majority of animals are kept for production. The
77 commercial producers often make a final cross between the terminal line sires and the
78 maternal line dams.

79 The need to generate huge numbers of production animals and the limited
80 number of progeny that a male can produce means that large numbers of nucleus
81 animals must contribute genetics to the subsequent layers and that one to several
82 generations are required for multiplication. These factors give rise to a genetic lag, a
83 difference in genetic mean between the nucleus and commercial layers. This lag can
84 also be represented with the number of years of genetic gain [9], e.g., ~4 years in a pig
85 breeding program. Surrogate sire technology would allow a single elite nucleus male
86 to give rise to very large numbers of commercial animals, by donating spermatogonial
87 stem cells to its commercial surrogates [1]. This could shorten the lag between the
88 nucleus, multiplication, and commercial layers.

89 Using surrogate sire technology in this way would require that animal
90 breeding programs identify elite donor males and create surrogate sires. This process
91 should take place in a sufficiently small amount of time so that the extra genetic gain
92 would not be significantly reduced by the extra time required for the identification of
93 donors and creation of surrogate sires.

94 A restructured animal breeding program with surrogate sire technology would
95 be conceptually similar to a plant breeding program that produces clonally propagated

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

96 individual lines or inbred lines or hybrid lines (Fig. 2). These programs seek: (i) to
97 identify the best individual (note: here we take individual to mean clonal, inbred or
98 hybrid lines), or a handful of individuals, from a population of individuals; and (ii) to
99 disseminate this individual very widely in the commercial layer [10]. To identify the
100 best individual, plant breeders typically use multiple stage testing and selection. As
101 the breeding program progresses through these stages the number of individuals being
102 tested is reduced and the precision of these tests increases. The small number of
103 individuals in the final stages are intensively tested in large replicated experiments
104 that are repeated across several environments and years. This ensures that the
105 commercially released individual is well characterized and carries a minimal risk of
106 major undetected weakness. This is necessary because this individual will have a huge
107 footprint in the commercial layer. Similar levels of evaluation would be needed with
108 surrogate sire technology in animal breeding programs.

109 The objective of this study was to develop a strategy for exploiting surrogate
110 sire technology in animal breeding programs. This strategy involved a subtle, but
111 important, reorganisation to combine components of traditional animal and plant
112 breeding programs. The reorganization is similar to the two-part breeding program
113 that we recently proposed for the incorporation of genomic selection into plant
114 breeding programs [11]. The reorganization involves an explicit partitioning of a
115 breeding program into a population improvement component and a product
116 development component. The population improvement component is similar to the
117 currently used recurrent genomic selection in many animal breeding nucleus
118 populations. The product development component is similar to traditional plant
119 breeding programs and involves a number of stages of testing to identify the elite
120 donors. The product development component could make use of testing for combining

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

121 ability, if that was appropriate for the particular species of interest.

122 With a focus on application in pig breeding, several alternative versions of the
123 reorganized breeding program were compared to different variants of a conventional
124 breeding program using simulation. The alternative versions varied: (i) the number of
125 stages of testing; (ii) the number of donor candidates tested at subsequent stages; (iii)
126 the accuracy of the genomic test at the first stage; and (iv) the accuracy of progeny
127 test in later stages. The results showed that using surrogate sire technology would
128 significantly increase the genetic merit of commercial sires, by as much as between
129 6.5 and 9.20 years' worth of genetic gain compared to different variants of a
130 conventional breeding program. The simulations suggested that an identification
131 strategy involving three stages (a genomic test followed by two subsequent progeny
132 tests) was the most effective of all the strategies tested. The use of one or a handful of
133 elite donors to generate the production animals would be very different to current
134 practice. While the results demonstrate the great potential of surrogate sire technology
135 there are considerable risks and these are discussed.

136

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

137 # Methods

138 Simulation was used to evaluate the impact of a strategy for exploiting
139 surrogate sire technology in animal breeding programs. One hundred replicates of
140 various scenarios were performed. Scenarios followed a common overall structure but
141 differed in the strategy used to identify elite donors and how these donors were used
142 (Fig. 3, 4).

143 Conceptually, the simulation scheme was divided into historical and future
144 phases. The historical phase represented historical evolution and recent animal
145 breeding efforts up to the present day, under the assumption that animal populations
146 have evolved for tens of thousands of years, followed by 22 recent generations of
147 modern animal breeding with selection on genomic breeding values in a nucleus
148 population. The future phase represented 20 future generations of modern animal
149 breeding, with selection on genomic breeding values in a nucleus population that
150 subsequently supplied genetic improvement to multiplication and commercial layers.
151 The historical animal breeding generations were denoted -21 to 0 and the future
152 animal breeding generations were denoted 1 to 20. The multiplier and commercial
153 layers were not explicitly simulated but were instead represented with the average
154 genetic merit of nucleus males that would give rise to multiplication and commercial
155 animals while accounting for the time lag. Specifically, we only focused on a breeding
156 program that produced terminal males in a scheme that closely resembled a pig
157 breeding program.

158 Simulations involved the following four steps:

159 (i) Generating genome,

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

160 (ii) Generating a quantitative trait and breeding values,

161 (iii) Generating an animal breeding program,

162 (iv) Selection and dissemination to the commercial layer with the conventional or
163 surrogate sires strategy.

164 Results are presented as the mean of one hundred replicates for each scenario
165 and encompass the genetic merit of nucleus males that would give rise to commercial
166 animals at a given time point.

167 **Genome**

168 Whole-genome sequences were generated using the Markovian Coalescent
169 Simulator (MaCS) [12] and AlphaSim [13] for 400 base haplotypes for each of ten 10
170 chromosomes. Chromosomes (each 100 cM long and comprising 10^8 base pairs) were
171 simulated using a per site mutation rate of 2.5×10^{-8}, a per site recombination rate of
172 1.0×10^{-8}, and an effective population size (N_e) that varied over time in accordance
173 with estimates that are representative of livestock populations [e.g., 14–17] as
174 follows: N_e was set to 100 in the final generation of the coalescent simulation, to N_e =
175 1256, 1000 years ago, to N_e=4350, 10,000 years ago, and to N_e=43,500, 100,000 years
176 ago, with linear changes in between these time-points. The resulting sequences had
177 approximately 540,000 segregating sites.

178 **Quantitative trait**

179 A quantitative trait was simulated by randomly sampling 10,000 causal loci
180 from the genome in the base population, with the restriction that 1,000 were sampled
181 from each of the 10 chromosomes. For these loci, the allele substitution effect was
182 randomly sampled from a normal distribution with a mean of 0 and standard deviation
183 of 0.01 (1.0 divided by the square root of the number of loci).

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

184 **Breeding values**

185 True breeding values were computed as a sum of effects at causal loci. To
186 simulate selection without the full computational burden and complexity of simulating
187 training sets and estimation with best linear unbiased prediction, we simulated pseudo
188 estimates of breeding values by adding a level of noise to true breeding values.
189 Different levels of noise were added to achieve a targeted accuracy. For the genomic
190 tests we simulated accuracies of 0.50, 0.70 and 0.90. For the progeny tests we
191 simulated accuracies as a function of the number of progeny [24] used in the different
192 scenarios (described below).

193 **Breeding program**

194 A pedigree of 42 generations for the nucleus population was simulated. Each
195 generation included 1,000 (**SmallScenario**) or 5,000 (**BigScenario**) individuals with
196 equal sex ratio. The different numbers of individuals were used to quantify impact of
197 nucleus population size on the benefit of surrogate sire technology. All females (500
198 for the SmallScenario or 2,500 for the BigScenario) and 50 males were selected as the
199 parents of each generation. This selection was based on a genomic test. In the first
200 generation of the recent historical animal breeding population (i.e., generation -22),
201 the chromosomes of each individual were sampled from the 400 base haplotypes. In
202 later generations (i.e., generations -21 to 20), the chromosomes of each individual
203 were sampled from parental chromosomes with recombination (assuming no
204 interference). A recombination rate of 1 Morgan per chromosome was used, resulting
205 in a 10 Morgan genome.

206 **Scenarios**

207 Two different strategies were used to identify males from the nucleus who

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

208 would give rise to commercial animals, either through conventional multiplication or
209 surrogate sires. The conventional multiplication strategy used the top 50, 200, or 500
210 males in both the SmallScenario and the BigScenario. Males were chosen based on a
211 genomic test. The surrogate sires strategy used multi-stage testing. Males were chosen
212 based on an initial genomic test (S0), followed by one or two subsequent progeny
213 tests (S1 and S2). As is the case with plant breeding programs, as the testing
214 progressed through the stages we reduced the number of tested individuals and
215 increased accuracy of tests. Based on the tests the best individual or set of individuals
216 were identified and used as elite donors of spermatogonial stem cells to surrogate
217 sires.

218 To quantify the impact of different amounts of testing resources and different
219 allocation of these resources we simulated different accuracies of the genomic test at
220 S0, different numbers of donor candidates tested with different number of progeny at
221 S1 and S2. At S0 we simulated a genomic test with an accuracy of 0.50, 0.70, and
222 0.90. To ensure that each breeding program had the same costs, we assumed that a
223 total of 14,000 progeny were available for progeny testing stages.

224 With single progeny test (S1) we used the 14,000 progeny to test 14 donor
225 candidates each with 1,000 progeny, 28 donor candidates each with 500 progeny,
226 56 donor candidates each with 250 progeny, 112 donor candidates each with
227 125 progeny, 224 donor candidates each with 63 progeny, or 448 donor candidates
228 each with 31 progeny.

229 With two progeny tests (S1 and S2) we used either 2,000, 4,000, or 6,000
230 progeny for the first test (S1) and the remaining 12,000, 10,000, or 8,000 for the
231 second test (S2). At S1 either 100, 200, or 400 donor candidates were tested. Thus,

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

232 when 2,000 progeny were used at S1 the 100, 200, or 400 donor candidates were each
233 tested with 20, 10, or 5 progeny respectively. When 4,000 progeny were used at S1
234 the 100, 200, or 400 donor candidates were each tested with 40, 20, or 10 progeny
235 respectively. When 6,000 progeny were used at S1 the 100, 200, or 400 donor
236 candidates were each tested with 60, 30, or 15 progeny respectively. At S2 we tested
237 either 10 or 20 donor candidates advanced from S1. When 12,000 progeny remained
238 to be used at S2 the 10 or 20 donor candidates were each tested with 1,200 or 600
239 progeny respectively. When 10,000 progeny remained to be used at S2 the 10 or 20
240 donor candidates were each tested with 1,000 or 500 progeny respectively. When
241 8,000 progeny remained to be used at S2 the 10 or 20 donor candidates were each
242 tested with 800 or 400 progeny respectively. From each of these testing strategies we
243 chose either 1 or 5 donors of spermatogonial stem cells for surrogate sires in the
244 commercial layer.

245 All of these different factors (two sizes of a breeding program [Small, Big],
246 three conventional strategy scenarios [50, 200, 500 males], six surrogate sires strategy
247 scenarios with two-stage testing, 18 surrogate sires strategy scenarios with three-stage
248 testing, and using one or five donors) gave 102 different scenarios for each level of
249 genomic test accuracy. The map of all these scenarios and used resources is
250 summarized in Fig. 4.

251 **Time assumptions**

252 The time taken to transfer germplasm from the nucleus to the commercial
253 layer was assumed to be 3.5 years for the conventional strategy (but see the note
254 below about "dilution"), 3.5 years for the surrogate sires strategy with two-stage
255 testing, and 4.5 years for the surrogate sires strategy with three-stage testing. The

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

256 different steps that underlie these time frames are presented in Fig. 3. We based our
257 parameters on pigs and assumed 6 months for a male to reach sexual maturity, 4
258 months for a successful gestation, and 8 months to collect terminal line phenotypes on
259 progeny. Based on these parameters we assumed 12 months to progeny test a sexually
260 mature male. When the donors are identified we assumed that it takes a further 12
261 months to produce surrogate sires from these. Finally, we assumed a 12 months for
262 the commercial progeny to pass through gestation and complete their growth. We
263 assumed that the conventional program involved two rounds of multiplication that
264 each take 12 months to complete.

265 Although we assumed that the genetic improvement with the conventional
266 strategy is delivered to the commercial population in 3.5 years, we assumed an
267 additional component of genetic lag, because the genetic merit of the sires entering
268 the multiplier layer is "diluted" by the lagged genetic merit of females in the
269 multiplier layer (i.e., we assumed no selection of females in the multiplier). Such a
270 dilution would not occur with the surrogate sires strategy, because the multiplication
271 layer does not arise. To account for this extra genetic lag in the conventional strategy
272 we "diluted" genetic merit of commercial sires as follows:

$$\bar{a}_t^d = 0.5\bar{a}_t + 0.5\sum_{i=1}^{6}\bar{a}_{t-i}w_i$$

273 where $\bar{a}_t$ is the average genetic merit of used nucleus males in generation t and w_i is
274 the relationship coefficient between the commercial sire and his maternal male
275 ancestor in the generation i. We only accounted for 6 generations with w_i ranging
276 from 0.5 in $t-1$ generation to 0.015625 in the $t-6$ generation. This "dilution"
277 increased genetic lag of the conventional strategy by an equivalent of ~1.04 years'

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

278 worth of extra genetic gain.

279 **Comparison of different scenarios**

280 To ensure that sufficient numbers of generations had been traversed for
281 "dilution", we chose to present the results in terms of the genetic merit of terminal
282 sires used in the commercial layer emerging from generation 11 and each subsequent
283 generation. We report genetic merit in units of the standard deviation of true breeding
284 values of the nucleus animals in the base generation (σ_b), i.e., as $(\bar{a}_t - \bar{a}_b)/\sigma_b$, where
285 $\bar{a}_t$ is the average true breeding value of the nucleus males that gave rise to
286 commercial sires in year t and $\bar{a}_b$ is the average true breeding value of nucleus
287 animals in the base generation. Calculating the genetic merit of commercial sires in
288 this way allowed the different strategies to be compared in terms of genetic merit of
289 the commercial sires at the same year. Finally, we have converted the standardized
290 genetic merit into years' worth of genetic gain by calculating the number of years it
291 takes the conventional breeding program when selecting the top 50 males to deliver
292 the same level of genetic merit to the commercial layer.

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

293 # Results

294 The surrogate sires strategy increased the genetic merit of terminal sires used in the
295 commercial layer. The genetic merit of commercial surrogate sires from the surrogate
296 sires strategy was as much as 6.5 to 9.2 years' worth of genetic gain higher than the
297 genetic merit of commercial sires from the conventional multiplication strategy. In
298 both the SmallScenario and BigScenario the three-stage testing strategy was the best
299 strategy for identifying elite donors. The best performing three-stage testing strategy
300 involved a genomic test at the first stage, 100 donor candidates tested each with 60
301 progeny at the second stage, and 20 donor candidates tested each with 400 progeny at
302 the third stage (see Table 1 for details). The benefit of surrogate sires strategy was
303 greatest when the genomic test accuracy was lowest and when the conventional
304 strategy required large proportions of the nucleus males to be used for multiplication.

305 In what follows the results are divided into three sub-sections for ease of
306 presentation: (i) comparison of the conventional strategy and the best performing
307 surrogate sires strategies; (ii) comparison of two-stage testing scenarios of the
308 surrogate sires strategy; and (iii) comparison of three-stage testing scenarios of the
309 surrogate sires strategy. To avoid clutter in the figures or tables we do not show
310 standard errors across the 100 replicates of the simulated scenarios because the
311 standard errors were small in all instances less than 0.009 YGG.

312 **Comparison of the conventional and the best performing surrogate sires**
313 **strategies**

314 Fig. 5 and S1 show the average genetic merit of commercial sires derived from
315 the best performing surrogate sires strategy scenario and the conventional strategy
316 against time, for three different genomic test accuracies (0.5, 0.7, and 0.9) and the

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

317 SmallScenario and the BigScenario. The conventional strategy used the top 50, 200,
318 or 500 males in multiplication. At all points in time and for all three genomic
319 accuracies commercial sires derived from the best performing surrogate sires strategy
320 scenario had a higher genetic merit than those derived from the conventional strategy.
321 This benefit was greater when more males were used for multiplication in the
322 conventional strategy. The benefit of using surrogate sires strategy decreased as the
323 genomic test accuracy increased. Across time the difference between the two
324 strategies was almost constant. These trends were common both in the SmallScenario
325 and the BigScenario, although with differences in magnitude.

326 Table 1 enumerates some of the main results than can be observed in Fig. 4
327 and S1. Across all scenarios tested the best performing surrogate sires strategy
328 scenario involved first a genomic test of all donor candidates followed by two
329 subsequent progeny tests and the use of a single elite donor. The benefit of surrogate
330 sires strategy above the conventional strategy was greater when more males were used
331 for multiplication with the conventional strategy. When the genomic test accuracy
332 was low (0.5) the best strategy was to first progeny test 100 candidates on 6,000
333 progeny and then to test 20 candidates on 8,000 progeny. This testing and subsequent
334 production of surrogate sires was assumed to take one additional year compared to the
335 conventional strategy. After accounting for this extra time and for the dilution in the
336 conventional multiplication process, we observed that in the SmallScenario the
337 surrogate sires strategy delivered on average between 6.5 and 9.2 years' worth of
338 extra genetic gain in commercial sires compared to the conventional strategy that uses
339 respectively between 50 and 500 males in multiplication. For the BigScenario the
340 equivalent values were between 2.7 and 4.1 years' worth of extra genetic gain.

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

341 When the genomic test accuracy was higher (> 0.5) the optimal allocation of
342 testing resources was slightly different. Instead of first progeny testing 100
343 candidates, as was the case when the genomic test accuracy was low, progeny testing
344 200 candidates was the best performing scenario. All other scenario parameters were
345 the same as when the genomic test accuracy was low. The benefit of surrogate sires
346 strategy decreased with the increasing genomic test accuracy and the magnitude of
347 benefit differed significantly between the SmallScenario and the BigScenario
348 (Table 1).

349 On average the surrogate sires strategy in SmallScenario delivered between
350 6.5 and 9.2 years' worth of extra genetic gain in commercial sires when the genomic
351 test accuracy was 0.5. When the genomic test accuracy was 0.7 these values reduced
352 to between 4.5 and 7.2 years and when the genomic test accuracy was 0.9 they further
353 reduced to between 2.4 and 5.0 years.

354 On average the surrogate sires strategy in BigScenario delivered between 2.7
355 and 4.1 years' worth of extra genetic gain in commercial sires when the genomic test
356 accuracy was 0.5. When the genomic test accuracy was 0.7 these values reduced to
357 between 2.1 and 3.5 years and when the genomic test accuracy was 0.9 they further
358 reduced to between 1.20 and 2.50 years.

359 The differences in the SmallScenario and the BigScenario were due to the
360 different proportions of males used in multiplication to give rise to commercial sires.
361 In the SmallScenario 10% to 100% of males were used while the in the BigScenario
362 2% to 20% of males were used.

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

363 For simplicity of presentation and based on the consistency of trends described
364 above, in the following sections we only present comparisons to the conventional
365 strategy in which 50 males were used in multiplication.

366 **Comparison of two-stage testing scenarios of the surrogate sires strategy**

367 Tables 2 and 3 show the performance of different two-stage testing scenarios
368 of the surrogate sires strategy. Performance is measured as the average years' worth
369 of extra genetic gain in the commercial sires delivered by the surrogate sires strategy
370 compared to the conventional strategy for both the SmallScenario (Table 2) and the
371 BigScenario (Table 3). Consistent with the results reported in the previous sub-section
372 the benefit of surrogate sires strategy was always lower when the genomic test
373 accuracy was higher. In some scenarios, the benefit was minimal. In all cases, there
374 was an intermediate optimum for the numbers of candidates tested. Using five elite
375 donors was always worse than using one. This behaviour was observed in both the
376 SmallScenario and the BigScenario although with some interesting differences. The
377 BigScenario showed a general shrinkage of years' worth of genetic gain compared to
378 the SmallScenario, resulting in a general increase in the number of scenarios that
379 showed a small benefit of the surrogate sires strategy.

380 At all levels of genomic test accuracy the best scenario was to screen
381 candidates based on genomic test, progeny test 112 candidates each with 125 progeny,
382 and use the best candidate as a single elite donor. With the genomic test accuracy of
383 0.5, 0.7, and 0.9 this scenario gave respectively 5.3, 3.6, or 2.2 years' worth of extra
384 genetic gain in commercial sires in the SmallScenario (Table 2) and respectively 2.5,
385 2.0 or 1.1 years' worth of extra genetic gain in commercial sires in the BigScenario
386 (Table 3).

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

387 Just as for the case of selecting one elite donor of spermatogonial cells for
388 surrogate sires, when selecting five elite donors, progeny testing 112 candidates each
389 with 125 progeny gave the highest benefit. With the genomic test accuracy of 0.5, 0.7,
390 and 0.9 this scenario gave respectively 3.6, 2.6 and 1.2 years' worth of extra genetic
391 gain in the SmallScenario (Table 2) and respectively 2.0, 1.1 and 1.0 in the
392 BigScenario (Table 2).

393 **Comparison of three-stage testing scenarios of the surrogate sires strategy**

394 Tables 4 and 5 shows the performance of different three-stage testing
395 scenarios of the surrogate sires strategy in the SmallScenario when either one or five
396 elite donors used. By varying several parameters, we tested 216 (108 for the
397 SmallScenario and 108 for the BigScenario) different scenarios of three-stage testing
398 with fixed total progeny testing resources. These resources were the same as for the
399 two-stage testing scenarios described in the previous sub-section. The parameters with
400 the three-stage testing scenarios were the genomic test accuracy for the first stage, the
401 split of resources between the two subsequent progeny tests, the number of tested
402 donor candidates, the number of progeny per tested donor candidate at each progeny
403 test stage, and the number of elite donors used for production of surrogate sires.

404 The three-stage testing gave a greater benefit than the two-stage testing. As for
405 the two-stage testing, using one elite donor for surrogate sires gave a greater benefit
406 than using five elite donors and the benefit of surrogate sires strategy was greater
407 when the genomic test accuracy was lower. A total of 14,000 progeny were split
408 across the two stages of progeny testing. Increasing the resources in the first progeny
409 test increased benefit of surrogate sires strategy. For example, with the SmallScenario
410 when the genomic test accuracy was 0.5, 6,000 progeny were used in the first progeny

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

411 test, 8,000 were used in the second progeny test, and when one elite donor was used in
412 the end, the benefit was 6.5 years' worth of extra genetic gain in commercial sires
413 above the conventional strategy that uses 50 nucleus males in multiplication. This was
414 a greater benefit than the 5.8 years' worth of extra genetic gain for the scenario that
415 split the 14,000 progeny into 4,000 for the first progeny test and 10,000 for the second
416 progeny test, which was in turn better than the 5.4 years' worth of extra genetic gain
417 for the scenario that split the 14,000 progeny into 2,000 for the first progeny test and
418 12,000 for the second progeny test. This trend of greater benefit when more progeny
419 were dedicated to the first progeny test was observed for almost all tested scenarios.

420 For the SmallScenario the difference between testing 100 or 200 donor
421 candidates at the first progeny test was not consistent. That said, when the genomic
422 test accuracy was 0.5, allocating 100 candidates to the first progeny test was usually
423 better than allocating 200, and allocating 200 candidates was usually better than
424 allocating 400. At higher genomic test accuracies, there were little differences
425 between allocating 100 or 200 candidates to the first progeny test, but both of these
426 sets of scenarios were usually better than allocating 400 candidates to the first
427 progeny test.

428 In the SmallScenario allocating 20 elite donor candidates to the second
429 progeny test was almost always better than allocating 10 candidates. A total of 54
430 scenarios were evaluated for SmallScenario. In only 6 of these scenarios allocating 10
431 candidates was better than allocating 20.

432 Overall for the SmallScenario, when the genomic test accuracy was 0.5, the
433 best three-stage testing scenario used 6,000 progenies in the first progeny test of 100
434 candidates each with 60 progeny, 8,000 progenies in the second progeny test of 20

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

435 candidates each with 400 progeny, and used a single elite donor for surrogate sires.
436 This scenario gave a benefit of 6.50 years' worth of extra genetic gain in commercial
437 sires compared to the conventional strategy. The same distribution of testing resources
438 was also the joint best when five, instead of one, elite donors were used for surrogate
439 sires.

440 The same trends as for the SmallScenario were observed also for the
441 BigScenario, but with smaller benefit of the surrogate sire strategy (See table S1 and
442 S2).

443

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

444 # Discussion

445 The results of this paper suggest that a surrogate sires strategy could be very
446 beneficial for the dissemination of genetic gain in animal breeding. In summary, our
447 results indicate that benefits of the as much as 6.5 to 9.2 years' worth of genetic gain
448 in commercial sires could be realized with surrogate sires compared to the
449 conventional multiplication. It was best to identify elite donors for surrogate sires via
450 a three-stage testing strategy involving a first screen with a genomic test followed by
451 two subsequent progeny tests. The benefits of a surrogate sires strategy were greater
452 when genomic test accuracy was low and when the conventional strategy used a large
453 proportion of males in multiplication. To discuss these results we divide the
454 discussion into four sections: (i) possible explanations for the observed trends; (ii)
455 justification and impact of assumptions; (iii) the potential impact of surrogate sires on
456 the redesign of animal breeding programs; and (iv) risks and opportunities of using
457 surrogate sires.

458 **Possible explanations for the observed trends**

459 That surrogate sire technology generates such a benefit in terms of years'
460 worth of genetic gain can be explained in the context of the breeders' equation. While
461 the surrogate sires strategy does not rely on the selection of the best individuals and
462 using them as parents of the next generation, it does rely on the identification of the
463 best individuals from a cohort and using them as donors of spermatogonial cells for
464 surrogate sires, which is another form of the selection problem. In any cohort, the best
465 few individuals will be some number of standard deviations above the cohort average.
466 For example, when surrogate sires technology delivered 6.5 years' worth of additional
467 genetic gain in commercial sires the best nucleus male was on average 2.7 standard

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

468 deviations above the cohort mean. In contrast, the best 50 nucleus males were 2.0
469 standard deviations above the cohort mean. Given that the breeding program
470 proceeded at a rate of genetic progress of 0.4 standard deviations per year, choosing
471 the best male as a donor for surrogate sires rather than the best 50 males produced
472 surrogate sires that were better for more than 8 years' worth of genetic gain. However,
473 accounting for the imperfect accuracy of identifying donors with the surrogate
474 strategy or the best 50 males for multiplication with the conventional strategy and the
475 time to generate commercial sires with either strategy the final result was 6.5 years'
476 worth of genetic gain.

477 With constant progeny test accuracies the benefit of the surrogate sires
478 strategy depended on the proportion of male candidates that the conventional strategy
479 used to give rise to commercial sires. If the breeding program needed to use a large
480 proportion of its nucleus male candidates (e.g., the best 200 or 500) the benefit of
481 surrogate sires strategy was greater than if it needed to use a few. Again, this result is
482 entirely consistent with the breeders' equation. Specifically, it can be explained in the
483 context of selection intensity being a nonlinear function of the percentage of selected
484 individuals, i.e., selection intensity increases almost linearly down to 20 or 10%
485 selected, but increases sharply (nonlinearly) thereafter. While both conventional and
486 surrogate sires strategies exploit the tail of distribution with high selection intensities,
487 the surrogate sires strategy also exploits the steeper part. This explains why the
488 benefit of surrogate sires was higher in the SmallScenario than in the BigScenario. In
489 the SmallScenario we had 500 candidates and selected 100 with the conventional
490 strategy (percentage selected 20% and selection intensity 1.4) or 1 with the surrogate
491 sires strategy (percentage selected 0.2% and selection intensity 3.2). In the
492 BigScenario we had 2,500 candidates and selected 100 with the conventional strategy

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

493 (percentage selected 4% and selection intensity 2.2) or 1 with the surrogate sires
494 strategy (percentage selected 0.04% and selection intensity 3.6). The same logic also
495 explains why selecting five as opposed to one donor for surrogate sires gave a lower
496 benefit.

497 The observed differences in the performance of different surrogate sire
498 strategies can also be explained in the context of the breeders' equation. When the
499 genomic test accuracy used in the first stage of testing was lower the benefit of
500 surrogate sires strategy was higher. Under the conventional strategy, the average
501 genetic merit of the nucleus males that gave rise to commercial sires was lower when
502 the genomic test accuracy was lower than when it was higher. With surrogate sires
503 strategy this reduction in genetic merit due to the low genomic test accuracy is
504 compensated by the subsequent progeny tests. This is in line with the analysis of
505 Dickerson and Hazel [18], who compared the use of progeny test as a supplement to
506 earlier culling on own or sibling performance. Their conclusion was that progeny
507 testing is warranted when heritability is low in which case accuracy of estimated
508 breeding values from own or sibling phenotypes (or genomic prediction in our study)
509 is low. Genomic selection can be thought of as a light touch first screen, the purpose
510 of which is to identify the top group of animals, which are then tested on many
511 progeny. The purpose of subsequent progeny tests is then a search for the best
512 individual within this group.

513 This same logic also explains why the three-stage testing was better than the
514 two-stage testing. Both schemes started with a genomic test that was followed by one
515 progeny test with the two-stage testing or two subsequent progeny tests with the three-
516 stage testing. With the three-stage testing the first progeny test serves to use a portion

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

517 of resources to evaluate many candidates relatively accurately in order to discard most
518 candidates. Then the second progeny test uses the remaining resources to even more
519 accurately identify the final candidate. In terms of the breeders' equation the first
520 progeny test maximizes selection intensity, while the second maximizes accuracy.
521 The three-stage testing appears to address both of these parameters more optimally
522 than the two-stage testing.

523 There is a substantial body of literature on multi-stage selection [19–23] which
524 the observed trends in this study are consistent with. It is well known that increasing
525 the number of progeny per candidate increases accuracy [24,25] and that the number
526 of candidates to be tested is important and the trade-off between the two must be
527 found. In our simulations, we found the optimum at progeny testing 112 candidates,
528 given a fixed amount of resources, in our case 14,000 progeny. This optimum was
529 consistent across the different levels of genomic test accuracy. However, the level of
530 genomic test accuracy heavily influenced the amount of extra genetic gain, because
531 higher accuracy directly translates to higher genetic gain. These trends are consistent
532 with the long-established multi-stage testing in plant breeding [9]. Most plant
533 breeding programs use multi-stage testing to identify elite single genotype (e.g.,
534 inbred line) that is then deliver to the commercial layer. Typically, these programs
535 initially screen many individuals imprecisely at the first stage. At each subsequent
536 stage they reduce the number of tested individuals, but the testing precision is
537 increased.

538 **Justification and impact of assumptions**

539 There is a huge range of possible strategies for the identification of donors for
540 surrogate sires and we only evaluated a small subset in this study. We choose the

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

541 tested range of scenarios because we believe they could demonstrate the properties of
542 surrogate sires strategy. They show that in some circumstances surrogate sires can
543 deliver a large benefit and in others small benefit. We chose the three levels of
544 genomic test accuracy as these levels reflect what might be possible in breeding
545 programs of various sizes. To ensure that all strategies used an equal set of resources
546 we set the total number of progeny involved in progeny testing to 14,000. We chose
547 this number as it was divisible in many ways and thereby enabled several strategies to
548 be compared and because this number was similar the 10,000 progeny that would be
549 used by an animal breeding program that each year tested 100 candidates each with
550 100 progeny, a scale of progeny testing that was not uncommon in some animal
551 breeding programs before the advent of genomic selection.

552 With the two-stage testing the total testing resources were distributed across
553 many or few candidates. As expected, testing an intermediate to high number of
554 candidates (i.e., 112 to 224) on a relatively small number of progeny (i.e., 125 to 63)
555 gave higher benefits than testing a few candidates (e.g., 14) on many progeny or a
556 very high number of candidates (448) on few progeny (31). These trends fit the
557 expectations from the breeders' equation and occur due to the interplay between
558 selection intensity and accuracy. However, when the chosen elite donors of
559 spermatogonial cells for surrogate sires are to be used to produce huge numbers of
560 progeny in the commercial layer, the risk of a donor carrying some major defect that
561 was not identified by the testing process must also be minimized. For this reason, it is
562 unlikely that a strategy in which donors are tested with a single stage of progeny
563 testing using a ~200 or less progeny would ever be used by a commercial breeding
564 program.

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

565 It was this logic that motivated us in our design of the three-stage testing
566 scheme. Our intuition was that the first progeny test would evaluate many candidates
567 with relatively low accuracy, while the second progeny test would evaluate a handful
568 of individuals with high accuracy, i.e., 10 or 20 candidates each with respectively 800
569 or 400 progeny. Using many progeny ensures high accuracy, but also a high degree of
570 certainty that the final donor(s) would not carry any major defects.

571 A major assumption of this study was the amount of time it took to identify
572 elite donors and then to make surrogate sires. It is likely that the different time
573 assumptions could be shortened or lengthened for both the conventional
574 multiplication strategy and the surrogate sires strategy in several ways and depending
575 on the assumed species. The benefit of surrogate sires strategy would change
576 accordingly.

577 Finally, we choose to model a pig breeding program in this study because this
578 is the species that we are most familiar with. The benefits may be greater or smaller
579 for other species. The benefits depend on the ratio of existing reproductive rates of
580 males versus that enabled by surrogate sire technology, the time and cost associated
581 with performing progeny tests, the levels of accuracy that can be obtained by genomic
582 prediction and the relative cost and technical possibilities of surrogate sire technology
583 itself in a particular species. Incidentally, in this study we did not account for the cost
584 aspects of surrogate sire technology itself. Undoubtedly developing the technology
585 itself will be hugely expensive and these costs of development may impact its
586 eventual commercial cost. That said, in time many biotechnologies which are initially
587 expensive become much cheaper (e.g., nowadays genotyping and animal cloning are
588 both relatively inexpensive compared to their former costs) and we anticipate that

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

589 surrogate sire technology will follow a similar pattern. However, given we have
590 ignored the cost component of surrogate sire technology its benefit may be
591 overestimated based on our results compared to a study which would account for such
592 costs.

593 **The potential impact of surrogate sires on the redesign of animal breeding**
594 **programs**

595 Animal breeding programs maximize the genetic merit of commercial animals
596 within the available financial, physical, technical, and physiological constraints.
597 Implicitly a breeding program has two objectives: (i) improving the mean of the
598 population; and (ii) delivering a product to the commercial producers. In dairy cattle
599 for example, before the advent of genomic selection, breeders used progeny testing
600 schemes that intensively evaluated relatively small numbers of candidate males and
601 used the best of these as parents to improve the population, but also as a commercial
602 and breeding product to be used by the commercial layer. In doing so, dairy cattle
603 breeders maximised selection accuracy, but were constrained in their ability to
604 increase selection intensity and decrease generation interval. However, commercial
605 producers used well tested sires and therefore an individual producer could rely on
606 using relatively few sires, who together could serve entire geographic regions. The
607 advent of genomic selection changed this paradigm. Under genomic selection progeny
608 testing of a small number of candidates has been replaced with a genomic testing of a
609 large number of candidates. Those with best predictions are used as parents to
610 improve the "open" nucleus population, but are also sold to commercial layer as a
611 team of sires product (i.e., a group of sires sold together rather than a single sire sold
612 on its own). In doing so, dairy cattle breeders increased selection intensity and

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

613 reduced generation interval, but are constrained in their ability to achieve very high
614 accuracy. Given that each candidate male has not had their merit assessed based on
615 phenotypes of their progeny, there is a risk that certain sires are not that good or may
616 carry mutations that are highly detrimental (e.g., a *de-novo* mutation that prevents
617 progeny from lactating) [26–28]. To overcome this risk, breeders recommend that
618 commercial producers use semen of a larger number (i.e., a team) of sires and limit
619 their use of any one sire.

620 A surrogate sires strategy would need to exploit aspects of both genomic and
621 progeny testing. Genomic testing can be used to drive the population improvement
622 and, as demonstrated in the present paper, to identify a set of candidates that could
623 enter a progeny testing scheme as part of the product development. The role of the
624 progeny testing is to ensure that the chosen elite donors that give rise to surrogate
625 sires released to the commercial sector are good animals, that they are not
626 significantly worse than it is predicted by a genomic test and that they do not carry
627 detrimental mutations. As demonstrated by the results of the present study two
628 subsequent progeny tests used resources more efficiently than a single progeny test.
629 Such multi-stage testing has a long history of use in plant breeding which also has a
630 long history delivering products to commercial producers in a way that is highly
631 analogous to what surrogate sires would enable for animal producers.

632 The majority of commercial producers for all of the major crops (maize,
633 wheat, rice) use inbred lines or their hybrids. These inbred or hybrid lines can be
634 grown on huge areas. Plant and animal breeding designs have diverged somewhat
635 over the years owing to differences in biology, economics, and technical possibilities.
636 Surrogate sire technology, combined with genomic selection, could result in a

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

637 coalescence of designs across plant and animal breeding. One such design that could
638 apply to both is the two-part scheme recently proposed by Gaynor et al. (2017) [11].
639 In this scheme, rapid recurrent selection based on genomic testing is used to increase
640 the mean of the population, while multi-stage testing (genomic and phenotypic) is
641 used to periodically extract, test, and develop a product from the population. The
642 population improvement component resembles the nucleus of animal breeding
643 programs, while the product development component resembles the multi-stage
644 testing to derive inbred or hybrid lines of plant breeding programs. The latter could
645 also be seen as an improved multiplication layer of animal breeding programs that
646 exploit breed complementarity to deliver a commercial product.

647 In the present work, we focused on the use of surrogate sires to produce
648 commercial animals (e.g., a terminal sire in a pig population). To do this, donors for
649 surrogate sires were chosen based on their general combining ability. The strategy
650 could also be extended to exploit specific combining ability to produce a relatively
651 homogenous set of females from a maternal line that are crossed with single terminal
652 male (via surrogate sires). Use of specific combining ability is widespread in hybrid
653 crops where it exploits complementarity of pairs of individuals and heterosis
654 generated by specific pairs of individuals. The surrogate sires strategy proposed in the
655 present paper could be extended to exploit specific combining ability by adding
656 additional stages that progeny tests specific crosses as is conducted in hybrid plant
657 breeding programs. Because in livestock the parents are outbred (compared to crops
658 where they are often inbred), a tiered strategy may be needed in the maternal line(s)
659 that homogenizes dam haplotypes. For example, using a single surrogate sire,
660 grandsire, and great-grandsire on the maternal population would give a pool of
661 females that carried one of two haplotypes for 87.5% (0.5 + 0.25 + 0.125) of their

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

662 genome. The terminal surrogate sire would be chosen based on a specific combining
663 ability to these haplotypes.

664 **Risks and opportunities of using surrogate sires**

665 Surrogate sires present risks and opportunities to commercial production. The
666 most obvious risk relates to the genetic homogeneity of commercial animals if a
667 single surrogate sire, or a set of very closely related surrogate sires were used. If a
668 disease emerged that this homogenous group of animals was susceptible to, it could
669 have a major detrimental impact on the commercial production. Having such large
670 groups of homogeneous animals would also increase the selection pressure on disease
671 pathogens to evolve pathogenicity to the group. Plant breeders and commercial crop
672 growers have extensive experience in managing the potential to have genetic
673 homogeneity across large segments of the production area. They have developed
674 strategies to minimize the risk of disease outbreaks and other failures such as crop
675 rotation, using multiple varieties on a farm, creating varietal blends consisting of
676 multiple genotypes, and taking holistic strategies to pathogen management [29]. Aside
677 from rotation, which is practically impossible in the animal sector, these strategies
678 might have important roles in ensuring the effectiveness of surrogate sires in
679 livestock.

680 A further risk of the homogenisation of the commercial population relates to
681 genetic diversity. The genetic diversity contained in current populations is potentially
682 a useful reserve of genetic diversity that could be used in breeding programs in case
683 the nucleus genetic diversity was to become inappropriate at some point in the future
684 (e.g., due to a disease catastrophe or because it became exhausted). Homogenisation
685 of the commercial population would remove this safety net requiring greater care to

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

686 be taken in the preservation of genetic diversity. Genebanks using frozen semen, eggs,
687 or embryos are well established ways to preserve genetic diversity. There are also new
688 ways which include the use of cultured primordial germ cells [30].

689 Undetected but highly deleterious mutations also pose a risk for the use of
690 surrogate sires. While it is unlikely that this would arise after sufficient testing, it is
691 not impossible. One such route could be through the occurrence of one or more such
692 mutations arising as somatic mutations after the animal had been tested, leading to a
693 mosaicism, which might affect sets of surrogate sires from the donor.

694 The most obvious opportunity emanating from surrogate sires also relates to
695 the genetic homogeneity of commercial animals and can also draw on practices that
696 are well established in crop production. In crops, management plans are supplied to a
697 farmer alongside the seed (e.g., https://catalog.extension.oregonstate.edu/em9004).
698 These plans are specifically tailored to the variety genotype based on extensive sets of
699 field trials. They include recommendations for target market, expected performance,
700 optimum sowing date, seeding rate, soil type and water, fertilizer, pesticide and
701 fungicide requirements. These management plans complement the genetics of the
702 variety and increase the benefit obtained from the genetic potential in a generic
703 environment. Similar management plans could be developed for surrogate sires and
704 the benefits would be similarly expected to exceed the benefit that was observed in
705 the present study for the genetics alone (e.g., 6.5 year's worth of genetic gain). The
706 phenotype data collected to development of the management plans would also serve
707 to further test and validate a particular donor.

708 Another obvious opportunity emanating from surrogate sires that also relates
709 to the genetic homogeneity is the potential for increasing the product homogeneity. In

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

710 animal production, product uniformity is an important topic. In meat animals, for
711 instance, uniformity has economic benefits because excessive variability in carcass
712 weight or conformation is penalized by slaughterhouses [31,32]. A genetically
713 homogeneous commercial population, achieved through the use of surrogate sires,
714 could aid product uniformity. However, if this was to be achieved, most of the
715 increase in uniformity would need to emanate from matching very specific
716 management plans to the homogenous genetics because homogeneous genetics in
717 itself has limited ability to increase phenotypic homogeneity. Van Vleck [33] showed
718 that in the context of cloned animals, if heritability is 25%, then the phenotypic
719 standard deviation among clones would be 87% of that of uncloned animals and only
720 if heritability is 100%, will clone mates have complete uniformity.

721 Compared to the conventional multiplication the surrogate sires strategy
722 enables shorter lag between nucleus and commercial layer and requires a smaller
723 number of parents contributing to the commercial layer. This offers several
724 advantages including the ability to rapidly change the entire genetics in the
725 commercial layer. This could be used to rapidly respond to sudden changes in
726 requirements such as pressure from a new disease or the emergence of a new market
727 for the product that has specific requirements (e.g., meat marbling).

728 The surrogate sire strategy would be costly to implement in practice because it
729 would require capacity in advanced molecular biology and infrastructure for progeny
730 testing. However, it presents other opportunities through which costs can be saved.
731 For example, multiplier populations to produce terminal sires would not need to be
732 maintained. This would free up resources for other investment in breeding programs,
733 such as more progeny testing of donor candidates.

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

734 The surrogate sires strategy presents breeding programs with an enhanced
735 opportunity to protect its intellectual property via limited release of males (thereby
736 limiting the access of competitors to the broader source germplasm) and by
737 exploitation of specific combining ability. This protection would give the breeding
738 companies incentive to invest more and help to avoid the commonly observed market
739 failure in some breeding industries. When intellectual property is properly protected,
740 breeding companies are anecdotally reported to share the benefits two-thirds to the
741 farmers and one third to the breeding company. Such sharing more than offsets the
742 purchase cost to a producer, while it also gives profit to the breeder. Perhaps the most
743 spectacular example of the benefits of such ways to reward investment in intellectual
744 property are seen in maize which has seen a 6-fold increase in productivity since
745 hybrid breeding was introduced in the 1930's [34]. By releasing hybrids breeding
746 organisations can protect the intellectual property that is their source germplasm. This
747 in turn enables them to invest heavily in breeding activities (e.g., technology, field
748 testing networks) that in turn drive accelerated genetic gains.

749 At least two barriers exist that may prevent the deployment of this technology
750 in in real livestock breeding program. Firstly, genome editing currently appears to be
751 the technology that is most likely to enable genome editing to be implemented in
752 practice [2]. Globally, the future of governmental regulation of genome editing
753 technology is currently uncertain which places uncertainty on the possibility for
754 practical implementation of surrogate sire technology in real livestock breeding
755 program. Secondly, effective deployment of surrogate sire technology will require
756 partitioning of animal breeding programs into population improvement and product
757 development parts. Product development will require deployment of extensive
758 progeny testing schemes. Over the past decade the advent of genomic selection has

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

759 removed progeny testing schemes from many breeding programs. Reinstating such
760 schemes would be costly and further work will be needed to demonstrate the exact
761 return on investment.

762 Finally, the results of this study raise an important question for existing
763 breeding programs that use artificial insemination for dissemination. As noted above,
764 genomic selection has led to the removal of progeny testing schemes from many
765 livestock breeding programs. Our results raise some doubts about the merit of this.
766 They show that when a breeding program releases a small number of individuals that
767 are deployed widely there is a benefit to progeny testing these individuals. The degree
768 of benefit depends on the accuracy of genomic selection, the number of individuals
769 released and their subsequent usage, and the accuracy and the number of stages in a
770 progeny testing scheme and the relative time taken to perform a progeny test.
771 Determining whether the removal of progeny testing schemes from genomic selection
772 driven livestock breeding programs was the right thing to do in retrospect is beyond
773 the scope of the present study but is an interesting question for future research.

774

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

775 # Conclusions

776 The results of this study showed that using the surrogate sires strategy could
777 significantly increase the genetic merit of commercial sires, by as much as 6.5 to 9.2
778 years' worth of genetic gain, compared to the conventional multiplication strategy.
779 The simulations suggest that identifying elite donors for surrogate sires should be
780 based on three stages, the first of which uses a genomic test followed by two
781 subsequent progeny tests. The use of one or a handful of elite donors to generate
782 surrogate sires that in turn give rise to all production animals would be very different
783 to current practice. While the results demonstrate the great potential of surrogate sires
784 strategy there are considerable risks as well as opportunities. Practical implementation
785 of surrogate sires strategy would need to account for these.

786

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

787 # Competing interests

788 The authors declare that they have no competing interests.

789 # Authors' contributions

790 JMH conceived the study. JMH, MB, and PG designed the study. PG performed the
791 analysis. JMH and PG wrote the first draft of the manuscript. GG, MB, RCG, JJ,
792 RRF, CBAW, AJM, and WOH helped to interpret the results and refine the
793 manuscript. All authors read and approved the final manuscript.

794 # Acknowledgements

795 The authors acknowledge the financial support from the BBSRC ISPG to The Roslin
796 Institute BBS/E/D/30002275, from Grant Nos. BB/N015339/1, BB/L020467/1,
797 BB/M009254/1, from Genus PLC and from Innovate UK. This work has made use of
798 the resources provided by the Edinburgh Compute and Data Facility (ECDF)
799 (http://www.ecdf.ed.ac.uk).

800

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

801 # References

802 1. Park K-E, et al. (2017) Generation of germline ablated male pigs by CRISPR/Cas9
803 editing of the NANOS2 gene. 7:40176.

804 2. Oatley Jon M. (2017) Recent advances for spermatogonial stem cell transplantation
805 in livestock. *Reproduction, fertility and Development 30:44-49*.
806

807 3. Zhang Z., Shao S., and Meistrich M. L. (2006) Irradiated mouse testes efficiently
808 support spermatogenesis derived from donor germ cells of mice and rats. *J.*
809 *Androl.***27**: 365–375.
810

811 4. Tsuda M., Sasaoka Y., Kiso M., Abe K., Haraguchi S., Kobayashi S., and Saga Y.
812 (2003) Conserved role of nanos proteins in germ cell development. *Science* **301**:
813 1239–1241.
814

815 4. De Roos APW. (2011) Recent trends in genomic selection in dairy cattle. 62nd
816 Annu. Meet. EAAP. Stavanger, Norway: Wageningen Academic Publishers: p.
817 Contribution 01–7, p. 4.

818 5. Pryce JE, Daetwyler HD (2012) Designing dairy cattle breeding schemes under
819 genomic selection: a review of international research. *Anim Prod Sci* 52(3):107–114.

820 6. Meuwissen T, Hayes B, Goddard M (2013) Accelerating Improvement of
821 Livestock with Genomic Selection. *Annu Rev Anim Biosci* 1(1):221–237.

822 7. Van Eenennaam AL, Weigel KA, Young AE, Cleveland MA, Dekkers JCM (2014)
823 Applied Animal Genomics: Results from the Field. *Annu Rev Anim Biosci* 2(1):105–
824 139.

825 8. Knol EF, Nielsen B, Knap PW (2016) Genomic selection in commercial pig
826 breeding. *Anim Front* :6-15.

827 9. Visscher P, Pong-Wong R, Whittemore C, Haley C (2000) Impact of biotechnology
828 on (cross)breeding programmes in pigs. *Livestock Production Science* 65(1):57–70.

829 10. Bernardo R (2014) Essentials of Plant Breeding. Stemma Press (2014), pp. 252,
830 £145.00 (pb). ISBN 13:9780972072427.

831 11. Gaynor RC, Gorjanc G, Bentley AR, Howell P, Jackson R, et al. (2017) A two-
832 part strategy for using genomic selection to develop inbred lines. *Crop Science*
833 57:2372-2386.

834 12. Chen GK, Marjoram P, Wall JD (2009) Fast and flexible simulation of DNA
835 sequence data. *Genome Research* 19(1):136–142.

836 13. Faux A-M, et al. (2016) AlphaSim: Software for Breeding Program
837 Simulation. *The Plant Genome* 9(3). doi:10.3835/plantgenome2016.02.0013.

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

838 14. Rodríguez-Ramilo ST, Fernández J, Toro MA, Hernández D, Villanueva B (2015)
839 Genome-Wide Estimates of Coancestry, Inbreeding and Effective Population Size in
840 the Spanish Holstein Population. *PLOS ONE*10(4):e0124157.

841 15. Hill WG (2009) Understanding and using quantitative genetic variation. *Philos*
842 *Trans R Soc Lond B Biol Sci* 365(1537):73.

843 16. Brotherstone S, Goddard M (2005) Artificial selection and maintenance of genetic
844 variance in the global dairy cow population. *Philos Trans R Soc Lond B Biol*
845 *Sci* 360(1459):1479.

846 17. Hayes BJ, Visscher PM, McPartlan HC, Goddard ME (2003) Novel Multilocus
847 Measure of Linkage Disequilibrium to Estimate Past Effective Population Size.
848 *Genome Res* 13(4):635–43.

849 18. G.E. Dickerson, Hazel LN. Effectiveness of selection on progeny performance as
850 a supplement to earlier culling in livestock. *Journal of Agricultural Research* 69:459.

851 19. Cunningham EP (1975) Multi-stage index selection. *Theoretical and Applied*
852 *Genetics* 46:55–61.

853 20. Saxton AM (1983) A comparison of exact and sequential methods in multi-stage
854 index selection. *Theoretical and Applied Genetics* 66(1):23–28.

855 21. Ducrocq V, Colleau J J (1989) Optimum truncation points for independent culling
856 level selection on a multivariate normal distribution, with an application to dairy
857 cattle selection. *Genet Sel Evol* 21(2):185–198.

858 22. Xie C, Xu S, Mosjidis JA (1997) Multistage selection indices for maximum
859 genetic gain and economic efficiency in red clover. *Euphytica* 98(1):75–82.

860 23. Mi X, Utz HF, Technow F, Melchinger AE (2014) Optimizing Resource
861 Allocation for Multistage Selection in Plant Breeding with R Package
862 Selectiongain. *Crop Science* 54(4):1413–1418.

863 24. Robertson A (1957) Optimum Group Size in Progeny Testing and Family
864 Selection. *Biometrics* 13(4):442–450.

865 25. Meuwissen THE, Goddard ME (1997) Optimization of progeny tests with prior
866 information on young bulls. *Livestock Production Science* 52(1):57–68.

867 26. Milan D, Jeon J-T, Looft C, Amarger V, Robic A, Thelander M, et al. (2000) A
868 Mutation in PRKAG3 Associated with Excess Glycogen Content in Pig Skeletal
869 Muscle. *Science* 288(5469):1248-51.

870 27. VanRaden PM, Olson KM, Null DJ, Hutchison JL (2011) Harmful recessive
871 effects on fertility detected by absence of homozygous haplotypes. *Journal of Dairy*
872 *Science* 94(12):6153–61.

873 28. Kadri NK, Sahana G, Charlier C, Iso-Touru T, Guldbrandtsen B, Karim L, et al.
874 (2014) A 660-Kb Deletion with Antagonistic Effects on Fertility and Milk Production
875 Segregates at High Frequency in Nordic Red Cattle: Additional Evidence for the

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

876 Common Occurrence of Balancing Selection in Livestock. *PLOS Genetics*
877 10:e1004049.

878 29. McDonald BA, Linde C (2002) The population genetics of plant pathogens and
879 breeding strategies for durable resistance. *Euphytica* 124:163–80.

880 30. Nandi S, Whyte J, Taylor L, Sherman A, Nair V, Kaiser P, et al. (2016)
881 Cryopreservation of specialized chicken lines using cultured primordial germ cells.
882 *Poultry Science* 95:1905–11.

883 31. Mulder HA, Bijma P, Hill WG (2007) Prediction of Breeding Values and
884 Selection Responses With Genetic Heterogeneity of Environmental Variance.
885 *Genetics* 175:1895.

886 32. Mulder HA, Bijma P, Hill WG (2008) Selection for uniformity in livestock by
887 exploiting genetic heterogeneity of residual variance. *Genetics Selection*
888 *Evolution* 40(1):37.

889 33. van Vleck LD (1999) Implications of cloning for breed improvement strategies:
890 are traditional methods of animal improvement obsolete? *J Anim Sci* 77 Suppl 2:111–
891 121.

892 34. Duvick DN (2005) Genetic Progress in yield of United States maize (Zea mays
893 L.). *Maydica* 50:193–202.

894

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

895 # Figures

896 **Fig. 1** Schematic depicting the possible application of spermatogonial stem cell
897 transplantation methodology in pig production (depiction inspired by Oatley et al.,
898 2018 [2])

899 **Fig. 2** Example animal (left) and plant (right) breeding schemes

900 **Fig. 3** Timeline of the different strategies to identify and disseminate genetic
901 improvement

902 **Fig. 4** Map of the scenarios used in the study

903 **Fig.5** Average genetic merit of commercial sires derived from the best performing
904 surrogate sire strategy scenario and the conventional strategy (top 50, 200 and 500
905 males) for SmallScenario (a and b) and BigScenario (c and d) plotted against time

906

907 # Tables

908 **Table 1** Average Years' worth of Genetic Gain (YGG) of the best performing
909 surrogate sire strategy scenario above the conventional strategy that uses either 50,
910 200, or 500 males

911 **Table 2** Average Years' worth of Genetic Gain (YGG) with the two-stage testing
912 scenarios of the surrogate sire strategy above the conventional strategy that uses 50
913 males (SmallScenario)

914

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

915 **Table 3** Average Years' worth of Genetic Gain (YGG) with the two-stage testing
916 scenarios of the surrogate sire strategy above the conventional strategy that uses 50
917 males (BigScenario)

918 **Table 4** Average Years' worth of Genetic Gain (YGG) with the three-stage testing
919 scenarios of the surrogate sire strategy with one elite donor above the conventional
920 strategy that uses 50 males (SmallScenario)

921 **Table 5** Average Years' worth of Genetic Gain (YGG) with the three-stage testing
922 scenarios of the surrogate sire strategy with five elite donors above the conventional
923 strategy that uses 50 males (SmallScenario)

924

925

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

926 # Supplementary material

927 **Fig. S1** Average genetic merit of commercial sires derived from the best performing
928 surrogate sire strategy scenario and the conventional strategy (top 50, 200 and 500
929 males) for SmallScenario (a) and BigScenario(b) plotted against time

930 ***Table S1*** Average Years' worth of Genetic Gain (YGG) with the three-stage testing
931 scenarios of the surrogate sire strategy with one elite donor above the conventional
932 strategy that uses 50 males (BigScenario)

933 **Table S2** Average Years' worth of Genetic Gain (YGG) with the three-stage testing
934 scenarios of the surrogate sire strategy with five elite donors above the conventional
935 strategy that uses 50 males (BigScenario)

936

937

938

939

940

941

942

943

944

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

945

946

947

948

949

950

Fig. 1 Schematic depicting the possible application of spermatogonial stem cell transplantation methodology in pig production (depiction inspired by Oatleyet al.,2018[2])

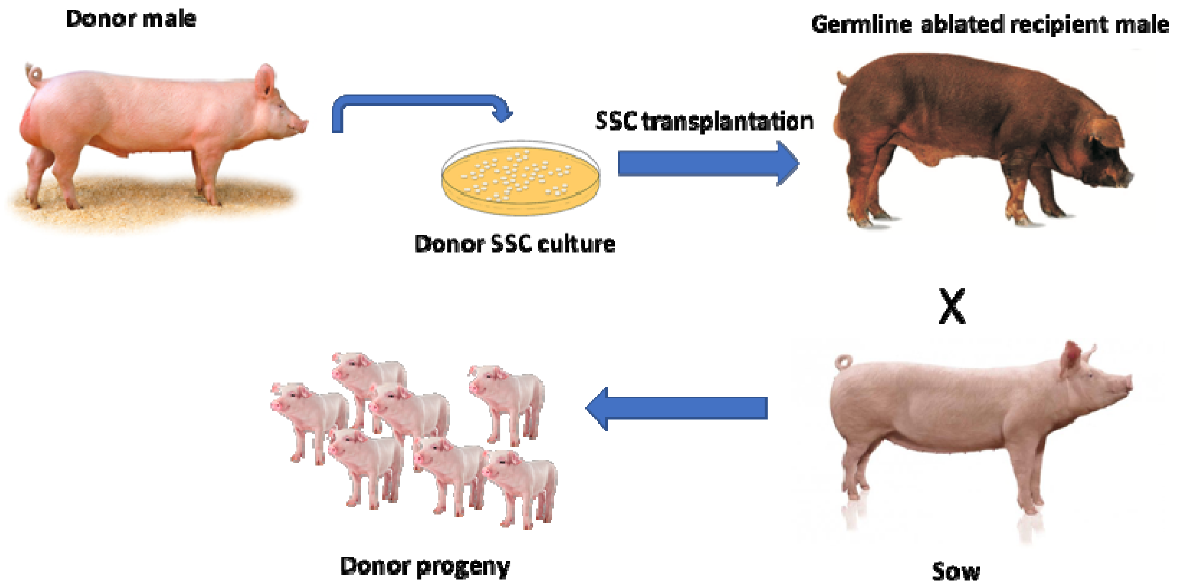
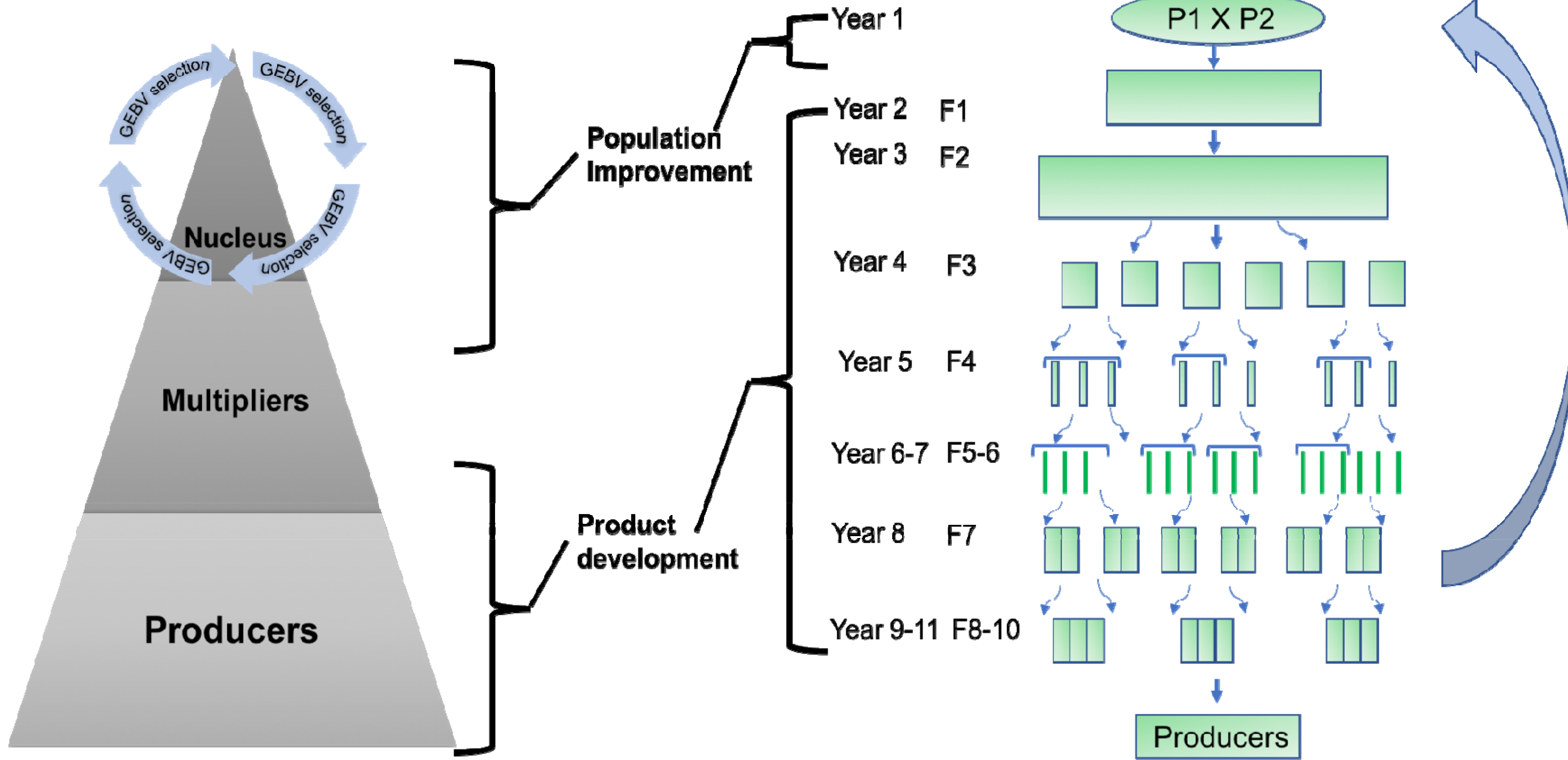


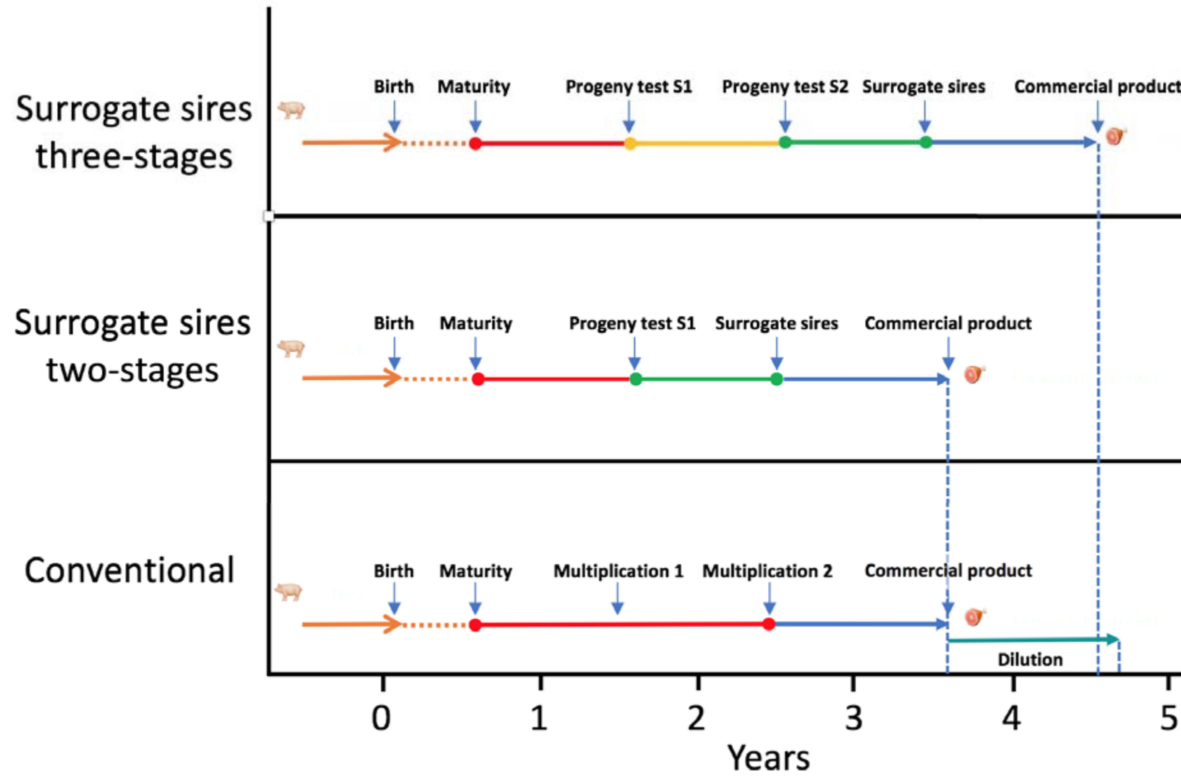
Fig. 2 Typical animal (left) and plant (right) breeding schemes

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

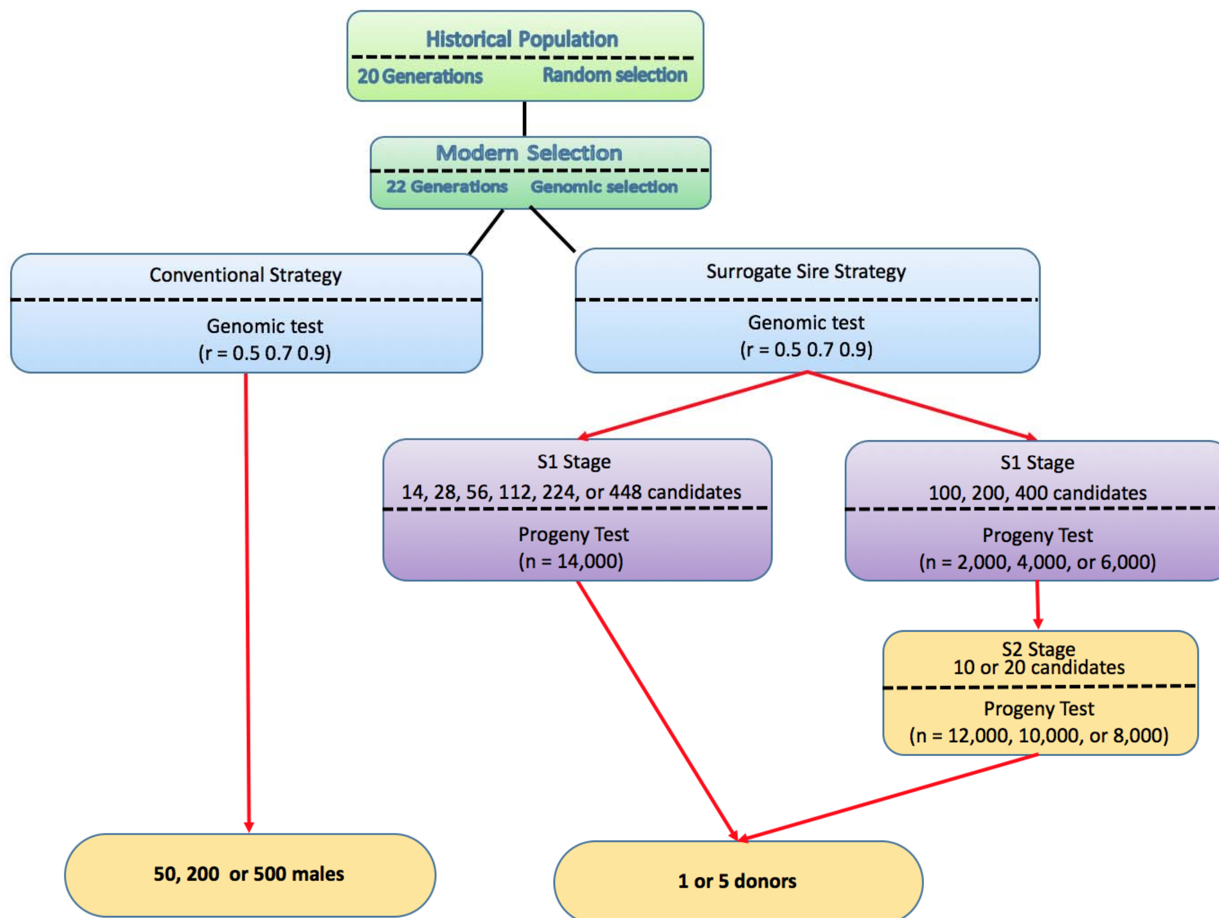
Fig. 3 Timeline of the different strategies to identify and disseminate genetic improvement



bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Fig. 4 Map of the scenarios used in the study

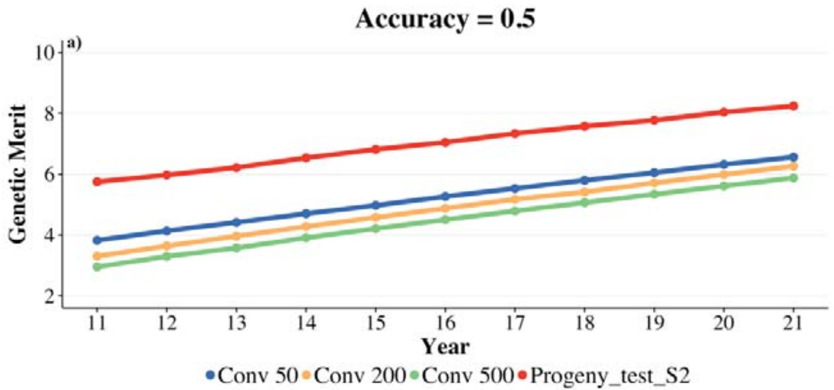
bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

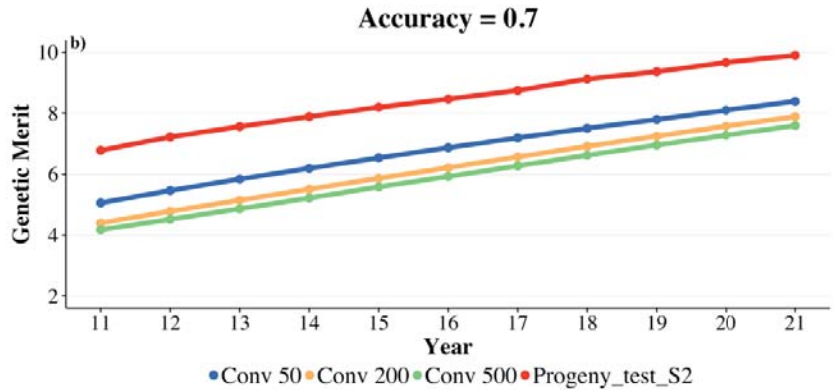


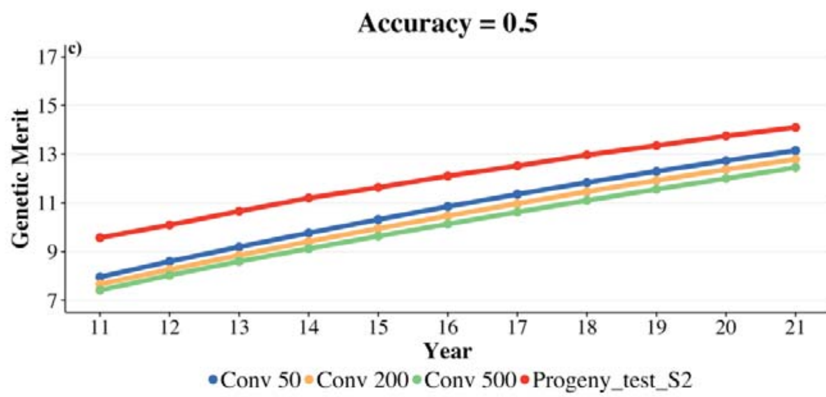
bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

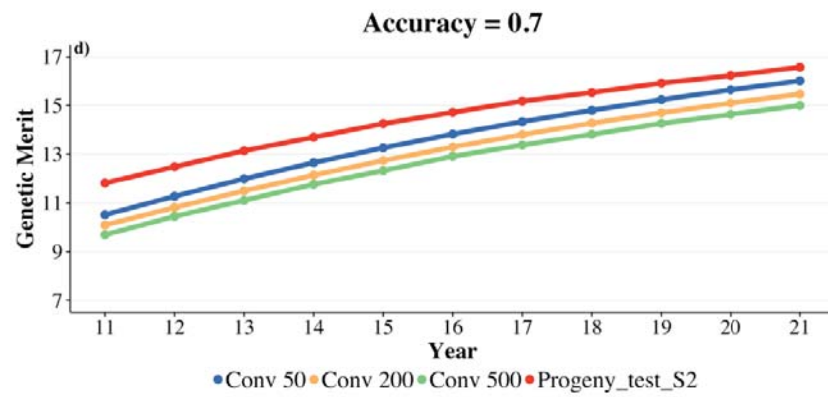
bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Fig. 5 Average genetic merit of commercial sires derived from the best performing surrogate sire strategy scenario and the conventional strategy (top 50, 200 and 500 males) for SmallScenario (a and b) and BigScenario (c and d) plotted against time



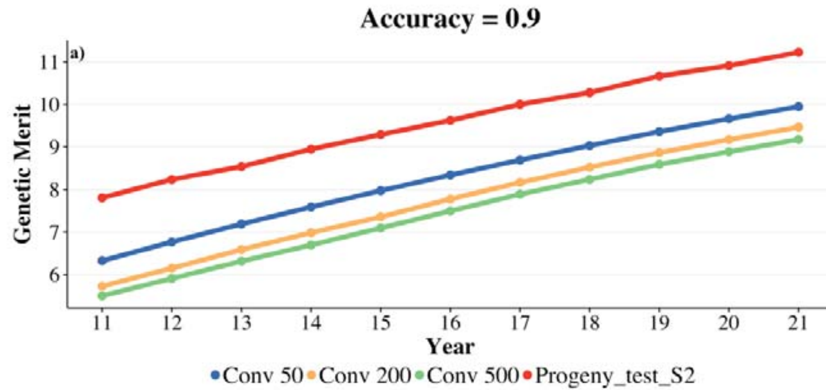


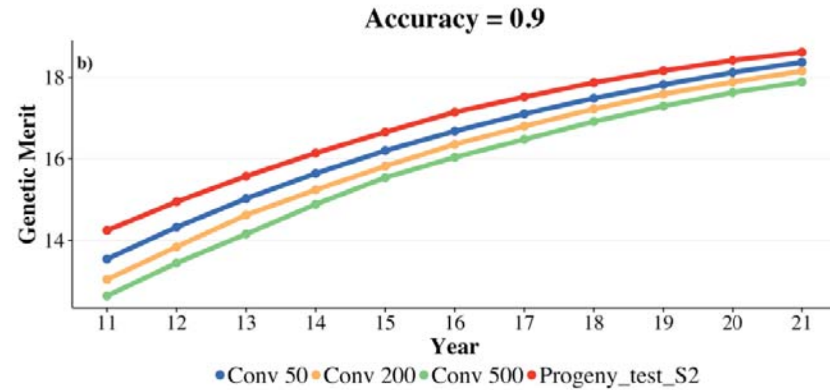




bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Fig. S1 Average genetic merit of commercial sires derived from the best performing surrogate sire strategy scenario and the conventional strategy (top 50, 200 and 500 males) for SmallScenario (a) and BigScenario(b) plotted against time





bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Table 1 Average Years' worth of Genetic Gain (YGG) of the best performing surrogate sire strategy scenario above the conventional strategy that uses either 50, 200, or 500 males

Genomic test accuracy	Males progeny tested S1	Males progeny tested S2	Progeny test resources[1]	Donors used	YGG_{50}	YGG_{200}	YGG_{500}
Small Scenario							
0.5	100	20	6000S1 / 8000S2	1	6.5	7.5	9.2
0.7	200	20	6000S1 / 8000S2	1	4.5	6.5	7.2
0.9	200	20	6000S1 / 8000S2	1	2.4	4.5	5.0
Big Scenario							
0.5	100	20	6000S1 / 8000S2	1	2.7	3.5	4.1
0.7	200	20	6000S1 / 8000S2	1	2.1	2.5	3.5
0.9	200	20	6000S1 / 8000S2	1	1.2	1.7	2.5

[1]*Total number of progeny allocated in the first progeny test (S1) and in the second progeny test (S2)*

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Table 2 Average Years' worth of Genetic Gain (YGG) with the two-stage testing scenarios of the surrogate sire strategy above the conventional strategy that uses 50 males (SmallScenario)

Males Tested	Progeny/Male	Donors used	$YGG_{0.5}$[1]	$YGG_{0.7}$[1]	$YGG_{0.9}$[1]
14	1000	1	4.1	3.0	1.8
28	500	1	4.7	3.0	1.2
56	250	1	5.1	3.5	2.2
112	125	1	5.3	3.6	2.2
224	63	1	4.8	2.8	1.3
448	31	1	3.8	2.1	1.1
14	1000	5	2.9	1.9	0.2
28	500	5	3.1	2.1	0.5
56	250	5	3.6	2.4	1.1
112	125	5	3.6	2.6	1.2
224	63	5	3.4	1.9	0.3
448	31	5	2.8	1.6	0.2

[1]*Genomic test accuracy at the initial stage (S0) 0.5, 0.7, and 0.9*

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Table 3 Average Years' worth of Genetic Gain (YGG) with the two-stage testing scenarios of the surrogate sire strategy above the conventional strategy that uses 50 males (BigScenario)

Males Tested	Progeny/Male	Donors used	$YGG_{0.5}$[1]	$YGG_{0.7}$[1]	$YGG_{0.9}$[1]
14	1000	1	2.3	1.7	0.7
28	500	1	2.4	1.9	0.8
56	250	1	2.5	2.0	1.0
112	125	1	2.5	2.0	1.1
224	63	1	2.0	1.8	0.8
448	31	1	1.9	1.5	0.4
14	1000	5	1.7	1.2	0.5
28	500	5	1.7	1.2	0.7
56	250	5	1.9	1.1	1.0
112	125	5	2.0	1.1	1.0
224	63	5	1.8	1.0	0.5
448	31	5	1.0	0.8	0.3

[1]*Genomic test accuracy at the initial stage (S0) 0.5, 0.7, and 0.9*

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Table 4 Average Years' worth of Genetic Gain (YGG) with the three-stage testing scenarios of the surrogate sire strategy with one elite donor above the conventional strategy that uses 50 males (SmallScenario)

Progeny test resources[1]	Males progeny tested S1	Progeny/Male S1	Males progeny tested S2	Progeny/Male S2	$YGG_{0.5}$[2]	$YGG_{0.7}$[2]	$YGG_{0.9}$[2]
2000S1/12000S2	100	20	10	1200	5.3	3.5	2.2
			20	600	5.4	3.6	2.4
	200	10	10	1200	4.9	3.2	2.2
			20	600	5.1	3.3	2.1
	400	5	10	1200	4.5	3.7	2.0
			20	600	4.7	2.7	1.4
4000S1/10000S2	100	40	10	1000	5.5	3.6	2.2
			20	500	5.8	4.0	2.3
	200	20	10	1000	5.3	3.5	2.4
			20	500	5.4	3.8	2.3
	400	10	10	1000	4.3	3.3	1.6
			20	500	4.5	3.5	1.4
6000S1/8000S2	100	60	10	800	5.9	4.1	2.0
			20	400	6.5	4.2	2.2
	200	30	10	800	5.3	4.2	2.1
			20	400	5.7	4.5	2.4
	400	15	10	800	5.0	3.4	1.6
			20	400	5.8	3.5	1.2

[1]*Number of total progeny allocated in the first progeny test (S1) and in the second progeny test(S2)*
[2]*Genomic test accuracy at the initial stage (S0) 0.5, 0.7, and 0.9*

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Table 5 Average Years' worth of Genetic Gain (YGG) with the three-stage testing scenarios of the surrogate sire strategy with five elite donors above the conventional strategy that uses 50 males (SmallScenario)

Progeny test resources[1]	Males progeny tested S1	Progeny/Male S1	Males progeny tested S2	Progeny/Male S2	$YGG_{0.5}$[2]	$YGG_{0.7}$[2]	$YGG_{0.9}$[2]
2000S1/12000S2	100	20	10	1200	4.1	2.1	1.1
			20	600	4.4	2.2	1.2
	200	10	10	1200	3.0	2.1	1.2
			20	600	3.7	2.5	1.3
	400	5	10	1200	2.2	1.5	1.0
			20	600	2.2	1.4	1.0
4000S1/10000S2	100	40	10	1000	4.4	2.4	1.3
			20	500	4.5	2.5	1.2
	200	20	10	1000	4.1	2.2	1.1
			20	500	4.1	2.7	1.2
	400	10	10	1000	4.2	1.7	1.0
			20	500	4.2	2.0	1.8
6000S1/8000S2	100	60	10	800	4.5	3.1	1.6
			20	400	5.0	3.2	1.8
	200	30	10	800	4.6	2.1	1.3
			20	400	5.0	2.2	1.4
	400	15	10	800	4.1	1.7	1.2
			20	400	4.6	2.2	1.2

[1]Number of total progeny allocated in the first progeny test(S1) and in the second progeny test(S2)
[2]Genomic test accuracy at the initial stage (S0) 0.5, 0.7, and 0.9

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Table S1 Average Years' worth of Genetic Gain (YGG) with the three-stage testing scenarios of the surrogate sire strategy with one elite donor above the conventional strategy that uses 50 males (BigScenario)

Progeny test resources[1]	Males progeny tested S1	Progeny/Male S1	Males progeny tested S2	Progeny/Male S2	$YGG_{0.5}$[2]	$YGG_{0.7}$[2]	$YGG_{0.9}$[2]
2000S1/12000S2	100	20	10	1200	2.2	1.5	0.9
			20	600	2.3	1.7	0.9
	200	10	10	1200	2.1	1.6	1.0
			20	600	2.2	1.1	0.9
	400	5	10	1200	2.2	1.3	0.9
			20	600	2.3	1.5	0.8
4000S1/10000S2	100	40	10	1000	2.2	1.6	0.8
			20	500	2.1	1.6	0.8
	200	20	10	1000	2.1	2.1	0.9
			20	500	2.2	2.1	1.0
	400	10	10	1000	2.3	1.7	0.9
			20	500	2.3	2.0	0.9
6000S1/8000S2	100	60	10	800	2.5	2.0	1.0
			20	400	2.7	2.1	1.1
	200	30	10	800	2.4	1.9	1.1
			20	400	2.6	2.1	1.2
	400	15	10	800	2.3	2.0	0.9
			20	400	2.4	2.0	0.9

[1]*Number of total progeny allocated in the first progeny test(S1) and in the second progeny test(S2)*
[2]*Genomic test accuracy at the initial stage (S0) 0.5, 0.7, and 0.9*

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Table S2 Average Years' worth of Genetic Gain (YGG) with the three-stage testing scenarios of the surrogate sire strategy with five elite donors above the conventional strategy that uses 50 males (BigScenario)

Progeny test resources[1]	Males progeny tested S1	Progeny/Male S1	Males progeny tested S2	Progeny/Male S2	$YGG_{0.5}$[2]	$YGG_{0.7}$[2]	$YGG_{0.9}$[2]
2000S1/12000S2	100	20	10	1200	1.7	1.2	0.6
			20	600	1.6	1.2	0.8
	200	10	10	1200	1.6	1.2	0.6
			20	600	1.5	1.0	0.3
	400	5	10	1200	1.6	1.1	0.7
			20	600	1.6	1.2	0.8
4000S1/10000S2	100	40	10	1000	1.7	1.3	0.7
			20	500	1.7	1.2	0.6
	200	20	10	1000	1.6	1.3	0.4
			20	500	1.6	1.4	0.6
	400	10	10	1000	1.3	1.3	0.3
			20	500	1.4	1.4	0.4
6000S1/8000S2	100	60	10	800	1.8	1.1	0.4
			20	400	1.8	1.4	0.7
	200	30	10	800	1.6	1.2	0.5
			20	400	1.6	1.2	0.5
	400	15	10	800	1.7	1.2	0.7
			20	400	1.6	1.2	0.9

[1]*Number of total progeny allocated in the first progeny test(S1) and in the second progeny test(S2)*
[2]*Genomic test accuracy at the initial stage (S0) 0.5, 0.7, and 0.9*

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

bioRxiv preprint doi: https://doi.org/10.1101/199893; this version posted November 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.