Genomic selection strategies for breeding adaptation and production in dairy cattle under climate change

Ismo Strandén[1], Juha Kantanen[1], Isa-Rita M. Russo[2], Pablo Orozco-terWengel[2], Michael W Bruford[2] and the Climgen Consortium[a]

[1] Natural Resources Institute Finland (Luke), Jokioinen, Finland

[2] School of Biosciences, Cardiff University, Sir Martin Evans Building, Museum Avenue, Cardiff, CF10 3AX, UK.

Corresponding author: Ismo Strandén

Mailing address: Natural Resources Institute Finland (Luke), Myllytie 1, FI-31600 Jokioinen, Finland

Phone: +358 29 532 6553

Fax: none

Email: Ismo.Stranden@Luke.fi

[a] The members of Climate Genomics Consortium are listed in Appendix.

RUNNING TITLE: Genomic selection for adaptation in dairy cattle

Abstract

Livestock production both contributes to, and is affected, by global climate change, and substantial modifications will be required to increase its climate resilience. In this context, reliance on dominant commercial livestock breeds, featuring small effective population sizes, makes current production strategies vulnerable if their production is restricted to environments which may be too costly to support under future climate scenarios. The adaptability of animal populations to future environments will therefore become important. To help evaluate the role of genetics in climate adaptation, we compared selection strategies in dairy cattle using breeding simulations, where genomic selection was used on two negatively correlated traits for production (assumed to be moderately heritable) and adaptation (assumed to have low heritability). Compared to within population breeding, genomic introgression produced a more positive genetic change for both production and adaptation traits. Genomic introgression from highly adapted but low production value populations into highly productive but low adaptation populations was most successful when the adaptation trait was given a lower selection weight than the production trait. Genomic introgression from highly productive population to highly adapted population was most successful when the adaptation trait was given a higher selection weight than the production trait. Both these genomic introgression schemes had the lowest risk of inbreeding. Our results suggest that both

adaptation and production can potentially be improved simultaneously by genomic introgression.

Keywords: adaptation, dairy cattle, introgression, simulation

Introduction

Adaptation of livestock to environmental challenges is becoming increasingly important in cost-effective animal production, especially as the climate becomes warmer, conditions for diseases are more favourable and production costs are set to rise (FAO, 2015; Phocas et al., 2016). Heat stress is one of the most pressing factors affecting livestock production (Niyas et al., 2015) and has been shown to cause a decrease in milk production (5-15%) and lower conception rates (Berman, 2011) in cattle. In addition, chronic stress triggers metabolic changes that result in stress-related disease and suppression of innate immunity (Das, 2016). However, selection can compensate for thermal stress with, for example, the slick hair coat phenotype being implicated in the thermoregulation of tropically adapted cattle breeds (Pitt et al., 2018).

In practice, adaptation comprises many traits including those influencing fitness, such as longevity and disease resistance. These traits typically have low heritability and have declined when milk production has increased (e.g. Mirkena et al., 2010). Genetic correlations have been estimated to range from -0.11 to -0.84 between milk yield and functional longevity (e.g. Sasaki, 2013; Pritchard et al., 2013a,b). Because adaptation to the local environment generally has a low heritability and possibly has antagonistic genetic correlations with milk production, long-term and efficient breeding strategies are required. In a rapidly changing environment, one approach is

to introgress locally adaptive genes found in autochthonous breeds into major production breeds or *vice versa* (Nardone et al., 2006; Hoffman, 2010; Berman, 2011; Hoffman, 2013). Thus, an efficient strategy may be needed to introduce adaptive traits quickly into commercial breeds or these breeds may need to be replaced with better adapted populations (Åby and Meuwissen, 2014).

Introgression strategies commonly assume that one or more alleles in genes of interest or associated markers have been located in a donor population but are missing in the recipient population (e.g., Visscher et al., 1996). The aim is to select these favourable alleles from individuals within the donor population and use backcrossing and selection to introduce them into the recipient population, such that the favourable allele becomes fixed in the recipient population with as small as possible proportion of the rest of the donor's genome included. However, the allele(s) of interest need to be known and to be of large effect, which is unfortunately seldom the case.

Knowing the location of important genes affecting traits of interest may be unnecessary for introgression to succeed. Ødegård et al. (2009a) simulated a fish breeding population where introgression was applied for a major quantitative trait locus (QTL) affecting disease resistance by backcrossing a production line with a resistant donor line. In their study, classical selection, i.e., without genomic information, was inefficient but genomic selection without specific knowledge of the target QTL was

usually effective in preserving favourable alleles. Gaspa et al. (2015) investigated introgression of polledness in Holstein Friesian cattle by using simulations and concluded that a single gene strategy, applying genomic selection helped to speed up the process of introgression while simultaneously increasing the genetic gain of other important traits and reducing inbreeding.

Adaptation and production can be both assumed to be polygenic traits. Ødegård et al. (2009b) simulated a fish breeding scheme where both production and disease resistance were polygenic with either low or high heritability. The authors concluded that in contrast to classical selection, genomic selection increased genetic gain in introgression backcrossing schemes, with the largest gain being for low heritability traits and on traits not recorded in selection candidates.

Åby and Meuwissen (2014) simulated two divergent populations according to production and fitness profiles of livestock with pure and crossbreeding in discrete generations. Both production and fitness were polygenic traits having moderate heritability but no genetic correlation. According to their results, selection using breeding values estimated by genomic best linear unbiased prediction (GBLUP) outperformed conventional BLUP in terms of genetic increase in production and fitness. In general, results from the simulation studies suggest that genomic selection can be effective in introgression of a lowly heritable trait to a

target population with high production when traits (i.e. the introgressed trait and production) are polygenic and uncorrelated.

The aim of this study was to evaluate breeding strategies for the selection of rapid environmental (e.g. temperature) adaptation and production using dairy cattle as a model, by stochastic simulation. Breeding strategies include pure breeding and crossbreeding schemes involving a poorly adapted but high production population and a well-adapted population but of low milk yield. Adaptation and production traits were assumed to be negatively correlated and governed by many loci. We assumed production to have a higher heritability than adaptation and use bivariate genomic BLUP to estimate breeding values to improve both adaptation and production. A variety of breeding strategies using selection and introgression were considered. Introgression schemes included selection from well-adapted to poorly adapted population and *vice versa* when selection strategies applied different selection weights to the traits.

Material and Methods

Simulation of breeding programs followed two phases. First, QMSim (Sargolzaei and Schenkel, 2009) was used to simulate an initial historic population (HP) and the subsequent selection of two divergent lines, a production line (PL) and an adaptation line (AL). Second, the final breeding animals from these populations were available as breeding animals for alternative selection schemes. Five replicates were simulated using QMSim,

i.e. five sets of final breeding animals. Within a replicate, the studied breeding schemes selected the first breeding animals from the same population although different schemes could use different sets of animals. Table 1 shows the simulation parameters used.

Simulation of two populations using QMSim

The HP in QMSim generates initial values of linkage disequilibrium (LD), mutation and drift. The simulated genome was assumed to have 30 chromosomes of 100 cM each. To mimic a commercial Bovine 54K SNP chip, the genome had 54,000 evenly distributed bi-allelic SNP markers with equal frequencies (0.5) for the two alleles in the base population (1,800 bi-allelic SNP markers per chromosome). The HP consisted of 1,000 individuals that were randomly mated for 95 generations (Figure 1). During the following five generations, the population was expanded to 12,000 individuals to allow selection of two populations.

QMSim allows selection of one trait in the populations descending from the HP, and here this trait was chosen to represent milk production in dairy cattle. The production trait was assumed to have 30 randomly positioned QTLs within each chromosome, i.e., in total 900 QTLs. The mutation rate used was 2.5×10^{-5} per generation (e.g. Solberg et al. 2008). The number of recombinations per Morgan was sampled from a Poisson distribution with a mean of one and the cross-overs were randomly placed on the chromosomes. QTL effects of the production trait were introduced in the

last generation of the HP for a trait with a heritability of 0.30 which is close to heritability of 305-day milk yield in dairy cattle. The allelic effects were sampled from a gamma distribution with a shape parameter of 0.4 as implemented in QMSim, so that the QTLs explained all genetic variation. QTL allelic effects were scaled in QMSim such that the phenotypic variance in the last HP generation was one.

Animals for two populations or breeds (AL and PL) were selected from the HP, and subsequent breeding in QMSim was carried out separately within these breeds (Figure 1). For both populations, 2,800 females and 200 males were selected from the last generation of the HP to be as breeding animals. The AL population animals were randomly selected but the PL population animals had the highest true breeding value for production. No animal was used in both populations. Within both populations each sire was mated to 14 dams, with each mating producing one offspring (50:50 birth sex ratio). This procedure yielded a total of 2,800 offspring (1,400 males and 1,400 females). In the following 10 generations in QMSim, selection of breeding animals in the AL population was random but in the PL population animals with highest pseudo EBVs (estimated breeding values) were selected for breeding.

Selection within the AL and PL populations allowed for overlapping generations. The breeding animals in each generation were selected from the current breeding animals and from the youngest mature generation. From

the 200 current breeding males, 60% (120 sires) were kept and 40% (80 sires) were replaced. In the AL population, the 80 new sires were randomly selected from the youngest mature generation, and they replaced a randomly selected 80 older breeding males. In the PL population, the selection used the pseudo EBVs calculated by QMSim based on 20 daughter records. The breeding animals replaced had the lowest pseudo EBV among the current breeding males, and the selected new males had the highest pseudo EBV in the youngest mature generation. Correspondingly, 20% (560 females) were replaced among the 2,800 breeding females. In the AL population, the selection of replacements and replaced females were random. In the PL population, the selection used the pseudo EBVs based on the cow's own information. Thus, the two populations had the same random drift base population from which the PL population was selected for higher production, but the AL population continued with random selection. The AL population depicts a random drift population where there has been no efficient selection for production or adaptation which reflects the prevailing situation for many small local breeds. The breeding scheme was continued for 10 generations separately within AL and PL to produce two genetically different populations.

After this set of simulations, QTL effects for the adaptation trait were further simulated as a correlated trait to production. In the simulations, all the production QTL positions and effects were used to generate a correlated

196 trait which had a genetic correlation of -0.30 and heritability of 0.10 for all
197 animals after the HP. Heritability of 0.10 is close to lowly heritable
198 adaptation traits such as functional longevity that have moderately negative
199 genetic correlation with milk production (Sasaki et al. 2013). In practice,
200 there can be adaptation traits having lower heritability and possibly weaker
201 or stronger genetic correlation with production, but use of different values
202 would affect only absolute values, not the observed trends. The adaptation
203 trait's QTL effects were scaled the same way as in QMSim for production in
204 order to get the wanted genetic correlation of -0.3 between production and
205 adaptation. However, the scaling of adaptation QTLs used information from
206 one generation later than for the production QTLs because that was the first
207 available generation with data from QMSim. Correlations of true breeding
208 values (TBV) between the traits indicated that the wanted correlation of -0.3
209 was realized. TBV of an animal was calculated using the true QTL values
210 and the QTL genotypes of animal.

211 **Breeding program simulations**

212 Different breeding schemes were simulated for an additional ten simulation
213 years after the QMSim simulation. These schemes used the final breeding
214 animals from the QMSim simulation (Figure 1) but due to restrictions in the
215 QMSim program, subsequent breeding program simulations were performed
216 using a different set of computer programs. In every simulation scheme, a
217 simulation year used current breeding animals to produce calves. A new set

218 of breeding animals were selected from the current breeding animals and
219 calves maturing from previous year according to their EBVs (Table 2), i.e.
220 breeding animals could be from a number of birth years. The basic
221 principles in the simulation followed those described for QMSim. However,
222 animals were simulated to mature after one year instead of immediately
223 being available as in QMSim.

224 Selection was based on EBVs that used genomic information and
225 observations from the final breeding animals in QMSim and subsequent
226 years. EBVs were calculated using a two trait SNP-BLUP where the
227 variance components were equal to the simulation parameters. This is
228 equivalent to using GBLUP (Strandén and Garrick, 2009). A variety of
229 breeding strategies were considered: (1) selection within a population; (2)
230 selection within a new synthetic breed made from the AL and PL
231 populations; (3) selection of females within a population but allowing
232 selection of males from another population, i.e., genomic introgression by
233 crossbreeding.

234 For the within population selection (strategies 1 and 2), selection of
235 breeding animals was from the current breeding animals and mature calves.
236 There were three types of breeding schemes denoted A, P and AP. Scheme
237 A used the current AL breeding animals, scheme P used the PL individuals,
238 and scheme AP used the combined AL and PL individuals. In AP, the first
239 generation males were selected according to the selection index from the F_1

240 individuals which were the offspring of male AL to female PL or male PL
241 to female AL matings. Because the number of female F_1 individuals was
242 only 2,800, another 2,800 cows were selected from the current AL and PL
243 population cows. Thus, the A and P schemes continued to use 200 males
244 and 2,800 females but the AP scheme doubled the breeding population to
245 400 males and 5,600 females keeping the same selection intensity.

246 Genomic introgression (strategy 3) was used to introduce favourable
247 alleles of a trait from a donor population to a target population (Figure 2). A
248 pure breed donor population was maintained along with the target
249 population. In introgression from the AL to PL population (AiP), a pure AL
250 population was maintained to allow selection of AL sires. Half of the
251 selected breeding sires used in the target population were from the AL
252 population, and the other half were from the current PL population. Thus,
253 after four years, some of the PL population individuals were backcrosses.
254 Note that the candidates for the next generation PL sires were the current PL
255 sires and the mature PL male calves which (in both cases) were not
256 necessarily pure PL breed animals either. The same logic was followed in
257 the introgression breeding scheme from the PL to AL population (PiA).
258 Note that in both introgression schemes selection in the pure line used the
259 same selection index weights as in the target population. However, in
260 practice the AL population is expected to continue its own selection scheme

in AiP. Thus, a genomic introgression scheme named rAiP was simulated where random selection was continued in the AL population.

Simulation and the statistics calculated

Each of the five QMSim simulations gave a set of final AL and PL breeding animals (Figure 1). The final breeding animals were available for the breeding schemes described. Unlike in the QMSim simulation schemes, the subsequent selection of animals in these breeding schemes was based on an index of EBVs by SNP-BLUP. Three alternative indices were used where the standardised EBVs of the adaptation and production traits were weighted differently. The EBVs were standardised trait wise by dividing by trait genetic standard deviations. The adaptation (A2P1) selection index had a weight 2 for adaptation and 1 for production EBVs. The equal weight (A1P1) selection index had equal weights for both of the EBVs. The production (A1P2) selection index weighted 1 for adaptation and 2 for production EBVs.

Two statistics were computed for both traits using the true breeding values of the progeny to breeding animals after the QMSim simulation: adjusted genetic level (G) and genetic change during the last nine simulation years (ΔG). Both statistics are reported in genetic standard deviation of trait. If selection is random or very mild, genetic change ΔG stays at zero. When two traits are studied, selection of one of the traits will change the other as well due to correlated response. Genetic change (ΔG) in the target

283 population was computed as the difference in the mean genetic level over
284 replicates between the final and the first year of the breeding scheme
285 simulation, i.e., genetic change during the last 9 years. Adjusted genetic
286 level (G) was computed as the average difference over replicates in genetic
287 level between the last simulation year in the target population and the
288 control level which was the last generation AL individuals of the QMSim
289 simulation. Genetic level in a simulation year was mean of TBVs of animals
290 born that year in the target population.

291 The adjusted genetic level (G) allowed comparison of the absolute
292 genetic level, which illustrates changes in production and adaptation levels
293 compared to the control level. After the QMSim simulation, the AL and PL
294 lines were on different genetic levels. Consequently, comparisons according
295 to the genetic change would provide a false impression on the short term
296 genetic consequences of selection. Because the control level is the same for
297 all schemes within a replicate, differences in absolute genetic level between
298 the replicates could be corrected. The control level calculated from the last
299 generation AL individuals of QMSim is from the random selection scheme,
300 and is close to the genetic level after HP. Thus, for example, in AiP, the
301 genetic change ΔG has been calculated using the target PL population
302 individuals only, but calculating the adjusted genetic level G we used the
303 genetic level values of the last generation PL population and the last
304 QMSim generation AL population. Note that the difference in G and ΔG

will not be the same in a single population (e.g. P) and introgression schemes (e.g., AiP) because the starting level in ΔG is based on the genetic level of the progeny compared to the selected individuals. In a single population, the parents of progeny come from one of the pure lines but in the introgression schemes the parents come from both of the lines.

The rate of inbreeding per generation (ΔF) describes the risk of a breeding scheme. Inbreeding rate was estimated using pedigree information following the method of Gutierrez et al. (2009). The method used pedigree inbreeding coefficients to calculate individual rates of inbreeding which are averaged to estimate the population inbreeding rate.

Existing and new programs were combined in several Linux shell scripts. The first phase of the script was based on the QMSim software which generated the two breeding lines to be used in the subsequent breeding scheme simulation. Then, each breeding scheme used a script that was a set of programs written in Perl, R (all steps needing random numbers) or Fortran. Existing programs included RelaX2 (Strandén and Vuori, 2006) for effective population size calculation and MiX99 (Strandén and Lidauer, 1999) for breeding value estimation by two trait SNP-BLUP.

Results

Genetic level in the final year of QMSim quantifies initial genetic level in the AL and PL populations before selection using the studied selection indices in the contrasted breeding schemes. In the final year of QMSim, the

327 average differences across replicates in adjusted genetic levels of adaptation
328 and production between the AL and PL populations were 1.10 and -2.99
329 genetic standard deviations of the traits, respectively. Because the difference
330 was positive for adaptation but negative for production, genetically the AL
331 population had higher adaptation but poorer production than in the PL
332 population. The genetic level of the AL population can be considered a
333 control level. Thus, adjusted genetic level values below 0 means that the
334 trait has not reached the same level as in the AL population at the time the
335 breeding scheme was started.

336 Genetic change and level were positive and high for adaptation in all
337 schemes when the selection index A2P1 was used (Table 3). The genomic
338 introgression and synthetic breeding schemes were able to achieve a higher
339 genetic level for adaptation than the within line selection schemes. Genetic
340 progress for adaptation was highest in the introgression scheme from AL to
341 PL (AiP). However, because genetic change for production trait was
342 negative in AiP, this scheme may be unrealistic in practice. In contrast,
343 introgression from PL to AL (PiA) gave high genetic progress for both
344 production and adaptation. Furthermore, the adjusted genetic level achieved
345 for production was high and was moderately high for adaptation. The
346 highest genetic increase in adaptation in the non-introgression schemes was
347 achieved by the combined AP scheme, with genetic progress being higher
348 than in the rAiP and PiA schemes. Similarly, genetic change and adjusted

genetic level were good in production in the AP scheme. The other non-introgression schemes showed low genetic increases in production. However, the P scheme, while giving the highest adjusted genetic level for production gave the lowest for adaptation.

When production and adaptation were weighted equally in the selection index, all schemes were able to make positive genetic change for both production and adaptation (Table 3) except rAiP, which showed a -0.20 change in production. So, in contrast to the A2P1 index selection, introgression from AL to PL (AiP) gave positive genetic change in both of the studied traits. The final genetic level of adaptation was not as high as with A2P1 but the genetic level for production was quite high. The positive genetic progress of adaptation in the AiP scheme was due to applying the same selection index in the donor AL population scheme. When the selection was random in the donor population (rAiP), genetic progress in production was negative. Introgression from PL to AL (PiA) gave lower genetic progress in adaptation than in AiP, but the genetic level reached was still positive. However, both the genetic progress and level for production were very high. Thus, as the relative weight of adaptation decreased in the selection index (from A2P1 to A1P1), genetic change and the level of adaptation decreased. In this context, genomic introgression did not prove to be the best strategy. Instead, continuing selection within the PL population

(P) gave a high genetic increase in adaptation and a high genetic level in production. Likewise, the joint population AP scheme was competitive.

When the selection index A1P2 was used, genetic change and level in adaptation tended to be very low or even negative. Scheme P achieved minor improvement in adaptation, but the final level of adaptation was very low. Scheme A was able to maintain reasonable increases in adaptation although at a low level. The introgression scheme from AL to PL (AiP) gave the highest genetic increase in adaptation with the final genetic level value of adaptation being almost 0. In AiP, the genetic progress in production was lower than in the non-introgression schemes. However, the final genetic level of production was still high due to the high genetic level of production at the beginning of the breeding scheme. Thus, when the index weight for production increased, scheme AiP became more favourable than PiA because genetic change in both production and adaptation was positive.

Performance of the breeding schemes under different selection indices is illustrated in Figures 4 to 7. The P and PiA introgression schemes had the highest genetic change in production (Figure 4). The PiA scheme seemed superior when adaptation was given high weight. The AiP introgression scheme had the highest genetic change in adaptation although different non-introgression schemes were often close (Figure 5). The synthetic AP scheme reached highest genetic level in production but the single population P and

392 the introgression PiA schemes were competitive when production was given
393 higher weight than adaptation (Figure 6). The A and AiP schemes achieved
394 the highest genetic level in adaptation (Figure 7).

395 The introgression schemes from the AL donor population (AiP) used AL
396 sires. The proportion of genes originating from the AL population depended
397 on the selection index (Figure 3). The more adaptation was weighted in the
398 selection index, the higher the influence of the AL population proved to be.
399 The continued use of random selection in the AL population decreased the
400 gene proportion of the original AL population considerably. In the last
401 simulation year, the proportion of the AL population genes in AiP (rAiP)
402 was 58% (39%) using A2P1, 44% (24%) using A1P1 and 34% (19%) using
403 A1P2. These numbers were similarly ranked in AiP and rAiP as the genetic
404 level G for adaptation in Table 3. When introgression was from the PL
405 donor population (PiA), the proportion of AL genes decreased from the
406 original 100% rapidly. In the final simulation year, the proportion of AL
407 genes in PiA was 36% using A2P1, 29% using A1P1 and 27% using A1P2.
408 For A2P1 and A1P1 these numbers were lower than those in AiP. The rate
409 of inbreeding per generation ΔF was highest in the within population
410 schemes A and P (Table 3). The AP scheme showed a lower ΔF and the
411 genomic introgression schemes had an even lower ΔF. These results can be
412 expected because the genomic introgression schemes introduced animals
413 with a lower than average relatedness to the target population, and the

414 combined breeding population AP had double the population size in
415 comparison to the within line selection schemes A and P. Differences in ΔF
416 within schemes using different selection indices were small.

417 ## Discussion

418 It should be anticipated that selection for a trait with high heritability will
419 produce a higher genetic change than selection for a trait with low
420 heritability when all the other conditions are the same. Thus, it should be
421 easier to change a highly heritable trait such as production in this study than
422 a less heritable trait such as adaptation. Consequently, PiA introgression
423 should be preferred over AiP because it should be easier to breed high
424 production into locally adapted animals than the other way around. Indeed,
425 we found that PiA crossbreeding was more successful than AiP, but only
426 when adaptation had a high weight in the selection index. In contrast, when
427 production had a high weight, AiP exceeded PiA because adaptation
428 increased in the former but decreased in the latter. This result is logical
429 because when production is given a high weight in the selection index, PiA
430 is efficient because adaptation is of less importance. However, when
431 production has a high weight, AiP will lead to selection of favourable
432 adaptive genes from the AL line, while maintaining a high genetic level of
433 production.

434 In our study, the synthetic breed scheme AP combined the two divergent
435 lines into one common population. This was the most successful non-

436 introgression strategy in terms of inbreeding rate and genetic change in
437 production. Overall genetic diversity decreased if only one of the original
438 populations was maintained. When the selection index weighted the traits
439 equally, the AP scheme reached similar genetic change in production and
440 higher genetic change in adaptation than the best introgression strategy
441 (PiA). In some cases, the AP scheme may be the only way to conserve
442 genes from both of the populations when either or both of the populations
443 have too low number of animals to make a viable population. Because the
444 potentially positive effects of heterosis were not considered in our study, the
445 AP scheme may show even more positive results. However, heterosis
446 effects would be expected to be lost rapidly in the AP scheme. In contrast,
447 heterosis would contribute for a longer duration in the crossbreeding
448 scheme, where each new generation would results in some heterosis because
449 a separate donor line is maintained.

450 The rate of inbreeding ΔF depends on the effective population size. The
451 synthetic breed scheme AP increased population size by combining two
452 breeds, leading to a lower rate of inbreeding than in the within breed
453 schemes, which is expected. Similarly, genomic introgression schemes
454 increase effective population size by incorporating genetic diversity from
455 another population. However, in the simulated introgression schemes, the
456 donor population remained separate during the simulation and allowed

457 continuous crossbreeding and thus featured a much lower rate of inbreeding
458 than in the AP scheme.

459 Repeated crossbreeding or backcrossing between a locally adapted breed
460 and a more productive breed has been used to breed locally adapted high
461 milk yield cows. For example, in Yakutia, north-eastern Siberia, the locally
462 adapted Yakutian cattle, which tolerate Siberian harsh cold environment,
463 has been crossed with Simmental and Russian Kholmgor breeds to establish
464 the Siberian Simmental and Siberian Kholmogor cattle populations,
465 respectively (Li et al., 2005). These crossbreds have been backcrossed with
466 the native Yakutian cattle to further develop two hybrid cattle populations.
467 Breeding selection can be used for adaptation to local warm climate (e.g.
468 Berman, 2011). In Ethiopia, the aim of crossbreeding native breeds to
469 Holstein is to produce cows with at least 25% native and at most 75%
470 Holstein breed proportions such that production is increased and adaptation
471 is maintained (Negussie et al., 1999). Simulation results in our study support
472 genomic introgression schemes from high production to adaptation (PiA)
473 breeds to be reasonable when local adaptation is of high importance.
474 However, the extent of crossbreeding varies in Africa even to such an extent
475 that there is a risk of introgression into indigenous populations, and
476 subsequent erosion of local genetic resources (Leroy et al., 2016). These
477 risks are supported by our simulation results, where proportion of the AL

478 population genes decreased more rapidly in PiA than they increased in AiP,
479 particularly when production was given a high weight.

480 In practice, any given farm is likely to have either high production or
481 high adaptation animals. This influences the type of introgression preferred
482 and possible. When a farm has high adaptation level animals and increasing
483 production is desired, then the PiA scheme could be used. However, while
484 increasing production and lowering adaptation may provide short term gains
485 it incurs risks that will present themselves later. This is another reason to
486 proceed cautiously using the PiA scheme. The use of AiP introgression is
487 likely to show lower production but has the long-term benefit of adaptation.
488 Climate change has direct and indirect effects on dairy cattle (Nardone et al.,
489 2010; Kantanen et al., 2015). Indirect effects may be apparent earlier via
490 extreme temperature or rainfall affecting feed production, which may
491 require population replacement. The choice may then favour robust cattle
492 that do not require high cost maintenance. The AL population had an
493 adaptation level of zero at the beginning of the breeding scheme
494 simulations. All schemes that showed a positive genetic level in adaptation
495 in the last simulation year achieved the same level as the AL population at
496 the beginning of the simulation. When adaptation was given a high weight,
497 the genetic levels in adaptation were always positive. Even the PL
498 population achieved high adaptation and was able to maintain high
499 production during the simulation time of ten years.

Our simulation assumed a two-year generation interval, and ten years of selection. In practice, the absolute genetic levels and genetic change are unlikely to be this favourable over such a short period. First, we assumed that all animals were genotyped and genomic evaluation is in use. High but less efficient genetic change is likely to be achieved using single-step genomic evaluation (Aguilar et al., 2010; Christensen and Lund, 2010) when only some of the animals, e.g. all bulls, are genotyped. However, a basic assumption for the results to be applicable in practice is that an efficient breeding scheme is in place. Second, both traits in the simulation were assumed to be observed after the animal become mature, i.e. at the age of two years. In practice, first lactation milk yield is at the earliest available at the age of three, while adaptation data are available much later, depending on the defined trait (e.g., longevity). When genomic evaluation is used, it is important that the genotyped reference animals and their progeny have a sufficient number of observations. Thus, the more animals that have been genotyped, the higher the accuracy of genetic evaluation will be, and individual record information is less important. Third, the importance of genomic evaluation extends to the rate of inbreeding as well. When breeding value estimation is based on genomic information instead of pedigree information, using an animal model, selecting animals from different families becomes more likely, which translates to a lower rate of inbreeding. Fourth, the use of young two year old animals to selection allows short

522 generation interval and is feasible due to the use of genomic evaluation.
523 However, calving at the age of 24 months may be too early. If a year in the
524 simulation was extended to be 15 months, then the first calving would be at
525 30 months. This would extend the simulation by 25% from ten to about 13
526 years. Finally, it was assumed that both of the populations had an equally
527 good recording system and that there were always breeding animals
528 available. In practice, conserved local breeds may have too low a number of
529 bulls with breeding values as accurate as in the major breed for an
530 introgression scheme to succeed as well as in the simulations.

531 Our study design is quite unique and genomic introgression simulations
532 including two selected traits have not been presented for dairy cattle.
533 However, our results are similar to those in Ødegård et al. (2009b) where
534 pure breed, synthetic breed and genomic introgression were simulated for
535 fish. In particular, genetic increase in the simulated 5 years in production
536 was higher (6.81 vs. 4.02) in the backcrossed scheme they used, similar to
537 our PiA scheme (3.50 vs. 1.63) than in the pure breed scheme, but the
538 opposite was the case for adaptation. Because each mating in fish produced
539 20 offspring, selection intensity was higher, and the absolute values for
540 genetic change were higher than achieved in our dairy cattle simulation.

541 All schemes featured a fairly low rate of inbreeding (Table 3). The major
542 reason for this result was that even for the within population schemes, the
543 number of selected males was quite high, and selection of females had low

544 intensity. However, the low rate of inbreeding also indicates moderate
545 accuracy of the estimated breeding values. Accuracy of genomic evaluation
546 by SNP-BLUP depends on the number of genotyped reference animals with
547 records. The simulated population had many reference animals but only
548 breeding bulls had a reasonable amount of information (20 daughter
549 records). In practice, the reference animals are likely to have more accurate
550 genomic evaluation which may give larger genetic differences between the
551 schemes and higher rates of inbreeding.

552 We assumed a genetic correlation of -0.3 between adaptation and
553 production. In addition, all loci affecting these traits were assumed to be
554 shared. In practice, this would not be the case. However, this allowed a
555 more realistic simulation than, for example, in Ødegård et al. (2009b) where
556 no pleiotropic effects for QTL were assumed. The use of non-zero genetic
557 correlations made selection to increase both the traits simultaneously more
558 challenging, which also contributed to the lower genetic responses we found
559 than were detected in the previous study of Ødegård et al. (2009b).

560 Changing genetic correlations and heritabilities will affect the absolute
561 values achieved in our study. In particular, if heritability for production
562 traits is higher and for adaptation traits is lower, increasing the genetic level
563 of adaptation becomes more difficult, especially if the genetic correlation is
564 more negative. However, the observed trends between the schemes under
565 different selection index weights and conclusions should remain. For

example, our results confirm the value of conserving genetic resources for the benefit of introgressing their favourable characteristics into commercial breeding programs when the production trait has a high value.

Our simulations show that accelerated progress was achieved most efficiently by genomic introgression from the locally adapted into the production populations when (as is likely to be the case) production had a high weight in the selection index. This approach led to the selection of favourable locally adapted genes, while still maintaining a high level of production. Practical application of this scheme, however, should proceed in caution, especially focusing on locally adapted genes, ensuring that they are preserved in a separate local population. Knowing some or all of the favourable alleles of genes affecting the selected traits and using this information in the genetic evaluation can increase selection accuracy but further research is needed to quantify the change in genetic level and the likely increase in rate of inbreeding.

Acknowledgements

This study is part of ClimGen ("Climate Genomics for Farm Animal Adaptation") project funded by ERA-NET Plus on Climate Smart Agriculture Initiative.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

588 Åby BA, Meuwissen THE (2014). Selection strategies utilizing genetic
589 resources to adapt livestock to climate change. 10th World Congress on
590 Genetics Applied to Livestock Production, August 17-22 2014,
591 Vancouver, Canada, poster 394.
592 Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ (2010).
593 Hot topic: A unified approach to utilize phenotypic, full pedigree, and
594 genomic information for genetic evaluation of Holstein final score. *J*
595 *Dairy Sci*, **93**:743-752.
596 Berman A (2011). Invited review: Are adaptations present to support dairy
597 cattle productivity in warm climates? *J Dairy Sci* **94**:2147-2158.
598 Christensen O, Lund MS (2010). Genomic prediction when some animals
599 are not genotyped. *Genet Sel Evol*, **42**:2.
600 Das R, Sailo L, Verma N, Bharti P, Saikai J, Imtiwati, Kumar R (2016).
601 Impact of heat stress on health and performance of dairy animals: a
602 review. *Vet World* **9**:260-268.
603 FAO (2015). Coping with climate change - the roles of genetic resources for
604 food and agriculture. Rome.
605 Gutierrez JP, Cervantes I, Goyache F (2009). Improving the estimation of
606 realized effective population sizes in farm animals. *J Animal Breed Genet*
607 **126**:327-332.

608 Gaspa G, Veerkamp R, Calus MPL, Windig JJ (2015). Assessment of
609 genomic selection for introgression of polledness into Holstein Friesian
610 cattle by simulation. *Livest Sci* **179**:86-95.
611 Hoffmann I (2010). Climate change and the characterization, breeding and
612 conservation of animal genetic resources. *Anim. Genet.* **41**(Suppl. 1):32-
613 46.
614 Hoffmann I (2013). Adaptation to climate change – exploring the potential
615 of locally adapted breeds. *Animal* **7**:346-362.
616 Kantanen J, Løvendahl P, Strandberg E, Eythorsdottir E, Li M-H, Kettunen-
617 Præbel A, Berg P, Meuwissen T (2015).Utilization of farm animal
618 genetic resources in a changing agro-ecological environment in the
619 Nordic countries. *Frontiers in Genetics* 6: 52.
620 Leroy G, Baumung R, Boettcher P, Scherf B, Hoffman I (2016). Review:
621 Sustainability of crossbreeding in developing countries; definitely not
622 like crossing a meadow…. *Animal* **10**:262-273.
623 Li MH, Nogovitsina E, Ivanova Z, Erhardt G, Vilkki J, Popv R, Ammosov
624 I, Kiselyova T, Kantanen J (2005). Genetic contribution of indigenous
625 Yakutian cattle to two hybrid populations, revealed by microsatellite
626 variation. *Asian-Aust J Anim Sci* **18**:613-619.
627 Mirkena T, Duguma G, Haile A, Tibbo M, Okeyo AM, Wurzinger M,
628 Sölkner J (2010). Genetics of adaptation in domestic fam animals: A
629 review. *Livestock Science* **132**:1-12.

630 Nardone A, Ronchi B, Lacetera N, Bernabucci U (2006). Climatic effects on
631 productive traits in livestock. *Vet. Res. Commun.* **30**(Suppl. 1), 75–81.

632 Negussie E, Brännäng E, Rottmann OJ (1999). Reproductive performance
633 and herd life of dairy cattle at Asela livestock farm, Arsi, Ethiopia. II:
634 Crossbreds with 50, 75 and 87.5% European inheritance. *J Animal Breed*
635 *Genet* **116**:225-234.

636 Niyas PA, Chaidanya K, Shaji S, Sejian V, Bhatta R, Bagtah M, Roa
637 GSLHVP, Kurien EK, Girish V (2015). Apadtatoin of livestock to
638 environmental challenges. *J of Vet Sci & Med Diag* **4**:3.

639 Ødegård J, Sonesson AK, Yazdi MH, Meuwissen THE (2009a).
640 Introgression of a major QTL from an inferior into a superior population
641 using genomic selection. *Genet Sel Evol* **41**:38.

642 Ødegård J, Yazdi MH, Sonesson AK, Meuwissen THE (2009b).
643 Incorporating desirable genetic characteristics from an inferior into a
644 superior population using genomic selection. *Genetics* **181**:737-745.

645 Phocas F, Belloc C, Bidanel J, Delaby L, Dourmad JY, Dumont B, Ezanno
646 P, Fortun-Lamothe L, Foucras G, Frappat B, González-García E, Hazard D,
647 Larzul C, Lubac S, Mignon-Grasteau S, Moreno CR, Tixier-Boichard M,
648 Brochard M (2016). Review: Towards the agroecological management of
649 ruminants, pigs and poultry through the development of sustainable
650 breeding programmes: I-selection goals and criteria. *Animal* **10**:1749-1759.

651 Pritchard T, Coffey M, Mrode R, Wall E (2013a). Genetic parameters for
652 production, health, fertility and longevity traits in dairy cows. *Animal*
653 **7**:34-46.
654 Pritchard T, Coffey M, Mrode R, Wall E (2013b). Understanding the
655 genetics of survival in dairy cows. *J Dairy Sci* **96**:3296-3309.
656 Sargolzaei M, Schenkel FS (2009). QMSim: A large-scale genome
657 simulator for livestock. *Bioinformatics* **25**:680-681.
658 Sasaki O (2013). Estimation of genetic parameters for longevity traits in
659 dairy cattle: A review with focus on the characteristics of analytical
660 models. *Animal Sci J* **84**:449-460.
661 Solberg TR, Sonesson AK, Woolliams JA, Meuwissen THE (2008).
662 Genomic selection using different marker types and densities. *J Animal*
663 *Sci* **86**:2447-2454.
664 Strandén I, Lidauer M (1999). Solving large mixed models using
665 preconditioned conjugate gradient iteration. *J Dairy Sci* **82**:2779-2787.
666 Strandén I, Vuori K (2006). RelaX2: pedigree analysis program. 8th World
667 Congress on Genetics Applied to Livestock Production, August 13-18,
668 2006, Belo Horizonte, MG, Brazil, Volume 27.30.
669 Strandén I, Garrick DJ (2009). Technical note: Derivation of equivalent
670 computing algorithms for genomic predictions and reliabilities of animal
671 merit. *J Dairy Sci* **92**:2971–2975.

672 Visscher PM, Haley CS, Thompson, R (1996). Marker-assisted
673 introgression in backcross breeding programs. *Genetics* **144**:1923-1932.

Figure Legends:

Figure 1.

QMSim simulation scheme.

Figure 2.

Repeated introgression backcrossing scheme of the production line (PL) with the adaptation line (AL) for the first three years. For simplicity of presentation, calves mature here after birth but in simulation after one year.

Figure 3.

Average proportion of AL population genes in cows by simulation year in different introgression schemes from AL to PL (AiP). In the target population selection index, adaptation and production were weighted by ratio 2:1 (black), 1:1 (blue), and 1:2 (red). Selection index in the donor population (AL) was the same as in the target population (solid line) or random (dashed line).

Figure 4. Genetic change of production in genetic standard deviation by selection index. The selection index weighted adaptation and production equally in A1P1, by 2:1 ratio in A2P1 respectively, and by 1:2 ratio in A1P2

694 respectively. The non-introgression schemes A (circle), AP (square) and P
695 (triangle) have solid lines, and the introgression schemes rAiP (cross), AiP
696 (circle), and PiA (triangle) have dashed lines.

697

698 Figure 5. Genetic change of adaptation in genetic standard deviation by
699 selection index. The selection index weighted adaptation and production
700 equally in A1P1, by 2:1 ratio in A2P1 respectively, and by 1:2 ratio in A1P2
701 respectively. The non-introgression schemes A (circle), AP (square) and P
702 (triangle) have solid lines, and the introgression schemes rAiP (cross), AiP
703 (circle), and PiA (triangle) have dashed lines.

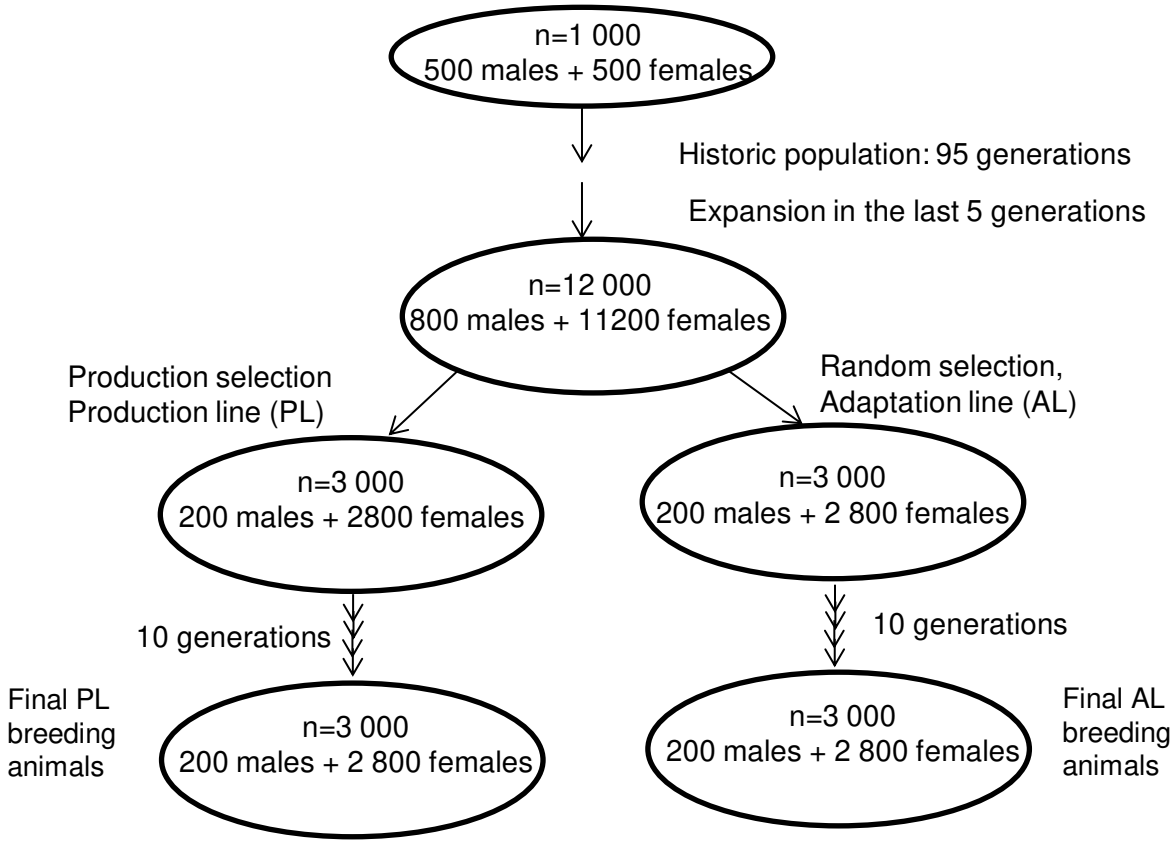
704

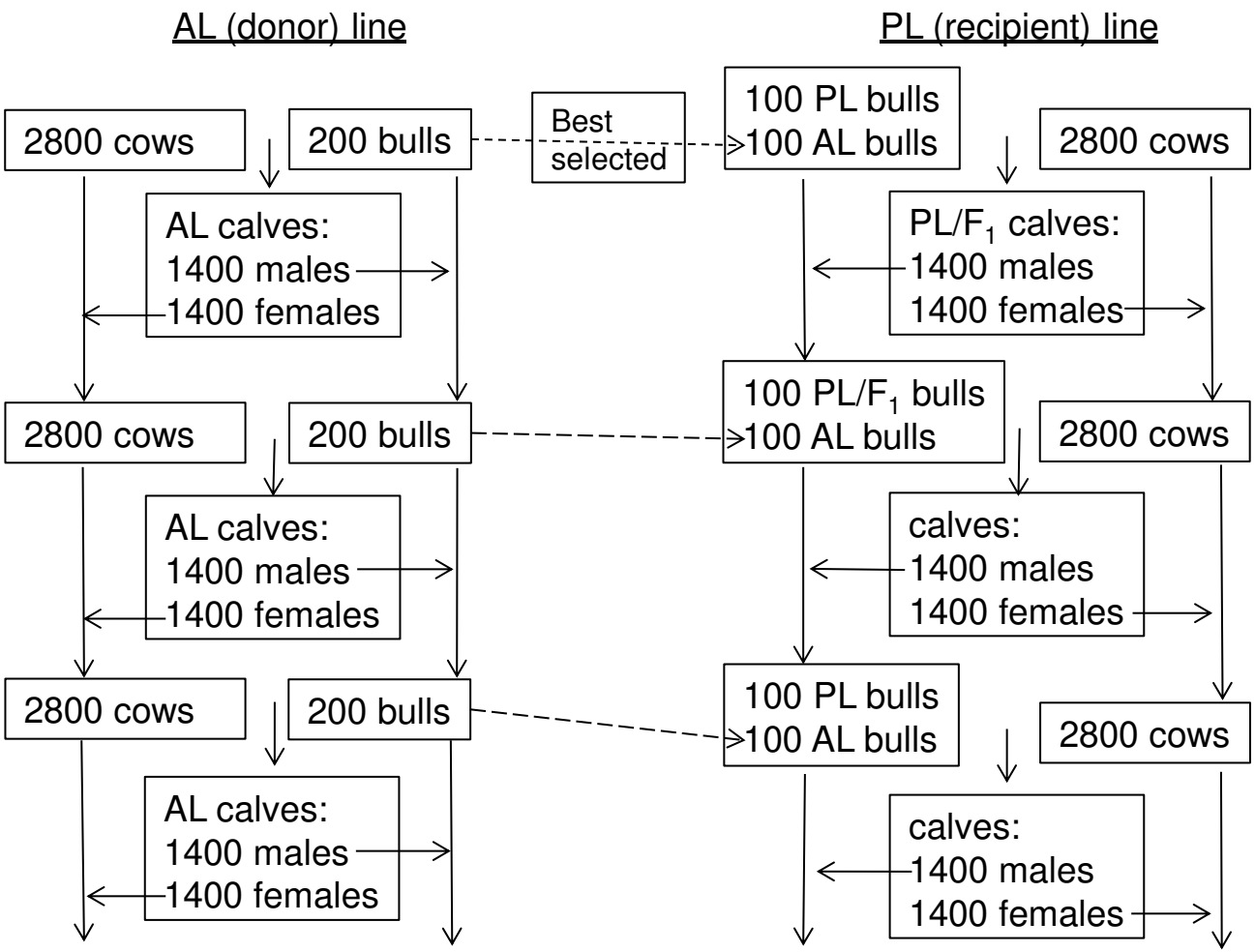
705 Figure 6. Genetic level of production in genetic standard deviation by
706 selection index. The selection index weighted adaptation and production
707 equally in A1P1, by 2:1 ratio in A2P1 respectively, and by 1:2 ratio in A1P2
708 respectively. The non-introgression schemes A (circle), AP (square) and P
709 (triangle) have solid lines, and the introgression schemes rAiP (cross), AiP
710 (circle), and PiA (triangle) have dashed lines.

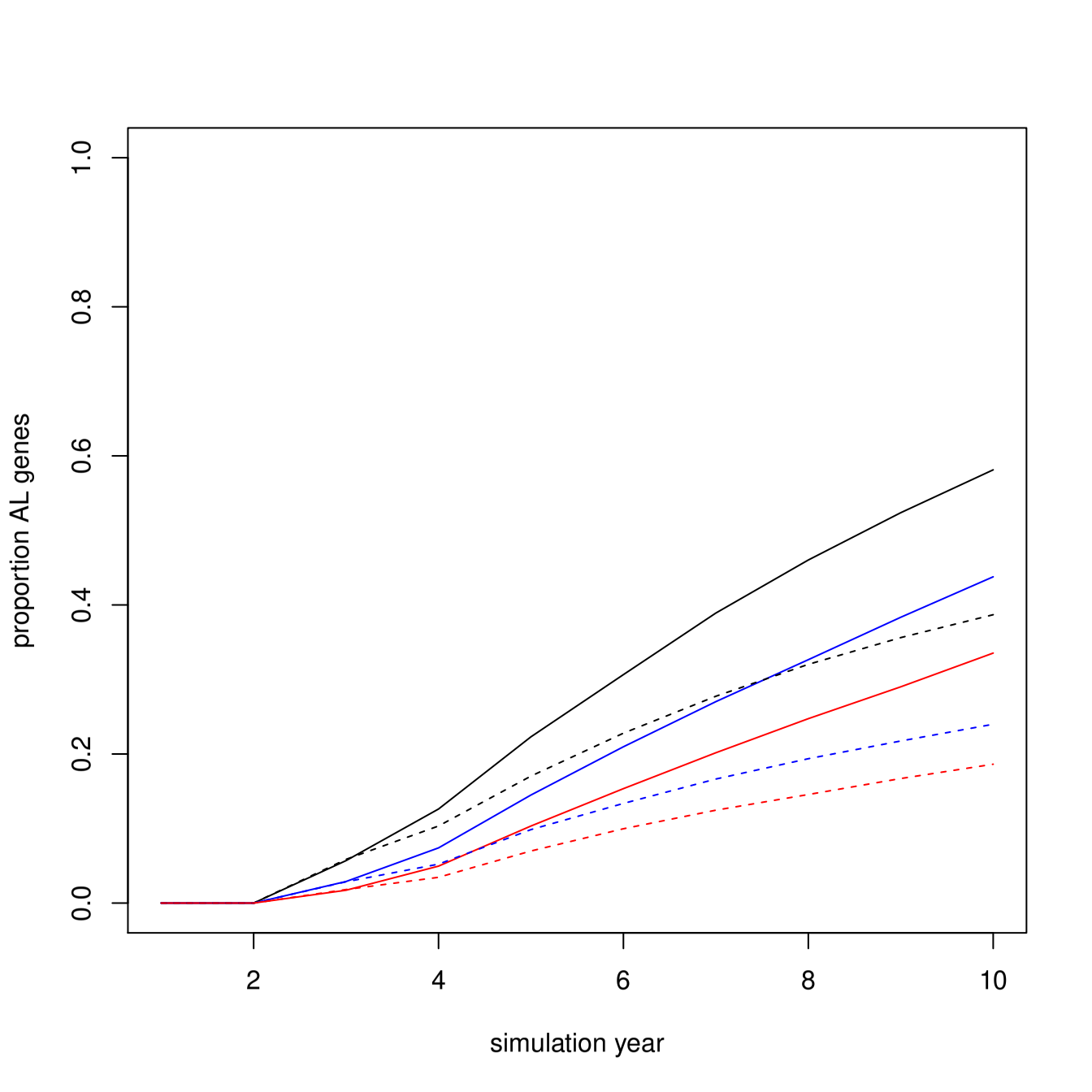
711

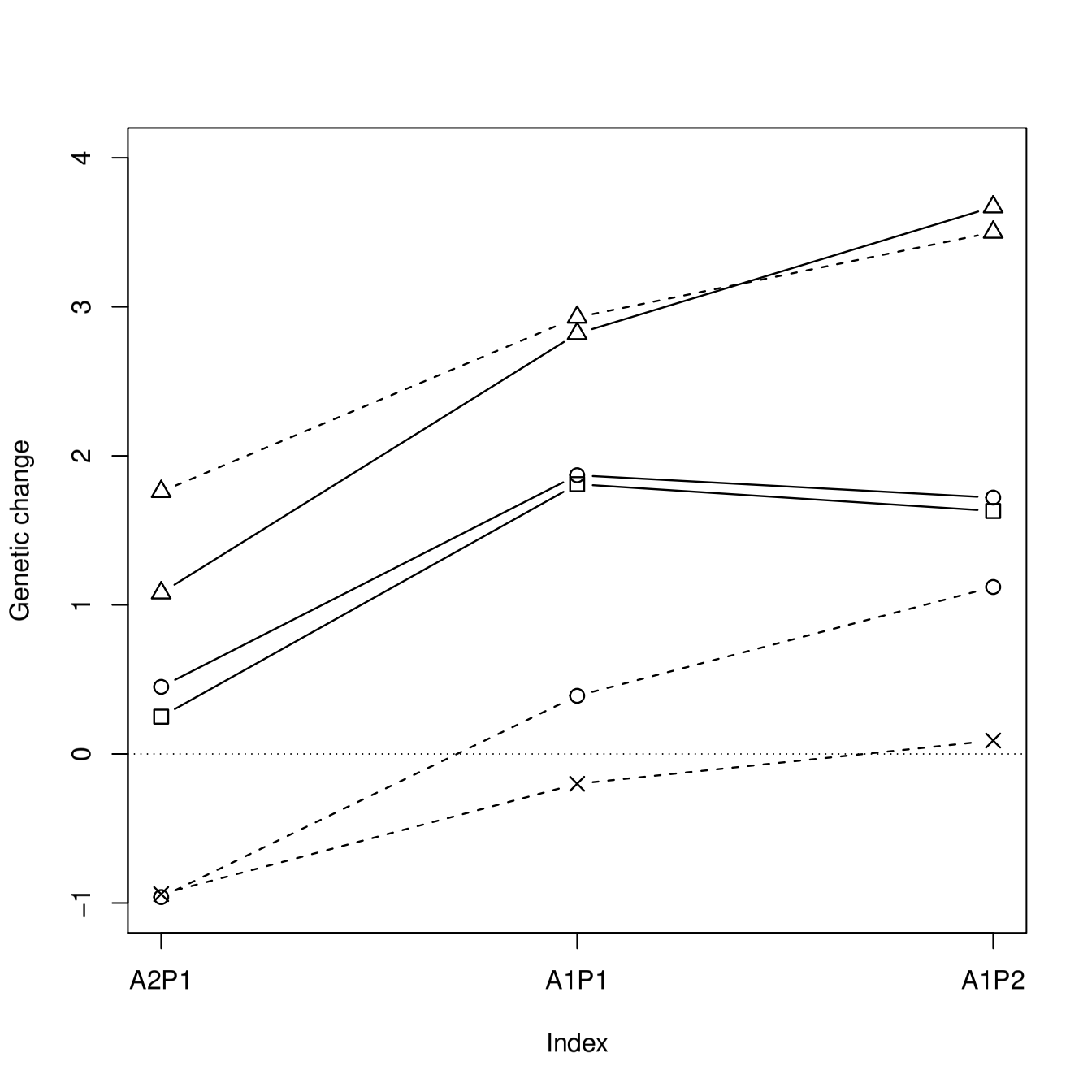
712 Figure 7. Genetic level of adaptation in genetic standard deviation by
713 selection index. The selection index weighted adaptation and production

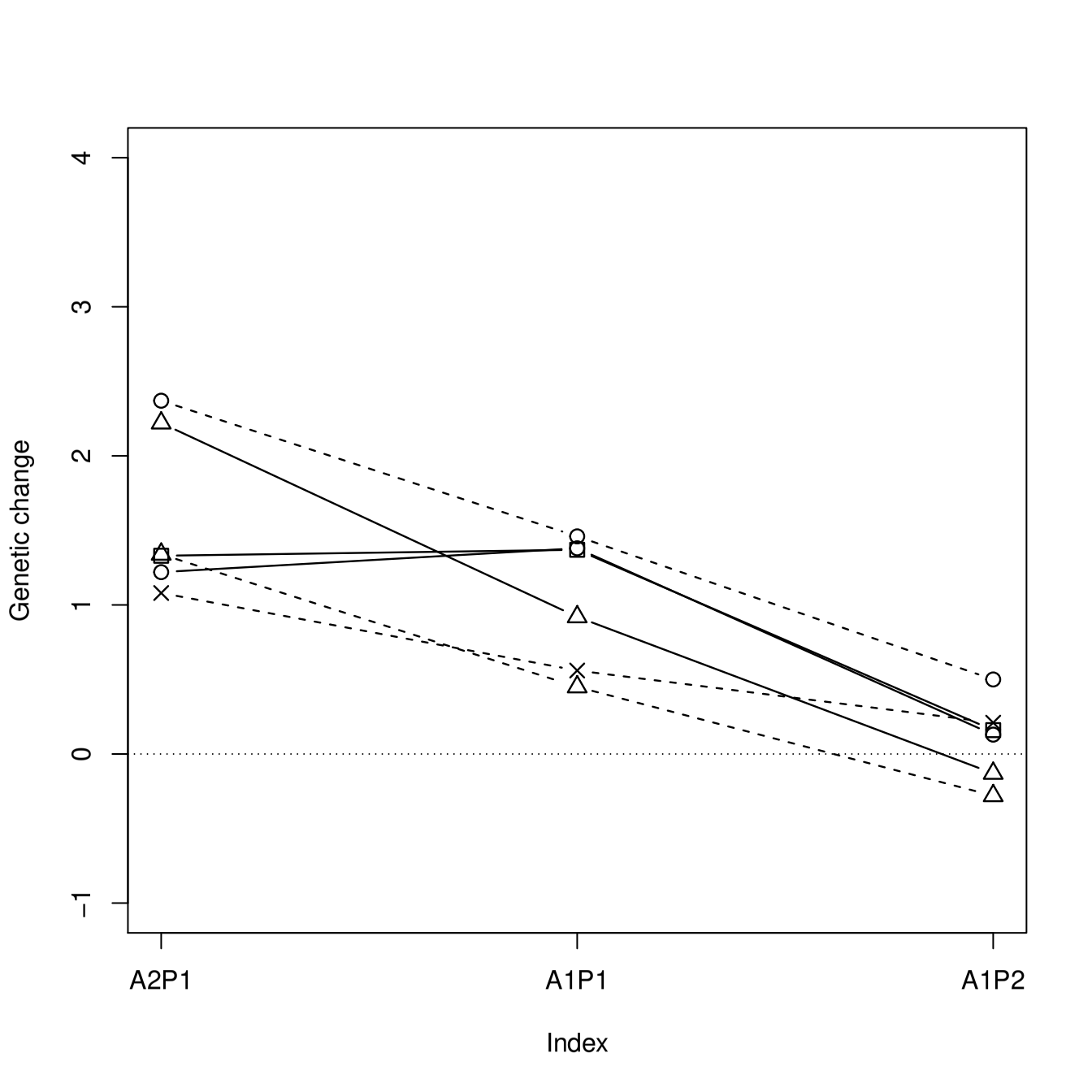
equally in A1P1, by 2:1 ratio in A2P1 respectively, and by 1:2 ratio in A1P2 respectively. The non-introgression schemes A (circle), AP (square) and P (triangle) have solid lines, and the introgression schemes rAiP (cross), AiP (circle), and PiA (triangle) have dashed lines.

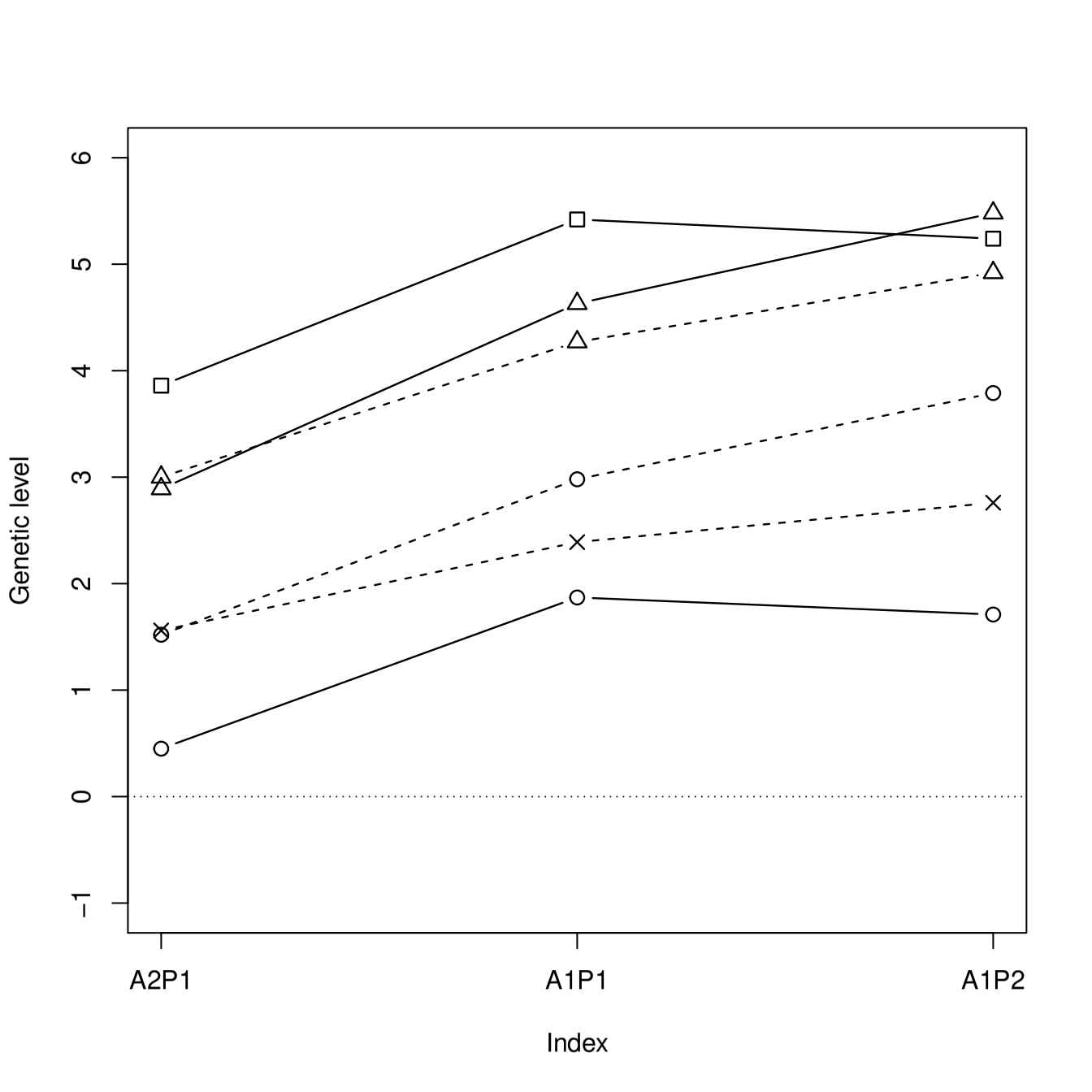

n=1 000
500 males + 500 females
Historic population: 95 generations
Expansion in the last 5 generations
n=12 000
800 males + 11200 females
Production selection
Production line (PL)
Random selection,
Adaptation line (AL)
n=3 000
200 males + 2800 females
n=3 000
200 males + 2 800 females
10 generations
10 generations
Final PL
breeding
animals
n=3 000
200 males + 2 800 females
n=3 000
200 males + 2 800 females
Final AL
breeding
animals


AL (donor) line
PL (recipient) line
2800 cows
200 bulls
Best selected
100 PL bulls
100 AL bulls
2800 cows
AL calves:
1400 males
1400 females
PL/F1 calves:
1400 males
1400 females
2800 cows
200 bulls
100 PL/F1 bulls
100 AL bulls
2800 cows
AL calves:
1400 males
1400 females
calves:
1400 males
1400 females
2800 cows
200 bulls
100 PL bulls
100 AL bulls
2800 cows
AL calves:
1400 males
1400 females
calves:
1400 males
1400 females


proportion AL genes
0.0
0.2
0.4
0.6
0.8
1.0
2
4
6
8
10
simulation year


Genetic change
4
3
2
1
0
−1
A2P1
A1P1
A1P2
Index


Genetic change
4
3
2
1
0
−1
A2P1
A1P1
A1P2
Index


Genetic level
6
5
4
3
2
1
0
−1
A2P1
A1P1
A1P2
Index

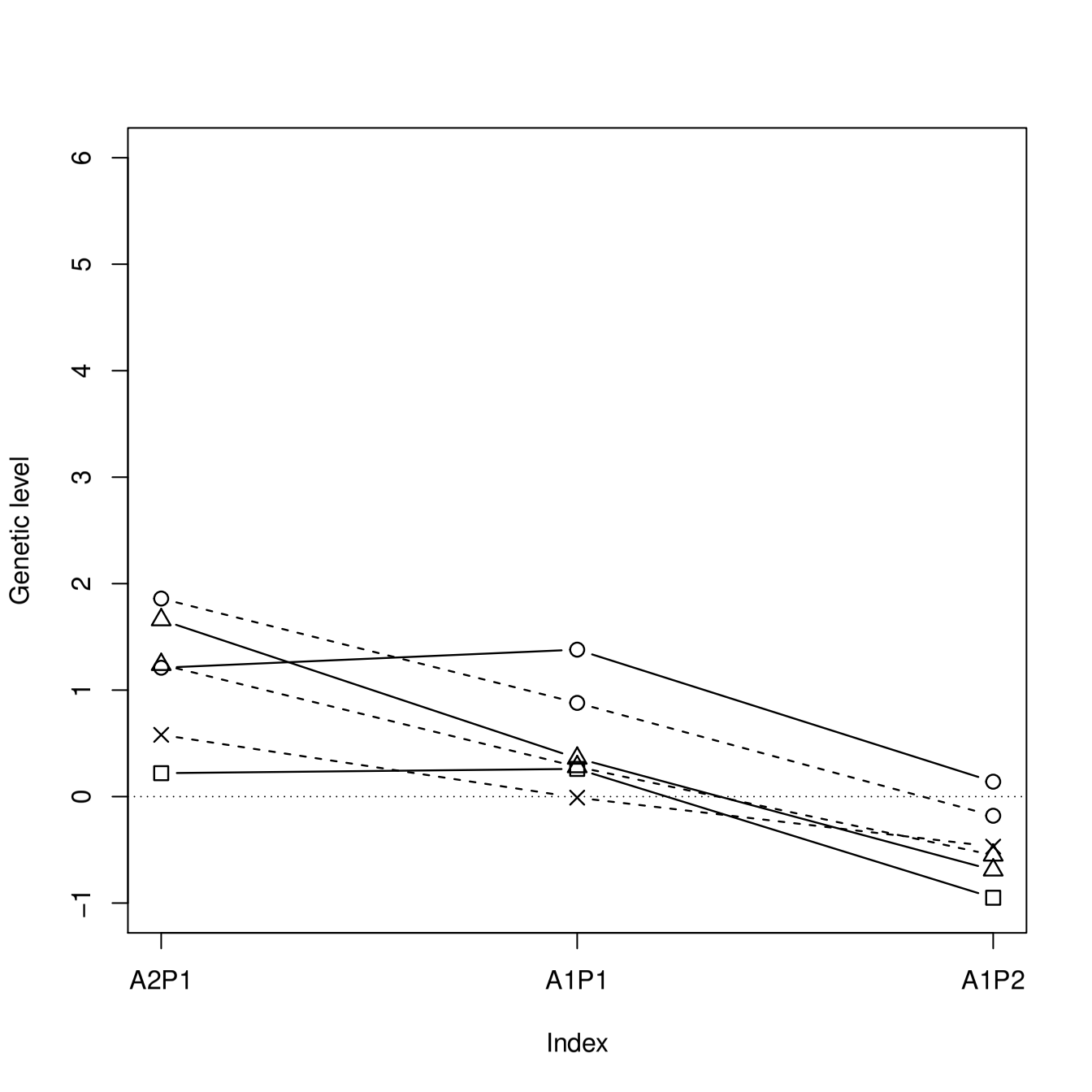

Genetic level
6
5
4
3
2
1
0
−1
A2P1
A1P1
A1P2
Index

Table 1. Parameters to simulation.

Parameter	Number
N breeding males	200
N breeding females	2,800
N progeny per mating	1
N females per male	14
h^2, adaptation	0.1
h^2, production	0.3
genetic correlation	-0.3
N chromosomes	30
Chromosome length	100 cM
N QTL simulated	900
N markers	54,000

Table 2. Steps in one simulation year after the QMSim simulation. Each step was a program in a script. In steps 1 to 5, current breeding animals are mated to produce calves. In steps 6 to 12, previous year calves mature and are included in genetic evaluation. These mature animals and current breeding animals are selection candidates for a new set of breeding animals.

Step	Operation
1)	Search genotypes of breeding animals from all animals
2)	Make random mating pairs to the breeding animals in 1)
3)	Generate random recombination positions for the mating pairs in 2)
4)	Offspring genotypes by mating pairs in 2) using recombination positions in 3)
5)	Calculate true breeding values to the new genotypes in 4)
6)	Generate phenotypes to the maturing cows
7)	Generate pseudo phenotypes to the maturing bulls
8)	Extract all phenotype data for genetic evaluation by SNP-BLUP
9)	Extract genotypes for animals in 8)
10)	Calculate estimated breeding values by SNP-BLUP using data from 8) and 9)
11)	Calculate index (A1P1, A2P1 or A1P2) using estimated breeding values
12)	Select new breeding animals from the current breeding and the mature animals

6

Table 3. Genetic change in last nine years (ΔG) and genetic level in the last year (G) within scheme measured in genetic standard deviation units of the trait (production or adaptation), and rate of inbreeding per generation (ΔF) in percentages with standard error (SE) over five replicates. The selection index weighted adaptation and production equally in A1P1, by 2:1 ratio in A2P1 respectively, and by 1:2 ratio in A1P2 respectively.

5

		Production		Adaptation		
Scheme*	Index	ΔG (SE)	G (SE)	ΔG (SE)	G (SE)	ΔF
A	A2P1	0.45 (0.05)	0.45 (0.05)	1.22 (0.04)	1.21 (0.03)	0.12 (0.004)
P	A2P1	0.25 (0.02)	3.86 (0.02)	1.33 (0.06)	0.22 (0.08)	0.17 (0.011)
AP	A2P1	1.08 (0.10)	2.89 (0.10)	2.22 (0.03)	1.66 (0.01)	0.07 (0.004)
rAiP	A2P1	-0.94 (0.06)	1.56 (0.05)	1.08 (0.05)	0.58 (0.03)	0.03 (0.002)
AiP	A2P1	-0.96 (0.09)	1.52 (0.07)	2.37 (0.04)	1.86 (0.03)	0.05 (0.004)
PiA	A2P1	1.76 (0.08)	3.00 (0.09)	1.34 (0.05)	1.24 (0.06)	0.04 (0.001)
A	A1P1	1.87 (0.06)	1.87 (0.06)	1.38 (0.05)	1.38 (0.05)	0.13 (0.006)
P	A1P1	1.81 (0.07)	5.42 (0.07)	1.37 (0.05)	0.26 (0.06)	0.20 (0.007)
AP	A1P1	2.82 (0.03)	4.63 (0.04)	0.92 (0.05)	0.36 (0.07)	0.09 (0.007)
rAiP	A1P1	-0.20 (0.04)	2.39 (0.01)	0.56 (0.03)	-0.01 (0.06)	0.04 (0.001)
AiP	A1P1	0.39 (0.06)	2.98 (0.07)	1.46 (0.06)	0.88 (0.07)	0.05 (0.003)
PiA	A1P1	2.93 (0.09)	4.27 (0.10)	0.45 (0.10)	0.28 (0.12)	0.04 (0.003)
A	A1P2	1.72 (0.08)	1.71 (0.07)	0.13 (0.04)	0.14 (0.04)	0.11 (0.009)
P	A1P2	1.63 (0.06)	5.24 (0.07)	0.16 (0.04)	-0.95 (0.10)	0.19 (0.007)
AP	A1P2	3.67 (0.04)	5.48 (0.05)	-0.13 (0.07)	-0.69 (0.10)	0.10 (0.005)
rAiP	A1P2	0.09 (0.05)	2.76 (0.03)	0.21 (0.03)	-0.47 (0.06)	0.04 (0.001)
AiP	A1P2	1.12 (0.04)	3.79 (0.04)	0.50 (0.03)	-0.18 (0.07)	0.05 (0.003)
PiA	A1P2	3.50 (0.05)	4.92 (0.07)	-0.28 (0.09)	-0.55 (0.12)	0.04 (0.003)

6

*The A, P, and AP schemes used only adaptation line (AL), production line (PL), or combined AL and PL synthetic line individuals, respectively. The AiP scheme used introgression from AL to PL, and the PiA scheme used introgression from PL to AL. In the rAiP scheme, there was no selection in the AL donor line used in introgression.