A simple strategy for managing many recessive disorders in a dairy cattle breeding program

John B. Cole

Animal Genomics and Improvement Laboratory, Agricultural Research Service,
United States Department of Agriculture, Beltsville, MD, USA

Email address:
john.cole@ars.usda.gov

Abstract

Background

High-density single nucleotide polymorphism genotypes have recently been used to identify a number of novel recessives that adversely affect fertility in dairy cattle, as well as to track other conditions such as red coat color and polled. Many current methods for mate allocation fail to consider that information, and it will be increasingly difficult to manage matings when a large number of recessives must be accounted for.

Methods

A modified version of Pryce's scheme that constrains inbreeding and accounts for the economic effects of Mendelian disorders on overall economic merit was developed and compared with random mating, truncation selection, and Pryce's method of constraining genomic inbreeding for several different scenarios, including one group of 6 hypothetical alleles and a second group of 12 recessives currently segregating in the US Holstein population.

Results

Pryce's method and the modified Pryce's method showed similar ability to reduce allele frequency, particularly for loci with frequencies greater than 0.30. The modified Pryce's method may outperform Pryce's method for low-frequency alleles with small economic values. Cumulative genetic gain for the selection objective was slightly higher using Pryce's method, but rates of inbreeding were similar across methods.

Conclusions

The proposed method appears to reduce minor allele frequencies for recessives with low frequencies faster than other methods, and can be used to maintain or increase the frequency of desirable recessives. It can easily be implemented in software for mate

allocation, and the code used in this study is freely available as a reference implementation.

Keywords: dairy cattle, genetic selection, mating programs, recessive disorders

Background

Recessive disorders have been present in livestock populations since modern animal breeding programs began, and hundreds are known to exist [1]. In the past, test matings were used to identify recessive disorders [2], but most recessives were identified after the carrier bull sired many daughters and had sons in AI (e.g., bovine leukocyte adhesion deficiency [3], complex vertebral malformation [4], and deficiency of uridine monophosphate synthase [5]). It also is possible for novel recessives to be spread through a population by popular bulls before routine screening is possible because such defects were not directly observable, such as occurred with Jersey haplotype 1 [6].

Several authors have proposed methods for including QTL information in breeding programs. Many approaches focus on the calculation of the additive genetic value of a QTL which is combined with other information using a selection index approach [7-10]. Shepherd and Kinghorn [11] have described how QTL information could be included in a look-ahead mate selection scheme, and they have suggested that it could be incorporated into a comprehensive mating service, such as Total Genetic Resource Management™ ([12]; http://www.xprime.com.au/products/tgrm/index.html) once efficient algorithms have been developed. Li et al. [13,14] reported that the use of QTL genotypes provides more benefit when utilized in mate selection rather than index selection for a variety of modes of inheritance under several breeding structures. Recently, Van Eenennaam and Kinghorn [15] extended the MateSel program [16] to permit selection against the total number of lethal alleles and recessive lethal genotypes.

Genomic tools have enabled the detection of many new recessives which have deleterious effects on fertility [17], many of which have effects early in gestation and could not previously be distinguished from failed breedings. As the number of recessives continues to grow, new tools are needed to consider that information when making mating decisions. However, many mate allocation tools do not consider carrier status when bulls and cows are paired, and few make use of DNA marker or haplotype information that is increasingly available for bulls and cows. When there are only a few recessives in a population it is easy to monitor individuals to avoid carrier-to-carrier matings, but that is considerably more difficult, or even impossible, when there are many harmful defects segregating in a population.

Pryce et al. [18] recently proposed a simple method for controlling the rate of increase in genomic inbreeding by penalizing parent averages (PA) for matings that produce inbred offspring. After PA are adjusted, the bull that will produce the highest PA when mated to a cow is selected in a sequential manner, and the number of matings permitted for each bull is constrained to prevent one bull from being mated to all cows. This method is straightforward to program, and effectively constrains genomic inbreeding at reasonable levels. The objectives of this research were to extend Pryce's method to include information about recessives, and to examine its use in simultaneously accounting for a large number of Mendelian disorders when allocating mates in dairy cattle breeding schemes by means of computer simulation. Managing genetic defects is a tradeoff between avoiding carrier matings in the short term and eliminating defects in the long run, so the simulation model will examine long-term changes in the population.

Methods

Base population

Base population cows had true breeding values (**TBV**) for lifetime net merit (**NM$**) that were randomly sampled from a normal distribution with a mean of $0 and a standard deviation of $200, which is similar to the genetic SD of lifetime net merit [19]. Bull TBV were sampled from a normal distribution with a mean of $300 (+1.5 genetic SD of NM$) and a standard deviation of $200. An animal's carrier status for each recessive was determined by randomly sampling sire and dam alleles using the minor allele frequencies (**MAF**) shown in Table 1. Recessives were assumed to be independent of one another, as though each locus was located on a different chromosome. A sex ratio of 0.5 was used, and base population animals were assigned a birth year from -9 to 0 (bulls) or -4 to 0 (cows) by sampling from a uniform distribution.

The base population in each scenario included 350 bulls and 35,000 cows distributed over 200 herds, and the population was permitted to grow to a maximum of 500 bulls and 100,000 cows over the 20 generations simulated. Bulls were permitted a maximum of 5,000 matings per year, and in the truncation selection scheme described later in this section only the top 10% of bulls based on TBV were retained for use as mates.

Descendants

The TBV for new calves were created by taking the parent average (**PA**) and adding a Mendelian sampling term:

$$\mathrm{TBV}_{\mathrm{calf}} = 0.5(\mathrm{TBV}_{\mathrm{sire}} + \mathrm{TBV}_{\mathrm{dam}}) + \mathrm{MS}$$

where $\mathrm{TBV}_{\mathrm{calf}}$, $\mathrm{TBV}_{\mathrm{sire}}$, and $\mathrm{TBV}_{\mathrm{dam}}$ are the TBV of the calf, its sire, and its dam, respectively. The Mendelian sampling term, MS, was drawn from a normal distribution with a mean of 0 and a variance of $\frac{1}{2}\left[1-\frac{1}{2}(f_S+f_D)\right]\sigma_a^2$, where f_S and f_D are coefficients of sire and dam inbreeding, respectively, and σ_a^2 is the additive genetic variance of NM\$ (\$40,000). Sex was assigned randomly with a 50:50 sex ratio. For each recessive in the scenario, an allele was sampled at random from each parent and used to construct the progeny genotype. If the recessive was lethal, an affected (aa) calf was created and marked as dead. Calves were born in the same herd as their dams, and cows did not move between herds. Allele frequencies were updated each generation by counting alleles.

Mating schemes

Four systems of mating, referred to hereafter as schemes, were simulated: random mating, truncation selection, the scheme proposed by Pryce et al. [18], and a modified version of Pryce's scheme that accounts for recessive alleles. In the random mating scheme, bulls were mated randomly to cows, with a parameter in the simulation used to limit the maximum number of matings permitted for each bull (5,000). In the truncation selection scheme, the top 10% of the bulls, based on TBV, were randomly mated to the cow population. Both lethal (e.g., DUMPS) and non-lethal (e.g., red coat color) recessives were included in the simulations.

In the Pryce scheme, matings were assigned as follows. For each herd, 20% of the bulls were randomly selected from the list of live bulls to simulate different groups of bulls available on different farms, and the top 50 bulls from that group were selected as herd sires based on TBV. This strategy is similar to that used by Pryce et al. [18]

for cows and bulls, but only sires were selected in this manner because cows and their offspring were assigned to fixed herds where they remained until death. A matrix of parent averages, $\mathbf{B_0}$, was then constructed with rows corresponding to bulls and columns corresponding to cows. The elements of $\mathbf{B_0}$ were computed as:

$$\mathbf{B}_{ij} = 0.5(\mathrm{TBV}_i + \mathrm{TBV}_j) - \lambda F_{ij}$$

where TBV_i is the TBV for NM$ of bull *i*, TBV_j is the TBV for NM$ of cow *j*, λ is the inbreeding depression ($) associated with a 1% increase in inbreeding, and F_{ij} is the pedigree coefficient of inbreeding of the calf resulting from mating bull *i* to cow *j*. Recessive genotypes are simulated without error, and it was necessary only to simulate genotypes for recessives because pedigrees are free of errors. The regression coefficient of NM$ on inbreeding (λ) was computed as the weighted average of inbreeding effects [20] on the traits in the index. This is similar to the $23.11 used by Weigel and Lin [21].

In the fourth scheme, recessives were accounted for by adjusting the elements of $\mathbf{B_0}$ to account for the recessives carried by the parents as:

$$B'_{ij} = B_{ij} - \sum_{r=1}^{n_r} P\,(aa)_r \times v_r$$

where n_r is the number of recessives in a scenario, $\mathbf{B}_{ij}'$ is the PA adjusted for all recessives in a scenario, $P(aa)_r$ is the probability of producing an affected calf for recessive *r*, and v_r is the economic value of recessive *r*. P(aa) will be either 0.25, for a mating of two carriers, 0.5, for a mating of an affected animal with a carrier, or 1, for a mating of two affected animals. Non-lethal recessives were entered into the simulation with an economic value of either 0 (in the case of recessives with no effect on lifetime profitability) or a negative number (which increases the PA for economically desirable recessives such as polled). The recessives used in each scenario are described in Table 1, which includes the minor allele frequency and the economic value assigned to each. For each recessive there is a correlation of F_{ij} with P(aa) that will result in some double-counting of the economic impact of each locus, and this may produce suboptimal rates of genetic gain. The relationship of these two quantities will be discussed in detail in the Results section.

Once the matrix of PA (**B** or **B'**, depending on the scenario) is constructed, a matrix of matings, **M**, is used to allocate bulls to cows. An element, $\mathbf{M}_{ij}$, is set to 1 if the corresponding value of $\mathbf{B}_{ij}$ is the greatest value in column *j* (that bull produces the largest PA of any bull available for mating to cow *j*), and all of the other elements of column *j* are set to 0. If the sum of the elements of row *i* is less than the maximum number of permitted matings for that bull then the mating can be allocated. Otherwise, the bull with the next-highest value of $\mathbf{B}_{ij}$ in the column is selected, and so-on, until each column has one and only one element in it with a value of 1. This approach overestimates genetic progress because it assumes equal selection accuracy among male and female breeding values, when in practice selection accuracy is lower in females, but it should permit a reasonable comparison of the Pryce and modified Pryce algorithms. All animals are assumed to be genotyped so that recessive status is known and reliabilities are similar for most animals in the population, unlike a traditional progeny test scenario in which cows and bulls have substantially different reliabilities.

Each step in the simulation represents 1 year of calendar time. New animals are born at the beginning of each year, affected calves die, and surviving animals are.

Generations overlapped, and bulls and cows could have offspring in multiple years. Bulls were culled first for age, with a maximum age of 10 years, and then on TBV (lowest-ranking animals culled first) to maintain a maximum population size. Cows were first culled for age, with a maximum age of 5 years. After age-related culling, animals were culled involuntarily. Finally, cows were culled at random to maintain population size, if necessary. Animals were not culled based on carrier status, and cows were not culled due to abortions or stillbirths.

Recessive scenarios

Several scenarios were used to characterize the performance of the proposed method, where the term scenario is used to refer to one or more recessives studied together.

Economic values. Each recessive was assigned an economic value based on the occurrence of embryonic or foetal loss during pregnancy (for lethals), or literature values for non-lethal conditions such as red coat color and horned status. Holstein haplotypes 1 through 5 (**HH1–HH5**) occur early in pregnancy, as does deficiency of uridine monophosphate synthase (**DUMPS**), and they were assigned a value of $40 based on reproductive costs included in the 2014 revision of the NM$ index [19]. Brachyspina and mulefoot result in stillbirths or calves that do not survive to adulthood, and they were assigned relatively high costs of $150, although actual losses could be higher. Complex vertebral malformation (**CVM**) results in late-term abortions, so a value intermediate to that of the Holstein fertility haplotypes and brachyspina/mulefoot was used. The high-cost scenario used 3 times the cost in the base scenario to assess the sensitivity of results to economic values. For the hypothetical recessives, values of either 0.10 ($20) or 1 ($200) genetic standard deviations of NM$ were used.

Holstein recessives. Twelve recessives currently segregating in the US Holstein population were grouped together in order to determine how the modified Pryce method will perform in a commercial livestock population: bovine leukocyte adhesion deficiency (**BLAD**), brachyspina, CVM, DUMPS, HH1–HH5, horned, mulefoot, and red coat color). Two scenarios that used the 12 Holstein recessives, but which differed in the economic value assigned to each locus, were used to determine the sensitivity of matings to different prices. In the normal scenario, prices were assigned based on the effect of the recessive and the timing of occurrence, as described above. In the high-cost scenario the prices used for the normal scenario were multiplied by 3. Allele frequencies for the 12 recessives were taken from [22].

Hypothetical recessives. The effect of initial allele frequency on response to selection under each strategy was examined using six scenarios, each of which included a single locus at low (0.01), medium (0.50), or high (0.90) frequency with either a low ($20) or high ($200) cost. In addition, a seventh scenario that included all of the hypothetical loci was examined.

Horned and other high-frequency non-lethal recessives. Not every recessive in a livestock population is lethal to homozygotes, one example being the horned locus in cattle. Because the horned condition in cattle is due to the action of a recessive allele [23], although it has a very high frequency, it was included in the simulation in place of polled with an allele frequency of 1 - 0.0071 = 0.9929. Based in part on the work of Widmar et al. [24], who calculated average expected costs for dehorning and polled genetics of $11.79 and $10.73, respectively, a value of $40 was assumed for horned to also account for breeders' preferences and premium marketing opportunities. Recall that a positive value reduces the PA in the modified Pryce method, resulting in this case in a lower frequency of horned.

Many recessives. Scenarios including 100 and 1,000 recessives were run in order to examine the relationship of inbreeding and the sum of the P(aa) terms using the modified Pryce method. Minor allele frequencies were sampled from a uniform distribution on the interval [0.01, 0.10], and the economic values from a random uniform distribution on the interval [-\$10, -\$50]. Correlations of the two terms were calculated, and separate regressions were computed for matings selected from among the available choices and those not selected.

Analysis

Results were averaged over each of the 10 replicates for each scenario. Observed changes in allele frequency were compared against expectations, where the expected allele frequency in each generation for lethals was calculated using an equation derived from Van Doormaal and Kistemaker [25]:

$$p_t = \frac{p_{t-1}^2 + p_{t-1}q_{t-1}}{2p_{t-1}^2 + p_{t-1}q_{t-1}}$$

$$q_t = \frac{p_{t-1}q_{t-1}}{2p_{t-1}^2 + p_{t-1}q_{t-1}}$$

where p_t is the frequency of the major allele at time *t*, q_t is the frequency of the minor allele at time *t*, and *t* is the time in years (ranging from 1 to 20). The minor allele frequency at time 0 was the value used in each scenario for each recessive, and the major allele frequency was calculated by subtracting the minor allele frequency from 1. Expected allele frequencies for non-lethals was calculated based on Hardy-Weinberg proportions [26] as:

$$p_t = p_{t-1}^2 + p_{t-1}q_{t-1}$$
$$q_t = q_{t-1}^2 + p_{t-1}q_{t-1}$$

For each recessive in each scenario, as well as for the expected frequencies, allele frequencies were regressed on generation using a linear model as implemented in the Python module Statsmodels version 0.5.0 ([27]; http://statsmodels.sourceforge.net/). For a given recessive, the slopes were extracted from the regression results and a two-sample *t*-test assuming unequal variances was used to compare the coefficients against each other, as well as against the expected frequencies. A Bonferroni adjustment was used to correct for multiple comparisons.

Computing environment

All simulations were performed on a Pogo Linux Atlas 1205 (Pogo Linux, Inc., Redmond, WA) computer with an 8-core AMD Opteron 6328 processor with a clock speed of 3.2 GHz, 64 GB of DDR3 1600 MHz RAM, and 64-bit CentOS Linux EL6 (Red Hat, Inc., Raleigh, NC), or a Thinkmate RAX QS6-4210 (Thinkmate, Inc., Waltham, MA) workstation with four 12-core AMD Opteron 6344 processors with a clock speed of 2.6 GHz, 256 GB of DDR3 1600 MHz RAM, and CentOS Linux EL7. Data analysis and visualization were performed on a MacBook Pro with two Intel Core i7 processors running at 2.9 GHz, 8 GB of DDR3 1600 MHz RAM, and Mac OS X 10.7.5 (Apple Inc., Cupertino, CA).

Computation time for the random mating scheme averaged 193 minutes per replicate in a one-recessive scenario (high frequency, high cost) and 215 minutes in a 12-recessive scenario (Holstein recessives). Considerably less time was required for the truncation selection scheme, averaging 34 minutes in the one-recessive scenario and 38 minutes in the Holstein scenario. The time needed for the Pryce and modified Pryce schemes averaged 206 minutes and 232 minutes in the one-recessive scenario, and 203 and 279 minutes in the Holstein scenario. These two schemes require the allocation of large arrays and the creation of large output files that are not part of the random mating or truncation selection schemes. If matings are done within herd, the memory used for 1 herd can be reused for the next to keep memory requirements low. The time required for processing 1 generation rather than 20 should be very reasonable.

Results and discussion

Holstein recessives

Normal costs. Observed allele frequency changes for 11 of the 12 recessives from the four mating schemes are shown in Figure 1. Horned is not shown because the allele frequency remained above 99% in all 4 schemes, and its inclusion in the plot obscured the changes in alleles at low frequency. The frequency of the 10 lethals generally decreased over time in all scenarios. The frequencies of HH1 and HH3 decreased significantly faster ($P < 0.05$ after a Bonferroni correction) under Pryce's method than under the modified Pryce's method. The rate of change in allele frequencies was similar under the Pryce's (Figure 2) and modified Pryce's (data not shown) schemes. An advantage of the modified Pryce approach is that it maintains the frequency of desirable recessives, such as red coat color, in the population. In the Pryce scheme, the frequency of red decreased over time because there is no mechanism in that scenario to balance undesirable economic effects of inbreeding against the desirable economic value of some recessives. In the modified Pryce scheme the positive economic value of red coat color offsets the inbreeding penalty and maintains a relatively constant gene frequency over time. Avoidance of genomic inbreeding limits homozygosity, but eventually the population should become homozygous for the favorable allele.

Average TBV for the total merit index under selection were similar among the schemes over time. The difference between cows in year 20 of the two schemes was $21, which was $2,966 versus $2,987 for Pryce and modified Pryce, respectively. Bulls in generation 20 differed by only $10 on average ($3,737 versus $3,747). These differences are relatively small when compared to the overall genetic gain in the population, which averaged approximately $148 per year in cows and $186 per year in bulls.

Average coefficients of inbreeding by birth year were very similar for cows and bulls, increasing by approximately 0.35% per year in both populations. The same general pattern was observed across all scenarios and mate allocation schemes (data not shown). A value for λ of $25 was used, which is similar to the $23.11 calculated by [21], and higher than the $12 reported by Smith et al. [28] and the AUS$5 value used by Pryce et al. [18].

High costs. In this scheme, the economic value of each recessive was increased by a factor of 3 over the base Holstein scheme. Results were similar to the base Holstein scenario (Figure 3); this is probably due to the use of the same constant to scale values for all traits. The frequencies of HH5 and horned decreased faster under the modified Pryce scenario than the Pryce scenario ($P < 0.05$), while HH1 decreased more quickly under the Pryce scenario.

Hypothetical recessives

High frequency, lethal recessives. The rate of allele frequency change was similar for both the low ($20; Figure 4) and high ($200; data not shown) value scenarios. This suggests that at minor allele frequency the change from generation to generation is driven principally by genotype frequencies, not by economic value. The fit of the observed to expected allele frequency changes was very good in both scenarios (data not shown).

Medium frequency, lethal recessives. Results for a minor allele with an initial frequency of 0.50 and an economic value of either $20 or $200 were very similar to those for the previous section. The economic values were again dwarfed by the allele frequency, and a different mate allocation strategy will be needed to decrease the allele frequency more quickly.

Low frequency, lethal recessives. The two low-frequency scenarios discussed in this section are representative of the recessives seen most commonly in livestock populations [22], harmful alleles with low frequencies (< 0.05). Both the Pryce and modified Pryce methods are successful at decreasing the allele frequency over time when the value of the recessive is high, and they do so more quickly than expected. However, the modified Pryce's method appears to be more effective than random mating, truncation selection, or Pryce's method schemes at lowering the allele frequency when the economic value of the recessive is low.

Six hypothetical, lethal recessives. All four systems of mate allocation produced similar changes in allele frequencies over time. Pryce's method and the modified Pryce's method do produce slightly lower frequencies for some of the alleles that had high or medium initial frequencies, but there was no apparent pattern based on the economic value of each locus. Observed allele frequencies showed much better fits to the predicted values than in the scenarios based on the actual Holstein recessives, but that is expected when alleles have initial frequencies greater than 0.20.

There was no apparent difference between the change in allele frequencies over time even though there was a tenfold difference between the high- ($200) and low-valued ($20) recessives. When the minor allele frequency is high, many of the potential mate pairs in the population will have their parent averages reduced, but the loci with large values will be decreased more than those with low values, which should result in few carrier-to-carrier matings. Increasing the economic value of recessives is not alone sufficient to increase the rate at which undesirable loci are eliminated from the population.

Horned and other high-frequency non-lethal recessives

The horned allele is present at a frequency greater than 99% in the US Holstein population, and there is increasing interest in reducing its frequency to improve animal welfare. A scenario including only the horned recessive was simulated to determine if the modified Pryce's scheme is an effective tool for reducing the frequency of horned (increasing the frequency of polled) in the population. A $40 value for horned was not effective in reducing the minor allele frequency, probably because the frequency of the polled allele is so low that carriers were unlikely to be one of the top-ranked bulls by TBV. The limitation of 5,000 matings per generation also limits the effect of a high TBV polled bull on the population. Increasing the value of horned from $40 to $400 also was unsuccessful in changing the allele. These results are consistent with those of the 12 recessive scenarios described above, in which there was no change in the frequency of horned. A more sophisticated approach for selecting mate pairs that will either produce polled offspring or heterozygotes, such as a scheme described by Li et al. [13,14] or Spurlock et al. [29] or the use of tools for non-meiotic allele introgression [30], will be needed to effectively increase the frequency of polled (decrease the frequency of horned) cows in the national dairy herd.

Mating schemes

As expected, there was negligible genetic trend under the random mating scheme except in scenarios in which lethals had initial minor allele frequencies greater than 20%. The results from the truncation selection scheme were generally similar to the Pryce's and modified Pryce's schemes for lethals, and to random mating for non-lethal recessives. This is reasonable because the allele frequency of the lethals is expected to decrease over time even if no additional selection pressure is imposed, and the threshold that retains the top 10% of bulls for breeding ensures that genetic trend is positive. The truncation selection scheme loosely resembles current mating strategies used on large commercial dairies in North America.

More affected calves were observed in the Pryce's and modified Pryce's schemes than in the random mating and truncation selection schemes. Figure 5 shows the proportion of simulated calves that were culled due to recessive genotypes averaged over replicates of the Holstein scenario; results were similar for the high value, high frequency and low value, low frequency scenarios (data not shown). This is expected because a bull can have genetic superiority over his contemporaries greater than the value assigned to the recessives he may carry. Selection for reduced allele number rather than reduced frequency of recessive genotypes could result in fewer embryonic losses [15]. There is a conflict between the goal of eliminating recessives from the population, which involves fixing associated haplotypes in a homozygous state, and minimizing inbreeding, which seeks to avoid such increases in homozygosity.

Relationships of inbreeding with recessive load

The relationship of inbreeding with the number of recessives carried by parents was examined by computing the correlation of f_{ij} with the sum of P(aa) for each possible mating in each generation (ΣP(aa)) for scenarios including 12 (the Holstein scenario discussed above), 100, or 1,000 recessives. Contrary to expectations, the correlation of f_{ij} with ΣP(aa) was near 0 in the Holstein scenario, and negative in the 100- and 1,000-recessive scenarios. The correlation was stronger for matings made than those not made, suggesting that the modified Pryce's method was successful in identifying matings that reduced the accumulation of recessives. Figure 6 shows the regressions for the matings evaluated in birth year 20 of replicate 1 of each scenario by mating category (0: mating not made, 1: mating made); results were similar across replicates. The final birth year in each scenario was chosen for plotting because they provided the most opportunity to generate correlations among inbreeding and the number of recessives carried by individuals.

The lack of large correlations may be due in part to the low allele frequencies used in most scenarios. When deleterious recessives have very low frequencies the probability that a mating will be affected by more than one recessive, and that such a mating will have a DGV extreme enough overcome the penalty, is extremely low. If the frequencies of the recessives were high a stronger relationship would probably be observed, but it is difficult to consider a situation in which several recessives would be at a high frequency in the population.

Although inbreeding is inevitable in finite populations under selection [e.g., 31], its deleterious effects can be managed if harmful alleles are eliminated from the population. There is evidence that such purging has occurred in cattle populations [32, 33], and affected animals in the 100- and 1000-recessive scenarios were eliminated from the population quickly. As the number of recessives in real populations increase it is more likely that affected matings will occur, resulting in purging consistent with the trends in Figures 6a and 6b.

Mate allocation

Mate allocation, the process of selecting mating pairs from a population of female and some portfolio of males, has a long history in animal breeding programs in both general [16, 18, 34, 35, 36] and trait-specific [37] applications. Many artificial insemination firms provide mating recommendations to their customers as part of their services, but the algorithms used are usually very simple. Sun et al. [38] recently showed that rates of genetic gain can be further increased when genomic relationships are used and matings are allocated using linear programming **(LP)**. This may be a more practical way to account for recessives than including them in selection indices because of the difficulty of obtaining the marginal cost of a recessive independent of all other costs already accounted for by the other traits in the index, although the possibility of double-counting costs remains.

An advantage of the modified Pryce method over Pryce's original method is that the former can be used to maintain the frequency of desirable recessives, such as red coat color, in the population. There are other recessives, such as slick hair coat [39], that are segregating in some lines of Holstein that are desirable to producers in subtropical regions, and the modified Pryce's method could be used to increase the frequency of that allele in the general population.

However, these two approaches suffer from order-dependence, that is, if the cows are reordered before bulls are allocated the mate pairs change. This is not a serious problem if the best bulls in the population have similar breeding values, but could be important if a small group of elite bulls has much higher breeding values than other active bulls. The use of LP could eliminate this problem, at the cost of some added complexity in the implementation phase. Sun et al. [38] found in simulation that expected progeny differences were slightly higher when using LP compared to Pryce's method. Progeny inbreeding also was slightly lower using LP. Similar gains were reported by Weigel and Lin [21].

Sequential allocation as used in the Pryce and modified Pryce algorithms cannot account for a situation in which the value of one mating is affected by other matings, which is common when matings on multiple farms are considered simultaneously or management of parental coancestry is desired. Van Eenennaam and Kinghorn [15] recently extended the MateSel program [16] to permit selection against the number of lethal alleles and recessive lethal genotypes. The genetic progress foregone to decrease the incidence of lethal homozygotes is dependent upon allele frequencies, the number of lethal loci, and the emphasis that is placed on avoiding embryonic deaths. Their approach theoretically more satisfying than the algorithm presented in this paper, but there is often considerable reluctance by breeding organizations in the US to modify their software. Because of this, ease-of-implementation is often accorded more importance than theoretically optimal properties, and it is better to have an imperfect mate allocation tool used than no tool at all.

Integration with on-farm systems

As of 27 July 2015 there were 1,059,438 genotypes in the National Dairy Database maintained by the Council on Dairy Cattle Breeding (Reynoldsburg, OH, USA), of which 854,766 were from females (https://www.cdcb.us/Genotype/cur_freq.html). The modified Pryce's method described in this paper can easily be integrated into existing herd management and mate planning software, where it can be used in combination with these genotypes to better inform culling decisions or identify matings to be avoided. In the case of some haplotypes, such as A2 beta-casein and polled, this may be a useful tool for increasing allele frequencies without sacrificing substantial cumulative genetic gain.

Tradeoffs and limitations

In the modified Pryce scenario it was possible to reduce but not eliminate embryonic mortality. As the relative weighting (economic value) of loci increases the foregone genetic progress also will increase. MacArthur et al. [40] recently estimated that human genomes contain approximately 100 loss-of-function mutations, and about 20 genes that are completely inactivated. While not all of those mutations are lethal, it suggests that the 100-locus scenario of Van Eenennaam and Kinghorn [15] represents a plausible limit to the selection problem. Segelke et al. [41] suggested that a genetic index including haplotypes of interest should be used when selecting females for mating, and breeding values should be used to select bulls in order to balance selection for specific alleles with genetic gain. As the number of recessives increases it will be increasingly difficult to assign proper weights to each of them, and the marginal value of each recessive will be difficult to calculate without double-counting.

Conclusions

A modified version of Pryce's method [18] that accounts for the economic effects of recessive conditions was developed and compared with random mating, truncation selection, and Pryce's method for several different scenarios, including hypothetical alleles as well as 12 recessives currently segregating in the US Holstein population. The new method appears capable both of reducing the frequency of undesirable recessives with low frequencies and maintaining or increasing the frequency of desirable recessives. The method can easily be implemented in software used for mate allocation, and the code used in this study is freely available for use as a reference implementation.

Competing interests

The author declares that he has no competing interests.

Acknowledgements

The author would like to thank Sophie Eaglen and Paul VanRaden for several suggestions about the study, Kristen Parker Gaddis for computational and editorial assistance, and Daniel Null for technical editing of the manuscript. The editor and two anonymous reviewers provided important advice for improving the quality of the paper. This research was supported by appropriated project 1265-31000-096-00, “Improving Genetic Predictions in Dairy Animals Using Phenotypic and Genomic Information”, of the Agricultural Research Service of the United States Department of Agriculture, and Agriculture and Food Research Initiative Competitive Grant No. 2013-68004-20365, “Improving Fertility of Dairy Cattle Using Translational Genomics”. Mention of trade names or commercial products in this article is solely

for the purpose of providing specific information and does not imply recommendation or endorsement by the US Department of Agriculture. The USDA is an equal opportunity provider and employer.

References

1. Nicholas F, Hobbs M: **Mutation discovery for Mendelian traits in non-laboratory animals: a review of achievements up to 2012**. *Anim Genet* 2014, **45**:157–170.
2. Robertson A, Rendel JM: **The use of progeny testing with artificial insemination in dairy cattle**. *J Genet* 1950, **50**:1–31.
3. Shuster DE, Kehrli ME Jr, Ackermann MR, Gilbert RO: **Identification and prevalence of a genetic defect that causes leukocyte adhesion deficiency in Holstein cattle**. *Proc Natl Acad Sci USA* 1992, **89**:9225–9229.
4. Agerholm J, Bendixen C, Andersen O, Arnbjerg J: **Complex vertebral malformation in Holstein calves**. *J Vet Diag Investig* 2001, **13**:283–289.
5. Shanks RD, Dombrowski DB, Harpestad GW, Robinson JL: **Inheritance of UMP synthase in dairy cattle**. *J Heredity* 1984, **75**:337–340.
6. Sonstegard TS, Cole JB, VanRaden PM, Van Tassell CP, Null DJ, Schroeder SG, et al.: **Identification of a nonsense mutation in CWC15 associated with decreased reproductive efficiency in Jersey cattle**. *PLOS ONE* 2013, doi:10.1371/journal.pone.0054872.
7. Fernando RL, Grossman M: **Marker assisted selection using best linear unbiased prediction**. *Genet Sel Evol* 1989, **21**:467–477.
8. Kinghorn BP, Kennedy BW, Smith C: **A method of screening for genes of major effect**. *Genetics* 1993, **134**:351–360.

9. Meuwissen THE, Goddard ME: **Estimation of effects of quantitative trait loci in large complex pedigrees**. *Genetics* 1997, **146**:409–416.

10. Nejati-Javaremi A, Smith C, Gibson JP: **Effect of total allelic relationship on accuracy of evaluation and response to selection**. *J Anim Sci* 1997, **75**:1738–1745.

11. Shepherd RK, Kinghorn BP: **Designing algorithms for mate selection when major genes or QTL are important**. *Proc Assoc Advmt Anim Breed Genet* 2001, **14**:377–380.

12. Shepherd RK: **Implementing look ahead mate selection**. *Proc Assoc Advmt Anim Breed Genet* 2005, **16**:298–301.

13. Li Y, Van Der Werf JHJ, Kinghorn BP: **Optimisation of crossing system using mate selection**. *Genet Sel Evol* 2006, **38**:1–36.

14. Li Y, Van der Werf JHJ, Kinghorn BP: **Optimal utilization of non-additive quantitative trait locus in animal breeding programs**. *J Anim Breed Genet* 2008, **125**:342–350.

15. Van Eenennaam AL, Kinghorn BP: **Use of mate selection software to manage lethal recessive conditions in livestock populations**. *Proc 10th World Congr Genet Appl Livest Prod* 2014, https://asas.org/docs/default-source/wcgalp-posters/408_paper_9819_manuscript_1027_0.pdf?sfvrsn=2. Accessed 27 Feb 2015.

16. Kinghorn BP: **An algorithm for efficient constrained mate selection**. *Genet Sel Evol* 2011, doi:10.1186/1297-9686-43-4.

17. VanRaden PM, Olson KM, Null DJ, Hutchison JL: **Harmful recessive effects on fertility detected by absence of homozygous haplotypes**. *J Dairy Sci* 2011, **94**:6153–6161.

18. Pryce JE, Hayes BJ, Goddard ME: **Novel strategies to minimize progeny inbreeding while maximizing genetic gain using genomic information**. *J Dairy Sci* 2012, **95**:377–388.

19. VanRaden PM, Cole JB: **AIP Research Report NM$5: Net merit as a measure of lifetime profit: 2014 revision**. *Animal Genomics and Improvement Laboratory, ARS, USDA* 2014, http://aipl.arsusda.gov/reference/nmcalc-2014.htm. Accessed 26 Feb 2015.

20. Council on Dairy Cattle Breeding: **December 2014 across breed base adjustment parameters**. 2014, https://www.cdcb.us/eval/summary/Bmean_bases_het.cfm. Accessed 23 Feb 2015.

21. Weigel KA, Lin SW: **Use of computerized mate selection programs to control inbreeding of Holstein and Jersey cattle in the next generation**. *J Dairy Sci* 2000, **83**:822–828.

22. Cole JB, VanRaden PM, Null DJ, Hutchison JL, Cooper TA, Hubbard SM: **AIPL Research Report GENOMIC3: Haplotype tests for recessive disorders that affect fertility and other traits**. *Animal Genomics and Improvement Laboratory, ARS, USDA* 2013, http://aipl.arsusda.gov/reference/recessive_haplotypes_ARR-G3.html. Accessed 5 Feb 2015.

23. Medugorac I, Seichter D, Graf A, Russ I, Blum H, Göpel KH, et al.: **Bovine polledness - an autosomal dominant trait with allelic heterogeneity**. *PLOS ONE* 2012, doi:10.1371/journal.pone.0039477.

24. Widmar NJO, Schutz MM, Cole JB: **Breeding for polled dairy cows versus dehorning: Preliminary cost assessments & discussion**. *J Dairy Sci* 2013, **96**(Suppl. 2):602.

25. Van Doormaal BJ, Kistemaker GJ: **Managing genetic recessives in Canadian Holsteins**. *Interbull Bull* 2008, **38**:70–74.

26. Falconer DS, MacKay FC: *Introduction to Quantitative Genetics. 4th ed*. New York: John Wiley & Sons; 1996.

27. Seabold JS, Perktold J: Statsmodels: **Econometric and statistical modeling with Python**. In: Proceedings of the 9th Python in Science Conference. 2010, http://conference.scipy.org/proceedings/scipy2010/pdfs/seabold.pdf. Accessed 5 Feb 2015.

28. Smith LA, Cassell BG, Pearson RE: **The effects of inbreeding on the lifetime performance of dairy cattle**. *J Dairy Sci* 1998, **81**:2729–2737.

29. Spurlock DM, Stock ML, Coetzee JF: **The impact of 3 strategies for incorporating polled genetics into a dairy cattle breeding program on the overall herd genetic merit**. *J Dairy Sci* 2014, **97**:5265–5274.

30. Tan W, Carlson DF, Lancto CA, Garbe JR, Webster DA, Hackett PB, et al.: **Efficient nonmeiotic allele introgression in livestock using custom endonucleases**. *Proc Natl Acad Sci USA* 2013, **110**:16526–16531.

31. Kristensen TN, Hoffmann AA, Pertoldi C, Stronen AV: **What can livestock breeders learn from conservation genetics and *vice versa*?** *Front Genet* 2015, http://dx.doi.org/10.3389/fgene.2015.00038.

32. McParland S, Kearney F, Berry DP: **Purging of inbreeding depression within the Irish Holstein-Friesian population.** *Genet Sel Evol* 2009, **41**:16.

33. Gulisija D, Crow JF: **Inferring purging from pedigree data.** *Evol* 2007, **61**: 1043–1051.

34. Jansen GB, Wilton JW: **Selecting mating pairs with linear programming techniques**. *J Dairy Sci* 1985, **68**:1302–1305.

35. Allaire FR: **Mate selection by selection index theory**. *Theoret Appl Genet* 1980, **57**:267–272.

36. Sonesson AK, Meuwissen THE: **Mating schemes for optimum contribution selection with constrained rates of inbreeding**. *Genet Sel Evol* 2000, **32**:231–248.

37. Dekkers JCM: **Optimal breeding strategies for calving ease**. *J Dairy Sci* 1994, **77**:3441–3453.

38. Sun C, VanRaden PM, O'Connell JR, Weigel KA, Gianola D: **Mating programs including genomic relationships and dominance effects**. *J Dairy Sci* 2013, **96**:8014–8023.

39. Huson HJ, Kim E-S, Godfrey RW, Olson TA, McClure MC, Chase CC, et al.: **Genome-wide association study and ancestral origins of the slick-hair coat in tropically adapted cattle**. *Front Genet* 2014, doi:10.3389/fgene.2014.00101.

40. MacArthur DG, Balasubramanian S, Frankish A, Huang N, Morris J, Walter K, et al.: **A systematic survey of loss-of-function polymorphisms in human protein-coding genes**. *Science* 2012, **335**:823–828.

41. Segelke D, Täubert H, Jansen S, Pausch H, Reinhardt F, Thaller G: **Management of genetic characteristics**. *Interbull Bull* 2014, **48**:85–88.

Figures

Figure 1 - Observed allele frequencies for Holstein recessives

Observed changes in minor allele frequencies for BLAD, brachyspina, CVM, DUMPS, HH1–HH5, mulefoot, and red coat color over 20 years under random selection, truncation selection, Pryce’s method for controlling genomic inbreeding, and Pryce’s method accounting for recessives.

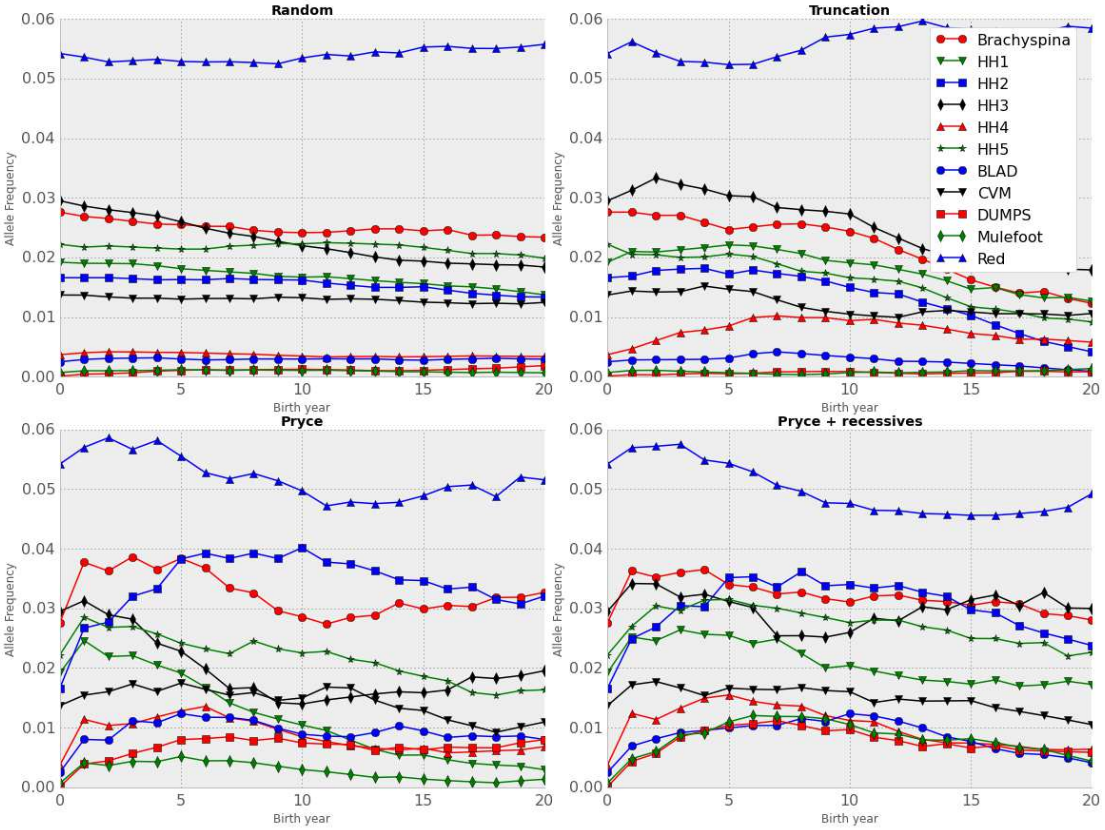


Figure 2 - Observed versus expected allele frequencies under the Pryce scenario

Observed versus expected allele frequencies under the Pryce scenario. Observed versus expected changes in minor allele frequencies for BLAD, brachyspina, CVM, DUMPS, HH1–HH5, horned, mulefoot, and red coat color over 20 years using Pryce's method for controlling genomic inbreeding. Note that the horned subplot is scaled differently on the y axis than the other subplots because of its allele frequency.

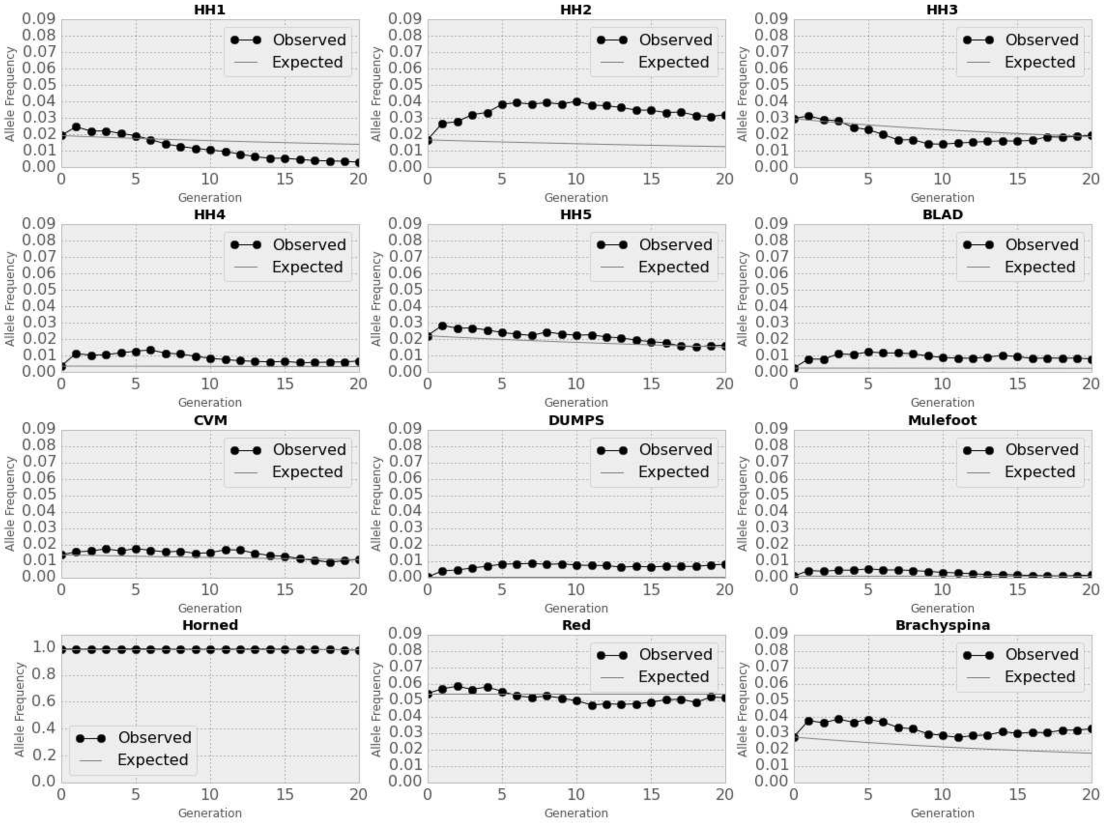


Figure 3 - Observed allele frequencies for Holstein recessives with high economic values

Observed changes in minor allele frequencies for BLAD, brachyspina, CVM, DUMPS, HH1–HH5, mulefoot, and red coat color over 20 years under random selection, truncation selection, Pryce’s method for controlling genomic inbreeding, and Pryce’s method accounting for recessives.

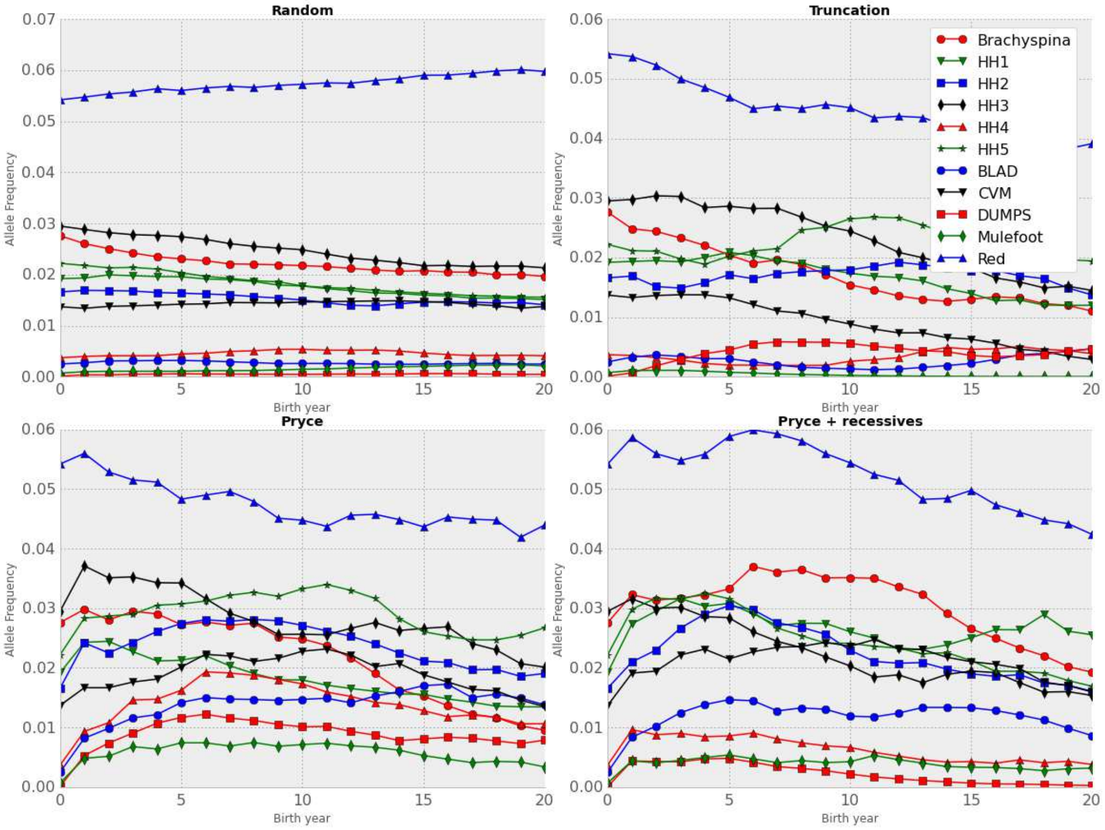


Figure 4 - Observed allele frequencies for a hypothetical recessive with a high frequency and low value

Observed changes in minor allele frequency for a hypothetical recessive with a starting frequency of 0.90 and an economic value of $20 over 20 years under random selection, truncation selection, Pryce's method for controlling genomic inbreeding, and a modified Pryce's method that accounts for recessives.

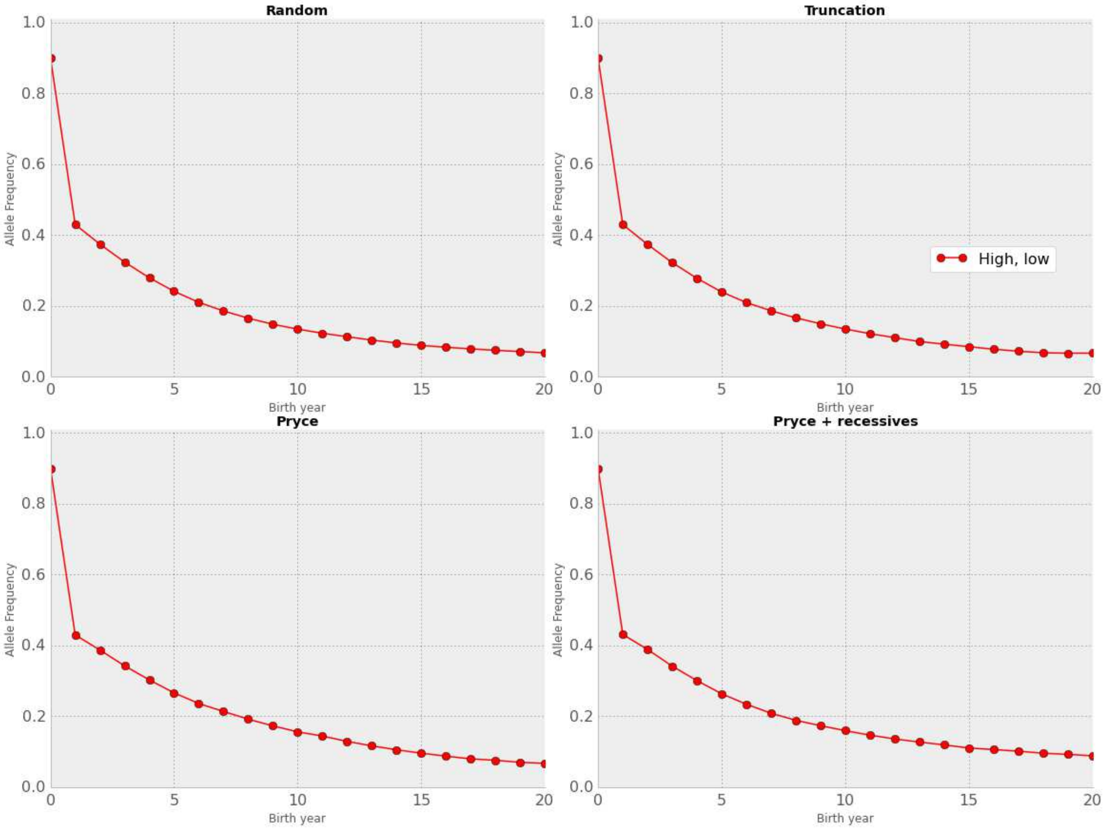


Figure 5 - Embryonic deaths by birth year

Proportion of embryos in each birth year that died due to the effects of recessive genotypes.

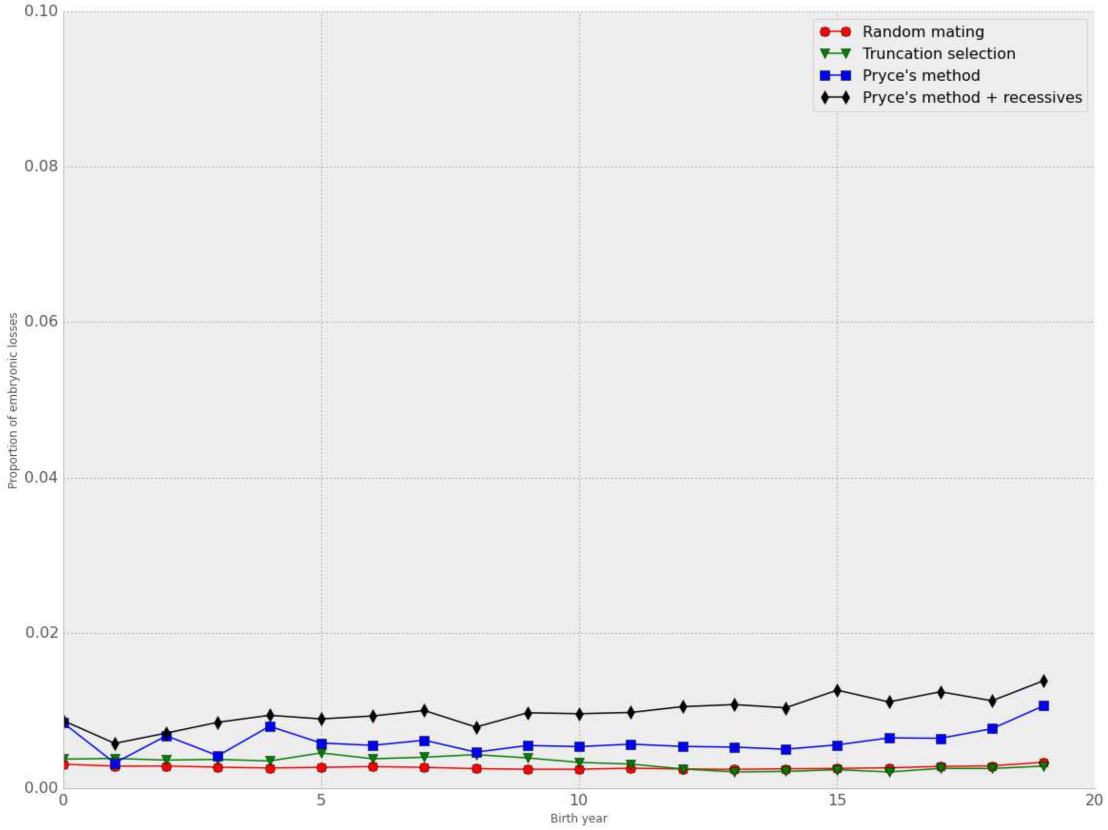
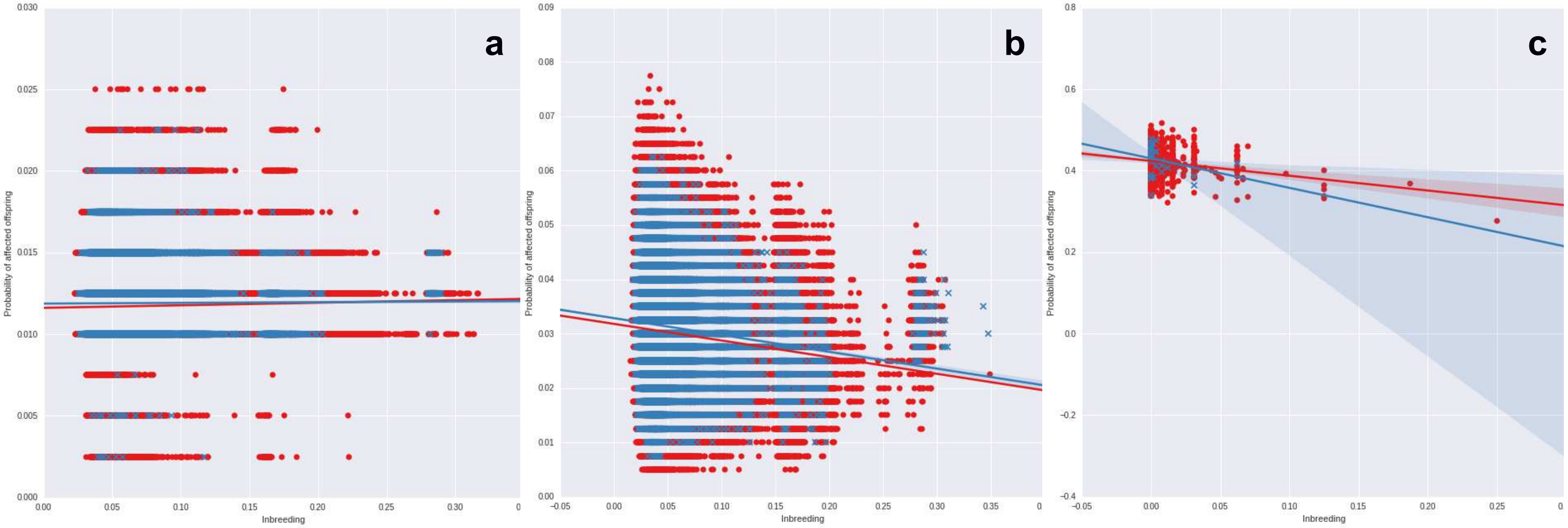


Figure 6 - Embryo inbreeding and probability of carrying recessives

Relationships of embryo inbreeding with the probability that the embryo will be affected by one or more recessive conditions for (a) Holstein recessives, (b) 100 simulated recessives, and (c) 1,000 simulated recessives. The mating variable distinguishes matings that were not made (red dots) from those that were made (blue crosses).



Tables

Table 1 - Properties of the recessives included in each scenario simulated

				Recessives		
Group	Scenario[1]	N[2]	Frequency	Value ($)[3]	Name	Lethal
Holstein	All recessives	12	0.0276	150	Brachyspina	Yes
			0.0192	40	HH1	Yes
			0.0166	40	HH2	Yes
			0.0295	40	HH3	Yes
			0.0037	40	HH4	Yes
			0.0222	40	HH5	Yes
			0.0025	150	BLAD	Yes
			0.0137	70	CVM	Yes
			0.0001	40	DUMPS	Yes
			0.0007	150	Mulefoot	Yes
			0.9929	40	Horned	No
			0.0542	-20	Red coat color	No
	All recessives, high cost	12	0.0276	450	Brachyspina	Yes
			0.0192	120	HH1	Yes
			0.0166	120	HH2	Yes
			0.0295	120	HH3	Yes
			0.0037	120	HH4	Yes
			0.0222	120	HH5	Yes
			0.0025	450	BLAD	Yes

			0.0137	210	CVM	Yes
			0.0001	120	DUMPS	Yes
			0.0007	450	Mulefoot	Yes
			0.9929	120	Horned	No
			0.0542	-60	Red coat color	No
Hypothetical	High frequency, low value	1	0.90	20	High. low	Yes
	High frequency, high value	1	0.90	200	High, high	Yes
	Medium frequency, low value	1	0.50	20	Medium, low	Yes
	Medium frequency, high value	1	0.50	200	Medium, high	Yes
	Low frequency, low value	1	0.01	20	Low, low	Yes
	Low frequency, high value	1	0.01	200	Low, high	Yes
	All recessives	6		As above.		
Horned	Horned, market value	1	0.9929	40	Horned	No
	Horned, high value	1	0.9929	400	Horned	No

[1] The specific scenario simulated for each trait or group of traits.

[2] The number of recessives in the scenario.

[3] Positive values are undesirable and negative values are desirable.