View metadata, citation and similar papers at core.ac.uk

Genet. Sel. Evol. 35 (2003) 353–368
© INRA, EDP Sciences, 2003
DOI: 10.1051/gse:2003028

Original article

Selection against genetic defects in conservation schemes while controlling inbreeding

Anna K. Sonesson*, Luc L.G. Janss, Theo H.E. Meuwissen

Institute of Animal Science and Health (ID-Lelystad), PO Box 65, 8200 AB Lelystad, The Netherlands

(Received 9 April 2002; accepted 15 January 2003)

Abstract – We studied different genetic models and evaluation systems to select against a genetic disease with additive, recessive or polygenic inheritance in genetic conservation schemes. When using optimum contribution selection with a restriction on the rate of inbreeding (ΔF) to select against a disease allele, selection directly on DNA-genotypes is, as expected, the most efficient strategy. Selection for BLUP or segregation analysis breeding value estimates both need 1–2 generations more to halve the frequency of the disease allele, while these methods do not require knowledge of the disease mutation at the DNA level. BLUP and segregation analysis methods were equally efficient when selecting against a disease with single gene or complex polygene inheritance, *i.e.* knowledge about the mode of inheritance of the disease was not needed for efficient selection against the disease. Smaller schemes or schemes with a more stringent restriction on ΔF needed more generations to halve the frequency of the disease alleles or the fraction of diseased animals. Optimum contribution selection maintained ΔF at its predefined level, even when selection of females was at random. It is argued that in the investigated small conservation schemes with selection against a genetic defect, control of ΔF is very important.

genetic defects / selection / inbreeding / conservation

1. INTRODUCTION

Many domesticated animal populations show heritable defects. Some defects are inherited by a single gene, *e.g.* complex vertebral malformation (CVM) in cattle [1]. Other diseases have a complex inheritance involving

* Correspondence and reprints
E-mail: Anna.Sonesson@akvaforsk.nlh.no
Current adress: AKVAFORSK (Institute of Aquaculture Research Ltd), PO Box 5010, 1432 Ås, Norway

multiple genes plus environmental effects, *e.g.* hip and elbow dysplasia in dogs [17].

One way to eliminate the disease from the population is to select against the disease in a breeding program. For diseases caused by an identified single gene, direct selection on DNA-genotypes against the disease allele is possible. This can be done irrespective of whether the disease is additionally affected by the environment (complete penetrance or not). For unknown genes, segregation analysis can be used to infer on the genotype probabilities of individual animals, using phenotypic records of the animal itself and relatives [2,5,7,13]. Segregation analysis can also be used to save genotyping costs when selecting on DNA-genotypes for known genes [15]. For diseases with complex inheritance (involving many genes), the assumption of normally distributed genetic effects seems more appropriate, leading to BLUP [12] or threshold model breeding value estimation [8]. However, the inheritance is unknown for many diseases and the breeding value estimation is not straightforward. We will here investigate the genetic models and evaluation methods to select against a disease of known [2,5,7,12,13] or unknown modes of inheritance.

Genetic drift increases the occurrence of heritable diseases. Genetic conservation schemes are often small and care has therefore to be taken to avoid high rates of inbreeding when selecting against the disease in such small populations. Increased inbreeding could for instance result from direct selection for a non-disease allele, detected by DNA genotyping, when the non-disease alleles come from a limited number of ancestral families. We will use a selection method that maximises genetic response with a restriction on the rate of inbreeding [10, 11,18,20]. The optimum contributions, which are translated to the optimum number of progeny will be calculated for each male selection candidate, assuming that female selection is at random. This reflects the situation, where every female is needed in a conservation scheme, or where there is little control over selection of the females.

The aim of this study was to find the best strategy for eliminating different kinds of genetic diseases, where the genetic evaluation method does not always agree with the true inheritance of the disease. We compared a threshold model, where many genes and environmental effects affect the liability of an animal to be diseased with a genetic model for a single gene. We also compared breeding values estimated from DNA-genotyping (for a known disease gene) to breeding values estimated by BLUP [12] or segregation analysis [2,5,7,13]. The disease trait is binary and is not (systematically) affected by the presence or absence of an infectious agent. Also, the disease is not genetically correlated to other traits under (natural or artificial) selection.

2. MATERIALS AND METHODS

2.1. Genetic model

2.1.1. Threshold model

The threshold genetic model assumes liabilities underlying the probability of having a diseased animal. The liability was assumed normally distributed. Genetic values for liability, g_i, of the base animals were sampled from the distribution $N(0, \sigma_a^2)$, where $\sigma_a^2 = 0.5$ is the base generation genetic variance. Environmental effects on liability, e_i, of base animals were sampled from the distribution $N(0, \sigma_e^2)$, where $\sigma_e^2 = 0.5$ is the environmental variance. Total liability was $x_i = g_i + e_i$. Later generations were obtained by simulating offspring genotypes from $g_i = 1/2g_s + 1/2g_d + m_i$, where s and d refers to sires and dams, respectively, and m_i is the Mendelian sampling component, sampled from $N\big(0, 1/2(1 - \bar{F})\sigma_a^2\big)$, where $\bar{F}$ is the average inbreeding coefficient of parents s and d. If x_i was higher than the threshold value, T, then the individual was diseased and $y_i = 0$. Healthy animals had $y_i = 1$. The threshold T was set to 0.0, which resulted in a disease incidence of 50% in the base generation. These phenotypic values, y_i, were used as input to estimate breeding values (EBV).

2.1.2. Single gene

For the base generation, two alleles of each animal were sampled, where allele A was sampled with probability q_0 and allele a was sampled with probability $(1 - q_0)$. For later generations, individual genotypes were sampled using Mendel rules. Animal i was diseased ($y_i = 0$) with probability $\big(P(y_i = 0|XX_i)\big)$, where $P(y_i = 0|XX_i)$ is the penetrance probability of having a diseased animal ($y_i = 0$) given genotype XX_i. When the inheritance was additive, the input values $P(y_i = 0|XX_i)$ were 0.0, 0.5 and 1.0 for genotypes XX_i = aa, Aa and AA, respectively. When the inheritance was recessive, these values were 0.0, 0.0 and 1.0 for genotypes aa, Aa and AA, respectively. The phenotypic disease records, y_i, which resulted from this sampling, were used as input for the genetic evaluation.

2.2. Genetic evaluation

2.2.1. BLUP

Phenotypic values from the threshold and single gene model were input to obtain EBV using a BLUP-breeding value estimation procedure [12]. This ignores the binary nature of the disease traits, but, when the fixed effect structure is as simple as here, where only an overall mean is fitted, linear BLUP-EBV

are almost as accurate as generalised linear mixed model EBV, which accounts for the binary nature of the disease trait [19].

For the threshold model [8], the animals are assumed to be diseased when a normally distributed liability trait is below a certain threshold, T, and the animals are assumed healthy when the trait is above T. For the estimation of BLUP breeding values, the heritability on the diseased scale, $h^2_{disease}$, is needed and obtained from [8]:

$$h^2_{disease} = f(T)^2 h^2_{liab}/[z(1-z)],$$

where z is the proportion of diseased animals when the threshold value is T, $f() =$ Normal density function and $h^2_{liab} =$ heritability of the liability trait. Here, $T = 0$, $z = 0.5$ and $h^2_{liab} = 0.5$, yielding $h^2_{disease} = 0.318$.

2.2.2. DNA genotyping

In this case, the disease was assumed to be due to a single known gene and only males were genotyped. When assigning the recessive genotype a value of 1, and the others a value of 0 (in Falconer and Mackay [6] notation $a = -d = 0.5$), it follows that the frequency of the disease genotype q^2 equals the disease incidence in the population [6]. Breeding values for the single gene were calculated as $EBV(aa) = 2q\alpha$, $EBV(Aa) = (q-p)\alpha$ and $EBV(AA) = -2p\alpha$, where α is the average effect of gene substitution, $\alpha = a + d(q-p)$ and d is the dominance deviation, $d = P(y_i = 0|Aa) - 0.5\big(P(y_i = 0|aa) + P(y_i = 0|AA)\big)$. These breeding values correspond to (twice the deviation of) disease incidences in progeny of the respective genotypes, and will be used as input for the selection algorithm to reduce disease incidence.

In the case of the threshold genetic model, the genetic effect is affected by many genes. We assume that not all genes are known, such that EBV from DNA genotyping cannot be calculated for the threshold genetic model.

2.2.3. Segregation analysis

The algorithm by Kerr and Kinghorn [14] was used to calculate genotype probabilities of each animal. It is an algorithm based on iterative peeling [2, 13] and it takes account of effects of selection.

Input for the segregation analysis is the probability that the phenotype was diseased given the genotypes XX_i, *i.e.* the penetrance probabilities. For an additive trait, the penetrance probabilities, $P(y_i = 0|XX_i)$ of a diseased animal i are 0.0, 0.5 and 1.0 for genotypes aa, Aa and AA, respectively. The probability of a non-diseased animal is $P(y_i = 1|XX_i) = 1 - P(y_i = 0|XX_i)$. For a recessive trait, $P(y_i = 0|XX_i)$ is 0.0, 0.0 and 1.0 for genotypes aa, Aa and AA, respectively, and again $P(y_i = 1|XX_i) = 1 - P(y_i = 0|XX_i)$. From these penetrance probabilities, the algorithm by Kerr and Kinghorn [14] calculates

the probability that the individual i has genotype XX, $P(XX)_i$. The $P(XX)_i$ are used to calculate EBV as:

$$EBV_i = P(aa)_i 2q\alpha + P(Aa)_i(q - p)\alpha - P(AA)_i 2p\alpha.$$

These EBV are input for the selection algorithms.

For the threshold genetic model, we estimated the penetrance probabilities as $P(y_i = 1|XX_i) = (\Sigma P(XX)_i y_i)/\Sigma P(XX)_i$ and $P(y_i = 0|XX_i) = 1 - P(y_i = 1|XX_i)$. Similarly, the initial allele frequencies were estimated as $q_o = \left[\sum_{base} P(AA)_i + \sum_{base} 1/2P(Aa)_i\right]/Nbase$, where Nbase is the number of base animals. Because these estimates of penetrance probabilities and initial frequencies depend on estimates of genotype probabilities $P(XX)_i$, which themselves depend on initial frequencies and penetrance probabilities, iteration was used to simultaneously estimate all these probabilities.

2.3. Optimum contribution selection method (OC)

Optimum contribution selection was used as proposed by Meuwissen [18]. This method maximises the genetic level of the next generation of animals, $\mathbf{G_{t+1}} = \mathbf{c_t'EBV_t}$, where $\mathbf{c_t}$ is the vector of genetic contributions of the selection candidates to generation $t + 1$ and $\mathbf{EBV_t}$ is the vector of estimated breeding values of the candidates for selection in generation t. The $\mathbf{c_t'EBV_t}$, is maximised for $\mathbf{c_t}$ under two restrictions: the first one is on the rate of inbreeding and the second one is on the contribution per sex. Rates of inbreeding are controlled by constraining the average coancestry of the selection candidates to $\bar{C}_{t+1} = \mathbf{c_t'A_tc_t}/2$, where $\mathbf{A_t}$ is a $(n \times n)$ relationship matrix among the selection candidates, $\bar{C}_{t+1} = 1 - (1 - \Delta F_d)^t$, and ΔF_d is the desired rate of inbreeding [10]. Note that the level of the restriction $\bar{C}_{t+1}$, can be calculated for every generation before the breeding scheme starts. Contribution of males (females) are constrained to 1/2, *i.e.* $\mathbf{Q'c_t} = \mathbf{1/2}$ where $\mathbf{Q}$ is a $(n \times 2)$ incidence matrix of the sex of the selection candidates (the first column yields ones for males and zeros for females, and the second column yields ones for females and zeros for males) and $\mathbf{1/2}$ is a (2×1) vector of halves. The selection algorithm presented in the Appendix of [18] optimised genetic contributions for each male selection candidate, $\mathbf{c_t}$, given that all dams had (*a priori*) equal contributions, *i.e.* there was no selection of females. In cases of single genes, at some point all selection candidates can have the desired genotype and a maximisation of genetic response is no longer relevant, in which case the algorithm switched to minimising inbreeding. What happens computationally is that the Lagrangian multiplier, λ_0, becomes zero when all animals have the same EBV and the equations for the optimal contributions cannot be solved (since they require dividing by λ_0). If this was the case, the simulation program called the minimisation routine presented in [22], which was modified here to handle discrete generations.

2.4. Mating

Random mating was applied. For each mating pair, a sire was randomly sampled with probabilities following the optimal contributions of the sires and a dam was randomly sampled from the available females. A mating pair always had two progeny, one female and one male.

2.5. Schemes

The general structure was that of a closed scheme with discrete generation structure. Recording of the disease was on both sexes before selection. The results were based on 100 replicated schemes with 60 or 100 selection candidates and on 50 replicated schemes for schemes with 200 selection candidates. Each replicate consisted of 15 generations of selection. Different constraints of ΔF per generation were considered. Firstly, ΔF was constrained to 0.010, which is considered as the maximum acceptable rate of inbreeding for a population to survive [3]. Secondly, for the larger schemes, the use of a more stringent ΔF constraint was simulated, with $\Delta F = 0.006$ and 0.003 for the schemes with 100 or 200 animals per generation, respectively. These more stringent ΔF constraints had the same ratio of N_e to N as the small schemes with 60 animals (0.833). We compared the evaluation models for the number of generations they needed to halve the frequency of the disease allele or the fraction of diseased animals for the single gene and threshold models, respectively.

3. RESULTS

3.1. Single gene model

For the genetic model with a single gene, the genetic evaluation was on DNA-genotype (GENO), BLUP EBV (BLUP) or on EBV based on genotype probabilities calculated by segregation analysis (SEGR).

As expected, GENO was the most efficient in reducing the frequency of the disease allele. BLUP and SEGR schemes always gave very similar results. For a scheme with 100 animals per generation and additive inheritance, GENO needed 2.0 generations to halve the frequency of the disease allele, whereas both BLUP and SEGR needed 3.0 generations (Fig. 1). As for a gene with additive inheritance, GENO also needed 2.0 generations to halve the frequency of the disease allele for a gene with a recessive inheritance, as expected (Tab. I). However, BLUP and SEGR needed more generations (4.0) than in the case of additive inheritance, because it is more difficult to identify and avoid selection of heterozygous animals, which have the same phenotype as non-diseased homozygotes, when inheritance is recessive.

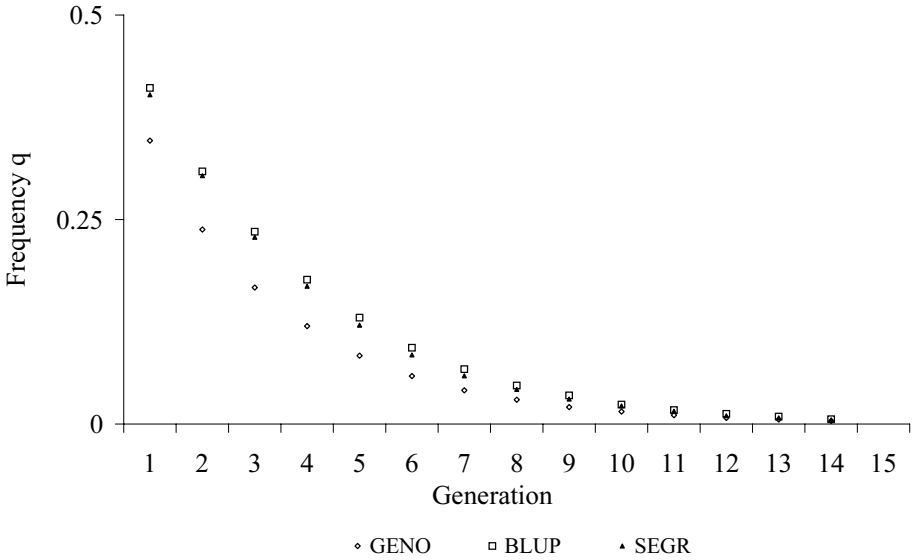

Figure 1. Single gene model. Frequency of disease allele (Frequency q) for schemes with 100 animals per generation and additive genetic effects. Genetic evaluation was done on DNA-genotype (GENO), BLUP EBV (BLUP) or on EBV based on genotype probabilities calculated by segregation analysis (SEGR).

Both BLUP and SEGR schemes achieved the restriction on ΔF of 0.010 during all generations (Fig. 2). The GENO scheme kept the restriction exactly until generation 3 (Fig. 2) and thereafter ΔF was lower than the maximum indicated by the restriction. This is because most animals have the non-disease genotype after three generations, and the simulation program switched to minimisation of ΔF, while still achieving the maximum selection response (selection of only homozygous non-disease genotypes).

In fact, the minimisation algorithm may already be used when many, but not all sires have the desirable genotype. In the latter situation, the selection algorithm leads to negative contributions for the disease allele carriers. The disease carriers will subsequently be eliminated from the list of selection candidates by the algorithm. In the resulting list of candidates, all animals have the desirable genotype and ΔF is minimised using these animals that are homozygous for the desirable allele (aa). EBV will differ somewhat in the BLUP and SEGR schemes, even if the gene frequency of the non-disease allele is 1.0. Selection among the candidates is then always possible, and the optimum contribution selection-algorithm will attempt to maximise EBV of the parents within the restriction on ΔF. Therefore, BLUP and SEGR kept the restriction on ΔF exactly and selected somewhat fewer sires than GENO (Tab. I).

Table I. Single gene model. Number of generations it took to halve the frequency of the disease allele (Halftime), number of selected sires (Nselsires) and accuracy of selection for schemes with ΔF restricted to 0.010, 0.006 or 0.003 per generation for schemes with 60, 100 or 200 animals per generation and additive or recessive inheritance of the single gene.

Genetic evaluation [1]	Halftime (gen)	Nselsires	Accuracy	Halftime (gen)	Nselsires	Accuracy
	Additive inheritance			Recessive inheritance		
	60 animals/generation, ΔF = 0.010					
GENO	3.0	22.3	1.000	3.0	21.8	1.000
BLUP	4.0	21.0	0.753	5.0	21.4	0.674
SEGR	4.0	21.0	0.757	5.0	21.6	0.699
	100 animals/generation, ΔF = 0.010					
GENO	2.0	23.6	1.000	2.0	23.9	1.000
BLUP	3.0	21.1	0.743	4.0	23.6	0.700
SEGR	3.0	19.9	0.743	4.0	20.9	0.712
	100 animals/generation, ΔF = 0.006					
GENO	3.0	35.5	1.000	3.0	36.5	1.000
BLUP	4.0	34.4	0.759	5.0	36.3	0.689
SEGR	4.0	34.4	0.757	5.0	35.6	0.704
	200 animals/generation, ΔF = 0.010					
GENO	1.5	38.3	1.000	1.5	37.4	1.000
BLUP	3.0	21.4	0.746	4.0	24.7	0.695
SEGR	3.0	18.8	0.744	4.0	23.8	0.705
	200 animals/generation, ΔF = 0.003					
GENO	3.0	73.0	1.000	3.0	73.7	1.000
BLUP	4.0	69.9	0.757	5.0	70.9	0.677
SEGR	4.0	69.9	0.769	5.0	72.3	0.703

[1] Genetic evaluation was done on DNA-genotype (GENO), BLUP EBV (BLUP) or on EBV based on genotype probabilities calculated by segregation analysis (SEGR).

For the small schemes with 60 animals per generation, GENO needed 3.0 and BLUP and SEGR 4.0 generations to halve the frequency of the disease allele for the gene with additive inheritance, *i.e.* smaller numbers of animals reduced the genetic response (Tab. I). For schemes with 200 animals per generation, GENO needed 1.5 and BLUP and SEGR 3.0 generations to halve the frequency of the disease allele. Hence, it takes a longer time to reduce gene frequency in smaller schemes, which is expected, because fewer selection candidates have the non-disease genotype.

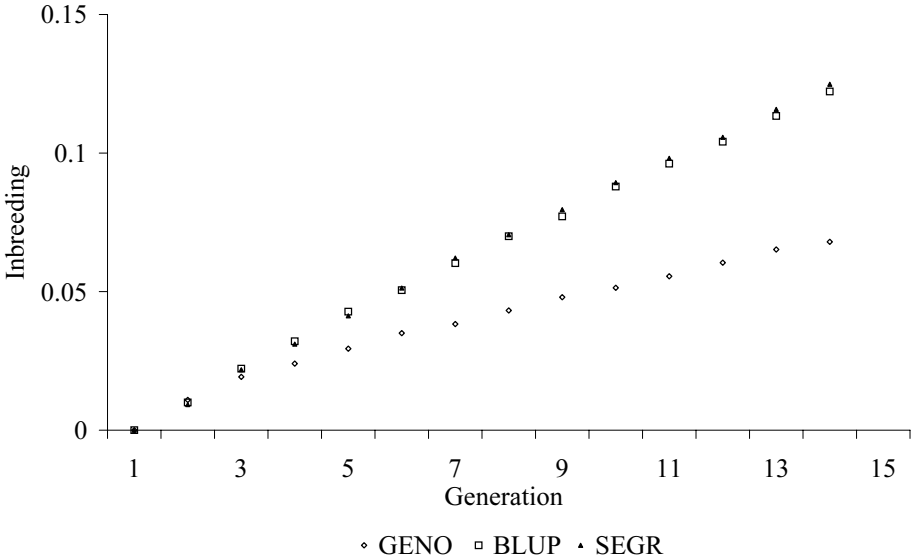

Figure 2. Single gene model. Level of inbreeding for schemes with 100 animals per generation and additive genetic effects. Genetic evaluation was done on DNA-genotype (GENO), BLUP EBV (BLUP) or on EBV based on genotype probabilities calculated by segregation analysis (SEGR).

Since it took more time to reduce the frequency of the disease allele for the smaller scheme with 60 animals per generation, ΔF was kept at the level of restriction for GENO for more generations (six for schemes with a gene that has an additive inheritance) than for the scheme with 100 animals per generation (not shown). Similarly, for the larger scheme with 200 animals per generation, ΔF was kept at the level of the restriction for GENO for only two generations. Thereafter, ΔF was minimised and thus lower than the restriction. For the BLUP and SEGR schemes, ΔF was kept at the restricted level during the whole period.

For the scheme with 200 animals per generation, GENO seemed in general to select more (about 38) sires than BLUP and SEGR (about 21), for the schemes with additive and recessive inheritance (Tab. I), because in later generations, the simulation program was able to minimise ΔF and still achieve a maximum selection response.

In order to investigate whether the higher genetic gain in the larger schemes is entirely due to their higher actual relative to effective population size, we also simulated single gene schemes, where the ratio of N_e over N was the same as for 60 animals. In those schemes, N_e over N was 0.833 and the rate of inbreeding was restricted to 0.006 and 0.003 per generation for schemes with 100 and 200 animals per generation, respectively. At constant N_e/N, the three

schemes with 60, 100 and 200 animals per generation indeed achieved a very similar selection response, *i.e.* GENO needed 3.0 generations and BLUP and SEGR 4.0 generations to halve the frequency of the disease allele for a gene with additive inheritance (Tab. I). For a gene with recessive inheritance, when compared at the same ratio of N_e to N, GENO needed 3.0 generations and BLUP and SEGR schemes 5.0 generations to halve the frequency of the disease allele for all three sizes of schemes. Thus, the ratio of N_e to N seems to determine the selection intensity of the scheme and also the genetic response.

For the scheme with 100 and 200 animals per generation, but the same ratio of N_e to N as the scheme with 60 animals per generation, GENO kept ΔF at the restricted level for 8 generations and thereafter ΔF was lower than the maximum indicated by the restriction for both genes with additive and recessive inheritance (not shown). BLUP and SEGR kept the restriction on ΔF during all generations.

There was an increase in the number of selected sires with an increasing effective population size. The number of selected sires was twice as many for the scheme with ΔF restricted to 0.003 (about 70) than for the scheme with ΔF restricted to 0.006 (about 35) (Tab. I). The same number of sires (about 21) was selected for schemes with the same ΔF (0.010), but with different actual population sizes.

For all BLUP and SEGR schemes, the accuracy of selection was between 0.67 and 0.76 (Tab. I).

3.2. Threshold model

For the threshold genetic model, where the genetic evaluation was either with BLUP or segregation analysis (SEGR), the fraction of diseased animals, which started at 0.50, was monitored. For schemes with 100 animals per generation, it took about 3.5 generations to halve the fraction of diseased animals to 0.25 for both BLUP and SEGR (Tab. II). Hence, even if the true genetic model involves many genes, but it is believed that the disease is determined by a single gene, SEGR selects animals with high disease resistance and reduces the fraction of diseased animals as fast as BLUP.

The restriction on ΔF of 0.01 was kept at the restricted level for both BLUP and SEGR (not shown).

The number of selected sires was also about the same (Tab. II) for both BLUP (22.7) and SEGR (21.7).

For schemes with 60 and 200 animals per generation, it took about 5.0 and 3.0 generations, respectively, to halve the fraction of diseased animals. Both BLUP and GENO achieved the restriction on ΔF. For schemes with 60 animals per generation, the number of selected sires was 21.2 for BLUP and 21.4 for SEGR (Tab. II). For schemes with 200 animals per generation, the number of selected sires was 25.6 for BLUP and 20.5 for SEGR (Tab. II).

Table II. Threshold model. Number of generations it took to halve the fraction of diseased animals (Halftime) and number of selected sires (Nselsires) for schemes with ΔF restricted to 0.010, 0.006 or 0.003 per generation for schemes with 60, 100 or 200 animals per generation.

Genetic evaluation [1]	Halftime (gen)	Nselsires
60 animals/generation, $\Delta F = 0.010$		
BLUP	5.0	21.2
SEGR	5.0	21.4
100 animals/generation, $\Delta F = 0.010$		
BLUP	3.5	22.7
SEGR	3.5	21.7
100 animals/generation, $\Delta F = 0.006$		
BLUP	5.0	35.5
SEGR	5.0	33.8
200 animals/generation, $\Delta F = 0.010$		
BLUP	3.0	25.6
SEGR	3.0	20.5
200 animals/generation, $\Delta F = 0.003$		
BLUP	5.0	71.0
SEGR	5.0	71.1

[1] Genetic evaluation was done on BLUP EBV (BLUP) or on EBV based on genotype probabilities calculated by segregation analysis (SEGR).

Schemes, which had the same ratio of N_e to N as the scheme with 60 animals per generations, were also investigated for the threshold model. For schemes with 60, 100 and 200 animals per generation, it took 5.0 generations for both BLUP and SEGR to halve the fraction of diseased animals (Tab. II). Hence, also for the threshold model scheme, the ratio of N_e to N seemed to determine the selection intensity and thus also genetic response.

For the threshold genetic model schemes, about the same numbers of sires were selected as for the single gene model.

4. DISCUSSION

4.1. General

For the single gene model, direct selection against the disease allele (GENO) was, as expected, the most efficient evaluation method to reduce the frequency of a disease allele of a known single gene. SEGR and BLUP needed about 1.0 to 2.0 more generations to halve the frequency of the disease allele. This period

was shorter for a gene with an additive inheritance and longer for a gene with a recessive inheritance. In the situation where the gene is unknown, an approach could be to first identify the gene and then apply GENO. A gene needs to be identified approximately within 1.0 (additive gene) to 2.0 (recessive gene) generations for this approach (identifying the gene and using GENO) to be as effective as SEGR and BLUP to halve the frequency of the disease allele.

For a threshold genetic model, SEGR and BLUP were equally efficient in reducing the fraction of diseased animals. Hence, when genes are unknown, either BLUP or SEGR can be used, although BLUP is a more natural choice, because it assumes many genes, which reflects the true genetic model here.

4.2. Genetic evaluation methods

When selecting against a disease allele, there is no steady state reached after some generations, as is the case when selection is for normally distributed traits. Instead, the number of non-diseased candidates changes with the frequency of the allele. This has implications for the achieved ΔF. Here we used the optimum contribution selection algorithm for discrete generations [18] to restrict the rate of inbreeding. BLUP and SEGR achieved the restriction for all schemes, because EBV generally differ somewhat such that selection can be made easily. Another selection method, *e.g.* truncation selection on EBV, would yield different selection intensities, but differences in selection accuracy between BLUP-EBV and SEGR-EBV are probably similar to those in the current study with optimum contribution selection. Hence, the differences in selection response between BLUP-EBV and SEGR-EBV are probably similar for truncation and optimum contribution selection.

In contrast with previous studies on optimum contribution selection, *e.g.* [10, 11, 18], we assumed no total control of the selection scheme. Males were sampled, as in previous studies, according to their optimum contribution, but all females were selected. Also in these schemes, optimum contribution selection resulted in an effective selection scheme, and the restrictions on inbreeding were kept.

The method of Kerr and Kinghorn [14] was used to calculate genotype probabilities for individuals in the SEGR scheme. This iterative method handles the many loops in the pedigree. For the single gene model, SEGR has information on both the mode of inheritance and penetrance probability. BLUP only has information on the level of heritability. However, SEGR and BLUP give very similar genetic response and accuracy of selection (0.743 for the scheme with additive inheritance and 100 animals per generation). In the case of disease genes with recessive inheritance, SEGR was expected to be more accurate, because it accounts for the dominance effects. There was hardly an effect on the selection response, probably because selection was for

EBV, *i.e.* the additive part of the SEGR model. For the threshold model, there was iteration on the genotypes and penetrance probabilities of the genotypes, which was used to estimate EBV of the animals, although there was no single gene and thus no need to estimate the genotypes or penetrance probabilities. Their estimated effects are thus an artefact of the segregation analysis model. However, the resulting EBV yield very similar selection response as those of the BLUP model. SEGR only showed a small reduction in genetic response for the threshold model when the frequency of diseased animals was very low (not shown). Possibly, this reduction in the long-term genetic response for SEGR was because the SEGR model expected genetic variance to be reduced faster than it occurred in the threshold model. Generally, BLUP-EBV yield good approximations to SEGR-EBV in the single gene model, and SEGR-EBV yield good approximations to BLUP-EBV in the polygenic threshold model.

The BLUP-EBV evaluation did not account for the binary nature of the disease trait. For ease of computation, we approximated these non-linear EBV of binary traits by linear BLUP-EBV, which yield very similar selection response when the fixed effect structures are rather simple as in the case considered here [19].

4.3. Relaxation of assumptions

All dams were selected and only selection of sires was optimised. For a scheme, where also the selection of dams was optimised, we would expect a faster decrease of the frequency of the disease allele, such that fewer animals, which are heterozygous carriers, would be selected at the same ΔF. Most practical schemes have hierarchical mating systems with more dams than sires, implying that only males would be genotyped, because it is too expensive to also genotype all females.

We used an initial frequency of the disease allele of 0.5. In practice, the initial frequency of the disease allele will often be lower. We also simulated a scheme with $q_0 = 0.2$. Then, the evolution of the allele frequency under selection was approximated by the part of the curve of Figure 1, where the allele frequency is smaller than q_0 (not shown). Similar results were found for the threshold model. This suggests that our approach of studying time to 50% reduction is also applicable to other starting frequencies.

The schemes simulated here had a discrete generation structure. An overlapping generation structure would mainly have an effect for the SEGR and BLUP schemes, which would increase the accuracy of selection due to the increased number of offspring for some individuals.

When the disease trait is determined by an infectious agent, the probability of getting the disease depends not only on genotypes, but also on epidemiological parameters, that is how infectious the disease is, how many other animals are diseased etcetera. Different models have to be used for such traits.

Biallelic single genes were considered here. The results would probably be similar for a tri-allelic gene, but it would take more generations to make the population homozygous for the best of the three alleles.

In the current study, selection was entirely against the disease with either a single gene or threshold model. In general, the results can also be applied to major genes in mixed inheritance models, where a major gene has a large effect on the trait, and many background genes have small effects [4]. Villanueva *et al.* [23] simulated schemes with a mixed model inheritance for BLUP optimum contribution selection (similar to our BLUP scheme). They found that the frequency of the favourable allele had reached 0.95 after eight generations for a gene with recessive inheritance. This finding is similar to the results in Figure 1, although the schemes differed considerably (the genotypic values (2.0 and dominance deviation was −0.5), initial frequency (0.15) of the non-disease allele, and size of the scheme (120 animals per generation)). In the case where a major (disease) gene and other (economic) polygenic traits are included in the breeding goal, the question arises how much weight should be given to the major gene relative to that of the polygenic traits in the short and long term [4, 9,16,21,23]. This question is beyond the scope of this paper and is currently under active investigation [24,25].

The disease gene that was selected against was not lethal, because also individuals with the homozygous disease alleles could be selected and they would also survive. A reduction of fitness due to natural selection was not accounted for here, but should be accounted for in cases where the artificial selection against a disease allele is weak relative to the natural selection.

We assumed random mating among the selected animals. When the gene is unknown, non-random mating can also be used to test whether an animal is a heterozygous carrier of the disease allele, by mating it to a known homozygous (non-carrier) in a progeny-testing scheme. This is, however, an expensive and time-consuming strategy to reduce the frequency of the disease allele.

5. CONCLUSIONS

1. Selecting against a known genetic defect (GENO) is more efficient than selection against an unknown genetic defect, although the difference in efficiency was rather small (GENO reduced the frequency of the disease allele only 1–2 generations faster than BLUP and SEGR). Thus, selection for a genetic defect that is not identified at the DNA level is only slightly less efficient.

2. Knowledge about the mode of inheritance of the disease was rather unimportant, because assuming the wrong mode of inheritance hardly reduced the efficiency of the selection scheme.

3. Optimum contribution selection was able to maintain ΔF at the predefined level even when it was only controlling the selection of the sires and females

were randomly selected. Control of ΔF is, however, very important in small selection and conservation schemes.

REFERENCES

[1] Agerholm J.S., Bendixen C., Andersen O., Arnbjerg J., Complex vertebral malformation in Holstein calves, J. Vet. Diag. Invest. 13 (2001) 283–289.

[2] van Arendonk J.A.M., Smith C., Kennedy B.W., Method to estimate genotype probabilities at individual loci in farm livestock, Theor. Appl. Genet. 78 (1989) 735–740.

[3] Bijma P., Long-term genetic contributions – Prediction of rates of inbreeding and genetic gain in selected populations, Dissertation Wageningen University, Universal Press, Veenendaal, The Netherlands, 2000.

[4] Dekkers J.C.M., van Arendonk J.A.M., Optimizing selection for quantitative traits with information on an identified locus in outbred populations, Genet. Res. 71 (1998) 257–275.

[5] Elston R.C., Stewart J., A general model for the genetic analysis of pedigree data, Hum. Hered. 21 (1971) 523–542.

[6] Falconer D.S., Mackay T.F.C., Introduction to quantitative genetics, 4th edn., Longman Sci. and Tech., Harlow, UK, 1996.

[7] Fernando R.L., Stricker C., Elston R.C., An efficient algorithm to compute the posterior genotypic distribution for every member of a pedigree without loops, Theor. Appl. Genet. 88 (1993) 573–580.

[8] Gianola D., Heritability of polychotomous characters, Genetics 93 (1979) 1051–1055.

[9] Gibson J.P., Short term gain at the expense of long term response with selection on identified loci, in: Proceedings of the 5th World Congress on Genetics Applied to Livestock Production, 7–12 August 1994, Vol. 21, University of Guelph, Guelph, pp. 201–204.

[10] Grundy B., Villanueva B., Woolliams J.A., Dynamic selection procedures for constrained inbreeding and their consequences for pedigree development, Genet. Res. 72 (1998) 159–168.

[11] Grundy B., Villanueva B., Woolliams J.A., Dynamic selection for maximizing response with constrained inbreeding in schemes with overlapping generations, Anim. Sci. 70 (2000) 373–382.

[12] Henderson C.R., Applications of linear models in animal breeding, University of Guelph Press, Guelph, Canada, 1984.

[13] Janss L.L.G., van Arendonk J.A.M., van der Werf J.H.J., Computing approximate monogenic model likelihood in large pedigrees with loops, Genet. Sel. Evol. 27 (1995) 567–579.

[14] Kerr R.J., Kinghorn B.P., An efficient algorithm for segregation analysis in large populations, J. Anim. Breed. Genet. 113 (1996) 457–469.

[15] Kinghorn B.P., Use of segregation analysis to reduce genotyping costs, J. Anim. Breed. Genet. 116 (1999) 175–180.

[16] Larzul C., Manfredi E., Elsen J.M., Potential gain from including major gene information in breeding value estimation, Genet. Sel. Evol. 29 (1997) 161–184.

[17] Lust G., An overview of the pathogenesis of canine hip dysplasia, J. Am. Vet. Med. Assoc. 210 (1997) 1443–1445.
[18] Meuwissen T.H.E., Maximising the response of selection with a predefined rate of inbreeding, J. Anim. Sci. 75 (1997) 934–940.
[19] Meuwissen T.H.E., Engel B., van der Werf J.H.J., Maximising selection efficiency of categorical traits, J. Anim Sci. 73 (1995) 1933–1939.
[20] Meuwissen T.H.E., Sonesson A.K., Maximizing the response of selection with a predefined rate of inbreeding: overlapping generations, J. Anim. Sci. 76 (1998) 2575–2583.
[21] Ruane J., Colleau J.J., Marker assisted selection for genetic improvement of animal populations when a single QTL is marked, Genet. Res. Camb. 66 (1995) 71–83.
[22] Sonesson A.K., Meuwissen T.H.E., Minimization of rate of inbreeding for small populations with overlapping generations, Genet. Res. Camb. 77 (2001) 285–292.
[23] Villanueva B., Pong-Wong R., Grundy B., Woolliams J.A., Potential benefit from using an identified major gene in BLUP evaluation with truncation and optimal selection, Genet. Sel. Evol. 31 (1999) 115–133.
[24] Villanueva B., Pong-Wong R., Woolliams J.A., Settar P., Maximising genetic gain with QTL information and control of inbreeding, in: Proceedings of the 7th World Congress on Genetics Applied to Livestock Production, 19–23 August 2002, Montpellier, CD-ROM communication No. 22–18.
[25] Villanueva B., Pong-Wong R., Woolliams J.A., Marker assisted selection with optimised contributions of the candidates to selection, Genet. Sel. Evol. 34 (2002) 679–703.

To access this journal online:
www.edpsciences.org