Managing inbreeding in selection and genetic conservation schemes of livestock

Promotor

Prof. dr. ir. E.W. Brascamp
Hoogleraar Fokkerij en Toegepaste Genetica
Wageningen Universiteit

Co-promotor

Dr. ir. T.H.E. Meuwissen
Senior Wetenschapper
Instituut voor Dierhouderij en Diergezondheid

Samenstelling promotiecommissie

Prof. dr. ir. J.A.M van Arendonk
Wageningen Universiteit

Prof. R.A. Cardellino
FAO, Italy

Dr. M.A. Toro
INIA, Spain

Dr. ir. L. Visser
Plant Research International

STELLINGEN

1. An optimal mix of a high response to selection with constrained increase of inbreeding by an optimal contribution of parents to the next generation can be further improved by a proper mating of parents (this thesis).
2. Well designed genetic conservation programs can conserve most of the genetic diversity in small populations (this thesis).
3. Despite the difficulties in defending the assumptions underlying linear models, such as BLUP breeding values for binary traits, these models are very useful and thus abundantly used.
4. The future key issue in animal breeding will be to change the focus from product quantity to product quality.
5. Diversity in research groups is at least as important for the fitness of the group as genetic diversity is important for animal breeding populations.
6. Stellige stellingen stellen zelden iets voor.

Stellingen behorende bij het proefschrift van Anna Sonesson
'Managing inbreeding in selection and genetic conservation schemes of livestock'
Wageningen 14 juni 2002

Managing inbreeding in selection and genetic conservation schemes of livestock

Anna Kristina Sonesson

Proefschrift

ter verkrijging van de graad van doctor
op gezag van de rector magnificus
van Wageningen Universiteit,
Prof. dr. ir. L. Speelman,
in het openbaar te verdedigen
op vrijdag 14 juni 2002
des namiddags te vier uur in de Aula.

Managing inbreeding in selection and genetic conservation schemes of livestock

Anna Kristina Sonesson, 2002

Ph.D. thesis Animal Breeding and Genetics group, Wageningen University, Wageningen and Division of Animal Science, Insitute for Animal Science and Health, Lelystad.

ISBN: 90-5808-663-1

Abstract. This thesis deals with the definition of selection and mating criteria for animal breeding populations under selection and for genetic conservation populations, especially emphasizing on populations with small effective size and with known pedigree. For populations under selection with overlapping generations, a method was developed that maximizes genetic response with a restriction on the rate of inbreeding (ΔF). At the same ΔF, the genetic response was up to 44% higher for the presented method than for truncation selection on BLUP breeding values. In combination with the presented selection method, non-random mating systems could further increase genetic response. For genetic conservation schemes, a method was developed that minimizes rate of inbreeding for populations with overlapping generations. For very small schemes, ΔF was 18-52% higher when simply selecting the oldest animals than for the proposed scheme. Use of frozen semen from sires of the oldest generations could reduce ΔF to zero in combined *in situ/ex situ* conservation schemes.

The research described in this thesis was financially supported by the Ministry of Agriculture, Nature Management and Fisheries, The Netherlands.

Preface

The work of this thesis has been conducted at the Institute of Animal Science and Health in collaboration with the Animal Breeding and Genetics group at Wageningen University. I would like to mention some of the persons that have stood behind or beside me along the road.

Thanks to Pim Brascamp for offering a helicopter perspective on my research, which often resulted in improved structure of the papers and thesis. Thanks to Kor Oldenbroek for giving me the opportunity to do this work. Thanks to Theo Meuwissen for the daily guidance and for being my best friend. From my guidance group, I would especially like to mention Piter Bijma for the genuine interest and all suggestions to the papers and thesis. Thanks to Ab Groen for comments on the papers and to John Woolliams for excellent suggestions and constructive discussions about inbreeding.

Thanks to all nice staff at the Animal Genetics and Breeding Unit in Armidale and at the Victorian Institute of Animal Science in Melbourne, especially to Mike Goddard who invited me to Australia.

Thanks for the introduction to Dutch life and "gezelligheid" by the former and current colleagues at 19:2 and the rest of my cluster.

And at last, thanks to my family and friends.

Anna

Contents

Chapter 1

General Introduction

INTRODUCTION

This thesis deals with the definition of selection and mating criteria for animal breeding populations under selection and genetic conservation populations, especially emphasizing on populations with small (approximately 50 (FAO, 1998)) effective sizes that have known pedigrees. The effective population size (N_e) is defined as $N_e = ½ΔF$, where $ΔF$ is the rate of inbreeding, such that the selection and mating criteria will deal with the management of $ΔF$. For selected populations, genetic gain will be maximized with a restriction on $ΔF$. For genetic conservation schemes, $ΔF$ will be minimized in order to maximize the effective population size.

The $ΔF$ needs to be managed in selection and genetic conservation schemes, because the genetic makeup changes irreversibly by two processes in finite and closed populations. Firstly, the frequency of heterozygote animals decreases on average by $2pqF$ at any locus with two alleles, where p and q are the frequencies of the two alleles in the base population and F is the average level of inbreeding of the population. A decrease of heterozygosity leads to inbreeding depression for traits affected by directional dominance. Since directional dominant alleles are often related to fitness traits, a lower fitness is expected when inbreeding increases. Secondly, random drift increases, which increases the probability of loosing favorable and unfavorable alleles. Random drift also induces variation of genetic levels for selected and non-selected traits. The only source of new variation in closed populations is by mutations. Mutational variation is however small, about .1 to .5% of the environmental variance and mutations are mostly detrimental with respect to fitness (Falconer and Mackay, 1996; Hill, 2000).

In randomly selected and mated populations with Poisson family size and discrete generations, $ΔF$ equals $1/(8N_s) + 1/(8N_d)$, where N_s and N_d refer to the number of selected sires and dams (Wright, 1931). However, the management of $ΔF$ in selected populations is more complicated than simply choosing N_s and N_d, because selection and mating are often non-random processes and generations are overlapping in most practical populations. In selected populations, $ΔF$ can exceed $1/(8N_s) + 1/(8N_d)$ by a factor 1-3 for schemes with high intensity of selection and in high family weights (Woolliams and Bijma, 2000).

Several methods have been proposed to reduce inbreeding in selection schemes. Some methods attempt to reduce the selection criterion of individuals, *e.g.* by reducing the weight of between-family information

(Dempfle, 1975; Grundy and Hill, 1993; Verrier *et al.*, 1993). However, these individual-based methods do not consider the selection candidate in relation to the other selected animals. Woolliams and Meuwissen (1993) introduced the group profit function that maximizes genetic gain with a cost factor on the prediction error variance of the average genetic level of the group. Wray and Goddard (1994) and Brisbane and Gibson (1995) maximized genetic gain with a cost factor on the average relationship of the selected animals. The above individual and group based methods do, however, not control the inbreeding, *i.e.* high inbreeding is still accepted if genetic gain is high. Meuwissen (1997) derived a group based method that maximizes genetic response while restricting the average coancestry of the selected animals, and thus the average inbreeding of the offspring, to a predefined rate for populations with discrete generations. The genetic contribution of each candidate was optimized each generation with the Lagrangian multiplier method for populations with discrete generations. The use of the Lagrangian multiplier method implies that the genetic contributions of the method of Meuwissen (1997) are close to optimum.

This thesis will extend the theory behind the method of Meuwissen (1997) to several applications that all consider the management of rate of inbreeding in selection and genetic conservation schemes.

OUTLINE OF THE THESIS

The thesis can be divided into four main parts. Firstly, Chapters 2 and 3 deal with selection algorithms that manage ΔF for populations under selection with overlapping generations and random mating. Secondly, Chapters 4 and 5 deal with non-random mating schemes in combination with selection algorithms for discrete and overlapping generation structures, respectively. Thirdly, Chapters 6 and 7 deal with algorithms that minimize ΔF for small and endangered populations with overlapping generations and when frozen semen of sires from the base population is available, respectively. Fourthly, Chapter 8 deals with alternative methods to select against genetic defects while restricting ΔF in populations with increased frequency of diseased alleles.

CHAPTERS 2 AND 3

In Chapter 2, the group selection index of Meuwissen (1997) is extended to account for populations with overlapping generations and progeny testing.

Populations with overlapping generations have selection candidates from several reproductive age-classes, and therefore also complex pedigrees. Thus, the selection algorithm has to take account of previous and future use of the selection candidates. Genetic gain is maximized with restrictions on ΔF and on the genetic contribution per sex. The genetic contribution of each selection candidate is optimized using Lagrangian multipliers. In Chapter 3, the method of Chapter 2 is compared to similar method by Grundy *et al.* (2000), where the contribution of each age-class is explicitly optimized.

CHAPTERS 4 AND 5

In Chapter 4, the effect of non-random mating methods on ΔF and ΔG are compared to random mating in combination with the selection method of Meuwissen (1997). In truncation selection schemes, non-random mating systems have been shown to improve the family structure, thereby reducing ΔF (Caballero *et al.*, 1996), but the selection method of Meuwissen (1997) restricts ΔF, such that no change in ΔF is expected. Considered non-random mating systems are minimum coancestry mating that minimizes the average coancestry of the mating pairs, compensatory mating that connects families with many selected animals to families with few selected animals (Santiago and Caballero, 1995), and factorial mating that restricts the number of full-sibs to zero or one (Woolliams, 1989). In Chapter 5, non-random mating systems are compared for the optimum contribution selection method with overlapping generations, presented in Chapter 2.

CHAPTERS 6 AND 7

In Chapter 6, an algorithm is presented that minimizes ΔF for small and endangered populations with overlapping generations, but without aiming for genetic gain. It minimizes the average coancestry of selection candidates in the form of a group index with a restriction on the genetic contribution per sex. The genetic contribution of each selection candidate is optimized with Lagrangian multipliers. The same method is used in Chapter 7 to minimize ΔF when frozen semen is available, which results in a combined *in situ* and *ex situ* genetic conservation scheme.

CHAPTER 8

In Chapter 8, populations with an increased frequency of a diseased allele are considered. The efficiency of selection against the disease allele is compared for alternative genetic models and evaluation methods, where the genetic evaluation method does not always agree with the true inheritance of the disease. A genetic model for a single gene is compared to a threshold model, where many genes and the environment affect the liability of an animal to become diseased. Selection on breeding values estimated with BLUP (Henderson, 1984) is compared to selection on breeding values estimated by segregation analysis (Elston and Stewart, 1971) and to selection on DNA-genotypes for a known disease gene. The optimum contribution selection method for discrete generations of Meuwissen (1997) is used to select against the disease.

LITERATURE

Brisbane, J.R. and Gibson, J.P. (1995) Balancing selection response and inbreeding by including predicted stabilized genetic contributions in selection decisions. Genet. Sel. Evol. 27: 541-549.

Caballero, A., Santiago, E. and Toro, M.A. (1996) Systems of mating to reduce inbreeding in selected populations. Anim. Sci. 62: 431-442.

Dempfle, L. (1975) A note on increasing the limit of selection through selection within families. Genet. Res. Camb. 24: 127-135.

Elston, R.C. and Stewart, J. (1971) A general model for the genetic analysis of pedigree data. Hum. Hered. 21: 523-542.

FAO (1998) Secondary guidelines for development of national farm animal genetic resources management plans, FAO, Rome, Italy.

Falconer, D.S. and Mackay, T.F.C. (1996) Introduction to quantitative genetics. 4th Ed. Longman Sci. and Tech., Harlow, UK.

Grundy, B. and Hill, W.G. (1993) A method for reducing inbreeding with Best Linear Unbiased Prediction. Anim. Prod. 56: 427.

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

Henderson, C.R. (1984) Applications of Linear Models in Animal Breeding, University of Guelph Press, Guelph, Canada.

Hill, W.G. (2000) Maintenance of genetic variation in animal breeding programmes. Livest. Prod. Sci. 63: 99-109.

Meuwissen, T.H.E. (1997) Maximising the response of selection with a predefined rate of inbreeding. J. Anim. Sci. 75: 934-940.

Santiago, E., and Caballero, A. (1995) Effective size of populations under selection. Genetics 139: 1013-1030.

Verrier, E., Colleau, J.J. and Foulley, J.L. (1993) Long-term effects of selection based on the animal model BLUP in a finite population. Theor. Appl. Genet. 87: 446-454.

Woolliams, J.A. (1989) Modifications to MOET nucleus breeding schemes to improve rates of genetic progress and decrease rates of inbreeding in dairy cattle. Anim. Prod. 49: 1-14.

Woolliams, J.A. and Meuwissen, T.H.E. (1993) Decision rules and variance of response in breeding schemes. Anim. Prod. 56: 179-186.

Woolliams, J.A. and Bijma, P. (2000) Predicting rates of inbreeding in population undergoing selection. Genetics 154: 1851-1864.

Wray, N.R. and Goddard, M.E. (1994) Increasing long-term response to selection. Genet. Sel. Evol. 26: 431-451.

Wright, S. (1931) Evolution in Mendelian populations. Genetics 16: 97-159.

Chapter 2

Maximizing the response of selection with a predefined rate of inbreeding- overlapping generations

T.H.E. Meuwissen and A.K. Sonesson

Institute for Animal Science and Health,
P.O. Box 65, 8200 AB Lelystad, The Netherlands

Published in Journal of Animal Science (1998) 76: 2575-2583
Reproduced by permission of American Society of Animal Science

ABSTRACT

In a breeding scheme, the aim is high rates of genetic gain with limited inbreeding. A dynamic selection rule is developed that maximizes selection response in populations with overlapping generations. The rule maximizes the genetic merit of selected animals while limiting the average relationship of the population after the current round of selection. The latter is shown to limit the contribution of the current population to the future inbreeding. The rule accounts for the selection of some candidates during previous selection rounds and for the expected future contributions of the selection candidates. Inputs for the rule are the BLUP breeding values and ages of selection candidates, the relationship matrix of all animals and contributions of animals during previous selection rounds. Output is the optimal number of offspring for each candidate. Computer simulations of dairy cattle nucleus schemes showed that predefined rates of inbreeding were actually achieved, without compromising long-term selection response, at least up to 20 yr of selection. At the same rates of inbreeding, the dynamic selection rule obtained up to 44% more genetic gain than direct selection for BLUP breeding values. The advantage of the dynamic rule over BLUP selection decreased with increasing population sizes and with higher predefined rates of inbreeding. Consequently, the dynamic rule should be especially useful in small selection schemes where relatively low rates of inbreeding are desired.

1. INTRODUCTION

Breeding schemes are usually designed to maximize genetic gain while limiting the rate of inbreeding. Rates of inbreeding are reduced by simply selecting more sires (and dams), by selection for EBV(h_a^2), where EBV are estimated using a heritability of h_a^2 and the true heritability is h^2 and $h_a^2 > h^2$ (Grundy and Hill, 1993), by increasing the weight of the within family component of the EBV (Verrier *et al.*, 1993), by using a selection index with a cost on the average relationships (Wray and Goddard, 1994; Brisbane and Gibson, 1995), and by selection with a constraint on the rate of inbreeding (Meuwissen, 1997). The latter two methods maximize genetic response given the rate of inbreeding that they achieve, and they are dynamic selection rules. A dynamic selection rule optimizes selection given the available selection candidates, whereas a static selection rule optimizes the breeding scheme beforehand (*i.e.*, for an average group of selection candidates) (Goddard and Howarth, 1994). Villanueva and Woolliams (1997) suggested a static rule for restricting rates of inbreeding by predefining the numbers of sires and dams selected and the selection index weights.

The dynamic selection rules of Wray and Goddard (1994), Brisbane and Gibson (1995) and Meuwissen (1997) are not strictly applicable to situations with overlapping generations. They reduce or constrain the relationships among selected parents, but, when generations overlap, some parents may have already produced many offspring during previous rounds of selection, and the dynamic rules do not account for this. The aim of this paper is to extend the above dynamic selection rules to situations with overlapping generations. In particular, the selection rule of Meuwissen (1997) will be extended to populations with overlapping generations, and, from this, the extension of the rules of Wray and Goddard (1994) and Brisbane and Gibson (1995) will be shown in the **DISCUSSION** section.

2. METHODS

2.1. The contribution of current to future relationships

In a population, inbreeding increases on average with half the increase of the average relationship. Following Wray and Goddard (1994), Brisbane and Gibson (1995), and Meuwissen (1997), inbreeding will be limited here by reducing the increase of average relationship. If relationships of an animal with itself are included, the average relationship is:

$$\overline{A} = \mathbf{1}'\mathbf{A}\mathbf{1}/n^2,$$

where $\mathbf{A}$ = matrix of additive genetic relationships, n = number of animals; $\mathbf{1}$ = vector of ones. If genetic variance is assumed 1, $Var(\mathbf{u}) = \mathbf{A}$. Thus $\overline{A} = Var(\overline{\mathbf{u}})$, where $\mathbf{u}$ = vector of true breeding values, and $\overline{\mathbf{u}} = \mathbf{1}'\mathbf{u}/n$ is the average true breeding value. Following Johnson (1977), a recursive formula for $\overline{\mathbf{u}}$ will be derived in order to obtain a formula for $\overline{A}$ in the situation where generations overlap.

At any time, t, there are animals of different ages in a population with overlapping generations. The animals will be divided into age classes, where an age class is defined as the time period between two consecutive rounds of selection. For convenience, this time period will be assumed to equal 1 yr here, but other time periods can be assumed without affecting the derivation. Let $\overline{\mathbf{u}}_t = [\overline{\mathbf{u}}_t(1)\ \overline{\mathbf{u}}_t(2)\ \overline{\mathbf{u}}_t(3)\ldots]'$ be a vector of mean true breeding values of the age classes, 1, 2, 3, ..., in year *t*, where $\bar{\mathbf{u}}_t$ (i) = average breeding value of age class *i* in year *t*. The following recursive relationship exists between $\bar{\mathbf{u}}_t$ and $\bar{\mathbf{u}}_{t+1}$ (Hill, 1974):

$$\bar{\mathbf{u}}_{t+1} = \mathbf{P}\bar{\mathbf{u}}_t + \mathbf{e}_{t+1}, \qquad [1]$$

where **P** = matrix of gene flow between the age classes (see Hill, 1974, for details), and $\mathbf{e}_{t+1}$ = sampling deviation from the expected breeding values, *i.e.*, the $\mathbf{P}\bar{\mathbf{u}}_t$ term. It will be assumed that the gene flow matrix, **P**, is the average gene flow matrix over years, and the derivation is simplified by considering only age classes instead of age × sex classes as in Hill (1974). The vector $\mathbf{e}_t$ represents the deviation from predictions based on this average gene flow matrix.

In the following, the contribution of relationships in year *t* to a future year *t+s* will be derived, where *s* is assumed to be large. Hence, the contribution of current relationships to long-term future relationships will be derived. From Equation [1], it follows that the vector of average true breeding values in year *t+s* is:

$$\bar{\mathbf{u}}_{t+s} = \mathbf{P}^s\bar{\mathbf{u}}_t + \mathbf{P}^{s-1}\mathbf{e}_{t+1} + \mathbf{P}^{s-2}\mathbf{e}_{t+2} + \ldots + \mathbf{e}_{t+s} = \mathbf{P}^s\bar{\mathbf{u}}_t + \sum_{i=1}^{s}\mathbf{P}^{s-i}\mathbf{e}_{t+i}\ , \qquad [2]$$

where $\mathbf{P}^s$ denotes **P** to the power *s*. Let $\bar{\mathbf{A}}_t = Var(\bar{\mathbf{u}}_t)$ be a matrix with the average relationships within age classes on the diagonals and between age classes on the off-diagonals. From Equation [2],

$$\bar{\mathbf{A}}_{t+s} = Var(\bar{\mathbf{u}}_{t+s}) = \mathbf{P}^s\bar{\mathbf{A}}_t\mathbf{P}'^s + \sum_{i=1}^{s}\mathbf{P}^{s-i}\mathbf{V}_e\mathbf{P}'^{s-i}\ , \qquad [3]$$

where $\mathbf{V}_e$ = variance of the $\mathbf{e}_t$ vectors. In Equation [3], the term $\mathbf{P}^s\bar{\mathbf{A}}_t\mathbf{P}'^s$ shows the contribution of average relationships in year *t* to those in year *t+s*, and the second term shows the contributions from later years to the relationships in year *t+s*. Hill (1974) showed that, for large *s*:

$$\underset{s\to\infty}{Lim}\,\mathbf{P}^s = \mathbf{1}\mathbf{r}',$$

where $\mathbf{r}$ = vector with element i equal to $r(i)=\left[\sum_{j=1}^{q} P(1,j)\right]/\overline{L}$, where q = number of age classes in **P**, (*i.e.*, the dimension of **P**), and $\overline{L}$ = average generation interval of the sires and the dams. In words, $r(i)$ equals the current plus future contributions of age class i until it dies, (*i.e.*, age class q is reached), divided by the generation interval. Hence, for large s, the term $\mathbf{P}^s \overline{\mathbf{A}}_t \mathbf{P}'^s$ in Equation [3] becomes

$$\lim_{s\to\infty} \mathbf{P}^s \overline{\mathbf{A}}_t \mathbf{P}'^s = \mathbf{1}\mathbf{r}'\overline{\mathbf{A}}_t\mathbf{r}\mathbf{1}' = \mathbf{1}\mathbf{1}' \times \mathbf{r}'\overline{\mathbf{A}}_t\mathbf{r},$$

where $\mathbf{1}\mathbf{1}'$ = $q \times q$ matrix of ones and $\mathbf{r}'\overline{\mathbf{A}}_t\mathbf{r}$ is a scalar that indicates the contribution of average relationships in year t to average relationships many years after year t. When selecting parents in year t-1, we want to limit the future increase of average relationships, which is limited by constraining the $\mathbf{r}'\overline{\mathbf{A}}_t\mathbf{r}$ term of year t.

2.2. A constraint on $\mathbf{r}'\overline{\mathbf{A}}_t\mathbf{r}$

In this section, it will be assumed that the distribution of the parents over the age classes is known, (*i.e.*, $\mathbf{r}$ is known). Optimization of $\mathbf{r}$ is considered in the next section. The population of year t is formed from that in year t-1. Age classes 2, 3,…, q of year t are formed by ageing, from the age classes 1, 2, ..., q-1, respectively, in year t-1. It is assumed here that an age class consists of both culled and nonculled animals, which implies that culling does not affect the number of animals nor relationships within an age class. Age class 1 is newly formed in year t, and is due to selection in year t-1. It is useful to partition the term $\mathbf{r}'\overline{\mathbf{A}}_t\mathbf{r}$ into components due to ageing that are not affected by the selection, and components that are determined by selection in year t-1:

$$\mathbf{r}'\overline{\mathbf{A}}_t\mathbf{r} = \mathbf{r}_a^2 \overline{A}_{t(a,a)} + 2\, r_a \overline{\mathbf{A}}_{t(a,b)}\, \mathbf{r}_b + \mathbf{r}_b' \overline{\mathbf{A}}_{t(b,b)}\, \mathbf{r}_b \quad , \qquad [4]$$

where the submatrices are defined by:

$$\mathbf{r} = \begin{bmatrix} r_a & \mathbf{r}_b \end{bmatrix} \text{ and } \overline{\mathbf{A}}_t = \begin{bmatrix} \overline{A}_{t(a,a)} & \overline{\mathbf{A}}_{t(a,b)} \\ \overline{\mathbf{A}}_{t(b,a)} & \overline{\mathbf{A}}_{t(b,b)} \end{bmatrix},$$

where subscript a refers to the element 1 and b to the elements 2, …, q. The $\bar{\mathbf{A}}_{t(b,b)}$ term in Equation [4] does not depend on the selected parents, because animals in age classes 2 to q in year t are identical to those in age classes 1 to q-1 in year t-1. Hence, $\bar{\mathbf{A}}_{t(b,b)}$ equals the (1:q-1; 1:q-1)-block of matrix $\bar{\mathbf{A}}_{t-1}$, and $\mathbf{r}_b'\bar{\mathbf{A}}_{t(b,b)}\mathbf{r}_b$ in [4] is calculated from $\bar{\mathbf{A}}_{t-1}$ and $\mathbf{r}_b$. The terms $r_a^2\bar{A}_{t(a,a)}$ and $r_a\bar{\mathbf{A}}_{t(a,b)}\mathbf{r}_b$ are affected by selection of parents, where the former term represents the average relationships among new progeny and the second term represents the average relationships of new progeny with other older animals (age classes 2 to q).

In Equation [4], the average relationships in age class 1, $\bar{A}_{t(a,a)}$, equal the average relationships of selected parents of year t-1, if the average relationships within a group do not include the relationships of an animal with itself, as will be assumed in the following. It follows that:

$$\bar{A}_{t(a,a)} = \mathbf{c}_{t-1}'\mathbf{A}_{t-1}\mathbf{c}_{t-1}, \qquad [5]$$

where $\mathbf{A}_{t-1}$ = relationship matrix among the animals in year t-1; $\mathbf{c}_{t-1}$ = vector of genetic contributions of animals in year t-1 to age class 1 in year t (contributions of animals not selected are 0; and male [female] contributions sum to ½). Also the term $\bar{\mathbf{A}}_{t(a,b)}$ in Equation [4] depends on selected parents in year t-1:

$$\bar{\mathbf{A}}_{t(a,b)} = \mathbf{c}_{t-1}'\mathbf{A}_{t-1}\mathbf{J}, \qquad [6]$$

where $\mathbf{c}_{t-1}'\mathbf{A}_{t-1}$ = row vector of average relationships of the selected group with every individual animal; and the $nq \times q$ matrix $\mathbf{J}$ averages these relationships of the individual animals within every age class, (*i.e.*, the j$^{\text{th}}$ column of $\mathbf{J}$ has the n elements that correspond to animals in age class j equal to $1/n$ and all other elements equal to zero, where n = number of animals per age class).

If the contribution of the population of year t to future relationships, (*i.e.*, $\mathbf{r}'\bar{\mathbf{A}}_t\mathbf{r}$) is to be constrained, we have to constrain (combining Equations [4], [5], and [6]):

nucleus. Genotypes, g_i, of the base animals were sampled from the distribution $N(0,.3)$. Dams had up to three records, where the j^{th} record was calculated as $y_{ij} = g_i + p_i + e_{ij}$, where p_i = permanent environmental effect sampled from $N(0, .2)$ (p_i was identical for all three records), and e_{ij} = temporary environmental effect sampled from $N(0, .5)$. Later years were obtained by simulating offspring genotypes from $g_i = \frac{1}{2}g_s + \frac{1}{2}g_d + m_i$, where s and d denote the sire and dam of the offspring i, respectively, and m_i = Mendelian sampling component which was sampled from $N(0, .15)$.

Table 1. Parameters of the closed nucleus breeding schemes.

Constraint on inbreeding	.50 or .25 %/yr
No. of new progeny per yr (males and females)	256 or 512
Size of unrelated base population	5 x (No. of new progeny)
No. of years evaluated	20
Involuntary culling rate of males and females	.3
Voluntary culling rate	negligible
Age at which females completed lactation records	2,3 and 4 yr [a]
No. of test daughters of bulls with unrelated cows outside the nucleus	0 or 100
Age at which progeny test became available	5 yr [a]
Reproductive rate of males and female within nucleus	unlimited
No. of sires and dams selected in BLUP selection schemes	equal such that inbreeding constraint holds
Genetic, and permanent and temporary environmenal variances	.3, .2, .5

[a] When the animals are selected for this information, the offspring are born 1 yr later (*i.e.*, the generation interval is 1 yr longer than the age at which the information becomes available)

Note that the variance of the Mendelian sampling effect was not reduced by inbreeding, (*i.e.*, genetic variance is not reduced by inbreeding). Hence, selection response is expected to asymptote to a constant value over years, which provides a test whether or not the OC method can maintain a constant selection response over time. If the Mendelian sampling variance was reduced by inbreeding, the selection response would decrease over time. This decrease of the selection response would be confounded with any decrease of the selection response, when the OC method would not be able to maintain its initial selection response. One of the goals of the simulation study was to detect any reduction in selection response over time of the OC method.

The number of offspring that a dam could produce was assumed unlimited, because of the use of new reproductive techniques such as ova pick up, *In vitro* maturation and fertilization (Kruip, *et al.*, 1994). Hence, one dam could produce a maximum of 256 or 512 offspring per yr, (*i.e.*, the total number of new progeny of the nucleus), but this number is reduced when more dams are selected because of the inbreeding restriction. A dam was mated to several sires to produce these offspring, and because mating was at random, a random sire was allocated to every offspring of the dam, where every sire had a probability of $2c_{t-1}(i)$ of being allocated, where $c_{t-1}(i)$ was the optimal contribution of animal *i*. In order to account for some variability in success of the female reproductive techniques, the number of offspring that a dam obtained varied around the optimal number, which is $2 \times c_{t-1}(i) \times$ (total number of new progeny in the nucleus). The variation in number of progeny of dams followed a multinomial distribution with $2c_{t-1}(i)$ being the probability that dam i produced an offspring.

For comparison, schemes were simulated with selection for BLUP-EBV, where every selected bull and dam had an equal probability of producing an offspring. The number of bulls and dams selected was assumed equal and was determined by trial and error such that the inbreeding constraint was also achieved in the BLUP schemes. In BLUP schemes, optimal unequal numbers of sires and dams can be selected, but finding the optimum BLUP scheme would require simulation of very many BLUP schemes, (*i.e.*, every possible combination of a number of sires and a number of dams selected). Because of the large amount of computer time involved, this optimization was not attempted. Equal numbers of sires and dams selected seems a reasonable starting point, because it is in between selecting fewer sires than dams, which is common in practice, and the selection of more sires than dams, as was found optimal in closed nucleus schemes with high female reproductive rates (De Boer *et al.*, 1994).

The inbreeding constraint was either .50 or .25% per yr for both the BLUP and OC schemes, (*i.e.*, the desired inbreeding coefficient after 20 yr was .10 or .05, respectively). The number of replicated simulations was 50, except for the large scheme (512 new progeny per yr) with OC selection where the number of replicated simulations was 25, because of the large amount of computer time needed for these schemes. Mean and standard errors of genetic levels and inbreeding levels after 20 yr of selection were calculated for 50 or 25 replicated simulations.

3. RESULTS

In Figure 1, the average genetic level and inbreeding coefficients of the new progeny are shown with the OC and BLUP methods for populations with 256 new progeny per year and no progeny-testing. In the first year, coancestry between new progeny was relatively high because the method constrained average coancestry of the entire population, and the rest of the population was still unrelated. This also allowed for the large selection response of the OC-method during the first year. When new progeny of yr 1, which had a high genetic level, were available for selection (as parents of new progeny of yr 3), few animals could be selected from them because they were so highly related. Hence, new progeny in yr 3 had approximately the same genetic level as those in yr 1. The same arguments hold for new progeny in yr 2 and 4, and this explains why the OC method yielded almost no selection response during yr 2 to 4. After these initial years, the method yielded a constant selection response over years, which shows that the long-term performance of the OC selection method is as good as its short-term performance. However, this constant response was only possible because the effect of inbreeding on the Mendelian sampling variances was ignored in the simulations. If this effect of inbreeding had not been ignored, the long-term response would decrease owing to the reduction of the genetic variance, but it would not decrease more than expected based on the reduced genetic variance.

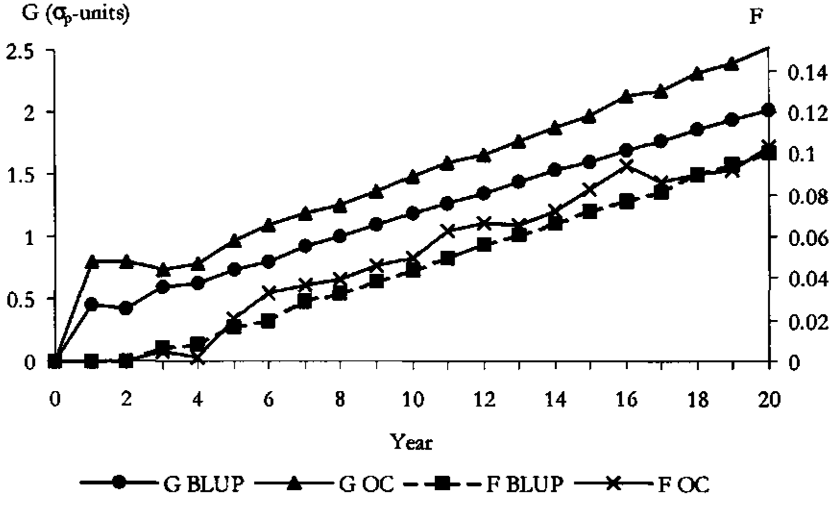

Figure 1. Genetic level (G) and inbreeding coefficient (F) for the optimal contribution (OC) and BLUP-EBV (BLUP) selection of 64 sires and 64 dams. Averages of 50 simulations with 256 new progeny per year without progeny-test of young bulls.

Table 2 shows the genetic level at yr 20 for the populations that were simulated with the OC and the BLUP methods. In all simulations, the realized inbreeding coefficient was around the value to which it was constrained. For the populations to which 256 animals were born and ΔF was constrained to .0050 per year, the OC method reached 25% higher genetic level than the BLUP method without progeny-tests and a 36% higher genetic level when the populations were progeny-tested. With a constraint on ΔF of .0025 per year, these figures were 37 and 44%, respectively. Note that these increases of genetic gain are partly because the BLUP schemes were not optimized for the number of sires and dams selected, and they should be seen as an upper limit for the increase of the response when starting from an optimized BLUP scheme. In schemes with discrete generations, Meuwissen (1997) found also that the advantage of the OC method increased when the restriction on ΔF became more stringent.

For the larger populations with 512 animals born and ΔF constrained to .0050 per year, the OC method yielded more genetic gain than did BLUP selection: 16 and 27% for schemes without or with progeny-testing, respectively. Hence, the advantage of OC over BLUP selection decreased with increasing population size, and increased when young bulls were progeny-tested. With ΔF constrained to .0025, these figures were 29 and 38%, respectively. Again, a more stringent constraint on the rate of inbreeding increased the extra gain of the OC method.

Table 3 shows the number of animals selected and the generation intervals with OC and BLUP selection. At the same levels of inbreeding, the OC method selected many fewer animals than BLUP selection. This indicated that OC selection achieved its higher genetic gains by realizing a higher selection differential at the same rate of inbreeding. Without progeny-testing, the OC method selected dams more intensely than sires, because this combined the higher intensity with the higher accuracy of selection of the cows. De Boer *et al.* (1994) found also that the selection intensity of the dams was higher than that of the sires when optimizing a nucleus scheme without progeny-testing of young bulls.

If young bulls were progeny-tested, the OC method selected mainly progeny-tested bulls, whereas the BLUP method continued to select mainly not progeny-tested bulls (Table 2). The latter occurred because the best young bulls had higher EBV than the best progeny-tested bulls in the BLUP schemes, in which the selection intensity of bulls was not very high and genetic gain was substantial, (*i.e.*, the average EBV of young bulls was substantially higher than that of progeny-tested bulls). In the OC schemes, the intensity of selection of the bulls could be and was increased, because the longer generation interval had a decreasing effect on the rate of inbreeding per year. Hence, the OC method combined the higher intensity with the higher accuracy of selection of progeny-tested bulls. Because the OC method took greater advantage of the availability of progeny-tested bulls than did BLUP selection, its superiority over BLUP selection was substantially higher when young bulls were progeny-tested (see Table 2).

Table 2. Genetic level (G) and inbreeding coefficients (F) at yr 20 when nucleus herds were selected with the optimal contribution method and with selection for BLUP-EBV. [a]

		Optimal Contribution		**BLUP**	
New borns per year (no of animals)	**Size of progeny test (no of records)**	**G (σ_p-units)**	**F**	**G (σ_p-units)**	**F**
	Constraint on ΔF per year = .0050				
256	**0**	2.52	.10	2.01	.10
	100	3.12	.08	2.30	.09
512	**0**	2.83	.11	2.43	.11
	100	3.46	.09	2.73	.09
	Constraint on ΔF per year = .0025				
256	**0**	2.24	.05	1.63	.05
	100	2.83	.04	1.97	.05
512	**0**	2.65	.05	2.05	.05
	100	3.27	.04	2.35	.05

[a] Results from 50 replicated simulations, except the case of 512 new progeny and optimal contribution selection when there were 25 replicates. Standard errors over replicated simulations were .13 to .29 for optimal contribution selection and .14 to .19 for BLUP schemes. For these schemes, the standard errors of *F* were .007 to .062 and .007 to .026, respectively

4. DISCUSSION

4.1. The inbreeding constraint

A dynamic selection rule was proposed for the selection of animals while restricting the rate of inbreeding in populations with overlapping generations. It was shown that a restriction of the $\mathbf{r}'\overline{\mathbf{A}}_t\mathbf{r}$ term restricted the future increase of average relationships in populations with overlapping generations. The elements of the vector **r** are proportional to the sum of the expected future contributions of an age class, which is the largest for young, immature age classes. Hence, the algorithm accounts for the relationships of the currently selected parents with the young, immature animals. For instance, consider a bull with a high EBV, which was selected during the previous years. When considering the bull again for selection, the $\mathbf{r}'\overline{\mathbf{A}}_t\mathbf{r}$ term accounts for the fact that the bull already has a lot of (immature) offspring and thus has already contributed much to the population. Also, consider an age class, which is still expected to contribute about 30% of the offspring next year and then die. The element of **r** will be proportional to 30% for this age class. Hence, the selection of the current parents accounts for the fact that this age class will also have to contribute genes in the next year. For example, if this age class has very high EBV animals, we cannot use them very heavily because we also have to use the same animals in the next year. Hence, the constraining of the $\mathbf{r}'\overline{\mathbf{A}}_t\mathbf{r}$ term seems to make sense when considering the selection of individual animals in populations with overlapping generations.

Table 3. Number of animals selected and generation intervals at yr 20 with the optimal contribution and BLUP selection, where numbers selected were chosen to achieve the inbreeding constraint. [a]

New progeny per year	Size of progeny test	Optimal Contribution		BLUP	
		Selected sires/dams	Gen. Interval sires/dams	Selected sires/dams	Gen. Interval sires/dams
(no of animals)	(no of records)	(no of animals)	(years)	(no of animals)	(years)
	Constraint on ΔF per year = .0050				
256	0	19.2 / 5.3	2.7 / 4.7	64 / 64	2.7 / 3.3
	100	2.8 / 4.5	6.2 / 4.5	64 / 64	3.2 / 3.1
512	0	24.6 / 5.3	2.9 / 4.5	80 / 80	2.5 / 3.2
	100	5.8 / 2.7	6.0 / 4.4	80 / 80	3.1 / 3.0
	Constraint on ΔF per year = .0025				
256	0	34.9 / 11.4	2.9 / 4.5	105 / 105	2.8 / 3.3
	100	7.6 / 10.1	6.0 / 4.4	95 / 95	3.1 / 3.1
512	0	42.9 / 14.0	2.6 / 4.5	130 / 130	2.6 / 3.2
	100	6.4 / 9.7	6.0 / 4.5	125 / 125	3.1 / 3.1

[a] Results from 50 replicated simulations, except the case of 512 new progeny and optimal contribution selection where there were 25 replicates

The dynamic selection rule restricted the increase of the inbreeding, $(F_t - F_{t-1})$. More precisely, we wanted to restrict the rate of inbreeding, $\Delta F = (F_t - F_{t-1})/(1 - F_{t-1})$, but the present level of inbreeding, F_{t-1}, was assumed to be small here. In the simulations, the level of inbreeding after 20 yr of selection was not very small, but this happened because of the long time period that was considered to test the long-term properties of the method. In practice, current levels of inbreeding are often small. Otherwise, the constraint C_t of equation [7B] has to be increased by $2\Delta F(1 - F_{t-1})$ in year *t*, (*i.e.*, $C_t = C_{t-1} + 2\Delta F(1 - F_{t-1})$), instead of by $2\Delta F$, in order to achieve the desired rate of inbreeding.

Further it was assumed that the annual inbreeding rate instead of the rate per generation had to be constrained. The latter does not affect the optimization when breeding schemes have fixed generation intervals. However, when the generation intervals are also optimized, a breeding scheme with a short generation interval and high annual inbreeding rates may have acceptable inbreeding rates per generation but not per year. In practice, a constraint on the annual rate of inbreeding seems more appropriate, because practical breeders want to limit the inbreeding depression, variance reduction due to inbreeding, and the risk of the breeding plan, until a time horizon, (*i.e.*, a fixed number of years and not a fixed number of generations).

From a theoretical and more long-term perspective, a limit on the inbreeding per generation may be more appropriate, because factors that counteract the detrimental effects of inbreeding are probably effective on a per-generation basis. For instance, the genetic variance generated by new mutations is probably constant per generation, at least if the number of new mutations is constant per generation number instead of per year, which may be questioned. Further, the response from natural selection that counteracts the inbreeding depression of fitness traits is probably constant per generation, although a longer generation interval may increase the selection differential of natural selection because animals have to survive until and reproduce at a later age.

The presented generation intervals are optimized with a restriction on the inbreeding per year. The inbreeding restriction favors long generation intervals, because with long generation intervals 1) the number of animals selected per generation tends to increase and 2) the inbreeding per generation is spread over more years. The latter argument holds only when the inbreeding per year is restricted. Hence, restricting the inbreeding per generation leads, generally, to shorter generation intervals than when annual inbreeding was restricted.

4.2. Other constraints

It was assumed that cows could produce as many offspring as required to achieve the optimized contributions. With new reproductive techniques, such as ova pick up, *in vitro* maturation and fertilization (Kruip *et al.*, 1994), this may be possible, at least in small nucleus herds. If reproductive limitations restrict the maximum contribution that a cow can achieve, additional constraints apply: $c_i < c_{\max}$ for all cows, where c_i is the optimal contribution, and $c_{\max}$ is the maximal contribution of a cow and is based on the maximum number of offspring from one cow. Meuwissen (1997) provided some approximations to deal with such additional constraints. The optimal solution requires allocating maximum contributions to some cows and smaller contributions to others, which is a large combinatorial optimization problem if the number of cows is large.

4.3. Cost factor methods and overlapping generations

The optimal contribution method can also be used to restrict the variance of the selection response, (*i.e.*, a component of the risk of the breeding scheme). In this case, the **A** matrix is replaced by the prediction error variance matrix of the EBV (Henderson, 1984). Average relationships of age classes are replaced by prediction error variances of average genetic merit of the age classes. If inbreeding and the variance of the selection response are to be restricted, the algorithm needs to be extended so that it can deal with two quadratic constraints instead of with one.

For discrete generations, the algorithms of Wray and Goddard (1994) and Brisbane and Gibson (1995) select the group of parents with maximum

$$\overline{\mathbf{EBV}} - k\overline{\mathbf{A}},$$

where $\overline{\mathbf{EBV}}$ and $\overline{\mathbf{A}}$ are the average EBV and relationship of the selected parents, respectively, and k is the cost factor of the average relationship. Meuwissen (1997) used the Lagrangian multiplier method to determine a cost factor k so that the desired rate of inbreeding is achieved. With overlapping generations, the cost factor would apply to the average relationships among the animals in the next year, (*i.e.* to the $\mathbf{r}'\overline{\mathbf{A}}_t\mathbf{r}$ term). Hence, the Wray and Goddard (1994) and Brisbane and Gibson (1995) algorithm is extended to overlapping generations by maximizing

$$\overline{\mathbf{EBV}} - k\,\mathbf{r}'\overline{\mathbf{A}}_t\mathbf{r}$$

of the selected parents, where $\mathbf{r}'\overline{\mathbf{A}}_t\mathbf{r}$ is decomposed as in Equation [4]. The distribution of selected parents over age classes is optimized as described in section 2.3.

4.4. General remarks

The computer time required was about 30 min CPU per year of a replicate on an Alpha 500 workstation, when the population size was 512 new progeny per year, which yielded about 1700 selection candidates. This was about 8 times longer than with 256 new progeny per year, (*i.e.*, half the size of 512 new progeny), and suggested that computer time was proportional to n^3, where n is the number of selection candidates. The matrix inversions that were required to calculate the optimal contributions, could explain this proportionality to n^3, because the computer time required to invert a matrix is proportional to n^3 (Press *et al.*, 1989). Hence, computer time may be large in practical schemes with many selection candidates, but it does not seem to be prohibitively large.

In a conference abstract, Grundy *et al.* (1997) presented a similar selection rule, but did not optimize the distribution of the parents over the age classes. Recently, they developed an alternative method for this optimization (B. Grundy, Personal Communication). The methods will be compared in a subsequent paper. The presented selection rule was compared with BLUP selection, in which the numbers of selected animals had to be predefined, (*i.e.*, BLUP selection is not a (completely) dynamic rule) (Goddard and Howard, 1994). The extra selection response at the same rate of inbreeding increased when populations became smaller and inbreeding restrictions became more stringent. Hence, the selection rule is especially useful when the population size is small relative to the required rate of inbreeding. Whether the optimal contribution selection is useful in practice depends on the size of the population, but at a nucleus size of 512 new progeny per year, (*i.e.*, ± 1700 nucleus animals), optimal contribution selected yielded 16 to 38% more genetic gain than BLUP selection. However, part of this superiority of the OC scheme is because the BLUP selection schemes did not select unequal (optimal) numbers of sires and dams, such that the increased responses should be seen as upper limits for the superiority of the OC schemes.

4.5. Implications

A dynamic selection rule was developed that maximized the selection response at a predefined rate of inbreeding in populations with overlapping generations. At the same rates of inbreeding, the dynamic rule yielded up to 44% more selection response than selection for BLUP-EBV. The superiority decreased with increasing population sizes and with increasing rates of inbreeding. The dynamic selection rule is most useful in small selection schemes where relatively low rates of inbreeding are desired.

ACKNOWLEDGEMENT

Helpful suggestions of Mike Goddard are gratefully acknowledged.

LITERATURE

Brisbane, J.R. and Gibson, J.P. (1995). Balancing selection response and inbreeding by including predicted stabilized genetic contributions in selection decisions. Genet. Sel. Evol. 27: 541-549.

De Boer, I.J.M., Meuwissen, T.H.E. and van Arendonk, J.A.M. (1994) Combining the genetic and clonal response in a closed dairy cattle nucleus scheme. Anim. Prod. 59: 345-358.

Goddard, M.E. and Howarth. J.M. (1994) Dynamic selection rules and selection of males to progeny test. Proc. 5th World Congress on Genetics Applied to Livestock Production. 18: 306-309.

Grundy, B. and Hill, W.G. (1993) A method for reducing inbreeding with best linear unbiased prediction. Anim. Prod. 56: 427.

Grundy, B., Villanueva, B. and Woolliams, J.A. (1997) Maximising response with a predefined rate of inbreeding for overlapping generation structures. Proc. British Society of Animal Science , p. 27.

Henderson C.R. (1984) Applications of Linear Models in Animal Breeding, University of Guelph Press, Guelph, Canada.

Hill, W.G. (1974) Prediction and evaluation of response to selection with overlapping generations. Anim. Prod. 18: 117-139.

Johnson, D.L. (1977) Variance-covariance structure of group means with overlapping generations. In: E. Pollak, O. Kempthorne, and T. B. Bailey (ed.) Proc. International Conference of Quantitative Genetics, pp. 851-858.

Kruip, Th.A.M., Boni, R. , Wurth, Y.A., Roelofsen, M.W.M. and Pieterse, M.C. (1994) Potential use of Ovum Pick-Up for embryo-production and breeding in cattle. Theriogenology 42: 675-684.

Meuwissen, T.H.E. (1997) Maximising the response of selection with a predefined rate of inbreeding. J. Anim. Sci. 75: 934-940.

Press, W. H., Flannery, B. P., Teukolsky, S. A. and Vetterling, W. T. (1989) Numerical Recipes. Cambridge Univ. Press, Cambridge, UK, pp 274-334.

Verrier, E., Colleau, J.J. and Foulley, J.L. (1993) Long term effects of selection based on the animal model BLUP in a finite population. Theor. Appl. Genet. 87: 446–454.

Villanueva, B. and Woolliams, J.A. (1997) Optimization of breeding programmes under index selection and constrained inbreeding. Genet. Res. Camb. 69: 145-158.

Wray, N.R. and Goddard, M.E. (1994) Increasing long-term response to selection. Genet. Sel. Evol. 26: 431-451.

APPENDIX

The Lagrangian multipliers λ and λ_0 in Equation [8] are derived by Meuwissen (1997) and, when adopted to the optimization problem of Equations [7], are:

$$\lambda = \left(\mathbf{Q}'\mathbf{A}_{t-1}\mathbf{Q}\right)^{-1}\left[\mathbf{Q}'\mathbf{A}_{t-1}^{-1}\ \left(\mathbf{EBV}_{t-1} - 2\lambda_0\mathbf{A}_{t-1}\mathbf{J}\,\mathbf{r}_\mathbf{b}\right) - 2\lambda_0\mathbf{s}\right]$$

$$\lambda_0^2 = \frac{\tfrac{1}{4}\,\mathbf{EBV}_{t-1}'\,\mathbf{R}\,\mathbf{EBV}_{t-1}}{K + \mathbf{r}_\mathbf{b}'\mathbf{J}'\mathbf{A}_{t-1}\mathbf{R}\mathbf{A}_{t-1}\mathbf{J}\mathbf{r}_\mathbf{b}r_a^2 - \mathbf{s}'(\mathbf{Q}'\mathbf{A}_{t-1}^{-1}\mathbf{Q})^{-1}\mathbf{s}r_a^2 - 2\mathbf{s}'(\mathbf{Q}'\mathbf{A}_{t-1}^{-1}\mathbf{Q})^{-1}\mathbf{Q}'\mathbf{J}\mathbf{r}_\mathbf{b}r_a},$$

where $\mathbf{R} = \left(\mathbf{A}_{t-1}^{-1} - \mathbf{A}_{t-1}^{-1}\mathbf{Q}\left(\mathbf{Q}'\mathbf{A}_{t-1}^{-1}\mathbf{Q}\right)^{-1}\mathbf{Q}'\mathbf{A}_{t-1}^{-1}\right)/r_a^2$ and $K = C_t - \mathbf{r}_\mathbf{b}'\overline{\mathbf{A}}_{t(b,b)}\,\mathbf{r}_\mathbf{b}$.

Chapter 3

Selection with control of inbreeding in populations with overlapping generations- a comparison of methods

A.K. Sonesson[†], B. Grundy[‡], J.A. Woolliams[§] and T.H.E. Meuwissen[†]

[†] Institute for Animal Science and Health,
P.O. Box 65, 8200 AB Lelystad, The Netherlands

[‡] Scottish Agricultural College, West Mains Road, Edinburgh EH9 3JG, UK

[§] Roslin Institute, Midlothian EH25 9PS, UK

Published in Animal Science (2000) 70: 1-8
Reproduced by permission of British Society of Animal Science

ABSTRACT

Methods that maximize genetic response in populations with overlapping generations while controlling rate of inbreeding by constraining average relationship among selection candidates were compared. Firstly, computer simulations of closed nucleus selection schemes showed that a two-stage optimization algorithm approach, where the distribution of parents within and thereafter over age classes was optimized resulted in different breeding schemes than an approach that performed iteration on this distribution. It yielded significantly lower annual genetic gain (.194 versus .223 σ_p units), fewer animals selected (21.9 versus 26.4) and longer generation intervals (2.38 versus 1.68 years), but maintained the rate of inbreeding closer to its constraint. In large schemes, iteration may be computationally the only feasible method for the optimization of parents across age classes. Secondly, the use of conventional relationships for constraining inbreeding was compared with that of augmented relationships, which do not depend on the level of inbreeding. Both relationships resulted in very similar breeding schemes, but the use of augmented relationships avoids correction of the current level of inbreeding. Thirdly, a constraint of the rate of inbreeding on a per year basis was compared with a constraint on a per generation basis. When optimising per generation, the generation interval was shorter compared with a scheme where an analogous annual restriction was in place (2.01 versus 2.38 years) and the annual rate of genetic gain was higher (.214 versus .194 σ_p units).

1. INTRODUCTION

Recently, selection methods that maximize genetic gain while inbreeding is constrained have been developed for discrete generations (Wray and Goddard, 1994; Brisbane and Gibson, 1995; Meuwissen, 1997; Grundy *et al.*, 1998). They are called Optimum contribution (OC) methods because they optimize genetic contribution of current generation to future generations (Meuwissen, 1997). The method of Meuwissen (1997) was extended to populations with overlapping generations by Meuwissen and Sonesson (1998), referred to as OC1, and by Grundy *et al.* (1997 and 2000), referred to as OC2. These two methods optimize genetic contributions over age classes. There are three main differences between OC1 and OC2. Firstly, OC2 uses Meuwissen's (1997) algorithm to find optimum contributions within age classes and the annealing algorithm (Press *et al.*, 1989) to find the global optimum contributions over age classes. OC1 uses Meuwissen's algorithm to find optimum contributions within and over age classes. With OC1, distribution of parents over age classes is calculated by iteration. Secondly, OC2 uses an augmented relationship matrix (Grundy *et al.*, 1998) instead of the conventional relationship matrix, which is used by OC1. Augmented relationships are larger than conventional relationships, because they do not account for reduction of Mendelian sampling variances due to inbreeding. Thus, augmented relationships do not depend on current level of inbreeding. This difference between the two relationship matrices affects how the constraint on inbreeding is implemented. OC1 constrains the absolute increase of inbreeding $(F_{t+1} - F_t)$, which can be adjusted to obtain a constant rate of inbreeding, *i.e.* $F_{t+1} - F_t = \Delta F(1 - F_t)$, where F_t is inbreeding in year t and ΔF is rate of inbreeding. Since the augmented relationships are independent of F_t, the $(1 - F_t)$ correction in the above constraint is not needed in OC2 and so ΔF is constrained directly. A third difference between the methods is that OC2 can be used to restrict ΔF per generation or per year, whereas OC1 restricts ΔF only per year.

The aim of this paper is to compare how differences between OC1 and OC2 affect breeding schemes, with particular emphasis on ΔF and rate of genetic gain (ΔG), generation intervals and number of animals selected in populations with overlapping generation structure.

2. MATERIAL AND METHODS

2.1. General optimization of contributions

The selection method OC1 was described in detail by Meuwissen and Sonesson (1998) and OC2 by Grundy *et al.* (2000) and here only an outline of OC1 and the main differences between OC1 and OC2 are given.

In a population, the annual increase in inbreeding, $\Delta F(1 - F_{t-1})$, equals half the increase of the average relationship. Hence, animals in year t should be selected such that the average relationship of the population in year $(t+1)$ does not exceed the constraint:

$$C_{t+1} = C_t + 2\Delta F(1 - \tfrac{1}{2} C_t) , \qquad [1]$$

where ΔF is the desired rate of inbreeding and $\tfrac{1}{2}C_t$ represents approximately the level of inbreeding, F_t. The average relationship of the population is defined as the average relationship between all possible pairs of individuals in the population including the relationship of each individual with itself.

In discrete generations, ΔF can be controlled by constraining the average increase in the relationship of new born progeny through constraining the average relationship of their parents. Relationships between the progeny have to be constrained because all future genes will derive from this progeny population. With overlapping generations, all animals up to and including maximum reproductive age have the potential to carry genes of future generations and so there is a need to consider also young age classes that are not yet reproducing when selecting. The long term contribution of age class i to the future gene pool, r_i, was derived by Hill's (1974) gene flow method. The r_i will serve as a weight for the relationships within that age class (Meuwissen and Sonesson, 1998). It equals $\mathbf{r}'\bar{\mathbf{A}}_t\mathbf{r}$, where $\mathbf{r}$ is a ($q \times 1$) vector of weights (q = number of age classes), $\bar{\mathbf{A}}$ is a ($q \times q$) matrix of the average relationship between the age classes. Now $\mathbf{r}'\bar{\mathbf{A}}_t\mathbf{r}$ of all age classes can be split into three terms:

$$r_1^2\bar{A}_{11} + 2r_1\bar{\mathbf{A}}_{12}\mathbf{r}_2 + \mathbf{r}_2'\bar{\mathbf{A}}_{22}\mathbf{r}_2, \qquad [2A]$$

where $\bar{\mathbf{A}} = \begin{bmatrix} \bar{A}_{11} & \bar{\mathbf{A}}_{12} \\ \bar{\mathbf{A}}_{21} & \bar{\mathbf{A}}_{22} \end{bmatrix}$ and $\mathbf{r} = \begin{bmatrix} r_1 \\ \mathbf{r}_2 \end{bmatrix}$

and subscripts 1 and 2 denote new progeny and older animals respectively (r_1 = weight of new progeny age class, $\mathbf{r}_2$ is a ((q-1) $\times$ 1) vector of weights of older age classes), $\bar{A}_{11}$ = average relationship of age class one, $\bar{\mathbf{A}}_{12}$ = average relationship between age class one and older age classes, $\bar{\mathbf{A}}_{22}$ = average relationship between older age classes. The first term $(r_1^2\bar{A}_{11})$ involves the average relationship between new progeny, the second term $(r_1\bar{\mathbf{A}}_{12}\mathbf{r}_2)$ involves the relationship between old animals and new progeny and the third term $(\mathbf{r}_2'\bar{\mathbf{A}}_{22}\mathbf{r}_2)$ between old animals. The $\mathbf{r}_2'\bar{\mathbf{A}}_{22}\mathbf{r}_2$ term is not affected by current selection since old animals are already born. The average relationship matrix among new progeny is:

$$\bar{\mathbf{A}}_{11} = \mathbf{c}'\mathbf{A}\mathbf{c} \qquad [2B]$$

and the average relationship of new progeny with older animals is:

$$\bar{\mathbf{A}}_{12} = \mathbf{c}'\mathbf{A}\mathbf{J}, \qquad [2C]$$

where $\mathbf{c}$ is a $(n \times 1)$ vector of contributions of selection candidates to new progeny (n= number of selection candidates); $\mathbf{A}$ is a $(n \times n)$ relationship matrix of the selection candidates and $\mathbf{J}$ is a $(n \times q)$ matrix that averages individual relationships to average relationship per age class.

Combining [1] and [2A to 2C] yields a restriction on the average relationships in a population with overlapping generations:

$$C_{t+1} \leq r_1^2\, \mathbf{c}_t^{'} \mathbf{A}_t \mathbf{c}_t + 2\, r_1 \mathbf{c}_t^{'} \mathbf{A}_t \mathbf{J}\, \mathbf{r}_2 + \mathbf{r}_2^{'} \overline{\mathbf{A}}_{22(t)} \mathbf{r}_2 \ , \qquad [3]$$

where the value of C_{t+1} is calculated by Equation [1] and subscript t refers to year t. The optimum always lies at the border of the solution space, so the smaller or equal sign in Equation [3] may be replaced by an equal sign.

The product of the vector of contributions for selection candidates in year t, $\mathbf{c}_t$, and the vector of breeding values in year t, represents the weighted average genetic merit of the selected group of candidates and thus of their progeny:

$$G_t = \mathbf{c}_t^{'} \mathbf{EBV}_t \ , \qquad [4]$$

where $\mathbf{EBV}_t$ is a vector of estimated breeding values of selection candidates in year t.

The algorithms optimize $\mathbf{c}_t$ such that G_t (in Equation [4]) is maximized under the restriction imposed by Equation [3] and Equation [5]:

$$\mathbf{Q}\mathbf{c}_t = \mathbf{s} \ , \qquad [5]$$

where, in OC1, $\mathbf{Q}$ is a $(2 \times n)$ design matrix indicating sex of the selection candidates and $\mathbf{s}^{'}$ is a (2×1) vector of halves; *i.e.* the restriction [5] ensures that for the optimization procedure, the sum of contributions of each sex actually sum to .5. For OC2, the contribution of each sex within each age class is restricted to a predefined value indicated in $\mathbf{s}$. The contribution of each sex within each age class, $\mathbf{s}$, is optimized with the annealing algorithm. Here, $\mathbf{Q}$ is a $(2q \times n)$ design matrix indicating sex within age class of every animal.

2.2. Annealing versus iteration method

Weights of the age classes, $\mathbf{r}_i$, are calculated from the distributions of parents over age classes and determine the long term contribution of the age class

(Hill, 1974). The distribution of parents over age classes is an input parameter for the OC algorithms that maximize Equation [4] given the restrictions in Equations [3] and [5].

OC1 optimizes by iteration the input distribution of parents over age classes, *i.e.* it calculates an output distribution of parents over age classes from the optimum contribution of each animal and uses this output distribution as the input distribution in the next iteration. This iteration is continued until input and output distributions of parents over age classes are sufficiently equal.

In OC2, extra restrictions were included in Equation [5], such that the output distribution of parents over age classes is forced to equal the input distribution. Thus, **Q** in Equation [5] is an incidence matrix indicating the sex by age class of the selection candidate and **s** is the sum of genetic contributions to which the sex by age class is restricted. OC2 uses the optimization technique of simulated annealing (*e.g.* Press *et al.*, 1989) to calculate optimum values of the constraints, **s**. The simulated annealing algorithm consists of four steps:

1. Set the current distribution of parents over age classes to a distribution where all age classes contribute an equal amount. Evaluate the current distribution using Meuwissen's algorithm and set EBV_c to the average EBV of the selected parents. Set initial "temperature" to T=.005.
2. Consider an alternative distribution of parents over age classes, where the contribution of age class i is decreased by 10% and that of j increased by 10%, with i and j being randomly sampled age classes. Evaluate the alternative distribution with Meuwissen's algorithm and set EBV_a to the average EBV of the selected parents.
3. If EBV_a is larger than EBV_c, replace the current distribution by the alternative distibution. Otherwise, replace the current distribution by the alternative distribution with a probability equal to $exp(-(EBV_c-EBV_a)/T)$, which decreases when EBV_a and/or T is small.
4. If the current distribution has been replaced five times by the alternative distribution or ten alternative distributions have been evaluated, decrease "temperature" T by 25%. If there were no accepted alternative distributions since the last reduction of T: finish (the algorithm will not find an improved distribution and T is too low to accept a lower EBV_a). Otherwise, go to Step 2.

For more details on the mechanisms of the simulated annealing algorithm, see Press *et al.* (1989).

2.3. Augmented versus conventional relationship matrix

When the augmented relationship matrix, $\mathbf{A}^*$, (Grundy *et al.*, 1998), is used, the augmented relationship of the population in year (*t+1*) should be restricted to:

$$C^*_{t+1} = C^*_t + 2\Delta F \qquad [6]$$

and in Equation [3], conventional relationships are replaced by augmented relationships

$$C^*_{t+1} \le r_1^2 \mathbf{c}_t' \mathbf{A}^*_t \mathbf{c}_t + 2 r_1 \mathbf{c}_t' \mathbf{A}^*_t \mathbf{J}\, \mathbf{r}_2 + \mathbf{r}_2' \mathbf{A}^*_{22(t)} \mathbf{r}_2 \qquad [7]$$

The augmented relationship matrix is computed as $\mathbf{A}^* = \mathbf{L}\mathbf{D}^*\mathbf{L}'$, where $\mathbf{L}$ is a lower triangular matrix with elements (*i*, *j*) equal to the genetic contribution of ancestor *j* to descendant *i* as in Quaas (1976); $\mathbf{D}^*$ is a diagonal matrix with ones for base animals and halves for non-base animals. Note that this differs from the original decomposition of the relationship matrix of Quaas, $\mathbf{A} = \mathbf{L}\mathbf{D}\mathbf{L}'$, where $\mathbf{D}$ is a diagonal matrix with ones for base animals and $\frac{1}{2}(1 - F_p)$ for non-base animals (F_p is the average inbreeding coefficient of the sire and dam of the animal). Hence, Equation [7] is obtained approximately by deleting $(1 - F)$ terms on both sides of Equation [3]. The $(1 - F)$ term on the left side of Equation [3] is shown in Equation [1] as $(1 - \frac{1}{2} C_t)$ and the $(1 - F)$ term on the right is shown in Quaas' description of the relationship matrix as $(1 - F_p)$. However, on the left side of Equation [3], the population average inbreeding is deleted against the individual inbreeding terms on the right side of Equation [3]. Hence, Equations [3] and [7] are not identical and Equation [7] might yield a better individual weighting of contributions of animals, because the approximate correction for the population average inbreeding on the left side of Equation [3], is not needed in Equation [7]. The underlying theory of the role of $\mathbf{A}^*$ in controlling inbreeding is given by Grundy *et al.* (1998). Schemes with the augmented relationship matrix and with the conventional relationship matrix are simulated with OC1 to compare the number of selected animals, generation interval, rate of genetic gain and rate of inbreeding.

2.4. Constraining inbreeding on a per year or per generation basis

In order to demonstrate the effect of a constant inbreeding on a per year or per generation basis, OC2 is used. This method was chosen because, although it was originally written to constrain ΔF on a per generation basis, it can also constrain ΔF on a per year basis. The constraint is always applied at the level of the cohort either being scaled to give ΔF per generation, in which case ΔF per year is the intended increment per generation divided by the generation interval, or applied directly for ΔF per year. Note that ΔG is always optimized on a per year basis. Some attempts to run OC1 with ΔF constrained on a per generation basis failed. In these schemes, the generation intervals fluctuated so much that the schemes did not yield an equilibrium selection response and rate of inbreeding.

2.5. The simulated breeding schemes

The simulated closed nucleus scheme is described in Table 1 and was described in detail by Grundy et al. (2000). The scheme is simulated for 30 years. Although the scheme is quite general, it might represent a pig nucleus with selection for growth rate as the trait has a heritability of .5 and is measured on both sexes before puberty. The number of new progeny born per year or generation is limited to 32 or 64 for computational reasons. Minimum generation interval is 1 year and maximum generation interval is 5 years, resulting in five reproductive age classes. The time between two selection rounds is one year. Note that there is no difference between males and females in fertility and age at which records become available, i.e. the breeding scheme is symmetric with respect to the sexes.

In order to simulate the number of progeny per sire (dam) in accordance with the optimal contributions of the sires (dams), calculated by OC1 or OC2, these contributions are multiplied with the total number of progeny to be born (32 or 64 in these schemes) and rounded to integers. The resulting numbers of progeny per sire (dam) were obtained by random mating of the parents.

Table 1. Parameters of the closed nucleus selection scheme.

Size of selection scheme	
Number of new progenies per selection round	32 or 64[‡]
Number of years of random selection before the optimisation methods starts	8
Total number of years over which breeding scheme was run	30
Number of replicated simulations	100
Parameters of trait	
Distribution of true breeding values of base animals	$N(0,.50)$
Distribution of true breeding values of offspring	$N(\frac{1}{2}g_s + \frac{1}{2}g_d; \frac{1}{4}(1-\frac{1}{2}F_s-\frac{1}{2}F_d))$[†]
Distribution of environmental effects	$N(0,.50)$
Culling rate	0
Record becomes available at age	0
Inbreeding constraint	
per year	.01 or .0025[‡]
per generation	.025

[†] g_s and g_d = true breeding value of sire and dam, F_s and F_d= inbreeding coefficient of sire and dam, N denotes normal distribution

[‡] Which of these figures was used is mentioned in the text

3. RESULTS

3.1. Annealing versus iteration method

Results from the comparison between the annealing and iteration methods, when using the augmented relationship matrix are shown in Table 2. ΔF was constrained to .01 per year and for both methods this level was not significantly exceeded. The increase in inbreeding and genetic level was approximately linear over time (results are not shown). Annual genetic gain, ΔG, was 15% higher for OC1 (.223 σ_p units) than OC2 (.194 σ_p units), but OC1 also had 10% higher ΔF. These results can also be transformed to a per generation basis and then ΔG per generation was 23% higher for OC2 than for OC1, which was achieved at 29% higher ΔF per generation for OC2 than for OC1. The latter effect is mainly caused by the longer generation intervals of OC2 than of OC1. The shorter generation interval of OC1 was compensated for by a higher number of animals selected in order to maintain a constant annual ΔF, *i.e.* there was a significantly higher number of selected animals with OC1 (13.3 males + 13.1 females = 26.4) than with OC2 (10.9 males + 11.0

females = 21.9). Within both schemes the number of sires and dams were approximately equal because the schemes were symmetric with respect to sexes.

Table 2. Number of sires plus dams selected (Sel. Anim.), generation interval (L), annual genetic gain (ΔG) and annual rate of inbreeding (ΔF) are presented for OC1 and OC2 selection methods using the augmented relationship matrix (A^*) and for OC1 using the conventional relationship matrix (A). The number of progeny born per year was 32 and ΔF was restricted to .01 per year. Average values of year 26 through 30 of 100 replicated simulations are given. †

	Sel. Anim. (no. sires+ dams)	L (years)	ΔG (σ_p-units/year)	ΔF (%/year)
		Augmented A matrix (A^*)		
OC1	26.4 (.29)	1.68 (.011)	.223 (.043)	.0113 (.0003)
OC2	21.9 (.25)	2.38 (.021)	.194 (.041)	.0103 (.0004)
		Conventional A matrix (A)		
OC1	27.7 (.30)	1.72 (.010)	.219 (.046)	.0108 (.0004)

† Standard errors are given between brackets

3.2. Augmented versus conventional relationship matrix

When the augmented relationship matrix was replaced by the conventional relationship matrix and restriction [3] was applied instead of [7] and OC1 was used, the annual ΔG and ΔF hardly changed (Table 2), although ΔF was somewhat closer to its constrained value. The generation interval was slightly but significantly longer (1.72 years) than when the augmented relationship matrix was used (1.68 years). There were somewhat more animals selected (27.7) than when the augmented relationship matrix was used (26.4). Hence, the augmented and conventional relationship matrices and their associated constraints gave similar results and the differences due to weighting with individual inbreeding levels, *i.e.* the augmented relationship method, versus weighting with population average levels, *i.e.* the conventional relationship matrix, seemed to be small. These results are consistent with those observed for discrete generations (Grundy *et al.* 1998).

Table 3. Number of sires plus dams (Sel. Anim.), generation interval (L), annual rate of genetic gain (ΔG) and rate of inbreeding per year or per generation (ΔF) are presented when rate of inbreeding was restricted on a per year or per generation basis. First the scheme with a restriction per year was simulated and thereafter the ΔF per generation from that simulation was used as the restriction per generation. The number of animals born each year was 32. Average values of year 26 through 30 of 100 replicated simulations are given. †

ΔF restriction	Sel. Anim. (no. sires+ dams)	L (years)	ΔG (σ_p-units/year)	ΔF (%/year)	ΔF (%/gen)
per year	21.9 (.25)	2.38 (.021)	.194 (.041)	.0103 (.0004)	.0250 (.0010)
per generation	21.6 (.21)	2.01 (.020)	.214 (.052)	.0127 (.0005)	.0255 (.0010)

† Standard errors are given between brackets

Table 4. Number of progeny born per year (Progeny), number of sires plus dams selected per year (Sel. Anim.), generation interval, annual genetic gain (ΔG) and annual rate of inbreeding (ΔF) are presented for two sizes of nucleus simulated with OC1 using the augmented relationship matrix. ΔF was restricted to .0025 per year. Average values of year 26 through 30 of 100 replicated simulations are given. †

Progeny (no. new-born per year)	Sel. Anim. (no. sires+ dams)	L (years)	ΔG (σ_p-units/year)	ΔF (%/year)
32 animals	46.4 (.28)	2.86 (.011)	.113 (.033)	.0027 (.0002)
64 animals	74.9 (.44)	2.20 (.008)	.169 (.024)	.0025 (.0001)

† Standard errors are given between brackets

3.3. Constraining inbreeding on a per year or per generation basis

Results from the comparison between constraining ΔF on a per year or per generation basis are shown in Table 3. Note that results of the scheme optimized on a per year basis shown in Table 3 are the same as for OC2 in Table 2. ΔF was constrained to .025 per generation, which equalled the rate of inbreeding per generation of the OC2 scheme in Table 2 (*i.e.* 2.38 × .0103). The constraint on ΔF of .025 per generation was achieved. Annual ΔG was significantly higher when ΔF was constrained on a per generation basis (.214 σ_p units) than when it was constrained on a per year basis (.194 σ_p units). The generation interval was significantly shorter when ΔF was constrained on a per generation basis (2.01 years) than when ΔF was constrained on a per year basis (2.38 years). Furthermore, the number of selected animals was about the same when ΔF was constrained on a per generation (21.6) or per year basis (21.9).

4. DISCUSSION

4.1. Annealing versus iteration method

The main conclusions from the comparison between annealing and iteration methods is that the annealing method yielded longer generation intervals, lower annual ΔG and tended to a somewhat lower annual ΔF (closer to the restriction).

Generation intervals. The relatively longer generation intervals of OC2 can be explained as follows. Because the annealing algorithm determines the contribution of each age class and the algorithm of Meuwissen (1997) has to realize these contributions per age class, OC2 restricts the contribution of each age class in restriction [5]. Restriction [5] is imposed by the Lagrangian multiplier method, *i.e.* the term $\boldsymbol{\lambda}'(\mathbf{Qc}-\mathbf{s})$ is added to Equation [4] that is to be maximized, where $\boldsymbol{\lambda}$ is a ($2q \times 1$) vector of Lagrangian multipliers. The Lagrangian multipliers are chosen such that the restriction in Equation [5] is fulfilled, and can be interpreted as cost factors on the contributions of that age class. OC2 uses the simulated annealing algorithm to optimize the contributions of the age classes, and thus it optimizes the cost factors per age class, λ_i. OC1 uses the algorithm of Meuwissen (1997) for the optimization of the contributions of both across and within age classes, which implies that there are only two restrictions in [5], namely those that restrict the contribution of each sex to .5. Again the restrictions [5] are imposed by cost factors, but here the cost factors apply across all age classes. The latter implies that a simple iteration scheme can be used to match the input distribution of parents across age classes to the output distribution that results from applying the algorithm. Because OC1 uses the same cost factors, $\boldsymbol{\lambda}$, across all age classes, whereas OC2 uses a different cost factor for each age class, OC2 has at least in theory more flexibility to maximize G_t in Equation [4].

In order to compare OC1 and OC2 on the same group of selection candidates, both optimization methods were used in one replicate of OC2. The different cost factors, $\boldsymbol{\lambda}$, that OC2 puts on the contribution of age classes were monitored in this replicate. In all selection rounds, OC2 set a high cost factor, λ, on younger age classes (results are not shown). The high cost factors of younger age classes result in more equal weights of relationships, r_i, over age classes. This increases the flexibility of the selection of animals over age classes, and an increased ΔG is expected. OC1 uses the same cost factor across all age classes and does therefore not have the same flexibility as OC2. Therefore OC2 is expected to get higher ΔG than OC1. The lower cost factors on old age classes also result in a longer generation interval.

Genetic gain. In the test replicate, OC2 mostly gave a solution with a higher selection differential (across age classes) of the selected animals than OC1, probably because the cost factors, λ, are adjusted per age class. The selection differential of selected animals was on average .053 higher for OC2 than for OC1. In the overall comparison of the two methods, annual ΔG of OC2 was however significantly lower (Table 2). This can be explained by the higher (although not significantly higher) annual ΔF of OC1 than of OC2, *i.e.* OC1 yielded 15% higher annual genetic gain at 10% higher rate of inbreeding. In a test run, the ΔF restriction of OC1 was set to the achieved ΔF of OC2 minus 15% (.0090). In this run, the realized ΔF of OC1 was the same as that of OC2 (.0103) and the resulting ΔG was still significantly higher than that of OC2 (12%). The difference in annual ΔG might also be due to second order effects. An example of these second order effects could be that OC2 might loose its advantage in the next generation of selection, when progeny of the current parents will be selected using the same cost factor since they are in the same age class. The latter will favour progeny of younger parents, and thus partly undo the previous selection of older parents.

Another factor that could have reduced ΔG of the OC2 schemes is that OC2 restricts the contribution of each sex within an age class to 50% of the contribution of that age class, whereas OC1 restricts the contribution of each sex across all age classes to 50%. Although the contribution of each sex within an age class is expected to equal 50% of the contribution of that age class, the optimum may differ from this expectation due to sampling in small breeding schemes, such as the present schemes.

Computer time. Although it depends on the efficiency of the programming, there was a large difference in CPU time required when comparing OC1 and OC2 (Table 2). With 100 replicated simulations of 30 years and 32 new progeny per year it took 37h 11min CPU time for the annealing method and 54 min for the iteration method. For practical schemes where only one single optimization is needed at a time, both methods have however an acceptable CPU time. In practice, the size of the optimization problem may be substantially larger than here, due to the larger number of selection candidates. OC1 has been used in breeding schemes with up to 1700 selection candidates (Meuwissen and Sonesson, 1998).

4.2. Augmented versus conventional relationship matrix

We expected that the schemes with the augmented relationship matrix would yield higher ΔG than schemes with the conventional relationship matrix, because of the correction for individual inbreeding levels in the augmented relationship matrix compared with correction at the population level with the conventional relationship matrix. This difference between the two relationship matrices did however not affect ΔF or ΔG. Hence, correcting elements of the relationship matrix for the population average inbreeding seems sufficiently accurate. There is hardly any difference in ease of implementation of both these relationship matrices. The difference in breeding schemes between the use of augmented and conventional relationship matrices was investigated with OC1, but similar results are expected when OC2 is used.

4.3. Constraining inbreeding on a per year or per generation basis

When ΔF was constrained on a per generation basis, a different scheme was obtained with higher annual ΔG and ΔF (Table 3), because the scheme with ΔF constrained on a per year basis was now suboptimal. Note that the increased annual ΔG is somewhat an artefact because if ΔF would first have been constrained on a per generation basis, a scheme with ΔF constrained per year would yield an increased annual ΔG at the same annual ΔF (since the scheme is optimized under a different constraint), but higher ΔF per generation and longer generation intervals. Generation intervals were shorter when ΔF was restricted on a per generation basis (2.01 years) than when ΔF was restricted on a per year basis (2.38 years), because $\Delta F_{yr} \times L$ was constant, where ΔF_{yr} is the annual rate of inbreeding which was increased and L is the generation interval. For OC1, a smaller increase in annual ΔG is expected than for OC2, because OC1 has already a short generation interval.

Practical breeders may find a constraint of inbreeding on a per year basis more appropriate, because they want to constrain inbreeding depression, variance reduction due to inbreeding and risk of the breeding scheme until a given time horizon, *i.e.* a fixed number of years and not a fixed number of generations.

Some costs of correcting genetic defects, which might occur when lethal recessive alleles increase in frequency with increasing inbreeding, are however per generation, which suggests that inbreeding should be constrained on a per generation basis. From a theoretical and long-term perspective, a constraint on ΔF per generation may be more appropriate, because some factors that counteract detrimental effects of inbreeding occur on a per generation basis. Examples of these factors are natural selection and mutations that occur during meiosis. However, natural selection may be somewhat stronger when generation interval increases since animals have to survive up to a higher age and some mutations occur on a per year basis *e.g.* mutations induced by radiation. In conclusion, the choice of constraining inbreeding on a per year or per generation basis will depend on if a long or short time perspective is taken, but it should be realized that a constraint of ΔF on a per generation basis results in shorter generation intervals. This might be important in practical breeding schemes since a constraint on a per generation basis would make it possible to cull animals at younger ages, which would decrease the costs of the schemes. Increasing longevity can however give positive economic returns because adult animals may be more profitable (see *e.g.* review of Essl, 1998).

4.4. General remarks

OC1 and OC2 have been compared in a small nucleus scheme, which could represent a pig nucleus with selection for growth. Although the methods are only compared for this scheme, the tendencies in generation interval, number of animals selected and rate of genetic gain and inbreeding are expected to hold for a broader range of schemes. For larger schemes and/or schemes with higher desired rate of inbreeding, the differences between OC1 and OC2 schemes are however expected to become smaller, because optimum contribution selection becomes more like best linear unbiased prediction (BLUP) selection as schemes become larger and when inbreeding restrictions are less stringent (Meuwissen and Sonesson, 1998). In general, generation intervals are expected to decrease as the population size increases (Table 4) because genetic gain increases with increased population size. This favours selection of younger animals. In many schemes information accumulates as the animals become older, through records on sibs and progeny as well as through repeated measurements of the same trait on an individual. This was not the case in the present scheme where animals obtained one record before puberty. If information accumulates, selection from some age classes may be particularly efficient because of a good age to accuracy of selection ratio. Both OC1 and OC2 may then predominantly select parents from these age classes. This would make the differences between the OC1 and OC2 schemes smaller for species where information accumulates over lifetime.

4.5. Conclusions

The resulting breeding schemes differed, with OC1 yielding a scheme with higher annual ΔG, shorter generation interval and more animals selected than OC2. OC1 had however difficulties in keeping the constraint on ΔF for these small schemes with 32 new born progenies per selection round, but with 64 new born progenies per selection round, OC1 managed to maintain ΔF at the constraint. The iterative optimization method used less CPU time than the annealing method and may therefore be the preferred method for larger population sizes. The augmented and conventional relationship matrix yielded similar results, but when augmented relationships are used, a correction of the current level of inbreeding is not necessary. Whether the inbreeding restriction is applied on a per year or per generation basis depends on if the breeder takes a short or long time perspective but the latter results in a shorter generation interval and is presently only feasible using OC2.

ACKNOWLEDGEMENTS

We would like to thank Beatriz Villanueva, Pim Brascamp and Ab Groen for their useful comments on the manuscript. BG grateful acknowledges funding from the Biotechnology and Biological Sciences Research Council (UK) and JAW from the Ministry of Agriculture Fisheries and Food (UK).

LITERATURE

Brisbane, J.R. and Gibson, J.P. (1995). Balancing selection response and inbreeding by including predicted stabilized genetic contributions in selection decisions. Genet. Sel. Evol. 27: 541-549.

Essl, A. (1998) Longevity in dairy cattle breeding: a review. Livest. Prod. Sci. 57: 79-89.

Grundy, B., Villanueva, B. and Woolliams, J.A. (1997) Maximising response with a predefined rate of inbreeding for overlapping generation structure. Proc. British Society of Animal Science, p.27.

Grundy, B., Villanueva, B. and Woolliams, J.A. (1998) Dynamic selection procedures for constrained inbreeding and their consequences for pedigree development. Genet. Res. Camb 72: 159-168.

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

Hill, W.G. (1974) Prediction and evaluation of response to selection with overlapping generations. Anim. Prod. 18: 117-139.

Meuwissen, T.H.E. (1997) Maximising the response of selection with a predefined rate of inbreeding. J. Anim. Sci. 75: 934-940.

Meuwissen, T.H.E. and Sonesson, A.K. (1998) Maximizing the response of selection with a predefined rate of inbreeding: overlapping generations. J. Anim. Sci. 76: 2575-2583.

Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T. (1989) Minimization or maximization of functions. In. Numerical recipes- the art of scientific computing, University of Cambridge Press, UK, pp. 274- 334.

Quaas, R.L. (1976) Computing the diagonal elements of a large numerator relationship matrix. Biometrics 32: 949-953.

Wray, N.R. and Goddard, M.E. (1994) Increasing long-term response to selection. Genet. Sel. Evol. 26: 431-451.

Chapter 4

Mating schemes for optimum contribution selection with constrained rates of inbreeding

A.K. Sonesson and T.H.E. Meuwissen

Institute for Animal Science and Health,
P.O. Box 65, 8200 AB Lelystad, The Netherlands

Published in Genetic Selection Evoultion (2000) 32: 231-248
Reproduced by permission of INRA, EDP Sciences

ABSTRACT

The effect of non-random mating on genetic response was compared for populations with discrete generations. Mating followed a selection step where the average coancestry of selected animals was constrained, while genetic response was maximized. Minimum coancestry (MC), Minimum coancestry with a maximum of one offspring per full-sib family (MC1) and Minimum variance of the relationships of the offspring (MVRO) mating schemes resulted in a delay in inbreeding of about two generations compared with Random, Factorial and Compensatory mating. In these breeding schemes where selection constrains the rate of inbreeding, ΔF, the improved family structure due to non-random mating increased genetic response. For schemes with ΔF constrained to 1.0% and 100 selection candidates, genetic response was 22% higher for the MC1 and MVRO schemes compared with Random mating schemes. For schemes with a less stringent constraint on ΔF or more selection candidates, the superiority of the MC1 and MVRO schemes was smaller (5-6%). In general, MC1 seemed to be the preferred mating method, since it almost always yielded the highest genetic response. MC1 mainly achieved these high genetic responses by avoiding extreme relationships among the offspring, *i.e.* full-sibs are avoided, and by making the contributions of ancestors to offspring more equal by mating least related animals.

1. INTRODUCTION

Selection and mating schemes attempt to reduce rates of inbreeding and/or increase genetic response. Optimum contribution selection methods (Toro and Nieto, 1984; Wray and Goddard, 1994; Brisbane and Gibson, 1995; Meuwissen, 1997; Meuwissen and Sonesson, 1998; Grundy *et al.* 1998; Grundy *et al.*, 2000; Sonesson *et al.*, 2000) maximise genetic response while constraining inbreeding by restricting the coancestry among selected parents. These authors assumed however random mating among selected parents, whereas non-random mating may improve family structure of the next generation, and thus affect the results of subsequent selection. Caballero *et al.* (1996) concluded that for phenotypic and BLUP selection, minimum coancestry and compensatory mating strategies generally have a small effect on the genetic response to selection, but that they reduce rates of inbreeding. The effect of non-random mating on rates of inbreeding may, however, be reduced with optimum contribution selection schemes, because the increase of the average coancestry of the population is restricted. Because optimum contribution selection does not use an improvement of family structure due to non-random mating to achieve a lower coancestry of the selected parents (which is restricted), it may use the improvement of family structure to increase selection differential, *i.e.* to increase genetic response. Three different effects of non-random mating in combination with optimum contribution selection may therefore be envisaged:

1. Genetic response may increase (Caballero *et al.*, 1996).
2. Level of inbreeding may decrease (Caballero *et al.*, 1996; Fernández and Toro, 1999; Toro and Pérez-Enciso, 1990).
3. Rates of inbreeding may decrease (Caballero *et al.*, 1996).

One step mating and selection strategies have been proposed (Toro *et al.*, 1988; Toro and Pérez-Enciso, 1990) that used the linear programming algorithm to optimize selection response, while restricting the inbreeding coefficients of the offspring. Restricting inbreeding coefficients of offspring does not control long-term inbreeding, because the increase of the average coancestry of the population, which equals the increase of the long-term inbreeding, may still increase too rapidly (Wray and Goddard, 1994; Brisbane and Gibson, 1995; Meuwissen, 1997; Grundy *et al.*, 1998; Meuwissen and Sonesson, 1998; Grundy *et al.*, 2000; Sonesson *et al.*, 2000). Furthermore, optimizing mating schemes by linear programming is rather computational intensive, which makes it unpractical for large populations (Fernández and Toro, 1999). One way to reduce the needed computer time is to optimize selection and mating in two separate steps so that when optimising the mating step, only the selected animals have to be considered, although this method may not result in maximum genetic response because all possible matings of selection candidates are not considered.

The aim of this paper is to investigate the rates of genetic response and rates of inbreeding for eight mating schemes that are combined with optimum contribution selection in a two step selection and mating optimization procedure. For all schemes, the number of offspring for each selected animal is given by the optimum contribution selection algorithm. The eight mating schemes were: 1) Random mating; 2) Compensatory mating, where sires, with the highest genetic contribution were mated to dams with the lowest genetic contribution, in sequence (Santiago and Caballero, 1995); 3) Compensatory mating, where sires, which were most related to all the selection candidates were mated to dams that were least related to all the selection candidates, in sequence (Santiago and Caballero, 1995); 4) Minimum coancestry mating, where the least related animals are mated; 5) A mating scheme which minimizes variance of coancestry among the progeny. For schemes 1, 3 and 4, a restriction of only having one offspring per mating pair was added (1b, 3b and 4b), *i.e.* paternal and maternal half-sib families were created instead of full-sib families, as in the factorial mating design of Woolliams (1989). The non-random mating mainly affects the family structure of the selection candidates, either by permitting only one offspring per mating pair (schemes 1b, 3b and 4b), by connecting certain families (schemes 2, 3, 3b, 4 and 4b) or by reducing extreme relationships (scheme 5).

2. MATERIAL AND METHODS

2.1. Selection method

The method of Meuwissen (1997) was used to select animals. This method maximizes the genetic level of the next generation of animals, $G_{t+1}=\mathbf{c}_t'\mathbf{EBV}_t$, where $\mathbf{c}_t$ is a vector of genetic contributions of the selection candidates to generation t+1 and $\mathbf{EBV}_t$ is a vector of BLUP estimated breeding values of the candidates for selection in generation t. Rates of inbreeding are controlled by constraining the average coancestry to $\overline{C}_{t+1} = \mathbf{c}_t'\mathbf{A}_t\mathbf{c}_t / 2$, where $\mathbf{A}_t$ is a ($n \times n$) relationship matrix among the n selection candidates, $\overline{C}_{t+1} = 1-(1-\Delta F_d)^{t+1}$, and ΔF_d is the desired rate of inbreeding (Grundy *et al.*, 1998). The constraint is based on the normal **A** matrix here instead of the augmented **A** matrix, but both matrices yield very similar results (Grundy *et al.*, 1998; Sonesson *et al.*, 2000). Note that the level of the constraint $\overline{C}_{t+1}$, can be calculated for every generation before the breeding scheme commences. For optimization of $\mathbf{c}_t$, it was also necessary to constrain the sum of the contributions of males (females) to ½, *i.e.* $\mathbf{Q}'\mathbf{c}_t = ½\mathbf{1}_2$, where **Q** is an ($n \times 2$) 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 ones.

In order to obtain the optimal $\mathbf{c}_t$ that maximizes G_{t+1}, Lagrangian multipliers are used, which yields the following quadratic index, I_t (Meuwissen, 1997):

$$I_t = \mathbf{c}_t'\mathbf{EBV}_t - \lambda_0\left(\mathbf{c}_t'\mathbf{A}_t\mathbf{c}_t - 2\overline{C}_{t+1}\right) - \left(\mathbf{c}_t'\mathbf{Q} - ½\mathbf{1}_2'\right)\lambda \quad , \qquad [1]$$

where λ_0 and $\boldsymbol{\lambda}$ are Lagrangian multipliers ($\boldsymbol{\lambda}$ = (2 × 1) vector of Lagrangian multipliers). The objective function, $\mathbf{c}_t^{'}\mathbf{EBV}_t$, is maximized for $\mathbf{c}_t$ under two restrictions; the first one is on the average relationship of the selection candidates and the second one is on the contribution per sex. The optimization procedure is explained in Meuwissen (1997). The output from the selection method is a vector with genetic contribution for each selection candidate, $\mathbf{c}_t$.

2.2. Random mating (R)

A summary of restrictions and objectives for the mating schemes is given in Table 1. For the R scheme, a sire and a dam are allocated at random for each new born progeny with a probability that is proportional to the genetic contribution that they received from the selection algorithm, $\mathbf{c}_t$.

2.3. Compensatory mating (C)

In the compensatory mating scheme, as originally described by Santiago and Caballero (1995), the selected offspring from families with largest contribution are mated to selected offspring from families with smallest contributions, in sequence. When optimum contribution selection is used, the contributions of selected animals differ, as indicated by their c_t value, and the compensatory mating scheme can be based on contributions of individuals instead of contributions of families. The number of progeny per sire (dam) is obtained from its optimum contribution, c_t, which is calculated by the selection algorithm by multiplying the contributions of the sire (dam) with the number of selection candidates and rounding to integers. Normally the truncation for rounding up versus down is .5, but if the total number of progeny does not sum up to the intended number, the truncation point is adjusted such that the intended number of progeny results. Thereafter, sires and dams are ranked according to their number of offspring. Sires with the highest number of progeny are mated to dams with the lowest number, in sequence. For example, if the highest ranked sire obtains ten progeny and the two lowest ranked dams three and eight progeny, this sire obtained three progeny with the lowest ranked dam and seven with the second lowest ranked dam. The second highest ranked sire will be the sire of the eighth progeny of the second lowest ranked dam.

2.4. Compensatory mating on relationships (CREL and CREL1)

For CREL, the original method (Santiago and Caballero, 1995) was modified as described in Caballero *et al.* (1996). Firstly, the average relationship of each selected animal to all other selected animals was calculated. Thereafter, sires and dams were ranked on this relationship and sires with the highest rank were mated to dams with the lowest rank, in sequence. Each candidate was given as many progeny as was optimized by the selection process. However, many full-sib relationships in a scheme result in either less genetic response or more inbreeding compared with the situation where the full-sib relationships are replaced by paternal and maternal half-sib relationships (Woolliams, 1989). A higher inbreeding in generation t can lead to problems for the selection algorithm to constrain the inbreeding in generation $t+1$. Hence, for CREL1, an additional constraint was imposed, namely that each mating pair should obtain only one offspring. However, this cannot always be achieved, *e.g.* when a sire should obtain more offspring than the number of dams that are selected, *i.e.* one offspring per dam does not suffice. In the latter case, the sire will obtain more than one offspring per dam.

2.5. Minimum coancestry mating (MC and MC1)

With minimum coancestry mating, MC, the average relationship of sires and dams and therefore also the inbreeding of their progeny is minimized. A matrix **F** of size ($N_s \times N_d$) is set up, where N_s (N_d) is the number of selected sires (dams) and element $\mathbf{F}_{ij}$ is the coefficient of coancestry of the selected animals, which is also the inbreeding coefficient of their progeny. MC schemes can result in many full-sib offspring, because a sire has only minimum coancestry with one dam. For MC1, an additional restriction was included, restricting the number of offspring per mating pair to a maximum of one.

Minimum coancestry matings were obtained by applying the simulated annealing algorithm (Press *et al.*, 1989). The implementation of the annealing algorithm is described in the **APPENDIX**. The alternative solutions that were evaluated by the annealing algorithm differed from the current optimal solution by replacing mating pairs according to the scheme in Figure 2. These conditions reduce the number of alternative solutions substantially, and thus reduce the parameter space for the annealing algorithm, which saves computer time.

2.6. Minimum variance of relationship of offspring mating (MVRO)

The MVRO mating method minimizes the variance of relationships of progeny of selected sires and dams, which reduces the number of extremely high relationships among the progeny. The latter also reduces the probability that two full-sibs are made because of their strong relationship. Note that the optimum contribution selection algorithm calculates the average relationship of the offspring, which can not be changed by the mating algorithm. For MVRO schemes, a symmetric matrix **V** is set up that yields the term that a pair of possible offspring would contribute to the variance of the relationship of offspring, *i.e.* $V_{ij} = (a_{ij} - \bar{a})^2$, for $i \neq j$ and $V_{ij}=0$ for $i=j$, where a_{ij} is the relationship of offspring i and j, and $i(j)$ denotes every possible offspring from all $N_s \times N_d$ mating combinations, *i.e.* **V** is of size $((N_s \times N_d) \times (N_s \times N_d))$ and $\bar{a}$ is the average relationship of selected animals. Note that $a_{ij} = ¼\left[a_{sisj} + a_{sidj} + a_{disj} + a_{djsj}\right]$, where s_i and d_i (s_j and d_j) denote the sire and dam of $i(j)$. Let the vector **m** contain the number of offspring from mating pair ij. Now, the variance of the relationships of the offspring is proportional to $\mathbf{m}'\mathbf{V}\mathbf{m}$. The minimization of $\mathbf{m}'\mathbf{V}\mathbf{m}$ is again done by the annealing algorithm (see **APPENDIX**). The alternative solutions that were evaluated by annealing are the same as for MC mating.

2.7. Factorial mating (R1)

For the factorial mating scheme (Woolliams, 1989), the simulated annealing algorithm is used as explained in the minimum coancestry mating section, but without an objective function, *i.e.* all suggested changes of matings (1000) are accepted in order to randomise the initial solution. This randomization by the annealing algorithm is preferred over simply sampling at random a sire and dam for each offspring according to the contributions of the sires and dams as in R, with the additional restriction that a sire and dam pair can only be sampled once. The latter often results in a not feasible solution, *i.e.* full sibs cannot be avoided anymore because of unfortunate sampling of earlier matings. Furthermore, it results (more often than by chance) in the mating of sires and dams with low contributions, since after some sampling of matings according to the contribution of the sires and dams, only sires and dams with low contribution matings are still available.

Table 1. Summary of restrictions and objectives for the mating schemes. For all mating schemes, the restriction applies that the number of progeny per selected animal are as calculated by the optimum contribution selection method.

Name	Abbrevation	Restriction	Objective
Random	R	None	Randomise matings
Factorial	R1	Max one progeny per full-sib family	As objective of R
Compensatory	CREL	None	Mate sires with highest relationship to other selected animals to dams with lowest relationship to other selected animals
Compensatory	CREL1	Max one progeny per full-sib family	As objective of CREL
Compensatory	C	None	Mate sires with highest genetic contribution to dams with lowest genetic contribution
Minimum coancestry	MC	None	Mate animals that are least related
Minimum coancestry	MC1	Max one progeny per full-sib family	As objective of MC
Minimum variance of relationship of offspring	MVRO	None	Mate animals such that the variance of coancestry among progeny is minimised

Table 2. Parameters of the closed nucleus scheme.

Size of selection scheme	
Number of selection candidates	100 or 200
Number of generations over which the breeding scheme was run	20
Number of replicated simulations	100
Parameters of trait	
Phenotypic variation, s^2_p	1.00
Heritability, h^2	.10, .25 or .50
Recording of trait	on both sexes, before selection
Inbreeding constraint	1.0 or 2.5% per generation

2.8. The simulated breeding schemes

The simulated breeding schemes are described in Table 2. The general structure is that of a closed nucleus scheme with discrete generation structure. The number of selection candidates per generation is 100 or 200. Genotypes, g_i, of the base animals were sampled from the distribution $N(0, \sigma_a^2)$, where σ_a^2 is base generation genetic variance of .10, .25 or .50. Later generations are obtained by simulating offspring genotypes from $g_i = ½g_s + ½g_d + m_i$, where s denotes the sire and d the dam of the offspring i, and m_i is the Mendelian sampling component, which was sampled from $N(0, ½\,(1-\bar{F})\,\sigma_a^2)$, where $\bar{F}$ is the average inbreeding coefficient of parents s and d. Phenotypes are simulated by adding an error term to the genotypes, which was sampled from $N(0, \sigma_e^2)$. The base generation phenotypic variance $\sigma_p^2 = \sigma_a^2 + \sigma_e^2$ is always equal to 1. Estimates of breeding values (EBVs) are obtained using the BLUP-breeding value estimation procedure (Henderson, 1984).

3. RESULTS

3.1. Inbreeding trend

For MC, MC1 and MVRO, there was a delay in inbreeding of about two generations compared with R, R1, C, CREL and CREL1 schemes (Figure 1). For the MC scheme, there was a somewhat lower F than for the MC1 scheme. The latter was probably due to the extra restriction of having only one offspring per mating in MC1, which reduced the opportunities for MC1 of mating animals that were least related, *i.e.* its opportunity to generate offspring with lowest inbreeding. In Figure 1, ΔF was constrained to 1.0%, the number of selection candidates was 100 and the trait had a heritability of .25, but similar results were obtained for all other schemes (results not shown).

For all schemes, the realized rates of inbreeding were close to the desired rates of inbreeding (Tables 3 and 4), which justifies a comparison of genetic response between the different schemes.

3.2. Genetic response in small schemes with low inbreeding

The genetic level at year 20, G_{20}, was significantly higher for the non-random mating schemes than for the random mating scheme when ΔF was constrained to 1.0% and the number of selection candidates was 100 (Table 3). For $h^2 = .25$, G_{20} was the highest for the MC1 (4.01 σ_p units) and MVRO

(4.02 σ_p units) schemes, which corresponds to 22.3 and 22.6% higher G_{20} than with the R scheme (3.28 σ_p units). The G_{20} was somewhat lower for the R1, CREL1 and MC schemes than for the MC1 and MVRO schemes, although this difference was not significant. The G_{20} for the C and CREL schemes was significantly lower than for the other non-random schemes. In general, for the other levels of heritability, the same pattern between the schemes was seen, although the superiority of the non-random schemes over the R schemes was slightly higher for h^2 = .10 and lower for h^2 = .50. The G_{20} for the CREL1 scheme was as high as for the MC1 and MVRO schemes for the lowest h^2 of .10, but for the higher heritabilies of .25 and .50, G_{20} was lower than for the MC1 and MVRO schemes, although this difference was not significant.

For all breeding schemes, there was a linear increase of genetic response over generations (results are not shown).

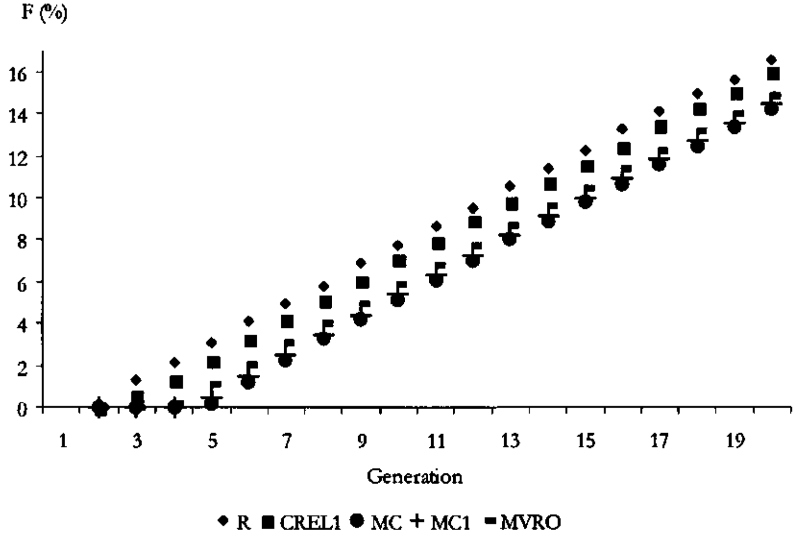

Figure 1. Coefficient of inbreeding (F) for schemes where ΔF was constrained to 1.0% per generation, the number of selection candidates was 100 and the heritability was .25. The F of the R1, C and CREL was always between F of the R (◆) and CREL1 (■) schemes and has been omitted to increase readability of the graph. R: Random, CREL1: Compensatory, MC and MC1: Minimum coancestry, MVRO= Minimum variance of relationship of offspring.

Table 3. Average rate of inbreeding (ΔF), genetic level at generation 20 (G_{20}), number of selected sires and dams and variance of relationship at generation 20 (V_{rel}) for schemes with ΔF constrained to 1% per generation and 100 selection candidates.

Mating scheme[1]	ΔF/gen [2] (%)	G_{20} (σ_p-units (s.e.))	Sel. sires/dams [2]	V_{rel}
		$h^2=.10$		
R	1.00	1.42 (.0177)	34.2/34.0	.00687
R1	1.00	1.72 (.0157)	27.1/27.4	.00447
C	0.97	1.70 (.0168)	30.3/30.2	.00620
CREL	1.01	1.69 (.0160)	31.7/31.6	.00571
CREL1	1.00	1.74 (.0168)	30.8/30.6	.00411
MC	0.99	1.69 (.0187)	27.5/27.4	.00542
MC1	1.00	1.74 (.0176)	27.9/27.7	.00331
MVRO	0.99	1.74 (.0166)	27.8/27.8	.00332
		$h^2=.25$		
R	1.00	3.28 (.0296)	32.2/32.1	.00701
R1	1.01	3.98 (.0249)	26.2/26.0	.00465
C	1.00	3.92 (.0249)	28.4/28.2	.00672
CREL	1.00	3.86 (.0264)	29.6/29.8	.00615
CREL1	1.01	3.96 (.0290)	29.1/28.7	.00441
MC	1.00	3.98 (.0263)	26.0/25.9	.00570
MC1	1.00	4.01 (.0266)	26.5/26.6	.00348
MVRO	1.00	4.02 (.0291)	26.2/26.4	.00351
		$h^2=.50$		
R	0.96	6.31 (.0376)	30.5/30.5	.00716
R1	1.00	7.53 (.0345)	24.5/24.7	.00500
C	0.96	7.50 (.0362)	26.1/26.1	.00728
CREL	1.01	7.28 (.0353)	27.9/28.1	.00650
CREL1	1.02	7.52 (.0398)	26.9/27.0	.00483
MC	1.00	7.53 (.0409)	24.4/24.4	.00607
MC1	1.00	7.61 (.0345)	25.0/25.2	.00372
MVRO	1.00	7.60 (.0332)	25.1/25.3	.00373

[1] See Table 1 for abbreviations of the mating schemes

[2] Average of generation 16-20

Table 4. Average rate of inbreeding (ΔF), genetic level at generation 20 (G_{20}), number of selected sires and dams and variance of relationship at generation 20 (V_{rel}) for schemes with ΔF constrained to 2.5% per generation and 100 selection candidates or ΔF constrained to 1.0% and 200 selection candidates.[1]

Mating scheme[1]	ΔF/gen [2] (%)	G_{20} (σ_p-units (s.e.))	Sel. sires/dams [2]	V_{rel}
		ΔF constrained to 2.5% and 100 selection candidates		
R	2.50	4.94 (.0396)	14.7/14.8	.00777
R1	2.48	5.18 (.0346)	13.3/13.4	.00637
C	2.48	4.92 (.0366)	13.7/13.8	.00937
CREL	2.49	4.94 (.0370)	15.7/15.6	.00711
CREL1	2.50	5.11 (.0412)	14.0/14.1	.00608
MC	2.49	5.15 (.0391)	13.5/13.6	.00608
MC1	2.49	5.28 (.0394)	13.9/13.6	.00440
MVRO	2.48	5.28 (.0355)	13.3/13.4	.00482
		ΔF constrained to 1.0% and 200 selection candidates		
R	1.00	5.07 (.0279)	36.5/36.1	.00595
R1	0.99	5.34 (.0229)	31.8/32.1	.00463
C	1.00	5.33 (.0277)	33.7/34.1	.00678
CREL	1.00	5.18 (.0269)	37.2/37.1	.00591
CREL1	0.99	5.37 (.0264)	34.9/34.7	.00435
MC	1.00	5.33 (.0307)	31.3/31.6	.0058
MC1	1.00	5.42 (.0266)	32.5/32.6	.00346
MVRO	1.00	5.43 (.0265)	32.8/33.1	.00353

[1] See Table 1 for abbreviations of the mating schemes

[2] Average of generation 16-20

Chapter 5

Non-random mating for selection with restricted rates of inbreeding and overlapping generations

A.K. Sonesson and T.H.E. Meuwissen

Institute for Animal Science and Health,
P.O. Box 65, 8200 AB Lelystad, The Netherlands

Published in Genetic Selection Evolution (2000) 34: 1-17
Reproduced by permission of INRA, EDP Sciences

ABSTRACT

Minimum coancestry mating with a maximum of one offspring per mating pair (MC1) is compared with random mating schemes for populations with overlapping generations. Optimum contribution selection is used, whereby ΔF is restricted. For schemes with overlapping generations and ΔF restricted to .25% per year, 256 animals born per year and heritability of .25, genetic gain increased with 18% compared with random mating. The effect of MC1 on genetic gain decreased for larger schemes and schemes with a less stringent restriction on inbreeding. Breeding schemes hardly changed when omitting the iteration on generation interval to find an optimum distribution of parents over age-classes, which saves computer time, but inbreeding and genetic merit fluctuated more before the schemes had reached a steady-state. When bulls were progeny tested, these progeny tested bulls were selected instead of the young bulls, which lead to increased generation intervals, increased selection intensity of bulls and increased genetic gain (35% compared to a scheme without progeny testing). The effect of MC1 decreased for schemes with progeny testing. MC1 mating increased genetic gain 11-18% for overlapping and 1-4% for discrete generations, when schemes with similar rate of inbreeding and genetic gain per generation were compared.

1. INTRODUCTION

For breeding schemes, the selection step determines the increase in average coancestry of the population, but the mating step can improve the genetic structure of the population for the next round of selection. Caballero *et al.* (1996) concluded that non-random mating decreased rate of inbreeding (ΔF), but had little effect on genetic response for BLUP and phenotypic selection. A reduction of ΔF is, however, not expected when selection is made with a restriction on ΔF, but instead the improved family structure due to non-random mating can be used to increase selection differential, *i.e.* to increase genetic response (Sonesson and Meuwissen, 2000). For selection schemes with a restriction on ΔF and discrete generations, minimum coancestry mating with only one progeny per mating pair (MC1) increased genetic response 5 through 22% compared with random mating when optimum contribution selection was used. The increase in genetic response was higher when schemes were small or when the restriction on rate of inbreeding was more stringent and can be explained by the effects of MC1 on the relationship structure of the population. Minimum coancestry mating connects unrelated families and avoids extreme relationships. The factorial mating schemes by Woolliams

(1989) avoid extreme relationships by exchanging full-sib relationships for (maternal) half-sib relationships, which is also done in the MC1 scheme by restricting the number of progeny per mating pair to zero or one. Both these effects result in a population with more homogenous relationships among animals. Because MC1 mating makes relationships more homogenous across families, the relationship among animals with the highest EBVs will be reduced. Hence, it will be easier for optimum contribution selection to select animals with the highest EBVs when MC1 mating is used instead of random mating. A third effect of non-random mating, especially minimum coancestry mating, are decreased inbreeding levels of the progeny and thus also of parents of the next generation. A larger Mendelian sampling variance leads to increased genetic variance and genetic gain.

The above schemes were tested for populations that had discrete generations. Most practical schemes have however overlapping generation structures. For schemes with overlapping generations, parents are selected from several age-classes, which makes their pedigrees more heterogeneous, so that the above mentioned effects of MC1 mating on the relationship structure of the population are advantageous. On the contrary, the benefits of MC1 decrease with increasing population size (Sonesson and Meuwissen, 2000) and schemes with overlapping generations can be larger than schemes with discrete generations. Hence, it is not clear how large the benefits of MC1 mating will be for schemes with overlapping generations.

The aim of this study was to compare genetic gain, generation interval and number of selected sires and dams for random and minimum coancestry mating with only one progeny per mating pair. Optimum contribution selection will be used for schemes with overlapping generations and with or without progeny testing of sires. For schemes with overlapping generations, the algorithm of Meuwissen and Sonesson (1998) iterates to obtain the optimum distribution of parents over age-classes. The iteration process is rather computer intensive, and we will also study the effect of omitting this iteration process and using the distribution of parents over age-classes of the previous year. Note that this does not result in a constant distribution of parents over age-classes, because the distribution defined by the contribution of each animal may deviate from the distribution of last year. Finally, we compared schemes with overlapping and discrete generations to see where minimum coancestry mating performs the best.

2. MATERIAL AND METHODS

2.1. Optimum contribution selection

The optimum contribution selection method for discrete generations was presented by Meuwissen (1997) and for overlapping generations by Meuwissen and Sonesson (1998). In principle, the genetic response is maximised with two restrictions. Firstly, there is a restriction on the contribution per sex, *i.e.* each sex have to contribute 50% of the genes to the next generation. Secondly, a constraint on the increase of average relationship of selected parents limits the inbreeding each selection round. For populations with overlapping generations, there are animals of different age-classes at year t from which parents are selected and the different age-classes get different weights, $\mathbf{r}$, which equal their long-term contributions and which are derived from the gene-flow theory of Hill (1974). The weights indicate how much age-classes are expected to contribute in the future. The algorithm iterates on these weights such that the contribution of each individual selection candidate, $\mathbf{c}_t$, and the contributions of each age-class, $\mathbf{r}$, are optimised. For discrete generations, only the contribution of each individual selection candidate is optimised.

This iteration process over age-classes is rather computer intensive because it results in many calculations of optimum $\mathbf{c}_t$. Therefore, we investigated the consequences of omitting the iteration and simply using the distribution of parents of the last year, $\mathbf{c}_{t-1}$, over age-classes to calculate $\mathbf{r}$. This does not result in a constant distribution of parents over age-classes, because the distribution defined by $\mathbf{c}_t$ may deviate from the distribution of last year, which was used to calculate $\mathbf{r}$. For year one, $\mathbf{r}$ was $1/q$, for all age-classes, where q is the number of age-classes.

2.2. Mating

Two mating schemes described in Sonesson and Meuwissen (2000) are used.

For random mating (RAND), sires and dams are allocated at random with a probability that is proportional to the genetic contribution that they received from the selection algorithm. Note that for the random mating scheme, the actual number of offspring deviated from the optimal contributions due to sampling of parents, which introduced some suboptimality in the schemes.

The non-random mating strategy minimises the average relationship of sires and dams, whereby the number of progeny per mating pair is restricted to about 50% more for schemes with overlapping generations than for discrete generations. They compared minimum coancestry mating without the restriction of zero or one progeny per mating pair and compensatory mating on the average coancestry of the candidates to all other selected animals (CMM) with random mating. CMM was more effective in reducing ΔF for less intense schemes (*e.g.* phenotypic selection or large populations) and minimum coancestry mating is more effective for schemes with more intense schemes (*e.g.* BLUP selection and small populations). We could confirm this conclusion in the study of non-random mating for discrete generations (Sonesson and Meuwissen, 2000). The comparison of Caballero *et al.* (1996) between discrete and overlapping schemes is, however, difficult to compare with our schemes. Firstly, in their study, they observed the effect of non-random mating in reduced ΔF, whereas ΔF is restricted in our study, such that the effect of non-random mating can be seen in increased genetic response. Secondly, they did not correct for the increased size of the overlapping scheme compared with the discrete scheme. Thirdly, their discrete schemes were smaller than our schemes and had fixed contributions of the two age-classes instead of optimum contributions of ten age-classes here. Fourthly, they used BLUP selection, for which a larger effect of non-random mating is expected than for optimum contribution selection, because with BLUP selection, progeny from successful families have themselves a higher probability of being selected due to their superior EBVs. A mating strategy, which mates animals from successful families (*i.e.* large) to animals from unsuccessful families avoids this cumulative effect of BLUP selection on the contributions of some families. Optimum contribution selection controls the contributions of families directly. Hence, the above advantage of non-random mating does not hold for optimum contribution selection.

For this study, we have compared different schemes with ten reproductive age-classes for males and females, and females obtained only three repeated records (in order to reduce computation of EBVs). The results of this study show that the older females were hardly selected, because their EBVs were much lower than the EBVs of younger females. With a more stringent restriction on inbreeding, the older females with high EBVs would probably be selected. Thus, the number of reproductive age-classes is especially important for small schemes with low inbreeding (stringent restriction), where the older animals are also competitive candidates, *i.e.* a more stringent ΔF restriction would reduce ΔG more if the older age-classes are not available.

4.2. Iteration on generation intervals

Schemes without iteration on generation intervals used much less computer time. For one replicate of 20 years with a constraint on ΔF of .25%, RAND needed 54 minutes and RAND_NIT needed 11 minutes on an Alpha 500 workstation.

After the breeding schemes had stabilised, there was very little difference between schemes with and without iteration on the distribution of parents over age-classes. However, during the first years of the scheme, when relationships in the population were not yet equalised, the level of inbreeding, and therefore the genetic level also fluctuated more for the scheme without iteration on generation interval than for the scheme with iteration on the generation interval (Figures 1 and 2). This makes the scheme without iteration on generation interval useful for steady-state breeding schemes, where the generation interval is stable and approximately known. If a new breeding scheme is implemented, iteration on the generation interval is desirable such that early fluctuations on inbreeding and genetic levels are reduced.

4.3. Practical schemes

Optimum contribution is a dynamic selection tool, *i.e.* it adapts to the current situation and makes the contribution of families more equal. This makes optimum contribution selection more efficient on maximizing genetic gain with a restriction on inbreeding than a static selection tool (Meuwissen and Sonesson, 1998; Grundy *et al.* 2000). For practical schemes, the actual contribution of each animal can only be aimed at, for example external factors may reduce or increase the actual number of progeny of a certain selected animal. However, in these cases, the next round of optimum contribution selection will adapt to the new situation in an optimal way, *e.g.* by increasing or decreasing the number of selected progeny in a family. Hence, there is correction of contributions over generations. Furthermore, additional (practical) restrictions can be added to the optimum contribution selection algorithm, *e.g.* restrictions on the maximum contribution (number of offspring) an animal can get (see **APPENDIX** of Meuwissen, 1997).

Two factors make optimum contribution selection best suitable for *e.g.* nuclei with elite breeding animals or smaller selection lines. Firstly, the optimum contribution algorithm requires recording of pedigree and control over selection. With non-random mating schemes, the matings also have to be controlled. Secondly, optimum contribution selection performs best for small populations (Meuwissen and Sonesson, 1998; Grundy *et al.* 2000, this study). Increasing size and less stringent restrictions on inbreeding lead to schemes that resemble BLUP selection schemes, although the number of selected animals (especially sires) and generation intervals are not optimized for regular BLUP schemes, as they are for optimum contribution selection.

Recommended ΔF restrictions are in the range of .50 through 1.3% per generation (see for example Goddard, 1992; Meuwissen and Woolliams, 1994; FAO, 1998), although it is difficult to generalize a figure for all situations (the ΔF restriction will also depend on the historical ΔF in the population, the attitude of the breeding company towards risk, the openness of the breeding scheme to animals from outside the scheme *etc*). An alternative dynamic selection method by Wray and Goddard (1994) maximizes genetic gain with a cost factor, *k*, on the average relationship (EBV-*k*A). The main advantages of a restriction on ΔF is that breeders have a better feel for ΔF than for cost factors on inbreeding and that cost factors on inbreeding are difficult to estimate (they include *e.g.* costs due to increased risk, inbreeding depression on (non)breeding goal traits, assessing time horizons). Any failure to account for all costs, results in schemes with a too high ΔF.

Recently, Wu and Schaeffer (2000) presented a study, where the methods of Grundy *et al.* (1994) and Verrier *et al.* (1993) to reduce inbreeding were compared with their new method. All three methods selected for modified EBVs with an artificially reduced effect of parent average on the animal model EBVs. The advantage of these methods is that they require very few changes to current selection methods, although computer simulations are needed to find the optimum weight of the parent average on the modified EBVs. There are, however, four drawbacks of selection for modified EBVs:

1. The number of selected males (and females) is not optimized by these methods, whereas this is an important tool to reduce inbreeding.
2. They do not consider the selected animals as a group, while the effect of an animal on future inbreeding depends on who else was selected.
3. These methods rely on truncation selection, *i.e.* animals above a certain threshold level obtain equal contributions, which may be more practical in some cases, but it is certainly not optimal (Bondesson, 1989; Grundy *et al.*, 1998).
4. If animals have no performance records and no progeny, their EBVs consist only of their parent averages and reducing the weight of the parent average would not affect the ranking of animals.

4.4. Conclusions

In conclusion, MC1 mating increased genetic response with 11 through 18% compared with RAND mating for optimum contribution selection and overlapping generations. The superiority of MC1 increased with a more stringent constraint on ΔF and with smaller sizes of the schemes. When progeny tested bulls were available, they were selected by the optimum contribution algorithm, which increased genetic response, but the effect of MC1 mating was, however, reduced. Omitting the iteration on distribution of parents over age-classes hardly affected the resulting breeding scheme when the scheme had reached its steady-state, but fluctuated more before their steady-state was reached, such that iteration on the generation interval seems, in general, more safe. MC1 mating increases genetic response more for schemes with overlapping generations than for schemes with discrete generations.

ACKNOWLEDGEMENTS

We are grateful to Piter Bijma and Pim Brascamp for helpful comments and suggestions.

LITERATURE

de Boer I.J.M., Meuwissen T.H.E., van Arendonk J.A.M. (1994) Combining genetic and clonal response in a closed dairy cattle nucleus scheme. Anim. Prod. 59: 345-358.

Bondesson L., Mathematical principles of optimal utilisation of clones under diversity constraints. Swedish University of Agricultural Sciences, Section of Forest Biometry, Report 27, Umeå, 1989.

Caballero, A., Santiago, E. and Toro, M.A. (1996) Systems of mating to reduce inbreeding in selected populations. Anim. Sci. 62: 431-442.

FAO (1998) Secondary Guidelines for development of national farm animal genetic resources management plans: Management of small populations at risk. Rome, Italy.

Grundy, B., Caballero, A., Santiago, E. and Hill, W.G. (1994) A note on using biased parameter values and non-random mating to reduce rates of inbreeding in selection programmes. Anim. Prod. 59: 465-468.

Grundy, B., Villanueva, B. and Woolliams, J.A. (1998) Dynamic selection procedures for constrained inbreeding and their consequences for pedigree development. Genet. Res. Camb 72: 159-168.

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

Goddard, M.E. (1992) Optimal effective population size for the global population of black and white dairy cattle. J. Dairy Sci. 75: 2902-2911.

Henderson C.R. (1984) Applications of Linear Models in Animal Breeding, University of Guelph Press, Guelph, Canada.

Hill, W.G. (1974) Prediction and evaluation of response to selection with overlapping generations. Anim. Prod. 18: 117-139.

Meuwissen, T.H.E. (1997) Maximising the response of selection with a predefined rate of inbreeding. J. Anim. Sci. 75: 934-940.

Meuwissen, T.H.E. and Sonesson, A.K. (1998) Maximizing the response of selection with a predefined rate of inbreeding: overlapping generations, J. Anim. Sci. 76: 2575-2583.

Meuwissen, T.H.E. and Woolliams, J.A. (1994) Effective sizes of livestock populations to prevent a decline in fitness. Theor. Appl. Genet. 89: 1019-1026.

Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T. (1989) Minimization or maximization of functions. In. Numerical recipes- the art of scientific computing, University of Cambridge Press, UK, pp. 274- 334.

Sonesson, A.K. and Meuwissen, T.H.E. (2000) Mating schemes for optimum contribution selection with constrained rates of inbreeding, Genet. Sel. Evol. 32: 231-248.

Verrier, E., Colleau, J.J. and Foulley, J.L. (1993) Long-term effects of selection based on the animal model BLUP in a finite population. Theor. Appl. Genet. 87: 446-454.

Woolliams, J.A. (1989) Modifications to MOET nucleus breeding schemes to improve rates of genetic progress and decrease rates of inbreeding in dairy cattle. Anim. Prod. 49: 1-14.

Wray, N.R. and Goddard, M.E. (1994) Increasing long-term response to selection. Genet. Sel. Evol. 26: 431-451.

Wu, L. and Schaeffer, L.R. (2000) Reducing the effect of parent averages from animal solutions in mixed model equations, J. Anim. Breed. Genet. 117: 361-374.

Chapter 6

Minimization of rate of inbreeding for small populations with overlapping generations

A.K. Sonesson and T.H.E. Meuwissen

Institute for Animal Science and Health,
P.O. Box 65, 8200 AB Lelystad, The Netherlands

Published in Genetical Research Cambridge (2001) 77: 285-292
Reproduced by permission of Cambridge University Press

ABSTRACT

We propose a method that minimizes the rate of inbreeding (ΔF) for small unselected populations with overlapping generations and several reproductive age-classes. It minimizes the increase in coancestry of parents and optimizes the contribution of each selection candidate. The carrying capacity of the population is limited to a fixed number of animals per year. When survival rate equalled 100%, only animals from the oldest age-class were selected, which maximized the number of parents per generation, slowed down the turn over of generations and minimized the increase of coancestry across sublines. However, the population became split into sublines separated by age-classes, which substantially increased inbreeding within sublines. Sublines were prevented by a restriction of selecting at least one sire and one dam from the second oldest age-class, which resulted in an L times lower ΔF, where L equals the average generation interval of sires and dams. Minimum coancestry mating resulted in lower levels of inbreeding than random mating, but ΔF was approximately the same. For schemes where the oldest animals were selected, ΔF increased with 18-52% compared with the proposed method.

1. INTRODUCTION

For small and endangered populations, it is important to avoid inbreeding as much as possible, because increased inbreeding leads to inbreeding depression, lower genetic variance and higher frequency of lethal genes. When the breeding of the population can be controlled, there are several methods to reduce or minimize the long-term rate of inbreeding, ΔF. One method equalizes family size (EFS) (Wright, 1931; Wang, 1997a). Recently, an extreme form of EFS was proposed by Wang and Hill (2000) that used marker-assisted selection (MAS) to reduce the variance of each contributed allele. In principle, the effective population size could increase infinitely with MAS, but in practice the amount of marker information, the genome size and the number of marker-genotyped offspring per family is limiting. Another method is the minimization of coancestry selection, where animals are selected such that the average coancestry of parents is minimized. This method is more practical than EFS, because it can handle temporary distortion of family sizes due to, for example, involuntary culling of animals. Finally, the average coancestry of mated animals and thus also the inbreeding level of the next generation can be minimized by minimum coancestry (MC) mating.

The above selection and mating methods minimize inbreeding for populations with discrete generations. In practice, however, generations are overlapping with complex pedigrees and high variance of family size for most species (Ballou and Lacy, 1995). For schemes with overlapping generation structures, there are selection candidates from several reproductive age-classes and an inbreeding minimization method has to account for earlier (unequal) contributions of parents. For overlapping generation structures, Wang *et al.* (1994) tested several heuristic selection indices, which mainly showed that prolonged generation intervals are effective in reducing ΔF.

Robertson (1964) concluded that the overall rate of inbreeding would be minimized if the population was split up in sublines and there was some mixing of the sublines. Within the sublines, inbreeding and genetic drift would be increased, but genetic drift of the entire population would be decreased in the long term. This becomes more apparent if we consider a single biallelic locus with alleles *A* and *a*. In a single population, one of the alleles *A* or *a* is eventually lost, whereas in some sublines the *A* allele will get fixed while in others the *a* allele, *i.e.* the subline structure will maintain both alleles. However, in practice, the sublines may become so highly inbred that they do not survive the inbreeding depression (Wang and Hill, 2000).

For random mating, Wray and Thompson (1990) showed that the variance of long-term contributions from ancestors to descendants, S_c^2, is related to the rate of inbreeding as $\Delta F = (1 + S_c^2)/4N$, where N is the number of breeding animals per generation (assumed to be constant over generations). Because S_c^2 equals zero when all ancestors have equal contributions to the descendants, rates of inbreeding are then minimized. Woolliams and Thompson (1994) extended the above expression to account for non-random mating (*e.g.* MC mating or sublining):

$$\Delta F = (1 + S_c^2)(1 - \alpha)/4N, \qquad [1]$$

where α is the correlation of alleles within individuals. For MC mating, where animals with least similar alleles are mated, animals will show a negative correlation between their alleles ($\alpha < 0$). Therefore, ΔF can actually be increased by MC mating, *i.e.* the term $(1 - \alpha) > 1$. However, MC mating also connects unrelated families with each other, such that the genetic contributions become more equal, *i.e.* it reduces S_c^2 and therefore also ΔF (Caballero and Toro, 2000).

The aim of this paper is to develop a method that minimizes inbreeding for small populations with overlapping generations and complex pedigrees. We test the effects of minimum coancestry selection and MC mating on inbreeding for overlapping generation structures. We consider a situation where the number of animals born per time period limits the population size. In practice, all mature females of endangered breeds will be mated, but since the carrying capacity of the local captive population is limited, the offspring from some females will be sold, moved to other populations or released into wild populations.

2. MATERIAL AND METHODS

2.1. Minimization of coancestry selection for overlapping generations (MCO)

The objective is to minimize the average coancestry of parents. We want to select parents and set up matings for the N new-born animals in the population of the next year. The problem is to find an optimum $\mathbf{c}$, which minimizes $\mathbf{c}'\mathbf{Ac}$ and thus also $\frac{1}{2}\mathbf{c}'\mathbf{Ac}$, where $\mathbf{c}$ is the contribution for each selection candidate and $\mathbf{A}$ is an ($n \times n$) matrix with relationship coefficients between selection candidates, which equal two times the coancestry coefficients (Meuwissen, 1997). For schemes with overlapping generations, animals are assumed to be in discrete yearly age-classes, but other time units can also be used. Animals up to and including the oldest reproductive age have the potential to carry genes of future generations, so also young age-classes that are not yet reproducing have to be considered. The average relationships in a population with overlapping generations are obtained by weighing the average relationships within and between age-classes by the long-term contribution of the age-classes ($\mathbf{r}$), which is obtained from Hill's (1974) gene-flow method (Meuwissen and Sonesson, 1998):

$$r_1^2\,\mathbf{c}'\mathbf{Ac} + 2r_1\,\mathbf{c}'\mathbf{AJ}\,\mathbf{r}_2 + \mathbf{r}_2'\,\overline{\mathbf{A}}_{22}\,\mathbf{r}_2\,, \qquad [2]$$

where subscripts 1 and 2 denote the youngest age-class (new progeny) and older age-classes respectively (r_1 = long term contributions of new progeny age-class from Hill's (1974) gene-flow method, $\mathbf{r}_2$ = ((q - 1) × 1) vector of long term contributions of older age-classes and q = total number of age-classes), $\overline{\mathbf{A}}_{22}$ = (q-1) × (q-1) matrix of average relationships between older age-classes. The ($\mathbf{c}'\mathbf{Ac}$) term equals the average relationship between new progeny, the ($\mathbf{c}'\mathbf{AJ}\,\mathbf{r}_2$) term equals the relationship between old animals and new progeny and the term $(\mathbf{r}_2' \overline{\mathbf{A}}_{22}\, \mathbf{r}_2)$ equals average relationships between old animals. The **J** matrix is a (n × (q-1)) matrix that averages individual relationship to average relationship per age-class, *i.e.* $J_{ij} = 1/m_j$ if animal i belongs to age-class j, and $J_{ij} = 0$ otherwise, where m_j is the number of animals in age-class j. The relationships between older animals $(\mathbf{r}_2' \overline{\mathbf{A}}_{22}\, \mathbf{r}_2)$ are not affected by the current selection, because relationship between old animals are already established.

Because the genetic contributions of all males and of all females must sum to ½, a restriction is added:

$$\mathbf{Q}'\mathbf{c} = ½\mathbf{1}_2 \,, \qquad [3]$$

where **Q** is a known incidence matrix for sex; the first column yields ones for males and zeros for females and the second column yields ones for females and zeros for males; $\mathbf{1}_2$ is a (2 × 1) vector of ones. In the **APPENDIX**, Equation [2] is minimized under the restriction [3], giving:

$$\mathbf{c} = \mathbf{A}^{-1}\mathbf{Q}\boldsymbol{\lambda}/2r_1^2 - \mathbf{J}\,\mathbf{r}_2/r_1 \,, \qquad [4]$$

where $\boldsymbol{\lambda}$ equals

$$\boldsymbol{\lambda} = 2r_1\left(\mathbf{Q}'\mathbf{A}^{-1}\mathbf{Q}\right)^{-1}\left(r_1\mathbf{s} + \mathbf{Q}'\mathbf{J}\,\mathbf{r}_2\right). \qquad [5]$$

In the previous section, the vector of long-term contributions, **r**, was assumed known. The vector is calculated from the gene flow matrix between age-classes, **P**, and depends on the contribution of each age-class to age-class 1 (Hill, 1974). These contributions depend however on the age of selected animals, *i.e.*, on **c**. Thus, iteration is needed to find optimal **c** and **r** simultaneously (Meuwissen and Sonesson, 1998). Alternatively, simulated annealing could be used to calculate optimum **r** (Grundy *et al.*, 1998). Both algorithms yielded very similar results on selection schemes (Sonesson *et al.*, 2000). The iteration algorithm was used here, because it requires less computer time.

2.2. Mating schemes

Minimum coancestry (MC) mating. Minimum coancestry seems a useful objective to determine the mating pairs, since it minimizes the inbreeding coefficients of the offspring and therefore avoids inbreeding depression as much as possible for the next generation (Wright, 1921). MC mating will be used here in combination with MCO selection schemes.

To find the mating structure that yields the minimum coancestry among the mating pairs, a matrix **F** of size ($N_s \times N_d$) is set up, where N_s(N_d) is the number of selected sires (dams) and element $\mathbf{F}_{ij}$ is the coefficient of coancestry of the selected individuals *i* and *j*, which is also the inbreeding coefficient of their progeny. The number of matings for each animal is obtained from its optimum contribution, **c**, and is calculated by the selection algorithm by multiplying the contributions of the sire (dam) by the number of selection candidates and rounding to integers. Normally, the truncation point for rounding up versus down is .5, but if the total number of progeny does not sum to the intended number, the truncation point is adjusted such that the intended number of progeny results. This is not guaranteed to give the optimum integer solution, but yields probably a reasonable approximation.

The MC mating structure is obtained using the simulated annealing algorithm (Press *et al.*, 1989). The implementation of the annealing algorithm is described in the **APPENDIX** of Sonesson and Meuwissen (2000).

Random (RAND) mating. For the random mating schemes (RAND), the simulated annealing algorithm is used as explained in the MC mating section above, but all changes were accepted and the total number of random changes was set to 10000. In this way, the number of progeny per sire (dam) was restricted to the numbers defined by MCO, while still achieving random mating (Sonesson and Meuwissen, 2000). In practice, more simple algorithms can be devised to achieve this, but note that simply sampling sires and dams according to the proportions defined by the **c** vector results in random deviations from the number of offspring defined by MCO.

2.3. Simulated schemes

For overlapping generation schemes, a closed nucleus scheme was simulated for 100 years. Schemes were symmetric with respect to the two sexes and there were no animals with limited reproductive capacity. One hundred replicated simulations were obtained for each scheme.

For convenience, it was assumed that one year equals one age-class. Animals started to reproduce at age 1 until the oldest reproductive age 9, resulting in 9 reproductive age-classes. Survival rate was 70%, which means that each individual has the probability of 70% to survive every selection round (year).

Although a survival rate of 100% is not realistic, schemes with 100% survival were also simulated in order to test whether the algorithm would create sublines of age-classes. There were 4 reproductive age-classes (1, 2, 3 and 4) for schemes with 100% survival. The variance of contributions of ancestors to descendants, S_c^2, was obtained by defining an entire generation of ancestors as a consecutively born group of animals, whose long-term contribution sum to 1 (Bijma and Woolliams, 1999). The correlation of alleles within an animal, α, was derived from Wright's (1969) F-statistics using

$$(1-F)=(1-\alpha)(1-k),$$

where F is the inbreeding of offspring (Wright's F_{IT}); k = average kinship among parents (equals inbreeding of offspring with random mating; Wright's F_{ST}) and α denotes a deviation of inbreeding due to non-random mating (Wright's F_{IS}).

The ΔF and increase of coancestry, ΔC, were calculated as an average over years 96-100. For overlapping generations and 70 or 100% survival, RAND and MC were compared for 6, 10 or 20 new-born animals per year.

3. RESULTS

3.1. 100% survival

For schemes with overlapping generations and 100% survival, the maximum generation interval was four years and four age-classes (1, 2, 3 and 4) were eligible for selection. For both RAND and MC mating, MCO selected only the oldest animals (Table 1), which maximizes the number of parents per generation and slows down the turn over of generations. However, it also results in complete sublining of age-classes as in the example of Table 2. Note that complete sublining requires that all parents come from the same age-class (age-class 4 in Table 2), which results in a small number of parents when few parents survive up to this age-class. Hence, a survival rate of 100% maximizes the probability of sublining. Complete sublining resulted in high ΔF (.0072 and .0069 for schemes with 10 new-born animals per year and RAND and MC mating, respectively) and ΔC (.0067 and .0070 per year for RAND and MC mating, respectively), because they are then calculated within age-classes. The aim of the MCO algorithm is, however, to minimize the overall ΔC, *i.e.* within and between age-classes, which is achieved if there is very little mixing

Table 1. Rate of inbreeding per year (ΔF), increase of coancestry per year (ΔC), generation interval of sires/dams (L), number of selected sires/dams (Nsel), number of male/female selection candidates (Ncand) for schemes with 6, 10 and 20 new-born animals per year and 100% survival.

Mating method*	ΔF†	ΔC†	L (years)	Nsel	Ncand
		6 animals			
RAND	.0117	.0119	4.0/4.0	3.0/3.0	12/12
MC	.012	.0123	4.0/4.0	3.0/3.0	12/12
		10 animals			
RAND	.0072	.0067	4.0/4.0	5.0/5.0	20/20
MC	.0069	.0070	4.0/4.0	5.0/5.0	20/20
		20 animals			
RAND	.0033	.0032	4.0/4.0	1.0/1.0	40/40
MC	.0034	.0033	4.0/4.0	1.0/1.0	40/40

* RAND = random mating, MC = minimum coancestry mating

† Average of year 96-100, standard errors are between .0000 and .0020

Table 2. An illustration of the occurrence of sublining in a breeding scheme where all animals have a generation interval of 4 years. The four sublines are denoted A, B, C and D and are separated by time.

Year	Age classes*			
	1	2	3	4
1	A	B	C	D
2	D	A	B	C
3	C	D	A	B
4	B	C	D	A
5	A	B	C	D
...	...	...	...	...

* When going from one year to the next, sublines in age-classes 1, 2 and 3 became one year older and the subline in age-class 4 reproduces to obtain offspring in age-class 1 of the next year.

of sublines (Robertson, 1964). Therefore, we included an extra restriction to restriction [3] on the distribution of parents over age-classes, such that MCO would not select only animals from the oldest age-class, but also one sire and one dam from the second oldest age-class. This is not the minimum possible mixing of sublines in a steady-state breeding scheme.

For these schemes with a restricted distribution of parents over age-classes, the generation interval was 3.89 years for both RAND and MC mating schemes with 10 new-born animals per year (Table 3). The ΔF (.0017 and .0018 per year for RAND and MC mating, respectively) and ΔC (.0016 and .0017 per year for RAND and MC mating, respectively) was similar for the two mating schemes. Note that ΔF was approximately L times smaller for this scheme compared with the scheme with sublines (Table 1), because of the L times larger population size.

In general, the two different mating schemes resulted in very similar breeding schemes: generation interval, number of selected animals and selection candidates were the same for both mating schemes. The correlation between alleles within individuals, α, and the variance of long-term contributions, S_c^2, were somewhat lower for the MC scheme compared with the RAND scheme, but these differences where not large enough to affect predicted rates of inbreeding, ΔF_p. The α equalled -.046 for RAND mating, which equals the expected value $-1/(2N-1)$ (Caballero, 1994), where N is the number of selected animals per year.

The S_c^2 and α were not given for the scheme with sublines (Table 1), because sublines did continuously drift further apart and thus S_c^2 and α did not reach stable values.

3.2. 70% survival

Similar to schemes with 100% survival, ΔF was about the same for the MC mating (.0046), and RAND mating (.0043) for the schemes with 70% survival (Table 4). The α was low for the MC scheme (-.074) compared with the RAND scheme (-.043), S_c^2 was about the same (2.55 and 2.65 for MC and RAND mating, respectively) and the predicted rate of inbreeding, ΔF_p, was about the same for both mating schemes.

For schemes with 6 and 20 new-born animals per year, results were similar to schemes with 10 new-born animals per year, except that the variation was larger for the small schemes with only three animals per sex, because of a larger sampling error (not shown) and the effect of culling, especially for RAND mating.

The results were compared with ΔF for populations, where simply the oldest (6, 10, and 20) animals were selected each year and there was random mating. The ΔC increased 31, 42 and 39% for schemes with 6, 10 and 20 new-born animals, respectively, for a survival rate of 50% compared with the MCO algorithm and random mating. For the survival rate of 70%, ΔC was 41, 36 and 52% higher and finally, for the survival rate of 90%, ΔC was 18, 31 and 50% higher for schemes with 6, 10 and 20 new-born animals, respectively. We looked at ΔC instead of ΔF, because of the smaller standard errors of ΔC.

There was no clear change of pattern of the superiority of the MCO schemes over the schemes that selected only the oldest animals when survival rate increased. This is possibly because a low survival rate results in most animals still being present in age-class 1, *i.e.* a scheme with an almost discrete generation structure, whereas a very high survival rate (*e.g.* 100%) also results in almost discrete generations. Intermediate survival rates result in schemes with different degrees of overlapping generations.

4. DISCUSSION

4.1. Minimum coancestry selection

A method has been presented that minimizes the rate of inbreeding for populations with overlapping generation structure and complex pedigrees. The algorithm operates at the level of coancestry and not inbreeding, because an initial increase in coancestry is more closely related to long-term inbreeding than the initial increase of inbreeding (Wray and Goddard, 1994).

If survival rate is 100%, selection from only the oldest age-class minimizes ΔC across age-classes, because ΔC is zero between age-classes, and the population splits up into L sublines separated by time, where L is the generation interval. Rate of inbreeding within a subline will be L times higher (Table 1) than the minimized rate of inbreeding that is achieved by the restriction that at least one sire and one dam comes from the second oldest age-class (Table 3). This restriction may seem somewhat *ad hoc*, but it should be noted that the animal that is selected from the second-oldest age-class will generally be avoided for selection next year, because this animal has already offspring. The result is that the variance of family size still is small and S_c^2 is only .033-.084, implying that ΔF is only 3.3-8.4% higher than its theoretical minimum value, which is achieved when $S_c^2=0$. The latter theoretical minimum can not be achieved, because the mixing of sublines results in $S_c^2 > 0$.

For practical schemes with survival rates less than 100%, the oldest age-classes contain too few animals and some younger animals have to be selected. The algorithm then weights the selection of more and consequently younger animals against the reduction of the generation interval, accounting for the pedigree of available animals. A simple scheme, where only the oldest animals were selected resulted in 18-52% higher ΔC and thus ΔF, than the MCO scheme.

Table 3. Rate of inbreeding per year (ΔF), increase in coancestry per year (ΔC), variance of longterm contributions (S^2_c), correlation between alleles within an individual (α), predicted ΔF (ΔF_p), generation interval of sires/dams (L), number of selected sires/dams (Nsel) and number of male/female selection candidates (Ncand) for schemes with 6, 10 and 20 newborn animals per year and 100% survival, but with restriction on distribution of parents over ageclasses.*

Mating method †	ΔF ‡	ΔC ‡	S^2_c	α	ΔF_p §	L (years)	Nsel	Ncand
			6 animals					
RAND	.0018	.0028	.035	-.0754	.0031	3.82/3.82	3.9/4.0	12/12
MC	.0029	.0030	.033	-.1090	.0032	3.83/3.82	3.9/3.9	12/12
			10 animals					
RAND	.0017	.0016	.084	-.0459	.0018	3.89/3.89	5.9/5.9	20/20
MC	.0018	.0017	.073	-.0814	.0019	3.89/3.89	5.9/5.9	20/20
			20 animals					
RAND	.0006	.0009	.042	-.0242	.0012	3.94/3.94	1.9/1.8	40/40
MC	.0009	.0009	.038	-.05129	.0012	3.94/3.94	1.9/1.8	40/40

* At least one sire and one dam are from the second oldest ageclass

† RAND = random mating, MC = minimum coancestry mating

‡ Average of years 96-100, standard errors are between .0001 and .0020

§ ΔF_p is predicted with Equation 1 using observed S^2_c, N and α

Table 4. Rate of inbreeding per year (ΔF), increase of coancestry per year (ΔC), variance of longterm contributions (S^2_c), correlation between alleles within an individual (α), predicted ΔF (ΔF_p), generation interval of sires/dams (L), number of selected sires/dams (Nsel), number of male/female selection candidates (Ncand) for schemes with 6, 10 and 20 newborn animals per year and 70% survival.

Mating method*	ΔF†	ΔC†	S^2_c	α	ΔF_p‡	L (years)	Nsel	Ncand
			6 animals					
RAND	.0077	.0079	2.71	-.0737	.0079	4.43/4.44	4.1/4.1	9.6/9.5
MC	.0083	.0078	2.8	-.1078	.0082	4.46/4.37	4.2/4.3	9.6/9.7
			10 animals					
RAND	.0043	.0044	2.65	-.0427	.0046	4.39/4.46	7.1/7.1	16.0/15.6
MC	.0046	.0045	2.55	-.0744	.0045	4.50/4.54	7.3/7.1	15.7/16.2
			20 animals					
RAND	.0021	.0021	2.57	-.0202	.0021	4.60/4.56	14.6/14.6	32.5/32.1
MC	.0022	.0022	2.48	-.044	.0022	4.56/4.45	14.7/14.6	31.6/32.3

* RAND = random mating, MC = minimum coancestry mating

† Average of year 96-100, standard errors are between .0001 and .0020

‡ ΔF_p is predicted with Equation 1 using observed S^2_c, N and α

4.2. MC mating

It is expected that in the short term, inbreeding levels would decrease, but that in the long term, ΔF would increase for MC mating compared with RAND mating (Kimura and Crow, 1963; Robertson, 1964; Ballou and Lacy, 1995; Wang, 1997b and others). The short and especially long-term effect of MC mating on inbreeding was however small here: the inbreeding level was lower for the MC mating than for the RAND mating scheme during all the 100 years (Figure 1), but the inbreeding of MC was catching up with the inbreeding of RAND, resulting in a slightly higher ΔF for MC mating (Table 3). At around year 300 (results not shown), when the inbreeding was relatively high, the graphs with inbreeding values in Figure 1 crossed over such that MC mating had marginally higher inbreeding values than RAND mating. Thus, although ΔF was slightly higher for MC mating, the actual inbreeding levels started lower for MC mating, caught up only at year 300 and were thereafter slightly higher.

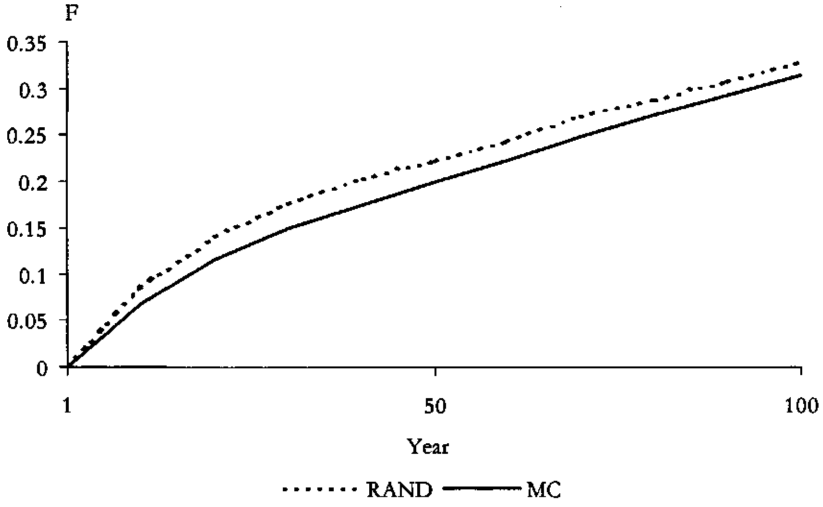

Figure 1. Inbreeding (F) for schemes with 6 new-born animals per year and random (RAND) or minimum coancestry (MC) mating and 100% survival.

4.3. Practical schemes

The different mating strategies hardly affected ΔF, but because the level of inbreeding was lower for a long time period (more than 100 years), MC mating is preferred over RAND mating for the present schemes. For schemes with, for example, short generation intervals, the inbreeding levels of MC and RAND mating might become similar earlier than in Figure 1, which makes inbreeding levels of MC and RAND mating schemes even more equal.

The schemes presented require full control of the family size and matings of animals. In practice, this may not be possible and the optimum contributions, **c**, can only be aimed for. Under these circumstances, the MCO algorithm seems most useful, because it adapts to the pedigree of the available animals (assuming the pedigree is known). For instance, if one family becomes too large by chance, its members will get lower contributions, and *vice versa.* However, ΔF may be substantially higher than in Tables 3 and 4, because of the reduced control over family size. Similarly, the MCO algorithm will adapt to possible distortions of the sex ratio of the populations. In situations where the population increases in size, MCO is expected to optimize selection across age-classes, where it balances selection from younger age-classes that contain more animals, against selection from old age-classes that contain few animals.

When a population is geographically split up into several subpopulations, there are other factors that have to be considered, *e.g.* transportation difficulties and costs and social structure of the populations. This study does not consider these factors, but the subpopulations can be seen as the sublines of Table 1 that can easily become highly related when there is too little mixing.

4.4. Conclusions

1. With overlapping generations and 100% survival, average coancestry across the whole population can be minimized by breeding only from the oldest age-class. This splits the population into sublines defined by age and increases inbreeding within sublines. The MCO algorithm does result in complete sublining with 100% survival.
2. With a restriction on the distribution of parents over age-classes, such that not only animals from the oldest age-class, but also one animal from the second oldest age-class were selected, ΔF became L times smaller than for the schemes with sublines, because the population was L times larger.
3. If survival is less than 100%, the algorithm weights the selection of more and consequently younger animals against the reduction of the generation interval, accounting for the pedigree of the available animals.
4. Based on these results, MCO and MC mating would be desirable for practical populations, because this resulted in lower levels of inbreeding for more than 100 years. Thereafter, inbreeding levels of RAND and MC mating were very similar. The ΔF was however somewhat higher for the MC mating schemes, because the inbreeding levels of the MC mating schemes caught up with the RAND schemes.
5. A scheme where only the oldest animals are selected led to 18-52% higher ΔC, and thus ΔF, than MCO selection.

ACKNOWLEDGEMENTS

We are grateful to Mike Goddard, Piter Bijma, Pim Brascamp, Ab Groen, Kor Oldenbroek and an anonymous referee for helpful comments and suggestions.

LITERATURE

Ballou, J.D. and Lacy, R.C. (1995) Identifying genetically important animals for management of genetic variation in pedigreed population. In. Population management for survival and recovery- analytical methods and strategies in small population conservation. Ed. J.D. Ballou, M. Gilpin and T.J. Foose. Columbia University Press, New York, USA.

Bijma, P. and Woolliams, J.A. (1999) Prediction of genetic contributions and generation intervals in populations with overlapping generations under selection. Genetics 151: 1197-1210.

Caballero, A. (1994) Developments in the prediction of effective population size. Heredity 73: 657-679.

Caballero, A. and Toro, M.A. (2000) Interrelations between effective population size and other pedigree tools for the management of conserved populations. Genet. Res. Camb. 75: 331-343.

Grundy, B., Villanueva, B. and Woolliams, J.A. (1998) Dynamic selection procedures for constrained inbreeding and their consequences for pedigree development. Genet. Res. Camb 72: 159-168.

Hill, W.G. (1974) Prediction and evaluation of response to selection with overlapping generations. Anim. Prod. 18: 117-139.

Kimura, M. and Crow, J.F. (1963) On the maximum avoidance of inbreeding. Genet. Res. Camb. 4: 399-415.

Meuwissen, T.H.E. (1997) Maximising the response of selection with a predefined rate of inbreeding. J. Anim. Sci. 75: 934-940.

Meuwissen T.H.E. and Sonesson A.K. (1998) Maximizing the response of selection with a predefined rate of inbreeding: overlapping generations. J. Anim. Sci. 76: 2575-2583.

Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T. (1989) Minimization or maximization of functions, in: Numerical recipes- the art of scientific computing. University of Cambridge Press, Cambridge, UK, pp. 274- 334.

Robertson, A. (1964) The effect of non-random mating within inbred lines on the rate of inbreeding. Genet. Res. 5: 164-167.

Sonesson, A.K. and Meuwissen, T.H.E. (2000) Mating schemes for optimum contribution selection with constrained rates of inbreeding. Genet. Sel. Evol. 32: 231-248.

Sonesson, A.K., Grundy, B., Woolliams J.A. and Meuwissen, T.H.E. (2000) Selection with control of inbreeding in populations with overlapping generations: a comparison of methods. Anim. Sci. 70: 1-8.

Wang, J.L., Xu, J.C., Song, J.Z. and Tang, W.S. (1994) Study on mating systems for controlling inbreeding levels in domestic animal populations with overlapping generations. Proc. 5th World Congress on Genetics Applied to Livestock Production 21: 536-539.

Wang, J.L. (1997a) More efficient breeding systems for controlling inbreeding and effective size in animal populations. Heredity 79: 591-599.

Wang, J.L. (1997b). Effect of excluding sib matings on inbreeding coefficient and effective size of finite populations. Biometrics 53: 1354-1365.

Wang, J. and Hill, W.G. (2000) Marker-assisted selection to increase effective population size by reducing mendelian segregation variance. Genetics 154: 475-489.

Woolliams, J.A. and Thompson, R. (1994) A theory of genetic contributions. Proc. of the 5th World Congress on Genetics Applied to Livestock Production, 19: 127-134.

Wright, S. (1921) Systems of mating. Genetics 6: 111-178.

Wright, S. (1931) Evolution in Mendelian populations. Genetics 16: 97-159.

Wright, S. (1969) Evolution and the Genetics of Populations, vol. 2, The Theory of Gene Frequencies. University of Chicago Press, Chicago, USA.

Wray, N.R. and Thompson, R. (1990) Prediction of rates of inbreeding in selected populations. Genet. Res. Camb. 55: 41-54.

Wray, N.R. and Goddard, M.E. (1994) Increasing long-term response to selection. Genet. Sel. Evol. 26: 431-451.

APPENDIX

Calculation of optimum contributions

The average relationship in Equation [2] will be minimized under the restriction of the genetic contributions per sex in Equation [3]. For the optimization algorithm, a Lagrangian multiplier, $\boldsymbol{\lambda}$, is used as a cost factor on the genetic contributions per sex and are calculated such that the contribution per sex actually becomes .5. The lagrangian function to be optimized, *H*, is:

$$H = r_1^2 \mathbf{c}' \mathbf{A} \mathbf{c} + 2 r_1 \mathbf{c}' \mathbf{A} \mathbf{J} \mathbf{r}_2 + \mathbf{r}_2' \overline{\mathbf{A}}_{22} \mathbf{r}_2 - (\mathbf{c}' \mathbf{Q} - \mathbf{s}') \boldsymbol{\lambda} . \quad [A1]$$

Equating the derivative of *H* with respect to the genetic contributions of the new-born progeny, **c**, to zero yields $2 r_1^2 \mathbf{A} \mathbf{c} + 2 r_1 \mathbf{A} \mathbf{J} \mathbf{r}_2 - \mathbf{Q} \boldsymbol{\lambda} = 0$.

Solving for **c** yields

$$c = \mathbf{A}^{-1} \mathbf{Q} \boldsymbol{\lambda} / 2 r_1^2 - \mathbf{J} \mathbf{r}_2 / r_1 . \quad [A2]$$

Equating the derivative of *H* with respect to the cost factor on the contribution per sex, $\boldsymbol{\lambda}$, to zero yields

$$\mathbf{s}' = \mathbf{c}' \mathbf{Q} . \quad [A3]$$

Substituting the solutions for **c** into Equation [A3] and solving for $\boldsymbol{\lambda}$, gives

$$\boldsymbol{\lambda} = 2 r_1 \left(\mathbf{Q}' \mathbf{A}^{-1} \mathbf{Q} \right)^{-1} \left(r_1 \mathbf{s} + \mathbf{Q}' \mathbf{J} \mathbf{r}_2 \right) . \quad [A4]$$

Finally, the solution for $\boldsymbol{\lambda}$ in Equation [A4] is substituted into Equation [A2] in order to get **c** for each selection candidate.

Chapter 7

The use of frozen semen to minimize inbreeding in small populations

Anna K. Sonesson [*], Mike E. Goddard [†] and Theo H.E. Meuwissen [*]

[*] Institute for Animal Science and Health, P.O. Box 65, 8200 AB Lelystad, The Netherlands

[†] Institute of Land and Food Resources, University of Melbourne, Parkville, Victoria 3052 and Victorian Institute of Animal Science, Attwood, Victoria 3049, Australia

Accepted by Genetical Research Cambridge
Reproduced by permission of Cambridge University Press

ABSTRACT

In this study, we compared the average coancestry and inbreeding levels for two genetic conservation schemes where frozen semen from a gene-bank is used to reduce the inbreeding in a live population. For a simple scheme where only semen of generation zero sires is used, the level of inbreeding asymptotes to $1/(2N)$, where N is the number of newborn sires in the base generation and rate of inbreeding goes to zero. However, when only sires of generation zero are selected, all genes will eventually descend from the founder sires and all genes from the founder dams are lost. We propose an alternative scheme where, next to the N sires from generation 0, also N sires from generation one have semen conserved, and the semen of generation 0 and 1 sires is used alternatively for dams of odd and even generation numbers, respectively. With this scheme, the level of inbreeding asymptotes to $1/(3N)$ and also genes of founder dams are conserved, because 50% of the genes of sires of generation one are derived from the founder dams. A computer simulation study shows that this is the optimum design to minimize inbreeding, even if semen from later generations is available.

1. INTRODUCTION

A common goal of genetic conservation schemes is to maintain a living population with minimum genetic drift and inbreeding. One way to reach this goal is to use old, less related males. This slows down the turn over of generations and thus reduces genetic drift and inbreeding.

Smith (1977) took this idea to the extreme and proposed to use only frozen semen of the least related sires from the founder generation (generation zero), in order to keep genetic drift to a minimum. However, if only sires from generation zero are selected, all genes will eventually descend from the founder sires, because in generation one, 50% of the genes come from the sires of generation zero, in generation two, 75%, in generation three, 87.5%, *etc.* Thereby, there will eventually be no genes of the females of generation zero (if frozen oocytes and embryos cannot be used). However, if sires from generation one as well as generation zero were used, genes of founder dams would be conserved in the progeny, because males from generation one have 50% of their genes from founder dams. Here, we propose to mate sires of generation zero to dams from odd generation numbers and sires of generation one to dams from even generation numbers.

The aim of this paper is to compare inbreeding and average coancestry, $\overline{C}_t$, of a scheme where only sires of generation zero are used (Smith, 1977) to a scheme where sires of generation zero and one are used. The average coancestry is a good measure of genetic diversity, because it accounts for both founder effects and effects of genetic drift with respect to contributions and variance of family size and it has a direct relationship with the expected heterozygosity (Lacy, 1995; Caballero and Toro, 2000).

2. MATERIAL AND METHODS

2.1. Derivation of average coancestry

The coancestry of two individuals is defined as the probability that two gametes taken at random, one from each individual, carry alleles that are identical by descent (Malécot, 1948). The average coancestry of animals in a population, $\overline{C}_t$, can be calculated as the average of all elements in a matrix of coancestry between all individuals. We can split up the total average coancestry $\overline{C}_t$ in three terms taken from the coancestry matrix

$$\overline{C}_t = 0.25\overline{C}_s + 0.25\overline{C}_d + 0.50\overline{C}_{sd}, \qquad [1]$$

where $\overline{C}_s$ and $\overline{C}_d$ are the average coancestry of sires and dams, respectively, and $\overline{C}_{sd}$ is the average coancestry between sires and dams.

Repeated use of frozen semen from sires of generation zero. For this scheme, N sires from generation zero are used every generation. Every sire is randomly mated to one of the N dams to produce one female offspring per mating pair.

The coancestry of generation 0 (G_0) sires, inclusive of the coancestry of each sire with itself $\mathbf{C}_s = \mathit{diag}(.5)$, is an $N \times N$ matrix where the diagonal term represents the coancestry of a sire with itself (.5) and the off-diagonal terms represents the coancestry between different G_0 sires, which is zero here. It follows that $\overline{C}_s = N \times 1/(2N^2) = 1/(2N)$.

The average coancestry between sires and dams, $\overline{C}_{sd}$, is the average of the $N \times N$ matrix $\mathbf{C}_s = \mathbf{J}\overline{F}$, where $\mathbf{J}$ = $N \times N$ matrix of ones, and the average coancestry of sires and dams equals, by definition, the average inbreeding of the population, $\overline{F}$, when there is random mating. It follows that $\overline{C}_{sd} = \overline{F}$.

$\overline{F}$ can be derived from $c_{x,y} = ½(c_{x,sy} + c_{x,dy})$, where $c_{x,y}$ denotes the coancestry of the sire *x* and dam *y*, and the sire and dam of *y* are denoted by *sy* and *dy*. At equilibrium, the expectation of $c_{x,sy}$ is $\overline{C}_s$ and the expectation of both $c_{x,dy}$ and $c_{x,y}$ is $\overline{F}$, because *x* is a sire and *y* and *dy* are dams and $\overline{F}$ has stabilized over generations. It follows that $\overline{F} = ½[1/(½N) + \overline{F}]$ and thus $\overline{F} = 1/(2N)$.

The average coancestry of different dams, inclusive of the coancestry of each dam with itself, $\overline{C}_d$, is the average of the *N* × *N* matrix

$$\mathbf{C_d} = \begin{bmatrix} ½ + ½\overline{F} & \dots & C_d^* \\ \dots & \dots & \dots \\ C_d^* & \dots & ½ + ½\overline{F} \end{bmatrix},$$

where C_d^* is the coancestry between two different dams. C_d^* can be derived from $c_{x,y} = ¼(c_{sx,sy} + c_{sx,dy} + c_{sy,dx} + c_{dx,dy})$, where $c_{x,y}$ denotes the coancestry between dam *x* and *y*. Because family sizes are equal in order to minimize inbreeding, every sire has one daughter, *i.e.* different daughters have different founder sires. It follows that $c_{sx,sy} = 0$. At equilibrium, the expectation of both $c_{sx,dy}$ and $c_{sy,dx}$ is $\overline{F}$ and the expectation of $c_{dx,dy}$ and $c_{x,y}$ is C_d^*. Hence, $C_d^* = ¼[0 + 1/N + C_d^*]$ and thus $C_d^* = 1/(3N)$.

Hence the average of the elements of the $\mathbf{C_d}$-matrix is:

$$\overline{C}_d = [N(½ + ½/(2N)) + N(N-1)/(3N)]/N^2 = [(5/6) - 1/(12N)]/N.$$

Finally, from Equation [1], the equilibrium average coancestry is:

$$\overline{C}_t = 7/(12N) - 1/(48N^2) = 7/(12N) - 1/(48N^2).$$

Alternating the use of sires from generation zero and one. For this scheme, *N* sires of generation 0, G_0, and *N* sires of generation one, G_1 are used; at odd generation numbers, G_0 sires, and at even generation numbers, G_1 sires are used. Every sire is randomly mated to one of the *N* dams to produce one male and one female offspring per mating pair. As for the previous scheme, there is random mating of sires and dams every generation.

The average coancestry of sires from different sires of the same generation, inclusive of the coancestry of each sire with itself, is $\overline{C}_s = 1/(2N)$, *i.e.* the same as for the previous scheme. The $\overline{C}_s$ is the same for the two schemes because G_1 sires are unrelated and non-inbred and have thus the same $\overline{C}_s$ as G_0 sires.

The average coancestry between sires and dams, $\overline{C}_{sd}$, is as in the previous scheme, equal to $\overline{F}$. The $\overline{F}$ can be derived from $\overline{F} = c_{x(0),y} = \tfrac{1}{2}\left(c_{x(0),sy(1)} + c_{x(0),dy}\right)$, where $c_{x(0),y}$ refers to the coancestry between a sire of generation zero, *x(0)*, and one dam, *y*, and where subscript *(0)* denotes generation number 0. At equilibrium, the expectation of $c_{x(0),sy(1)}$ is $1/(4N)$, where ¼ represents the coancestry between a father of G_0 and a son of G_1. The $c_{x(0),dy} = \tfrac{1}{2}\left(c_{x(0),sdy(0)} + c_{x(0),ddy}\right)$, where *sdy(0)* refers to the G_0 sire of *dy* and *ddy* refers to the maternal grand-dam of *y*. The expectation of $c_{x(0),sdy(0)}$ is $\overline{C}_s$, because both sires are from generation zero. The expectation of $c_{x(0),ddy}$ is $\overline{F}$, because *x(0)* is a sire and *ddy* is a dam. It follows that $\overline{C}_{sd} = \overline{F} = \tfrac{1}{2}\left[1/(4N) + \tfrac{1}{2}\left(1/(2N) + \overline{F}\right)\right]$ and thus $\overline{F} = 1/(3N)$.

The average coancestry of different dams, inclusive of the coancestry of each dam with itself, $\overline{C}_d$, is the average of the $N \times N$ matrix

$$\mathbf{C_d} = \begin{bmatrix} \tfrac{1}{2} + \tfrac{1}{2}\overline{F} & \dots & C_d^{*} \\ \dots & \dots & \dots \\ C_d^{*} & \dots & \tfrac{1}{2} + \tfrac{1}{2}\overline{F} \end{bmatrix},$$

where C_d^{*} is the coancestry between two different dams. C_d^{*} can be derived from $c_{x,y} = \tfrac{1}{4}\left(c_{sx,sy} + c_{sx,dy} + c_{sy,dx} + c_{dx,dy}\right)$, where $c_{x,y}$ denotes the coancestry between females *x* and *y*. At equilibrium, the expectation of $c_{sx,sy}$ is 0, because females *x* and *y* will be from different sires *sx* and *sy*, and every sire gets one female offspring in order to equalize family sizes. The expectation of both $c_{sx,dy}$ and $c_{sy,dx}$ is $\overline{F}$ and the expectation of $c_{dx,dy}$ and $c_{x,y}$ is C_d^{*}. It follows that $C_d^{*} = \tfrac{1}{4}\left(0 + 2\overline{F} + C_d^{*}\right)$ and thus $C_d^{*} = (2/3)\overline{F}$. And using $\overline{F}$ as calculated above,

$$\overline{C}_d = \left[N\left(\tfrac{1}{2} + \tfrac{1}{2}\overline{F}\right) + N(N-1) \times (2/3)\overline{F}\right] / N^2 = \left(13 - (1/N)\right)/(18N).$$

From Equation [1], we get the equilibrium coancestry for the scheme where G_0 and G_1 sires are used alternatively:

$$\overline{C}_t = 17/(36N) - 1/(72N^2).$$

$\overline{C}_t$, $\overline{C}_s$, $\overline{C}_{sd}$, $\overline{C}_d$ and $\overline{F}$ were calculated for schemes with 6, 10 or 20 male plus female animals born per generation, *i.e.* N= 3, 5 or 10.

2.2. Simulation study

We also conducted a simulation study, where the average coancestry of the selected parents, and thus the inbreeding of the offspring, was minimized each generation (Sonesson and Meuwissen, 2001), but where a discrete generation structure was considered. Female candidates for selection were females from the current generation, t whereas male candidates were either G_0 sires only or sires from all generations (G_0, G_1, ..., G_t). The latter scheme tests whether younger sires than G_0 and G_1 sires will also be used when average coancestry is minimized.

3. RESULTS AND DISCUSSION

We attempted here to minimize inbreeding in genetic conservation schemes that combine the conservation of live animals with the use of a semen bank. A scheme proposed by Smith (1977) where sires of generation zero was used on all later generations of dams resulted in steady state average coancestry of .192, .116 and .058, for schemes with 6, 10 and 20 animals per generation, respectively (Table 1). The total average coancestry was about 23% higher for the scheme where only sires of generation zero were used compared to the schemes where sires of generations zero and one were used. Similarly, the average inbreeding was 50% higher for the scheme where only sires of generation zero were used compared to the schemes where sires of generations zero and one were used.

Selection of only G_0 sires or alternative use of G_0 and G_1 sires are only two of many possible designs. For instance, we could make equal use of G_0, G_1 and G_2 sires. To study the optimum design we compared the predicted results with a selection algorithm that minimized the average coancestry each generation. It is assumed that this also minimizes the average coancestry in the longer-term. Either G_0 or all sires were selection candidates. The average coancestry and inbreeding levels from the simulated schemes were nearly identical to the two schemes of Table 1, respectively (results not shown). The simulated scheme with all sires as selection candidates resulted in a scheme where G_0 and G_1 sires were selected alternatively, *i.e.* the scheme that we propose here. The simulation results indicate that our proposed scheme is optimal, because the minimization of coancestry every generation is expected to result in the minimum possible coancestry. Males of later generations are not used in the optimum design because they are related to the current females through intermediate generation dams as well as through the founder sires and dams.

Table 1. Inbreeding ($\bar{F}$), coancestry among sires ($\bar{C}_s$), among dams ($\bar{C}_d$), between sires and dams ($\bar{C}_{sd}$) and in total ($\bar{C}_t$) at equilibrium for two schemes where only sires from generation zero (Gen 0) can be selected or where sires from generations zero and one (Gen 0 & 1) are used alternatively.

Scheme	$\bar{F}$	$\bar{C}_s$	$\bar{C}_d$	$\bar{C}_{sd}$	$\bar{C}_{sd}$
	6 newborn animals per generation				
Gen 0	.167	.167	.269	.167	.192
Gen 0 & 1	.111	.167	.235	.111	.156
	10 newborn animals per generation				
Gen 0	.100	.100	.163	.100	.116
Gen 0 & 1	.067	.100	.142	.067	.094
	20 newborn animals per generation				
Gen 0	.050	.050	.083	.050	.058
Gen 0 & 1	.033	.050	.072	.033	.047

The proposed method is developed for schemes where the population is maintained for several generations when the number of sires is less than or equal to the number of dams and when the number of sires is limited. When the purpose of the use of gene bank material is to restore a large live population, the same principle holds, namely that more genes are conserved when semen of sires from both generations zero and one is stored. Note that, if the number of generation zero sires is larger than the number that can be stored, it is always better to first select all the sires of generation zero, because they contain more genes from the founder generation than sires from generation one. Also, if founder female gametes or embryos can be frozen, this will reduce average coancestry more than the scheme proposed here, where only semen could be frozen.

A drawback of the conservation schemes presented is that they will allow no further evolution (genetic adaptation and accumulation of mutations) in the conserved population. Further research is needed to optimize combined *in situ* and *ex situ* conservation schemes that do allow for evolution of the conserved population. For practical schemes, the selection strategies as presented here are probably over-simplified, but the general principle that storage of semen from two generations of sires conserves more genetic diversity than that of only one generation of sires is expected to hold also in more complicated situations.

LITERATURE

Caballero, A. and Toro, M.A. (2000) Interrelations between effective population size and other pedigree tools for the management of conserved populations. Genet. Res. Camb. 75: 331-343.

Lacy, R. (1995) Clarification of genetic terms and their use in the management of captive populations. Zoo Biology 14: 565-578.

Malécot, G. (1948). Les mathématiques de l'hérédité. Masson, Paris, France.

Smith, C. (1977) Use of frozen semen and embryos to measure genetic trends in farm livestock. Z. Tierzucht. Züchtungsbiol. 94: 119-227.

Sonesson, A.K. and Meuwissen, T.H.E. (2001) Minimisation of rate of inbreeding for small populations with overlapping generations. Genet. Res. 77: 285-292.

Chapter 8

Selection against genetic defects in small populations under controlled inbreeding

A.K. Sonesson, L.L.G. Janss and T.H.E. Meuwissen

Institute for Animal Science and Health,
P.O. Box 65, 8200 AB Lelystad The Netherlands

Submitted to Genetic Selection Evolution

ABSTRACT

We have studied different genetic models and evaluation systems to select against a genetic disease with additive, recessive or polygenic inheritance in small populations. When using optimum contribution selection with a restriction on rate of inbreeding (ΔF) to select against a single gene, selection directly on DNA-genotypes (GENO) is, as expected, most efficient. GENO needed 2.0 generations to half the frequency of the disease allele with additive effects in a population with discrete generations, 100 new-born animals and ΔF restricted to .01 per generation. When only phenotypic records were available, selection on BLUP or on genotype probabilities calculated by segregation analysis (SEGR) needed 1.0 or 2.0 generations longer to half the frequency of the disease allele when allele effects were additive or recessive, respectively. For a threshold polygenic model, BLUP and SEGR both needed 2.5 generations to half the fraction of diseased animals. Smaller schemes or schemes with a more stringent restriction on ΔF needed more generations to half the frequency of diseased allele or the fraction of diseased animals. SEGR and BLUP were approximately equally efficient under both single gene and polygenic inheritance models, suggesting that efficient selection against a disease is possible without knowing its mode of inheritance.

1. INTRODUCTION

Many populations show heritable defects. Some defects are inherited by a single gene, *e.g.* complex vertebral malformation (CVM) in cattle (Agerholm *et al.*, 2001). Other diseases have a complex inheritance involving multiple genes plus environmental effects, *e.g.* hip and elbow dysplasia in dogs (Lust, 1997).

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 environment (complete penetrance or not). For unknown genes, segregation analysis can be used to make an inference of the genotype probabilities of individual animals, using phenotypic records of the animal itself and relatives (Elston and Stewart, 1971; van Arendonk *et al.*, 1989; Fernando *et al*, 1993; Janss *et al.*, 1995). Segregation analysis can also be used to save genotyping costs when selecting on DNA-genotypes for known genes (Kinghorn, 1999). For diseases with complex inheritance (involving multiple genes) and environmental effects, BLUP breeding value estimation seems more suitable.

However, the inheritance is unknown for many diseases and the breeding value estimation is not straightforward. We will here investigate genetic models and evaluation methods to select against a disease of known or unknown inheritance.

Especially in small populations, genetic drift increases the occurrence of heritable diseases. Hence, the populations we will be dealing with are often small and care has therefore to be taken to avoid high rates of inbreeding when selecting against the disease in 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 optimum contribution selection, which maximises genetic response with a restriction on the rate of inbreeding (Meuwissen, 1997; Grundy *et al.*, 1998; Meuwissen and Sonesson, 1998; Grundy *et al.*, 2000).

The aim of this study is to investigate alternative genetic models and evaluation methods for different kind of diseases, where the genetic evaluation method does not always agree with the true inheritance of the disease. We will compare a genetic model for a single gene to a threshold model, where many genes and environmental effects affect the liability of an animal to be diseased. We will compare breeding values estimated by BLUP to breeding values estimated by segregation analysis and to DNA-genotyping (for a known disease gene). 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. MATERIAL AND METHODS

2.1. Genetic model

Threshold model. The threshold genetic model assumes liabilities underlying the probability of having a diseased animal. The liability was assumed normally distributed. Genotypes for liability, g_i, of the base animals were sampled from the distribution $N(0,\sigma_a^2)$, where $\sigma_a^2 = .5$ is base generation genetic variance. Environmental effects on liability, e_i, of base animals were sampled from the distribution $N(0,\sigma_e^2)$, where $\sigma_e^2 = .5$ is environmental variance. Total liability was $x_i = g_i + e_i$. Later generations were obtained by simulating offspring genotypes from $g_i = \frac{1}{2}g_s + \frac{1}{2}g_d + m_i$, where s and d refers to sires and dams, respectively, and m_i = Mendelian sampling component, sampled from

$N(0, \frac{1}{2}(1-\overline{F})\sigma_a^2)$, where $\overline{F}$ 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, 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).

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's rules. Animal *i* was diseased ($y_i = 0$) with probability ($P(y_i = 0|XX_i)$), where $P(y_i = 0|XX_i)$ is the penetrance probability of having a diseased animal ($y_i = 0$) given genotype XX_i. The input values $P(y_i = 0|XX_i)$ were .0, .5 and 1.0 for genotypes XX_i = aa, Aa and AA, respectively. If inheritance was recessive, these values were .0, .0 and 1.0 for genotypes aa, Aa and AA, respectively, when inheritance was recessive. The phenotypic disease records, y_i, that resulted from this sampling were used as input for the EBV evaluation.

2.2. EBV estimation

BLUP. Phenotypic values from the threshold and single gene model were input to obtain EBVs using BLUP-breeding value estimation procedure (Henderson, 1984). 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 (Meuwissen *et al.*, 1995).

For a threshold model, heritability of the liability of the trait on the underlying scale, h^2_{under}, was .5. A correction on the heritability is done for the upper diseased scale, h^2_{upper}, to estimate BLUP breeding value estimates. The correction is (Gianola, 1979):

$$h^2_{upper} = f(T)^2\, h^2_{under} / [z(1-z)]\,, \qquad [1]$$

where z is the proportion diseased animals when the threshold value is *T*, and $f()$ = Normal density function.

DNA genotyping. In this case, the disease was assumed to be due to a single known gene, all males were genotyped. When assigning the recessive genotype a value of 0, and the others a value of 1 (in Falconer and Mackay (1996) notation $a=d=.5$), it follows that the frequency of the diseased genotype q^2 equals the disease incidence in the population (Falconer and Mackay, 1996). 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 the dominance deviation, $d = P(y_i = 1 | \mathrm{Aa}) - .5(P(y_i = 1 | \mathrm{aa}) + P(y_i = 1 | \mathrm{AA}))$. These breeding values correspond to (twice the deviation of) disease incidences in progeny of the respective genotypes, and will optimise the selection to reduce disease incidence when used as input for the selection algorithm.

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 EBVs from DNA genotyping can not be calculated for the threshold genetic model.

Segregation analysis. The algorithm by Kerr and Kinghorn (1996) was used to calculate genotype probabilities of each animal. It is an algorithm based on iterative peeling (van Arendonk *et al.*, 1989; Janss *et al.*, 1995) 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, .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, 1.0 and 1.0 for genotype 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 (1996) calculates the probability that individual i has genotype XX, $P(XX)_i$. The $P(XX)_i$ are used to calculate EBVs as

$$EBV_i = P(\mathrm{aa})_i 2q\alpha + P(\mathrm{Aa})_i (q-p)\alpha - P(\mathrm{AA})_i 2p\alpha \; . \qquad [2]$$

These EBVs are input for the selection algorithms.

For the threshold genetic model, we estimated the penetrance probabilities as $P(y_i = 1 | XX_i) = (\sum P(XX)_i y_i) / \sum P(XX)_i$ and $P(y_i = 0 | XX_i) = 1 - P(y_i = 1 | XX_i)$. Similarly, initial allele frequencies were estimated as $q_o = \left[\sum_{\mathrm{base}} P(\mathrm{AA})_i + \sum_{\mathrm{base}} \tfrac{1}{2} P(\mathrm{Aa})_i\right] / Nbase$, 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 (1997). This method maximises the genetic level of next generation of animals, $G_{t+1}=\mathbf{c}_t'\mathbf{EBV}_t$, where $\mathbf{c}_t$ is a vector of genetic contributions of the selection candidates to generation $t+1$ and $\mathbf{EBV}_t$ is a vector of estimated breeding values of the candidates for selection in generation t. The objective function, $\mathbf{c}_t'\mathbf{EBV}_t$, 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 $\overline{C}_{t+1}=\mathbf{c}_t'\mathbf{A}_t\mathbf{c}_t/2$, where $\mathbf{A}_t$ is a $(n \times n)$ relationship matrix among the selection candidates, $\overline{C}_{t+1}=1-(1-\Delta F_d)^{t+1}$, and ΔF_d is the desired rates of inbreeding (Grundy *et al.*, 1998). Note that the level of the restriction $\overline{C}_{t+1}$, can be calculated for every generation before the breeding scheme commences. Contribution of males (females) are constrained to ½, *i.e.* $\mathbf{Q}'\mathbf{c}_t=\frac{1}{2}$ 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 $\frac{1}{2}$ is a (2×1) vector of halves. The optimization procedure was explained in Meuwissen (1997). The output from the selection method is a vector with genetic contributions for each selection candidate, $\mathbf{c}_t$. In case of single genes, at some point all selection candidates can have the desired genotype and a maximization of genetic response is no longer relevant, in which case the algorithm switched to minimizing inbreeding.

2.4. Mating

Random mating was applied. A progeny got a sire assigned by random sampling a sire with sampling probabilities following the optimal contributions of the sires. A mating pair always got two progeny, one female and one male.

2.5. Schemes

The simulated breeding schemes are described in Table 1. The general structure is that of a closed scheme with discrete generation structure. Recording of the disease was on both sexes before selection. All females were selected for one mating producing a male and a female offspring, *i.e.* there was no selection differential in the female selection path. The selection of sires was optimised by the selection algorithm, such that sires could get unequal numbers of progeny. Results were based on 100 replicated schemes for schemes with 60 and 100 selection candidates and on 50 replicated schemes for schemes with 200 selection candidates. ΔF was restricted to .01 per generation, which is considered as the maximum acceptable rate of inbreeding (Bijma, 2000). For the scheme with 100 and 200 new-born animals per generation, also some schemes with the same N_e over N ratio as the scheme with 60 new-borns were simulated. We compared the models on the number of generations they needed to half the frequency of the disease allele or the fraction of diseased animals for the single gene and threshold models, respectively.

Table 1. Parameters of closed scheme.

Size of scheme	
Number of selction candidates	60,100 or 200
Number of generations	15
Inbreeding constraint	.01, .003 or .006 per generation
Parameters of trait	
Phenotypic variation, σ^2_p	1
Heritability single gene	.5
Threshold model	
threshold	.0
penetrance (aa, aA/Aa, AA)	
additive model	.0, .5, 1.0
recessive model	.0, 1.0, 1.0
heritability, h^2_{under}	.5

3. RESULTS

3.1. Single gene model

For the single gene model, the genetic evaluation was on DNA-genotype (GENO), BLUP EBVs (BLUP) or on EBVs based on genotype probabilities calculated by segregation analysis (SEGR). As expected, GENO was most efficient to reduce the frequency of the disease allele. BLUP and SEGR schemes always gave very similar results. For a scheme with 100 new-born animals per generation and additive inheritance, GENO needed 2.0 generations to half the frequency of the disease allele, whereas both BLUP and SEGR needed 3.0 generations (Figure 1). GENO used the same number of generations to half the frequency of the disease allele for a gene with recessive inheritance as for a gene with additive inheritance, as expected. 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-disease homozygotes, when inheritance is recessive.

Both BLUP and SEGR schemes achieved the restriction on ΔF of .01 during all generations (Figure 2). The GENO scheme achieved the restriction until generation 3 (Figure 2), but thereafter, the achieved ΔF was lower than the restriction. This could be explained by that, after 3 generations, most animals have the non-disease genotype, such that effective population size could be higher than the restriction, while still achieving the maximum selection response (selection of only homozygous non-disease genotypes). For the BLUP and SEGR schemes, EBVs will differ somewhat, even if the 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 EBVs of the parents within the restriction on ΔF. Therefore, BLUP and SEGR maintained the restriction on ΔF and selected somewhat fewer sires than GENO (Table 2).

For the small schemes with 60 new-born animals per generation, GENO needed 3.0 and BLUP and SEGR 4.0 generations to half the frequency of the disease allele for the gene with additive inheritance, *i.e.* smaller numbers of animals reduce the genetic response (Table 2). For schemes with 200 new-born animals per generation, GENO needed 1.5 and BLUP and SEGR 3.0 generations to half the frequency of the disease allele. Hence, it takes longer time to reduce allele frequency in smaller schemes, which is expected, because fewer selection candidates have the non-disease genotype.

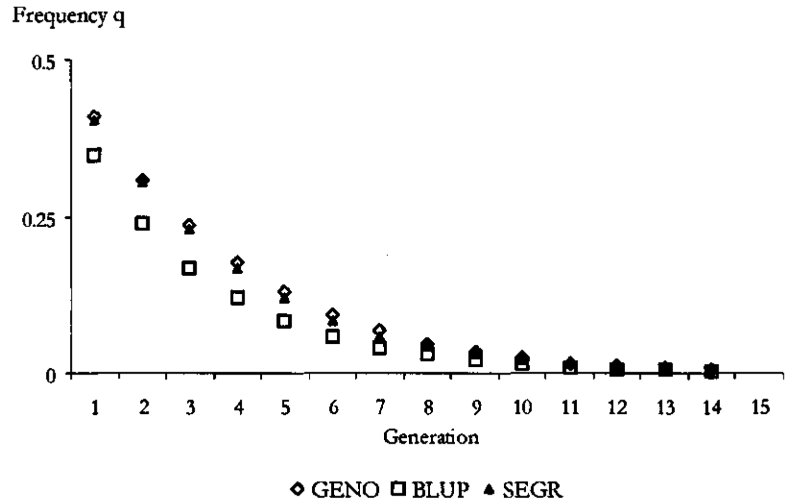

Figure 1. Frequency of disease allele (Frequency q) for schemes with 100 new-born animals per generation, with a single gene model with additive genetic effects and genetic evaluation was done on BLUP EBVs (BLUP) or on EBVs based on genotype probabilities calculated by segregation analysis (SEGR).

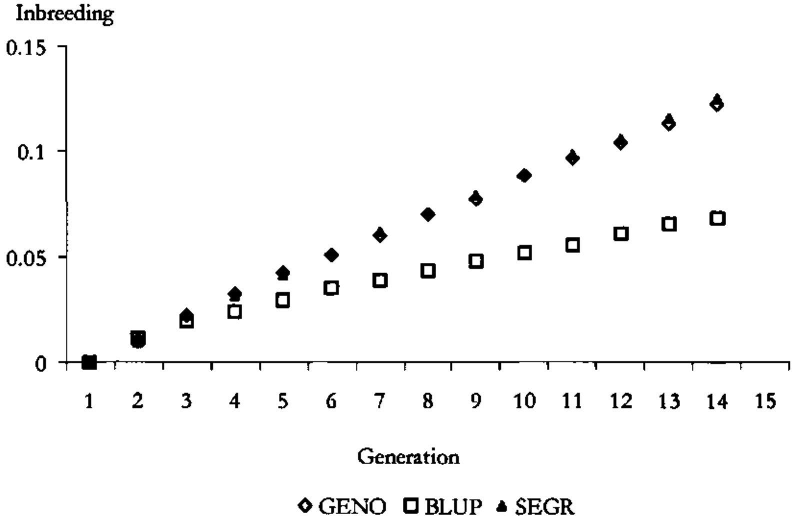

Figure 2. Level of inbreeding for schemes with 100 new-born animals per generation, with a single gene model with additive genetic effects and genetic evaluation was done on BLUP EBVs (BLUP) or on EBVs based on genotype probabilities calculated by segregation analysis (SEGR).

Since it took longer time to reduce the frequency of the disease allele for the smaller scheme with 60 new-born animals per generation, ΔF was kept at the level of the restriction for GENO for more generations (6 for schemes with a gene that has additive inheritance) than for the scheme with 100 new-born animals per generation (not shown). Similarly, for the larger scheme with 200 new-born animals per generation, ΔF was kept at the level of the restriction for GENO for only two generations. Thereafter, ΔF was 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 new-born animals per generation, GENO seemed in general to select more sires than the other two schemes (Table 2), because in later generations, more sires had the desirable genotype.

Larger schemes were also simulated, where the ratio of N_e over N was the same as for 60 new-born animals, *i.e.* .833 and the rate of inbreeding was restricted to .006 and .003 per generation for schemes with 100 and 200 new-born animals per generation, respectively. For schemes with 60, 100 and 200 new-born animals per generation, GENO needed 3.0 generations and BLUP and SEGR 4.0 generations to half the frequency of the disease allele for a gene with additive inheritance (Table 2). For a gene with recessive inheritance, GENO needed 3.0 generations and BLUP and SEGR schemes 5.0 generations to half the frequency of the disease allele for all three sizes of schemes, when compared at the same ratio of N_e over N. Thus, the ratio of N_e over N seems to determine the selection intensity of the scheme and also the genetic response. This corresponds to Hill (2000), in which response to selection in a single gene model is proportional to N_e/N if new mutations are ignored.

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

For these schemes, there was an increase in the number of selected sires with increasing population size. For all BLUP and SEGR schemes, the accuracy of selection was between .67 and .76 (Table 2).

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 .50, was monitored. For schemes with 100 new-born animals per generation, it took about 2.5 generations to half the fraction of diseased animals to .25 for both BLUP and SEGR (Figure 3). 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 .01 was kept at the restricted level for both BLUP and SEGR (Figure 4).

The number of selected sires was also about the same (Table 3) for both BLUP (19.9) and SEGR (19.2). As for the single gene model, input to the algorithm was that all dams were selected.

For schemes with 60 and 200 new-born animals per generation, it took about 4.0 and 2.0 generations, respectively, to half the fraction of diseased animals. Both BLUP and GENO achieved the restriction on $\Delta \mathrm{F}$. For schemes with 60 new-born animals per generation, the number of selected sires was 21.4 for BLUP and 21.8 for SEGR (Table 3). For schemes with 200 new-born animals per generation, the number of selected sires was 15.8 for BLUP and 14.8 for SEGR (Table 3). Thus, BLUP and SEGR selected similar numbers of sires.

Schemes, which had the same ratio of N_e over N as the scheme with 60 new-born animal per generations, were also investigated for the threshold model. For schemes with 60, 100 and 200 new-born animals per generation, it took 4.0 generations for both BLUP and SEGR to half the fraction of diseased animals (Table 3). Hence, also for the threshold model scheme, the ratio of N_e over N seems to determine the selection intensity and thus also genetic response.

There was an increase in the number of selected sires with increasing population size. The number of selected sires was, as expected, twice as many for the scheme with 200 new-born animals per generation than for the scheme with 100 new-born animals per generation (Table 3).

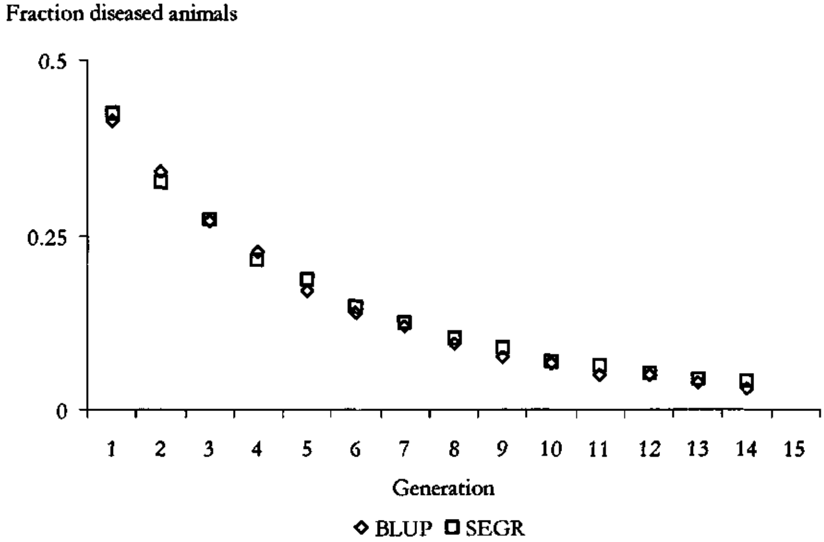

Figure 3. Fraction of diseased animals for schemes with 100 new-born animals per generation, with a threshold genetic model and genetic evaluation was done on BLUP EBVs (BLUP) or on EBVs based on genotype probabilities calculated by segregation analysis (SEGR).

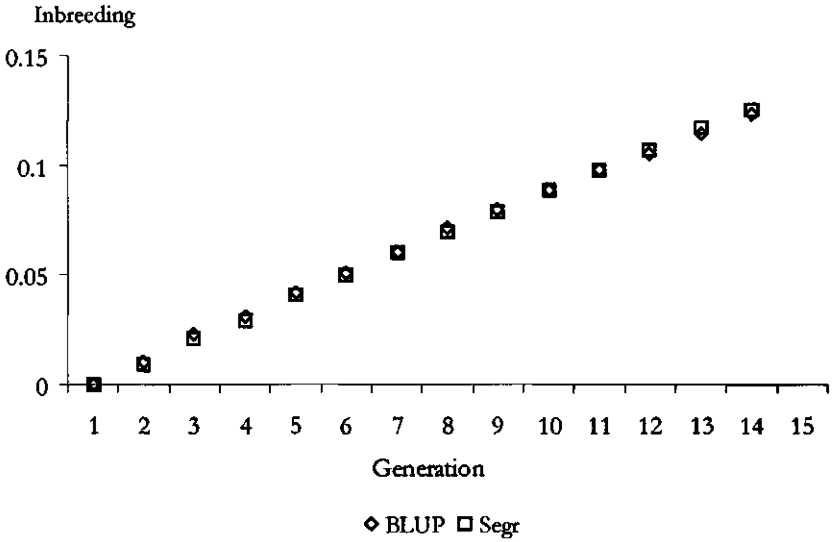

Figure 4. Level of inbreeding for schemes with 100 new-born animals per generation, with a threshold genetic model and genetic evaluation was done on BLUP EBVs (BLUP) or on EBVs based on genotype probabilities calculated by segregation analysis (SEGR)

4. DISCUSSION

4.1. General

For the single gene model, direct selection against the disease allele (GENO) is, 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 half the frequency of the disease allele. This period was shorter for a gene with additive inheritance and longer for a gene with recessive inheritance. Hence, a gene needs to be identified approximately within 1.0 (additive) to 2.0 (recessive) generations for GENO to be as effective as SEGR and BLUP to half 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, which 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 optimum contribution selection for discrete generations (Meuwissen, 1997) to restrict the rate of inbreeding. BLUP and SEGR achieved the restriction for all schemes, because EBVs generally differ somewhat such that selections can easily be made. Another selection method, *e.g.* truncation selection on EBVs, 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-EBVs and SEGR-EBVs are probably similar for truncation and optimum contribution selection.

Table 2. Number of generations to reduce the frequency of the disease allele by 50% (Halftime), the number of selected sires (Nselsires) and the accuracy of selection for schemes with ΔF restricted to .01, .006 or .003 per generation for schemes with 60, 100 or 200 new-born animals per generation and additive or recessive inheritance of the single gene.

EBV evaluation [1]	Halftime (gen)	Nselsires	Accuracy	Halftime (gen)	Nselsires	Accuracy
			Additive inheritance			
	60 newborn animals /generation, ΔF = 0.01					
Genotypic	3.0	22.3	1.000			
BLUP	4.0	21.0	0.753			
Segregation	4.0	21.0	0.757			
	100 newborn animals /generation, ΔF = 0.01			100 newborn animals /generation, ΔF = 0.006		
Genotypic	2.0	23.6	1.000	3.0	35.5	1.000
BLUP	3.0	21.1	0.743	4.0	34.4	0.759
Segregation	3.0	19.9	0.743	4.0	34.4	0.757
	200 newborn animals /generation, ΔF = 0.01			200 newborn animals /generation, ΔF = 0.003		
Genotypic	1.5	38.3	1.000	3.0	73	1.000
BLUP	3.0	21.4	0.746	4.0	69.9	0.757
Segregation	3.0	18.8	0.753	4.0	69.9	0.769
			Recessive inheritance			
	60 newborn animals /generation, ΔF = 0.01					
Genotypic	3.0	21.8	1.000			
BLUP	5.0	21.4	0.674			
Segregation	5.0	21.6	0.699			
	100 newborn animals /generation, ΔF = 0.01			100 newborn animals /generation, ΔF = 0.006		
Genotypic	2.0	23.9	1.000	3.0	36.5	1.000
BLUP	4.0	23.6	0.700	5.0	36.3	0.689
Segregation	4.0	20.9	0.712	5.0	35.6	0.704
	200 newborn animals /generation, ΔF = 0.01			200 newborn animals /generation, ΔF = 0.003		
Genotypic		37.4	1.000		73.7	1.000
BLUP	4.0	24.7	0.695	5.0	70.9	0.677
Segregation	4.0	23.8	0.710	5.0	72.3	0.703

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

Table 3. Number of generations to reduce the fraction of diseased animals by 50% (Halftime), the number of selected sires (Nselsires) and the accuracy of selection for schemes with ΔF restricted to .01, .006 or .003 per generation for schemes with 60, 100 or 200 new-born animals per generation.

Genetic evaluation [1]	Halftime (gen)	Nselsires	Halftime (gen)	Nselsires
	60 newborn animals /generation, $\Delta F = 0.01$			
BLUP	5.0	21.2		
Segregation	5.0	21.4		
	100 newborn animals /generation, $\Delta F = 0.01$		100 newborn animals /generation, $\Delta F = 0.006$	
BLUP	3.5	21.5	5.0	33.8
Segregation	3.5	21.7	5.0	35.5
	200 newborn animals /generation, $\Delta F = 0.01$		200 newborn animals /generation, $\Delta F = 0.003$	
BLUP	3.0	25.6	5.0	71.1
Segregation	3.0	20.5	5.0	71.0

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

The method of Kerr and Kinghorn (1996) 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 model 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 (.743 for the scheme with additive inheritance and 100 new-born 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 genotype and penetrance probabilities of the genotypes, which estimates EBVs of the animals, although there was no single gene and thus no genotype or penetrance probabilities to be estimated. Their estimated effects are thus an artefact of the segregation analysis model. However, the resulting EBVs yield very similar selection response as those of the BLUP model (Figure 1). SEGR only shows a small reduction in genetic response for the threshold model when the frequency of diseased animals was very low (Figure 3). Possibly, this reduction in 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-EBVs in the single gene model and SEGR-EBV yield good approximations to BLUP-EBVS 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-EBVs, which yield very similar selection response when the fixed effect structures are rather simple as is the case here (Meuwissen *et al.*, 1995).

4.3. Relaxation of assumptions

All dams were selected and only selection of sires was optimised. For a scheme, where also 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 .5. In practice, the initial frequencies of the disease allele will often be lower. We also simulated a scheme with q_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). Similarly, under the threshold model, an initial disease frequency will be changed under selection as predicted by the part of the curve of Figure 3, where the frequency is smaller than the initial disease frequency. This validates our approach to look at half-times to be applicable also for other starting frequencies.

We have only treated single trait selection against the disease. In a practical selection program, there would sometimes be another breeding goal next to the disease trait. This would require a weighing of the EBV for disease against the EBV of the remainder of the breeding goal. The weights may be obtained from the marginal effect of both traits (disease and remainder of breeding goal) on total profit of the animals. However, in some situations, public concern may increase the weight for the disease trait.

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 *e.g.* an infectious agent, the probability of getting the disease depends not only on DNA-genotypes, but also on epidemiological parameters, *e.g.* how infectious the disease is, how many other animals are diseased *etc.* Different models have to be used for those traits.

The biallelic nature of the single gene is the simplest case. The obtained results would probably be valid for a tri-allelic gene, but it would take more generations for BLUP and SEGR to get the population homozygous for the best of the three alleles.

We have assumed that there was only selection for the disease with either a single gene or threshold model. In general, results can also be applied to major genes in mixed inheritance models, which is a model in between these two genetic models we have used here, and where a major gene has a large effect on the trait, and many background genes have small effects (Dekkers and van Arendonk, 1998). Villanueva *et al.* (1999) simulated schemes with a mixed model inheritance for BLUP optimum contribution selection (similar to our BLUP scheme). Villanueva *et al.* (1999) found that the frequency of the favourable allele had reached .95 after 8 generations for a gene with recessive inheritance. This is similar to results in Figure 1, despite that the schemes differed considerably (the genotypic values (2.0 and dominance deviation was -.5), initial frequency (.15) of the non-disease allele, and size of the scheme (120 new-born animals per generation)).

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 have assumed random mating among the selected animals. For the SEGR scheme, non-random mating would reduce the frequency of the disease allele and increase accuracy of selection. 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 a expensive and time consuming strategy to reduce the frequency of the disease allele.

LITERATURE

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

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

Bijma, P. (2000) Long-term genetic contributions- Prediction of rates of inbreeding and genetic gain in selected populations. PhD thesis, Wageningen University, The Netherlands.

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

Elston, R.C. and Stewart, J. (1971) A general model for the genetic analysis of pedigree data. Hum. Hered. 21: 523-542.

Falconer, D.S. and Mackay, T.F.C. (1996) Introduction to quantitative genetics. 4th Ed. Longman Sci. and Tech., Harlow, UK.

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

Gianola, D. (1979) Heritability of polychotomous characters, Genetics 93: 1051- 1055.

Grundy, B., Villanueva, B. and Woolliams, J.A. (1998) Dynamic selection procedures for constrained inbreeding and their consequences for pedigree development. Genet. Res. Camb 72: 159-168.

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

Henderson C.R. (1984) Applications of Linear Models in Animal Breeding, University of Guelph Press, Guelph, Canada.

Hill, W.G. (2000) Maintenance of quantitative genetic variation in animal breeding programmes, Livest. Prod. Sci. 63: 99-109.

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

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

Kinghorn. B. P. (1999) Use of segregation analysis to reduce genotyping costs, J. Anim. Breed. Genet. 116: 175-180.

Lust, G. (1997) An overview of the pathogenesis of canine hip dysplasia., J. Am. Vet. Med. Assoc. 210: 1443-1445.

Meuwissen, T.H.E. (1997) Maximising the response of selection with a predefined rate of inbreeding. J. Anim. Sci. 75: 934-940.

Meuwissen, T.H.E. Engel, B. van der Werf, J.H.J. (1995) Maximising selection efficiency of categorical traits, J. Anim Sci. 73: 1933-1939.

Meuwissen T.H.E., Sonesson A.K. (1998) Maximizing the response of selection with a predefined rate of inbreeding- overlapping generations, J. Anim. Sci. 76: 2575-2583.

Villanueva, B., Pong-Wong, R., Grundy, B. and Woolliams, J.A. (1999) Potential benefit from using an identified major gene in BLUP evaluation with truncation and optimal selection. Genet. Sel. Evol. 31: 115-133.

Chapter 9

General Discussion

This thesis deals with the definition of selection and mating criteria for animal breeding populations under directional selection and genetic conservation populations. Emphasis is on populations with small effective sizes that have known pedigrees.

When selecting for BLUP estimated breeding values (Henderson, 1984), which currently is the most used selection criterium, the probability of co-selecting related animals is high, because the estimates are higher correlated than the true breeding values. Co-selection of related individuals results in higher rates of inbreeding (ΔF) of the population. One way of reducing inbreeding is to maximize the genetic response with a cost factor on inbreeding (Wray and Goddard, 1994; Brisbane and Gibson, 1995). This method was further developed by Meuwissen (1997) into optimum contribution selection which controls the rate of inbreeding. However, in the study by Meuwissen (1997), only discrete generation structures were considered. The first part of this thesis dealt with optimum contribution selection for schemes with overlapping generations, which is common in practical selection schemes. When optimizing the genetic contributions of the selection candidates for schemes with overlapping generations, account has to be taken of previous and future use of the candidates, such that there is optimization between and within age-classes.

The second part of this thesis dealt with non-random mating schemes, which have been shown to reduce inbreeding for truncation selection on phenotypes or BLUP estimated breeding values (*e.g.* Caballero *et al.*, 1996). However, because inbreeding is restricted in optimal contribution selection schemes, non-random mating schemes improve the family structure, so that genetic response can further increase.

The aim of many genetic conservation schemes is to minimize ΔF, *i.e.* to maximize effective population size in order to get the populations out of a genetic bottleneck. The ΔF is minimized when contributions of individuals are equal for schemes that are symmetric with respect to the numbers of sires and dams (Wright, 1931). In the third part of this thesis, an optimum contribution algorithm was presented that minimizes ΔF for populations with overlapping generations, where it is more complicated to equalize the contribution of single families than simply selecting the same number of progeny per mating pair.

Increased inbreeding leads to increased frequencies of heritable diseases. The fourth part of this thesis dealt with alternative methods to select against heritable diseases of a small population, while controlling inbreeding.

In conclusion, populations under selection aim at maximizing the genetic response with a control of inbreeding, whereas the aim of many genetic conservation schemes is to minimize rates of inbreeding. The management of genetic contributions is, however, an important tool to manage inbreeding in both types of schemes.

In this General Discussion, I will firstly discuss the theory behind selection and conservation schemes. Thereafter, I will discuss the relevance of this thesis for theory and practice of selection and conservation schemes in a broader perspective than in the papers comprising the thesis.

1. THEORY

1.1. The effect of selection on ΔF

The ΔF increases due to selection, because families with high genetic values get higher contributions to future generations than families with low genetic values, resulting in a higher probability of having common ancestors. For this process, Robertson (1961) introduced the concept of inherited selective advantage, which would half every generation and which sums to approximately two. However, the derivation of Robertson (1961) results in an overpredicted ΔF (Hill, 1985), because it ignores the reduction of between family variance caused by the gametic disequlibrium over generations due to selection (Bulmer, 1971), such that the selective advantage is more than halved every generation (Wray and Thompson, 1990; Santiago and Caballero, 1995).

In most closed selection schemes, the genetic variances find equilibrium values after all alleles have mixed during the initial generations (when ΔF is low). Wray and Thompson (1990) showed that the contributions of ancestors do not change anymore after approximately five to seven generations with random mating and discrete generations. The rate of inbreeding is related to these long-term genetic contributions of ancestors, r, as $\Delta \mathrm{F} = \frac{1}{4}\sum r^2$ (Wray and Thompson, 1990; Woolliams and Bijma, 2000). Note that although the genetic contributions of ancestors converge, the combination of alleles from the base animals still changes due to Mendelian sampling, such that genetic gain is achieved. The long-term genetic contributions are related to genetic gain as $\Delta G = \sum ra$, where a is the Mendelian sampling term of an ancestor (Woolliams and Thompson, 1994). Recent work of Woolliams and Bijma (*e.g.* 2000) show linear approximations to predict r and thus ΔF and ΔG under truncation selection. Woolliams *et al.* (1999) show that this derivation of ΔG is consistent with the classical theory. Inbreeding and genetic gain are thus related to each other. In a selection scheme, we want to achieve that $\Delta \mathrm{F} = \frac{1}{4}\sum r^2$ is not too high and $\Delta G = \sum ra$ is maximized. In a conservation scheme, we often want to minimize $\sum r^2$.

1.2. Managing ΔF with optimum contributions

In the following we will consider selection methods that manage, *i.e.* control, ΔF while genetic gain is maximised. The group selection method of Bondesson (1989) showed that genetic contributions, c, which are optimised such that genetic level is maximal while the diversity of the clones of unrelated trees is at a predefined level, are linear above a certain threshold genetic value. Trees with a genetic value below the threshold will not be used at all. He maximized $\sum cg$ under the restrictions $\sum c = 1$ (*i.e.* the total contribution of all individuals is 1), $\sum c^2$ = constant, where $\sum c^2$ is a measure of diversity (*i.e.* the diversity of the clones is constrained to a constant) and $c \geq 0$ (*i.e.* contributions must be non-negative), where g is the genetic value. The tree clones used by Bondesson (1989) were however assumed to be unrelated. Woolliams and Meuwissen (1993) changed the restriction on diversity to a cost factor on the matrix of prediction error of variances. Wray and Goddard (1994) and Brisbane and Gibson (1995) restricted ½ the average relationship, *i.e.* they accounted for relationship between selection candidates. Their objective function to be maximised was $\mathbf{c}'\mathbf{EBV} - k\mathbf{c}'\mathbf{Ac}$, where **c** is an (1 × n) vector of genetic contributions, **EBV** is an (1 × n) vector with BLUP estimated breeding values and **A** is an (n × n) matrix with additive relationships of the n selection candidates; and k denotes the costs of inbreeding. They did, however, not optimize the cost factor, k, on the average relationship. Meuwissen (1997) replaced k by a Lagrangian multiplier to optimize **c** at a fixed ΔF. The selection index that resulted is quadratic in form and is a group index, *i.e.* the optimum contribution of an individual depends not only on its own EBV and inbreeding coefficient but also on the relationship of that individual to the other selected individuals. In contrast, for a linear index, the value of an individual for a selected group is additive and depends only on the value of the individual itself.

Grundy *et al.* (1998) showed that with the above so-called optimum contribution selection, the regression coefficient of c on r is close to 1.0. With this method of selection, the immediate contribution of an animal, c, is an unbiased prediction of its long-term contribution, which is influenced by future selection decisions. The fact that there are no systematic differences between c and r shows that optimum contribution selection sets the immediate contribution, c, of the animals as good as is possible at the time of selection. It may be noted that because $\Delta G = \sum ra$ and $\Delta F = \frac{1}{4}\sum r^2$, the problem of optimizing r is posed in the same terms as Bondesson's tree cloning problem, since the true Mendelian sampling terms, a, are unrelated (Wray and Thompson, 1990). However, optimum contribution selection results only in an approximate linear relationship between r and a (Grundy *et al.*, 1998), *i.e.* the long-term contributions deviate from their optimal values that maximize $\Delta G = \sum ra$. There are two reasons for this (Grundy *et al.* (1998)):

1. Although the Mendelian sampling effects are (implicitly) estimated by BLUP-EBV, they are not estimated without error.
2. Optimum contribution selection can not set the contribution of individual Mendelian sampling terms to any desired value, because it can only change the contributions of animals which contain the Mendelian sampling terms of themselves and all their ancestors, ie. the Mendelian sampling term contributions cannot be independently set to a desired value. The genetic values of current animals consist of the Mendelian sampling terms of all their ancestors, *i.e.* optimum contribution selection has to act upon packages of Mendelian sampling terms instead of individual terms.

In Chapter 2, an optimum contribution selection method was developed for populations with overlapping generations, which will be discussed in section 2.1 of this General Discussion.

1.3. Minimizing ΔF

For genetic conservation schemes, minimum coancestry selection maximizes effective population size for schemes with discrete generations and equal numbers of males and females (Ballou and Lacy, 1995; Caballero and Toro, 2000). Minimum coancestry selection methods include both founder and non-founder effects (*i.e.* genetic drift), as opposed to other tools that only include founder effects (effective number of founders (Lacy, 1989), founder genome equivalents (Lacy, 1989) or only non-founder effects (effective number of non-founders (Caballero and Toro, 2000)). At equilibrium, minimum coancestry results if a dam is replaced by her daughter and a sire by his son, *i.e.* family sizes are equalized for schemes that are symmetrical with respect to the

number of sires and dams (Wright, 1931; Caballero and Toro, 2000). For schemes with random selection, the variance of long-term genetic contributions is proportional to the variance of family size (Caballero *et al.*, 1996). An extreme form of equal family size is when the contribution of single alleles is equalized with marker assisted selection (Wang and Hill, 2000).
In hierarchical schemes, *e.g.* where each sire is mated to n_d dams, sire families and dam families are of different size. However, the size of the families in the sire to sire path is equalised by replacing each sire by one of his sons, and the sizes of the families in the dam to dam path are equalised by replacing each dam by one of her daughters (Gowe *et al.*, 1959). This will also equalise family sizes in the sire to dam paths (each sire obtains n_d selected daughters), but family sizes in the dam to sire path are not equalised because, within a sire family, only one of the dams gets a selected son. Wang (1997) improved this scheme: a dam, which produced a son, should not also produce a daughter, because she already has a large contribution through her son. Instead, one of the other dams should produce a second daughter. Sanchez Rodriguez *et al.* (2002) have further improved Wang's (1997) method to reduce the variance of long-term genetic contributions for schemes with a higher mating ratio, which have an even larger variance of long-term genetic contributions. The use of positive assortative mating (Falconer and Mackay, 1996) could further reduce ΔF in the scheme of Sanchez Rodriguez *et al.* (2002).

The minimum coancestry selection methods above are developed for populations with discrete generations. In Chapter 6, a minimum coancestry method was developed for genetic conservation schemes with overlapping generations and survival rates < 100%. These schemes will be discussed in section 3.1 of the General Discussion.

1.4. The role of mating in selection schemes

Random mating implies that any selected individual has equal chance of mating with any other selected individual in the population. Strictly, this is not possible in sexually reproducing populations, but matings may still be approximately at random. For populations with pedigree, several non-random mating systems have been proposed as a means of reducing inbreeding in selected populations (Toro *et al.*, 1988; Woolliams, 1989; Santiago and Caballero, 1995; Caballero *et al.*, 1996).

Woolliams and Thompson (1994) extended the formula for ΔF by Wray and Thompson (1990) to account for non-random mating:

$$\Delta F = \tfrac{1}{4}(1-\alpha)\sum r^2 \qquad [1]$$

where α is the correlation between the alleles within an animal. Non-random mating affects both $(1-\alpha)$ in [1], which affects within family drift, and $\sum r^2$, which mainly affects between family drift (Caballero *et al.*, 1996; Bijma, 2000).

In the following, several mating schemes and their effects will be described for truncation selection schemes. The effects of these mating schemes under optimum contribution selection will be discussed in section 2.2.

Minimum coancestry (MC) mating is a system, whereby the least related animals are mated, which leads to increased frequency of heterozygous animals. This leads to increased within family drift and is represented by $(1-\alpha)$ in [1], where $\alpha < 0$ is due to mating of the least related animals. MC also decreases the between family drift, $\sum r^2$, which has a decreasing effect on ΔF (Caballero, 1994). In mass selection schemes, these two opposite effects of MC mating approximately cancel against each other when the variance of long-term contributions was 2/3 in the random mating scheme (Caballero *et al.*, 1996). If the variance of long-term contributions exceeds 2/3, which is mostly the case in selection schemes, MC mating will reduce ΔF and *vice versa*. In mass and BLUP truncation selection schemes, the variance of long-term contributions is sufficiently decreased by MC mating to have a large decreasing effect on ΔF (Toro *et al.*, 1988; Caballero *et al.*, 1996).

Compensatory mating is a system that mates animals with the smallest number of selected sibs to animals with the largest number of selected sibs (Santiago and Caballero, 1995). It mainly reduces the cumulative selection advantage, because successful families are mated to unsuccessful families (Caballero *et al.*, 1996), which couples their genetic contributions together. The latter reduces the variance of the long-term genetic contributions and thus $\sum r^2$. Other varieties on compensatory mating based on equalizing the contributions of ancestors to descendants have been presented in Caballero *et al.* (1996).

Factorial mating (Woolliams, 1989), which is a non-random mating system that reduces ΔF by replacing strong full-sib relationships by those of (maternal) half-sibs, such that sires (and dams) would mate with more than one dam (sire). If dam x has only offspring with sire y, the contributions of x and y are coupled through their offspring and thus are correlated. Such correlations among long-term contributions will increase their variance and thus $\sum r^2$. Factorial mating avoids that dam x has only offspring with sire y and thus decreases these correlations and $\sum r^2$ and ΔF.

1.5. The role of mating when minimizing ΔF

In genetic conservation schemes, the variance of long-term contributions may be close to zero, which changes the effects of non-random mating systems on ΔF compared to genetic improvement schemes, which have variance of the long-term contribution > 2/3. The MC mating has mainly an increasing effect on ΔF, because in Equation [1], $(1-\alpha)>1$ due to the mating of least related parents. The theoretical best solution for conservation schemes would thus be to mate highly related animals (Robertson, 1964). This solution is however not desired, because the population would be split into highly inbred sublines, which increases the probability of loss of sublines and thus of the entire population. A second effect of MC mating is to reduce F, because mating of the least related animals results in less inbred offspring. Thus, even if ΔF is higher for MC than random mating scheme, F would be lower for at least some critical initial generations of the selection scheme.

In Chapter 6, we compared random and minimum coancestry mating schemes in genetic conservation schemes. We did not include compensatory and factorial mating strategies, because they are both set up to reduce $\sum r^2$, which often already is small in genetic conservation schemes. Simulation results of schemes that combine MC mating and minimum coancestry selection are described in Chapter 6 and section 3.2.

2. OPTIMUM CONTRIBUTION SELECTION

In Chapters 2 and 3, extensions of the optimum contribution method of Meuwissen (1997) to populations with overlapping generations are presented.

2.1. Extension of optimum contribution selection to overlapping generations

Populations with overlapping generations have selection candidates from many reproductive age classes and some candidates have already been used in previous selection rounds or will be used in later selection rounds. Therefore, the expected long-term contributions of the age-classes need to be predicted. The current selection decisions affect the relationships of the new-born age-class, which is one of the age-classes whose long-term contribution is predicted. In Chapter 2, the long-term contributions of the age-classes are predicted by Hill's (1974) gene-flow method. Bijma and Woolliams (1999) improved this gene-flow to account for the selective advantage of good families, but concluded that the method of Hill is a good approximation when there is truncation selection across age-classes. Since the long-term contributions do not deviate systematically from the immediate contributions with optimum contribution selection, it may be expected that there are no systematic differences in probabilities of being selected among the offspring within an age-class. Therefore, Hill's and Bijma and Woolliams' gene-flow methods are expected to predict similar long-term contributions for optimum contribution selection.

In Chapter 3, the optimum contribution method of Chapter 2 was compared with the optimum contribution selection method of Grundy *et al.* (2000). Using Hill's (1974) gene-flow method, Grundy *et al.* (2000) optimized the long-term genetic contributions of each sex-age class with a simulated annealing algorithm. In Chapter 2, these contributions were optimized simultaneously with the individual genetic contributions across all age-classes using an iterative method. The derivations in Chapter 2 indicate that both optimization methods are expected to be optimal given that the predicted gene-flow is correct. The annealing method has more flexibility when optimizing the long-term contributions of age-classes. However, the methods yielded similar breeding schemes, although the generation interval was somewhat longer for the method of Grundy *et al.* (2000). The method of Grundy *et al.* (2000) adhered more closely to the inbreeding restriction. The method of Chapter 2 was however computationally considerably faster, partly because of a simpler optimization problem, and would therefore be preferred in situations where speed is important. The method of Grundy *et al.* (2000) would be applied when a restriction on ΔF per generation is needed instead of per year, because the iterative optimization method did not converge to equilibrium when a restriction on ΔF per generation instead of per year was used. A restriction on ΔF per generation is more appropriate for the long-term survival of the population. In the longer term, inbreeding depression depends on how many generations it took to reach the level of inbreeding, *i.e.* if F is attained with a longer generation interval inbreeding depression is higher than when it took more generations to reach the same F (see review of Wang, 2000). A restriction per year is more appropriate when the time horizon is shorter and the costs and risks due to inbreeding are the main consideration for the inbreeding restriction.

2.2. Mating schemes for optimum contribution selection

In the case of optimum contribution selection, non-random mating systems did not reduce ΔF, because ΔF was restricted, but increased ΔG instead (Chapter 4). When ΔF is restricted as in the optimum contribution selection schemes (Meuwissen, 1997), we saw that MC1 mating (a combination of MC and factorial mating, *i.e.* the number of full-sib offspring per mating pair is restricted to zero or one), increased ΔG by up to 22% compared with random mating for populations with discrete generations. The effect of non-random mating was larger in small schemes or when the restriction on ΔF was more stringent. In populations with overlapping generations, the increasing effect of MC mating on ΔG was larger than for the populations with discrete generations, after correction for increased generation interval and population size (Chapter 5).

Compensatory mating did hardly increase genetic gain relative to schemes with random mating (Chapter 4). It aims at breaking the selective advantage of successful families (Caballero *et al.*, 1996), which is implicitly optimized by the optimum contribution selection method. Hence, the positive effect of compensatory mating on BLUP selection schemes was already taken care of by the optimum contribution selection.

A mating system, which minimizes the variability of the relationships among the offspring that resulted from the mating (MVRO) was also considered. Although computationally more demanding, it yielded very similar genetic gain as MC1 mating. An advantage of MVRO is that it is based on one concept, namely equalizing the relationships among the offspring as much as possible, whereas MC1 is based on two concepts, the mating of the least related parents, and the avoidance of full sib offspring. However, both these concepts result in more equal relationships among the offspring, where the avoidance of full sib offspring implies avoiding the most extreme relationships as much as possible. Thus, the success of MC1 mating may be based on the concept of MVRO mating, *i.e.* equalizing relationships among the offspring, but achieving this in an approximate and computationally less intensive way.

2.3. Design of breeding programs

The design of breeding programs changes when using the optimum contribution selection method presented in Chapter 2, which optimizes how many sires and dams are selected from each age-class, using specific information on the selection program, *e.g.* restriction on ΔF, culling rate, number of male and female selection candidates. With other methods, those parameters have to be optimized *a priori* for each selection scheme (*e.g.* Villanueva and Woolliams, 1997). A strict restriction on ΔF led to an increased generation interval in dairy cattle schemes (Chapter 2). Despite the long generation interval, genetic gain was still high for the schemes in Chapter 2, because few progeny tested sires are selected and because the high accuracy of selection in schemes with progeny tests. However, the advantage of selecting progeny tested sires is reduced for schemes with a less stringent restriction on ΔF, larger schemes and when the trait under selection can be measured in both sexes. In the latter case, accuracy of EBVs of young animals are already quite high and awaiting progeny test results will lengthen the generation interval too much.

For schemes without progeny tested sires, dams were more intensely selected than sires, because of their higher accuracy of selection and lower sib correlation (Chapter 2; de Boer *et al.*, 1994). Selection of dams was based on estimates of their own Mendelian sampling term, because they have a record of their own. Selection of sires was based on the Mendelian sampling terms of their parents. There are however only few parents, *i.e.* selection intensity of the sires will be greatly reduced. Hence, due to the dynamic nature of the optimum contribution selection method, it could yield up to 44% higher ΔG per generation than BLUP selection at the same rate of inbreeding (Chapters 2 and 3).

Optimum contribution selection is especially well suited for relatively small populations with a stringent restriction on ΔF, *e.g.* selection lines of pigs and poultry or nuclei of dairy cows. It resulted in relatively more genetic gain compared with BLUP truncation selection for small populations with a stringent restriction on ΔF, whereas larger schemes and schemes with a less stringent restriction on ΔF resemble BLUP truncation selection schemes. Thus, the same gain and same risk can be achieved in smaller (cheaper) breeding schemes. For modern animal breeding schemes, restrictions on contributions due to reproductive limitations will probably continue to decrease in the future due to further development of reproductive techniques such as Ovum Pick-up (Kruip *et al.*, 1994). Additional constraints can be included in the objective function, *e.g.* a constant number of animals could be selected with a constraint on the maximum contribution per selected animal. Meuwissen (1997) and Chapter 3 show how linear additional constraints can be included in the objective function. Similarly, if a selection candidate had been selected before, its original contribution should be accounted for.

All schemes that we have simulated assume that schemes are closed. In open schemes, migration of animals from other population occurs. If the populations are unrelated, F will be substantially reduced, *e.g.* F will be reduced by approximately 20% with 10% migration. With related populations, the reducing effect of migration on F will however be smaller. Note that genetic gain will be reduced if migration comes from an genetically inferior population.

In open nucleus breeding schemes, selection is practiced in two tiers: the nucleus and the base. The genetic level of the base generally lags x generations behind the nucleus, x depending on selection intensity in the base. Despite this genetic lag, base animals can still contribute to the genetic and inbreeding level of the nucleus, because of the high selection intensity that can be practiced in the base. For optimum contribution selection in both nucleus and base, the selection is problem is identical to that of Chapter 2, where the population was structured in say 2 age classes. When selecting in the base, optimum contribution selection has to account for the selection decisions that have been made in the nucleus and the relative long-term contributions of both tiers, and *vice versa*. Thus, the method of Chapter 2 can be applied to selection in open nucleus schemes in a straightforward manner.

However, in open nucleus schemes the control of matings in the base is often limited. This problem also occurs in other diffuse breeding schemes, *i.e.* in schemes where there is not a clear nucleus population that generates the genetic gain. For example in a commercial dairy cattle breeding schemes, the selection of females may be under the control of individual farmers, and the selection differential in the female selection path is often negligible due to reproductive limitations and due to other practicalities. Hence, optimum contribution selection is in these schemes limited to the selection of the sires, but should of course account for the relationship among the females and between the sires and the females. One way to achieve the optimum contributions of the sires may be to only take the sire off the list of sires available for insemination after the sire has obtained the desired number of matings. This may however be a difficult strategy in that sometimes a popular sire has to be taken off the list. If the realized contributions deviate from the optimum values, optimum contribution selection will adjust the contributions in the next selection, such that the inbreeding restriction is (expected to be) achieved. The latter will come at the expense of genetic gain.

3. GENETIC CONSERVATION SCHEMES

3.1. Minimization of ΔF for populations with overlapping generations

The methods that minimize average coancestry by equalizing family size, and thus the long-term contributions that were presented in section 1.3. were derived for populations with discrete generations. Most practical populations have, however, overlapping generations. In Chapter 6, a method to minimize coancestry for overlapping generations, based on the group index theory with optimum contributions, resulted in 3.3-8.4% higher ΔF than the theoretical minimum level. The theoretical minimum is achieved when the variation of family size is zero (Woolliams and Thompson, 1994), which results in $\Delta F = \frac{1}{4}N$. Thus, the minimization of coancestry method for overlapping generations was not optimal, because the variation of contribution of families and the correlation of alleles within an individual could not be reduced to zero. However, in a population with overlapping generations, it is impossible to have animals with exactly the same generation interval if sublining of the population should be avoided, which makes it impossible to reduce the variance of the long-term contributions to zero every generation. With a survival rate < 100%, variance of the long-term contributions increases, because it is impossible to predict which animals will be culled before the next selection round. Hence, the method presented in Chapter 6, is an close to optimal method that minimizes average coancestry for practical breeding schemes with overlapping generations and survival rate < 100%.

3.2. Mating schemes to minimize ΔF

In Chapter 6, non-random mating was compared with random mating for schemes that minimized ΔF in small, unselected populations.

MC1 mating reduced the correlation of alleles within individuals with 11% for the scheme with 6 new-born animals per year, which were simulated in Chapter 6, thus increasing ΔF. Since MC1 reduces the average inbreeding of the progeny, MC1 reduced F, but because ΔF was higher, the inbreeding curves in Figure 1 of MC and random mating crossed-over at around year 300. Hence, MC mating would be preferred over random mating for the schemes simulated in Chapter 6, and probably in most other schemes, because the $(1-\alpha)$ effect in [1] is generally small and the delay of the inbreeding due to MC mating is substantial (usually about one generation of inbreeding (Chapters 4-6)).

In practical genetic conservation schemes, family sizes can often not be equalized. Thus, the variance of long-term contributions will often be > 2/3, implying that MC mating will have a decreasing effect on ΔF (Caballero *et al.*, 1996, section 1.5 in General Discussion).

3.3. Combined in situ and ex situ schemes

The group selection indices presented in Chapter 6 can be used to reduce ΔF for *in situ* genetic conservation schemes of small and endangered populations. *In situ* conservation of breeds, whereby the population is kept in its own original environment is preferred over *ex situ* conservation, whereby breeds are kept outside their own original environment for example at a zoo or in a gene bank (Convention on Biodiversity, 1992). If animals are kept in their own environment, they will continue to evolve in this environment, whereas in *e.g.* a zoo, they will adapt to the artificial zoo environment. In Chapter 7, we have seen the advantage of using frozen semen of older less related sires in order to reduce ΔF to zero, *i.e.* to combine *in situ* and *ex situ* schemes for populations of the size 6-20. The level of average coancestry depended on the number of sires selected, because there were fewer sires than dams selected, *i.e.* repeated use of the sires sets the level of average coancestry. We have also seen that the frozen semen can be used to select with a restriction on inbreeding in these very small populations with substantial genetic gain compared with larger populations. Genetic gain was only five times lower than in a ten times larger scheme, although the generation interval was six times longer and there was low selection intensity in the dam path (dams were replaced by their daughters in a scheme with discrete generations (Sonesson and Meuwissen, 2002)). Hence, ΔF can be reduced to zero by using frozen semen from a gene-bank or can be reduced substantially while still achieving genetic gain.

3.4. Required population size

Selection and genetic conservation schemes often have small N_e. For animal breeding populations, the effective population size needed has been estimated to be around 25 through 50 per generation for populations under selection to operate without a too high risk (Goddard, 1992) and without reduced fitness for endangered population (*e.g.* Meuwissen and Woolliams, 1994). Not only the current effective size is important, but also the change of effective population size over time. Populations, which have survived for a long time at a small effective population size are expected to survive at this low effective population size, whereas larger populations contain more detrimental alleles, which may drift to high frequencies, when the effective population size is reduced. If not all pedigree is known, the effective population size can be calculated based on uncertain relationships (Pérez-Enciso, 1995). If the pedigree is known, information on genetic markers may be used to estimate the effective population size (Hill, 1981).

However, not only the effective population sizes, also the actual population sizes determine the probability of survival of a population. A larger actual population size prevents the population from extinction in a physical sense rather than genetically, *i.e.* for two populations with the same effective population size, the population with smaller actual size has a lower probability of surviving in case of for example a disease outbreak. Due to these external factors, it is better to physically split up the populations in smaller groups, but keep them genetically as one group.

3.5. Selection against a genetic defect

In populations with small effective sizes, there is a higher probability that unfavourable alleles increase in frequency due to chance (due to drift). In Chapter 8, we tested different methods to select against these unfavorable alleles using the optimum contribution selection method of Meuwissen (1997) for populations with discrete generations. When the single gene was known, it was most efficient to select for the favorable genotype. For schemes with 100 new-born animals per generation and ΔF restricted to .01 per generation, it took 2.0 generations to half the frequency of the unfavorable allele with additive and recessive inheritance. For a gene with smaller penetration probabilities and thus also smaller genetic values, the generations needed to half the frequency of the unfavourable allele would be longer, but the disease would then also be less detrimental for the population. Optimum contribution selection using BLUP EBVs (Henderson, 1984) was as efficient as using EBVs from a segregation analysis (Elston and Stewart, 1971) in reducing the frequency of unfavorable alleles for genetic defects determined by single or multiple genes. Hence, optimum contribution selection on BLUP EBVs against genetic defects is an efficient method in reducing the frequency of unfavorable alleles for genetic defects determined by single or multiple genes.

4. FURTHER DEVELOPMENTS

4.1. Mate selection

Generally, mate selection aims more at obtaining high producing offspring, *e.g.* with high dominance effect or low inbreeding depression, than the methods presented here, which aim at maximizing additive genetic gain. Mate selection achieves the objective by optimizing selection and mating in one step. In theory, mate selection is as good as or better than methods that first optimize selection and next optimize the mating in two steps (Kinghorn, 1986). However, mate selection has until now been optimized with a computationally demanding linear programming method, proposed by Jansen and Wilton (1985), such that it is impossible to use for practical breeding schemes without reduced optimality (*e.g.* Kinghorn, 1998). When mate selection was used to maximize genetic gain, Kinghorn *et al.* (1999) suggested the objective:

$$\mathbf{c}'\mathbf{EBV} - b_1\mathbf{c}'\mathbf{Ac} - b_2\overline{F},$$

where $\overline{F}$ = the average inbreeding of the offspring, and b_1 and b_2 are user defined cost factors on average relationships (reflecting long-term inbreeding) and inbreeding of progeny (reflecting short-term inbreeding), respectively. In many situations, b_1 and b_2 are chosen such that the term $\mathbf{c}'\mathbf{EBV} - b_1\mathbf{c}'\mathbf{Ac}$ is much larger than $b_2\overline{F}$, which implies that mate selection first makes sure that $\mathbf{c}'\mathbf{EBV} - b_1\mathbf{c}'\mathbf{Ac}$ is maximized, and second minimizes the inbreeding coefficients of the offspring (Kinghorn *et al.*, 1999). In this case, mate selection is equivalent to the two step procedure of optimum contribution selection and minimum coancestry mating in Chapter 4. Kinghorn *et al.* (1999) presented more subtle choices of b_1 and b_2, which in some cases resulted in more gain at the same level of inbreeding or in less inbreeding at the same genetic gain. Although the non-random methods used in Chapter 4 increased genetic gain substantially, they are probably not optimal, because their derivation is not based on an attempt to maximize genetic gain. Instead, mate selection with optimum b_1 and b_2 would result in matings that maximize genetic gain. The actual optimization of b_1 and b_2 needs further research.

4.2. Dominance effects

Optimum contribution selection as presented in Chapter 2 is a method that maximizes additive genetic response. However, dominance deviations, which are interactions between alleles within an individual, will not contribute to additive response, because these interactions are not inherited (*e.g.* Hoeschele and VanRaden, 1991; Falconer and Mackay, 1996). The changes in allele frequency due to dominance, are not accounted for by the infinitesimal model. Still, using the infinitesimal model, dominance deviations have been used to increase genetic level of cross-breds through mating of single animals. Hayes and Miller (2000) included within and between breed dominance and additive variances and covariances in the objective function of mate selection, and maximized genetic gain with linear programming. Total genetic gain in the progeny was up to 12.5% higher than truncation selection with random mating. However, they only considered one generation of selection. Toro (1998) used one breeding population (generation 1) and one commercial population that is not further selected (generation 2) and maximized total genetic gain in the commercial population, *i.e.* two generations ahead. However, the full genetic model that considers the reduction of base dominance variance, the increase of dominance variance of completely inbred animals and the covariance between additive and dominance effects (Smith and Mäki-Tanila, 1990) or an approximation of the full genetic model (de Boer and Hoeschele, 1993) was not included in the study by Toro (1998). This scheme was also tested for only few generations, such that the longer-term effects have to be tested in further studies. For optimum contribution selection, the within breed dominance effects could be optimized by the simulated annealing algorithm of Chapter 4. This would increase the total genetic level of the offspring generation.

4.3. Hierarchical mating schemes

All the schemes simulated in this thesis had (on average) the same number of sires as dams selected. In practice, due to lower female reproductive rates, hierarchical schemes, whereby one sire is mated to several dams, are more common. The reason for hierarchical schemes is that females are often mated only once or a very limited number of times. For optimum contribution selection this implies that the maximum contribution of females is limited. Meuwissen (1997) showed how optimum contribution selection could be adapted such that contributions do not exceed a maximum value. For MC1 mating, the restriction on the number of matings per selected pair to zero or one, the factorial mating part of MC1, would often be violated in practical dairy cattle or pig schemes. Further research on MVRO mating schemes that include the restrictions due to hierarchical mating schemes seems promising, because such a mating scheme would still avoid extreme relationships, as MC1 mating does, as much as possible.

4.4. Using DNA-markers to restrict ΔF

In this thesis, probabilities of inbreeding are calculated based on the pedigree information of the animals. The reason that only probabilities of inbreeding can be calculated is that the Mendelian sampling of alleles is a stochastic process of which the result is unknown. DNA-markers can be used to improve the predicted probabilities of inbreeding because they can be used to monitor the Mendelian sampling of alleles (Toro *et al.*, 1999). These DNA-marker based F values will however differ from one gene to the next, and when averaged over many genes they are expected to closely resemble the pedigree based F values. The correlation between pedigree and molecular coancestry was .89 for a genome size of 5 Morgan, and higher for larger genome sizes in the study of Fernández *et al.* (2000). Thus, if the genome is sufficiently large, the pedigree and marker based relationship matrices will give very similar results in optimum contribution selection schemes.

In this thesis it was assumed that inbreeding was measured at neutral loci, *i.e.* loci that are not under selection and are also not linked to loci under selection. Note that selection can increase inbreeding even for neutral loci, because it tends to select animals from the same superior families. Changes of allele frequencies and thus inbreeding at loci under selection, *i.e.* a Quantitative Trait Loci (QTL) are desired. Unless the selective value of the locus is very small, the changes of allele frequencies will be in the right direction. However, the allele frequencies of closely linked loci will also change due to the hitch-hiking effect (Maynard Smith and Haigh, 1974). This hitch-hiking effect will lead to substantially increased inbreeding at the loci linked to the locus under selection. The latter effect will probably be more severe in marker assisted selection schemes, where genetic markers are used to enhance changes of allele frequencies at the QTL. To quantify this effect, let us consider a rare QTL allele, which is fixed within 5 generation in a MAS scheme. The inbreeding at a locus 40 cM away from the QTL is $F = (1-.275)^{5/2} = .45$, where .275 is the recombination rate using the Haldane mapping function (Haldane, 1919) and 5/2 is assumed to be the average number of generations since the most recent common ancestor of the QTL allele. This demonstrates that optimum contribution selection is also very pertinent in MAS schemes, where the loss of genetic variation at positions next to the QTL, which is monitored by marker based relationship matrices, should be restricted.

LITERATURE

Ballou, J.D. and Lacy, R.C. (1995). Identifying genetically important animals for management of genetic variation in pedigreed population. In. Population management for survival and recovery- analytical methods and strategies in small population conservation. Ed. J.D. Ballou, M. Gilpin and T.J. Foose. Columbia University Press, New York, USA.

Bijma, P. (2000) Long-term genetic contributions- Prediction of rates of inbreeding and genetic gain in selected populations. PhD thesis, Wageningen University, The Netherlands.

Bijma, P. and Woolliams, J.A. (1999) Prediction of genetic contributions and generation intervals in populations with overlapping generations under selection. Genetics 151: 1197-1210.

de Boer, I.J.M. and Hoeschele, I. (1993) Genetic evaluation methods for population with dominance and inbreeding. Theor. Appl. Genet. 86: 245-258.

de Boer I.J.M., Meuwissen T.H.E. and van Arendonk J.A.M. (1994) Combining genetic and clonal response in a closed dairy cattle nucleus scheme. Anim. Prod. 59: 345-358.

Bondesson, L. (1989). Mathematical principles of optimal utilization of clones under diversity constraints. Swedish University of Agricultural Sciences, Section of Forest Biometry, Report 27, Umeå, Sweden.

Brisbane, J.R.. and Gibson, J.P. (1995). Balancing selection response and inbreeding by including predicted stabilized genetic contributions in selection decisions. Genet. Sel. Evol. 27: 541-549.

Bulmer, M.G. (1971) The effect of selection on genetic variability. Am. Nat. 105: 201-211.

Caballero, A. (1994) Developments in the prediction of effective population size. Heredity 73: 657-679.

Caballero, A, Santiago, E and Toro, M.A. (1996) Systems of mating to reduce inbreeding in selected populations. Anim. Sci. 62: 431-442.

Caballero, A. and Toro, M.A. (2000) Interrelations between effective population size and other pedigree tools for the management of conserved populations. Genet. Res. Camb. 75: 331-343.

Convention on Biological Diversity (1992), www.biodiv.org, United Nations, Rio.

Elston, R.C. and Stewart, J. (1971) A general model for the genetic analysis of pedigree data. Hum. Hered. 21: 523-542.

Fernández, B.; Santiago, E; Toro, M.A. and Caballero, A. (2000) Effect of linkage on the control of inbreeding in selection programmes. Genet. Sel. Evol 32: 249-264.

Grundy, B., Villanueva, B. and Woolliams, J.A. (1998) Dynamic selection procedures for constrained inbreeding and their consequences for pedigree development. Genet. Res Camb 72: 159-168.

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

Goddard, M.E. (1992) Optimal effective population size for the global population of black and white dairy cattle. J. Dairy Sci. 75: 2902-2911.

Gowe, R.S., Robertson, A. and Latter, B.D.H. (1959) Environment and poultry breeding problems. 5. The design of poultry control strains. Poultry Sci. 38: 462-471.

Haldane, J.B.S. (1919) The combination of linkage values, and the calculation of distance between loci of linked factors. J. Genetics 8: 299-309.

Hayes, B.J. and Miller, S.P. (2000) Mate selection strategies to exploit across– and within-breed dominance variation. J. Anim. Breed. Genet. 117: 347-359.

Henderson C.R. (1984) Applications of Linear Models in Animal Breeding, University of Guelph Press, Guelph, Canada.

Hill, W.G. (1974) Prediction and evaluation of response to selection with overlapping generations. Anim. Prod. 18: 117-139.

Hill, W.G. (1981) Estimation of effective population size from data on linkage disequilibrium. Genet. Res. 38: 209-216.

Hill, W.G. (1985) Fixation probabilities of mutant genes with artificial selection. Genet. Sel. Evol. 17: 351-358.

Hoeschele, I. and VanRaden, P.M. (1991) Rapid inversion of dominance relationship matrices for non-inbred populations by including sire by dam subclass effects. J. Dairy Sci. 74: 557-569.

Jansen, G.B. and Wilton, J.W. (1985) Selecting mating pairs with linear programming techniques. J. Dairy Sci. 68: 1302-1305.

Kinghorn, B. P., Shepherd, R. K. and Woolliams, J. A. (1999) An index of estimated breeding value, parental coancestry and progeny inbreeding to help maximise genetic gains. In. Proc. Association for the Advancement of Animal Breeding and Genetics 13: 412-415.

Kinghorn, B.P. (1986) On computer strategies for mate allocation. J. Anim. Breed. Genet. 104: 12-22.

Kinghorn, B.P. (1998) Mate selection by groups. J. Dairy Sci. 81: 55-63.

Lacy, R.C. (1989) Analysis of founder representation in pedigrees: founder equivalents and founder genome equivalents. Zoo Biology 8: 111-124.

Kruip, Th.A.M., Boni, R. , Wurth, Y.A., Roelofsen, M.W.M. and Pieterse, M.C. (1994) Potential use of Ovum Pick-Up for embryo-production and breeding in cattle. Theriogenology 42: 675-684.

Maynard Smith, J. and Haigh, J. (1974) The hitch-hiking effect of a favourable gene. Genet. Res. 23: 23-35.

Meuwissen, T.H.E. and Woolliams, J.A. (1994) Effective sizes of livestock populations to prevent a decline in fitness. Theor. Appl. Genet. 89: 1019-1026.

Meuwissen, T.H.E. (1997) Maximising the response of selection with a predefined rate of inbreeding. J. Anim. Sci. 75: 934-940.

Nicholas, F.W. and Smith, C. (1983) Increased rates of genetic change in dairy cattle by embryo transfer and splitting. Anim. Prod. 36: 341-353.

Pérez-Enciso, M. (1995) Use of the uncertain relationship matrix to compute effective population size. J. Anim. Breed. Genet. 112: 327-332.

Robertson, A. (1961) Inbreeding in artificial selection schemes. Genet. Res Camb. 2: 189-194

Robertson, A. (1964) The effect of non-random mating within inbred lines on the rate of inbreeding. Genet. Res. 5: 164-167.

Sanchez Rodriguez (2002). Submitted.

Santiago, E. and Caballero, A. (1995) Effective size of populations under selection. Genetics 139: 1013-1030.

Smith, S.P. and Mäki-Tanila, A. (1990) Genotypic covariance matrices and their inverses for models allowing for dominance and inbreeding. Genet. Sel. Evol. 22: 65-91.

Sonesson, A.K. and Meuwissen, T.H.E. (2002) Selection in small populations where ex situ conservation is used to reduce inbreeding. Submitted to 7th World Congress Applied to Livestock Production, Montpellier, France.

Toro, M.A., Nieto, B. and Salgado, C. (1988) A note on minimization of inbreeding in small-scale selection programmes. Livest Prod. Sci. 20: 317-323.

Toro, M.A. (1998) Selection of grandparental combinations as a procedure designated to make use of dominance genetic effects. Genet. Sel Evol. 30: 339-349.

Toro, M.A. Silió, L., Rodrigañez, J., Rodriguez, C. and Fernández, J. (1999) Optimal use of genetic markers in conservation programmes. Genet Sel. Evol. 31: 255-261.

Villanueva, B. and Woolliams, J.A. (1997) Optimization of breeding programmes under index selection and constrained inbreeding. Genet. Res. Camb. 69: 145-158.

Wang, J.L. (1997) More efficient breeding systems for controlling inbreeding and effective size in animal populations. Heredity 79: 591-599.

Wang, J. (2000) Effects of population structures and selection strategies on the purging of inbreeding depression due to deleterious mutations. Genet. Res. Camb. 76: 75-86.

Wang, J. and Hill, W.G. (2000) Marker-assisted selection to increase effective population size by reducing mendelian segregation variance. Genetics 154: 475-489.

Woolliams, J.A. (1989) Modifications to MOET nucleus breeding schemes to improve rates of genetic progress and decrease rates of inbreeding in dairy cattle. Anim. Prod. 49: 1-14.

Woolliams, J.A., Bijma, P. and Villanueva, B. (1999) Expected genetic contributions and their impact on gene flow and genetic gain. Genetics 153: 1009-1020.

Woolliams, J.A. and Bijma, P. (2000) Predicting rates of inbreeding in population undergoing selection. Genetics 154: 1851-1864.

Woolliams, J.A. and Meuwissen, T.H.E. (1993) Decision rules and variance of response in breeding schemes. Anim. Prod. 56: 179-186.

Woolliams, J.A. and Thompson, R. (1994) A theory of genetic contributions. Proc. of the 5th World Congress on Genetics Applied to Livestock Production 19: 127-134.

Wray, N.R. and Thompson, R. (1990) Prediction of rates of inbreeding in selected populations. Genet. Res. Camb. 55: 41-54.

Wray, N.R. and Goddard, M.E. (1994) Increasing long-term response to selection. Genet. Sel. Evol. 26: 431-451.

Wright, S. (1931) Evolution in Mendelian populations. Genetics 16: 97-159.

Summary

This thesis deals with the definition of selection and mating criteria for animal breeding populations under selection and for genetic conservation populations, especially emphasizing on populations with small effective sizes that have known pedigrees.

The thesis can be divided into four main parts. Firstly, Chapters 2 and 3 deal with selection algorithms that manage ΔF for populations under selection with overlapping generations and random mating. Secondly, Chapters 4 and 5 deal with non-random mating schemes in combination with selection algorithms for discrete and overlapping generation structures, respectively. Thirdly, Chapters 6 and 7 deal with algorithms that minimize ΔF in small and endangered populations. In Chapter 6, ΔF is minimized for populations with overlapping generations. In Chapter 7, ΔF is further reduced by using frozen semen of sires from the less related base population. Fourthly, Chapter 8 deals with methods to select against genetic defects while restricting ΔF in populations with increased frequency of diseased alleles.

CHAPTERS 2 AND 3

In Chapter 2, a method is presented that maximizes the genetic merit of the selected animals while limiting the average coancestry of a population with overlapping generations after the current round of selection. For populations with overlapping generations, account has to be taken for previous and future use of animals of certain age-classes. Contributions within and over age-class were found by iteration. Inputs are Best Linear Unbiased Predicted (BLUP) breeding values of the selection candidates, and the relationship matrix of all animals. Output is the optimal number of offspring of each candidate. Computer simulations of dairy cattle nucleus schemes showed that the predefined rate of inbreeding was achieved. At the same rates of inbreeding, the dynamic selection method obtained up to 44% more genetic gain than truncation selection for BLUP breeding values. The advantage of the dynamic method over BLUP selection decreased with increasing population size and with less stringent restriction on inbreeding. In Chapter 3, the method of Chapter 2 was compared to a similar method that firstly optimized the distribution of parents within and thereafter over age classes per sex. It yielded significantly lower annual genetic gain, fewer animals selected and longer generation intervals, but maintained the rate of inbreeding closer to its constraint. The use of conventional relationships and of augmented relationships, which do not depend on the level of inbreeding, resulted in very similar breeding schemes, but the use of augmented relationships avoids correction of the current level of inbreeding. When optimising per generation, the generation interval was shorter compared to a scheme where an analogous annual restriction was in place and the annual rate of genetic gain was higher.

CHAPTER 4 AND 5

In Chapter 4, the effect of non-random mating on genetic gain was compared for populations with discrete generations. Mating followed a selection step where the average coancestry of selected animals was constrained, while genetic gain was maximized. Minimum coancestry (MC), Minimum coancestry with a maximum of one offspring per full-sib family (MC1) and Minimum variance of the relationships of the offspring (MVRO) mating schemes resulted in a delay in inbreeding of about two generations compared to Random, Factorial and Compensatory mating. At the same ΔF, genetic gain was up to 22% higher for the MC1 and MVRO schemes compared to Random mating schemes. The effect of non-random mating was largest for small schemes or for schemes with a stringent restriction on ΔF. MC1 yielded the highest genetic gain in almost all selection schemes, with a lower computational cost than MVRO. In Chapter 5, MC1 mating scheme was compared with random mating schemes for populations with overlapping generations and a restriction on ΔF. When sires were progeny tested, these progeny tested bulls were selected instead of the young bulls, which lead to increased generation intervals, increased selection intensity of bulls and increased genetic gain (35% compared to a scheme without progeny testing). The effect of MC1 decreased for schemes with progeny testing. MC1 mating increased genetic gain 11-18% for overlapping and 1-4% for discrete generations, when schemes with similar rate of inbreeding and genetic gain per generation were compared.

CHAPTER 6 AND 7

In Chapter 6, a method that minimizes the increase of coancestry of parents and optimizes the contribution of each selection candidate for populations with overlapping generations is presented. When survival rate equalled 100%, only animals from the oldest age class were selected, which maximized the number of parents per generation, slowed down the turn over of generations and minimized the increase of coancestry across sublines. However, the population became split into sublines separated by age classes, which substantially increased inbreeding within sublines. Sublines were prevented by a restriction of selecting at least one sire and one dam from the second oldest age class, which resulted in an L times lower ΔF, where L equals the average generation interval of sires and dams. Minimum coancestry mating resulted in lower levels of inbreeding than random mating, but ΔF was approximately the same or somewhat higher. For schemes where only the oldest animals were selected, ΔF increased with 18-52% compared with the proposed method. In Chapter 7, the advantage on the average coancestry level of selecting not only the least related sires from the oldest age-class, but also sires from the second oldest age-class is presented. By selecting sires from generation zero only, all genes will eventually descend from the founder sires and all genes from the founder dams are lost. By selecting sires from generation zero and one alternatively, also some genes of the founder dams will be conserved and the average coancestry level was approximately 20% lower than for a scheme, where only the oldest sires were used. The ΔF was zero at equilibrium for both schemes. Dams could be used for one generation and sires unlimited, because the amount of frozen semen very large relative to the small population sires. Population size was 6, 10 and 20 and the schemes were symmetric with respect to the sexes.

CHAPTER 8

Increased inbreeding will result in increased frequency of detrimental alleles. In Chapter 8, different genetic models and evaluation systems to select against a genetic disease in populations with discrete generations are compared. When using optimum contribution selection with a restriction on ΔF of 1.0% to select against a single gene, selection directly on DNA-genotypes needed 2.0 generations to half the frequency of the disease allele with additive effects and a population with 100 new-born animals. When only phenotypic records were available, selection on BLUP or on genotype probabilities calculated by segregation analysis (SEGR) needed 1.0 or 2.0 generations longer to half the frequency of the disease allele when allele effects were additive or recessive, respectively. Smaller schemes or schemes with a more stringent restriction on ΔF needed more generations to half the frequency of the diseased allele or the fraction of diseased animals. SEGR and BLUP were approximately equally efficient under both single gene and polygenic inheritance models, suggesting that efficient selection against a disease is possible without knowing its mode of inheritance.

In conclusion, by taking account of the relationships of the selected group of selected animals, inbreeding is controlled in the selection and mating criteria presented in this thesis. This principle was extended to populations with overlapping generations, which makes the methods useful for practical selection and genetic conservation populations with known pedigree. Only in well-controlled breeding schemes, the optimum contributions of each selection candidate will be realized, some deviations may be corrected in later rounds of selection. The optimized contributions and thus also the restriction on ΔF affected the structure of the breeding schemes, *e.g.* whether progeny tested animals were selected or not. In general, older and more parents were selected with a more stringent selection on ΔF. Due to such dynamic adaptations of the breeding schemes, genetic gain reduced only little when ΔF was lowered. Non-random mating can improve the family structure of the population under selection, thereby further increase the genetic gain. For the genetic conservation schemes, the contributions per family could not become completely equalized, because of the overlapping structure of the generations, resulting in a level of average coancestry being somewhat higher than the theoretical minimum. Use of frozen semen from sires of the oldest generations could reduce ΔF to zero and reduce the average level of coancestry in combined *in situ/ex situ* conservation schemes.

Samenvatting

INLEIDING

In dit proefschrift worden selectie en paringscriteria gedefinieerd voor twee soorten fokprogramma's voor landbouwhuisdieren: selectie-programma's en genetische conserverings-programma's. Het doel van een selectie-programma's is om de erfelijke aanleg voor een gedefinieerd fokdoel te verbeteren. De verbetering van de erfelijke aanleg wordt bereikt door in elke generatie de dieren met de hoogste fokwaarde te selecteren als ouders voor de volgende generatie. Fokwaarden zijn gebaseerd op fenotypische waarnemingen van kenmerken aan selectiekandidaten en/of aan familieleden. Het doel van genetische conserverings-programma's is om de bloedvoering van de populatie zo breed mogelijk te houden. Er wordt in genetische conserverings-programma's minder op het verbeteren van specifieke eigenschappen gelet. In sommige conserverings-programma's wil men echter rastypische kenmerken behouden, b.v. kleurtekeningen, waardoor het programma meer op een selectie-programma lijkt. Omgekeerd is in selectie-programma's voldoende breedte van bloedvoering vaak wel degelijk punt van aandacht.

Inteelt is een begrip dat centraal staat in zowel selectie- als genetische conserverings-programma's. Nakomelingen die ontstaan uit een paring tussen verwante dieren zijn ingeteeld. In een populatie zijn er hogere verwantschappen tussen dieren (*b.v.* volle broers en sussen) en lagere (*t.g.v.* een voorouder die vele generaties geleden leefde), zodat in een gesloten populatie inteelt nooit volledig kan worden vermeden. Inteelt heeft nadelige effecten op de fitness van een populatie, verlaagt de genetische variatie (diversiteit) en verhoogt de kans op een verhoogde frequentie van erfelijke aandoeningen. Doordat de laatste jaren bij de berekening van fokwaarden steeds meer ook familie-informatie wordt gebruikt (zogenaamde BLUP-fokwaardes), komen de beste dieren vaker uit een beperkt aantal families wat in de fokpopulaties leidt tot hogere inteelt. Tegelijkertijd zijn het juist de combinaties van dieren met een hoge fokwaarde die leiden tot hogere genetische vooruitgang. Selectie-programma's moeten daarom dieren met de hoogste fokwaardes combineren en tegelijkertijd inteelt zien te vermijden.

In zijn algemeenheid kun je zeggen dat inteelt hoog is als de fokpopulatie klein is. In selectie-programma's kan de inteelt hoog zijn terwijl de populatieomvang toch groot is, omdat slechts een klein aantal families wordt gebruikt voor het voortbrengen van de volgende generatie. In genetische conserverings-programma's is daarentegen vaak de fysieke populatie omvang klein. Een bedreigde populatie heeft vaak een lagere commerciële waarde omdat andere rassen beter (geacht worden te) produceren, wat de interesse voor en daarom ook de grootte van die populatie verlaagt. Ecologische en socio-economische veranderingen zijn twee andere factoren die bijdragen tot een kleinere populatieomvang. Inteelt wordt verlaagd door het inbrengen van onverwante dieren in de populatie. Echter, voor een commerciële fokpopulatie is het introduceren van onverwante dieren uit een vreemde populatie vaak geen optie, omdat de erfelijke aanleg voor het beoogde fokdoel in de vreemde populatie lager is. Bovendien verdwijnen er steeds meer populaties waardoor het vinden van een goed passende vreemde populatie steeds moeilijker wordt.

Dit proefschrift gaat erover hoe je de erfelijke vooruitgang kunt maximaliseren met een beperking van de inteelt-toename in selectie-programma's (hoofdstukken 2-5), en hoe je inteelt kunt minimaliseren in genetische conserverings-programma's (hoofdstukken 6-8). In het volgende zal ik de zeven hoofdstukken, die de kern van dit proefschrift vormen, kort samenvatten.

DIT PROEFSCHRIFT

In hoofdstuk twee wordt een methode gepresenteerd, die de genetische vooruitgang maximaliseert met een limiet op hoeveel de verwantschapsgraad en dus inteelt elke selectieronde mag toenemen. Deze methode is ontwikkeld voor populaties met overlappende generaties, waarbij een dier voor meerdere selectie-ronden beschikbaar is als selectiekandidaat. De methode houdt rekening met eerder en komend gebruik van dieren en optimaliseert het aantal nakomelingen, die elke selectie-kandidaat moet krijgen gegeven de beoogde limiet van de inteelt-toename. De methode moet balanceren tussen het selecteren van oudere, minder verwante dieren met een lagere fokwaarde (omdat de genetische vooruitgang steeds voortgaat) en jongere, meer verwante dieren met een hogere fokwaarde. In het algemeen kun je zeggen dat de hoogte van de limiet op inteelt bepaalt hoe oud de geselecteerde dieren zullen zijn: een hoge inteelt-limiet maakt het mogelijk om jonge dieren met hoge fokwaardes te selecteren, terwijl een lage inteelt-limiet betekent dat oudere minder verwante dieren geselecteerd zullen worden. Selectie van oudere dieren leidt tot een lang generatie interval hetgeen twee voordelen heeft: 1) dezelfde inteelt-toename per generatie geeft een lagere inteelt-toename per jaar; 2) een zelfde aantal geselecteerde dieren per jaar geeft meer dieren per generatie, hetgeen resulteert in een lagere gemiddelde inteelt. Bij gelijke inteelt-toename was de genetische vooruitgang tot 44% hoger vergeleken met de nu veel gebruikte selectie op BLUP-fokwaardes. De superioriteit van de methode neemt toe in meer strikte selectie-programma's waarbij de inteelt-limiet laag is en waar het aantal selectie-kandidaten klein. De methode is goed te gebruiken in praktische selectie-programma's omdat die vaak een overlappende generatie-structuur hebben.

In hoofdstuk vier wordt het effect van parings-schema's op de genetische vooruitgang onderzocht in populaties met discrete generaties, waarbij de dieren slechts één keer selectiekandidaat zijn. Eerdere literatuur liet zien dat specifieke paringen de inteelt-toename kunnen verlagen, maar dat ze nauwelijks effect op de genetische vooruitgang hadden. Echter, bij gebruik van de nieuwe selectie methode, die de inteelt-toename limiteert, leidde een parings-schema, dat probeert de minst verwante dieren met elkaar te paren en tevens probeert slechts één nakomeling per ouderpaar te realiseren, tot een 22% hogere genetische vooruitgang dan een niet-gestructureerd 'random' parings-schema. De specifieke paringen leidden tot een hogere genetische vooruitgang bij meer strikte selectie-programma's. Hoofdstuk vijf laat zien dat het effect van parings-schema's op genetische vooruitgang groter is in selectie-programma's met overlappende generaties (11-18%) vergeleken met programma's met discrete generaties (1-4%) bij dezelfde inteelt-toename en genetische vooruitgang per generatie.

In een *in situ* genetisch conserverings-programma wordt de populatie in stand gehouden in zijn originele productie-omgeving en kan zich dus blijven aanpassen aan die omgeving. In hoofdstuk zes wordt een methode gepresenteerd, die de gemiddelde verwantschap en dus inteelt minimaliseert in een *in situ* genetische conserverings-programma. Deze methode optimaliseert het aantal nakomelingen per selectie-kandidaat voor populaties met overlappende populaties waarbij dieren selectie-kandidaten zijn in meerdere selectie-ronden. Een theoretisch programma, waarbij alle kandidaten overleven tot volgende selectie-rondes tot een maximale leeftijd, leidt tot hoog ingeteelde sub-populaties doordat steeds de oudste, minst verwante dieren geselecteerd worden. In practische programma's zijn er echter te weinig dieren in de oudste leeftijdklasse en zullen ook jongere dieren geselecteerd worden. Dit laatste leidt tot overlap tussen generaties en voorkomt dus de vorming van sub-populaties. Toch is selectie van de oudste dieren in een populatie vaak een goed uitgangspunt als je inteelt wil minimaliseren, omdat oudere dieren gemiddeld minder verwant zijn dan jongere dieren. Selectie van alleen de oudste dieren leverde echter een inteelt-toename op, die 18-52% hoger was dan de in hoofdstuk zes gepresenteerde methode.

In een *ex situ* conserverings-programma wordt de populatie uit zijn originele omgeving gehaald en gehouden in een vreemde omgeving. Een extreem voorbeeld van een *ex situ* programma is het bewaren van het genetisch materiaal in een genenbank. Echter, materiaal uit een genenbank kan heel goed worden gebruikt om de inteelt-toename te verlagen in een *in situ* programma. In hoofdstuk zeven wordt onderzocht van welke generatie mannelijke dieren het best kunnen worden geselecteerd zodat de gemiddelde verwantschap van de *in situ* dieren wordt geminimaliseerd. De aanname was dat zaad van alle mannelijke dieren is ingevroren in een genen-bank, terwijl de vrouwelijke dieren alleen één generatie konden worden gebruikt (de genenbank bevatte alleen genetisch materiaal van mannelijke dieren). De gemiddelde verwantschapsgraad was 23% hoger voor een programma, waarbij alleen de oudste mannelijke dieren (uit generatie 0; bijv. de huidige generatie) worden gebruikt, vergeleken met een optimale programma waarbij ook mannelijke dieren, die een generatie jonger zijn worden geselecteerd. Omdat de gemiddelde verwantschap een maat is voor hoeveel genen er verloren zijn gegaan, worden er meer genen bewaard als ook mannelijke dieren uit generatie 1 worden geselecteerd. Deze extra bewaarde genen zijn in feite genen van de vrouwelijke dieren uit generatie 0. Zij zijn immers de moeders van de mannelijke dieren uit generatie 1. Het resulterende genetische conserverings-programma is dus een combinatie van levende dieren van een *in situ* programma en ingevrorene genetisch materiaal van een ex situ programma, waarbij zaad van 2 (oude) generaties van mannelijke dieren wordt ingevroren (generaties 0 en 1).

Een van de effecten van inteelt is een hogere kans op een verhoogde frequentie van ziekte-allelen. Zo'n verhoogde frequentie leidt ertoe dat een aanzienlijk deel van de dieren de zieke allelen (in homozygote vorm) bezit, en dus dat we een erfelijk gebrek waarnemen. In hoofdstuk acht wordt gekeken naar de beste manier om tegen een erfelijk gebrek te selecteren. Voor een erfelijk gebrek, dat alleen wordt bepaald door één gen, is directe selectie tegen het zieke-allel het meest effectief. Echter, in de meeste gevallen is het gen niet bekend en moet een andere manier gezocht worden om tegen het erfelijk gebrek te selecteren. Aangezien erfelijke gebreken veel in kleine populaties voorkomen, is een beperking van de inteelt-toename, en dus van de in hoofdstuk 2 ontwikkelde methode, gewenst. Deze methode maakt echter gebruik van een fokwaarde schatting. Twee fokwaarde-schattings-methoden zijn gebruikt 1) BLUP, waarbij wordt aangenomen dat het defect door veel genen en door milieu omstandigheden wordt bepaald; 2) segregatie-analyse, waarbij wordt aangenomen dat het defect slechts door één gen wordt bepaald. In de praktijk weten we niet hoeveel genen het defect bepalen, maar simulatie resultaten wezen uit dat in beide omstandigheden (één of meerdere genen) BLUP en segregatie analyse ongeveer gelijke resultaten gaven. Aangezien BLUP-fokwaarde-schatting meestal gemakkelijker is toe te passen dan segregatie analyse heeft het gebruik van BLUP fokwaarde schattingen een lichte voorkeur (ongeacht de vererving).

Al met al laat het proefschrift zien dat er in selectie-en parings-programma's reële mogelijkheden bestaan om genetische vooruitgang te combineren met een aanzienlijke beperking van de toename aan inteelt. Het proefschrift laat ook zien dat via het toepassen van doordachte conserverings-programma's ook bij kleine populaties behoud van genetische diversiteit in belangrijke mate mogelijk is.

Curriculum Vitae

I, Anna Kristina Sonesson, was born on March 13th 1970 and grew up in Helsingborg, Sweden, where I finished Naturvetenskaplig gymnasielinje at Olympiaskolan in 1988. Thereafter, I took a break from school, living an exciting life in Germany, working with good horses. Between 1991 and 1996, I studied Animal Science at the University of Agicultural Sciences in Uppsala. In 1997, I got a grant from the Royal Swedish Academy of Agriculture and Forestry, to work on a meat quality project at the Department of Animal Breeding and Genetics ID-DLO in Lelystad. In 1998, I started the project, which resulted in in this PhD. thesis. From August 1st 2002, I will work at the Research Division of Genetics and Breeding at AKVAFORSK, Norway.