Novel Methods for Improved Tree Breeding

Jon Hallander

Faculty of Forest Science
Department of Forest Genetics and Plant Physiology
Umeå

Doctoral Thesis
Swedish University of Agricultural Sciences
Umeå 2009

Cover: an improved Slash pine in North Carolina, USA
(photo: Jon Hallander)

Novel methods for improved tree breeding

Abstract

The development and implementation of statistical tools to improve inference in sustainable forest tree breeding are presented here. By combining classical quantitative genetic theory and novel statistical methods, a number of parameters are optimized. The results obtained are compared to those achieved by traditional methods for visualizing improvements to genetic parameters. The methods are tested on both simulated data and on a real Scots pine pedigree.

Modeling non-additive gene action using a finite loci model indicates that the development of the additive variance component does not decay initially as the underlying theory predicts. This phenomenon is shown for different sets of genetic components and models. In addition, variable numbers of loci were used so that different numbers of interactions could be captured.

To draw inferences about the genetic parameters, a powerful Bayesian Markov chain Monte Carlo method was developed. The method utilizes transformation of the genetic covariance structure to improve computational speed. By combining two different Bayesian Gibbs samplers, a useful hybrid sampler was developed; this was found to enhance convergence statistics and computational speed.

A method that finds the number of trees and their respective mating proportions that will maximize genetic gain was implemented and modified to handle a large number of selection candidates. When testing the selection method on a real pedigree an increase in genetic gain of up to 30 % was found compared to traditional methods in which the same restrictions were placed on relatedness. In order to provide a long-term breeding perspective, the selection method was combined with various mating schemes to examine the development of genetic parameters. A modified minimum coancestry mating scheme resulted in a level of genetic gain closest to the theoretically achievable limit while reducing the level of inbreeding in the population.

Author's address: Jon Hallander, Department of Forest Genetics and Plant Physiology, Swedish University of Agricultural Sciences, 901 83 Umeå, Sweden. *E-mail:* jon.hallander@genfys.slu.se

Till Annie

4

Contents

6

List of Publications

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:

I   Hallander, J. & Waldmann, P. (2007). The effect of non-additive genetic interactions on selection in multi-locus genetic models. *Heredity* 98, 1-11.

II Waldmann, P., Hallander, J., Hoti, F. & Sillanpää, M.J. (2008). Efficient MCMC implementation of Bayesian analysis of additive and dominance genetic variances in non-inbred pedigrees. *Genetics* 179, 1101-1112.

III Hallander, J. & Waldmann, P. (2009). Optimum contribution selection in large general tree breeding populations with an application to Scots pine. *Theoretical and Applied Genetics* [online] 1432, 2242. Available from: http://springerlink.com/content/t831115126071v62/fulltext.pdf [Accessed 31st January 2009].

IV Hallander, J. & Waldmann, P. Optimization of selection contribution and mate allocation in tree breeding populations (manuscript).

Papers **I-III** are reproduced with the kind permission of the publishers.

8

1 Introduction

1.1 Quantitative genetics and breeding

The field of quantitative genetics originates from the pioneering work of Fisher (1918), in which he demonstrated the connection between genetic variation within populations and covariances between relatives. By looking at many individuals at the population level, phenotypic trait variation can be partitioned into different genetic and non-genetic components which can be estimated by statistical analysis. The traits that are typically of interest are continuous ones where the underlying genetic architecture consists of an unknown number of genes which produce distributions of phenotypic expressions. In breeding, genetic variation in a population must be predicted to determine the potential of a selection program. Lush (1947) showed that the efficiency of breeding could be enhanced by deriving information from relatives. By tracking as many ancestors as possible that had transmitted copies of genes to their descendants in the current population, a pedigree could be created. One key question, therefore, is how to combine phenotypic records and pedigree information as efficiently as possible in order to increase production output from breeding programs.

With the introduction of the individual model (otherwise known as the animal model), Henderson (1963; 1973) resolved this key question. In the individual model, all individuals in the pedigree are connected via their additive genetic relationships using the mixed model equations (MME). Both phenotypic records and genetic covariances between individuals are

combined in an evaluation procedure which ensures that all information is utilized very efficiently. For example, the breeding value of one particular tree depends not only on its own phenotype, but also on the performances of its full and half-siblings and more distant relations such as cousins, parents etc. Since covariances between all individuals in the pedigree are taken into consideration, the gene flow within the pedigree is monitored, making the model robust with respect to selection and genetic drift situations (Kennedy *et al.*, 1988). This property is very important since most breeding populations are influenced by selection. Moreover, if a multiple trait analysis is performed, information from correlated traits is incorporated into the genetic evaluation procedure (Henderson and Quaas, 1976), which facilitates estimation of all included parameters. These features of the individual model highlight its power and flexibility and, during the last two decades, it has been the preferred choice in animal breeding when conducting genetic evaluations.

It is well known in quantitative genetics that the genetic variance underlying a complex trait can be divided into additive and non-additive components. The non-additive portion can be further decomposed into a dominance component due to within-locus interactions between alleles and an epistatic component resulting from interactions among alleles at different loci (Falconer and Mackay, 1996). To estimate non-additive variance in a population, complex pedigrees including multiple covariance terms between relatives are needed. For example, if both full and half-sibs are represented in the pedigree, both additive and dominance effects can be estimated in the statistical analysis. Traditionally in breeding programs, almost all attention has been devoted to the additive variance component since it determines the heritability of the breeding population and, hence, the possible response to selection via the standard quantitative genetic formula for gain (i.e. the "breeders equation"; Falconer and Mackay 1996). Non-additive variance, on the other hand, has generally been ignored since non-additive effects are difficult to utilize in breeding programs (Walsh, 2005). In addition, to accurately predict non-additive genetic parameters in a population, quality pedigrees and large computational recourses are needed (Ovaskainen *et al.*, 2008).

In general pedigrees, the standard additive model is not sufficient to describe accurately the covariance structure, since the residual variance is assumed to be homogenous (Lynch and Walsh, 1998). When analysing pedigrees that include a large proportion of full-sib families, for example in

10

diallel crossing designs that are often used in tree breeding populations, dominance relationships (covariances) arise between the individuals included, and these might complicate genetic parameter evaluations. As a result, if dominance deviations are omitted from the statistical analysis, both the additive variance and the heritability may be inflated, sometimes considerably so (e.g. **II**). Hence, by including genetic dominance effects in the statistical model as a location parameter, a more complete model should be obtained. This has been reported in animal breeding situations, where there is a large proportion of full-sibs in the pedigrees (Misztal and Besbes, 2000; Pante *et al*., 2002; Serenius *et al*., 2006). As a practical result, the ranking of selection candidates in the population might be changed when dominance is included; this may consequently increase the accuracy of selection (Pante *et al*., 2002; Wall *et al*., 2005; Serenius *et al*., 2006; **II**).

In closed populations (i.e. when there is no migration into the population), there will be an inevitable rise in inbreeding. Some individuals that share copies of genes inherited from common ancestors will mate; this will increase the proportion of homozygous loci in the population. For inbred populations, non-additive effects will play a major role in trait expression and will complicate the covariance structure (e.g. Harris, 1964; Smith and Mäki-Tanila, 1990; Edwards, 2008). As a result, the levels of both heterozygosity and within-family additive variance will be reduced; this has major effects on the breeding population. Reduction in within-family additive variance (c.f. Mendelian segregation variance) may lead to reduction in the future accumulation of genetic merit. A reduction in Mendelian segregation variance will decrease each individual's unique contribution to the population (i.e. the Mendelian sampling term) which will, in turn, reduce the total additive variance. Woolliams and Thompson (1994) demonstrated that the expectation of the rate in genetic merit in the breeding population is proportional to the Mendelian sampling term of the selected individuals within a pedigree. Similar conclusions have been drawn in investigations of the importance of the Mendelian sampling term to genetic gain (Grundy *et al*., 1998; Avendaño *et al*., 2004; 2005). Finally, as the level of inbreeding increases in the population, the risk of trees suffers from harmful inbreeding depression also increases. It has been proved that inbreeding depression is closely connected to fitness-related traits, since these are expected to account for a large portion of non-additive genetic variation (Crnokrak and Roff, 1995; Falconer and Mackay, 1996; Ovaskainen *et al*., 2008), which may lead to reduced reproductive capacity and decreased survival in forest tree populations (e.g. Williams and

Savolainen, 1996). To summarize, non-additive gene effects play a significant role in populations undergoing selection and/or genetic drift where inbreeding occurs, so the level of inbreeding needs to be carefully managed in selected populations.

From a theoretical point of view, simulation studies have shown that if non-additive genetic variance is present in a breeding population and if there is a finite number of loci coding for the trait of interest, the level of additive genetic variance might remain constant or even increase depending on the direction of the dominance and epistatic effects (Fuerst *et al*., 1997, Carter *et al*., 2005, **I**). This phenomenon has been verified empirically in genetic drift situations using several model species (e.g. Lindholm *et al*., 2005, Biggs and Goldmann, 2006; van Heerwaarden *et al*., 2008) and in long term poultry selection experiment (Carlborg *et al*., 2006). Similar conclusions have been drawn from simulations of other genetic drift situations, such as population bottlenecks and founder effects (e.g. Barton and Turelli, 2004).

A cornerstone of forest tree breeding theory is the selection procedure. For example, the breeder has to choose which trees to select from all the available candidate trees; those chosen will contribute to future generations and effective size of the breeding population. In addition, the mating proportion of each selected tree needs to be decided. A breeder generally wants to select the best-ranked trees according to their estimated breeding values (EBV), because the offspring of these trees are expected to perform in a similar way to their parents. Hence, by selecting the best parents, the selection differential is likely to be increased in comparison to randomly selected trees in the breeding population. There is, however, a problem with this strategy, since the top-ranked trees are likely to be related, particularly if the heritability of the target trait is low. The reason for this problem is that more family information is taken into account in genetic evaluations, which results in a higher correlation of EBV between relatives. If too many related trees are selected, the genetic variability of the population becomes severely reduced, which could reduce the long term response to selection. On the other hand, due to the long generation times in forest tree breeding in general, the short term response is very important. Therefore, a healthy compromise between short and long-term responses to selection is needed.

There have been massive efforts to develop efficient selection algorithms to increase selection response. Toro and Nieto (1984) introduced the concept

of unbalanced contribution of selected individuals; this increased response to selection compared to a method with equal contributions (see also Lindgren, 1991). Meuwissen (1997) introduced a dynamic optimization method for finding the number of candidates, and their respective mating proportions that maximizes genetic merit while restricting the relatedness (coancestry) of the population. This method, known as optimum contribution (OC), maximizes the selection differential by using quadratic indices (i.e. quadratic optimization functions). When compared to static methods, such as truncation selection, at the same level of relatedness, OC has been shown to be superior in simulations (Meuwissen, 1997; Grundy *et al*., 1998; Fernandez and Toro, 1999; Avendaño *et al*., 2004), in deterministic predictions (Villanueva *et al*., 2006) and in various practical applications with real pedigrees in animal breeding (e.g. Colleau *et al*., 2004; Kearney *et al*., 2004; Koenig and Simianer, 2006).

1.2 Limitations in traditional tree breeding

Traditionally, long-term forest tree breeding has used breeding cycles, or recurrent selection (e.g. Zobel and Talbert, 1984; White, 1987; Namkoong *et al*., 1998) to increase the genetic merit of the breeding population. Normally, one cycle includes the establishment of field trials to evaluate the performances of parent trees (selection candidates), followed by genetic evaluations of the trials, and finally, selecting which trees should contribute to the next generation. These steps are repeated cyclically in a long term breeding strategy and unfortunately, in general, not all of the information from earlier generations is incorporated into the current breeding cycle. For example, EBV of the parents are obtained using the performance of the offspring without taking complete pedigree information into consideration in the genetic evaluation procedure (i.e. family based models). Since a considerable proportion of the tree breeding programs worldwide have reached advanced cycles of breeding (e.g. Loblolly pine: McKeand and Bridgwater, 1998; Slash pine: White *et al*., 2003; Douglas fir: Jayawickrama *et al*., 2004; Maritime pine*:* Gaspar *et al*., 2008), information on previous generations might not be included in genetic evaluations and consequently, breeding values will be inaccurately estimated. Individual tree breeding programs will differ in details but the general framework is based on cyclical breeding. Moreover, family based models have mainly been used in the genetic evaluation procedure, and this is known to include higher

order interactions in the general combining ability term (Lynch and Walsh, 1998). These higher order effects are generally neglected and, hence, heritability predictions may be inflated. When simulating a population undergoing selection, Sorensen and Kennedy (1984) found that by using a sire model (i.e. a family based model), the predicted additive genetic variance was biased regardless of the estimation method used. Although the individual model was introduced to the field of forest tree improvement in the middle of last decade by Borralho (1995), family based models are still used to some extent when genetic analyses are performed for forest tree populations (i.e. Gaspar *et al*., 2008; Ward *et al.,* 2008).

Since selection is expected to increase relatedness in the population, some restrictions on selecting relatives are needed if long term selection response is to be maintained at a high level. Traditionally, to deal with this issue in forest tree breeding, a number of siblings are selected from all or some of the families represented (e.g. Zobel and Talbert, 1984). Either, an equal number of siblings are selected from all chosen families or the siblings are selected so that families that are known to perform better are over represented; the choice of approach depends on the severity of the restrictions on relatedness among the selected cohort of trees. The selected trees are then allowed to make equal contributions to the next generation, i.e. each tree participates in an identical number of crosses and an equal number of siblings from each family are tested (i.e. restricted truncation selection is applied). To improve selection among the candidate trees, Lindgren and Mullin (1997) proposed a method which optimizes the genetic improvement while weighting the coancestry of the selected set of trees. In tree improvement situations, the method has been widely applied, both in simulations (e.g. Rosvall and Andersson, 1999; Stoher *et al.*, 2007) and to real pedigrees (Olsson *et al.*, 2000). In these studies, the mating proportions (contributions) of selected trees were equal. Routsalainen and Lindgren (2001) varied the contribution of founder trees and found enhanced levels of genetic improvement in the next generation of the breeding population. Similar conclusions have been arrived at by using stochastic simulations of multiple generation tree breeding (Rosvall *et al.*, 2003; Lstiburek *et al.*, 2004; 2005), where boosts to both genetic merit in the breeding and deployment populations have been reported.

The selection methods described above generally do not adjust to the specific situation of the breeding population when deciding which trees should be chosen. For example, the sample size of the selected population

14

would typically be decided on beforehand, regardless the result of the genetic evaluation procedure. If the heritability is low, the predicted error variance of EBV would therefore be high, giving a less precise ranking of the candidate trees, which is then incorporated into selection decisions. Therefore, the aforementioned selection methods are static. Furthermore, these studies do not simultaneously select candidate trees and their individual contributions in an optimal way. Kerr *et al.* (1998) introduced the OC method developed by Meuwissen (1997) to a tree improvement context by adjusting the algorithm to account for monoecious species. This quadratic selection algorithm was tested by Fernandez and Toro (2001) and in **III** on real pedigrees of Eucalyptus and Scots pine, respectively, resulting in an apparent improvement in increased genetic gain or reduced coancestry compared to standard methodology. These studies are the only ones, however, applying a quadratic optimization method to tree breeding situations although the benefits are well proven in the breeding literature (e.g. Villanueva *et al.*, 2006; Woolliams, 2007).

While the methodology for selecting individuals in the breeding population has received much attention in the literature, the impact of the mating scheme on genetic parameters is less well studied. The mating between selected candidates in the breeding population influences the selection method in the next generation (when selecting the offspring of the mated parents) and thereby distinguishes between the genetic parameters depending on which mating scheme is employed. In the forest tree breeding literature, most effort has been devoted to comparing the performance of random mating (RM) with positive assortative mating (PAM) with respect to selection parameters, both in breeding and deployment populations (e.g. Rosvall and Mullin, 2003; Lstiburek *et al.*, 2004; 2005). When applying PAM to characteristic forest tree breeding scenarios, the aforementioned studies found an increase in additive genetic variance in the breeding population, compared to corresponding levels obtained by RM, when performing within-family selection (i.e. all families are represented within the set of selected trees). The enhanced additive variance was exploited in the deployment population when selecting the top-ranked trees according to EBV, producing a boost in genetic gain. The genetic improvement in the breeding population was, on the other hand, unaffected in comparison with the performance of RM. A strict PAM scheme might, however, result in a higher rate of inbreeding in the breeding population, particularly if the heritability is low (Rosvall and Mullin, 2003). Therefore, combining a strict PAM with avoiding the mating of close relatives could be a viable

alternative to strict PAM. In addition, one possible option would be to allocate crosses so that pairwise coancestry between mates is minimized (i.e. minimum coancestry mating; MCM). Consequently, inbreeding of the offspring would be minimized which would reduce the long term risk to the breeding program (i.e. reduced inbreeding in offspring). Kerr *et al.* (1998) applied minimum coancestry mating (MCM) in a forest tree breeding scenario and compared the obtained rate of inbreeding to the corresponding rate when applying RM to allocate mates among selected trees. Inbreeding was found to be delayed by one or two generations when using MCM even though the rate of inbreeding was equal between the schemes after the initial round of selection, as has been reported in animal breeding studies (e.g. Toro *et al.*, 1988; Sonesson and Meuwissen, 2000; 2001). Kerr *et al.* (1998) did not, however, compare the level of genetic improvement resulting from the different mate allocation strategies at the same restricted rate of inbreeding.

1.3 Objectives

The overall objective of the work described in this thesis was to improve the methodology for a traditional tree breeding program so that production output is optimized in both the short and long terms, given constraints on relatedness in the breeding population. In more detail, the aim of the thesis were as follows

➢ Since directed selection acts on a breeding population, non-additive effects and inbreeding will play an important role (e.g. Fuerst *et al.*, 1997). Investigation is warranted, therefore, into the role of non-additive effects on genetic parameters in the breeding population. What are the consequences of the direction of the non-additive effects, i.e. does negative or positive dominance result in varying release of additive variance and thereby different responses to selection? In addition, how does the number of simulated loci affect the release of additive genetic variation in the selected population? This is an issue that has been overlooked in the traditional breeding literature and requires further attention.

➢ I wished to develop and test methods that include pedigree information and dominance genetic effects so that genetic parameters

16

could be more accurately estimated. These parameters could then be used in selection decisions, thus increasing the accuracy of the genetic evaluations. The methods developed for estimating genetic parameters were applied to tree breeding pedigrees and compared to traditional methods used in forest tree breeding.

➢ Both short and long term responses to selection were optimized using real pedigrees and simulated while constraining the accumulation of relatedness in the breeding population. Selection of trees and their respective contribution to the next generation of the breeding population were performed simultaneously. This selection method is compared to traditional methods of selection (i.e. restricted truncation selection).

➢ I evaluated different methods of mating for creating future generations of the breeding population and combined these with optimized selection. If selection is applied to the breeding population over multiple generations, how are the levels of inbreeding and the long term contributions of the founders controlled? How can recombination events between parents be allocated for maximum response to selection? To examine the impact of the mating scheme on the breeding population, classic population genetic theory was incorporated into the analysis.

18

2 Theoretical considerations

2.1 Genetic models

In order to dissect the genetic architecture controlling complex traits, different models have been suggested within a quantitative genetic framework (e.g. Fisher, 1918; Cockerham, 1954). A standard tool for breeders in all disciplines is the infinitesimal genetic model, in which an infinite number of genes are assumed to contribute to trait expression with each gene having a small effect (Fisher, 1918; Falconer and Mackay, 1996). However, the rapid development of molecular techniques in the last decades has increased the possibility of utilizing molecular marker information in estimations of genetic parameters. For instance, the chances of identifying quantitative trait loci (QTL) that are partly responsible for the genetic variation of economically important traits in breeding have been greatly enhanced. This has given support to various forms of finite loci and mixed inheritance models, where one or several QTL are assumed to contribute to trait expression (e.g. Lstiburek *et al.*, 2005; Zeng *et al.*, 2005). Although in recent years much attention has been devoted to molecular methods in breeding, such as marker aided selection, the infinitesimal model is still the preferred choice, worldwide, when evaluating genetic parameters in most breeding programs for outcrossing species (e.g. Walsh and Henderson, 2004; Bernardo, 2008).

The main advantage of using a finite loci model as the genetic model is that changes in allele frequencies can be incorporated; this is a good feature in, for example, genetic drift studies. An additional advantage is that non-

additive interactions within and between loci are relatively simple to simulate, which is important when simulating QTL-effects, for example (Zeng *et al*., 2005). Moreover, if inbred pedigrees are considered, the covariance structure is more easily implemented using a finite locus model (de Boer and Hoeschele, 1993). One obvious drawback, however, is that results achieved using finite loci models are very difficult to generalise, since the results depend on the number of loci used in the model and how the genetic effects are defined. Linkage disequilibrium is, in general, assumed to be absent in most finite loci applications although Wang *et al*. (1998) presented a way of taking account of linkages when deriving genetic variance components.

The finite loci model considered in **I** was first described by Cockerham (1954) and further developed by several authors (i.e. Zeng *et al*., 2005). The basic idea is that, by using orthogonal contrast scales, the genetic variance of a population can be partitioned into variance components of different orders. The advantages of using orthogonal contrast is that the variance components will be orthogonal, which is a convenient property in the model since no bias will be introduced due to the covariance terms between different variance components. In the model of Cockerham (1954), the frequency at one locus is assumed to be uncorrelated with the frequency at any other locus and linkage disequilibrium is not taken into consideration. One positive feature of this model is that it is relatively straightforward to include higher order epistatic interaction terms between multiple loci, while most other finite locus models only consider pairwise interactions between two loci (for example, the model used by Fuerst *et al*., 1997). Furthermore, free recombination is assumed for loci both within and between chromosomes and the effect of inbreeding on variance components is not taken into account. Interested readers are invited to consult Zeng *et al*. (2005) or **I** for further details on the finite loci model considered here.

In the infinitesimal model, selection induces linkage disequilibrium which temporarily decreases the additive genetic variance. If the selection pressure is relaxed, the population would, after a number of generations of random mating, regain all of its lost additive variance by recombination. The non-additive variance is not affected by selection and hence, remains at the same level over the generations (Bulmer, 1985). It should be mentioned that the infinitesimal model does not try to describe the biological reality of complex traits since an infinite number of genes are assumed. If changes in allele frequencies are minor during one generation of applied selection and

if selection is acting on a large number of loci, however, the model has been shown to give very good approximations in selection experiments (e.g. Sorensen and Hill, 1983; Martinez *et al.*, 2000).

In **IV**, we made use of the infinitesimal model. Additive genetic effects are assumed to be inherited from parents through copies of genes, so that each individual receives the average additive value of its parents plus a Mendelian segregation term. For example, to generate the additive effect (i.e. true breeding value), $\mathbf{bv}_t$, for all trees in the current breeding population at generation t we use

$$\mathbf{bv}_t \sim \text{MVN}\left(\overline{\mathbf{bv}}_p, \mathbf{A}_t \sigma_a^2\right),$$

where $\overline{\mathbf{bv}}_p$ is a vector containing the average additive effect of the parents of the candidate trees (subscript p denotes parents at generation t-1), $\mathbf{A}_t$ is the additive relationship matrix between candidates, σ_a^2 is the additive genetic variance in the population. Note that by using $\mathbf{A}_t$, the reduction in the Mendelian segregation variance due to inbreeding is taken into account. Phenotypes were created for each simulated tree by adding the additive effect and a normally distributed environmental effect.

2.2 Statistical linear model

A key issue when analyzing breeding data is the linear model used to draw inferences, although non-linear models have been tested with various degree of success (e.g. Lynch and Walsh, 1998; Mrode and Thompson, 2005). For example, to separate the phenotypic variance in pedigrees into the underlying genetic effects, a well defined linear model is needed. Observed phenotypes need to be corrected with respect to various systematic effects, such as environmental block effects, so that obtained EBV are as accurate as possible. Furthermore, the structure of the data needs to be examined and taken into consideration when performing a genetic analysis. The framework for drawing inferences about parameters in a linear model is the mixed model equations (MME). The MME are a flexible tool in which a large number of parameters can be included, both genetic and non-genetic. See Searle *et al.* (1992) for a comprehensive investigation into how the MME are derived.

21

2.2.1 Individual tree model for additive and dominance genetic effects for non-inbred pedigrees

Normally, analysis of general tree breeding pedigrees consists of phenotypic measurements from some of the trees. Using the approach of Henderson (1985), both additive and dominance genetic parameters are included in the analysis, which utilizes the covariance structures of each effect when there is no inbreeding. The individual model uses a mixed model framework under Gaussian assumptions

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Za} + \mathbf{Zd} + \mathbf{e} \,,$$

where $\mathbf{y}$ is the individual's phenotypic record, $\mathbf{a}$ and $\mathbf{d}$ are location vectors of individual additive and dominance genetic values respectively, $\mathbf{X}$ and $\mathbf{Z}$ are known incidence matrices and $\mathbf{e}$ is a vector containing the individual residual errors around $\mathbf{X\beta} + \mathbf{Za} + \mathbf{Zd}$. Here, $\mathbf{a}$, $\mathbf{d}$ and $\mathbf{e}$ are assumed to follow a joint multivariate normal distribution. Henderson (1985) presented the MME including both additive and non-additive covariance structures according to

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z} + k_a \mathbf{A}^{-1} & \mathbf{Z'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z} & \mathbf{Z'Z} + k_d \mathbf{D}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \\ \hat{\mathbf{d}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \\ \mathbf{Z'y} \end{bmatrix} \,,$$

where $\mathbf{A}$ and $\mathbf{D}$ are the additive and dominance relationship matrices respectively, k_a is the ratio of residual and additive variance and k_d is the ratio of residual and dominance variance. In addition, Henderson (1985) showed how to obtain $\mathbf{D}$ for non-inbred pedigrees. Henderson and Quaas (1976) extended the individual model to include multiple traits in the analysis by including covariance between traits in the MME. This gives rise to a more complex and higher dimensional covariance structure although the same principles that underpin the univariate case remain. Multivariate analysis was used in **III**, where two traits were included (stem height and diameter at breast height).

2.3 Bayesian inference in breeding

22

During the last two decades, the awareness of Bayesian analysis has grown in the field of breeding genetics (e.g. Wang *et al.*, 1993; Sorensen and Gianola, 2002; Waldmann and Ericsson, 2006). Generally, Bayesian inference is considered to be a flexible and powerful tool, with the potential to deal with complex problems if there are a large number of parameters to be considered in the genetic analysis. The goal of Bayesian inference is to achieve the target distributions (i.e. marginal posterior distributions) of all parameters included in the statistical model. Here, a probabilistic framework is adopted, which reflects the uncertainty or the degree of belief in the parameters. This property generally makes the output of Bayesian analysis (i.e. marginal posterior distributions) easy to interpret. Moreover, another convenient feature of Bayesian inference is its ability to incorporate prior information into the analysis. If a breeder has a good knowledge of the parameters of interest (i.e. by performing an initial analysis), more precise priors can be adopted, thus facilitating inference.

In general, to obtain the marginal posterior distribution in Bayesian analysis, computationally demanding integration of the highly dimensional joint probability distribution of parameters and data has to be performed. One more convenient alternative is to apply a Markov chain Monte Carlo (MCMC) method to sample from the conditional posterior distribution, thereby approximating the marginal posterior distribution. Perhaps the most frequently used MCMC method is the Gibbs sampler (Geman and Geman, 1984; Sorensen and Gianola, 2002), in which parameters are repeatedly sampled from their conditional posterior distribution to form the Markov chain from which inference can be drawn (Sorensen and Gianola, 2002). In cases where the conditional posterior distribution cannot be obtained directly (i.e. the distribution does not have an identifiable form), one suitable alternative is the Metropolis-Hastings algorithm. Bayesian inference have been used frequently in animal breeding but less often in forest tree applications, although the interest has grown recently (e.g. Zeng *et al.*, 2004; Waldmann and Ericsson, 2006; Cappa and Cantet, 2006; 2008; **II**).

2.3.1 Hybrid Gibbs sampler

In Gibbs sampling, the location parameters can be drawn either element by element (single site sampler; Sorensen and Gianola, 2002) or in a single block (block sampler; García-Cortés and Sorensen, 1996). The main advantage of the single site sampler is the speed of completing one iteration, since the sampling of location parameters is from standard normal distributions; the more computer intensive blocked Gibbs sampler, samples from multivariate normal distributions. Since the number of trees included in the analysis is likely to be very large, computational time required to perform block updating of parameters could, unfortunately, be massive. Even for moderately large data sets, such as the pedigree analyzed in **II** and **III**, block updating could result in a computational time of several weeks on a small workstation. Another major advantage of the single site sampler is that there is no need to invert and update the entire coefficient matrix, **C**, only its diagonal.

On the other hand, by performing a block update of the parameters, mixing of the Markov chain is enhanced and better convergence properties are obtained. The single site sampler suffers from poor mixing, due to high correlation of the conditional posterior distributions as the parameters are sampled one by one.

One credible alternative to the fast, but poorly mixing single site sampler and the slow but efficient block sampler would be to combine these two into a hybrid sampler. Such a combination would typically involve using the single site sampler, but every 50th iteration would use the block sampler. By doing this the mixing of the Markov chain would be enhanced compared to the single site sampler, while the computational time required by the hybrid sampler would be considerably less than for the block sampler. For further details, see García-Cortés and Sorensen (1996), Sorensen and Gianola (2002) and **II**.

2.3.2 Variable transformation of genetic effects

Mrode and Thompson (1989) suggested a variable transformation of the covariance structures included in the MME. The purpose was to decrease the computational burden, thus making it possible to solve a higher dimensional equation system (i.e. increasing the size of the analyzed pedigree) by making the transformed parameters statistically independent. The greatest saving in computational time when solving MME is that the

24

transformed genetic parameters make it possible to sample from diagonal covariance matrices. As a result, instead of sampling from highly dimensional multivariate normal distributions, which have a computation time requirement proportional to the number of parameters to the power of three, sampling can be from simple univariate normal distributions. Moreover, if multiple covariance structures are included in the model (i.e. if the order of the coefficient matrix is high), the reduction in computational time can be very great compared to that required to evaluate untransformed models. For example, for the Scots pine pedigree analyzed in **II**, the reduction in computational time was at least fivefold.

2.4 Selection strategies

2.4.1 Restricted and unrestricted selection

In unrestricted truncation selection, all candidate trees are ranked according to EBV and the highest-ranking trees are selected without taking into account the genetic relationships between the trees. This strategy is expected to increase relatedness rapidly, particularly if heritability is low, since the best linear unbiased predictions (BLUP) procedure uses more family information to obtain EBV. In any tree breeding program worldwide, however, restrictions on relatedness are imposed since otherwise the long term response to selection would probably be greatly reduced. Therefore, restricted truncation selection has been the most common selection method (e.g. Zobel and Talbert, 1984). Here, we define a restricted selection strategy by first ranking families according to their average EBV. Then a number of the highest-ranked individuals within each of the best families are selected for further breeding, resulting in a future breeding population of the size specified by the breeding manager. To maintain a reasonable level of genetic variability, each parent tree is not allowed to contribute in more than a certain number of crosses; this maintains a large number of founder alleles within the breeding population. The contribution of the parent trees can easily be made unequal by selecting more trees from higher-ranked families, as described in **III**.

2.4.2 Optimized contribution dynamic selection

The OC algorithm uses Lagrange multipliers to solve the quadratic object function with an iterative procedure. First, the additive relationship matrix, $\mathbf{A_t}$ between all selection candidates at generation t is obtained from the additive relationship matrix between all individuals in the pedigree. In addition, a vector $\mathbf{a_t}$ containing the EBV of the selection candidates is constructed based on the BLUP evaluations. To maximize the level of genetic gain in the offspring at generation t+1, G_{t+1}, the following linear relationship is used

$$G_{t+1} = \mathbf{c'_t a_t},$$

where $\mathbf{c_t}'$ is a vector containing the mating proportion of each candidate in the current breeding population. The restriction on group coancestry is quadratic resulting in

$$C_{t+1} = \mathbf{c'_t A_t c_t}/2,$$

which holds true if the increase in group coancestry is small between generations (Meuwissen, 1997). Additional constraints are: 1. the sum of all mating proportions must be one; 2. an individual tree must not transfer a negative contribution to the next generation. The optimisation problem is solved by introducing two Lagrange multipliers, λ_0 and λ_1, and by maximizing the corresponding objective function

$$f(\mathbf{c_t}) = \mathbf{c'_t a_t} - \lambda_0 (\mathbf{c'_t A_t c_t} - 2C_{t+1}) - \lambda_1 (\mathbf{c'_t 1} - 1).$$

The individual that has the most negative mating proportion after performing the optimisation process above is removed from the process and the system is solved again. This iteration procedure continues until all remaining individuals have contributions that are not negative suggesting that the final solution has been obtained.

If the number of candidate trees available for selection is large, the inversion of the additive relationships between candidates, $\mathbf{A_t^{-1}}$, might be very time consuming and memory demanding, when solving the quadratic object function in each iteration. Hinrichs et al. (2006) derived an algorithm to compute $\mathbf{A_t^{-1}}$ by expressing the additive relationship between two

candidates based on the relationships between their parents. By implementing this algorithm, Hinrichs *et al.* (2006) were able to use the OC method on a pedigree containing 39214 individuals. I implemented the approach in **III** and **IV** with good outcomes.

One problem is that the optimal solution obtained, $\mathbf{c}_t$, is given as a proportion (a real number), although a breeding manager is more likely to want actual numbers of crosses per individual tree instead (an integer). There are several options for how to transform $\mathbf{c}_t$ into a number of crosses, however. In **III** and **IV** slightly different procedures were employed although, in general, the number of crosses for each tree was obtained by first multiplying $\mathbf{c}_t$ by twice the number of desired families N_{fam}, and then $2N_{fam}\mathbf{c}_t$ was rounded down to the integer below the actual value. The tree with the highest deviation between the integer and real value of $2N_{fam}\mathbf{c}_t$ had its number of crosses increased by 1 until the total number of families summed to N_{fam}. The size of each family was kept constant in **IV**.

2.5 Allocation of recombination events

There are several options for allocating matings between parents in a non-random fashion, that have been shown to improve genetic parameters in breeding situations (e.g. Sonesson and Meuwissen, 2000; Rosvall and Mullin, 2003). Two of the most frequently used mating schemes are: Minimized coancestry mating (MCM; Toro *et al.*, 1988); and Positive assortative mating (PAM; Baker, 1973). Additional rules can be applied to these mating schemes to improve the structure of the pedigree, such as minimizing variance in both family size and in pairwise coancestry between parents (to avoid matings between highly related parents, for example). These additional rules have been shown to facilitate selection parameters such as an increased rate of genetic gain (Kerr *et al.*, 1998; Sonesson and Meuwissen, 2000). I implemented and evaluated the following non-random mating schemes in **IV** combined with OC selection:

1. PAM combined with minimization of variance in family size (PAM)
2. PAM combined with minimization of variance in pairwise coancestry of parents (PAMCM)
3. Regular MCM with no additional constraints (MCM1)

27

4. MCM combined with minimization of variance in family size (MCM2)
5. MCM combined with minimization of family size (MCM3)
6. minimization of variance in both family size and pairwise coancestry of parents (MCM4)

All these schemes were optimized by using the simulated annealing algorithm (Kirkpatrick *et al*., 1983), which has been used successfully for calculating of optimal mating scenarios in various breeding and conservation programs (e.g. Kerr *et al*., 1998; Sonesson and Meuwissen 2000; 2001; Fernandez *et al*., 2003; Sanchez *et al*., 2008).

3 Summary of results

3.1 Influence of non-additive genetic effects on selection parameters

3.1.1 Variance components

When applying directional truncation selection on the base population during multiple generations, changes in allele frequencies and genetic variances were observed. Figure 1 show the trajectories of the variance components over time when a genetic network of two loci coded the trait of interest. If only the additive genetic effects of each allele were included in the genetic architecture of the simulated trait, V_A decayed as predicted by the underlying theory. If non-additive effects were included, however, an increase in V_A was obtained during the initial rounds of selection.
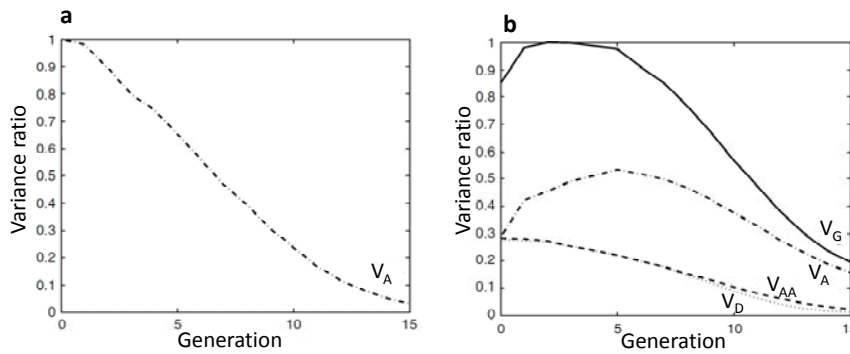
Figure 1. Ratio of genetic variance components with two loci. (**a**) pure additive gene action and (**b**) non-additive gene action.

The observed increase in V_A was largest after the first round of selection. Furthermore, the direction of the non-additive effects played an important role, since positive overdominance (i.e. heterozygote advantage) tended to reduce the increase in V_A, while negative overdominance acted in the opposite way. Although some fluctuations were noted in the case of four loci, in general, the non-additive variance components decreased when selection was applied (**I**).

3.1.2 Selection response

Clearly, the response to selection was greatly enhanced when non-additive gene action was included in the genetic architecture (Figure 2). The difference in selection response was greater between the purely additive and the non-additive genetic architectures if four loci contributed to trait expression. The most plausible reason is that if more loci are included in the genetic model, a higher number of interactions between alleles and loci can contribute to the trait expression. Consequently, a greater rate of genetic gain is obtained when more interaction terms contribute.
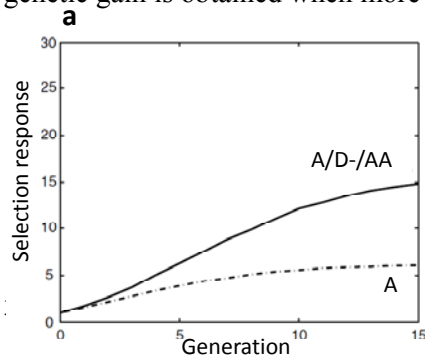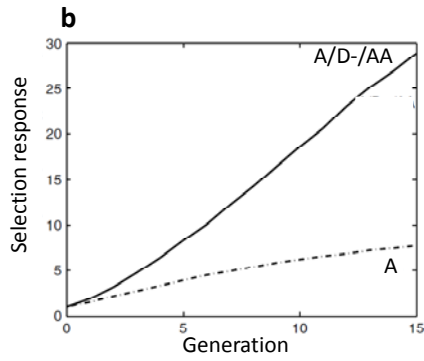

Figure 2. Response to selection over 15 generations measured as increase in phenotypic standard units; (**a**) two loci; (**b**) four loci.

b

3.2 Bayesian inference of additive and dominance variance components

3.2.1 Genetic parameters

The dominance variation, V_D, was greater than V_A for tree diameter, while the opposite was found for tree height (Table 1). This is also confirmed by the low and high heritability for diameter and height, respectively.

Table 1. *Achieved posterior mode and 95 % highest probability density (HPD) interval of the variance components, narrow sense heritability h^2 and dominance proportion d^2 of height and diameter in Scots pine.*

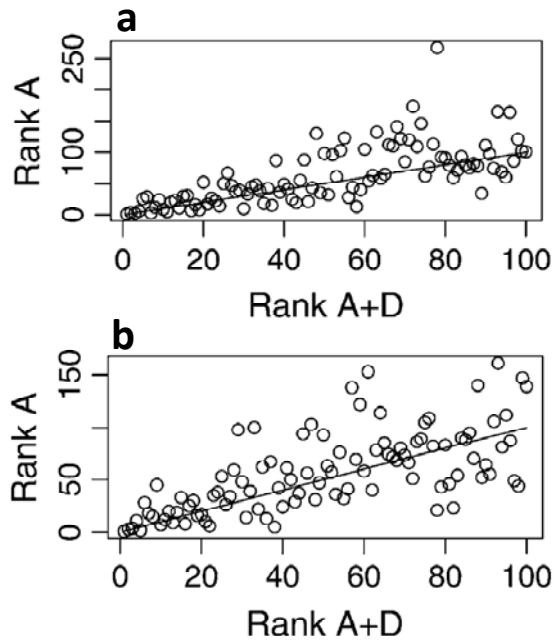|	Height			Diameter		
Parameter	Mode	HPD2.5	HPD97.5	Mode	HPD2.5	HPD97.5
V_A	32.01	18.89	50.56	54.70	27.67	103.7
V_D	16.27	8.632	26.46	82.88	39.70	142.2
V_E	103.8	90.01	114.0	722.2	665.3	776.8
h^2	0.210	0.130	0.314	0.063	0.033	0.117
d^2	0.105	0.055	0.170	0.094	0.046	0.162

Figure 3. Rank of EBV for the additive (A) and additive plus dominance (A + D) model of Scots pine data for (**a**) height and (**b**) diameter. The y-axis indicates the position in the A model of the 100 highest ranked individuals in the A + D model. Consistency in ranking between the two models is indicated by the straight line.

When the rank of the top 100 trees based on EBV was examined for both the additive and additive + dominance (full) model (Figure 3), different individuals were selected in 21 and 13 out of 100 trees for height and diameter, respectively. Another way to view this difference in rank is to focus on the correlation between the rank positions of the 100 highest-ranked trees; these were 0.696 and 0.668 for height and diameter, respectively.

3.3 Optimizing selection under a given constraint on relatedness

3.3.1 Comparison of selection methods

Table 3 shows the results of the different selection algorithms on the Scots pine data. OC outperformed the restricted selection approach in terms of genetic gain at the same level of coancestry for both height and diameter, yielding 8–30 % more gain. The biggest impact of OC was found for diameter at the lower level of coancestry. The number of selected trees, the number of families represented by means of the selected cohort of trees and

the number of suggested crosses were all considerably higher when using the OC algorithm.

Table 3. *Obtained levels of genetic merit (G) using restricted selection (RES) and optimized contribution (OC). Two levels of coancestry (C) where tested for RES and OC, the number of selected trees (N), the number of represented families (N_{fam}), the number of founders (N_f) and the number of crosses suggested by the selection algorithm (N_c).*

	Height						Diameter					
	C [%]	G	N	N_{fam}	N_f	N_c	C [%]	G	N	N_{fam}	N_f	N_c
RES	2.1	6.7	50	25	35	50	2.1	6.1	50	25	34	50
	2.6	7.8	50	25	35	50	2.6	7.6	50	25	34	50
OC	2.1	7.8	141	74	41	200	2.1	7.9	254	77	41	450
	2.6	8.4	105	56	36	150	2.6	9.0	188	44	31	500

3.4 Implication of mating scheme on population structure and selection parameters

3.4.1 Pedigree development and efficiency of mating scheme

Table 4 shows the residual variance of the regression of long term genetic contribution of founders (r) on the respective Mendelian sampling term; MCM2 produced the lowest residual variance while PAMCM produced the highest variance. The population structure obtained by MCM2 gives a lower sum of squared r which, in turn, results in lower inbreeding. The opposite was found for PAMCM, reflecting the high level of inbreeding in the population at generation 7.

Table 4. *Impact of mating scheme on residual variation of long term contribution on Mendelian sampling term (σ_e^2), sum of squared long term contribution ($\Sigma r^2/4$), level of inbreeding at generation 7(F) and genetic merit at generation 7 (G) when $\Delta C = 1\%$ and $h^2 = 0.05$. Standard errors are given in parenthesis.*

	$\sigma_e^2 [\cdot 10^{-4}]$	$\Sigma r^2/4 [\cdot 10^{-3}]$	F [%]	G
RM	4.27	6.76 (0.08)	7.1 (0.1)	5.62 (0.09)
PAMCM	5.74	7.05 (0.27)	10.3 (0.2)	5.48 (0.09)
MCM 2	3.86	6.27 (0.04)	5.0 (0.0)	6.01 (0.08)

3.4.2 Impact on selection parameters

Allocating matings according to minimum pairwise coancestry combined with various options to improve pedigree structure (i.e. lower sum of squared r) resulted in the highest level of accumulated gain at generation 7 in most scenarios. For example, in Table 4, gains obtained by MCM2 are higher than the corresponding level for RM. PAMCM schemes resulted in a similar level of gain as RM, although for some scenarios the gain associated with the former was somewhat lower.

The level of inbreeding obtained by RM, PAMCM and MCM2 is presented in Figure 4. PAMCM always resulted in the highest level of inbreeding after seven generations, while MCM2 reduced inbreeding compared to RM noticeably in all scenarios. MCM2 reduced the level of inbreeding during the initial rounds of selection, but the same rate of inbreeding as RM was reached after a delay of two generations.
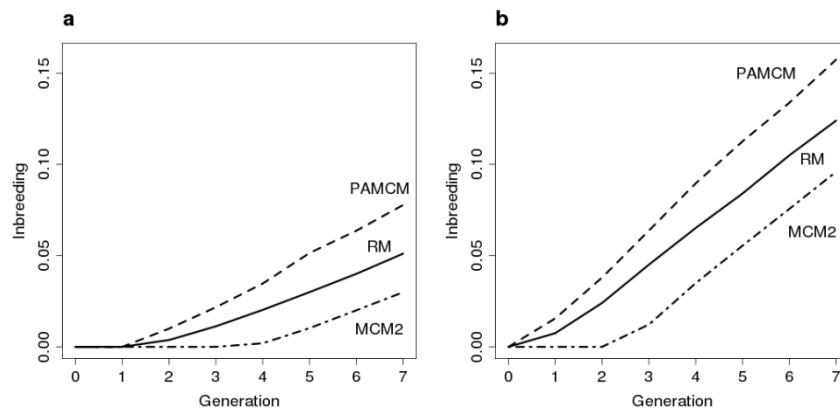

Figure 4. Accumulated level of inbreeding in the breeding population when $h^2 = 0.05$ for RM, PAMCM and MCM2 (**a**) $\Delta C = 1 \%$ (**b**) $\Delta C = 2 \%$

34

4 Discussion

4.1 The role of non-additive gene action for the release of additive variance during selection

The work underlying this thesis, has shown that the assumption from traditional quantitative genetic theory that a loss of genetic variation is to be expected in a population undergoing directional selection does not hold if non-additive genetic interactions and a finite number of loci are present in the genetic architecture. Similar conclusions have recently been reached by Carter *et al*. (2005) and Hansen *et al*. (2006). When using the finite loci model proposed by Zeng *et al*. (2005), multiple higher order interaction terms between alleles and loci can be included in the genetic model when calculating genotypic values. As a result, if more loci are simulated in the model, a higher number of interaction terms can be captured, resulting in an increased release in additive variance and higher response to selection (see Figure 2; Lynch and Walsh, 1998). Even though the contribution of each single interaction might be small, all combinations together could have a large impact (Walsh, 2005). In traditional finite loci models were only pairwise interactions are used, the additional release in additive variance will be a linear function related to the number of pairs of loci (e.g. Fuerst *et al*., 1997). In that respect, the model used in this thesis might be better for predicting the effect of higher order interactions. However, Turelli and Barton (2006) showed that a high level of inbreeding is required for maximal conversion of higher order components to contribute to the release of additive variance.

The direction of the non additive effects has been shown to influence the trajectory of the additive variance component and thereby the response to selection. It was shown in **I** that negative dominance (i.e. heterozygote disadvantage) increases the release in additive variance while positive dominance reduces the release in additive variance (see also Fuerst *et al.* 1997). Hansen *et al.* (2006), when examining the effects of epistasis on the response to selection, discovered that the direction of epistasis influenced evolvability in the additive variance. Moreover, when the initial levels of genetic variance components are equal, additive-by-additive epistasis seems to cause a greater release in additive variance than in dominance (López-Fanjul *et al.*, 2002; **I**). Clearly, non-additive effects influence selection response and hence, understanding the genetic composition that codes complex traits is an important issue in various areas of quantitative genetics (e.g. Aylor and Zeng, 2008). However, these issues need to be examined further using real and simulated data within a pedigree-based framework.

4.2 Incorporating dominance genetic effects into genetic parameter evaluations

In breeding, utilizing a dominance covariance structure in genetic evaluation procedures is uncommon, although Henderson (1985) derived the MME and included individual dominance effects. In both **II** and **III**, it was shown that by including dominance in the statistical model, a better fit to the data is obtained, both for a real pedigree and using simulated data. Based on the pedigree analyzed in **II** and **III**, when ranking the top 100 trees according to EBV in **II** and according to mating proportions obtained from OC in **III**, positions of individual trees differed vastly depending whether dominance was included or not. Hence, by excluding dominance from the genetic evaluation procedure, unfavorable trees might be selected to contribute to the next generation of the breeding population. Recall, from Table 1, that diameter had a higher ratio of dominance and additive variance than height, although the dominance proportion was higher for the latter due to the relatively low residual variance. The difference between the pure additive and full models among the top 100 rank positions was greater for height than for dominance. Hence, even in pedigrees for traits where the dominance variance component is thought to be relatively low, it

36

might still be important to incorporate individual dominance merit in the genetic evaluations. Similar conclusion was drawn by Ovaskainen et al. (2008) in an evolutionary context using simulations.

To my knowledge, no study has been published examining the impact of dominance on the accuracy of the estimated genetic parameters in forest tree pedigrees. It has been reported that in other species that have a large proportion of full-sibs in the analyzed pedigree, like the diallel design examined in **II** and **III**, estimations of the additive variance and heritability are changed when a dominance term is included. Misztal and Besbes (2000) reported that a standard additive model was inappropriate for analyzing traits related to egg production in poultry and that the goodness of fit was 20-77 % lower than for a model that took dominance into account. In rainbow trout populations, when analyzing harvest body weight, Pante *et al*. (2002) found decreased estimates of heritability and additive variance with a reduction of as much as 70 % compared to a pure additive model, when incorporating a dominance component. Similar conclusions have been drawn by Serenius *et al*. (2006) examining sow longevity in the Finnish pig breeding population. All these results point to the importance of including a dominance term for estimating genetic parameters when the pedigree contains a large portion of full-sibs.

4.3 Optimizing selection differential at a predefined level of relatedness

In **III**, a substantial increase in genetic gain was found when optimizing the selection differential compared to corresponding results obtained by applying restricted selection at the same level of group coancestry. The benefit of the OC algorithm was greatest when strict restrictions were placed on coancestry and at low heritability; in this case a 30 % increase in gain was obtained after one generation of selection, which agrees well with results of deterministic predictions (Villanueva *et al*., 2006). There are several reasons why OC outperformed the standard approach to selection. First, the highest proportion of the increased level of gain can be devoted to unequal mating proportions among the selected set of trees. Why should a better tree that is predicted to greatly enhance the genetic merit participate in the same number of crosses as an average tree, given that the same rate in group coancestry is maintained in the breeding population? Second, OC

selects more trees from a larger number of families (Table 3), so that good genotypes are not wasted and a higher number of trees are selected from the very best families (i.e. the algorithm utilizes the genetic variation efficiently). As shown by Grundy *et al.* (1998) and Avendaño *et al.* (2004) the selective advantage in an OC framework is the Mendelian sampling term, i.e. each individual tree's unique contribution to the population. By using the best available estimate of the Mendelian sampling term for each individual tree when computing the mating proportion, the OC algorithm ensures that the selection differential between families is maximized (Avendaño *et al.*, 2004; Woolliams, 2007).

By definition, implementing a quadratic selection algorithm will maximize the selection differential and hence, the genetic gain in the next generation. Consequently, if any other algorithm to select candidates in the breeding population is used, sub-optimal levels of gain will be achieved (Woolliams, 2007). All these results demonstrate the importance of optimizing selections in breeding programs, particularly if strict restrictions on relatedness are required and if heritability is low for the trait of interest.

4.4 Effect of different mating schemes when applying optimization of contributions

One key question in long term breeding is how to manage the contribution of all ancestors in an efficient way so that genetic gain is maximized in the future breeding population. Woolliams and Thompson (1994) found that the rate of genetic gain in the population is proportional to the covariance between long term genetic contributions and the Mendelian sampling term of all individuals in the pedigree (see also Woolliams *et al.*, 1999). The rate of gain is therefore, connected to the pedigree by means of the long term genetic contributions. Hence, the pedigree structure has a large impact on selection parameters and, consequently, it is very important to choose a mating scheme that carefully manages the long term contribution of ancestors in the pedigree over medium and long term breeding.

One way to examine the impact of the mating scheme on the genetic gain is to perform a regression of long term genetic contribution of ancestors on their respective Mendelian sampling term. Grundy *et al.* (1998) showed that in the ideal case (i.e. obtaining the highest possible level of genetic

gain), OC allocates long term contributions to the Mendelian sampling term in a linear fashion. The variance around the regression line (i.e. the residual variance) is the deviation from the highest attainable level of genetic gain and will therefore be a measure of the noise that does not contribute to gain (Avendaño *et al.*, 2004). In **IV**, MCM2 produced the lowest residual variance of the regression line with respect to the long term contribution of founders on their respective Mendelian sampling term compared to corresponding results obtained by RM and PAMCM. This suggests that MCM2 gives a genetic gain closer to the ideal theoretical limit and, thus, is better at managing the long term contributions within the pedigree. At the same time, the level of inbreeding was dramatically reduced in all scenarios by MCM2 due to the low sum of squares for the long term contribution of founders (Wray and Thompson, 1990; Woolliams and Thompson, 1994).

Worth noting is that PAM might have other beneficial effects if selection to the production population is taken into account as reported in literature (e.g. Rosvall and Mullin 2003; Lstiburek et al., 2004; 2005). A natural extension of **IV** would be to take the genetic parameters of the deployment population into consideration.

5 Conclusions

This thesis covers a range of different aspects of a typical tree breeding scenario within a quantitative genetic framework. By introducing novel methodology in tree breeding, improvements over traditional methods have been documented here, with respect to both simulated data and a real pedigree. If parts of the benefits shown here could be realized in deployment populations, the economic gain could be enhanced. Furthermore, the conceptual framework of the long term genetic contribution of ancestors is a suitable way to control the development of the pedigree better and to understand the mechanisms behind quadratic optimization of contributions (OC) in long term breeding. Powerful computer programs have been developed to achieve the results presented in this thesis, and these might be useful for breeding managers. For example, by using the OC algorithm to select clones and their respective proportions for deployment in seed orchards, enhanced levels of genetic gain might be achieved.

5.1 Future research

The field of quantitative genetics has expanded greatly in recent decades, particularly with respect to statistics and computation. Powerful MCMC Bayesian methods allow inference of complex statistical models to draw conclusions about parameters without making assumptions about sample size for hypothesis testing and significance levels. As statistical models become more and more highly dimensional, Bayesian methods would certainly help with sampling from relevant models, for example by using

probabilistic model selection techniques. More effort is warranted in this area of quantitative genetics and breeding.

The genetic architecture of quantitative traits remains a subject of debate in genetics and genomics. In general, most complex traits seem to be under the control of a large number of genes (e.g. Valdar *et al*., 2006), and these are the traits of most interest in breeding situations (Bernardo, 2008). These findings contradicts earlier studies in molecular genetics, which suggested that relatively few genes might explain a large portion of the trait variation that could be utilized in breeding programs (see citations in Kearsey and Farquhar, 1998, e.g.). One recent approach is genome wide selection (GWS; Meuwissen *et al*., 2001; Xu, 2003; Schaeffer, 2006) which has received much attention in the breeding literature. Whether GWS will become a feasible option in tree breeding remains to be seen. Nonetheless, the traditional quantitative genetics' view of genetic architecture of complex traits is still a robust tool for improving productivity through breeding.

References

Avendaño, S., Woolliams, J.A. & Villanueva, B. (2004). Mendelian sampling terms as a selective advantage in optimum breeding schemes with restriction on the rate of inbreeding. *Genetical Research* 83, 55-64.

Avendaño, S., Woolliams, J.A. & Villanueva, B. (2005). Prediction of accuracy of estimated Mendelian sampling terms. *Journal of Animal Breeding Genetics* 122, 302-308.

Aylor, D. L. & Zeng, Z. B. (2008). From classical genetics to quantitative genetics to systems biology: modeling epistasis. *PLoS Genetics* 4.

Baker, R. J. (1973). Assortative mating and artificial selection. *Heredity* 31, 231-238.

Barton, N.H. & Turelli, M. (2004). Effects of genetic drift on variance components under a general model of epistasis. *Evolution* 58, 2111-2132.

Bernardo, R. (2008). Molecular markers and selection for complex traits in plants: learning from the last 20 years. Crop Science 48, 1649-1664.

de Boer, I.J.M. & Hoeschele, I. (1993). Genetic evaluation methods for populations with dominance and inbreeding. *Theoretical and Applied Genetics* 86, 245-258.

Borralho, N.M.G. (1995). The impact of individual tree mixed models (BLUP) in tree breeding strategies. In: Potts BM, Borralho NMG, Reid J, Cromer R, Tibbits W, Raymond C (eds) Proceedings of the CRC–IUFRO conference on eucalypt plantations: improving fibre yield and quality, Hobart, Australia, 19–24 February 1995. CRC Temperate Hardwood Forestry, Hobart, pp 141-144.

Briggs, W.H. & Goldman, I.L. (2006). Genetic variation and selection response in model breeding populations of *Brassica rapa* following a diversity bottleneck. *Genetics* 172, 457-465.

Bulmer, M.G. (1985). *The mathematical theory of quantitative genetics*. Oxford: Clarendon Press.

Cappa, E.P. & Cantet, R.J.C. (2006). Bayesian inference for normal multiple-trait individual-tree models with missing records via full conjugate Gibbs. *Canadian Journal of Forest Research* 36, 1276-1285.

Cappa, E.P. & Cantet, R.J.C. (2008). Direct and competition additive effects in tree breeding: Bayesian estimation from an individual tree mixed model. *Silvae Genetica* 57, 45-56.

Carlborg, Ö., Jacobsson, L., Åhgren, P., Siegel, P. & Andersson, L. (2006). Epistasis and the release of genetic variation during long-term selection. *Nature Genetics* 38, 418-420.

Carter, A.J.R., Hermisson, J. & Hansen, T.F. (2005). The role of epistatic gene interactions in the response to selection and the evolution of evolvability. *Theoretical Population Biology* 68, 179-196.

Cockerham, C.C. (1954). An extension of the concept of partitioning hereditary variances for analysis of covariances among relatives when epistasis is present. *Genetics* 39, 859-882.

Colleau, J.J., Moureaux, S., Briend, M. & Bechu, J. (2004). A method for the dynamic management of genetic variability in dairy cattle. *Genetics Selection Evolution* 36, 373-394.

Crnokrak, P. & Roff, D.A. (1995). Dominance variance – association with selection and fitness. *Heredity* 75, 530-540.

Edwards, J.W. (2008). Predicted genetic gain and inbreeding depression with general inbreeding levels in selection candidates and offspring. *Crop Science* 48, 2086-2096.

Falconer, D.S. & Mackay, T.F.C. (1996). *Introduction to quantitative genetics*. New York, Longman.

Fernandez, J. & Toro, M.A. (1999). The use of mathematical programming to control inbreeding in selection schemes. *Journal of Animal Breeding Genetics* 116, 447-466.

Fernandez, J. & Toro, M.A. (2001). Controlling genetic variability by mathematical programming in a selection scheme on an open-pollinated population in *Eucalyptus globulus*. *Theoretical and Applied Genetics* 102, 1056-1064.

Fernandez, J., Toro, M.A. & Caballero, A. (2003). Fixed contributions designs vs. minimization of global coancestry to control inbreeding in small populations. *Genetics* 165, 885-894.

Fisher, R.A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh* 52, 399-433.

Fuerst, C., James, J.W., Sölkner, J. & Essl, A. (1997). Impact of dominance and epistasis on the genetic make-up of simulated populations under selection: a model development. *Journal of Animal Breeding Genetics* 114, 165-175.

García-Cortés, L. A. & Sorensen, D.A. (1996). On a multivariate implementation of the Gibbs sampler. *Genetics Selection Evolution* 28, 121-126.

Gaspar, M.J., Louzada, J.L., Silva, M.E., Aguiar, A. & Almeida, M.H. (2008). Age trends in genetic parameters of wood density components in 46 half-sibling families of Pinus pinaster. *Canadian Journal of Forest Research* 38, 1470 -1477.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions and Bayesian restoration of images. *IEEE Trans. of Pattern Analysis and Machine Intelligence* 6, 721-741.

Grundy, B., Villanueva, B. & Woolliams, J.A. (1998). Dynamic selection procedures for constrained inbreeding and their consequences for pedigree development, *Genetical Research* 72, 159-168.

Hansen, T.F., Alvarez-Castro, J.M., Carter, A.J.R., Hermisson, J., Wagner, G.P. (2006). Evolution of genetic architecture under directional selection. *Evolution* 60, 1523-1536.

Harris, D.L. (1964) Genotypic covariances between inbred relatives. *Genetics* 50**,** 1319-1348.

van Heerwaarden, B., Willi, Y., Kristensen, T.N. & Hoffmann, A.A. (2008). Population bottlenecks increase additive genetic variance but do not break a selection limit in rain forest Drosophila. *Genetics* 179, 2135-2146.

Henderson, C.R. (1953). Estimation of variance and covariance components. *Biometrics* 9, 226-252.

Henderson, C. R. (1963). Selection index and expected genetic advance. In W. D. Hanson and H. F. Robinson (Ed.). *Statistical genetics and plant breeding*. National Academy of Sciences-National Research Council, Washington, DC, Pub. 982.

Henderson, C. R. (1973). Sire evaluation and genetic trends. Proc. of the Anim. Breeding and Genetics Symp. in Honor of Dr. J. L. Lush, A.S.A.S. and A.D.S.A., Champaign, IL. p. 10.

Henderson, C. R. & Quaas R. L. (1976). Multiple Trait Evaluation Using Relatives' Records. *Journal of Animal Science* 43, 1188-1197.

Henderson, C. R. (1985). Best linear unbiased prediction of nonadditive genetic merits in noninbred populations. *Journal of Animal Science* 60, 111-117.

Hinrichs, D., Wetten, M. & Meuwissen, T.H.E. (2006). An algorithm to compute optimal genetic contributions in selection programs with large numbers of candidates. *Journal of Animal Science* 84, 3212-3218.

Jayawickrama, K.J.S., Johnson, G.R. & Ye, T. (2004). Cooperative advanced-generation breeding and testing of coastal Douglas-fir and Western Hemlock ‑ strategy and implementation. In: Proceedings: IUFRO joint conference of division 2. Forest genetics and tree breeding in the age of genomics: progress and future. B. Li and S. McKeand (eds). November 1-5, 2004. Charleston, South Carolina, USA. p. 101-103.

Jensen, J., Wang, C.S., Sorensen, D.A. & Gianola, D. (1994). Bayesian inference on variance and covariance components for traits influenced by maternal and direct genetic effects using the Gibbs sampler. *Acta Agriculturae Scandinavica* 44, 193-201.

Kearney, J.F., Wall, E., Villanueva, B. & Coffey, M.P. (2004). Inbreeding trends and application of optimized selection in the UK Holstein population. *Journal of Dairy Science* 87, 3503-3509.

Kearsey, M.J. & Farquhar, A.G.L. (1998). QTL analysis in plants; where are we now? *Heredity* 80, 137‑142.

Kennedy, B.W., Schaeffer, L.R. & Sorensen, D.A. (1988). Genetic properties of animal models. *Journal of Dairy Science* 71(Suppl. 2), 17-26.

Kerr, R.J., Goddard, M.E. & Jarvis, S.F. (1998). Maximising genetic response in tree breeding with constraints on group coancestry. *Silvae Genetica* 47, 165-173.

Kirkpatrick, S., Gelatt, C.D. & Vecchi, M.P. (1983). Optimization by simulated annealing. *Science* 220, 671-680.

Koenig, S. & Simianer, H. (2006). Approaches to the management of inbreeding and relationship in the German Holstein dairy cattle population. *Livestock Science* 103, 40-53.

Lindgren, D. (1991). Optimal utilization of genetic resources. *Forest Tree Improvement* 23, 49-67.

Lindgren, D. & Mullin, T.J. (1997). Balancing gain and relatedness in selection. *Silvae Genetica* 46, 124-129.

Lindholm, AK, Breden F, Alexander HJ, Chan, W.K., Thakurta, S.G. & Brooks, R. (2005). Invasion success and genetic diversity of introduced populations of Guppies *Poecilia reticulata* in Australia. *Molecular Ecology* 14, 3671-3682.

López-Fanjul, C., Fernandez, A. & Toro, M.A. (2002). The effects of epistasis on the excess of the additive and nonadditive variances after population bottlenecks. *Evolution* 56, 865-876.

Lstiburek, M., Mullin, T.J., Lindgren, D. & Rosvall, O. (2004). Open-nucleus breeding strategies compared   with population-wide positive assortative mating. *Theoretical and Applied Genetics* 109, 1169-1177.

Lstiburek, M., Mullin, T.J., Mackay, T.F.C., Huber, D. & Li, B. (2005). Positive assortative mating with family size as a function of predicted parental breeding values. *Genetics* 171, 1311-1320.

Lush, J.L. (1947). Family merit and individual merit as bases for selection. *American Naturalist* 81, 241-261.

Lynch, M., & Walsh, B. (1998). *Genetics and analysis of quantitative traits*. Sunderland, Sinauer Associates.

Martinez, V., Bunger, L. & Hill, W.G. (2000). Analysis of response to 20 generations of selection for body composition in mice: fit to infinitesimal model assumptions. *Genetics Selection Evolution* 32, 3-21.

McKeand, S.E. & Bridgwater, F.E. (1998). A strategy for the third breeding cycle of loblolly pine in the Southeastern US. *Silvae Genetica* 47, 223-234.

Meuwissen, T.H.E. (1997). Maximizing the response of selection with a predefined rate of inbreeding. *Journal of Animal Science* 75, 934-940.

Meuwissen, T.H.E., Hayes, B.J. & Goddard, M.E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819-1829.

Misztal, I. & Besbes, B. (2000). Estimates of parental-dominance and full-sib permanent environment variances in laying hens. *Animal Science* 71, 421 -426.

Mrode, R. & Thompson, R. (1989). An alternative algorithm for incorporating the relationships between animals in estimating variance components. *Journal of Animal Breeding Genetics* 106, 89-95.

Mrode, R. & Thompson, R. (2005). *Linear models for the prediction of animal breeding values*. Wallingford, CABI Publishers.

Namkoong, G., Kang, H. & Brouard, J.S. (1998). *Tree breeding: Principles and Strategies*. New York, Springer Verlag.

Olsson, T., Lindgren, D. & Ericsson, T. (2000). A comparison of group merit selection and restricted selection among full-sib progenies of Scots pine. *Forest Genetics* 7, 137-144.

Ovaskainen, O., Cano, J.M. & Merilä, J. (2008). A Bayesian framework for comparative quantitative genetics. *Proceedings of the Royal Society B* 275, 669-678.

Pante, M.J.R., Gjerde, B., McMillan, I. & Misztal, I. (2002). Estimation of additive and dominance genetic variances for body weight at harvest in Rainbow trout, *Oncorhynchus mykiss*. *Aquaculture* 204, 383-392.

Rosvall, O. & Andersson, E. (1999). Group-merit selection compared to conventional restricted selection for trade-offs between genetic gain and diversity. *Forest Genetics* 6, 11-24.

Rosvall, O. & Mullin, T.J. (2003). Positive assortative mating with selection restrictions on group coancestry enhances gain while conserving genetic diversity in long-term forest tree breeding. *Theoretical and Applied Genetics* 107, 629-642.

Rosvall, O., Mullin, T.J. & Lindgren, D. (2003). Controlling parent contributions during positive assortative mating and selection increases gain in long-term forest tree breeding. *Forest Genetics* 10, 35-53.

Ruotsalainen, S. & Lindgren, D. (2001). Number of founders for a breeding population using variable parental contribution. *Forest Genetics* 8, 57-67.

Sanchez, L., Yanchuk, A.A. & King, J.N. (2008). Gametic models for multitrait selection schemes to study variance of response and drift under adverse genetic correlations. *Tree Genetics and Genomes* 4,201-212.

Searle, S.R., Casella, G. & McCulloch, C.E. (1992). *Variance Components*. New York, Wiley.

Serenius, T., Stalder, K.J. & Pounti, M. (2006). Impact of dominance effects on sow longevity. *Journal of Animal Breeding Genetics* 123, 355-361.

Schaeffer, L.R. (2006). Strategy for applying genome-wide selection in dairy cattle. *Journal of Animal Breeding Genetics* 123, 218-223.

Smith, S.P. & Mäki-Tanila, A. (1990). Genotypic covariance matrices and their inverses for models allowing dominance and inbreeding. *Genetics Selection Evolution* 22, 65-91.

Sonesson, A.K. & Meuwissen, T.H.E. (2000). Mating schemes for optimum contribution selection with constrained rates of inbreeding, *Genetics Selection Evolution* 32, 231-248.

Sonesson, A.K. & Meuwissen, T.H.E. (2001). Minimization of rate of inbreeding for small populations with overlapping generations. Genetical Research 77, 285-292.

Sorensen, D.A. & Hill, W.G. (1982). Effect of short term directional selection on genetic variability: experiments with *Drosophila melanogaster. Heredity* 48, 27-33.

Sorenson, D.A. & Hill, W.G. (1983). Effects of disruptive selection on genetic variance. *Theoretical and Applied Genetics* 65, 173-180.

Sorensen, D.A. & Kennedy, B.W. (1984). Estimation of genetic variances from unselected and selected populations. *Journal of Animal Science* 59, 1213-1223.

Sorensen, D. & Gianola, D. (2002). *Likelihood, Bayesian and MCMC methods in quantitative genetics*. Berlin, Springer.

Stoehr, M., Yanchuk, A., Xie, C.-Y. & Sanchez, L. (2007). Gain and diversity in advanced generation coastal Douglas-fir selections for seed production populations. *Tree Genetics and Genomes* 4, 193-200.

Toro, M.A. & Nieto, B. (1984). A simple method for increasing the response to artificial selection. *Genetical Research* 44, 347–349.

46

Toro, M.A., Nieto, B. & Salgado, C. (1988). A note on minimization of inbreeding in small-scale selection programs. *Livestock Production Science* 20, 317-323.

Turelli, M. & Barton, N.H. (2006). Will population bottlenecks and multilocus epistasis increase additive genetic variance? *Evolution* 60, 1763-1776.

Valdar, W., Solberg, L.C., Gauguier, D., Burnett, S., Klenerman, P., Cookson, W.O., Taylor, M.S., Rawlins, J.N.P., Mott, R. & Flint, J. (2006). Genome-wide genetic association of complex traits in heterogeneous stock mice. *Nature Genetics* 38, 879-887.

Waldmann, P. & Ericsson, T. (2006). Comparison of REML and Gibbs sampling estimates of multi-trait genetic parameters in Scots pine. *Theoretical and Applied Genetics* 112, 1441-1451.

Wall, E., Brotherstone, S., Kearney, J.F., Woolliams, J.A. & Coffey, M.P. (2005). Impact of nonadditive genetic effects in the estimation of breeding values for fertility and correlated traits. *Journal of Dairy Science* 88, 376-385.

Walsh, B. & Henderson, D. (2004). Microarrays and beyond: What potential do current and future genomics tools have for breeders? *Journal of Animal Science* 82, 292-299.

Walsh, B. (2005). The struggle to exploit non-additive variation. *Austrailian Journal of Agricultural Research* 56, 873-881.

Wang, C.S., Rutledge, J.J. & Gianola, D. (1993). Marginal inference about variance components in a mixed linear model using Gibbs sampling. *Genetics Selection Evolution* 21, 41-62.

Wang, J., Caballero, A. & Hill, W.G. (1998). The effect of linkage disequilibrium and deviation from Hardy-Weinberg proportions on the changes in genetic variance with bottlenecking. *Heredity* 81, 174-186.

Ward, S.E., Wightman, K.E. & Santiago, B.R. (2008). Early results from genetic trials on the growth of Spanish cedar and its susceptibility to the shoot borer moth in the Yucatan Peninsula, Mexico. *Forest Ecology and Management* 255, 356-364.

White, T.L. (1987). A conceptual framework for tree improvement programs. *New Forests* 4, 325-342.

White, T. L., Huber, D. A. & Powell, G. L. (2003). Third-cycle breeding strategy for Slash pine by the cooperative forest genetics research program. 27th Southern Forest Tree Improvement Conference, Stillwater, OK, USA.

Villanueva, B., Avendaño, S. & Woolliams, J.A. (2006). Prediction of genetic gain from quadratic optimisation with constrained rates of inbreeding. *Genetics Selection Evolution* 38, 127 -146.

Williams, C.G. & Savolainen, O. (1996). Inbreeding depression in conifers: implications for breeding strategy. *Forest Science* 42, 102-117.

Woolliams, J.A. & Thompson, R. (1994). A theory of genetic contributions. Proceedings, 5th World Congress on Genetics Applied to Livestock Production, University of Guelph, Guelph, Ontario, Canada, 7-12 August 1994. Volume 19. Selection and quantitative genetics; growth; reproduction; lactation; fish; fiber; meat.

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

Woolliams, J.A. (2007). Genetic contribution and inbreeding. In: K. Oldenbroek (ed.) *Utilisation and Conservation of Farm Animal Genetic Resources*. AE Wageningen, Wageningen Academic Publishing, pp. 147-165.

Wray, N.R. & Thompson, R. (1990). Prediction of rates of inbreeding in selected populations. *Genetical Research* 55, 41-54.

Xu, S. (2003). Estimating polygenic effects using markers of the entire genome. *Genetics* 163, 789-801.

Zeng, W., Gosh, S. & Li, B. (2004). A blocking Gibbs sampling method to detect major genes with phenotypic data from a diallel mating. *Genetical Research* 83, 143-154.

Zeng, Z.-B., Wang, T. & Zou, W. (2005). Modeling quantitative trait loci and interpretation of models. *Genetics* 169, 1711-1725.

Zobel, B. & Talbert, J.T. (1984). *Applied forest tree improvement*. New York, John Wiley & Sons.

Acknowledgements

All those associated with the research school in forest genetics and breeding. Thanks for all the fun moments, for example when we were ordering food at a restaurant in Raleigh… It is always great to meet all of you.

All my roommates at UPSC, in particular Robert, Sara and Vaibhav who survived all five years with me. Thanks to Juha, Daniel, Benjamin, Aurora, Mattia, Nils, Chat, Sultana, Abdul and everyone who has shared a cup of coffee in the lunch room with me.

Eva, Leif, Ingrid, Daniel, Margareta, Ebba and Veide who always support and help me massively. Thanks a lot! Per, Lena and Ida for your generosity and help. Also thanks to all friends for the fun moments during these years.

My dear Annie for the tremendous support you have shown me. Without you, finishing this thesis would not have been possible. Now I am really looking forward to meeting the new member of our family this summer!

Version 1, February 2008.