

## ABSTRACT

**Lstibůrek, Milan.** Population response to positive assortative mating in forest tree breeding.  
(Under the direction of Dr. Timothy J. Mullin.)

Positive assortative mating (PAM) may substantially enhance genetic variance in a breeding population (BP). This creates potential for additional genetic gains available through production populations (PP) to forest plantations. Open-nucleus strategies (NB) have been incorporated in forest tree breeding programs. In NB, the BP is subdivided into two hierarchical levels, a nucleus and a main population, and can be considered a less rigorous form of PAM. First, PAM was compared to NB by stochastic simulation considering jointly genetic gain and diversity within the framework of a long-term breeding program. Test effort was either assumed constant throughout the entire BP or was redirected according to the rank of each mate. The simulation revealed that PAM results in larger gains in the PP compared to NB under both situations and at any target PP diversity. Second, the test effort during PAM was redirected by varying family sizes as a linear function of mid-parent BLUP values. The actual distribution of mid-parent BLUP values was standardized by a constant value, which was varied in simulation scenarios to cover the entire range of the distribution of family sizes. When equal numbers of progenies were selected per family and the variation in family sizes was maximized, only a minimal reduction in BP diversity was observed, compared to cases with constant family sizes. Under such favorable conditions, the redistribution of resources increased genetic response and variance in the BP, causing substantially greater genetic response in the PP. These conclusions were verified under a mixed-inheritance model

with a major-gene locus contributing to variation in a quantitative trait. Finally, the investigation of PAM was extended by considering correlated traits within the framework of a clonal forestry program. The success of somatic embryogenesis in families generated by crossing elite genotypes developed in the breeding program was either considered exponentially distributed or constant. The distribution of success caused non-significant differences in genetic gain of PP. These conclusions were verified over a range of correlation, heritabilities and economic weights of traits.

**POPULATION RESPONSE TO POSITIVE ASSORTATIVE MATING  
IN FOREST TREE BREEDING**

by

**Milan Lstibůrek**

A dissertation submitted to the Graduate Faculty of North Carolina State University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

**FORESTRY**

Raleigh, NC

2005

Approved by:

Dr. Timothy J. Mullin, Chair of Advisory Committee

Dr. Dag Lindgren, Co-chair of Advisory Committee

Dr. Trudy F. C. Mackay

Dr. Bailian Li

## **DEDICATION**

This work is dedicated to my wife, Marie, whose love, support, and understanding gave me the ability to continue the research that is embodied here. I am grateful to her for believing in me and my work. Thank you, Marie.

## **BIOGRAPHY**

Milan Lstibůrek was born in Pilsen, Czech Republic on June 19, 1977. In 2000, he received a MSc. degree majoring in biology/forest genetics from the Swedish University of Agricultural Sciences in Umeå. In 2001, Milan completed the requirements for his Ing. degree in Forestry from the Czech University of Agriculture, Prague. While finishing those studies Milan worked at the Forestry and Game Management Research Institute of the Czech Republic from September, 2000 until May, 2001. In early 2001, he was accepted to the graduate program at North Carolina State University (NCSU), Department of Forestry. Milan was honored to be a recipient of the Namkoong Family Fellowship in 2004 recognizing his academy achievements and the outcome of his research.

## ACKNOWLEDGEMENTS

I wish to thank Dr. Timothy J. Mullin, my primary advisor, for the support and patience he provided during my studies at NCSU. Dr. Mullin has been an exemplary teacher and mentor, who always found the time to discuss my projects despite the demanding duties he faced as director of the North Carolina State University – Industry Cooperative Tree Improvement Program. The development of Dr. Mullin’s breeding simulator, POPSIM<sup>TM</sup>, has given me great insight into the world of genetics and breeding that no textbook could ever have accomplished.

I also wish to thank, with equal gratitude, the other members of my committee including Dr. Dag Lindgren who has been a never ending source of innovative research ideas and who was a great mentor during my undergraduate studies in Sweden, raising my interest in quantitative genetics. Dr. Trudy Mackay provided valuable advice about underlying genetic models and Dr. Bailian Li encouraged and directed me to think more thoroughly about simulation assumptions and the practical consequences of my studies.

My appreciation also goes out to Dr. Ola Rosvall, Dr. Dudley Huber, Dr. Ross Shepherd, Dr. Jennifer Myszewski, Dr. Fikret Isik, Dr. Floyd Bridgwater, Dr. Gary Hodge, Dr. Arthur Gilmour, Dr. Hossein Jorjani, and Salvador Gezan for their valuable assistance with my research projects. The graduate courses at NCSU were of the highest quality, and I want to offer many thanks to my teachers Dr. David Dickey, Dr. Eugene Eisen, Dr. Gary Hodge, Dr. Jackson Jett, Dr. Bailian Li, Dr. Trudy Mackay, Dr. Steven McKeand, Dr. Timothy Mullin, Dr. Jason Osborne, Dr. Michael Purugganan, Dr. Stephen Roberts, and

Dr. Robert Rodman. I thank Ms. Peggy Arnold, Mrs. Deborah Leistikow, Mrs. Robin Hughes, Mrs. Kathie Zink, Dr. Steven McKeand and Dr. Robert Abt for providing help with formal student issues. I also want to recognize the advice I received from many anonymous people who shared their programming ideas with me at the online Intel Forum.

My interest in forest genetics was initiated during my undergraduate studies at the Czech University of Agriculture in Prague and I want to thank Dr. Jaroslav Koblíha, a member of that University's faculty, who taught my first course in forest genetics and raised my interest in the field. I also acknowledge friendship working environment, while I was an employee at the Forestry and Game Management Research Institute of the Czech Republic.

The dissertation was financially supported by the Föreningen Skogsträdsförädling (Swedish Tree Breeding Association), CellFor Inc., and the North Carolina State University-Industry Cooperative Tree Improvement Program. Their contribution to my studies and research is greatly appreciated. I also thank Genesis Forest Science Canada Inc. for giving me free access to proprietary software.

My studies would not have been feasible without the love and support of my wife Marie and that of our parents and other relatives. Marie and I have been delighted and blessed to have great friends including Paul and Lynn, David and Jana, Daniel, Missy and Kelsey, Cathy, Jason and Jenny, Jennifer and many others. Because of all of them our life abroad was much easier. Thank you all!

# TABLE OF CONTENTS

<b>LIST OF TABLES</b>	<b>viii</b>
<b>LIST OF FIGURES</b>	<b>ix</b>
<b>CHAPTER 1 – Review and summary</b>	<b>1</b>
Introduction .....	2
Positive assortative mating .....	4
Computer modeling of forest tree breeding programs .....	7
Objectives .....	16
Summary of results .....	19
Practical implications .....	21
Suggestions for future research .....	24
References .....	26
<b>CHAPTER 2 – Open-nucleus breeding strategies compared with population-wide positive assortative mating I. Equal distribution of testing effort</b>	<b>32</b>
Abstract .....	33
Introduction .....	34
Materials and methods .....	38
Results .....	42
Discussion .....	44
Conclusion .....	49
Acknowledgements .....	50
References .....	51
<b>CHAPTER 3 - Open-nucleus breeding strategies compared with population-wide positive assortative mating II. Unequal distribution of testing effort</b>	<b>62</b>
Abstract .....	63
Introduction .....	64

Materials and methods .....	67
Results .....	71
Discussion .....	74
Conclusion.....	80
Acknowledgements .....	81
References .....	82

**CHAPTER 4 - Positive assortative mating with family size as a function of parental**

<b>breeding values</b> .....	<b>93</b>
Abstract .....	94
Introduction .....	95
Methods.....	98
Results and discussion.....	104
Acknowledgements .....	112
References .....	113

**CHAPTER 5 – The impacts of positive assortative mating, differential success of somatic embryogenesis, and multi-trait selection on the outcome of clonal**

<b>forestry programs</b> .....	<b>123</b>
Abstract .....	124
Introduction .....	125
Methods.....	127
Results .....	133
Discussion .....	135
Acknowledgements .....	139
References .....	140

# LIST OF TABLES

## CHAPTER 4

1. Description of individual simulation scenarios.....	117
--------------------------------------------------------	-----

## LIST OF FIGURES

### CHAPTER 2

1. Breeding alternatives compared in this study. ....	56
2. Average additive effect, additive variance, and status number in the breeding population.....	57
3. Average inbreeding coefficient and status number in the breeding population.....	58
4. Development of between-group coancestry over five breeding cycles.....	59
5. Average additive effect and status number in the seed orchard.....	60
6. Average additive effect in the seed orchard at variable nucleus sizes .....	61

### CHAPTER 3

1. Breeding alternatives compared in this study .....	86
2. 3:2:1 mating design.....	87
3. Average additive effect, additive variance, inbreeding coefficient, and status number in the breeding population (1 <sup>st</sup> comparison).....	88
4. Average additive effect, additive variance, inbreeding coefficient, and status number in the breeding population (2 <sup>nd</sup> comparison).....	89
5. Average additive effect and status number in the seed orchard (1 <sup>st</sup> comparison)....	90
6. Average additive effect and status number in the seed orchard (2 <sup>nd</sup> comparison)...	91
7. Average additive effect and status number in the seed orchard (3 <sup>rd</sup> comparison) ...	92

### CHAPTER 4

1. Schematic description of the conceptual breeding strategy .....	118
2. Distribution of family sizes in generation five.....	119
3. Breeding population in generation six .....	120
4. Frequency of a major gene in generations .....	121
5. Production population in generation six.....	122

## CHAPTER 5

1. Schematic description of the conceptual breeding strategy .....	144
2. Example of embryogenesis success (family size), as derived from the truncated exponential distribution .....	145
3. Additive genetic response and status number in the production population.....	146
4. Additive genetic response, additive variance, inbreeding coefficient, and status number in the breeding population with variable heritability .....	147
5. Additive genetic response in the production population with variable genetic correlation between traits .....	148
6. Additive genetic response in the production population with variable economic weights on traits.....	149

# **CHAPTER 1**

## **Review and summary**

## **Introduction**

Recurrent selection systems in forest tree breeding involve repeated cycles of selection and mating. In simple mass selection, seed is collected from the best phenotypes in a stand, with no control of pollen source, and used for the regeneration of stands. Since the breeder selects only female parents, the expected gain of this implementation is half of the simple recurrent selection, in which breeder selects both male and female parents and performs controlled crosses. The most common form of simple recurrent selection is based on general combining ability, where the primary source for genetic improvement is the additive genetic variance. Such a selection system is considered throughout this dissertation. Alternatively, when a large component of genetic variance is non-additive, recurrent systems for specific combining ability have been developed and are used primarily in hybrid breeding (Namkoong et al. 1988).

The recurrent selection in this dissertation is generally implemented as follows. Founder parents (the “founder population”) are sampled from a population of unrelated and non-inbred individuals (natural stands). Founder trees are progeny tested and crossed. Progeny tests are established and results are evaluated. All progenies in the test are referred to as the “recruitment population”. Progenies in the test are candidates for selection of individuals used for the next generation of breeding, the “breeding population” (BP). Throughout all generations, the size of the BP is constant and generations are discrete (non-overlapping). In every breeding generation, top-ranking genotypes are selected out of the BP. These are symbolized as the “production population”, (PP) and used for the large-scale deployment of improved genetic material for the reforestation.

This dissertation is organized in five chapters. The first (current) chapter provides an introductory overview of positive assortative mating (PAM), which is the main object of the investigation. The basic methodology implemented in the dissertation was computer simulation, and it is reviewed in the following section of the chapter. The objectives for individual studies are developed next, followed by the summary of main results, their practical implications and suggestions for future research.

The studies included in chapters two and three are devoted to quantitative evaluations of open-nucleus breeding schemes that have been recently incorporated into forest tree breeding programs in different countries throughout the world. Both studies suggest breeding schemes that are genetically more efficient than open-nucleus. While Chapter two deals with comparisons under the balanced case and manifests in a simpler form the underlying principle, imbalance is considered in Chapter three, where more resources are allocated proportionally to the rank of trees in the BP. Chapter three is more appealing to tree breeders, as it describes the conventional implementation in tree breeding programs. In Chapter four, more elaborate assessment of individual tree performance is implemented and utilized in more efficient implementation of PAM under various degrees of imbalance. A strategy is proposed for tree breeders, which further enhances the beneficial effects of PAM. Finally, in Chapter five, a clonal forestry program is considered, where the form of deployment was by somatic embryogenesis. The impact of variation in propagation success on the PP is assessed under the framework of correlated traits and PAM.

## **Positive assortative mating**

Under random mating, individuals are paired with no regard to the phenotypic expression in a particular trait and the expected correlation between mates ( $r$ ) is therefore zero. PAM is delimited by  $0 < r \leq 1$ , while negative assortative (dissortative) mating by  $-1 \leq r < 0$ . The extent to which non-random mating alters genetic properties of a population depends upon the primary cause of preferences among mates. If the sole cause is the environment, there is no genetic mechanism underlying the phenotypic similarities and consequently no genetic change in the population attributable to non-zero  $r$ . If on the other hand selection of mates was driven primarily by similarities reflecting pedigree relationships (as often practiced in breeding of domestic animals), the genetic consequences would be those of inbreeding. Usually though, in commercial breeding programs, mates are paired primarily on the basis of their phenotypic records. The genetic consequence is then due to correlation between breeding values of mates, whose magnitude depends on the narrow-sense heritability ( $h^2$ ), i.e., to what extent similarities among breeding values reflect similarities among phenotypic measurements. Since many traits are correlated, the assortment on one trait may result in assortment on another trait, depending upon the correlation between traits, but it is also possible that mating is random with respect to one trait and assortative with respect to another (Crow 1966; Falconer and Mackay 1996).

The appealing effect of PAM to breeders is the enhancement of additive variance (Crow 1986). Consider a simple quantitative trait, where the variance is due to two biallelic loci of additive effect and equal allelic frequencies ( $p$  and  $q$ ) at both loci are considered (Crow and Kimura 1970). The equilibrium variance following random mating is then  $4pq$ .

Inbreeding would increase the frequency of homozygous genotypes and lead to the equilibrium variance  $8pq$ , double that of random mating. Inbreeding though does not systematically influence the frequency of gamete types. Assortative mating increases the coupling of genes with similar effects ( $A_1B_1$  and  $A_0B_0$ ), i.e., induces gametic-phase disequilibrium (Lynch and Walsh 1998). The equilibrium frequency of genotypes  $A_1A_1B_1B_1$  and  $A_0A_0B_0B_0$  following PAM is  $p$  and  $q$ , respectively and the equilibrium variance is  $16pq$ , much larger than that under random mating or inbreeding. Using a more general approach and assuming  $n$  unlinked loci under continuous PAM, additive variance will asymptotically approach a stable equilibrium, at which the equilibrium variance ( $\hat{V}_A$ ) is

$$\hat{V}_A = \frac{V_{A0}}{1 - r\left(1 - \frac{1}{2n}\right)}$$

(Wright 1921), where  $V_{A0}$  is the additive variance under random mating at gametic-phase equilibrium (before PAM was initiated). An equal expectation holds when loci are linked and with arbitrary effects; the effective number of loci then replaces  $n$  (Crow and Felsenstein 1968). This general formulation shows that relatively large  $r$  and  $h^2$  are prerequisites for any larger enhancement of the variance. Usually, the assortment of mates is not perfect. Under such “partial” PAM, the equilibrium state is far from the case of zero heterozygosity, thus PAM leads usually to a less dramatic increase in homozygosity, compared to inbreeding, specifically when larger numbers of loci influence variation in the trait. Under a majority of practical circumstances, the effect of PAM on the dominance and environmental variance is negligible (Fisher 1918). So far, the theory underlying PAM has not considered epistasis (Lynch and Walsh 1998). The main message is that inbreeding has a greater effect on

homozygosity, while PAM causes a greater enhancement of variance (Li 1955; Wright 1969; Nagylaki 1992).

In the context of a realistic breeding program, the effect of PAM does not stand alone, but is combined with effects of other factors. Based on theoretical assumptions, the expectation is that PAM will increase additive variance and, consequently,  $h^2$ . When selection is practiced, the mean of selected individuals is increased, because population is more variable (Crow 1986). A comprehensive review of empirical and simulation studies dealing with the response to selection under PAM is provided by Jorjani (1995). Although results from analytical and simulation models support the enhancing effect of PAM on the response to selection (review of studies specific to forest-tree breeding is provided in the next section), more variable responses were found in various empirical studies. Jorjani mentions large fluctuations in phenotypic correlations in some studies, when in fact random mating was assumed. This is called “unconscious” assortative mating and is primarily a concern under small population sizes.

## **Computer modeling of forest tree breeding programs**

Computer simulation was used as a primary methodology throughout this dissertation. Simulation in general involves models of systems. Computer simulation refers to methods for studying models by numerical evaluation with software imitating the characteristics of a system. It can be particularly useful when dealing with complex systems, where the components can be identified and quantified, where there is some interdependence among these components, or when a new system is being developed with little knowledge about the possible outcome, thus there is a high investment risk associated with the actual implementation. Simulation can provide valuable insight to the system performance and dynamics (Kelton et al. 2002; Carson 2004).

The simulations used in this dissertation were stochastic in nature, meaning there is a source of randomness in the model. Although a deterministic, pseudo-random number generator is used to provide such random input, the generator used here passed many complex statistical tests for both the uniformity and independence of the output (L'Ecuyer et al. 2002). Stochastic models are useful for representing systems that are too complicated to be expressed analytically. Examples in genetics could be the long-term response to selection under non-random mating, unbalanced mating schemes, or unequal family sizes, all of which were considered throughout this dissertation. The computer simulation software POPSIM<sup>TM</sup> (Mullin and Park 1995) was used as a basis and was further developed to accommodate the objectives of individual studies. To perform the first two studies (Chapters 2 and 3), a model describing an open-nucleus breeding scheme with two levels of hierarchy (tiers) was developed. The model assumed unrestricted migration between tiers in every cycle of

breeding and utilized “group-merit selection”, in which relative importance of both gain and average relatedness for selected group of trees can be declared (Lindgren and Mullin 1997). In the third study (Chapter 4), procedural changes were made to the program structure to accommodate unequal family sizes. Restricted maximum likelihood and best linear unbiased prediction were then incorporated into the model through an external call to the ASReml software (Gilmour et al. 2002). Finally, the mixed-inheritance model (based on the study by Gomez-Raya and Klemetsdal 1999) was added as an alternative to the existing infinitesimal model. The simulation of a single trait was generalized to multiple traits in the last study (Chapter 5). More specific descriptions of changes made to the model are provided in the methodology sections of individual studies, along with the model scope and underlying assumptions. POPSIM<sup>TM</sup> is built on a general-purpose application development platform with mixed-language programming, offering the advantage of flexibility and customization, although greater care was required in the verification phase of development. Since the objective of all studies was to compare alternative breeding scenarios, special attention was given to variance-reduction techniques, such ensuring that the same sequence of random numbers was used for modeling the same processes in comparable simulation runs, avoiding unnecessary randomness in the output (L’Ecuyer 1994). To construct the model range of accuracy, standard univariate statistical techniques were used to obtain confidence intervals for each output statistic (average of simulation iterations). The number of simulation replicates (iterations) in each study was predicted from sample variances of selected output variables, the level of statistical significance and the desired lengths of confidence intervals (Law and Kelton 2000). The breeding program for Norway spruce in Sweden described and

modeled by Rosvall et al. (1998) served as a conceptual breeding strategy throughout this dissertation, although the focus has been placed only on a single 48-tree BP. In reality, there are a larger number of such functional populations, geographically distributed along main environmental gradients in Sweden. Generation advancement is by balanced, within-family selection, which minimizes the loss of gene diversity or, equivalently, an increase in group coancestry (Hill et al. 1996; Lindgren et al. 1996). Concentrating effort on the recruitment population enables substantial selection differentials from selection within families that are further facilitated by the use of clonally replicated testing. This results in attractive genetic gains in the PP available for today's reforestation, while more conservative management of the BP secures long-term response under uncertain environmental conditions.

In forest-tree breeding, simulating the long-term impact of population management is particularly useful, because generation intervals can be extremely long, and so can be the influence of poor decisions (when such are based on guessing and "rules of thumb"). A number of simulation studies investigating forest tree breeding programs have been published. Some of them are focused on a specific problem, while others deal with more general questions. Simulation of breeding is based on two broad types of models, "parametric" and "finite locus". In parametric models, the value of a tree is sampled from a distribution of effects. A large (strictly infinite) number of loci, each of small effect, are assumed to influence the variation in quantitative traits. Parametric models offer the advantage of relative simplicity due to well-established theory underlying the inheritance of polygenic traits. Unlike parametric models, finite-locus models imply a finite number of loci that are individually characterized in the computer program. These models offer more

realistic evaluation of a specific genetic architecture; however a larger number of parameters is involved in the model (locus position, number of alleles, their frequencies, genotypic effects). Finally, a mixed-inheritance model is a combination of both approaches. Examples of stochastic simulation studies dealing with forest tree breeding programs are provided below, classified according to the modeling approach used. Some examples of studies based on deterministic approaches (where prediction of response to selection is described algebraically) are those of Wei et al. (1998); Wei and Lindgren (2001); Appiolaza and Garrick (2001); and Danusevicius and Lindgren (2004).

### **Parametric (infinitesimal) models**

King and Johnson (1993) performed a comparison of five different mating schemes and reported both genetic gain and effective population size over five breeding cycles. Random mating resulted in the largest gains per unit effective population size. Applying unbalanced designs created only a slight advantage over completely balanced schemes with random mating.

The decision-support software tool POPSIM<sup>TM</sup> was presented and described by Mullin and Park (1995). The software was flexible to accommodate various inputs specifying the population sizes, initial variances, mating schemes and forms of deployment. An example was provided along with the software description.

The effects of various population sizes, mating designs and selection strategies on status number, inbreeding coefficient and genetic gain were studied by Gea et al. (1997), using POPSIM<sup>TM</sup>. They found that if population size is too small, inbreeding may become

too severe, thus larger population sizes are more suitable for long-term breeding. Larger populations also provided greater genetic response, although they were slightly less efficient in preserving status number compared to smaller populations.

A comparison between phenotypic and combined-index selection was performed by Andersson et al. (1998) with a later version of POPSIM<sup>TM</sup>. When comparing genetic response at the same effective population size, selection based on phenotype resulted in gains similar to those under combined-index selection. This was true when  $h^2$  was high or when it was low and family sizes were small. It was also found that phenotypic selection preserves additive variance to a greater extent.

The robustness and efficiency of a long-term tree breeding strategy for Norway spruce in Sweden was investigated by Rosvall et al. (1998), mimicking closely a realistic management plan. POPSIM<sup>TM</sup> has been further developed to accommodate the objective of the study. It was found that balanced within-family selection is suitable for maintenance of long-term gains in the BP, while more gain can be achieved in the PP, primarily due to higher efficiency of clonal testing. The breeding program was found to be sustainable in providing long-term response.

Breeding based on discrete generations was compared with rolling-front breeding strategies (overlapping generations) over a period of 40 years (Borralho and Dutkowski 1998). It was concluded that the rolling-front strategy resulted in 25-35% greater additive genetic response per year, primarily due to shortening the generation intervals.

The influence of a number of parental combinations on reselection in seed orchards was investigated by Johnson (1998), focusing on a Douglas-fir breeding program in the

Pacific Northwest. A small number of crosses per mate (three) seemed sufficient to provide enough information for reselecting in the orchard and for reducing the variation of gain estimates.

Jayawickrama and Jefferson (1999) described software for simulating multiple traits utilizing a sequential culling method. The software considered only one cycle of breeding and did not include a calculation of inbreeding.

Rosvall and Andersson (1999) analyzed the effect of group-merit selection with POPSIM<sup>TM</sup>. Group-merit selection was found to be always more efficient than conventional restricted selection, except at the two theoretical limits, where both methods provided the same outcome. The efficiency of group-merit selection increased under low  $h^2$ . Higher weighting on group-coancestry was recommended when the variation among families is low.

Thirty generations of selection were modeled by Rodríguez (2000). Random mating was followed by balanced within-family selection, which was compared to selection with slight imbalance (unbalanced parental contributions to the next generation). Results showed that when genetic diversity is given a high priority, scenarios with slight imbalance may provide larger genetic gains than completely balanced selection scenarios.

Myszewski (2003) performed a comparison of selection and breeding strategies to incorporate wood properties into loblolly pine breeding program. Correlated response to selection was simulated in mainline and elite populations. Within the framework of studied parameters, the selection for high growth resulted in lower wood quality in the mainline population, and when wood quality was selected in the elite population, a decrease in growth

was observed. Adjustments were made to the selection index were alternatively used and it was demonstrated that a decrease in the value of the secondary trait could be avoided.

The effect of positive assortative mating was evaluated in detail by Rosvall and Mullin (2003), considering both genetic gain and diversity during selection. Positive assortative mating resulted in substantial increase of additive variance in the BP and consequently larger genetic gains in the PP. Although the effective population size was not affected, positive assortative mating resulted in greater inbreeding. These results were extended for the case of unequal parental contributions by Rosvall et al. (2003). Unequal parental contributions generated additional enhancement of additive variance in the BP over the case with equal contributions and further affected inbreeding, as well as effective population size. It was suggested that unbalanced mating followed by balanced within-family selection is a more suitable strategy to support seed orchards than balanced mating followed by unbalanced selection.

### **Finite-locus models**

Mahalovich (1990) studied the effect of positive assortative mating in recurrent selection programs. Both main-line and elite populations were considered. Positive assortative mating resulted in 1-12% larger gains than random mating when an additive model was assumed. Inbreeding varied, depending on the size of the elite population. Less consistent results were obtained under partial-dominance model.

Bridgwater et al. (1993) investigated the effects of sublining to manage inbreeding. It was concluded that random or disassortative assignment of founders to sublines will result in

greater variation within sublimes, thus contributing to larger genetic gains. This allelic model was adopted to support the decision-making in the planning of a strategy for third breeding cycle of loblolly pine in the Southeastern USA (McKeand and Bridgwater 1998).

Purging deleterious genes under different breeding schemes was compared by Fu et al. (1998). Progeny survival and the probability of extinction were studied under a multiplicative fitness model. Continuous half-sib mating was found to be as effective as selfing in purging deleterious alleles, but under a lower risk of population extinction (this is particularly important in programs developed for endangered species). Fu (1999) extended the scope of the study by considering a synergistic fitness model. Purging was increased under high genetic load and low inbreeding, high dominance and strong synergism (i.e., greater effect of inbreeding on fitness than was achieved under the multiplicative fitness model).

Kerr et al. (2004a) developed a computer model, which enables comparison of programs targeted at hybrid forest tree species. Additive, dominance and epistatic gene action is considered and the model is very flexible to accept characterization of a specific genetic architecture as an input. Although the theory underlying hybridization is more complex than that for single-species breeding, the model is a valuable tool to evaluate hybrid tree improvement programs. A comparison among four specific hybrid tree breeding strategies using this model was performed by Kerr et al. (2004b).

### **Mixed-inheritance models**

The effect on genetic response of marker-assisted within-family selection in radiata pine was evaluated by Kumar and Garrick (2001). The model considered a QTL linked to a marker

locus and a large number of genes, each with small effect. In the deployment of full-sib families, marker-aided selection provided an additional 4-8% genetic gain and in the deployment of clones, the extra advantage was 2-3%, although considerable more time was required for the evaluation of each family.

## Objectives

The efficient application of positive assortative mating by tree breeders requires assessment of alternative population management strategies that attempt to optimize program deliverables to forest plantations over the long term. The specific objective of the first two studies was to explore quantitatively open-nucleus breeding schemes in the framework of a forest tree breeding program. Open-nucleus schemes were first developed by animal breeders (Roden 1994) and later incorporated into forestry programs in various countries throughout the world (references are given in Chapter 2). The two-tier hierarchical assortment of the BP into elite and main subpopulations in the basic nucleus scheme may enhance the additive genetic response and variance in the BP (compared to a population managed without such assortment), thus increasing the potential for selecting more superior genotypes for the PP. However, as Kinghorn et al. (2000) suggested, assortment across the whole BP (population-wide PAM) could be genetically more efficient than open-nucleus breeding scheme, since in PAM, each parental cross comprises a distinct hierarchical level.

The aim of the first study (presented in Chapter 2) was to quantitatively compare a two-tier open-nucleus breeding scheme with population-wide PAM in a long-term forest tree breeding program. A simple scheme with balanced distribution of test resources was considered as a starting point. The balanced distribution means that each mate was involved in an equal number of parental combinations, and an equal number of progenies was generated for each family in both open-nucleus and PAM breeding strategies.

The next logical step was to perform a similar comparison, but under unbalanced distribution of test resources, where more effort could be allocated to parents of higher rank.

This provided a more realistic comparison, since breeders usually suggest incorporation of open-nucleus schemes with more effort concentrated to the “elite” part of the population, at the expense of less effort directed to the remaining part of the population (Cotterill 1989). The important point [which was quantitatively demonstrated by Rosvall et al. (2003)] is that the effect of population-wide PAM could also be substantially enhanced if more effort is concentrated to higher-ranking mates. Thus, the objective of the second study (presented in Chapter 3) was to compare population-wide PAM with an open-nucleus breeding strategy under unbalanced distribution of test effort, with more resources allocated to trees of higher rank.

The objective of the study that follows in Chapter 4 was developed based on the main conclusions derived out of the first two studies, that population-wide PAM is genetically more superior to open-nucleus schemes and that the favorable genetic effect of parental assortment on the BP, and especially the PP, which is of greatest interest to the tree breeder, can be greatly enhanced by the redistribution of test resources. Although the second study showed that PAM with unbalanced distribution of test effort generated significantly larger gains in the BP and PP at any target level of diversity compared with open-nucleus breeding schemes, the redistribution of resources among mates (as done in Chapter 3) was not performed ideally, as it was based on simplified schemes not fully utilizing the potential of mated trees. The objective of the study presented in Chapter 4 was therefore to develop methodology for more optimal reallocation of resources among mates, directly being a function of their predicted breeding values. The secondary objective was to utilize mixed-inheritance model methodology to investigate whether the presence of a single major-gene

locus (symbolizing a QTL contributing to the variation in a quantitative trait), would alter significantly output from the BP or PP, utilizing the mixed-inheritance model methodology.

In the last study (presented in Chapter 5), the simulation of response to positive assortative mating was extended to correlated traits. A tree breeding program was assumed, where somatic embryogenesis was incorporated for the deployment of top-ranking genotypes. The primary objective was to assess the influence of the variation in success of embryogenesis among individual lines on the output of the program. The joint influence of the assortment of mates, genetic correlation,  $h^2$  and economic importance of traits was evaluated.

## **Summary of results**

The first study (Chapter 2) revealed that under completely balanced situations, PAM on a population-wide basis provides higher genetic response in the PP than population management with only two distinct levels of hierarchy, open-nucleus (NB). This is because PAM is more efficient in enhancing the additive genetic response and variance in the BP, which generates more superior selection candidates for the PP. This conclusion was true over the entire range of target PP diversity, studied over five breeding generations. Comparison between NB and PAM were performed, such that the size of the nucleus population (elite individuals in BP) in the NB strategy was optimized to provide maximum genetic response in the PP. The optimum size of the nucleus population varied with the desired level of PP diversity; much larger sizes were optimum when desired diversity was high.

The same conclusion holds when the distribution of test effort is unbalanced, with more resources allocated to higher-ranking mates at the expense of the remaining part of the population, as revealed in the second study (Chapter 3). Both PAM and NB strategies were enhanced by the redistribution of test effort and provided greater genetic response in the PP (compared to balanced scenarios in the first study), but PAM was more efficient in generating additional PP gains compared to NB strategies. In this study though, the imposed imbalance in the PAM strategy was done based on parental ranks, rather than actual predicted breeding values.

In the third study (Chapter 4), restricted maximum-likelihood estimation and best linear unbiased prediction (BLUP) methodologies were incorporated to the simulation. It was then possible to deal efficiently with any unbalanced situations, as required by the study

objective. The test resources were redirected by varying the number of progenies (family sizes) as a linear function of mid-parent BLUP values under a simple balanced mating design (single-pair mating). In the developed methodology, the actual distribution of mid-parent BLUP values was standardized by a constant value, which was varied in simulation scenarios to cover the entire range of the distribution of family sizes between its theoretical limits. Interestingly, the simulation revealed that constraining to an equal number of progenies selected from each family (balanced within-family selection), while maximizing the variation in family sizes, caused only a little change in the effective population size and average inbreeding in BP compared to case where family sizes were constant. Under such favorable conditions, the redistribution of resources increased significantly the effect of PAM on the additive genetic response and variance in the BP and provided large additional gains in the PP. The mixed-inheritance model was also incorporated in this study and it was verified that these conclusions hold when a major gene locus contributes to the variation in a quantitative trait, in addition to the variation explained by a large number of genes with small effects.

In the last study (Chapter 5), scenarios with unequal embryogenic success among lines resulted in non-significant differences in genetic response of selected clonal mixtures (PP) compared to scenarios where embryogenesis in every line was equally successful. These conclusions were verified over a range of input parameters, including correlation,  $h^2$  and economic importance of simulated traits. Higher  $h^2$  enhanced the effect of PAM, resulting in larger genetic responses in the BP and PP, and a relative reduction in the effective population size of the PP. The effective size and inbreeding in the PP were not significantly affected by genetic correlation and economic importance of traits.

## **Practical implications**

Breeders are encouraged to use quantitative models during the design and evaluation of breeding programs, rather than relying on subjective reasoning. Decisions regarding population structuring and management will fundamentally influence the amount of genetic improvement and the diversity carried forward in the reproductive material, which has a tremendous impact on the socioeconomic value of future forest ecosystems.

As demonstrated in the first two studies, there is a genetic advantage of population-wide PAM over breeding strategies with open-nucleus. Given this theoretical conclusion, more site-specific analysis could be performed involving factors like timing, relative cost of breeding operations or species reproductive biology. The computer model presented is flexible enough to provide such a detailed assessment for any particular program. Furthermore, redirecting more test effort to higher-ranking mates in either PAM or open-nucleus strategies may provide additional genetic response in the PP at little or no loss of genetic diversity. The exact implementation of such redirection will again be best assessed on a program-specific basis.

The main reasons for implementing open-nucleus breeding schemes in animal breeding are probably not applicable to forest trees, where it seems more practical to perform simultaneous evaluation of all selection candidates and where implementation of population-wide PAM is practical. The effects of PAM in forest tree breeding could be further enhanced by larger family sizes and the use of clonal replication. Animal breeding textbooks (see Bourdon 1997) consider nucleus schemes more as alternatives to large-scale genetic

evaluations. It is recommended that tree breeders consider population-wide PAM as a more powerful alternative to nucleus breeding schemes.

The third study provides a methodology for redirecting test resources efficiently, according to actual predicted breeding values and the theoretical advantage of such implementation is demonstrated in detail. Forest tree breeders are encouraged to consider this proposed strategy under practical conditions. If PAM is feasible under practical settings and the genetic response is primarily due to selection within families [as described for Norway spruce in Sweden by Rosvall et al. (1998)], spending an equal amount of resources on each family seems very inefficient considering the additional gains that can accumulate over generations as a result of imbalance. The addition of imbalance can be performed very simply, by maximizing the variation in family size (proportional to mid-parent BLUP values), while maintaining equal parental contributions. This will lead to only negligible reduction of effective population size and negligible increase in average inbreeding coefficient in the BP, but leading to substantial augmentation of gains in the PP, exceeding 10%. This can mean a tremendous added value to the forest industry and society.

Finally, the main conclusion of the last study regarding no significant impact of variation in embryogenic success on direct and correlated responses in selected clonal mixtures is of immediate interest to those in the forest industry who wish to integrate somatic embryogenesis deployment as a system to maximize gain from each cycle of breeding. Further progress in somatic embryogenesis may lead to the normalization of responses among genotypes. One may then intuitively allocate higher testing effort to families of higher expected value. Our results indicate that this would not necessarily lead to higher genetic

gains in deployed clonal mixtures, because the majority of genetic variation in elite crosses is available within families and the impact of such reallocation was only marginal.

## **Suggestions for future research**

There are further questions that were raised while working on this dissertation and that should be addressed in future research. The breeding strategy advocated in the third study should be tested under a wider range of constraints on relatedness during the selection, in order to test how this algorithm behaves between extremes: where the genetic value is the only component of the selection criterion on one hand, and where the average relatedness of a selected group of trees is given maximum importance on the other.

In this dissertation, it was assumed that the rank of each tree (which was used as the basis for assortment prior to mating) is determined purely by its predicted breeding value. Assuming balanced, within-family selection, one could incorporate a penalty to the pair-wise coancestry of mated trees, which would influence the average inbreeding in the following generation's BP. Some point of balance might be achieved, in which the assortative mating still provides a desirable effect, while some avoidance of mating among relatives causes lower reduction of within-family portion of additive variance and lower inbreeding depression in the generation that follows.

The mixed-inheritance model should also be explored more systematically as a basis for studying more specifically the incorporation of marker-aided selection into classical tree-breeding programs. Various theoretical models combining information on molecular markers (linking to a QTL) with phenotypic observations can then be tested.

Only a simplified model of genetic correlation was implemented in this study, under the infinitesimal model. A detailed study incorporating a finite locus model would also be suitable to test the hypothesis in Chapter 5. In many aspects, there is great potential to

perform a number of studies exploiting the possibilities offered by the incorporation of genetic evaluation methods to the simulation model.

## References

- Andersson EW, Spanos KA, Mullin TJ, Lindgren D (1998) Phenotypic selection compared to restricted combined index selection for many generations. *Silvae Fennica* 32:111-120
- Apiolaza LA and Garrick DJ (2001) Breeding objectives for three silvicultural regimes of radiata pine. *Can J For Res* 31:654-662
- Borrvalho MG and Dutkowski GW (1998) Comparison of rolling front and discrete generation breeding strategies for trees. *Can J For Res* 28:987-993
- Bourdon RM (1997) *Understanding animal breeding*. Prentice Hall, Upper Saddle River
- Bridgwater FE, Woodbridge WC and Mahalovich MF (1993) Computer modeling of a sublining breeding system. *Proc South For Tree Improv Conf* 22:327-333
- Carson JS (2004) Introduction to modeling and simulation. *Proceedings of the 2004 Winter Simulation Conference*, R. G. Ingalls et al. (eds.), pp. 9-16
- Cotterill PP (1989) The nucleus breeding system. *Proc South For Tree Improv Conf* 20:36-42
- Crow JF (1966) *Genetics notes*. Burgess Publishing Company, Minneapolis
- Crow JF (1986) *Basic concepts in population, quantitative, and evolutionary genetics*. W. H. Freeman and Company, New York
- Crow JF and Felsenstein (1968) The effect of assortative mating on the genetic composition of a population. *Eugenics Quart* 15:85-97
- Crow JF and Kimura M (1970) *An introduction to population genetics theory*. Burgess Publishing Company, Minneapolis

- Danusevicius D and Lindgren D (2004) Progeny testing preceded by phenotypic pre-selection - Timing considerations. *Silvae Genet* 53:20-26
- Falconer DS and Mackay TFC (1996) Introduction to quantitative genetics. Longman, Harlow
- Fisher RA (1918) The correlation between relatives on the supposition of Mendelian inheritance. *Trans Roy Soc Edinb* 52:399-433
- Fu YB, Namkoong G and Carlson JE (1998) Comparison of breeding strategies for purging inbreeding depression via simulation. *Conservation-Biology* 12:856-864
- Fu YB (1999) Patterns of the purging of deleterious genes with synergistic interactions in different breeding schemes. *Theor Appl Genet* 98:337-346
- Gea LD, Lindgren D, Shelbourne CJA, Mullin TJ (1997) Complementing inbreeding coefficient information with status number: implications for structuring breeding populations *NZ J For Sci* 27:255-271
- Gilmour AR, Gogel BJ, Cullis BR, Welham SJ and Thompson R (2002) ASReml User Guide Release 1.0 VSN International Ltd, Hemel Hempstead, HP11ES, UK
- Gomez-Raya L and Klemetsdal G (1999) Two-stage selection strategies utilizing marker-quantitative trait locus information and individual performance. *J Anim Sci* 77:2008-2018
- Hill WG, Caballero A and Dempfle L (1996) Prediction of response to selection within families. *Genet Sel Evol* 28:379-383
- Jayawickrama KJS and Jefferson PA (1999) Stochastic simulation of genetic advancement of multiple traits using sequential culling. *NZ J For Sci* 29:25-43

- Johnson GR (1998) Parental GCA testing: How many crosses per parent? *Can J For Res* 28:540-545
- Jorjani H (1995) Genetic studies of assortative mating in selected and unselected populations. PhD dissertation. Swedish University of Agricultural Sciences, Uppsala
- Kelton DW, Sadowski RP, Sadowski DA (2002) *Simulation with Arena*. McGraw-Hill, New York
- Kerr RJ, Dieters MJ, Tier B and Dungey HS (2004a) Simulation of hybrid forest tree breeding strategies. *Can J For Res* 34:195-208
- Kerr RJ, Dieters MJ, and Tier B (2004b) Simulation of the comparative gains from four different hybrid tree breeding strategies. *Can J For Res* 34:209-220
- King JN, Johnson GR (1993) Monte Carlo simulation models of breeding-population advancement. *Silvae Genet* 42:68-78
- Kinghorn B, Werf J, Ryan M (2000) *Animal breeding: use of new technologies: a textbook for consultants, farmers, teachers and for students of animal breeding*. University of Sydney, Sydney
- Kumar S and Garrick DJ (2001) Genetic response to within-family selection using molecular markers in some radiata pine breeding schemes. *Can J For Res* 31:779-785
- Law AM and Kelton WD (2000) *Simulation modeling and analysis*. McGraw-Hill, New York
- L'Ecuyer P (1994) Efficiency improvement and variance reduction. *Proceedings of the 1994 Winter Simulation Conference*, J. D. Tew et al. (eds.), pp. 122-132

- L'Ecuyer P, Simard R, Chen EJ and Kelton WD (2002) An objected-oriented random-number package with many long streams and substreams. *Operations Research* 50:1073-1075
- Li CC (1955) *Population Genetics*. The University of Chicago Press, Chicago
- Lindgren D, Gea LD and Jefferson PA (1996) Loss of genetic diversity monitored by status number. *Silvae Genet* 45:52-59
- Lindgren D and Mullin TJ (1997) Balancing gain and relatedness in selection. *Silvae Genet* 46:124-129
- Lynch M and Walsh B (1998) *Genetics and analysis of quantitative traits*. Sinauer Assocs., Inc., Sunderland, Massachusetts
- Mahalovich MF (1990) Modeling positive assortative mating and elite populations in recurrent selection programs for general combining ability. PhD dissertation. North Carolina State University
- McKeand SE and Bridgwater FE (1998) A strategy for the third breeding cycle of loblolly pine in southern U.S. *Silvae Genet* 47:223-234
- Mullin TJ, Park YS (1995) Stochastic simulation of population management strategies for tree breeding: a new decision-support tool for personal computers. *Silvae Genet* 44:132-141
- Myszewski JH (2003) A comparison of selection and breeding strategies for incorporating wood properties into a loblolly pine (*Pinus taeda* L.) elite population breeding program. PhD dissertation. Texas A&M University
- Namkoong G (1988) *Tree breeding: Principles and strategies*. Springer-Verlag, New York

- Nagylaki T (1992) Introduction to theoretical population genetics. Springer-Verlag, New York
- Roden JA (1994) Review of the theory of open nucleus breeding systems. Anim Breed Abstr 62:151–157
- Rodríguez LS (2000) Balanced vs. slightly unbalanced selection. Heredity 84:685-691
- Rosvall O, Lindgren D, Mullin TJ (1998) Sustainability robustness and efficiency of a multi-generation breeding strategy based on within-family clonal selection. Silvae Genet 47:307-321
- Rosvall O, Andersson EW (1999) Group-merit selection compared to conventional restricted selection for trade-offs between genetic gain and diversity. For Genet 6:1-14
- Rosvall O, Mullin TJ (2003) Positive assortative mating with selection restrictions on group coancestry enhances gain while conserving genetic diversity in long-term forest tree breeding. Theor Appl Genet 107:629-642
- Rosvall O, Mullin TJ, Lindgren D (2003) Controlling parent contributions during positive assortative mating and selection increases gain in long-term forest tree breeding. For Genet 10:35–53
- Wei RP, Hansen CR, Dhir NK and Yeh FC (1998) Genetic gain with desired status number in breeding programs: a study on selection effects. Can J For Res 28:1861-1869
- Wei RP and Lindgren D (2001) Optimum breeding generation interval considering buildup of relatedness. Can J For Res 31:722-729
- Wright S (1921) Systems of mating. III. Assortative mating based on somatic resemblance. Genetics 6:144-161

Wright S (1969) Evolution and the genetics of populations. Volume 2, The theory of gene frequencies. The University of Chicago Press, Chicago

## **CHAPTER 2**

### **Open-nucleus breeding strategies compared with population-wide positive assortative mating**

#### **I. Equal distribution of testing effort**

This article was originally published in Theoretical and Applied Genetics.

Authors: M. Lstibůrek, T. J. Mullin, D. Lindgren, O. Rosvall

Volume: 109

Page Numbers: 1196-1203

Year of Publication: 2004

The article is reproduced with a written permission from the publisher.

© Springer-Verlag Berlin Heidelberg

## **Abstract**

Positive assortative mating (PAM) can enhance the additive genetic variance in a breeding population (BP). This increases the potential for gains in the production population (PP, selected subset of the BP) for recurrent selection programs in forest trees. The assortment of mates can be either: (1) by individual tree rank across the whole BP (PAM), or (2) trees of similar rank can be merged into larger hierarchical groups and then mated randomly within group ("open"-nucleus breeding, NB). The objective of this study was to compare PAM and NB in quantitative terms. The NB simulation model assumed two tiers (nucleus, main) with unrestricted migration between the tiers. Clonal tests were used to predict breeding values and test resources per mate were kept constant for all mates. Both gain and diversity were combined into a single selection criterion, "group-merit selection". Alternatives were compared over five breeding cycles by considering genetic gain and diversity in a selected PP established in a seed orchard. The assortment of mates in both alternatives enhanced additive variance and increased the additive effect in the BP, leading to additional gain in the PP. Gains generated under PAM always exceeded gains under NB. Thus, the main message from this study is that PAM in both the short- and long-term results in more gain at any target level of diversity in the PP (the breeder's target) than is achieved by the NB alternative. The optimum size of the nucleus varies with the desired level of seed orchard diversity. At lower target diversity, smaller nucleus sizes are favorable, while larger sizes result in more gain when seed orchard diversity is considered more important.

## Introduction

Two basic forms of mating can be described in terms of the correlation between phenotypic values of mated individuals. One is random mating (RM), where the expected correlation is zero, and the other is non-RM, where the correlation differs from zero. When the correlation is positive, mating is positive assortative, and when the correlation is negative, mating is negative assortative (Crow and Kimura 1970; Jorjani 1995). The advantage of positive assortative mating (PAM) for practical breeding is the potential for expansion of the additive variance (Breese 1956). This expansion is mainly influenced by the phenotypic correlation of mates,  $h^2$  and the effective number of loci affecting a trait of interest (Lynch and Walsh 1998). The actual benefit of assortative mating for practical breeding programs under selection is complex to evaluate. Computer simulation can be used to combine simultaneously quantitative models with other important factors. For example, Rosvall and Mullin (2003) evaluated the effect of PAM in long-term forest tree breeding programs. They found a large expansion of additive variance, which was attributable to the effects of PAM and restricted among-family selection. The increased additive variance enhanced gains in deployed genetic material (forest plantations) derived from a selected seed orchard production population (PP), while conserving genetic diversity in the breeding population (BP).

Nucleus breeding (NB) can be described as a form of PAM where selected individuals are allocated into a few distinct hierarchical levels (tiers) prior to mating within tiers. The nucleus tier (also known as the elite population) is the uppermost tier and can either be closed (no gene flow into the nucleus) or open [the gene pool in the nucleus is periodically enriched

by migration from lower tier(s) or external populations]. Allocation to the different tiers is done after ranking individuals for breeding values, so it can lead to an expansion of additive variance. The concept of NB was originally developed and used in animal breeding programs. The first open-NB schemes for sheep were developed in the mid 1960s in Australia and New Zealand. The concept became popular in these two countries in the late 1960s and the 1970s. A detailed review of the development is provided by del-Bosque González (1989) and Roden (1994).

A number of simulation models have been developed to investigate the optimum tier structure under NB schemes. James (1977) introduced the first general open-NB model. Genetic progress was expressed deterministically as a function of selection differentials, migration rates and genetic lag between tiers. Among the main findings was that a constant equilibrium in genetic differentiation between tiers and annual gain is achieved after repeated cycles of selection. At the equilibrium, the rate of gain in both tiers is equal and depends only on selection and migration. James (1977) concluded that migration in an open-nucleus system reduces the rate of inbreeding to approximately one half of that in a closed-nucleus system (when the size of nucleus is small). James (1978) derived a simple formula for effective population size with an open-NB system and Hopkins and James (1978) introduced a modified model for overlapping generations. All of these models assumed constant genetic variances. Later, Mueller and James (1983) improved the model and showed that the rate of genetic gain could be overestimated if the loss of variance due to selection (Bulmer 1971) was not considered. Shepherd (1991) and Shepherd and Kinghorn (1992) investigated schemes with multiple-tiers and proposed examining NB schemes as "structured" assortative

mating. They concluded that additional genetic gain was generated with the addition of extra tiers in the open-nucleus system due to the introduction of additional between-tier assortative mating. Roden (1995) compared three alternative schemes and found that the mean annual gain was highest and the rate of inbreeding lowest with the open-nucleus system.

The NB concept was introduced into forestry by Cotterill (1989) and incorporated into breeding plans for radiata pine (*Pinus radiata* D. Don) in Australia (White et al. 1999) and bluegum eucalypt (*Eucalyptus globulus* Labill.) in Portugal (Cotterill et al. 1989). Cotterill et al. (1989) and Cotterill (1989) summarized some advantages of NB over classic schemes in tree improvement, including the potential gains in the nucleus population as a result of concentrating elite breeding individuals in the nucleus, so that the majority of investment is on outstanding genetic material, and the versatility of the nucleus in terms of the potential transfer of selections from outside the breeding program. The latter concept fits neatly within the structure of breeding cooperatives in forestry, where NB has been considered in Finland (Mikola 2002), the Southeastern United States (White et al. 1993; McKeand and Bridgwater 1998), Denmark (Wellendorf et al. 1994) and South Africa (Hagedorn 1991).

Few simulation studies have investigated the NB concept specifically in forest trees. Most of these have focused on the BP as the primary target, rather than the PP (e.g., seed orchard). Mahalovich and Bridgwater (1989) investigated the advantage of an elite population for forest tree breeding and found that gains in the random-mating elite population exceeded those in the main population by 4–45% over 12 generations. In an alternate scenario, the elite population was periodically enriched by replacing the poorest trees with the best trees from the main population to offset inbreeding in the nucleus.

Inbreeding was reduced, but gain in the elite population was not greater than without enrichment. This was explained as due to low narrow-sense heritability; and progeny testing of replacements was suggested. King and Johnson (1993) compared an open-nucleus scheme with four other mating schemes. Making additional crosses among elite genotypes in each generation offered more potential for genetic gain, but the effective population size was reduced as more selections shared the same parents.

Kinghorn et al. (2000) suggested that assortment across the whole BP (PAM) could be genetically more efficient than the open-nucleus system. It therefore seems interesting to compare in detail a two-tiered open-nucleus system with PAM across the whole BP (where each parental cross comprises a distinct hierarchical level). The aim of the current study was to quantitatively compare open-NB to population-wide PAM in a long-term program with recurrent selection for general combining ability in forest trees. Situations with a balanced distribution of testing resources are considered in this paper; and more complex scenarios follow in a companion paper (part II, this issue). Under balanced testing efforts, the hypothesis is that open-NB leads to a lower expansion of additive variance in the BP and thus to lower genetic gains in the PP than is the case with population-wide PAM.

## Materials and Methods

Three breeding alternatives were compared in this study (Figure 1). Under NB, the BP was structured into two distinct tiers, the nucleus (elite) and main (base). Under the population-wide PAM and non-hierarchic RM alternatives, the BP was not divided into tiers. In PAM, parental assortment was done at the individual-tree level and applied across the entire BP. Mates were paired at random in the RM alternative. RM in this study means that mates were randomized, but duplicate and self-crosses were excluded.

A discrete-generation model was developed for stochastic simulation of an open-NB system relevant for forest trees. Each scenario (unique set of input parameters) was replicated by 800 independent runs (iterations). Parameter means across all iterations were calculated for each scenario and presented along with 95% confidence intervals (CI). Generation intervals were assumed to be constant in all scenarios.

The breeding simulation software "POPSIM", developed by Mullin and Park (1995), served as the basis for the stochastic simulation. The model structures and parameters were initially those used by Rosvall et al. (1999) to model the current Norway spruce [*Picea abies* (L.) Karst.] breeding program in Sweden, which features the use of clonal testing, although much of their study is relevant to a broader range of species and programs. Both Mullin and Park (1995) and Rosvall et al. (1999) discuss underlying assumptions of the simulation model, the majority of which are relevant to this study.

A founder population of 48 unrelated, non-inbred trees ( $N = 48$ ) was constructed by random sampling from normal distributions of genetic and environmental effects. A single polygenic trait in a diploid organism was assumed (although this trait may be a complex

index of many observations). The trait was under the influence of a large number of autosomal loci, each with a small effect (infinitesimal model). Initial variance components were set to approximate growth traits in conifer species according to Rosvall et al. (1999): additive variance ( $V_A$ ) = 100, dominance variance ( $V_D$ ) = 25 and environmental variance ( $V_E$ ) = 375 (giving  $h^2 = 0.2$ ) with a phenotypic mean = 100. In the NB alternative, the nucleus was initialized by sampling  $N_N$  best phenotypes from the founder population. The remaining trees ( $48 - N_N$ ) were allocated to the main tier. The nucleus size was varied by increments of eight trees ( $N_N = 8, 16, 24, \dots$ ). Single-pair RM was used within both tiers to generate a recruitment population of 24 full-sib families,  $N_N/2$  of which were generated in the nucleus. A total testing effort of 2,400 seedlings (genotypes) was assumed to be distributed uniformly in the recruitment population, i.e., equal family size. Each seedling was clonally replicated by ten ramets. Clonal replication of genetic tests is current practice for Norway spruce in Sweden (Karlsson and Rosvall 1993) and has been recommended for applications in other tree breeding programs (e.g., Danusevicius and Lindgren 2002; Isik et al. 2003).

The phenotypic value of each tree was the sum of independent additive, dominance and environmental effects. Epistatic interactions and maternal effects were not considered. The mean additive effect for each full-sib family  $a_{FS}$  was calculated as the mid-parent additive effect. The additive effect for each offspring within each family was then generated by random sampling from Normal  $\{a_{FS}, 0.5V_A [1 - 0.5(F_f + F_m)]\}$ , accounting for the expected reduction of within-family additive variance due to inbreeding of the female  $F_f$  and male  $F_m$  parents, respectively. The mean dominance effect for each full-sib family  $d_{FS}$  was drawn randomly from Normal  $(0, 0.25V_D)$ . The dominance effect for each offspring within each

family was then generated by random sampling from Normal  $\{d_{FS}, 0.75V_D [1 - 0.5(F_f + F_m)]\}$ . Inbreeding depression was not considered. The environmental effect for a clonal copy (ramet) of each genotype within each family was drawn randomly from Normal  $(0, V_E)$ ; and it was assumed that there was no additional source of variation associated with clonal replication. Genotype-environment covariance and interaction were not considered. Natural selection, migration (closed BP) and mutation were further assumed absent.

The breeding value of each progeny genotype was predicted using the combined index of individual clonal mean and family mean (average of clonal means within a family). The general approach for combined index calculation was presented by Baker (1986, pp. 110–111). Relatedness was described by "group coancestry",  $\Theta$ , a concept introduced by Cockerham (1967). Group coancestry is the average of all pairwise coancestries, including self-coancestry and reciprocals. It is the probability that two genes taken at random from the gene pool, with replacement, are identical by descent. Status number,  $N_S$ , is a measure of effective population size: it is the census size of an equivalent population composed of unrelated, non-inbred individuals, where the probability to draw two genes identical by descent is the same as for the population under study (Lindgren et al. 1996). Status number was calculated from group coancestry ( $N_S = 1/2 \Theta$ ).

Group-merit selection (Lindgren and Mullin 1997) considering both breeding value and relatedness was applied to select 48 progeny genotypes to form the next cycle of the BP. The algorithm utilized in this model maximized iteratively the population merit:  $B_\omega = \bar{g}_\omega - c\Theta_\omega$  where  $B_\omega$  is the group merit of a selected set  $\omega$ ,  $\bar{g}_\omega$  is the average breeding value of the set,  $\Theta_\omega$  is the group coancestry of the set and  $c$  is a weighting constant. The

weighting constant was varied between zero, the extreme when only breeding value was considered, and infinity (approximated by  $1.0 \times 10^7$ ), when only group coancestry was considered in selecting among families. In this way it was possible to study a wide range of alternatives varying from low to high restrictions on relatedness. Trees in the selected BP were then ranked by their respective breeding values (combined index). The top-ranking selections were allocated to the nucleus tier and the remaining members formed the main tier.

Group coancestry between the nucleus and main tier,  $\Theta_{NM}$ , was calculated in each cycle to express the average relatedness of trees in the nucleus tier with those in the main tier. Since trees were not duplicated, this was equal to the average of all pairwise coancestries between the two tiers.

In each cycle, the six genotypes with the highest breeding values were selected from the BP to contribute to a PP (seed orchard). Group coancestry was not considered as a selection criterion in selecting the orchard. This was done to test the ability of the BP to support seed orchards and determine what gene diversity these best clones could carry. The number of genotypes used in the orchard (six) is somewhat arbitrary (Rosvall et al. 1999); and it is assumed that other sources (adjacent BPs) of selected genotypes would contribute to any given seed orchard.

Selection and breeding were repeated, as described above, for five generations. The NB alternative was compared with PAM and RM alternatives, using the same BP size (48), and under the same testing effort. Positive assortative mating in the PAM alternative was done such that all trees in the BP were sorted according to their combined-index values before mating.

## Results

The main effect of selection and selection restrictions on the BP structure is demonstrated for the three mating alternatives in Figure 2. The values for BP  $N_S$  vary from low at low weight on group coancestry in selecting the BP (low  $N_S$  resulting from unbalanced selection scenarios) to high at a high weight (high  $N_S$  resulting from balanced selection scenarios). A nucleus size of 16 is used for the results presented in the figures unless otherwise specified. Both the population-wide assortment of mates (PAM) and the assortment in a group sense (NB) resulted in changes in the average additive effect ( $A$ ),  $V_A$  and average inbreeding coefficient ( $F$ ) in the BP (Figure 3). The average dominance effect and the dominance variance in the BP were not significantly altered by any parental assortment.

Breeding population  $A$  was enhanced by the positive assortment of mates (NB, PAM) compared with RM when the BP was selected with a low weight on group coancestry, resulting in a low BP  $N_S$  (Figure 2). The difference is more pronounced in the later breeding cycles (see graph corresponding to breeding cycle 5 in Figure 2). Compared with RM, the PAM alternative resulted in the greatest enhancement in BP  $A$ , followed by the NB alternative. With higher weight on group coancestry (higher BP  $N_S$ ), there were no differences in BP  $A$  among the three alternatives.

Breeding population  $V_A$  was reduced in the subsequent breeding cycles under the RM alternative (compare  $V_A - RM$  in breeding cycles 1 and 5 in Figure 2), while PAM and NB continuously enhanced BP  $V_A$  in all breeding cycles, with the largest enhancement under more balanced selection scenarios (high BP  $N_S$ ). The maximum enhancement again was

observed under the PAM alternative, with the open-nucleus alternative resulting in values approximately intermediate to those of the RM and PAM alternatives.

The PAM alternative resulted in the highest BP  $F$  of the three alternatives, while the RM alternative resulted in the lowest BP  $F$  (Figure 3).

Group coancestry between the nucleus and main tier in the NB alternative expressed accumulated relatedness between the tiers as a result of recurrent selection and breeding (Figure 4). Low weight on group coancestry when selecting the BP resulted in higher group coancestry between tiers.

The most important results of this study are presented in Figure 5, where the additive effect of the selected seed orchard, i.e., genetic gain (PP  $A$ ) in the first and fifth cycles, is plotted against the PP  $N_S$  (census number = 6). The greatest PP  $A$  of the three alternatives was achieved with the PAM alternative. NB resulted in a PP  $A$  intermediate between the RM and PAM alternatives. The position of a maximum value for PP  $A$  suggests that restricted selection provides the greatest gain in a long run.

The effect of varying the size of the nucleus in the NB alternative on PP  $A$  in the fifth cycle seed orchard at low and high PP  $N_S$  respectively is shown in Figure 6. Nucleus sizes of zero and 48 correspond to the nonhierarchical, randomly mated population "RM" with no population subdivision; and results for both cases are therefore equivalent. Smaller nucleus sizes were more favorable when lower levels of seed orchard diversity were accepted, while larger sizes resulted in greater cumulative gains when diversity was considered important.

## Discussion

The simulation is based on the infinitesimal model for genetic variation in a quantitative trait that considers genetic sampling and the effects of selection and inbreeding on genetic variance. This model is suitable for showing generally applicable principles relevant to the majority of growth traits in conifer species (for references, see Mullin and Park 1995; Andersson 1999; Rosvall 1999). A BP of 48 individuals has been shown to be reasonable for studying the sustainable effect of assortative mating relevant to long-term forest tree breeding (Rosvall and Mullin 2003). Simulation iterations may be thought of as independent replicates of the same breeding program. Conclusions were drawn across these replicates. This study provides generally applicable recommendations under balanced, albeit somewhat ideal, conditions. Under some circumstances, operational constraints in a real breeding program may present difficulties to complete all desired assortative mating. In these cases, the advantage of such mating would be somewhat reduced according to the mating success rate.

The advantage of population management schemes modeled in this study is that both gain and diversity are considered simultaneously when selecting the BP. Furthermore, the gene transfer between tiers is simply directed by the breeding value of available candidates in the recruitment population and the weight placed on group coancestry when the BP is selected. The population structure is optimized by maximizing genetic gain at any target level of diversity in the PP for any value of initial parameters. This is a powerful tool for comparing breeding alternatives.

Positive assortative mating enhanced the additive variance when a high weight was placed on group coancestry when selecting the BP (high BP  $N_S$  resulting from more balanced

selection scenarios) and increased the potential for gains in the PP. It has been described earlier that larger variance enhancement due to assortative mating is expected at lower selection intensities (Baker 1973; De Lange 1974; Jorjani 1995), since the enhancement in variation among families in the recruitment population is utilized when the BP is selected. The enhancement of additive variance due to PAM is greater compared with the effect of inbreeding, which may at most only double the additive variance in the absence of epistasis (Falconer and Mackay 1996; Lynch and Walsh 1998). PAM may cause greater expansion of additive variance, provided the number of loci is large (Crow and Felsenstein 1968). Assortative mating does not affect dominance and environmental variance, under the assumptions described by Crow (1986), and therefore the increased BP  $V_A$  by PAM also increases heritability.

Under unbalanced selection scenarios (low BP  $N_S$ ), most BP trees were selected within a limited number of superior crosses. The extra gain ( $A$ ) in the BP by PAM compared with the RM alternative under these conditions was due to the higher values of these best crosses. The increased inbreeding in the BP by PAM (as more related trees have a tendency to be more closely ranked) had an adverse effect on gain by decreasing the within-family portion of additive variance. This loss was, however, greatly overcompensated by the increased additive effect and additive variance among families in BP (compare Figures 2 and 3). Rosvall and Mullin (2003; Figure 4) attempted to express the decline in the mean phenotype with increasing homozygosity in the population. When inbreeding depression was present, PAM still resulted in enhanced PP gains compared with a RM population (1% inbreeding depression per 0.01 increase in the inbreeding coefficient was considered in their model).

However, modeling inbreeding depression by a simple reduction in the individual dominance effect and assuming that all trees with the same inbreeding coefficient are subjected to the same inbreeding depression seems the most unrealistic point in their infinitesimal simulation model (Rosvall 1999, p. 48). The biology underlying the inbreeding depression is much more complex (Williams and Savolainen 1996; Lynch and Walsh 1998) and no single recommendation can be generalized for numerous species under various management schemes. Under real situations, selection candidates that are highly inbred are likely avoided. When mates are ranked (PAM), some avoidance of crosses with close relatives can be applied, thus utilizing the advantage of increased phenotypic correlation of mates described in this study and at the same time reducing the negative impact of inbreeding on the mean phenotype in the recruitment and, consequently, the BP and PP.

Assortment of mates thus generated a potential for more PP gain due to the increased additive variance in the BP under more balanced selection scenarios and the increased additive effect in BP in more unbalanced selection scenarios (depending upon the weight applied to group coancestry when the BP was selected). The increased selection precision with clonal replication makes it possible to effectively explore the within-family portion of the additive variance (which is reproduced by genetic sampling in each new generation) under these balanced scenarios. The conserved among-family variance can then be exploited when selecting the PP. In this way, diversity is conserved in the BP, while PP gain is maintained at a high level. This beneficial effect of increased additive variance by assortative mating under restricted family selection was emphasized by Rosvall and Mullin (2003) for long-term forest tree breeding.

The key result of this study (presented in Figure 5), that the assortment of mates in a group sense (open nucleus) is a less powerful version of the individual population-wide parental assortment, is in agreement with Kinghorn et al. (2000). Since there was no effect of NB on the BP or the PP, other than the one described for assortative mating, it can be concluded that open-NB would only provide genetic advantages under circumstances that favor population-wide assortative mating. However, under such circumstances, a population-wide assortment of mates would always provide more gain regardless of the target-level of production–population diversity. It can be noted from Figure 5 that splitting the BP into just two hierarchical levels (open nucleus) has a large effect on the PP gain. The diminishing return from adding more and more tiers to the existing hierarchy was noted earlier by Shepherd (1991).

In this two-tier model, both gain and diversity in PP were influenced by the nucleus size. In a long-term perspective, smaller nucleus sizes (approximately 20% of the BP size) were more favorable (resulted in higher PP gain) when lower values of seed orchard diversity were accepted (low PP  $N_S$ ). This was due to the increased correlation among mates within the small nucleus, which generated extra genetic response (increased  $A$  in the selected PP). The optimum nucleus size increased to approximately 50–70% of the BP size when higher values of orchard diversity were the target (high PP  $N_S$ ). The fact that relatively large nucleus sizes provided maximum PP long-term gain is a point to be noted for open-NB application in forest trees, where the financial investment is focused on relatively smaller nucleus sizes (White et al. 1999; Mikola 2002). In earlier animal-breeding studies, the optimum nucleus sizes were often underestimated due to some simplified assumptions in their models (e.g.,

genetic variance was assumed to be constant throughout breeding cycles). Later studies revealed that optimum nucleus sizes are larger, especially when the effects of inbreeding and selection on genetic variance are accounted for (Roden 1994).

In forest tree breeding, Rosvall et al. (2003) investigated PAM combined with an increased number of selections from the best BP members and found similarities to open-NB. The present study is the first in making a direct comparison of both approaches. Distributing the testing effort equally to BP members (as done in this study) is a simple approach to implementing such a scheme, but real-life breeding programs often propose a more complex allocation of effort. In the next part of this study (Lstibůrek et al. 2004, this issue), breeding alternatives are evaluated where more effort is concentrated on parents of higher rank (unequal distribution of testing effort).

## **Conclusion**

Applying population-wide PAM in the BP is more efficient in supporting the PP than is open-NB for any desired PP gene diversity. Both PAM and NB enhance the additive variance and the additive effect of the BP by mating closely ranked parents and thus the top clones used in the PP are improved compared with those used under unstructured breeding, but PAM is more effective. Larger nucleus sizes in the NB alternative result in higher PP gain when the desired level of PP diversity is high. Smaller nucleus sizes can be marginally more efficient when PP diversity is considered less important.

## **Acknowledgements**

We gratefully acknowledge financial support from the Föreningen Skogsträdsförädling (Swedish Tree Breeding Association) and the North Carolina State University–Industry Cooperative Tree Improvement Program. We also thank Dr. Jennifer Myszewski and Dr. Floyd Bridgwater for their valuable comments on the manuscript and Dr. Ross Shepherd for his generous help with the animal breeding terminology.

## References

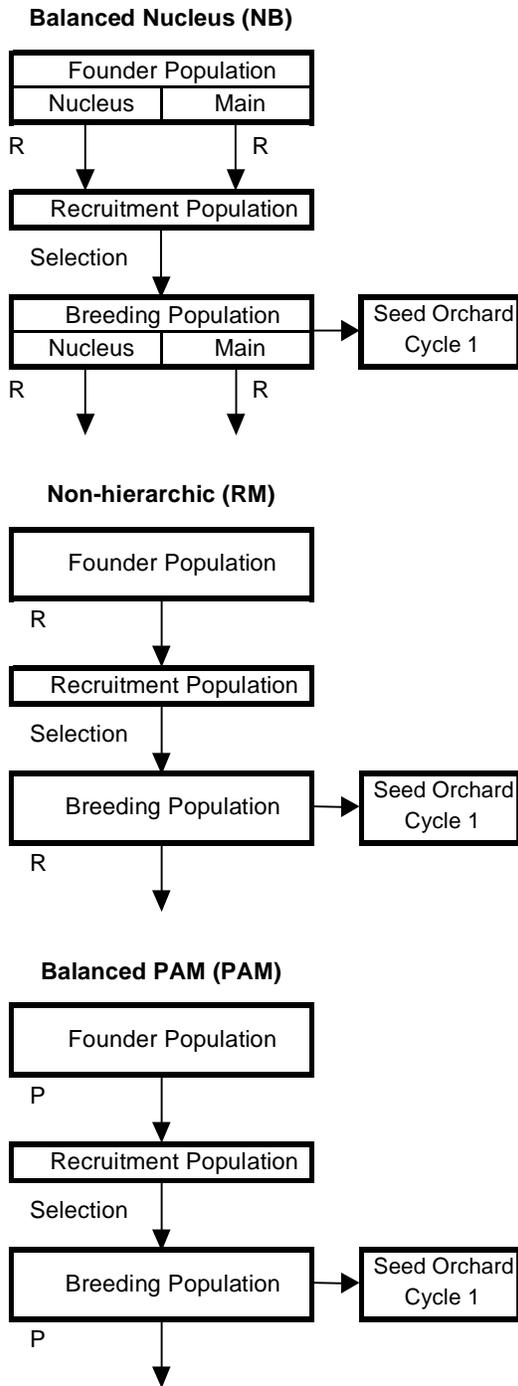
- Andersson EW (1999) Gain and diversity in multi-generation breeding programs. PhD thesis, Swedish University of Agricultural Sciences, Umeå
- Baker RJ (1973) Assortative mating and artificial selection. *Heredity* 31:231–238
- Baker RJ (1986) Selection indices in plant breeding. CRC Press, Boca Raton
- Breese EL (1956) The genetical consequences of assortative mating. *Heredity* 10:323–343
- Bulmer MG (1971) The effect of selection on genetic variability. *Am Nat* 105:201–211
- Cockerham CC (1967) Group inbreeding and coancestry. *Genetics* 56:89–104
- Cotterill PP (1989) The nucleus breeding system. *Proc South For Tree Improv Conf* 20:36–42
- Cotterill PP, Dean CA, Cameron J, Brindbergs M (1989) Nucleus breeding: a new strategy for rapid improvement under clonal forestry. In: Gibson GL, Griffin AR, Matheson AC (eds) *Breeding tropical trees: population structure and genetic improvement strategies in clonal and seedling forestry*. (Proceedings of a conference held in Pattaya, Thailand, 28 November–3 December 1988) Winrock International, Arlington, pp 39–51
- Crow JF (1986) Basic concepts in population, quantitative, and evolutionary genetics. Freeman, New York
- Crow JF, Felsenstein J (1968) The effects of assortative mating on the genetic composition of a population. *Eugen Q* 15:85–97
- Crow JF, Kimura M (1970) *An introduction to population genetics theory*. Harper and Row, New York

- Danusevicius D, Lindgren D (2002) Efficiency of selection based on phenotype, clone and progeny testing in long-term breeding. *Silvae Genet* 51:19–26
- del-Bosque González AS (1989) Simulations of nucleus breeding schemes for wool production. PhD thesis, University of New England, Armidale
- Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics, 4th edn. Longman, Harlow
- Hagedorn SF (1991) Breeding *Pinus patula* for plantation establishment in the Natal midlands. (Annual research report 1991) Institute for Commercial Forestry Research, Pietermaritzburg, pp 204–207
- Hopkins IR, James JW (1978) Theory of nucleus breeding schemes with overlapping generations. *Theor Appl Genet* 53:17–24
- Isik F, Li B, Frampton J, Goldfarb B (2003) Efficiency of seedlings and rooted cuttings for testing and selection in *Pinus taeda*. *For Sci* 50:44–53
- James JW (1977) Open nucleus breeding systems. *Anim Prod* 24:287–305
- James JW (1978) Effective population size in open nucleus breeding schemes. *Acta Agric Scand* 28:387–392
- Jorjani H (1995) Genetic studies of assortative mating in selected and unselected populations. PhD thesis, Swedish University of Agricultural Sciences, Uppsala
- Karlsson B, Rosvall O (1993) Breeding programmes in Sweden 3. Norway spruce. In: Lee SJ (ed) Progeny testing and breeding strategies. (Proceedings of the Nordic group of tree breeding, October 1993, Edinburgh) Forestry Commission, Edinburgh, pp 135–141

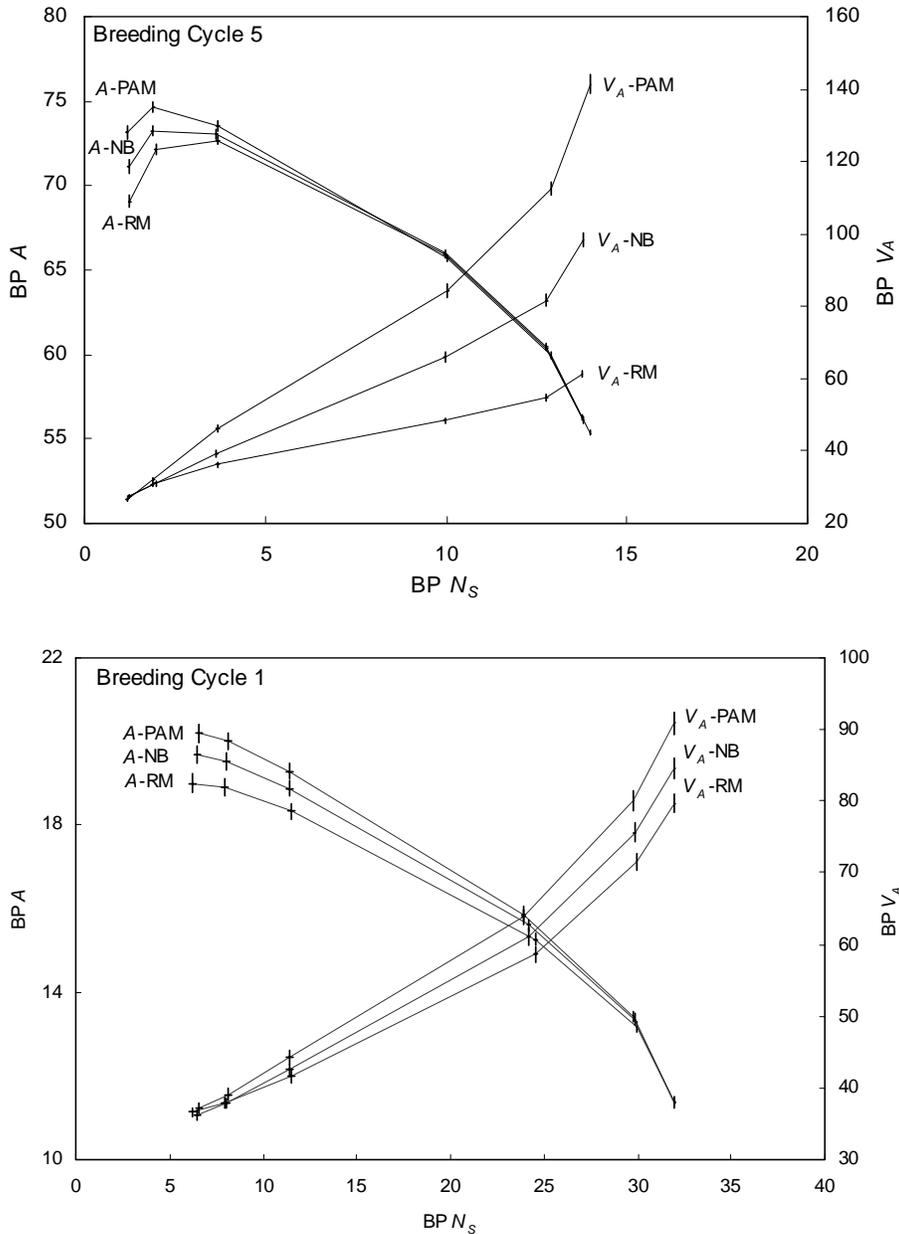
- King JN, Johnson GR (1993) Monte Carlo simulation models of breeding-population advancement. *Silvae Genet* 42:68–78
- Kinghorn B, Werf J, Ryan M (2000) *Animal breeding: use of new technologies: a textbook for consultants, farmers, teachers and for students of animal breeding*. University of Sydney, Sydney
- De Lange AO (1974) A simulation study of the effects of assortative mating on the response to selection. In: Garsi (ed) *First world congress on genetics applied to livestock production, 7–11 October 1974. 3. Contributed papers*. Madrid, pp 421–425
- Lindgren D, Mullin TJ (1997) Balancing gain and relatedness in selection. *Silvae Genet* 46:124–129
- Lindgren D, Gea LD, Jefferson PA (1996) Loss of genetic diversity monitored by status number. *Silvae Genet* 45:52–59
- Lstibůrek M, Mullin TJ, Lindgren D, Rosvall O (2004) Open-nucleus breeding strategies compared with population-wide positive assortative mating. II. Unequal distribution of testing effort. *Theor Appl Genet* 109:1169–1177
- Lynch M, Walsh B (1998) *Genetics and analysis of quantitative traits*. Sinauer, Sunderland, Massachusetts
- Mahalovich MF, Bridgwater FE (1989) Modeling elite populations and positive assortative mating in recurrent selection programs for general combining ability. *Proc South For Tree Improv Conf* 20:43–49
- McKeand SE, Bridgwater FE (1998) A strategy for the third breeding cycle of loblolly pine in the Southeastern US. *Silvae Genet* 47:223–234

- Mikola J (2002) Long-term tree breeding strategy in Finland: integration of seed production and breeding. In: Haapanen M, Mikola J (eds) Integrating tree breeding and forestry. (Proceedings of the Nordic group for management of genetic resources of trees, Mekrijärvi, March 2001. Research papers 842) Finnish Forest Research Institute, Mekrijärvi, pp 12–13
- Mueller JP, James JW (1983) Effects of reduced variance due to selection in open nucleus breeding systems. *Aust J Agric Res* 34:53–62
- Mullin TJ, Park YS (1995) Stochastic simulation of population management strategies for tree breeding: a new decision-support tool for personal computers. *Silvae Genet* 44:132–141
- Roden JA (1994) Review of the theory of open nucleus breeding systems. *Anim Breed Abstr* 62:151–157
- Roden JA (1995) A simulation study of open nucleus and closed nucleus breeding systems in a sheep population. *Anim Sci* 60:117–124
- Rosvall O (1999) Enhancing gain from long-term forest tree breeding while conserving genetic diversity. PhD thesis, Swedish University of Agricultural Sciences, Umeå
- Rosvall O, Mullin TJ (2003) Positive assortative mating with selection restrictions on group coancestry enhances gain while conserving genetic diversity in long-term forest tree breeding. *Theor Appl Genet* 107:629–642
- Rosvall O, Lindgren D, Mullin TJ (1999) Sustainability robustness and efficiency of a multi-generation breeding strategy based on within-family clonal selection. *Silvae Genet* 47:307–321

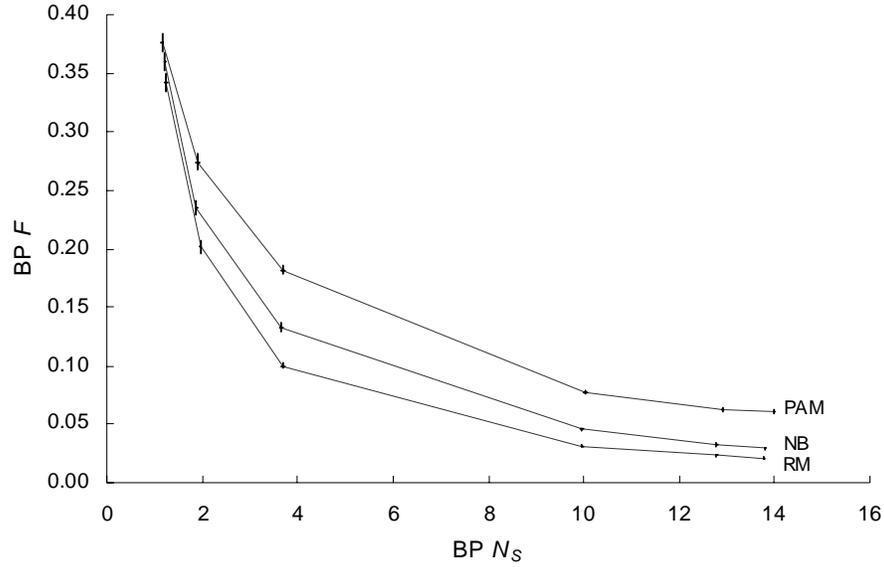
- Rosvall O, Mullin TJ, Lindgren D (2003) Controlling parent contributions during positive assortative mating and selection increases gain in long-term forest tree breeding. *For Genet* 10:35–53
- Shepherd RK (1991) Multi-tier open nucleus breeding systems. PhD thesis, University of New England, Armidale
- Shepherd RK, Kinghorn BP (1992) Optimising multi-tier open nucleus breeding schemes. *Theor Appl Genet* 85:372–378
- Wellendorf H, Skov E, Kjaer E (1994) Suggested updating of improvement strategy for Danish-grown Norway spruce, vol 25. *Forest Tree Improvement*, Arboretet Horsholm, pp 1–12
- White TL, Hodge GR, Powell GL (1993) An advanced-generation tree improvement plan for slash pine in the Southeastern United States. *Silvae Genet* 42:359–371
- White TL, Matheson AC, Cotterill PP, Johnson GR, Rout AF, Boomsma DB (1999) A nucleus breeding plan for radiata pine in Australia. *Silvae Genet* 48:122–133
- Williams CG, Savolainen O (1996) Inbreeding depression in conifers: implications for breeding strategy. *For Sci* 42:102–117



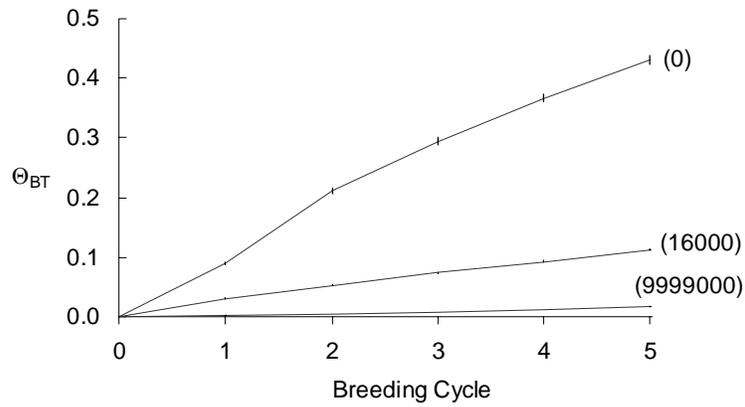
**Figure 1.** Breeding alternatives compared in this study. Mating was either random (R) or positive assortative (P) as indicated below the population box.



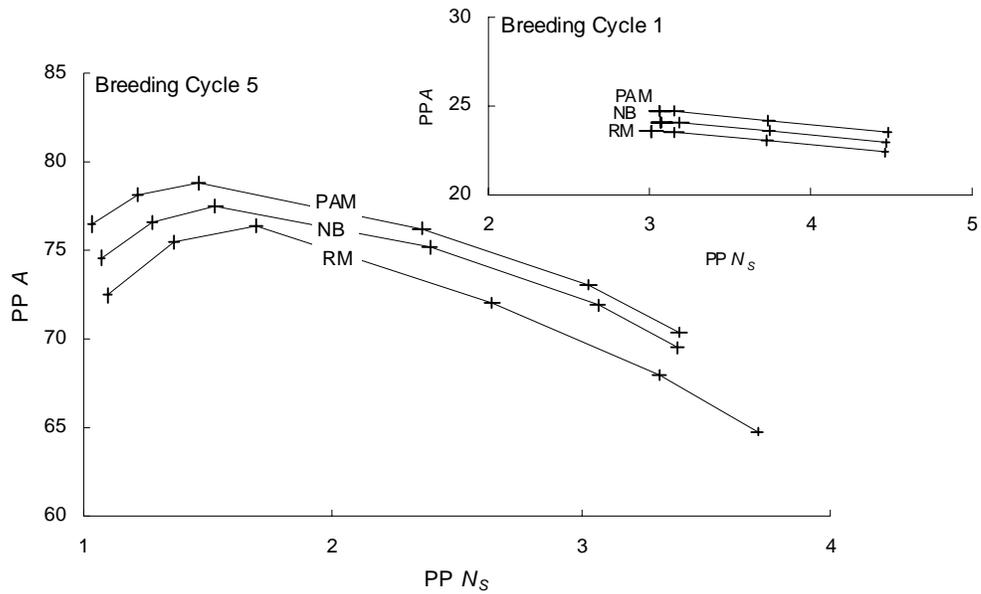
**Figure 2.** Average additive effect  $A$  and additive variance  $V_A$  in the BP after one and five breeding cycles for a set of scenarios with variable weight on group coancestry in the selection criterion, resulting in different BP status numbers ( $N_S$ ). The three lines connect scenarios for: non-hierarchical RM, NB with nucleus size  $N_N = 16$  and PAM. CI (95%) for BP  $A$ ,  $V_A$  and  $N_S$  are based on 800 iterations of the simulation.



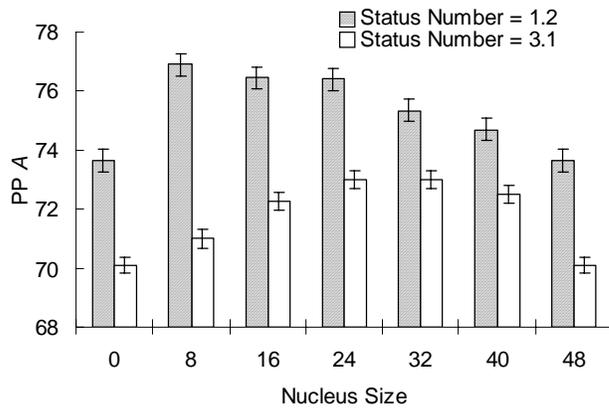
**Figure 3.** Average inbreeding coefficient ( $F$ ) in the BP after five breeding cycles for a set of scenarios with variable weight on group coancestry in the selection criterion, resulting in different BP status numbers ( $N_S$ ). The three lines connect scenarios for: non-hierarchic RM, NB with nucleus size  $N_N = 16$  and PAM. CI (95%) for BP  $F$  and  $N_S$  are plotted based on 800 iterations of the simulation.



**Figure 4.** Development of between-group coancestry ( $\Theta_{BT}$ ) over five breeding cycles in the NB alternative. The result of three scenarios with different weights (given in parentheses) on group coancestry while selecting the BP are illustrated. The nucleus size  $N_N = 16$ . CI (95%) for  $\Theta_{BT}$  are based on 800 iterations of the simulation.



**Figure 5.** Average additive effect  $A$  in the seed orchard (PP) established after one and five breeding cycles for a set of scenarios with variable weight on group coancestry in the selection criterion, resulting in different PP status numbers ( $N_S$ ). The three lines connect scenarios for: non-hierarchic RM, NB with nucleus size  $N_N = 16$  and PAM. CI (95%) for PP  $A$  and  $N_S$  are based on 800 iterations of the simulation.



**Figure 6.** Average additive effect  $A$  in the seed orchard (PP) established after five breeding cycles at variable nucleus sizes in the NB alternative interpolated at lower (1.2) and higher (3.1) PP status numbers, corresponding to low and high weight on group coancestry in selecting the BP. CI (95%) for PP  $A$  are based on 800 iterations of the simulation.

## **CHAPTER 3**

### **Open-nucleus breeding strategies compared with population-wide positive assortative mating**

#### **II. Unequal distribution of testing effort**

This article was originally published in Theoretical and Applied Genetics.

Authors: M. Lstibůrek, T. J. Mullin, D. Lindgren, O. Rosvall

Volume: 109

Page Numbers: 1169-1177

Year of Publication: 2004

The article is reproduced with a written permission from the publisher.

© Springer-Verlag Berlin Heidelberg

## **Abstract**

This study compares population-wide positive assortative mating (PAM) with open-nucleus breeding with an elite and main population when more effort is allocated to parents of the elite. A companion study showed that PAM is advantageous when testing effort is independent of parental value. In the present study, unbalanced testing was imposed by varying the number of crosses or the number of genotypes per cross. These unbalanced alternatives are compared with PAM, where the testing effort was varied so that better parents were mated more frequently. More effort allocated to parents of higher rank increased the additive effect and the additive variance and only slightly altered the group coancestry and inbreeding in the breeding population (BP) compared with completely balanced scenarios. Of particular interest to the breeder, large enhancement of the additive variance in the BP contributed to higher gains in the production population (PP). These simulations demonstrate that population-wide PAM leads to higher genetic gains compared with open-nucleus alternatives at any desired target level of diversity in the PP. This is true for both balanced (part I) and unbalanced distribution of testing effort (part II).

## **Introduction**

This is the second part of a simulation study comparing open-nucleus breeding and population-wide positive assortative mating (PAM) in long-term forest tree breeding based on recurrent selection for general combining ability. Part I illustrated comparisons under a relatively simple set of situations (Lstibůrek et al. 2004, this issue). In reality, breeding plans are often more complex (White et al. 1993; Danell 1995; McKeand and Bridgwater 1998; White et al. 1999; Li et al. 2000). These may combine both hierarchical structuring and sublining of the breeding population (BP). Hierarchical structuring means that a mating hierarchy is imposed on the BP, either at an individual level (PAM) or in groups (nucleus breeding). The main advantage of imposing a hierarchy for mating is the enhancement of additive variance in the BP, which can be utilized in the selected production population (PP) (Rosvall and Mullin 2003). Sublining means that a BP is divided into a number of subpopulations that are kept isolated and provide unrelated genotypes for the PP, e.g., a seed orchard, in each breeding cycle. In this way, inbreeding in the material deployed to forest plantations can be avoided (McKeand and Beineke 1980; Ruotsalainen and Lindgren 2000). "Multiple populations" is a related concept in forest tree breeding (Namkoong et al. 1988; White 1992), referring to a set of populations managed for different breeding objectives (Burdon and Namkoong 1983).

In the first part of this study, it was demonstrated quantitatively that population-wide PAM is superior to open-nucleus breeding for a variety of selection alternatives (unrestricted to restricted selection) when the distribution of testing effort across the recruitment population is balanced (Lstibůrek et al. 2004, this issue). In this second part, more complex

situations are evaluated, where the distribution of testing effort is not balanced. It has often been suggested to direct more effort on higher-ranking parents in the BP (e.g., Cotterill 1989). Focusing more resources on the elite part of a BP (nucleus tier) can be done in a number of ways (Smith 1988; White 1992). One way is to apply less expensive open pollination to regenerate the main tier and to reallocate financial resources to advance the nucleus tier by means of control crosses. More parental combinations or more progenies per parent can be made in the nucleus tier compared with the main tier. In addition, parents can be individually ranked by their values in the nucleus before mating takes place (Cotterill et al. 1989; Hagedorn 1991; White et al. 1999; Mikola 2002).

Similar options of resource reallocation are available when the BP is not managed in distinct hierarchical layers, but rather parental assortment is done on a population-wide basis. In theory, it is expected that the optimum contribution of parents to the next generation is linearly related to their respective breeding values (Wei and Lindgren 1995). Similarly, one might expect that the optimum resource allocation is related to parental breeding values and, since the distribution of these values is continuous, so might be the resource allocation. An effort to study more matings among better parents was done by Rosvall et al. (2003), but in their study the effect of resources is not separated from the effect of selection.

The objective of this study was to compare population-wide PAM with open-nucleus breeding under unbalanced distribution of testing effort, with more resources allocated to parents of higher rank. In the first part of this study, it was demonstrated that open-nucleus breeding does not provide any additional genetic benefit beyond that achieved by allocating the mates into two distinct hierarchical layers (assortment in a group sense). Thus, the

hypothesis under the unbalanced set of conditions was that the population-wide assortment of mates would again provide more gain in the PP compared with the open-nucleus breeding.

## **Materials and Methods**

The simulation in this study was based on the infinitesimal model for genetic variation in a quantitative trait that considers genetic sampling and the effects of selection and inbreeding on genetic variance. The breeding simulation software "POPSIM" developed by Mullin and Park (1995) was modified to model an open-nucleus breeding system with two hierarchical levels (nucleus and main tiers). Main model structures and assumptions are described in our earlier study dealing with the balanced distribution of testing effort (Lstibůrek et al. 2004, this issue). Figure 1 presents the main breeding alternatives compared in this study.

Each alternative started by generating 48 unrelated, non-inbred founders. In each breeding cycle, a BP was selected from a clonally replicated recruitment population (progenies of parents from the previous breeding cycle) by group-merit selection, combining average breeding value and group coancestry into a single selection criterion (Lindgren and Mullin 1997). Different scenarios were run for each breeding alternative, using a wide range of weights on group coancestry to produce results across a range of effective population sizes. In each breeding cycle, six top-ranking genotypes were selected from the BP to contribute to a PP (seed orchard). Breeding alternatives differed in the hierarchical structuring of the BP as follows.

### **Balanced nucleus**

Founder and BP were divided into two tiers: nucleus and main. The nucleus was composed of the 16 top-ranked parents in the BP (a third of the total BP). Mating within tiers was random. Single-pair mating was used with eight crosses made among nucleus tier parents and

16 crosses among main tier parents. A total of 100 genotypes (seedlings) were generated for each parental combination. Thus, there were 800 progenies in nucleus-tier families and 1,600 progenies in main-tier families. This alternative was identical to the balanced nucleus (NB) alternative in the first part of this study and was used here as a basis for comparisons.

### **Unbalanced nucleus**

At unbalanced nucleus alternatives (NUB), the nucleus was composed of the 16 top-ranked parents (as in the NB alternative). Three options for concentrating more effort on nucleus parents were evaluated.

#### *Unbalanced nucleus with double-pair mating*

Unbalanced nucleus with double pair mating (NDP) means that double-pair random mating was performed in the nucleus tier (16 crosses) and single-pair random mating in the main tier (16 crosses). The size of all families in the recruitment population was 75 genotypes, giving a total of 2,400 genotypes in the recruitment population.

#### *Unbalanced nucleus with bigger families*

Unbalanced nucleus with bigger families (NBF) means that single-pair mating was applied at random in both tiers (eight crosses in nucleus, 16 crosses in main tier) with larger families generated in the nucleus and smaller families in the main tier: 150/75 (2×); 200/50 (4×) and 250/25 (10×); giving a total of 1,200/1,200; 1,600/800 and 2,000/400 genotypes in the nucleus/main families.

### *Unbalanced nucleus with positive assortment of mates*

At unbalanced nucleus with positive assortment of mates (NPAM), the nucleus parents were sorted by breeding values before mating took place. Main tier parents were mated randomly. Double-pair mating was used in the nucleus tier (16 crosses) and single-pair mating in the main tier (16 crosses). A total of 75 seedlings were generated for each family, resulting in total of 2,400 genotypes in the recruitment population.

### **Non-hierarchic random mating**

This alternative functioned as a baseline [identical to the random mating (RM) alternative in the first part of this study]; the founder and BP were not subdivided into tiers (no hierarchical structure), and assignment of mates was at random. Single-pair mating resulted in a total of 24 crosses. Each family had a uniform size of 100 genotypes, giving a total of 2,400 genotypes in the recruitment population.

### **Unbalanced population-wide positive assortative mating**

In this alternative, population-wide PAM was modified to distribute more effort to higher-ranking parents and less to lower-ranking parents. In this modified PAM alternative, parents were first ranked by their breeding (combined-index) values. The top third of the parents were then each involved in three crosses, the middle third in two crosses, and the lowest third of parents in one cross—PAM 3:2:1 (PAM321) design (Ruotsalainen and Lindgren 2001; Rosvall et al. 2003). Thus, there were, in total, 48 crosses. A total of 50 genotypes were generated per each family, resulting in 2,400 genotypes in the recruitment population. The

like-with-like mating design was optimized such that the phenotypic correlation of mates was maximized (Figure 2).

All test genotypes (seedlings) were clonally replicated by ten ramets, resulting in total testing effort (size of recruitment population) of 24,000 test plants in each breeding alternative. The simulation was conducted for five cycles of selection and breeding. Each simulation scenario was replicated by 800 iterations.

## Results

### More crosses in the nucleus

Figure 3 shows the average additive effect ( $A$ ), additive variance ( $V_A$ ) and average inbreeding coefficient ( $F$ ) in the BP after five breeding cycles for unbalanced nucleus (NDP, NPAM), the PAM321, the PAM, and the RM alternative.

The values for status number ( $N_S$ ) on the x-axis vary from low at low weight on group coancestry in selecting the BP (low  $N_S$  resulting from unbalanced selection scenarios) to high at high weight (high  $N_S$  resulting from balanced selection scenarios). In all alternatives, the maximum BP  $A$  was found in the lower end of the BP  $N_S$  range and the minimum at the maximum BP  $N_S$  (balanced within-family selection). BP  $A$  at the lower end of the BP  $N_S$  range was greatest under the PAM321 alternative, followed by the nucleus and RM alternatives. In this range of BP  $N_S$ , NPAM generated more gain than the corresponding alternative with random mating in the nucleus, NDP. When the weighting on group coancestry resulted in BP  $N_S$  approximately equal to 10, all alternatives achieved similar BP  $A$ . With greater restriction on diversity, the ranking of alternatives was inverted compared with ranks under low BP  $N_S$ , but the range among alternatives was smaller. There was negligible enhancement of BP  $V_A$  due to parental assortment at low BP  $N_S$ . The difference in BP  $V_A$  among alternatives increased with higher weight on group coancestry (higher BP  $N_S$ ). Maximum enhancement of BP  $V_A$  was obtained by the PAM321 alternative. The two nucleus alternatives (NDP, NPAM) were similar to each other. When there was low or no weight on group coancestry (low BP  $N_S$ ), nucleus alternatives resulted in the greatest BP  $F$ , followed by

RM and PAM321 alternatives. With greater weights (higher BP  $N_S$ ), the PAM321 alternative resulted in the greatest BP  $F$  followed by the nucleus and RM alternatives.

### **Larger families in the nucleus**

Increasing family sizes in the nucleus can also lead to enhancement of  $A$  and  $V_A$  in the BP. An NBF (two times and ten times more genotypes per family in the nucleus compared with the main tier) is compared with NB, PAM321, PAM and RM alternatives in Figure 4.

When there was low or no weight on group coancestry (low BP  $N_S$ ), larger family sizes in the nucleus resulted in greater  $A$  in the BP. Similar to alternatives with larger numbers of crosses in the nucleus (Figure 3), scenarios with larger family sizes in the nucleus also resulted in approximately the same  $A$  in BP at  $N_S$  close to ten. The exception was when family size in the nucleus greatly exceeded the family size in the main tier ( $A$ -NBF10 $\times$ ). This extreme alternative achieved the least additive improvement of all the alternatives at the moderate to high BP  $N_S$  range. The ranking of nucleus alternatives in the high BP  $N_S$  range was again inverted, compared with ranks under low BP  $N_S$ . Greatest BP  $V_A$  enhancement was achieved with the NBF alternative with family sizes in the nucleus ten times that in the main tier. Other alternatives followed as indicated in Figure 4. Nucleus alternatives produced the greatest BP  $F$  at lower BP  $N_S$  range, followed by the RM and PAM321 alternatives. The rank of alternatives under higher BP  $N_S$  changed (as in Figure 3), with the PAM321 alternative producing the greatest BP  $F$ , followed by the NBF and RM alternatives.

## **Production population**

Out of all the alternatives compared in this study, PAM321 produced the greatest genetic gain at any level of PP diversity (Figures 5, 6, 7). Doubling the number of crosses in the nucleus tier substantially increased PP gain in all cycles at all levels of target diversity (Figure 5).

Increasing family sizes in the nucleus also delivers more PP gain. Larger family sizes in the nucleus, compared with those in the main tier, are favorable at lower PP  $N_S$ , while more balanced conditions are favorable at higher diversity levels (Figure 6). Vertical lines in Figure 6 depict the space of solutions covering a range of unbalanced conditions.

Positive assortative mating in the nucleus further increases PP gain, but leads eventually to lower PP  $N_S$  in PP established in later cycles—at high weights on group coancestry when selecting the BP (Figure 7).

## **Discussion**

This study demonstrates the advantage of allocating resources according to parental breeding values. This advantage generally supports suggestions made by Cotterill (1989) in the case of an open-nucleus breeding system and those by Rosvall et al. (2003) in the case of a population-wide PAM. This study quantitatively compares open-nucleus breeding with population-wide PAM at both balanced (Lstibůrek et al. 2004, this issue) and unbalanced distribution of testing effort (current study). The results support the genetic advantage of population-wide PAM over open-nucleus breeding under both situations.

### **Genetic advantage: breeding population**

Generally, allocating more effort to better parents has the potential to enhance both  $A$  and  $V_A$  in the BP, which contributes to the beneficial effect of PAM per se described in this study and elsewhere (Jorjani 1995; Rosvall and Mullin 2003). By using group-merit selection to create a range of population effective sizes, it was shown that these conclusions are valid for a range of selection options that may occur under practical conditions.

When the desired level of genetic diversity is low (unbalanced selection scenarios), the allocation of more resources to the elite part of the BP produces additional gain in the BP. This is because unrestricted selection is made from a few superior parental combinations. Under this diversity objective, it is not worthwhile to spend resources on poor crosses that will not contribute to the next breeding cycle; it is more meaningful to reallocate these resources to better parental combinations (lower  $N_S$ , Figures 3, 4). At higher target diversity levels (balanced selection scenarios), parental combinations are forced to contribute more

equally to the next breeding cycle and allocating fewer resources to the poorer part of the BP can result in lower BP  $A$ . In this case, the reduction in gain is a consequence of unbalanced resource allocation (higher  $N_S$ , Figures 3, 4). Under the circumstances in this study, a balance existed when  $N_S$  equaled about 10. At this point, BP  $A$  was constant for all alternatives unless the magnitude of imbalance was extreme (NBF10 $\times$  in Figure 4). In the PAM321 design, the allocation of resources to parents is closer to the expected optimum (more continuously distributed). This "close-to-optimum" distribution of effort produced gains exceeding those produced by all other alternatives at lower target levels of genetic diversity. Since less effort goes to the poorest part of BP in this design, BP  $A$  shrinks when the desired diversity is high. The results suggest that if BP gain is the target, then it would not be worthwhile to reallocate resources to better parents, but rather to distribute them equally. Under balanced selection strategies, it is the additional enhancement of BP  $V_A$  due to the unbalanced allocation of effort that is beneficial.

The peak BP  $V_A$  enhancement was reached under balanced within-family selection (maximum desired level of genetic diversity). Here, the unequal distribution of effort was combined with equal parental contributions to the next breeding cycle. The resulting recruitment population was more heterogeneous under unequal distribution of effort and when the selection scenarios were more balanced (high BP  $N_S$ ); this further enhanced  $V_A$  of the selected BP in the next cycle.

The PAM321 alternative resulted in lower BP  $F$  at lower BP  $N_S$ . This was because of a more conveniently structured recruitment population, with more evenly distributed values of superior crosses (selection was made from a greater number of parental combinations

compared with nucleus or RM alternatives). Under more restricted selection (higher desired level of diversity), the PAM321 alternative resulted in the highest BP  $F$  over all alternatives. Here, the selection was made more evenly out of 48 crosses. The majority of these crosses were allocated among the top third of parents. Therefore, the group coancestry of selected trees was higher compared with alternatives where the distribution of parental contributions to the recruitment population was more balanced.

### **Genetic advantage: production population**

Improved genetic material is delivered to forest plantations through the PP. Thus, the PP is the ultimate target for maximizing genetic gain. The consequences for practical breeding are given here, where the discussion focuses on the PP, which is a subset of top-ranking parents passing their genes through regeneration material to forest plantations. This is where the comparison of breeding alternatives is most relevant. Higher gene diversity in the BP gives more room for a more intensive selection sacrificing a larger share of the gene diversity when selections are made to the PP. Therefore, preservation of gene diversity in the BP will appear more important if the PP is considered than if progress in the BP is the only goal.

The PAM321 alternative produced the greatest gains at all target levels of diversity in the PP, in all breeding cycles. This alternative was superior due to the favorable balance between  $A$ ,  $V_A$ , and  $F$  in BP. The PAM321 alternative is an easy scheme to implement, but there is probably a more optimum design where the allocation of effort is even more continuous and more adapted to the targets set. The advantage of 3:2:1 design as implemented in this study is that it clearly demonstrates the superiority of allocating more

effort to better parents under population-wide PAM rather than spending more resources in just one tier, as in case of open-nucleus breeding. Our future research will investigate alternatives to PAM 3:2:1 design, with more continuous allocation of testing effort and where the degree of imbalance is controlled to a higher degree.

The nucleus breeding schemes with more resources spent on better parents resulted in additional PP gain compared with the NB alternative. Making more crosses in the nucleus compared with the main tier resulted in a wider pool of extreme candidates for the next cycle, which was reflected in larger genetic gains in seed orchards. PAM within the nucleus tier increased genetic gain in the seed orchard, but led eventually to a reduction of PP  $N_S$  in later cycles under balanced selection schemes. This occurred because the recruitment population became more heterogeneous and led to fewer families contributing to the PP. Increasing the family size in the nucleus generated a larger pool of extreme candidates for selection. It is not worthwhile to produce very large differences between family sizes in two tiers. An optimum in this study was where families in the nucleus were two to four times larger compared with the main tier (depending upon the target PP  $N_S$ ).

### **Open-nucleus population in forest tree breeding?**

Animal breeders first adopted the concept of nucleus breeding primarily as it offered a convenient way to cooperate. There are some practical factors that made nucleus breeding suitable for animal breeders in some countries (often rather small breeding organizations; the breeding stock mostly owned by private persons). Animal breeders are aware of some

potential problems associated with nucleus breeding systems, and genotype-by-environment interaction is considered as the major problem (del-Bosque González 1989; Willis 1998).

The majority of the animal breeding-simulation studies have compared closed- versus open-nucleus breeding strategies (Roden 1994), as these were the alternatives considered under their practical circumstances. A direct comparison with animal breeding studies is hampered by key characteristics of animals, in particular, distinct sexes and low female reproductive rates. When compared with animals, the situation in forest tree breeding seems more practically suited to simultaneous evaluation of all candidates and the use of population-wide assortative mating, effects which can be further enhanced by the use of clonal replication in progeny testing.

In fact, nucleus schemes in animal breeding are considered more as an alternative to large-scale genetic evaluation, which entails detailed assessment of large populations, typically entire breeds (Bourdon 1997). Animals are evaluated objectively, utilizing all available information (BLUP analysis), which allows direct comparison of animals located in different herds or flocks. This leads to a larger pool of candidates for selection, which may in turn result in greater genetic progress.

The advantage of open nucleus breeding observed in this study is the potential for more gain in the PP due to the assortment in a group sense (when compared with the nonhierarchic RM population). However, when compared with population-wide PAM, the genetic advantage of open-nucleus for forest tree breeding seems less attractive.

In this study, the BP can be thought of as a single subline, and there might be a number of parallel sublimes in a breeding program (Rosvall et al. 1999). There was no attempt in this

study to evaluate all possible options of structuring BP, as the single-subline case demonstrates the main principle. The model could be tailored to a more specific case, if desired.

There might be some other advantages of open nucleus through factors not considered in this study. Functioning of breeding cooperatives or a shifting to smaller BP are examples of such factors that may favor the use of open-nucleus in forest tree breeding under specific conditions.

## **Conclusion**

This study demonstrates that population-wide PAM provides more gain (short and long term) in the PP than open-nucleus breeding schemes. This is true for both balanced (Lstibůrek et al. 2004, this issue; first part of the study) and unbalanced distribution of effort (this paper; second part of the study). These results apply to any target level of genetic diversity in the PP. Unbalanced distribution of testing effort results in a substantial increase in genetic gain at any reasonable target value of diversity. Our proposal to forest tree breeders is to consider population-wide PAM with more effort concentrated in higher-ranking parents as a more powerful alternative to open-nucleus breeding schemes.

## **Acknowledgements**

We gratefully acknowledge financial support from the Föreningen Skogsträdsförädling (Swedish Tree Breeding Association) and the North Carolina State University-Industry Cooperative Tree Improvement Program. We also thank Dr. Jennifer Myszewski and Dr. Floyd Bridgwater for their valuable comments on the manuscript and Dr. Ross Shepherd for his generous help with the animal breeding terminology.

## References

- Bourdon RM (1997) Understanding animal breeding. Prentice Hall, Upper Saddle River
- Burdon RD, Namkoong G (1983) Short note: multiple populations and sublines. *Silvae Genet* 32:221–222
- Cotterill PP (1989) The nucleus breeding system. *Proc South For Tree Improv Conf* 20:36–42
- Cotterill PP, Dean CA, Cameron J, Brindbergs M (1989) Nucleus breeding: a new strategy for rapid improvement under clonal forestry. In: Gibson GL, Griffin AR, Matheson AC (eds) *Breeding tropical trees: population structure and genetic improvement strategies in clonal and seedling forestry*. (Proceedings of a conference held in Pattaya, Thailand, 28 November–3 December 1988) Winrock International, Arlington, pp 39–51
- Danell Ö (1995) Breeding programs in Sweden. 1. General approach. SkogForsk. Arbetsrapport nr 302, pp 1–4
- del-Bosque González AS (1989) Simulations of nucleus breeding schemes for wool production. PhD Dissertation, University of New England, Armidale
- Hagedorn SF (1991) Breeding *Pinus patula* for plantation establishment in the Natal midlands. (Annual research report 1991) Institute for Commercial Forestry Research, Pietermaritzburg, pp 204–207
- Jorjani H (1995) Genetic studies of assortative mating in selected and unselected populations. PhD thesis, Swedish University of Agricultural Sciences, Uppsala
- Li B, McKeand S, Weir R (2000) Impact of forest genetics on sustainable forestry—results from two cycles of loblolly pine breeding in the US. *J Sustain For* 10:79–85

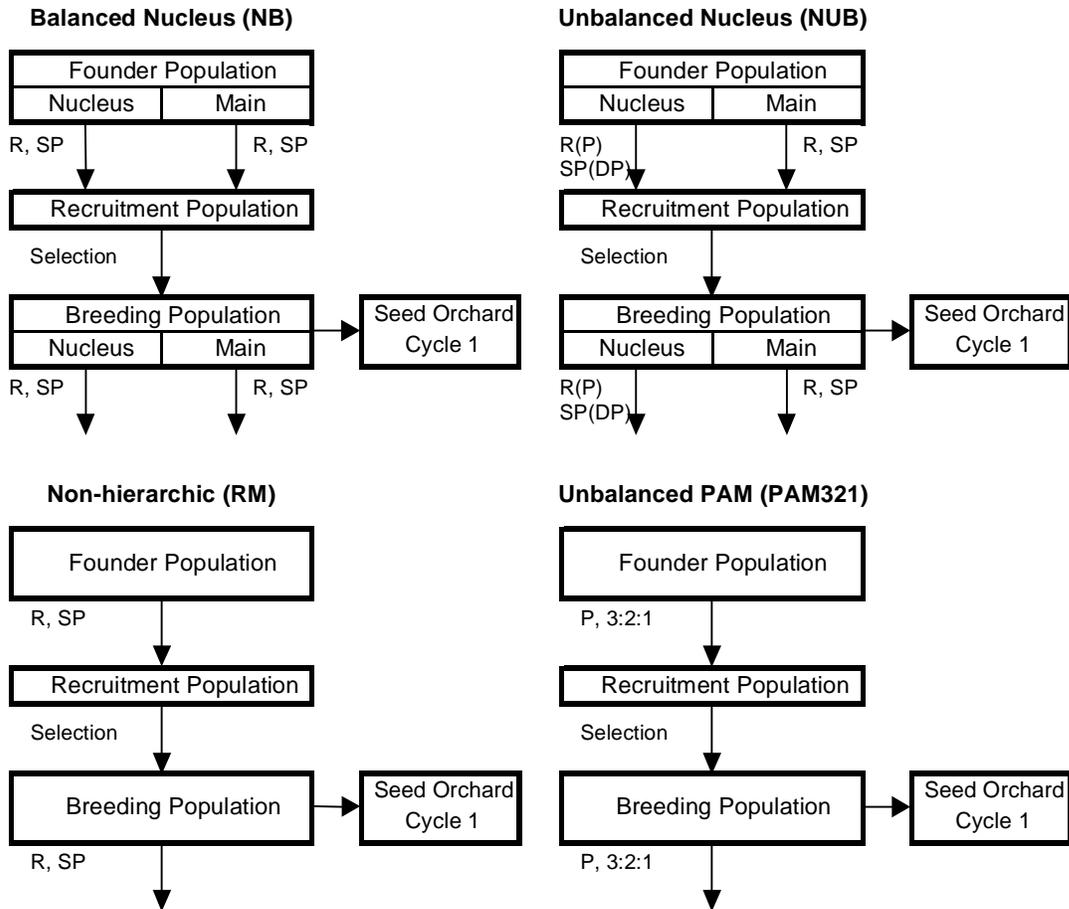
- Lindgren D, Mullin TJ (1997) Balancing gain and relatedness in selection. *Silvae Genet* 46:124–129
- Lstibůrek M, Mullin TJ, Lindgren D, Rosvall O (2004) Open-nucleus breeding strategies compared with population-wide positive assortative mating I. Equal distribution of testing effort. *Theor Appl Genet.* 109:1196-1203
- McKeand E, Beineke F (1980) Sublining for half-sib breeding populations of forest trees. *Silvae Genet* 29:14–17
- McKeand SE, Bridgwater FE (1998) A strategy for the third breeding cycle of loblolly pine in the southeastern US. *Silvae Genet* 47:223–234
- Mikola J (2002) Long-term tree breeding strategy in Finland: integration of seed production and breeding. In: Haapanen M, Mikola J (eds) Integrating tree breeding and forestry. (Proceedings of the Nordic group for management of genetic resources of trees, Mekrijärvi, March 2001. Research papers 842) Finnish Forest Research Institute, Mekrijärvi, pp 12–13
- Mullin TJ, Park YS (1995) Stochastic simulation of population management strategies for tree breeding: a new decision-support tool for personal computers. *Silvae Genet* 44:132–141
- Namkoong G, Kang HC, Brouard JS (1988) Tree breeding: principles and strategies. Monographs on theoretical and applied genetics. Springer, Berlin Heidelberg New York
- Roden JA (1994) Review of the theory of open nucleus breeding systems. *Anim Breed Abstr* 62:151–157

- Rosvall O, Mullin TJ (2003) Positive assortative mating with selection restrictions on group coancestry enhances gain while conserving genetic diversity in long-term forest tree breeding. *Theor Appl Genet* 107:629–642
- Rosvall O, Lindgren D, Mullin TJ (1999) Sustainability robustness and efficiency of a multi-generation breeding strategy based on within-family clonal selection. *Silvae Genet* 47:307–321
- Rosvall O, Mullin TJ, Lindgren D (2003) Controlling parent contributions during positive assortative mating and selection increases gain in long-term forest tree breeding. *For Genet* 10:35–53
- Ruotsalainen S, Lindgren D (2000) Stratified sublining: a new option for structuring breeding populations. *Can J For Res* 30:596–604
- Ruotsalainen S, Lindgren D (2001) Number of founders for a breeding population using variable parental contribution. *For Genet* 8:57–67
- Smith C (1988) Genetic improvement of livestock, using nucleus breeding units. *World Anim Rev* 65:2–10
- Wei RP, Lindgren D (1995) Optimal family contributions and a linear approximation. *Theor Popul Biol* 48:318–332
- White TL (1992) Advanced-generation breeding populations: size and structure. In: *Resolving tropical forest resource concerns through tree improvement, gene conservation and domestication of new species. (Proc IUFRO Conf Breeding Tropical Trees held in Cartagena and Cali, Colombia, 9–18 October 1992)*, pp 208–222

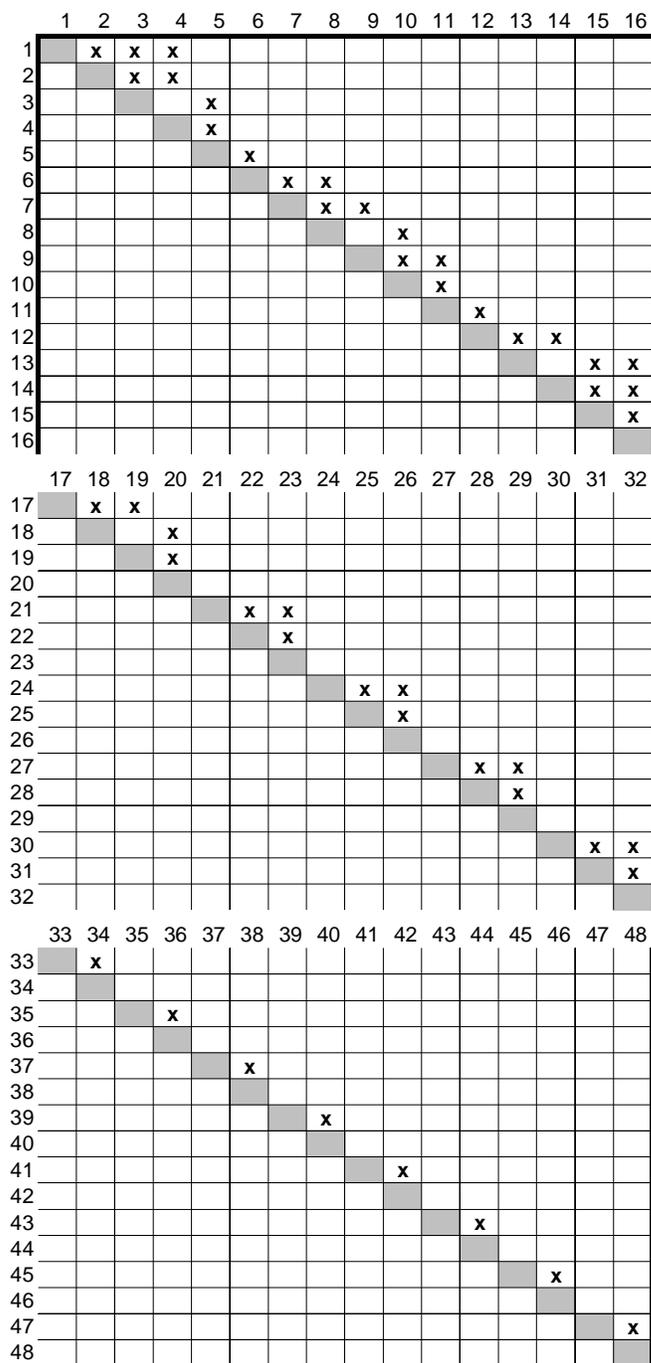
White TL, Hodge GR, Powell GL (1993) An advanced-generation tree improvement plan for slash pine in the southeastern United States. *Silvae Genet* 42:359–371

White TL, Matheson AC, Cotterill PP, Johnson GR, Rout AF, Boomsma DB (1999) A nucleus breeding plan for radiata pine in Australia. *Silvae Genet* 48:122–133

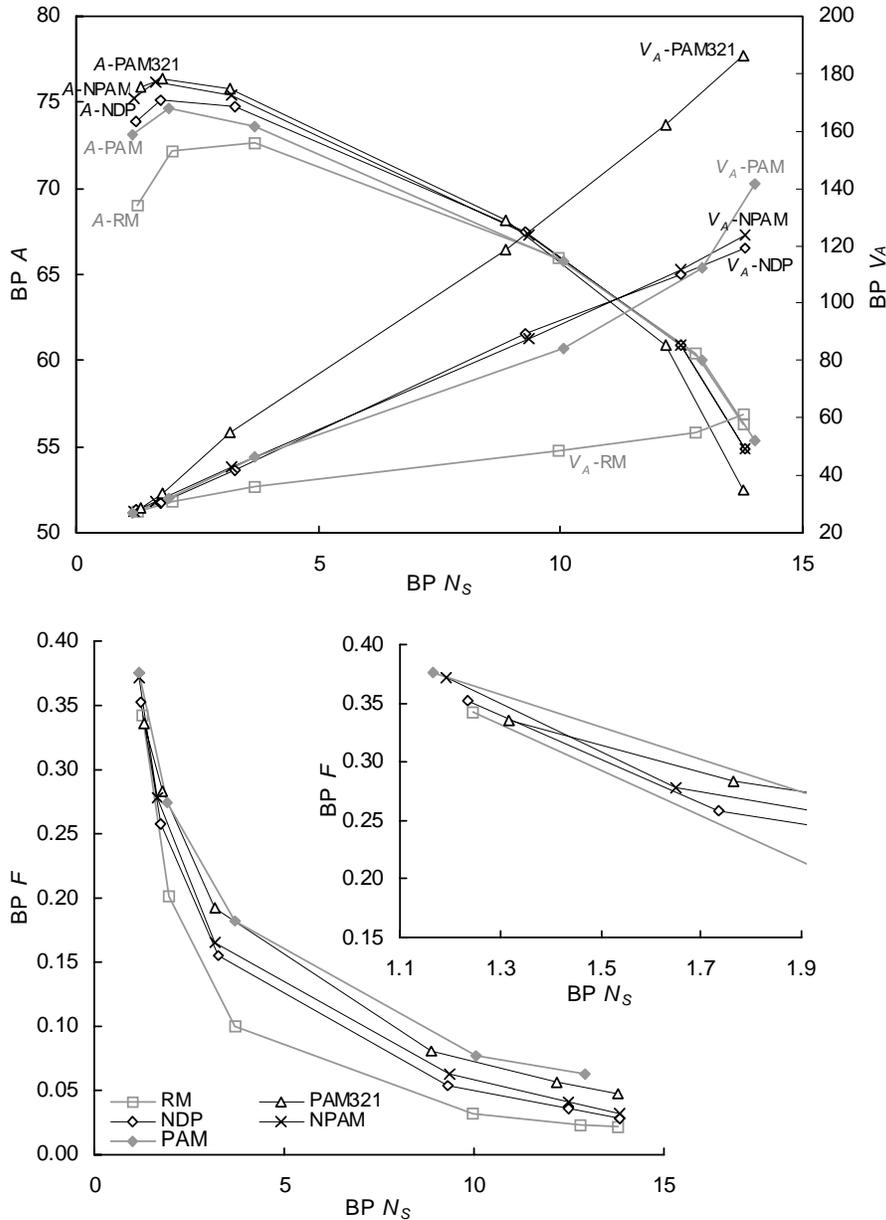
Willis MB (1998) Dalton's introduction to practical animal breeding. Blackwell, Malden



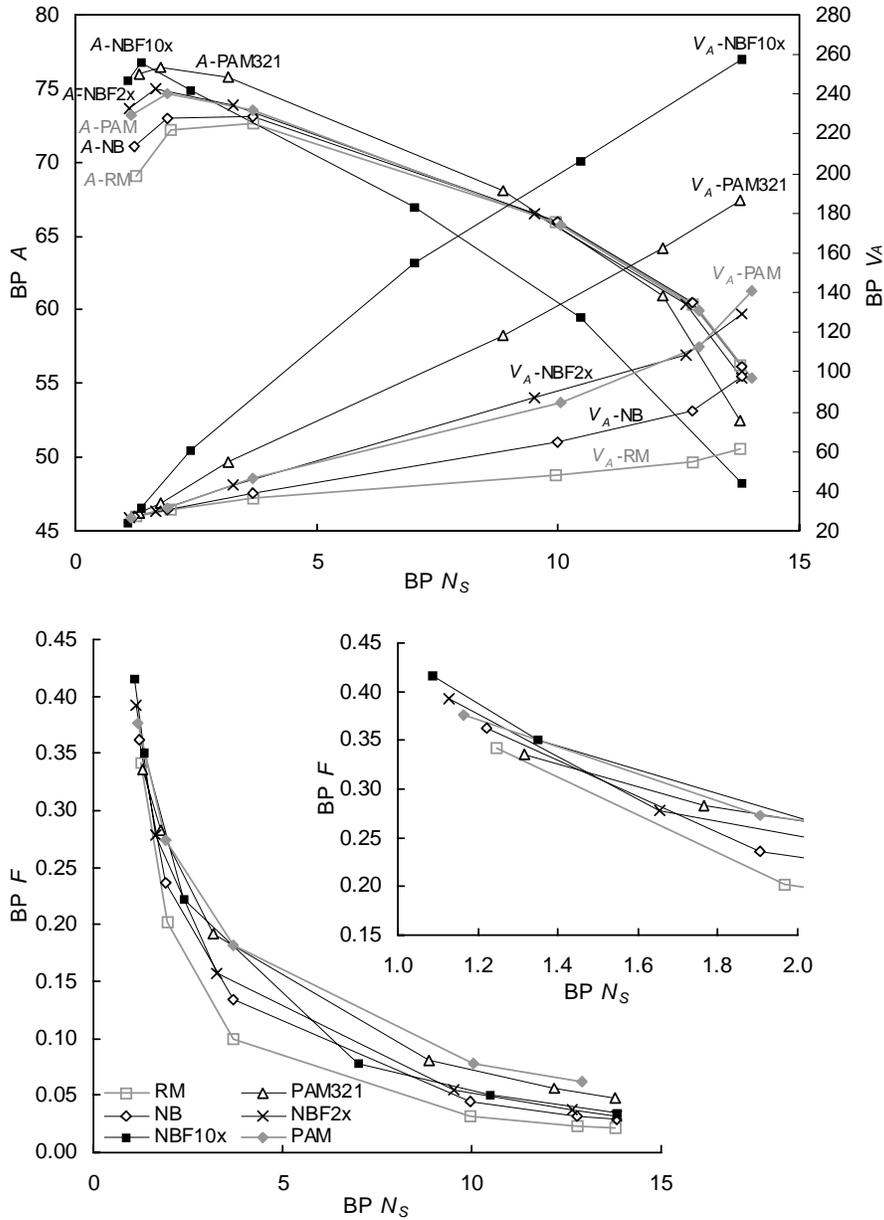
**Figure 1.** Breeding alternatives compared in this study. Mating was either random (R) or positive assortative (P) and either single-pair (SP), double-pair (DP), or 3:2:1 design.



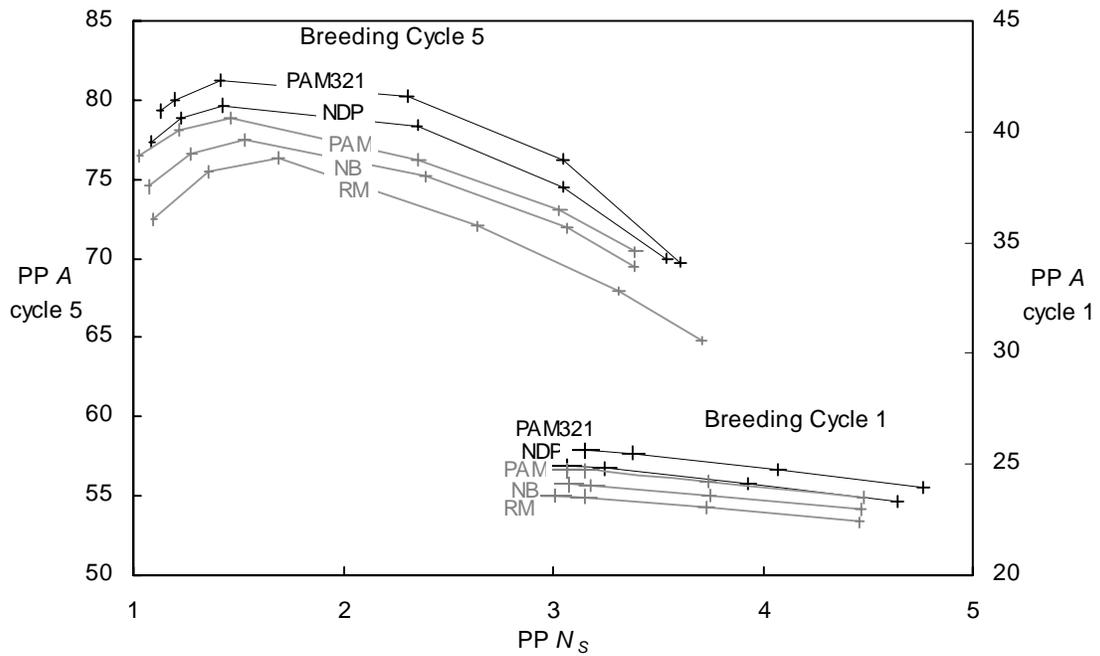
**Figure 2.** 3:2:1 mating design. Parents are denoted by their rank. Top ranking third of parents (1–16) were each involved in three crosses; middle ranking third of parents (17–32) in two crosses; the lowest ranking third of parents (33–48) in one cross.



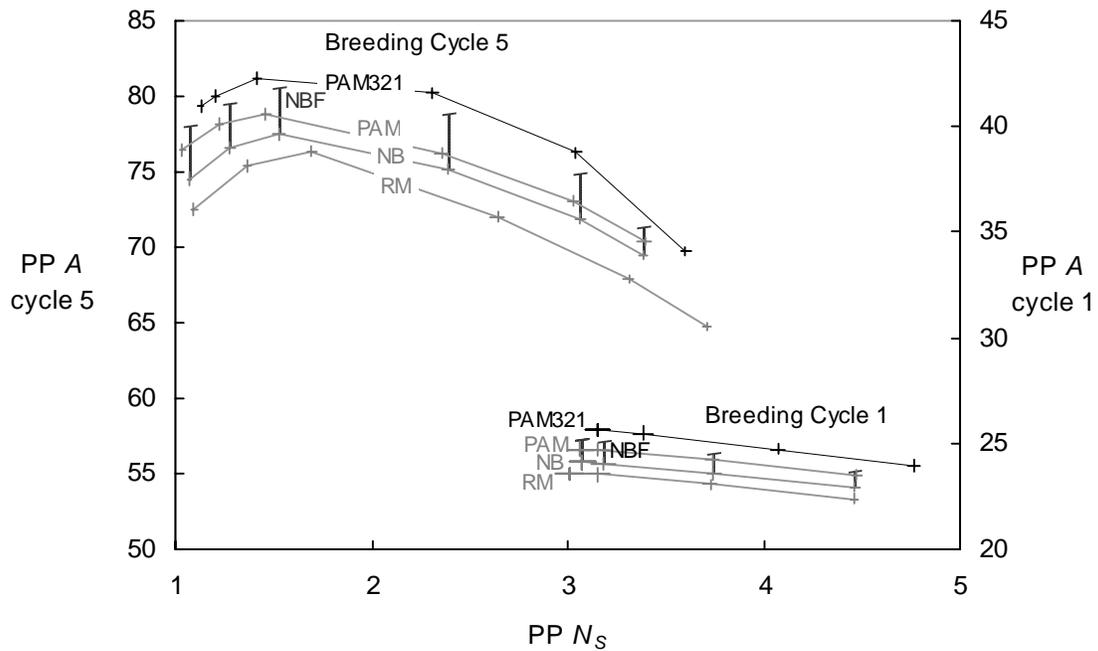
**Figure 3.** Upper figure shows average additive effect ( $A$ ) and additive variance ( $V_A$ ) in the breeding population (BP) after five breeding cycles for a set of scenarios with variable weight on group coancestry in the selection criterion resulting in different BP status numbers ( $N_S$ ). Lower figure shows corresponding average inbreeding coefficient ( $F$ ) in BP. The small figure shows the upper left corner of the diagram at higher resolution. The lines connect scenarios for nonhierarchic random mating (RM), unbalanced nucleus with double-pair mating (NDP), unbalanced nucleus with positive assortment of mates (NPAM), positive assortative mating (PAM), and unbalanced population-wide positive assortative mating (PAM321). Light lines depict comparison alternatives from the first part of this study (I. Equal distribution of testing effort).



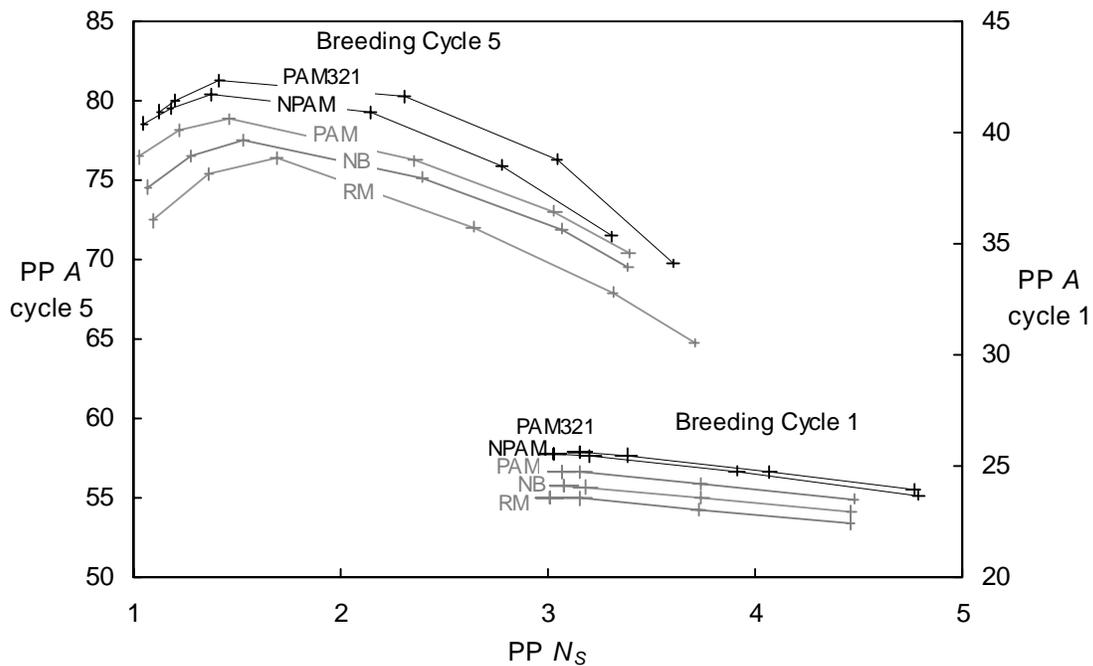
**Figure 4.** Upper figure shows  $A$  and  $V_A$  in the BP after five breeding cycles for a set of scenarios with variable weight on group coancestry in the selection criterion resulting in different  $BP N_s$ . Lower figure shows corresponding average  $F$  in BP. The small figure shows the upper left corner of the diagram at higher resolution. The lines connect scenarios for RM, balanced nucleus (NB), unbalanced nucleus with two times (NBF2 $\times$ ) and ten times (NBF10 $\times$ ) bigger families in nucleus, PAM, and PAM321. Light lines depict comparison alternatives from the first part of this study (I. Equal distribution of testing effort).



**Figure 5.**  $A$  in the seed orchard (PP) established after one and five breeding cycles for a set of scenarios with variable weight on group coancestry in the selection criterion resulting in different  $PP N_s$ . The lines connect scenarios for RM, NB, NDP, PAM, and PAM321. Light lines depict comparison alternatives from the first part of this study (I. Equal distribution of testing effort). Confidence intervals of 95% for  $PP A$  and  $PP N_s$  are based on 800 iterations of the simulation.



**Figure 6.** A in the PP established after one and five breeding cycles for a set of scenarios with variable weight on group coancestry in the selection criterion resulting in different PP  $N_S$ . Alternatives compared in this figure are RM, NB, PAM, and PAM321. Vertical lines depict space of solutions for studied range of larger family sizes for nucleus families (NBF). Light lines depict comparison alternatives from the first part of this study (I. Equal distribution of testing effort). Confidence intervals of 95% for PP A and PP  $N_S$  are based on 800 iterations of the simulation.



**Figure 7.** A in the PP established after one and five breeding cycles for a set of scenarios with variable weight on group coancestry in the selection criterion resulting in different PP  $N_s$ . The lines connect scenarios for RM, NB, NPAM, PAM, and PAM321. Light lines depict comparison alternatives from the first part of this study (I. Equal distribution of testing effort). Confidence intervals of 95% for PP  $A$  and PP  $N_s$  are based on 800 iterations of the simulation.

## **CHAPTER 4**

### **Positive assortative mating with family size as a function of parental breeding values**

Authors: M. Lstibůrek, T. J. Mullin, T. F. C. Mackay, D. Huber, and B. Li  
(submitted manuscript)

## **Abstract**

While other investigations have described benefits of positive assortative mating (PAM) for forest tree breeding, the allocation of resources among mates in these studies was either equal, or was varied using schemes corresponding only to parental rank (i.e., more resources invested in higher-ranking parents). In this simulation study, family sizes were proportional to mid-parent BLUP values. The distribution of mid-parent BLUP values was standardized by a constant, which was varied to study the range of distributions of family size. Redistributing progenies from lower- to higher-ranking families to a point where an equal number of progenies were still selected out of each family to the next generation caused minimal change in group coancestry and inbreeding in the breeding population (BP), while the additive genetic response and variance in the BP were both greatly enhanced. This generated additional genetic gains for forest plantations by selecting more superior genotypes from the BP (compared to PAM with equal family sizes) for production of improved regeneration materials. These conclusions were verified for a range of heritability under a polygenic model and under a mixed-inheritance model with a QTL contributing to the trait variation.

## **Introduction**

Earlier studies by computer simulation have demonstrated that positive assortative mating (PAM) applied in a long-term forest tree breeding program has the potential to generate extra genetic improvement in forest plantations [e.g., dissertations by Mahalovich (1990) and Rosvall (1999)]. This is due to the enhancement of the additive genetic response and variance in the breeding population (BP), enabling the selection of more extreme genotypes from the BP. These genotypes can be established in orchards to produce seeds for production of reforestation nursery stock, or alternatively can be vegetatively propagated and directly planted as clones. When genetic gain and diversity are both considered in a single selection criterion, PAM does not much alter the effective population size compared with that achieved under random mating, but increases the average inbreeding in the BP as a consequence of mating among more related individuals (Rosvall and Mullin 2003).

Ideally, the contribution of individuals in the BP to the next generation should correlate with their breeding values (Lindgren 1986). Various weighting schemes have been proposed to assign mating frequencies to individuals in the BP. These are built on different assumptions about the distribution of breeding values and the function used to assign the corresponding mating frequencies (Kang and Namkoong 1988; Kang 1989; Wei and Lindgren 1995).

Two approaches for controlling parental contributions during PAM were investigated by Rosvall et al. (2003), using a stochastic model of a forest-tree breeding program. In the first approach, balanced mating (each individual involved in an equal number of combinations) was followed by unbalanced selection (more progenies selected from higher-

ranking families and fewer from lower-ranking families). In the second, unbalanced mating (individuals of higher rank mated more frequently than lower-ranking trees) was followed by balanced within-family selection. Even though the second approach led to less additive genetic response in the BP, this was overcompensated by a larger expansion of additive variance, resulting in greater genetic gains from the very best genotypes deployed in plantations. In the unbalanced mating scheme, individuals in the BP were ranked by breeding value and subdivided into three distinct hierarchical groups of equal size. The number of mating combinations per individual within a group varied among groups: more in the best group, and fewer in the lowest-ranking group. Unbalanced mating followed by either balanced or unbalanced selection was also advocated for forest tree breeding as a better alternative to nucleus-breeding strategies (Lstibůrek et al. 2004).

A weakness of these earlier studies is that parental contributions did not follow the actual distributional patterns of the predicted breeding values. The number of matings assigned to each individual followed its rank, rather than its predicted breeding value, which imposes an unrealistic assumption that the correlation between rank and breeding value is one. Furthermore, the number of combinations was constant within each hierarchical group, such that the highest-ranking individual was involved in the same number of combinations as the lowest-ranking member of the same group.

The objective of this study was to dynamically control parental contributions, allowing the importance assigned to each mate be guided by a linear function of predicted BLUP values. In this study, we quantify the effect of this approach on genetic response and diversity in the BP, and on the actual benefit for forest plantations derived from the BP. We

also test the sensitivity of the approach with respect to a single major-gene locus contributing to variation in the quantitative trait of interest.

## Methods

The computer simulation program POPSIM<sup>TM</sup> (Mullin and Park 1995) was modified as described below. The situation modeled represents a single breeding population managed over six discrete (non-overlapping) generations (Figure 1). The strategy was based on the actual breeding program implemented for Norway spruce [*Picea abies* (L.) Karst.] in Sweden (Karlsson and Rosvall 1993), although some of the components were modified or eliminated to provide general recommendations over a wider range of forest tree breeding programs. In many such programs, the genetic component of observed variation in a quantitative trait is attributable primarily to additive gene action, and the basic breeding strategy employed based on “recurrent selection for general combining ability” (Namkoong et al. 1988).

### Genetic model

A single quantitative trait was described by the genetic model for a population of diploid, monoecious individuals, with an initial population mean of 100. The trait was influenced by a large number of loci, segregating independently, each with a small effect (polygenic model). In the polygenic model, the phenotypic value ( $P$ ) of each individual was composed of an independent additive polygenic ( $A_p$ ) component and an environmental deviation ( $E$ ), such that  $P = A_p + E$ . Correspondingly, the phenotypic variance ( $V_p$ ) was  $V_p = V_{A_p} + V_E$ . The  $A_p$  was adjusted for individual sets of simulation scenarios such that initial narrow-sense heritability  $h^2$  was 0.1, 0.3, or 0.5, while initial  $V_p$  remained constant at 500 (Rosvall et al. 1999).

## Founder Population

Forty eight founder genotypes were sampled randomly from a population of unrelated, non-inbred individuals. The additive effect was sampled from  $\text{Normal}(0, V_{AP})$  and the environmental deviation was sampled from  $\text{Normal}(100, V_E)$ . Each founder was progeny tested with a pollen mix (polycross test). To generate this test, each founder was mated at random to a common pool of 20 unique unrelated, non-inbred individuals (representing the mixture of tester pollen). These individuals were sampled from the same distribution of effects as the founder population. One hundred and fifty polycross-test progenies were generated for each founder. The additive effect for each test progeny was randomly sampled from  $\text{Normal}(a_{FS}, 0.5V_{AP})$ , where  $a_{FS}$  is the mid-parent additive effect. The environmental effect for each progeny genotype was drawn randomly from  $\text{Normal}(100, V_E)$ . The best linear unbiased prediction (BLUP) value of each founder was then calculated from the test using the animal model (Mrode 1996; Lynch and Walsh 1998) implemented in the ASReml software package (Gilmour et al. 2002). In this totally balanced case of unrelated founders, the initial breeding values could be as well calculated as simple deviations from family means. In later generations (following the introduction of imbalance), founders must also be included in the genetic evaluation. Therefore, to simplify programming, we incorporated generalized BLUP analysis from the initial generation. Founders were then sorted by BLUP values and mated in a single-pair mating scheme (total number of families  $N_f = 24$ ) in order to generate progenies that are selection candidates for the breeding population in the next generation.

### **Derivation of family sizes**

The number of full-sib progenies in each family (family size) was determined as a linear function of the mid-parent BLUP value (average of parental BLUP values). Mid-parent BLUP values were standardized to the arithmetic mean  $a_A$ . Following this standardization, all negative values were converted to zero and all positive values were converted to a relative scale. Family sizes were then determined as a product of the recruitment population size  $N_{RP}$  (total number of test progenies within each generation, equal to 720) and these relative values. The average family size was equal to 30, i.e., 720 progeny genotypes / 24 families. The variance in family sizes ( $V_f$ ) can have two extremes: (1), all family sizes are equal to 30 ( $V_f = 0$ ), and (2), all progenies (720) are generated in the top-ranking family, and all remaining families have family size zero ( $V_{fmax} = N_{RP}^2 / N_f = 21,600$ ). The desired proportion of variance in family sizes relative to the maximum variance  $V_f / V_{fmax}$  is then bounded within the interval [0, 1]. This desired proportion was specified as a simulation input parameter and was varied in individual simulation scenarios to cover the entire interval (Table 1). Parameter  $a_A$  was then determined iteratively from  $V_f / V_{fmax}$  within each generation, according to the actual distribution of mid-parent BLUP values.

### **Generation of recruitment progenies**

The additive effect of each progeny was randomly sampled from  $\text{Normal}\{a_{FS}, 0.5V_{Ap}[1-0.5(F_f + F_m)]\}$ , where  $F_f$  and  $F_m$  are inbreeding coefficients of a female and male parents, respectively. Each progeny genotype was replicated by eight clonal copies

(ramets), where the environmental effect for each ramet was drawn randomly from  $\text{Normal}(100, V_E)$ . A constant testing environment was assumed throughout all generations, and the cause of all variability among ramets was assumed to be environmental.

### **Selection**

After generation of the recruitment population, input files for the ASReml software were updated (to include information on all individuals in the pedigree, including founders, their polycross progenies and all selection candidates in all generations). ASReml was then run to predict BLUP values of selection candidates (current recruitment population). Group-merit selection (Lindgren and Mullin 1997) was then performed to select the next-generation breeding population from the recruitment population. The selection algorithm maximized iteratively the population merit:  $B_\omega = \bar{g}_\omega - c\Theta_\omega$ , where  $B_\omega$  is the group merit of a selected set  $\omega$ ;  $\bar{g}_\omega$  is the average BLUP value of the set;  $\Theta_\omega$  is the group coancestry of the set; and  $c$  is a weighting constant. Group coancestry is the average of all pair-wise coancestries, including self-coancestry and reciprocals. It is the probability that two genes taken at random from the gene pool, with replacement, are identical by descent (Cockerham 1967). The weighting constant was set to a very large value, forcing the group-merit selection algorithm to minimize group coancestry.

### **Production population**

Genetically improved planting stock is commonly derived from a production population, such as a seed orchard, selected as a sub-set of the breeding population. The six trees with

highest BLUP values irrespective of their coancestry were selected for the production population. The purpose was to test the ability of the breeding population to support a production population (the breeder's target) and to determine the proportional gene diversity of these best clones.

### **Mixed-inheritance model**

Additional simulation scenarios were run using a “mixed-inheritance” model. The purpose of this model was to investigate whether the main conclusions of this study would be significantly altered by the presence of a single, biallelic, major-gene locus, contributing up to 10% of the additive genetic variance in the quantitative trait. The simulation was implemented as described by Gomez-Raya and Klemetsdal (1999). The phenotypic value ( $P$ ) was composed of three independent components  $P = A_M + A_p + E$ , where  $A_M$  is the additive effect of a major gene,  $A_p$  is the additive effect due to a large number of polygenic loci, each with a small effect (polygenic background), and  $E$  is the residual deviation. The phenotypic variance ( $V_p$ ) can be expressed as  $V_p = V_{A_M} + V_{A_p} + V_E$ , where  $V_{A_M}$  and  $V_{A_p}$  are the additive variances (referring to the base unselected population) due to the major-gene component [ $V_{A_M} = 2p(1-p)\alpha^2$ ] and the polygene component, respectively ( $p$  is the initial gene frequency of allele  $M_1$  at the major-gene locus). The effect of a major gene ( $\alpha$ ) was

calculated as:  $\alpha = \sqrt{\frac{\lambda h^2 V_p}{2p(1-p)}}$ , where  $\lambda$  is the proportion of the additive variance explained

by the major-gene component [ $\lambda = V_{A_M} / (V_{A_M} + V_{A_p})$ ],  $h^2$  is the narrow-sense heritability

$[h^2 = (V_{A_M} + V_{A_P})/V_P]$ . Simulation scenarios were run at  $\lambda = 0.1$ , where it was assumed that  $\lambda$  was constant through all generations.  $V_{A_P}$  was calculated as  $V_{A_P} = h^2V_P - V_{A_M}$ . To generate the founder population, alleles at the major-gene locus were sampled randomly from the allelic pool with a frequency of the  $M_1$  allele  $p = 0.1$ . The additive effect of the major gene was then  $\alpha$  for  $M_1M_1$  and  $-\alpha$  for  $M_2M_2$  genotypes. To generate progenies, alleles at the major-gene locus were sampled from parental genotypes with the probability 0.5.

### **Simulation and evaluation of results**

The process described above was repeated over six generations for each simulation scenario. Each scenario (unique variance in family sizes, presented in Table 1) was replicated by 400 independent runs (iterations). Parametric means across all iterations were calculated for each scenario along with 95% confidence intervals. The random number generator used in this simulation was “MRG32k3a” (L’Ecuyer et al. 2002). Parametric means for each scenario in generation six (average additive response, variance and inbreeding) were plotted as a function of the resulting group coancestry in both the breeding and production populations.

## Results and Discussion

In this study, the redistribution of test resources among families by varying their sizes under fixed total test resources in this study magnified the effect of PAM, which by itself provides a large enhancement of  $V_{A(BP)}$ , particularly under low selection intensity scenarios (Baker 1973; De Lange 1974; Jorjani 1995; Rosvall and Mullin 2003). The enhancement of  $V_{A(BP)}$  observed in this study exceeds the magnitude reported earlier under unbalanced mating schemes (Rosvall et al. 2003; Lstibůrek et al. 2004). This conclusion holds under infinitesimal model assumptions, as well as under the mixed-inheritance model with a single major-gene locus, contributing 10% of the additive genetic variance.

The average distribution of family sizes in the fifth generation for scenarios one through nine (as presented in Table 1) is depicted in Figure 2. This general trend in the distribution of family sizes was also observed in earlier generations. For clarity in further discussion, scenarios one through five are referred to as “balanced” and all remaining scenarios (six through nine) as “unbalanced”. The term “balance” in this context refers to balanced within-family selection (equal number of individuals selected from each family), as opposed to balanced distribution of test resources. Balanced within-family selection was facilitated by a high weighting on  $\Theta$  during group-merit selection.

The reallocation of test effort among families had a significant effect on the BP.  $\Theta_{BP}$  increased as a result of redistributing more progenies to higher-ranking families (Figure 3). The marginal increase in  $\Theta_{BP}$  under balanced scenarios ( $0.0353 \leq \Theta_{BP} \leq 0.0463$ ) agrees with the finding of Rosvall et al. (2003) that keeping equal parent contributions by means of balanced mating and selection maintained the lowest  $\Theta_{BP}$ . After this point was exceeded

(scenario six), some parents did not contribute to the next generation (family size equal to 0), which resulted in more progenies being selected from fewer remaining families and consequently in a more rapid increase of  $\Theta_{BP}$  (approx.  $0.0463 < \Theta_{BP} \leq 0.6745$ ). Balanced scenarios also resulted in a much lower increase in  $\Theta_{BP}$  at generation shifts, compared to more unbalanced scenarios ( $\Theta_{BP} = 0.0104$  in the first generation).  $\Theta_{BP}$  was not significantly altered when a mixed-inheritance model was considered (dotted lines in Figure 3).

The average inbreeding in the BP ( $F_{BP}$ ) in generations one (founder population) and two was equal to 0. Following the mating of individuals in the second-generation breeding population,  $F_{BP}$  reached a value of approximately 0.05 in the third generation under balanced scenarios and up to 0.25 in scenario nine. The increase in  $F_{BP}$  when progressing through generations was again more pronounced under unbalanced scenarios and minimal under balanced scenarios. A small reduction in  $F_{BP}$  when progressing from scenarios one through five (visible in Figure 3) is an interesting consequence of the assortment of mated individuals with no avoidance of mating among full-sibs, as performed in this study. The likelihood of mating among selected full-sibs is reduced to a certain point (compare to the completely balanced scenario) as a result of smaller sizes of lower-ranking families and increased variation among the top two selections from these families. Although  $F_{BP}$  was marginally higher in the mixed-inheritance model, the difference was not significant.

The average additive effect in the BP ( $A_{BP}$ ) increased considerably under balanced scenarios ( $A_{BP} = 60.92$  in scenario one up to  $A_{BP} = 69.17$  in scenario five) and even more under unbalanced scenarios. The general increase in genetic response due to PAM is expected from the theory, because PAM induces gametic-phase disequilibrium, i.e., expands

additive variance (the among-family component), which creates opportunities for additional genetic response to selection (e.g., Crow 1986). In this model, the added imbalance induces a positive correlation between the expected family value and corresponding family size. Thus, higher selection differentials are achieved within families of higher expected (mid-parent additive) values. This induces further expansion of additive variance in generations that follow, but also more efficient conversion of variance into genetic gains. The trend for the increase of  $A_{BP}$  due to added imbalance up to the sixth generation was very similar to that in earlier generations. Interestingly,  $A_{BP}$  in earlier generations under unbalanced scenarios reached, and potentially exceeded, gains in later generations generated under more balanced scenarios. Thus, for example in the fourth generation,  $A_{BP}$  was equal to 69.98 (scenario seven), exceeding  $A_{BP}$  in the sixth generation ( $A_{BP} = 69.1654$  in scenario five). This, of course, assumes complete absence of inbreeding depression, which could adversely affect the additive response under unbalanced scenarios, depending upon the distribution and magnitude of the depression (Williams and Savolainen 1996).

The additional expansion of additive variance in the BP ( $V_{A(BP)}$ ) due to redistribution of family sizes was most pronounced under balanced scenarios (one through five). The maximum enhancement of  $V_{A(BP)}$  was observed in scenario four ( $V_{A(BP)} = 444.64$ , compared with  $V_{A(BP)} = 278.70$  in scenario one); this trend was true in all generations, starting in the second. Added imbalance (higher  $V_f$  in scenarios six through nine) caused a sharp reduction in  $V_{A(BP)}$ . In these unbalanced scenarios, some parents would not contribute to subsequent generations, and therefore, the variation of expected family means sharply drops for each unit of increase in  $V_f$  (Bulmer 1985). Larger variation in  $V_{A(BP)}$  among individual simulation

iterations (assessed by confidence intervals) was observed at lower values of  $\Theta_{BP}$ , while greater run-to-run similarity was achieved at a higher  $\Theta_{BP}$ . Simulation showed that the use of clonal replication in progeny testing was the primary factor giving rise to more accurate assessment of breeding values, facilitating high within-family selection differentials and a stronger effect by PAM on expansion of additive variance. The effect of clonal testing was more important than the inclusion of the multi-generation relationship data set to the genetic evaluation by BLUP. This supports the benefits of clonal assessment in progeny testing for forest tree breeding programs described by other authors (e.g., Shaw and Hood 1985; Russell and Loo-Dinkins 1993; Mullin and Park 1994; Danusevicius and Lindgren 2002; Isik et al. 2003).

There was a significant effect from the major gene on both  $A_{BP}$  and  $V_{A(BP)}$  (Figure 4, dotted lines).  $A_{BP}$  was up to 10% lower (the difference was similar in all scenarios), while  $V_{A(BP)}$  was up to 20% higher in the mixed-inheritance model. The difference in  $V_{A(BP)}$  was significant only for more balanced scenarios (the differences in scenarios seven through nine were not significant). The frequency of the favorable allele in the BP for each generation is presented in Figure 4. Higher  $V_{A(BP)}$  is due to the presence of a segregating major-gene locus with an effect that is substantially higher than individual effects of remaining polygenic loci. Elevated  $V_{A(BP)}$  under balanced scenarios is then due to the variation of allelic frequencies at this locus responding to selection and drift (e.g., Falconer and Mackay 1996). The probability of eventual fixation of the favorable allele was higher in scenarios with greater imbalance (Figure 4), thus the difference in  $V_{A(BP)}$  due to the major gene locus almost disappeared with added imbalance. Since information on the major gene was not extracted (e.g., by genetic

marker analysis) and incorporated into the genetic evaluation (which may actually happen in real breeding programs, particularly for QTLs associated with smaller effects), and since there was zero correlation between the additive value due to a major-gene locus and the corresponding value due to polygenic loci, there was no additional genetic response due to the presence of a major gene.

Narrow-sense heritability ( $h^2$ ) is one of the main factors influencing the effectiveness of PAM (Falconer and Mackay 1996; Lynch and Walsh 1998) and its impact was further enhanced in this study by the use of clonal replication in progeny testing. Results described so far were obtained for scenarios where the initial value of  $h^2$  was set to 0.3. The maximum observed enhancement of  $V_{A(BP)}$  when  $h^2 = 0.3$  (in scenario four) was 157% of that in the completely balanced scenario (scenario one). The corresponding maximum enhancement was 133% when  $h^2 = 0.1$  and 180% when  $h^2 = 0.5$ . For the range of  $h^2$  studied, the maximum absolute value of  $V_{A(BP)}$  achieved in scenario four ( $V_{A(MAX)}$ ) can be described as a linear function of  $h^2$  ( $V_{A(MAX)} = 1563.5h^2 - 26.707$ ) with a high confidence ( $R^2 = 0.9998$ ). As for  $V_{A(BP)}$ ,  $A_{BP}$  also increased with  $h^2$  ( $A_{BP} = 24.45$  when  $h^2 = 0.1$ ,  $A_{BP} = 60.92$  when  $h^2 = 0.3$ , and  $A_{BP} = 89.40$  when  $h^2 = 0.5$ ). The increase in  $A_{BP}$  at the limit of balanced scenarios (scenario five) was 117% ( $h^2 = 0.1$ ), 114% ( $h^2 = 0.3$ ), and 111% ( $h^2 = 0.5$ ) of that in the completely balanced scenario (scenario one).  $F_{BP}$  presented in Figure 3 was not significantly altered by lower ( $h^2 = 0.1$ ) or higher ( $h^2 = 0.5$ ) initial  $h^2$  values.

Given that only the effect of PAM on the enhancement of  $V_{A(BP)}$  would be of interest to the breeder, this variable family-size approach would seem very attractive. Nevertheless, this extra enhancement of variance in the BP increases the frequency of mating among selected

full-sibs, following the assortment of the BP, which of course has an adverse effect on  $F_{BP}$ . This consequently lowers the within-family portion of additive variance, resulting in reduced intensity of within-family selection and an unavoidable reduction in additive response in the BP.

Similar to the BP,  $\Theta_{PP}$  was only slightly influenced by the variation in family sizes up to the limit of balanced scenarios (refer to Figure 2, scenarios one through five).  $\Theta_{PP}$  increased more progressively at higher levels of imbalance (Figure 5). The average additive effect (genetic gain) in the PP ( $A_{PP}$ ) responded greatly to the increase in variance in family sizes;  $A_{PP}$  increased from 85.2 in scenario one, to 96.1 in scenario five, and up to 102.5 in scenario six. In practical terms, this means that the reallocation of test effort in the BP resulted in an extra 12.7% and 20.3%  $A_{PP}$  at  $h^2 = 0.3$ , in scenarios five and six, respectively. Although higher values of  $A_{PP}$  are observed in scenarios six through nine, these are of less practical interest due to the rapid increase of  $\Theta_{BP}$  and  $F_{BP}$ . At  $h^2 = 0.1$ , there was 11.9% and 21.8% additional  $A_{PP}$ , in scenarios five and six, respectively. At  $h^2 = 0.5$ , the additional  $A_{PP}$  was 13.5% and 21.1%, in scenarios five and six, respectively. Thus, for the entire range of  $h^2$  studied, there was over 11% extra  $A_{PP}$  in scenario five and over 20% extra  $A_{PP}$  in scenario six, due to the reallocation of testing effort in BP, with minimal increase of  $\Theta_{BP}$  and  $F_{BP}$ . These results hold also for the mixed-inheritance model, where the advantage in  $A_{PP}$  due to added imbalance was over 11% in scenario five and over 18% in scenario six. In general, lower gains in the PP in the presence of a major gene are due to the lower  $A_{BP}$ .

Alternatively, other forms of deployment of superior genetic material selected from BP could be considered (e.g., deployment of full-sib families or clones). Our attempt though was

to provide more general description of the accumulated additive response and the average relatedness among these best clones, without imposing assumptions on the actual form of their deployment. In reality, a number of populations could contribute to any given production population (McKeand and Beineke 1980; Rosvall et al. 1999; Ruotsalainen and Lindgren 2000), resulting in a mixture of selections with a higher census number and with a lower level of relatedness than presented in this study.

An additional comparison was performed with the unbalanced mating strategy with equal family sizes (UM) proposed by Rosvall et al. (2003). The variable-family-size approach presented in this study did not outperform UM strategy. In the PP, both  $A_{PP}$  and  $\Theta_{PP}$  were of a similar magnitude. We attribute this similarity primarily to a lower frequency of mating among selected full-sibs in the UM strategy (within the distinct hierarchical groups), as all families are of equal size, resulting in greater similarity among selected groups of full-sibs, and consequently greater dispersion of individuals, following their assortment. Thus, even though the enhancement of  $V_{A(BP)}$  was more efficient using the variable-family-size approach and lower in the UM strategy, the lower  $F_{BP}$  in the latter led to a smaller reduction in the within-family portion of additive variance, resulting in about 4% extra additive response in BP. On the other hand, the variable-family-size approach led to a lower value of  $\Theta_{BP}$ , by retaining a larger share of lower-ranking individuals in the BP contributing to subsequent generations (Figure 2, scenario six).

This study suggests that allocating resources during PAM according to mid-parent breeding values may greatly enhance genetic gains in forest plantations while causing a minimal increase of  $\Theta_{BP}$  and  $F_{BP}$ . To reach a reasonable point of balance between  $\Theta_{BP}$ ,  $F_{BP}$

and  $A_{PP}$  in the BP, we suggest minimizing the parameter  $a_A$  (i.e., maximizing  $V_f$ ) under a restriction where an equal number of progenies are selected from each family (balanced within-family selection; the size of lowest-ranking family is at least two under single-pair mating). Reducing  $a_A$  below this limit would provide an additional boost of  $A_{PP}$ , but would also be accompanied by a much greater increase in  $\Theta_{BP}$  and  $F_{BP}$  per unit of increase of  $A_{PP}$ .

The group-merit selection method used in this study was constrained to maximize gene diversity in the BP, and genetic progress was primarily a consequence of strong within-family selection, facilitated by clonal replication of the recruitment population. Future research may demonstrate a larger genetic response per unit of gene diversity when this selection constraint on BP diversity is slightly relaxed (Rodríguez 2000).

## **Acknowledgements**

We gratefully acknowledge financial support from the Föreningen Skogsträdsförädling (Swedish Tree Breeding Association), CellFor Inc., and the North Carolina State University-Industry Cooperative Tree Improvement Program. We thank Dr. Dag Lindgren and Dr. Ola Rosvall for their valuable comments during the development of the study, and Dr. Arthur Gilmour, Dr. Gary Hodge and Salvador Gezan for their generous assistance with the ASReml software.

## References

- Baker RJ (1973) Assortative mating and artificial selection. *Heredity* 31:231-238
- Bulmer MG (1985) *The mathematical theory of quantitative genetics*. Clarendon Press, Oxford
- Cockerham CC (1967) Group inbreeding and coancestry. *Genetics* 56:89-104
- Crow JF and Kimura M (1970) *An introduction to population genetics theory*. Burgess Publishing Company, Minneapolis
- Danusevicius D and Lindgren D (2002) Efficiency of selection based on phenotype, clone and progeny testing in long-term breeding. *Silvae Genet* 51:19-26
- Falconer DS and Mackay TFC (1996) *Introduction to quantitative genetics*. Longman, Harlow
- Gilmour AR, Gogel BJ, Cullis BR, Welham SJ, and Thompson R (2002) *ASReml User Guide Release 1.0* VSN International Ltd, Hemel Hempstead, HP11ES, UK
- Gomez-Raya L and Klemetsdal G (1999) Two-stage selection strategies utilizing marker-quantitative trait locus information and individual performance. *J Anim Sci* 77:2008-2018
- Isik F, Li B, Frampton J, and Goldfarb B (2003) Efficiency of seedlings and rooted cuttings for testing and selection in *Pinus taeda*. *For Sci* 50:44-53
- Jorjani H (1995) *Genetic studies of assortative mating in selected and unselected populations*. PhD thesis, Swedish University of Agricultural Sciences, Uppsala
- Kang H (1989) Inbreeding effective population size under some artificial selection schemes. II. Normal distribution of breeding values. *For Sci* 35:303-318

- Kang H and Namkoong G (1988) Inbreeding effective population size under some artificial selection schemes. I. Linear distribution of breeding values. *Theor Appl Genet* 75:333-339
- Karlson B and Rosvall O (1993) Breeding programmes in Sweden 3. Norway spruce. *In*: S. J. Lee (ed.) Progeny testing and breeding strategies. Proceedings of the Nordic Group of Tree Breeding, Oct. 1993, Edinburgh. Forestry Commission. pp. 135-141
- De Lange AO (1974) A simulation study of the effects of assortative mating on the response to selection. *In*: Garsi (ed.) First World Congress on Genetics Applied to Livestock Production, 7-11 October 1974, Madrid. Vol. III., pp. 421-425
- L'Ecuyer P, Simard R, Chen EJ, and Kelton WD (2002) An objected-oriented random-number package with many long streams and substreams. *Operations Research* 50:1073-1075
- Lindgren D (1986) How should breeders respond to breeding values? *In*: Proc. IUFRO Conf. Joint Meet. Working Parties on Breeding Theory, Progeny Testing, Seed Orchards, Oct 13-17, Williamsburg, VA. N.C. State University-Industry Cooperative Tree Improvement Program, Raleigh, NC. pp. 361-371
- Lindgren D and Mullin TJ (1997) Balancing gain and relatedness in selection. *Silvae Genet* 46:124-129
- Lstibůrek M, Mullin TJ, Lindgren D, and Rosvall O (2004) Open-nucleus breeding strategies compared to population-wide positive assortative mating. II. Unequal distribution of testing effort. *Theor Appl Genet* 109:1169-1177

- Lynch M and Walsh B (1998) Genetics and analysis of quantitative traits. Sinauer Assoc., Inc., Sunderland, Massachusetts
- Mahalovich MF (1990) Modelling positive assortative mating and elite populations in recurrent selection programs for general combining ability. PhD thesis, North Carolina State University, Raleigh
- McKeand S and Beineke F (1980) Sublining for half-sib breeding populations of forest trees. *Silvae Genet* 29:14-17
- Mrode RA (1996) Linear models for the prediction of animal breeding values. CAB International. Wallingford, Oxon, UK
- Mullin TJ and Park YS (1994) Genetic parameters and age to age correlations in clonally replicated test of black spruce. *Can J For Res* 24:2330-2341
- Mullin TJ and Park YS (1995) Stochastic simulation of population management strategies for tree breeding: a new decision-support tool for personal computers. *Silvae Genet* 44:132-141
- Namkoong G, Kang HC, and Brouard JS (1988) Tree breeding: principles and strategies. Monographs on theoretical and applied genetics 11. Springer-Verlag, New York
- Rodríguez LS (2000) Balanced vs. slightly unbalanced selection. *Heredity* 84:685-691
- Rosvall O (1999) Enhancing gain from long-term forest tree breeding while conserving genetic diversity. PhD thesis, Swedish University of Agricultural Sciences, Umeå
- Rosvall O, Lindgren D, and Mullin TJ (1999) Sustainability robustness and efficiency of a multi-generation breeding strategy based on within-family clonal selection. *Silvae Genet* 47:307-321

- Rosvall O and Mullin TJ (2003) Positive assortative mating with selection restrictions on group coancestry enhances gain while conserving genetic diversity in long-term forest tree breeding. *Theor Appl Genet* 107:629-642
- Rosvall O, Mullin TJ, and Lindgren D (2003) Controlling parent contributions during positive assortative mating and selection increases gain in long-term forest tree breeding. *For Genet* 10:35-53
- Ruotsalainen S and Lindgren D (2000) Stratified sublining: a new option for structuring breeding populations. *Can J For Res* 30:596-604
- Russell JH and Loo-Dinkins JA (1993) Distribution of testing effort in cloned genetic tests. *Silvae Genet* 42:98-104
- Shaw DV and Hood JV (1985) Maximizing gain per effort by using clonal replicates in genetic tests. *Theor Appl Genet* 71:392-399
- Wei R-P and Lindgren D (1995) Optimal family contributions and a linear approximation. *Theor Pop Biol* 48:318-332
- Williams CG and Savolainen O (1996) Inbreeding depression in conifers: implications for breeding strategy. *For Sci* 42:102-117

**Table 1.** Description of individual simulation scenarios (one through nine). Each scenario corresponds to a specified variance in family sizes ( $V_f$ ) relative to the maximum possible variance in family sizes ( $V_{fmax}$ ). Total number of progenies ( $N_{RP}$ ) was equal to 720 in all scenarios. The distribution of family sizes was varied, bounded by two extremes:

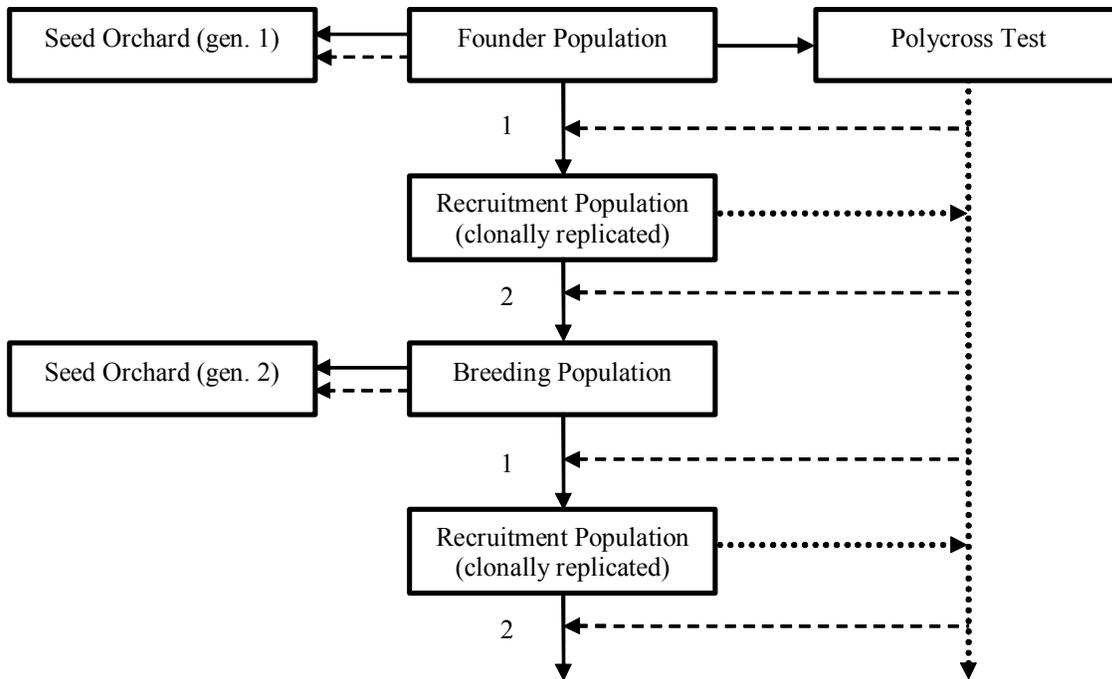
<sup>a</sup> Size of each family equal to 30 ( $N_{RP}/N_f$ ).

<sup>b</sup> Size of top-ranking family equal to 720 ( $N_{RP}$ ), size of remaining families equal to 0.

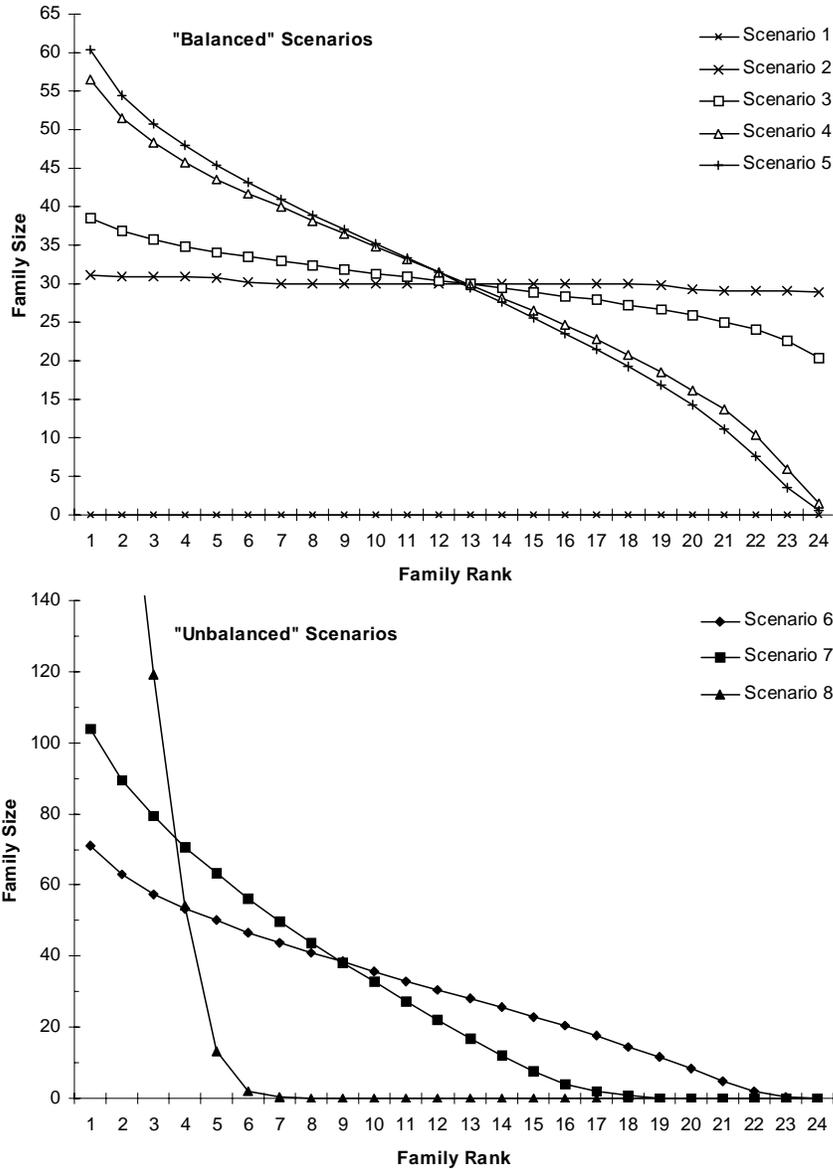
The particular values of  $V_f/V_{fmax}$  simulated were selected to cover the entire range of variation in family size.

Variance of family sizes		
Scenario	$V_f/V_{fmax}$	Note
1	0	Equal family size <sup>a</sup>
2	$2 \times 10^{-5}$	
3	$1 \times 10^{-3}$	
4	$1 \times 10^{-2}$	
5	$1.25 \times 10^{-2}$	
6	$2 \times 10^{-2}$	
7	$5 \times 10^{-2}$	
8	$3 \times 10^{-1}$	
9	1	Maximum $V_f^b$

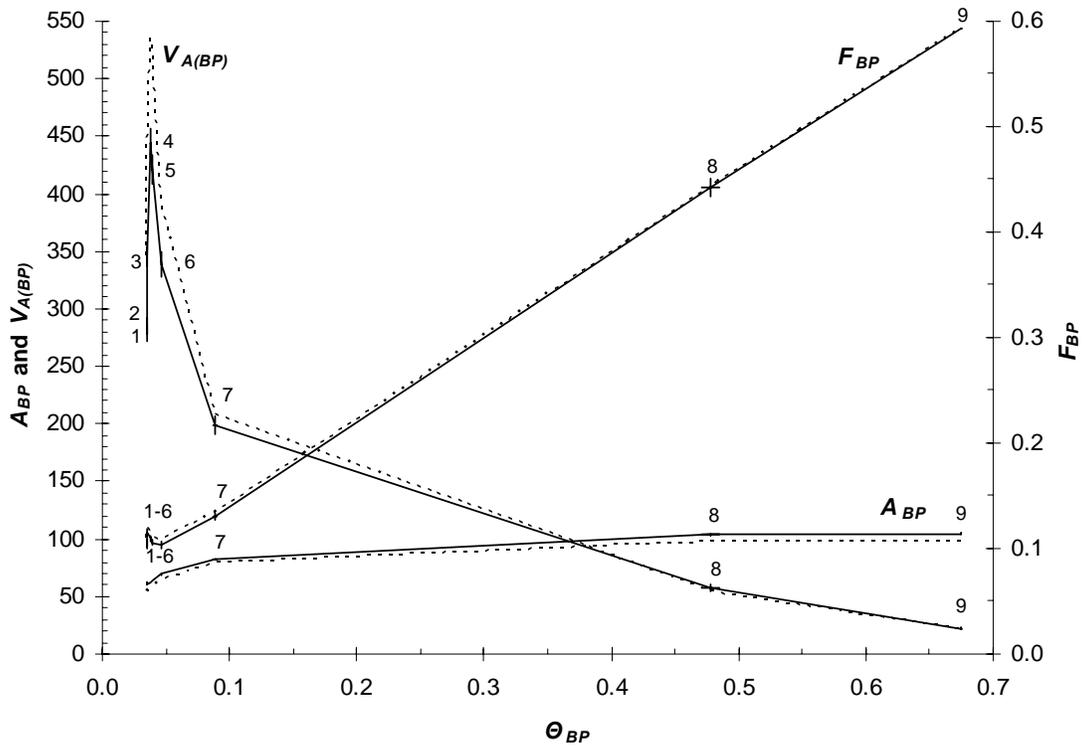
↓  
↓ imbalance ↓  
↓



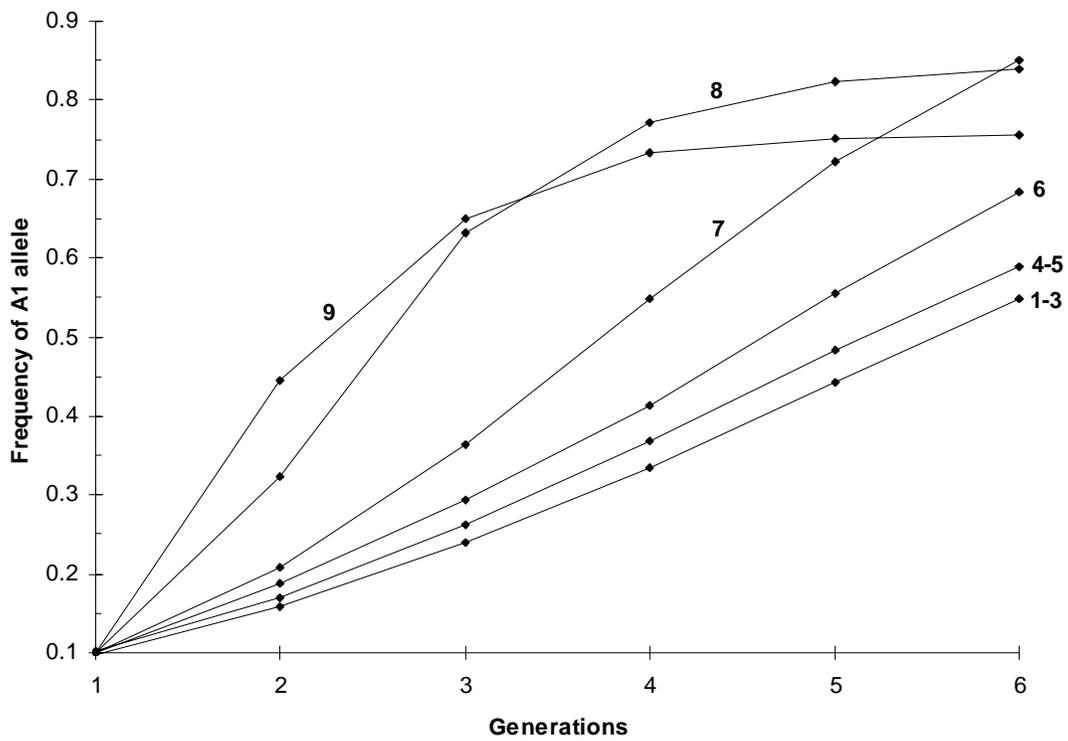
**Figure 1.** Schematic description of the conceptual breeding strategy, where the symbols are: “1” = positive assortative mating, “2” = group-merit selection. The addition of information on relatives for genetic evaluation is depicted by dotted lines, while broken lines designate the use of predicted BLUP values. Only the first two generations are depicted; subsequent generations follow an identical plan.



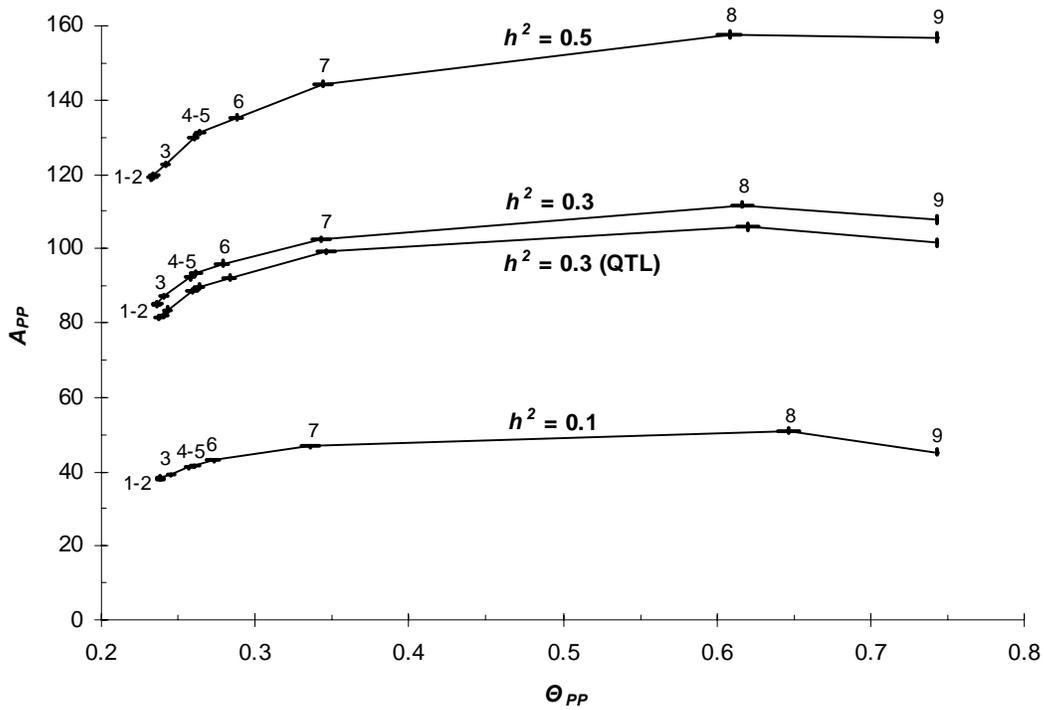
**Figure 2.** Distribution of family sizes in generation five. Individual lines were drawn for each scenario (one through nine) presented in Table 1. The X-axis shows the rank of each family; families are sorted by mid-parent BLUP value from the left (highest rank) to the right (lowest rank). Each line connects family sizes within an individual simulation scenario, averaged over 400 iterations. Scenario eight had an average size of family 23 equal to 196.9 and of family 24 equal to 334.1 (outside the scale of in the figure). The extreme scenario corresponding to  $V_f = V_{fmax}$  (scenario nine) is not presented in the figure, where all 720 progenies were allocated to family of rank one. The figure corresponds to the value of  $h^2$  equal to 0.3, and  $\lambda$  equal to 0.



**Figure 3.** Breeding population (BP) in generation six. The figure presents the average additive effect ( $A_{BP}$ ) and additive variance ( $V_{A(BP)}$ ) on the Y axis, and the average inbreeding coefficient ( $F_{BP}$ ) on the secondary Y axis as a function of group coancestry ( $\Theta_{BP}$  on the X axis). Lines connect individual scenarios described in Table 1, such that the variation in family sizes  $V_f$  increases (in individual scenarios) from the left to the right in the graph. The number of each scenario is presented at each point in the figure. The figure corresponds to an initial value of  $h^2$  equal to 0.3, and  $\lambda$  equal to 0. Identical scenarios but with  $\lambda$  equal to 0.1 are represented by dotted lines. Confidence intervals at the 95% level are presented around each average of 400 simulation iterations.



**Figure 4.** Mixed-inheritance model, giving the frequency of the  $B_1$  allele in the founder (generation one) and breeding populations (generations two through six) in individual simulation scenarios, one through nine.



**Figure 5.** Production population (PP) in generation six. The figure presents average additive effect ( $A_{PP}$ ) as a function of group coancestry ( $\Theta_{PP}$ ). Lines connect individual scenarios described in Table 1, the variation in family sizes  $V_f$  increases (in individual scenarios) from the left to the right, in the graph. The number identifying each scenario is presented at each point in the figure. The figure corresponds to values of  $h^2$  equal to 0.1, 0.3, and 0.5 at  $\lambda$  equal to 0, and  $h^2$  equal to 0.3 at  $\lambda$  equal to 0.1. Confidence intervals at the 95% level are presented around each average of 400 simulation iterations.

## **CHAPTER 5**

# **The impacts of positive assortative mating, differential success of somatic embryogenesis, and multi-trait selection on the outcome of clonal forestry programs**

Authors: M. Lstibůrek and T. J. Mullin

## **Abstract**

A breeding program where deployment progeny from elite-parent genotypes was by somatic embryogenesis was evaluated by simulation. Differences in the success of somatic embryogenesis among progenies of these genotypes were considered, and their impact on direct and correlated genetic response in selected clonal mixtures was evaluated. The results revealed that the variation in success of clonal propagation does not necessarily lead to a reduction of genetic gains in selected clonal mixtures. This can be explained by relatively little variation among elite families as opposed to large within-family variation, which provides potential for selecting superior progeny genotypes, even though they may have not originated from the crosses among the very best parents. This conclusion holds for a range of heritability, correlation between traits and their respective economic importance.

## **Introduction**

Since the first published reports on conifer somatic embryogenesis (SE) in 80's (El-Nil 1980; Chalupa 1985; Hakman et al. 1985), SE has been accomplished in many species and hybrids, primarily of the family Pinaceae. Work with some species has already evolved into large-scale clonal selection programs [for review, see Cyr and Klimaszewska (2002), pp. 42-43]. Such programs focus on elite genotypes developed in breeding programs with the goal to obtain embryogenic cultures of their progenies, recover viable plants from these cultures that are planted into clonal tests, followed by mass propagation and deployment of superior, tested clones.

The implementation of SE may realize larger gains compared to seed orchard technologies due to higher selection intensity and capture of non-additive genetic effects. The propagation effort can focus on a few elite parents and additional improvement can be achieved by practicing within-family selection in test families (Sutton 2002, Figure 3). The testing stage can take a considerable amount of time, during which the embryogenic cultures must be maintained in cryostorage (Park 2002). This technology offers potential for further combination with marker-aided selection and genetic transformation. However, factors such as genetic effects, developmental stage of embryos, seed manipulation and storage, and culture medium may all result in yield losses. As a result, only a fraction of genotypes entering the laboratory culture phase of the program may result in successful embryogenic lines, and propagation success may not correlate with the commercial value of these lines (Park et al. 1993; Timmis 1998; Klimaszewska and Cyr 2002; Niskanen et al. 2004).

The objective of this study was to evaluate the influence of the variation in success of SE among individual lines on the outcome of a clonal selection program, where the deployment is by SE. In forest-tree breeding, the improvement of multiple characters is usually of interest (e.g., Danell 1995; McKeand and Bridgwater 1998; Mikola 2002). Two traits (productivity and quality) are therefore considered and the impact of their covariance and economic importance on the study's objective is evaluated.

## Methods

A schematic description of the simulated breeding strategy is provided in Figure 1. Forest tree breeding simulation tool POPSIM<sup>TM</sup> (Mullin and Park 1995) was used and modified as described in later sections below. The phenotypic value of each tree was composed of independent additive genetic component and environmental deviation. Two traits were considered throughout this study: productivity and quality.

### Simulating multiple traits

When multiple traits are considered in a parametric simulation model, a set of random variates is drawn from a multivariate normal distribution. The expected variances and covariance of these variates were supplied as an input in matrix  $\mathbf{V}$ . Triangular matrix  $\mathbf{R}$  was then generated by the Cholesky Decomposition such that  $\mathbf{RR}' = \mathbf{V}$  (Gentle 1998). The lower triangular elements of matrix  $\mathbf{R}$  for the case of two variates can be expressed as:

$$\mathbf{R} = \begin{vmatrix} \sigma_1 & \\ \sigma_2 \sigma_{1,2} & \sigma_2 \sqrt{1 - \sigma_{1,2}^2} \end{vmatrix}$$

where  $\sigma_1$  and  $\sigma_2$  are desired standard deviations of variates one and two, respectively; and  $\sigma_{1,2}$  is the desired covariance between the two variates. Since the task was to generate variates with expected variances equal to one, the matrix  $\mathbf{V}$  was:

$$\mathbf{V} = \begin{vmatrix} 1 & \sigma_{1,2} \\ \sigma_{1,2} & 1 \end{vmatrix}$$

and consequently the triangular matrix  $\mathbf{R}$  further simplified into:

$$\mathbf{R} = \begin{vmatrix} 1 & \\ \sigma_{1,2} & \sqrt{1 - \sigma_{1,2}^2} \end{vmatrix}$$

The generation of input random samples consisted of generating independent random numbers from the uniform distribution  $U(0, 1)$ . The random number generator implemented in this study was “MRG32k3a” (L’Ecuyer et al. 2002). These random numbers were then transformed as described in Press et al. (1992, pp. 280) to normal variates  $r$ :

$$r \sim N(0, 1).$$

Additive effects of founder genotypes ( $\mathbf{A}_F$ ) were then sampled from the multivariate normal distribution, such as:

$$\mathbf{A}_F \sim N(\mathbf{0}, \mathbf{V}); \mathbf{V} = \begin{vmatrix} \sigma_{A1}^2 & \sigma_{1,2} \\ \sigma_{1,2} & \sigma_{A2}^2 \end{vmatrix}$$

where  $\sigma_{A1}^2$  and  $\sigma_{A2}^2$  are initial additive variances for traits one and two, respectively. To generate the additive effect of each trait  $n$  in each  $i^{\text{th}}$  founder genotype, the following expression was used:

$$A_{Fni} = \sum_{j=1}^{n_{\text{trait}}} \mathbf{R}_{nj} r_{ji} \sigma_{An}$$

Additive effects of progeny genotypes ( $\mathbf{A}_P$ ) generated in each breeding generation were sampled from the multivariate normal distribution, such as:

$$\mathbf{A}_P \sim N(\mathbf{MPA}, \mathbf{V}); \mathbf{V} = \begin{vmatrix} 0.25(2 - F_f - F_m) \sigma_{A1}^2 & \sigma_{1,2} \\ \sigma_{1,2} & 0.25(2 - F_f - F_m) \sigma_{A2}^2 \end{vmatrix}$$

where  $\mathbf{MPA}$  are mid-parent additive values for full-sib families corresponding to trait  $n$  and  $F_f$  and  $F_m$  are inbreeding coefficients for female and male parents, respectively. The additive

effect of each trait  $n$  in each  $i^{\text{th}}$  progeny genotype was generated [modified from del-Bosque González (1989)] as:

$$A_{\text{pni}} = \frac{A_{\text{fn}} + A_{\text{mn}}}{2} + \sum_{j=1}^{\text{ntrait}} \mathbf{R}_{\text{nj}} r_{\text{ji}} \sqrt{\frac{1 - 0.5(F_f + F_m)}{2}} \sigma_{An}^2$$

Each progeny genotype in the recruitment population was clonally replicated by ten ramets. The environmental effect of each ramet for each trait  $n$  was randomly sampled from  $N(100, \sigma_{En}^2)$ , thus the environmental deviation did not contribute to the correlation between traits. The initial variances ( $\sigma_{An}^2, \sigma_{En}^2$ ) were supplied as a simulation input parameters, such that the phenotypic variance for both traits was equal to 500 and the narrow-sense heritability for the two traits set to  $h_1^2 = 0.1$  and  $h_2^2 = 0.2$ , or  $h_1^2 = 0.3$  and  $h_2^2 = 0.6$  for productivity (trait 1) and quality (trait 2), respectively. The correlation between traits was varied in individual simulation scenarios.

### **Hypothetical breeding strategy**

The breeding was initialized by selecting 100 founder genotypes (founder population) from a population of unrelated and non-inbred individuals (natural stands). Two-trait index value was calculated for each tree in the founder population. In this calculation, it was assumed that the true genetic variances and covariances were known. In reality, only estimates of these parameters are available, but the index seems to be robust in this perspective (Harris 1964). Let  $\mathbf{P}$  be the variance-covariance matrix of phenotypic records,  $\mathbf{C}$  is the matrix of covariances between the sources of information in selection index and additive genetic values in aggregate breeding value and  $\mathbf{a}$  is a vector of economic values for both traits. The vector of

index weights  $\mathbf{b}$  (partial regression coefficients) was then calculated as  $\mathbf{b} = \mathbf{P}^{-1}\mathbf{Ca}$ . Index values were calculated in the same way for recruitment populations and for candidate test genotypes for clonal deployment (discussed below).

Trees in the founder population were sorted by this index value and single-pair (positive assortative) mating was performed to generate 50 controlled crosses. For each of these crosses, 50 progeny genotypes were generated (as described above) and each one of them was clonally replicated by ten ramets. These progenies are referred to as the “recruitment population”.

Following the assessment of all progenies (selection candidates) and the calculation of selection index values, the two top-ranking (based on their two-trait index value) progenies were selected from each family (balanced within-family selection) to form the next generation’s breeding population (BP). The breeding-testing-selection process was repeated over six generations.

### **Production population**

The clonal deployment implemented in this study was basically identical to that illustrated by Park (2002, Figure 1). In every generation, additional “elite” crosses were performed among the ten top-ranking trees in the BP using a half-diallel mating design with 45 crosses in total (selfing was excluded). The purpose of these elite-parent crosses was to generate candidate clones for the mass deployment to clonal forest plantations. The distribution of embryogenesis success (family sizes) for these elite crosses was derived as follows. First, five families were randomly selected and assigned family size equal to zero (representing

families whose embryogenic success in the laboratory was a complete failure). Second, 40 random variates  $t$  were sampled from the exponential distribution  $t \sim \text{Exponential}(\lambda)$ , using “MRG32k3a” random number generator (L’Ecuyer et al. 2002) and transformed to obtain exponentially distributed values (Casella and Berger 2002, pp. 247). These exponential random variates were then randomly assigned to individual families. The maximum size of each family was limited to 150 genotypes, sizes exceeding this maximum were truncated. Considering this restriction, parameter  $\lambda$  was iterated, such that the overall number of progenies in all embryogenic families reached 2,000, averaging 50 genotypes per family. An example of such a distribution of family sizes is provided in Figure 2. Each progeny genotype was clonally replicated by ten clonal copies (ramets) as described earlier. The ten highest-ranking clones were then selected from these families (the selection was based on two-trait index values); and relatedness was not considered when selecting these clones. Two additional sets of scenarios were run for the comparison. In the first set, the generation of family sizes was as described above, but exponential random variates were not assigned to families at random, but rather followed the ranking based on mid-parent index value. A second set of scenarios was run with equal size of each family (exactly 50 genotypes in each of 40 families).

### **Simulation and evaluation of results**

The simulation was replicated by 400 independent iterations. The effective population size used for the evaluation of results was measured by “status number”, which is the census size of an equivalent population composed of unrelated, non-inbred individuals, where the

probability to draw two genes identical by descent is the same as for the population under study (Lindgren et al. 1996).

## Results

Throughout all scenarios, the additive genetic response from the selected clonal mixture was only marginally affected by the alternative distributions of embryogenic success. An example illustrating the point is provided in Figure 3. The difference between the exponentially distributed embryogenic success with random allocation of family sizes [E(RM)] and constant (C) embryogenic success in genetic response was not statistically significant ( $p > 0.05$ ). While there appeared to be a tendency for higher values of  $N_S$  under C, under the majority of cases these differences were not significant. In addition, there was considerable run-to-run variability in  $N_S$  (assessed by confidence intervals), so for any two iterations of the same scenario, the relationship between E(RM) and C in  $N_S$  is essentially not predictable. The observed similarity between E and C distributions in individual generation outcomes was independent on other input parameters in this study that are described separately below. When the exponential allocation of embryogenic success was proportional to the expected family values [E(BV)], there was approximately 1% increase in additive genetic response, accompanied by a significant decrease in  $N_S$  [ $N_S$  in generation 6 was 2.7 under E(RM), but 2.3 under E(BV)].

The effect of narrow-sense heritabilities of both traits in the BP and PP is illustrated in Figure 4. Higher  $h^2$  provided stronger genetic responses in both traits in the BP at the same  $N_S$  (balanced within-family selection). In the PP, higher  $h^2$  led to greater gains as well lower  $N_S$ , although this reduction in  $N_S$  was statistically significant only in later generations (five and six).

After the initial decline,  $V_{A(BP)}$  increased over generations, due to positive assortative mating. The value of  $V_{A(BP)}$  in the quality trait in generation six was  $V_{A(BP)} = 420.8$  at  $h^2 = 0.6$  and  $V_{A(BP)} = 144.0$  at  $h^2 = 0.2$ . The increase of  $V_{A(BP)}$  in the productivity trait was lower (due to lower  $h^2$ ),  $V_{A(BP)} = 167.3$  at  $h^2 = 0.3$  and  $V_{A(BP)} = 57.1$  at  $h^2 = 0.1$ .  $F_{BP}$  was zero in generations one and two and accumulated in later generations up to  $F_{BP} = 0.033$  in generation six (census size of BP equal to 100) at the lower  $h^2$  scenario. For higher case  $h^2$ ,  $F_{BP}$  was increased by 30%. In the PP,  $V_{A(PP)}$  at the higher  $h^2$  was over double that of lower  $h^2$  ( $V_{A(PP)} \approx 80-90$  vs.  $V_{A(PP)} \approx 30-40$ ). There was no apparent change in  $V_{A(PP)}$  over generations.  $F_{PP}$  in generation six was 0.064 (lower  $h^2$ ) and 0.082 (higher  $h^2$ ).

The effect of genetic correlation ( $r$ ) between productivity and quality traits on the PP  $A$  and  $N_S$  is presented in Figure 5. The scenarios depicted in the figure are for equal economic weights applied to both traits. Generally,  $r$  did not affect significantly PP  $N_S$  ( $N_S$  is not shown in the figure), although some fluctuation was observed, specifically in earlier generations. At high  $r$ , the residual difference in responses of both traits was due only to the difference in heritabilities.  $F_{PP}$  in generation six was between 0.07 and 0.085 at different  $r$  values, but the differences were not statistically significant. Finally, the impact of economic weights on PP is demonstrated in Figure 6, where  $r$  is kept constant at  $r = 0.1$ . The response in quality is much larger compared to productivity, as a result of higher  $h^2$ . In the majority of cases, the change in economic weights did not significantly alter  $N_S$  and  $F_{PP}$ .

## Discussion

Of great interest to those involved in SE development and deployment is that variation in embryogenic success among individual progeny genotypes does not appear to reduce genetic gains in selected clonal mixtures. The clonal program focuses on progeny of elite individuals in the BP. These individuals represent one extreme of a distribution of aggregate index breeding values and the variance among them can be considerably lower than that of the entire BP. When a truncated normal distribution is considered, the phenotypic variance of the top 10% of individuals in BP represents only about 17% of the phenotypic variance of the entire BP and the additive genetic variance is then between 17 to 100% of the additive variance of the BP, depending on  $h^2$  (the “Bulmer effect”, Bulmer 1985). Since these top individuals in the BP were selected based on their index value, which was calculated from the average clonal performance (average of 10 ramets) assessed in both traits, a greater reduction of additive variance is expected compared to the case when the assessment would be performed on a single phenotypic observation. The expected breeding value of an individual is the average of parental breeding values (Falconer and Mackay 1996). The actual value of a progeny deviates from this expectation due to random selection of copies of parental genes (Mendelian sampling), giving rise to within-family variance. Thus, even though there is a reduced variation among family means by focusing on only the top 10% of BP, the contribution of Mendelian sampling within each family is unchanged, depending only on the additive genetic variance in the unselected population and inbreeding of the parents. Therefore, it is possible to find superior genotypes for selecting the clonal mixture

provided that a proportion of families is highly successful in SE and that a large number of genotypes in these families are tested.

The above notwithstanding, there remains some motivation for sharing the test effort among a larger number of families (rather than focusing on only few). The delay in time due to the testing phase of cloned progenies was not assessed, even though a weak correlation between early and later performance may hinder accurate selection at early ages, thus reevaluation of the field tests at later ages would be necessary and would lead to revised compositions of selected clonal mixtures (Park 2002). Maintenance of genetic diversity imposes a restriction on the maximization of genetic gain in the deployed clonal plantations. In reality, this dilemma has often been resolved by deploying a certain number of clones (rather than single) and expecting that the diversity of such mixture would provide enough protection against plantation failure due to various catastrophic events (Sutton 2002). Based on the results in this study, it can be inferred that the effective population size of a periodically revised selected clonal mixture may fluctuate greatly from one time to another, even though the same census number of clones is selected each time. The solution to this can be to select with the goal to maximize genetic gain, but to restrict the outcome of such selection by a minimum effective population size, specified by the forest manager. Group-merit selection (Lindgren and Mullin 1997) can be used to provide such solution. Maintaining larger numbers of families and clones in the test may be useful for two reasons: first, it would provide more choices for performing such selection; and second, as the clonal test matures the rank among clones in the combined index value may change considerably

(economic weights in the index may change as well) and more opportunities would be available for the future.

SE technology has evolved considerably over the years. Merkle and Dean (2000) point to the short-coming of this technology as the “low frequency of regeneration for many of the most desirable clones”. We believe, based on our simulation study, that this problem may not be as severe as it might first appear. As the SE technology further develops, the normalization of responses among genotypes in individual steps is expected (Klimaszewska and Cyr 2002). The results of this study indicate that following such normalization and redirecting resources according to expected family values (as shown in Figure 3, lines “E(BV)”) would not necessarily lead to an important increase in genetic gains, if the testing effort is fixed. A positive effect of such redirection might appear under a much lower test budget than that considered here, when the average family size is considerably reduced and the variation among family means becomes more influential.

Higher values of  $h^2$  for both traits were reflected in larger genetic gains in the BP. Since balanced within-family selection was practiced for the population advancement, higher  $h^2$  did not influence the effective population size of the BP. However, when a clonal mixture was selected (with no restrictions on relatedness), higher  $h^2$  led to greater accuracy of selection among families, resulting in lower effective population size in the PP. In this model,  $h^2$  of quality trait was set to double that of the productivity trait. Even under slight correlation ( $r = 0.1$ ), selection focused primarily on productivity resulted in a substantial correlated response in quality, attributable primarily to the effect of higher  $h^2$  of the quality trait. Some alternative way of combining the value of both traits may be considered due to

the observed sensitivity of responses to  $h^2$ , genetic correlation and economic weights. An example could be the Desired Gain Index where instead of using economic weights, the amount of desired response is specified for each trait, relative to each other (Pesek and Baker 1969; Harville 1975).

Information regarding the detailed location of individual loci, their particular genetic effects and interactions in commercial traits like productivity is still not yet within practical reach for forest tree species, thus limiting us from building more accurate simulation models of the actual biological system. Animal and forest-tree breeders have often used infinitesimal assumptions for modeling traits where the quantitative evidence indicated larger number of genes with relatively small effects and there was small or no experimental indication of a few genes having major effects (del-Bosque González 1989; Rosvall 1999; Jorjani 2002). In the current study, the objective was very general, applicable to a larger number of programs. The use of infinitesimal model was more suitable than the finite-locus one; since less information regarding particular genetic mechanism was available thus fewer parameters were tested in the simulation. The changes in population structure were considered over six generations only. If the study expanded to large number of generations, a more elaborate genetic model would be required to study the change in variances and covariances. When the evidence suggests that a few major-gene loci contribute to the observed variation, higher number of parameters (characterizing pleiotropic effects and linkage) can be tested in order to study in higher detail the correlated response to selection (e.g., Lascoux 1997).

## **Acknowledgements**

Authors acknowledge financial support from CellFor Inc., and the North Carolina State University-Industry Cooperative Tree Improvement Program. We thank Dr. El-Kassaby for his initial ideas that stimulated this research.

## References

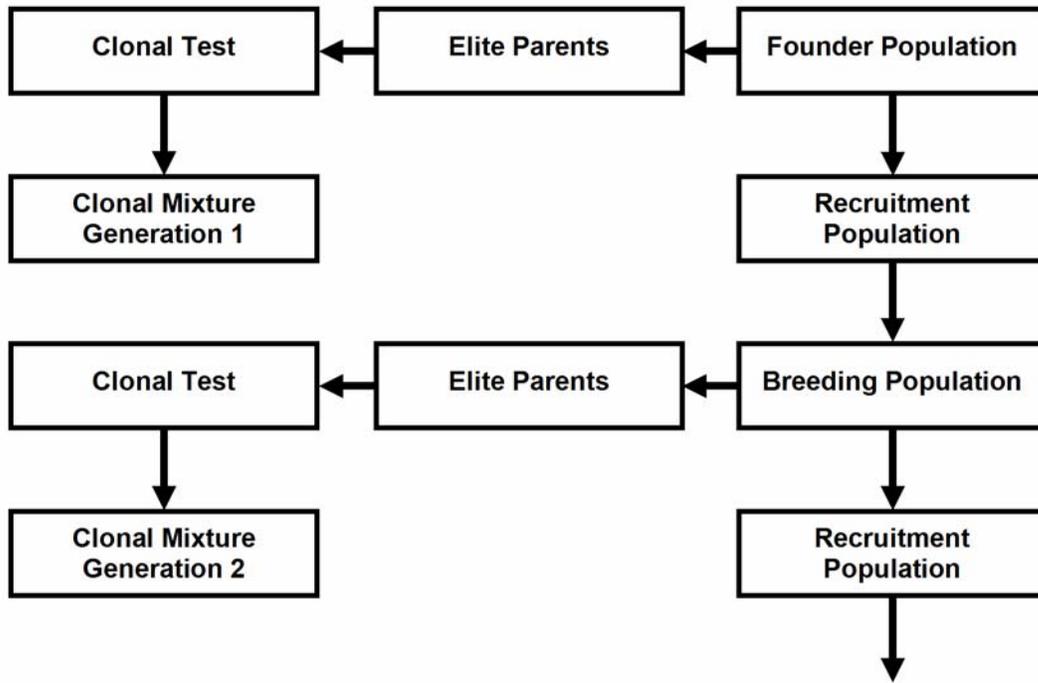
- Bulmer MG (1985) The mathematical theory of quantitative genetics. Clarendon Press, Oxford
- Casella G and Berger RL (2002) Statistical inference. 2<sup>nd</sup> edition, Duxbury, Pacific Grove
- Chalupa V (1985) Somatic embryogenesis and plantlet regeneration from cultured immature and mature embryos of *Picea abies* (L.) Karst. Comm Inst For Czechosloveniae 14:57-63
- Cyr DR and Klimaszewska (2002) Conifer somatic embryogenesis: II. Applications. Dendrobiology 48:41-49
- Danell Ö (1995) Breeding programs in Sweden. 1. General approach. SkogForsk. Arbetsrapport nr 302, pp 1–4
- del-Bosque González AS (1989) Simulations of nucleus breeding schemes for wool production. PhD Dissertation, University of New England, Armidale
- El-Nil A (1980) Embryogenesis of gymnosperm forest trees. US Patent 4,217,730
- Falconer DS and Mackay TFC (1996) Introduction to quantitative genetics. Longman, Harlow
- Gentle JE (1998) Cholesky Factorization. §3.2.2 in Numerical Linear Algebra for Applications in Statistics. Berlin: Springer-Verlag, pp. 93-95
- Hakman I, Fowke LC, von Arnold S, Eriksson T (1985) The development of somatic embryogenesis in tissue cultures initiated from immature embryos of *Picea abies* (Norway spruce). Plant Science 38:53-59

- Harris DL (1964) Expected and predicted progress from index selection involving estimates of population parameters. *Biometrics* 20:46-72
- Hartville DA (1975) Index selection with proportionality constraints. *Biometrics* 31:223-225
- Jorjani H (2002) Comparison of simulation strategies: the infinitesimal model and finite locus model. In *Proceedings of the 7th World Congress on Genetics Applied to Livestock Production, Montpellier, France, August 2002*, 1-4
- Klimaszewska K and Cyr DR (2002) Conifer somatic embryogenesis: I. Development. *Dendrobiology* 48:31-39
- Lascoux M (1997) Unpredictability of correlated response to selection: linkage and initial frequency also matter. *Evolution* 51:1394-1400
- Lindgren D, Gea LD and Jefferson PA (1996) Loss of genetic diversity monitored by status number. *Silvae Genet* 45:52-59
- L'Ecuyer P, Simard R, Chen EJ and Kelton WD (2002) An objected-oriented random-number package with many long streams and substreams. *Operations Research* 50:1073-1075
- Lindgren D and Mullin TJ (1997) Balancing gain and relatedness in selection. *Silvae Genet* 46:124-129
- McKeand SE and Bridgwater F (1998) A strategy for the third breeding cycle of loblolly pine in the southeastern U.S. *Silvae Genet* 47:223-234
- Merkle SA and Dean JFD (2000) Forest tree biotechnology. *Current opinions in biotechnology* 11:298-302

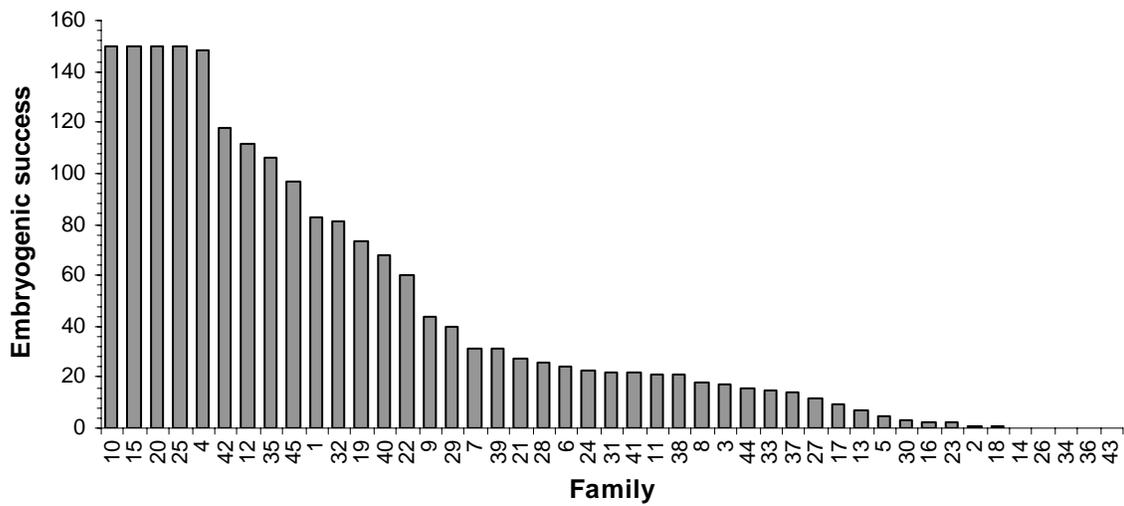
- Mikola J (2002) Long-term tree breeding strategy in Finland: integration of seed production and breeding. In Proceedings of the Nordic Group for Management of Genetic Resources of Trees, Mekrijärvi, Finland, March 2001. Finnish Forest Research Institute, Research Papers 842:12-13
- Mullin TJ, Park YS (1995) Stochastic simulation of population management strategies for tree breeding: a new decision-support tool for personal computers. *Silvae Genet* 44:132–141
- Niskanen AM, Lu J, Seitz S, Keinonen K, Weissenberg KV and Pappinen A (2004) Effect of parent genotype on somatic embryogenesis in Scots pine (*Pinus sylvestris*). *Tree Physiology*, 24:1259–1265
- Park YS (2002) Implementation of conifer somatic embryogenesis in clonal forestry: technical requirements and deployment considerations. *Ann For Sci* 59:651-656
- Park YS, Pond SE, Bonga JM (1993) Initiation of somatic embryogenesis in white spruce (*Picea glauca*): genetic control, culture treatment effects, and implications for tree breeding. *Theor Appl Genet* 86:427-436
- Pesek J and Baker RJ (1969) Desired improvement in relation to selection indices. *Can J Plant Sci* 49:803-804
- Press WH, Teukolsky SA, Vetterling WT, and Flannery BP (1992) Numerical recipes in FORTRAN. The art of scientific computing. Cambridge University Press, Cambridge
- Rosvall O (1999) Enhancing gain from long-term forest tree breeding while conserving genetic diversity. PhD thesis, Swedish University of Agricultural Sciences, Umeå

Sutton B (2002) Commercial delivery of genetic improvement to conifer plantations using somatic embryogenesis. *Ann For Sci* 59:657-661

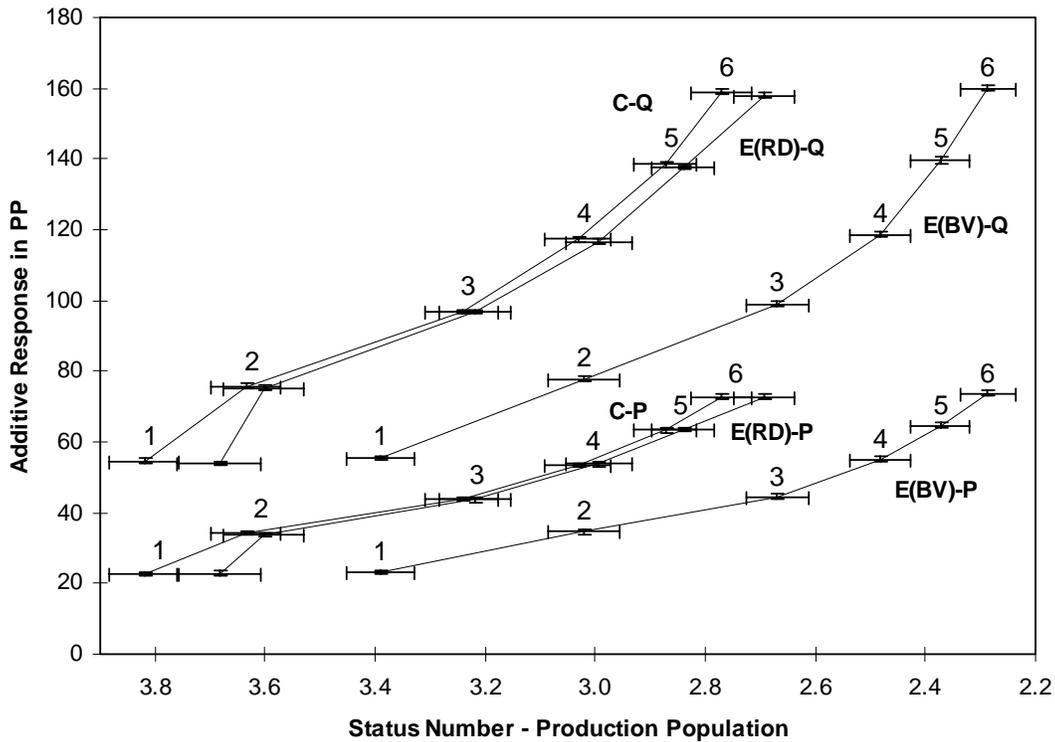
Timmis R (1998) Bioprocessing for tree production in the forest industry: conifer somatic embryogenesis. *Biotechnol Prog* 14:156-166



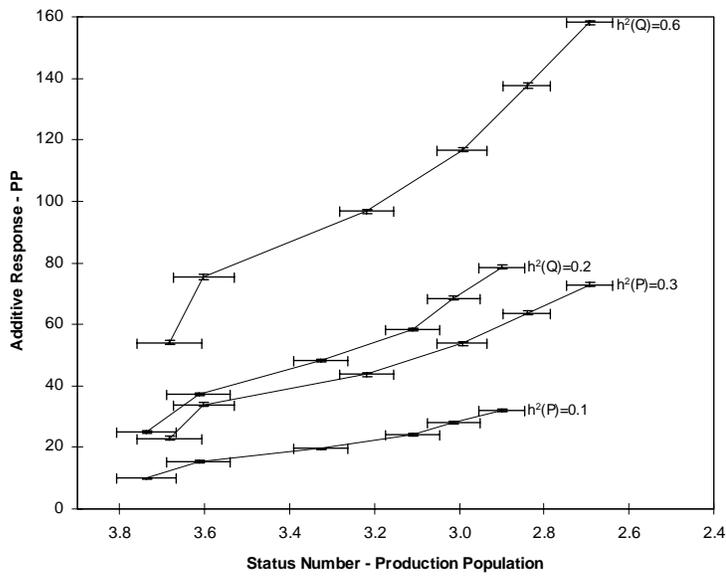
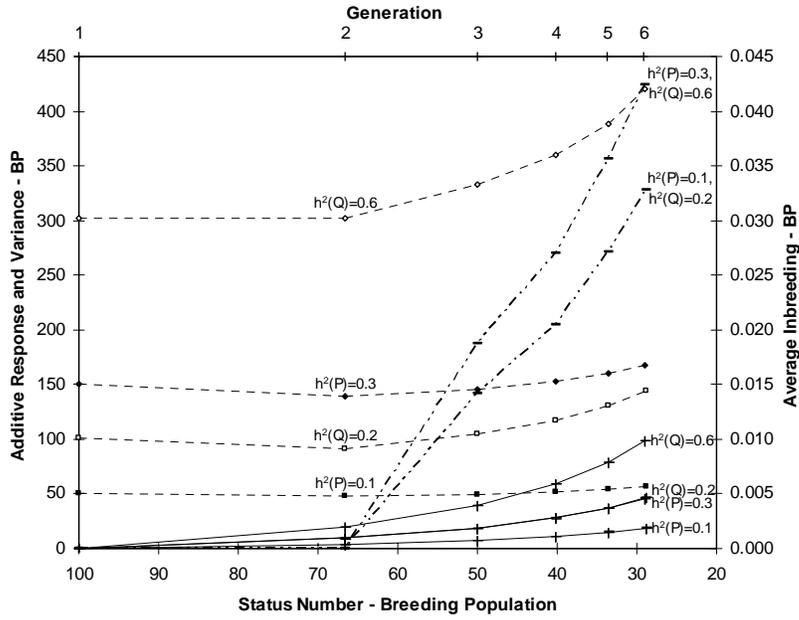
**Figure 1.** Schematic description of the conceptual breeding strategy. First two generations are illustrated; subsequent generations follow an identical plan.



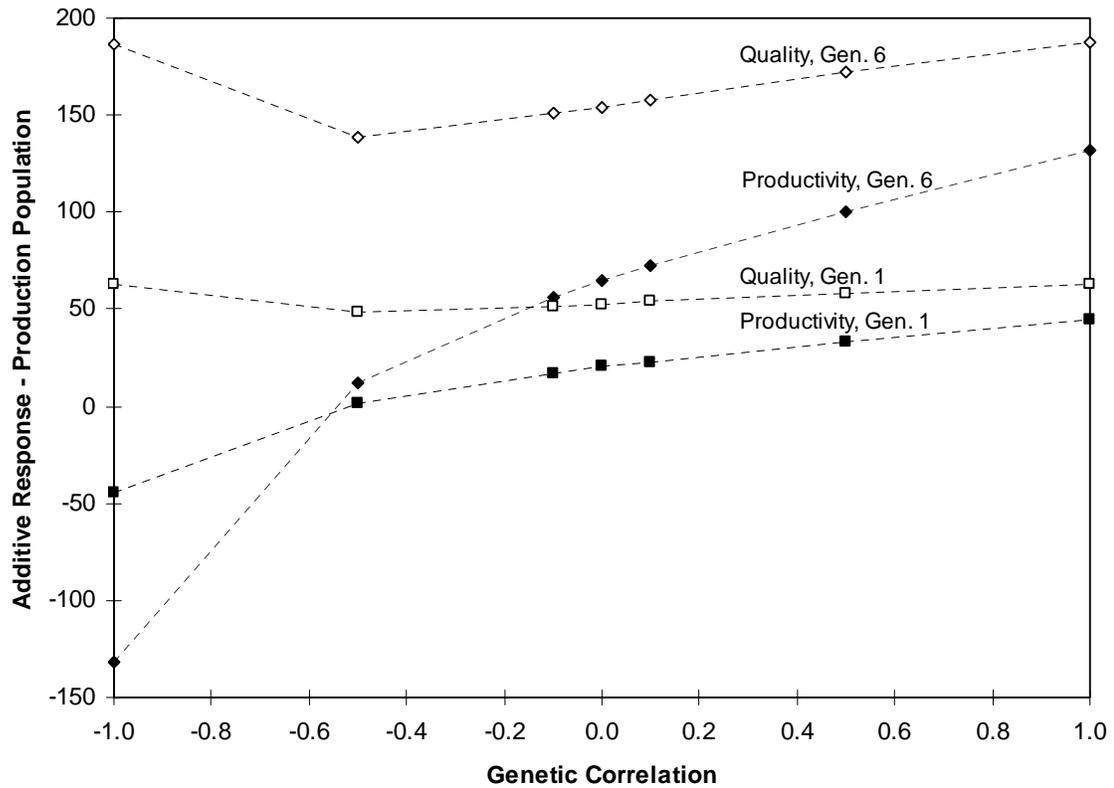
**Figure 2.** Example of embryogenesis success (family size), as derived from the truncated exponential distribution. Rank number of each family is provided on X-axis, families were assigned the success at random (presented in this figure) or alternatively proportional to expected family values. Parameter  $\lambda$  was iterated such as the total number of progenies reached 2,000, and satisfying the restriction of maximum family size equal to 150. Different values were obtained with different sequences of random numbers in individual generations and iterations of simulation scenarios.



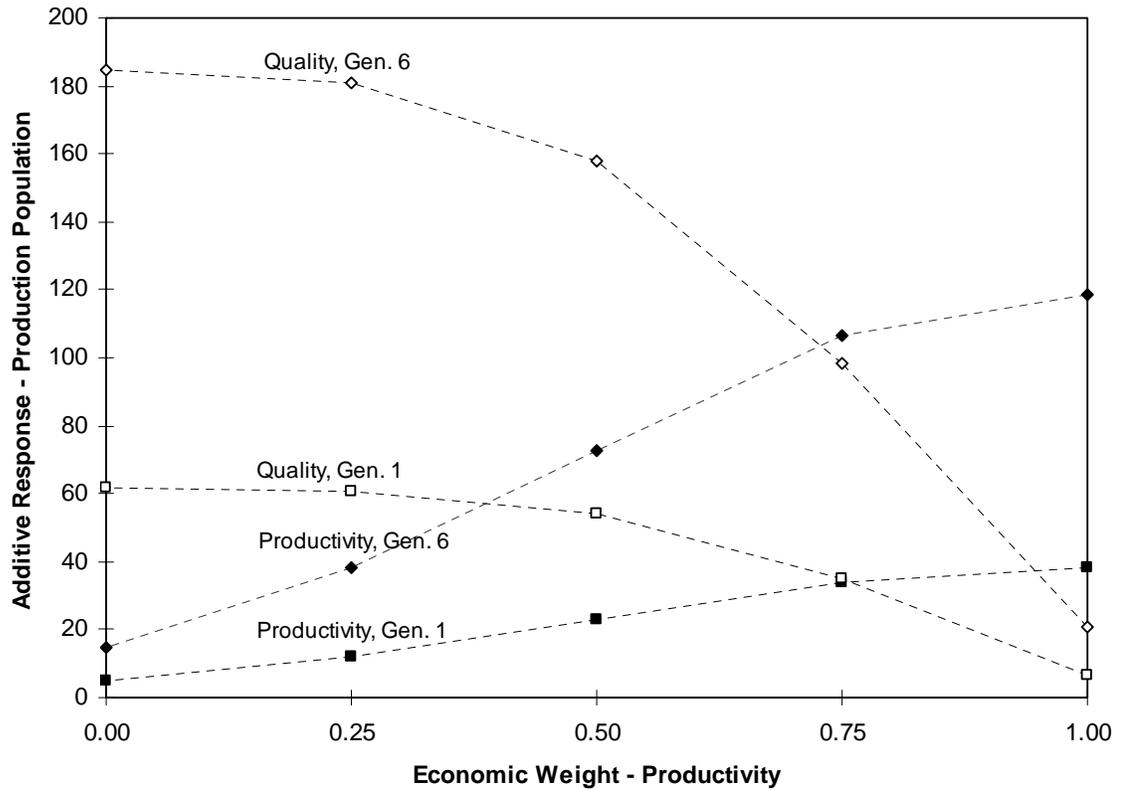
**Figure 3.** Additive genetic response and effective population size (status number) in the PP in individual generations (labeled one through six). Three distributions of embryogenic success are compared in the graph: constant (C); exponential with random allocation of success to elite families [E(RD)]; and exponential with success proportional to the expected family value [E(BV)]. Symbols P and Q refer to the productivity and quality traits, respectively. Lines connect average values of 400 simulation iterations of a given scenario, plotted with 95% confidence intervals. The comparison corresponds to the case where the heritability for productivity [ $h^2(P)$ ] = 0.3 and that for quality [ $h^2(Q)$ ] = 0.6. Correlation between traits ( $r$ ) was 0.1 and economic weights on both traits were equal ( $a_1 = a_2 = 0.5$ ).



**Figure 4.** Additive genetic response (—), additive variance (-----), average inbreeding coefficient (---), and effective population size (status number, X-axis) in the BP (upper graph) and PP (lower graph) in individual generations. Different values of heritability for productivity [ $h^2(P)$ ] and quality trait [ $h^2(Q)$ ] are compared in the figure. Lines connect average values of 400 simulation iterations of a given scenario, plotted with 95% confidence intervals. Comparison corresponds to the correlation among traits  $r = 0.1$ , equal economic weights on both traits ( $a_1 = a_2 = 0.5$ ) and exponentially distributed embryogenic success.



**Figure 5.** Additive genetic response for studied range of genetic correlation between productivity and quality traits in the PP in generations one and six. Figure corresponds to the heritability of productivity  $[h^2(P)] = 0.3$  and the heritability of quality  $[h^2(Q)] = 0.6$ . Equal economic weights on both traits ( $a_1 = a_2 = 0.5$ ) and exponentially distributed embryogenic success were assumed in presented scenarios.



**Figure 6.** Additive genetic response at different economic weights on productivity trait ( $a_1$ ) in the PP in generations one and six. The economic weight on quality ( $a_2$ ) is equal to  $1 - a_1$ . Figure corresponds to the heritability of productivity [ $h^2(P)$ ] = 0.3 and the heritability of quality [ $h^2(Q)$ ] = 0.6. Correlation among traits ( $r$ ) was equal to 0.1 and exponentially distributed embryogenic success was assumed in presented scenarios.