

- Magnolia latahensis* (Magnoliaceae) and an rbcL sequence of *Persea pseudocarolinensis* (Lauraceae). *Amer. J. Bot.* **91**: 615–620.
- LINDAHL, T. (1993): Instability and decay of the primary structure of DNA. *Nature* **362**: 709–715.
- LINDAHL, T. (1997): Facts and artifacts of ancient DNA. *Cell* **90**: 1–3.
- LIPSCHITZ, N., G. BIGER, G. BONANI and W. WOOLPHI (1997): Comparative dating methods: Botanical identification and ¹⁴C dating of carved panels and beams from the Al-Aksa Mosque in Jerusalem. *J. Archaeolog. Sci.* **24**: 1045–1050.
- MA, L. J., S. O. ROGERS, C. CATRANIS and W. T. STARMER (2000): Detection and characterization of ancient fungi entrapped in glacial ice. *Mycologia* **92**: 286–295.
- MILLER, N. F. (1999): Interpreting ancient environment and patterns of land use: Seeds, charcoal and archaeological context. *TUBA-AR II*: 15–29.
- MILLER, N. F. (1998): The macrobotanical evidence for vegetation in the Near East, c 18000/16000 BC to 4000 BC. *Plaeorient.* **23/2**: 197–207.
- PÄÄBO, S., H. POINAR, D. SERRE, V. JAENICKE-DESPRES, J. HEBLER, N. ROHLAND, M. KUCH, J. KRAUSE, L. VIGILANT and M. HOFREITER (2004): Genetic analyses from ancient DNA. *Ann. Rev. Genet.* **38**: 645–679.
- PÄÄBO, S. (1993): Ancient DNA. *Sci. Amer.* **267**: 60–66.
- PÄÄBO, S. (1988): Molecular cloning of Ancient Egyptian Mummy DNA. *Nature* **314**: 644–645.
- PARDUCCI, L. and R. J. PETTIT (2004): Ancient DNA – unlocking plants’ fossil secrets. *New Phytol.* **161**: 335–339.
- ROGERS, S. O. (1994): Phylogenetic and taxonomic information from herbarium and mummified DNA, pp. 47–67 in *DNA Utilization, Intellectual Property and Fossil DNA*, edited by R. P. ADAMS, J. MILLER, E. GOLENBERG, and J. E. ADAMS, Missouri Botanical Gardens Press, St. Louis/MO.
- ROGERS, S. O. and A. J. BENDICH (1985): Extraction of DNA from milligram amounts of fresh, herbarium and mummified plant tissues. *Plant Mol. Biol.* **5**: 69–76.
- ROGERS, S. O. and A. J. BENDICH (1988): Extraction of DNA from plant tissues, pp. A6:1–10. *In: Plant Molecular Biology Manual*, edited by S. B. GELVIN and R. A. SCHILPEROORT, Kluwer Academic Publishers, Boston/MA.
- ROGERS, S. O. and A. J. BENDICH (1994): Extraction of total cellular DNA from plants, algae, and fungi, pp. D1:1–8. *In: Plant Molecular Biology Manual*, 2nd ed., edited by S. B. GELVIN and R. A. SCHILPEROORT, Kluwer Academic Publishers, Dordrecht, The Netherlands.
- ROGERS, S. O., L. MA, Y. ZHAO, C. M. CATRANIS, W. T. STARMER and J. D. CASTELLO (2005): Recommendations for elimination of contaminants and authentication of isolates from ancient ice cores, pp. 5–21. *In: Life in Ancient Ice*, edited by J. D. CASTELLO and S. O. ROGERS, Princeton University Press, Princeton/NJ.
- ROGERS, S. O., S. REHNER, C. BLEDSOE, G. J. MUELLER and J. F. AMMIRATI (1989): Extraction of DNA from Basidiomycetes for ribosomal DNA hybridizations. *Can. J. Bot.* **67**: 1235–1243.
- ROGERS, S. O., V. THERAISNATHAN, L. J. MA, Y. ZHAO, G. ZHANG, S. G. SHIN, J. D. CASTELLO and W. T. STARMER (2004): Comparison of protocols to decontaminate environmental ice samples for biological and molecular examinations. *Appl. Environ. Microbiol.* **70**: 2540–2544.
- SAVOLAINEN, V., P. CUÉNOUD, R. SPICHTIGER, M. D. P. MARTINEZ, M. CRÉVECOEUR and J.-F. MANEN (1995): The use of herbarium specimens in DNA phylogenetics – evaluation and improvement. *Plant Syst. Evol.* **197**: 87–98.
- SUYAMA, T., K. KAWAMURO, I. KINOSHITA, K. YOSHIMURA, Y. TSUMURA and H. TAKAHARA (1996): DNA sequence from a fossil pollen of *Abies* spp. from Pleistocene peat, in <http://www.madsci.org/> (University of Washington, Medical School): The MadScientist Network.
- SWOFFORD, D. L. (2000): PAUP: Phylogenetic Analysis Using Parsimony, Version 4.03, Sinauer Associates, Sunderland/MA.
- TANI, N., Y. TSUMURA and H. SATO (2003): Nuclear gene sequences and DNA variation of *Cryptomeria japonica* samples from the postglacial period. *Molec. Ecol.* **12**: 859–868.
- TAUTZ, D. and M. RENZ (1983): An optimized freeze-squeeze method for recovery of DNA from agarose gels. *Anal. Biochem.* **132**: 14–19.
- WAYNE, R. K., J. A. LEONARD and A. COOPER (1999): Full of sound and fury: The recent history of ancient DNA. *Ann. Rev. Ecol. Syst.* **30**: 457–477.

Stepwise Penalty Index Selection from Populations with a Hierarchical Structure

By R.-P. WEI^{1,*} and D. LINDGREN²

(Received 5th August 2005)

¹) South China Agricultural University, Wushan, Guangzhou 510642, China, and Sino-Forest Corporation, 3815-29, 38/F., Sun Hung Kai Centre, 30 Harbour Road, Wanchai, Hong Kong.

²) Department of Forest Genetics and Plant Physiology, Swedish University of Agricultural Sciences, S-901 83 Umeå, Sweden.

* Correspondence author and address: RUN-PENG WEI, Sino-Forest Corporation, 3815-29, 38/F., Sun Hung Kai Centre, 30 Harbour Road, Wanchai, Hong Kong. Tel.: +852-2514 2124. Fax +852-2877 0062. Email: runpeng-wei@sinoforest.com

Abstract

By adding a penalty to a candidate’s breeding value for its relationship with the selected individuals, two indexes were constructed as criteria for stepwise selection of superior individuals from populations with a hierarchical structure. The relationship was expressed in terms of either family contribution or group coancestry. One of the indexes was derived from an optimal

selection model. A stepwise procedure that screened superior individuals one by one was introduced to make selection based on these indexes possible. Two penalty selection methods exclusively maximized gain at given coancestry. Both methods produced all identical solutions in most of the populations simulated, and were nearly equivalent in the remaining populations, particularly when heritability was high and the population structure was simple. A better balance between gain and coancestry following penalty index selection can be obtained by avoiding the two extreme solutions: combined-index and within-family selection, and using simple mating designs rather than complex ones.

Key words: breeding value, penalty index, family contribution, gain, group coancestry.

Introduction

Selection has two major consequences that concern the success of a breeding program: genetic gain, and relatedness or group coancestry (simplified to coancestry) among selected individuals. Increase in coancestry is associated with low genetic diversity and high inbreeding, which will suppress the improvement and realization of gain in current and later generations (ROBERTSON, 1961; BULMER, 1971; BURROWS, 1984; CABALLERO, 1994; SANTIAGO and CABALLERO, 1995; WILLIAMS and SAVOLAINEN, 1996; WANG et al., 2004). Efficient methods have been developed to obtain maximum gain such as selection based on combined-index or breeding value (LUSH, 1947; OSBORNE, 1957; HENDERSON, 1984; FALCONER and MACKAY, 1996) or minimum coancestry such as within-family or minimum coancestry selection (ASKEW and BURROWS, 1983; BURROWS, 1984; WEI, 1995). It was not until the mid 1980s, however, that the concept and methodology of simultaneously considering high gain and low coancestry were advocated and developed (e.g. LINDGREN and MATHESON, 1986).

Selection methods that combine gain, genotype contribution, relatedness or effective population size have frequently been proposed recently (e.g. TORO and PEREZ-ENCICO, 1990; QUINTON et al., 1992; VERRIER et al., 1993; WRAY and GODDARD, 1994; BRISBANE and GIBSON, 1995a, 1995b; WEI, 1995; WEI and LINDGREN, 1995; LINDGREN and MULLIN, 1997; ZHENG et al., 1997; FERNÁNDEZ et al., 2003; ROSVALL and MULLIN, 2003). These methods may differ at comparable conditions, e.g. comparing gain at the same selection intensity and coancestry (LINDGREN et al., 1993; WEI and LINDGREN, 1995; WEI and YEH, 2000). FERNÁNDEZ and TORO (1999) compared some of these methods by using mathematical programming technique. A desirable method for a single generation of selection is to increase gain at the same selection intensity and coancestry or to decrease coancestry at the same selection intensity and gain.

Allowing for selection intensity, gain and effective family number (relatedness or coancestry), LINDGREN et al. (1993) and WEI and LINDGREN (1995) derived optimal selection method for populations with unrelated families of large sizes. WEI and YEH (2000) further extended the optimal method to populations with a hierarchical structure. This method ensures that maximum gain is

achieved at given selection intensity and coancestry, and is optimal in balancing gain and coancestry (WEI and LINDGREN, 1995). However, the derived model demands cumbersome search for family contributions. The search work quickly becomes difficult when the hierarchical structure becomes complex (WEI and YEH, 2000).

In this study, we investigate two dynamic indexes that are valid only for selection of one individual each time, and the corresponding selection procedure that differs from traditional one that truncates the top-ranking individuals according to a certain criterion (e.g. phenotypic or breeding value). One index is derived from optimal model developed by WEI and YEH (2000), and another is the linear combination of an individual's breeding value and the coancestry of selected group (WRAY and GODDARD, 1994; BRISBANE and GIBSON, 1995a; LINDGREN and MULLIN, 1997). The main objective is to simplify algorithms that optimise selection in terms of selection intensity, gain and coancestry.

Populations and Basic Models

A population of N non-inbred individuals has a balanced hierarchical structure with k tiers of family constellations (simplified to family hereon unless otherwise stated) (*Fig. 1*). Besides the clustering tree, the legends in the *Fig. 1* also describe the hierarchical structure of the population: 1) j ($=1,2,\dots,k$), the tier of a family nested within any family at tier $j-1$; 2) s_j , the equal number of families that belong to a family at tier $j-1$; 3) s_w , the equal number of individuals that belong to a family at tier k ; 4) ji , the rank (i) of a family at tier j , ranging from 1 to s_j ; 5) wi , the rank (i) of an individual within a family at tier k , ranging from 1 to s_w ; and 6) $1i2i\dots ji$ ($j=1,2,\dots,k,w$), a family or individual at given tier or its relationship with other ordered tiers of families or individuals in the hierarchy.

A j th-tier family is identified by the genetic relationship (r_j) between individuals from different families at tier $j+1$, and the whole population is obtained when no genetic relationship is considered, i.e. $r=0$ (*Fig. 1*). A family constellation becomes small as j and r_j increase ($r_1 < r_2 < \dots < r_j < \dots < r_k$). A nested mating population contains two tiers of families, i.e. half- ($r_1=0.25$) and full-sibs ($r_1=0.5$); a population generated under single-pair or open-pollinated mating is a special or limiting case, having only half- ($r_1=0.25$) or full-sibs ($r_2=0.5$).

Given a character, the phenotypic value of an individual from a population can easily be decomposed into the independent family means at different tiers as deviations from the average of all families in the same tier, and the independent within-family deviation at tier k . Assume all gene effects are small and exclusively additive. Further variance analysis can provide sufficient parameters such as individual, family and within-family heritabilities for best predicting the breeding value ($g_{1i\dots kiwi}$) of any individual involved (WEI and YEH, 2000).

Let n denote the total number of selections from the population, and $n_{1i\dots ji}$ the number of selections contributed by a family at tier j , then,

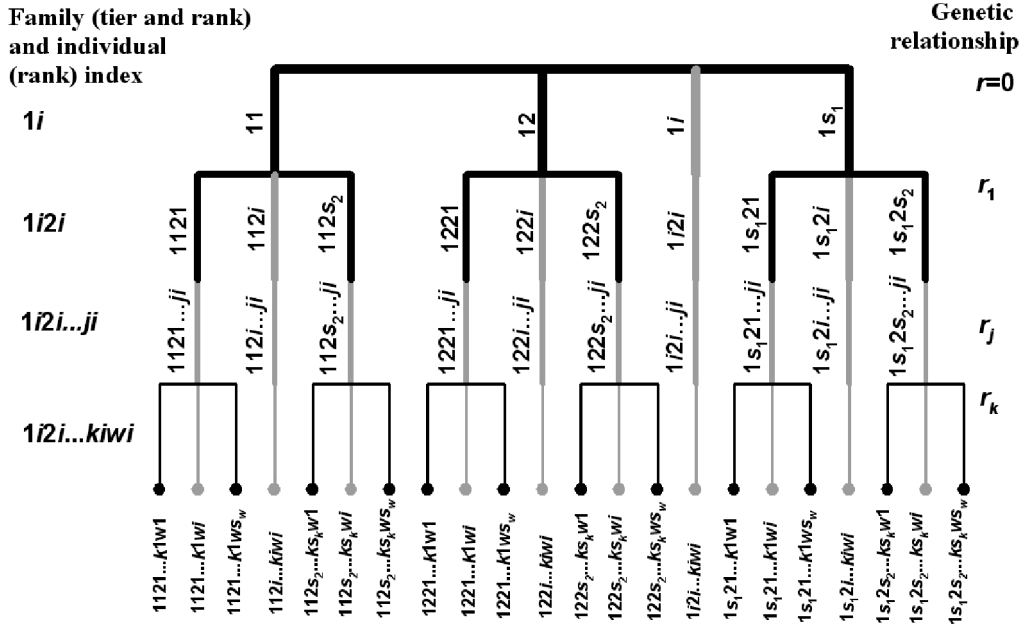


Figure 1. – A diagram to describe the structure of a balanced hierarchical population with k tiers of families clustered by using genetic relationship, r_j ($j=1, \dots, k$, the tier of a family), and $1i2i \dots ji$ ($i=1, \dots, s_j$), a family at tier j or $1i2i \dots kiwi$, a member within a family at tier k . All families at tier j have the same number of members (s_j or s_w).

$$\Sigma \Sigma \dots \Sigma n_{1i2i \dots ki} = \dots = \Sigma \Sigma \dots \Sigma n_{1i \dots ji} = \dots = \Sigma \Sigma n_{1i2i} = \Sigma n_{1i} = n, \quad (1)$$

where $n_{1i \dots ji} = \Sigma n_{1i \dots ji(j+1)}$. Genetic gain is obtained by averaging $g_{1i \dots kiwi}$ over n selected individuals distributed among different tiers of families as expressed in (1),

$$\Delta G = n^{-1} \Sigma \Sigma \dots \Sigma g_{1i \dots kiwi} \quad (2)$$

The group coancestry of n selections is obtained by averaging the pair-relatedness over n^2 pairs of individual combinations including self-pairs in the form (WEI and YEH, 2000),

$$C = 0.5n^{-2} \{n + r_1 \Sigma n_{1i} (n_{1i} - 1) + \dots + [r_j - r_{(j-1)}] \Sigma \Sigma \dots \Sigma n_{1i \dots ji} (n_{1i \dots ji} - 1) + \dots + (r_k - r_{(k-1)}) \Sigma \Sigma \dots \Sigma n_{1i \dots ki} (n_{1i \dots ki} - 1)\}, \quad (3)$$

which can be used to predict the inbreeding level, or to derive the inbreeding effective population size of the offspring generated under random mating.

Stepwise Penalty Index Selection Models

An individual has a certain breeding value but a vague contribution to group coancestry that depends on the interactions among all individuals involved. Taking both gain and coancestry into account is therefore much more complicated than considering gain alone (e.g. FERNÁNDEZ and TORO, 1999). The idea that the contribution to coancestry of an individual is considered in a scale compatible with its breeding value underlies penalty index selection.

Penalty Index I

An optimal selection model considering selection intensity, gain and coancestry from populations with a hierarchical structure was derived for a single generation in the form (WEI and YEH, 2000),

$$g_{1i \dots kiwi} - [d_{1i \dots kiwi} - t(n_{1i \dots ki})] h_w^2 - \lambda_1 - \lambda_2 \{r_1 n_{1i} + [r_2 - r_1] n_{1i2i} + \dots + (r_k - r_{(k-1)}) n_{1i \dots ki} - 0.5 r_k\} = g_{1i \dots kiwi} - [d_{1i \dots kiwi} - t(n_{1i \dots ki})] h_w^2 - \lambda_1 - \lambda_2 \mathbf{r} \mathbf{U}^T = 0 \quad (4)$$

where $g_{1i \dots kiwi}$ is the breeding value of individual $1i \dots kiwi$; $d_{1i \dots kiwi}$, the within-family deviation of individual $1i \dots kiwi$ at tier k ; $t(n_{1i \dots ki})$, the within-family truncation point at tier k , above which $n_{1i \dots ki}$ individuals were selected; h_w^2 , the within-family heritability; λ_1 and λ_2 , the constant parameters; $\mathbf{r} = [r_1, r_2 - r_1, \dots, r_k - r_{(k-1)}, r_k]$; $\mathbf{U} = [n_{1i}, n_{1i2i}, \dots, n_{1i \dots ji}, \dots, n_{1i \dots ki}, 0.5]$; and T , the transpose of a vector. A pair of λ_1 and λ_2 values decides a solution with specified selection intensity, gain and coancestry (WEI and YEH, 2000). The simplest situation with populations of unrelated families that was first studied by LINDGREN et al. (1993) and WEI and LINDGREN (1995), is the limiting case of (4).

Based on (4), a penalty index for individual $1i \dots kiwi$ is constructed in the form,

$$I_{1i \dots kiwi} = g_{1i \dots kiwi} - \beta \mathbf{r} \mathbf{U}^T. \quad (5)$$

Selection based on (5) is denoted by PIS1. An individual's index depends on its breeding value, the coefficient (β) that decides the scale of a family-contribution-dependent deduction or penalty, and the sum of the weighted contributions of families that the individual belongs to. The value of β is confined to $[0, \infty]$ so that one limiting case with $\beta=0$ is exactly combined-index selection that globally maximises gain, and another with $\beta \rightarrow \infty$ is minimum coancestry or within-family selection that globally minimizes coancestry (LINDGREN and WEI, 1994; WEI and LINDGREN, 1995; WEI and YEH, 2000). Increase of β relatively decreases the weight of breeding value in (5), which in turn reduces coancestry in selection.

Penalty Index II

A similar index was early proposed to select superior individuals one by one by imposing a coancestry-dependent deduction on an individual's breeding value (WRAY and GODDARD, 1994; BRISBANE and GIBSON, 1995a, 1995b; LINDGREN and MULLIN, 1997; ZHENG et al., 1997). Coancestry (\hat{C}) is calculated for the individuals obtained during the process of one by one selection rather than the whole set of selected individuals. For the present situation, the index is expressed in the form,

$$J_{1i\dots kiwi} = g_{1i\dots kiwi} - \gamma \hat{C} \quad (6)$$

where γ , as β in (5), decides the scale of a coancestry-dependent deduction or penalty. Selection based on (6) is denoted by PIS2. An individual's index thus depends on its own breeding value, penalty scale and the coancestry (\hat{C}) of itself and selected individuals. Similarly, the domain of γ needs to be confined to $[0, \infty)$. However, β and γ differ in value due to the distinction between the sum of weighted family contributions and the coancestry of the candidate and selected individuals.

Stepwise Selection Procedure

A candidate for selection may have different penalty index values due to the varying β and $r\mathbf{U}^T$ in (5) or γ and \hat{C} in (6) at different steps because one or more individuals are involved. Therefore, a penalty index is dynamic, and could not directly be used as a criterion for truncation selection. To complete selection using these penalty indexes, a 'stepwise' procedure that calculates penalty index and makes selection in a way different from conventional truncation selection is given as follows.

1. Shortlist a small set of potential candidates from the k th-tier families. With one or more unselected individuals, each family at tier k provides one candidate with the highest breeding value. At the start of the procedure, all individuals are unselected.
2. Calculate penalty indexes for the shortlist at a preset value of β or γ . For a candidate, the family contributions ($n_{1i\dots ji}$) that involve the candidate and selected individuals are given by (1). By knowing $n_{1i\dots ji}$ ($1 \leq j \leq k$), $r\mathbf{U}^T$ and $I_{1i\dots kiwi}$ in (5) or \hat{C} and $J_{1i\dots kiwi}$ in (6) are thus obtained.
3. Sort the candidates by their penalty indexes, and select the top one.
4. Repeat from step 1 to 3 until the preset number (n) of individuals is selected.

This stepwise procedure is similar to that used for (6) by WRAY and GODDARD (1994) and LINDGREN and MULLIN (1997). We use the term 'stepwise' to simply emphasize that selection based on a dynamic penalty index is a step-by-step process. Each step of calculation of penalty index and selection screens a single individual with the highest penalty index. The number of the steps involved corresponds to the number of the total selections. It is easy to apply this procedure even though a population is large. Other approach such as mathematical programming (FERNÁNDEZ and TORO, 1999) may also be used to obtain penalty index selection.

Choice of β in (5) or γ in (6) is easy and arbitrary, as long as confined to the domain $[0, \infty)$. A certain value of β or γ turns out specific gain and coancestry at a preset number of selections. Therefore, a predefined coancestry or gain at given number of selections can be obtained by using simple iteration technology for searching corresponding values of β or γ . It should be noted that for finite populations, a predefined coancestry or gain may not exist, but a close value is always available. As β or γ is continuous but the number of combinations of selections with specific coancestry or gain is finite, a range of β or γ values may correspond to the same set of coancestry and gain. Both penalty index methods are valid for any population with a hierarchical structure. In this study, however, we exemplified them by only considering populations generated under open-pollinated, single-pair and nested mating.

Simulation Study

For populations generated under single-pair or open-pollinated mating, there were s_1 unrelated families, each of s_w members. For populations generated under nested mating, there were s_1 male parents, each of them mated to s_2 female parents. Thus, there were s_1 half-sib families, and $s_1 s_2$ full-sib families, each of s_w members. The variance components and heritability parameters for these populations were summarized in details by WEI and YEH (2000). A population has zero mean and unit total phenotypic variance. Given census population structure and individual heritability (h^2), normal random variables or order statistics were used to simulate the family means and within-family deviations, respectively. Order statistics provides approach to study finite populations without random error.

When single-pair ($r_1 = 0.5$) or open-pollinated ($r_1 = 0.25$) mating was considered, there was only one family tier ($k = 1$), and (5) reduced to,

$$I_{1iwi} = g_{1iwi} - \beta r_1 (n_{1i} - 0.5), \quad (7)$$

and (6) to,

$$J_{1iwi} = g_{1iwi} - \gamma \hat{C}. \quad (8)$$

Eqn (7) was equivalent to the penalty index developed for unrelated families by WEI et al. (1998). In a population generated under nested mating, there were two tiers ($k = 2$) of families (half-sibs, $r_1 = 0.25$, and full-sibs, $r_2 = 0.5$). Eqns (5) and (6) became,

$$I_{1i2iwi} = g_{1i2iwi} - \beta [r_1 n_{1i} + (r_2 - r_1) n_{1i2i} - 0.5 r_2], \quad (9)$$

and,

$$J_{1i2iwi} = g_{1i2iwi} - \gamma \hat{C}. \quad (10)$$

respectively. Obviously, a candidate's penalty index depends on its breeding value and family contributions (n_{2i} and/or n_{2i1i}) or the coancestry (\hat{C}) of individuals involved.

In simulating selection based on (7) to (10), we mainly investigated: 1) the superiority and nature of stepwise penalty selection, 2) the difference between two penalty

indexes, and 3) the relationship between gain (ΔG) and coancestry (C) following penalty index selection. To make 1) and 2) simple but sensible, small populations were considered: $s_1=6$ and $s_w=6$ for single-pair or open-pollinated mating; $s_1=3$, $s_2=2$ and $s_w=6$ for nested mating; $h^2=0.25$. The selected number (n) was set to 6. We compared PIS1 and PIS2 solutions with all comparable combinations of the same number (6) of individuals. For each population, there are C_{36}^6 possible combinations (nearly 2 millions!) in total. Each combination has specific ΔG and C . A PIS1 or PIS2 solution should correspond to the combination with the highest ΔG at given C or the lowest C at given ΔG . For a realistic but valid comparison, we can downgrade C_{36}^6 combinations to a small portion (462) of 6-subsets by simply: 1) treating any number (1 to 6) of the top-ranking individuals in breeding value from a family as a single element, and 2) picking unordered elements without repetition from 6 full-sib families. This portion of subsets includes all combinations with highest ΔG at given C .

For each type of mating design, 100 populations at $h^2=0.25$ were simulated to examine PIS1 and PIS2 solutions and 462 6-subsets of individuals. As an example, ΔG against C was plotted on plane figure for PIS1 and PIS2 solutions as well as 462 subsets from three typical populations generated under nested mating (Fig. 2). Additional simulations were run for these populations at different h^2 . Besides, the equivalence and difference between PIS1 and PIS2 was investigated in terms of the proportions of simulated populations for 3 mating designs and $h^2=0.01, 0.05$ and $0.1(0.1)1$ (Fig. 3). For each mating design and h^2 , 10,000 populations were simulated.

We only considered PIS1 and two types of populations in examining the relationship between ΔG and C : single-pair mating, $s_1=50$ and $s_w=200$, and nested mating, $s_1=25$, $s_2=3$ and $s_w=133$. Individual heritability (h^2) was set to 0.1 and 0.25, and the selected number was set to 100, which was equal to the number of parents for each mating design. Family means and within-family deviations for a population were mimicked by order statistics. Varying β in eqns (7) and (9) resulted in a series of solutions with different ΔG and C including breeding value and within-family selection, which exhibited a special pattern of relationship (Fig. 4).

Results

A PIS1 or PIS2 solution was always the combination of individuals with the highest gain (ΔG) at the given coancestry (C) or the lowest C at given ΔG (e.g. Fig. 2). However, PIS1 and PIS2 did not result in solutions at all C where there existed combinations with the highest ΔG . PIS1 or PIS2 might exclusively produced one or few solutions in a population (Fig. 2b, c). All additional simulations exhibited the same patterns of results presented in Fig. 2. PIS1 and PIS2 generated identical solutions in most of the populations simulated, and most identical solutions in the remaining populations (Fig. 3).

Some combinations of individuals with maximum ΔG at given C were not PIS1 or PIS2 solutions, depending

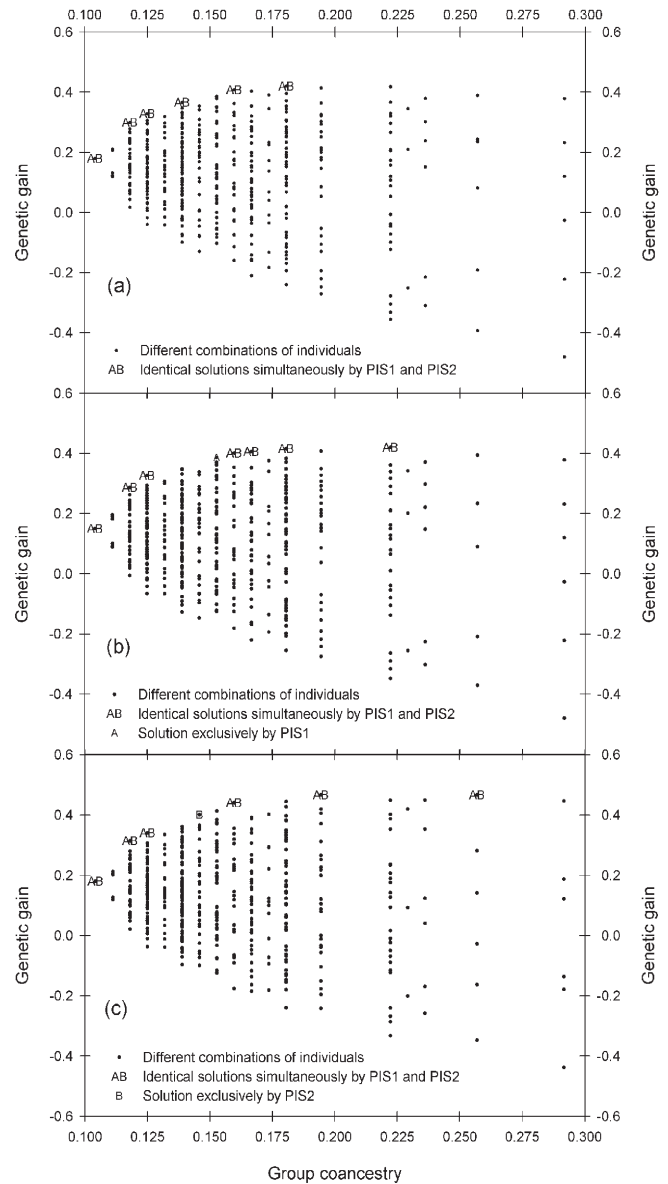


Figure 2. – Gain against group coancestry for PIS1 and PIS2 solutions and 462 comparable combinations of 6 individuals from 3 nested mating ($s_1=3$, $s_2=2$ and $s_w=6$) populations (a, b and c respectively) at $h^2=0.25$. The later included those with the highest gain at given coancestry. The comparison illustrated the superiority (a, b and c), and the equivalence and/or minor discrepancy of PIS1 and PIS2: 1) all identical solutions (a), 2) 7 identical solutions and 1 exclusively by PIS1 (b), and 3) 6 identical solutions and 1 exclusively by PIS2 (c).

on populations (Fig. 2). They might have ΔG close to the norm of PIS or significantly lower than a PIS solution at lower adjacent C . Penalty index selection solution with $\beta < 0$ or $\gamma < 0$ resulted in both lower ΔG and higher C than combined-index selection (WEI and LINDGREN, 1995). PIS1 with $\beta < 0$ and PIS2 with $\gamma < 0$ were therefore not efficient, and were excluded from further investigation below unless otherwise stated.

PIS1 and PIS2 exhibited high equivalence in producing selection solutions. While both methods were equivalent in producing identical solutions in more than 93.5%, 83.9% and 61.1% of 10,000 populations simulated at specific h^2 and under open-pollinated, single-pair

and nested mating, respectively (Fig. 3a), they were also very close in the remaining populations (WEI and YEH, 2000). They yielded most identical solutions, e.g. 7 and 6 identical solutions with one exception from the populations presented in Fig. 2b, c, respectively. The exceptional solution exclusively produced by one method was actually highly close to the norm of the solutions by the other (Fig. 2b, c). PIS1 (PIS2) produced additional solutions in less than 9.8% (29.1%), 2.2% (14.0%) and 0.3% (6.3%) of the populations generated under nested, single-pair and open-pollinated mating, respectively (Fig. 3b, c). Increase in h^2 enhanced the equivalence between PIS1 and PIS2 (Fig. 3).

Genetic gain changed in a typical pattern over the range of C following PIS1 from populations with different structures (Fig. 4). Starting at the lowest ΔG associ-

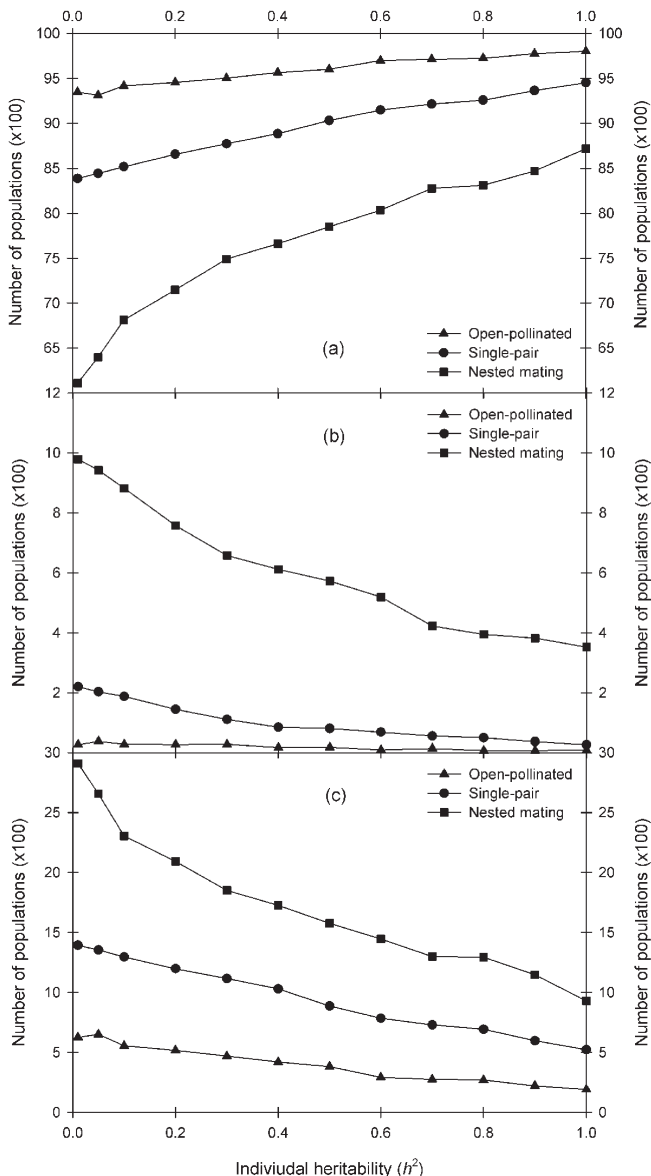


Figure 3. – The numbers of populations where PIS1 and PIS2 produced all identical solutions (a), and most identical solutions except for one or few solutions exclusively produced by PIS1 (b) or PIS2 (c) at different h^2 for open-pollinated ($s_1=6$ and $s_w=6$), single-pair ($s_1=6$ and $s_w=6$) and nested mating design ($s_1=3$, $s_2=2$ and $s_w=6$). Six individuals were selected.

ated with minimum C , it rose dramatically at low C , increased with decreasing rate as C increased, and reached the highest value associated with combined-index selection (WEI et al., 1998). Penalty index solutions associated with globally minimum C and maximum ΔG were identical to within-family and combined-index selection, respectively. It appeared that change in ΔG over C was rather smooth (Figs. 2 and 4).

Single-pair mating resulted in a wider range of C as well as ΔG with both lower and higher limits than nested mating at the same parent number and similar population sizes (Fig. 4). Single-pair mating always yielded higher ΔG than nested mating at given C . The superiority was compiled in Table 1 for some C values including those associated with within-family and combined-index selection from populations with nested mating. The superiority first rapidly decreased from the highest value at minimum C to the bottom at a certain value of C , and then trivially bounced up as C approached to its upper limit associated with combined-index selection (Fig. 4, Table 1). At high h^2 , the absolute difference was large but the relative superiority was small. The superiority of single-pair over nested mating was enlarged as the male number of nest mating was reduced (results not presented).

Discussion and Conclusions

Optimality and/or Approximation

Although PIS1 is based on an optimal model (WEI and YEH, 2000) and PIS2 are suggested to maximize gain (ΔG) at given coancestry (C) (e.g. WRAY and GODDARD, 1994; LINDGREN and MULLIN, 1997), we could not explicitly prove their optimality in the present study. All of our simulation results however confirmed that PIS1 and PIS2 always maximized ΔG at given C or minimized C at given ΔG (e.g. Fig. 2). The superiority of PIS1 over other selection methods was investigated in an early study considering populations of unrelated families (WEI et al., 1998).

PIS1 and PIS2 generally yielded the same results, but might slightly differ in one or few exceptions (see below). We were not clear how this difference arose, but it might imply bias from optimality of PIS1 or PIS2 at corresponding C . We incline to infer that both PIS1 and PIS2 generally generate optimal solutions, but PIS2 may occasionally mistake one or few combinations of individuals with highest ΔG at certain C , or miss one or few real optimal solutions. Yet, this inference lacks direct and strong proofs. A PIS2 solution, if not optimum, is still very close to being optimal (Fig. 2), and can thus be treated as a good approximation of optimal selection (WEI and YEH, 2000). On the other hand, missing solution at prefixed C is an undesirable feature although the possibility is low (Fig. 3b).

Local Optimum Solution

For all combinations of individuals from a population with specific size and family structure, there exist a certain number of C values. For example, there are 18 values for the simulation presented in Figs. 2. A combina-

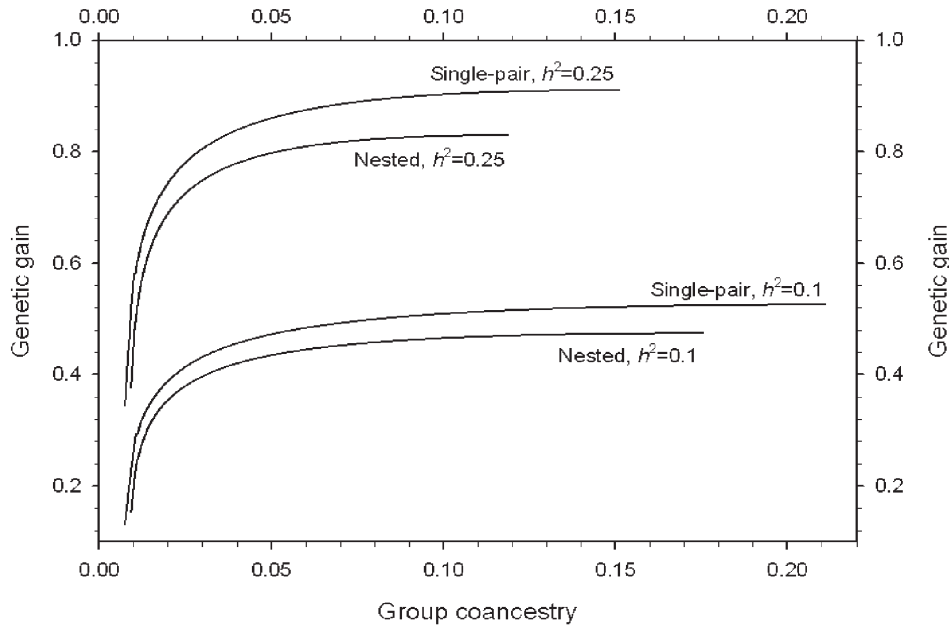


Figure 4. – Relationship between genetic gain and group coancestry following PIS1 from populations under single-pair ($s_1 = 50$ and $s_w = 200$) and nested mating ($s_1 = 25$, $s_2 = 3$ and $s_w = 133$) at $h^2 = 0.1$ and 0.25 . One hundred individuals were selected.

Table 1. – Superiority of single-pair ($s_1 = 50$ and $s_w = 200$) over nested mating ($s_1 = 25$, $s_2 = 3$ and $s_w = 133$) at $n = 100$ for balancing gain and coancestry in penalty index selection (PIS1).

h^2		Group coancestry (C)					
		0.0094 ^a	0.050	0.100	0.1188 ^b	0.150	0.1756 ^c
0.1	Extra ΔG	0.079	0.042	0.043	0.044	0.044	0.047
	Superiority (%)	51.41	9.58	9.34	9.31	9.28	9.90
0.25	Extra ΔG	0.145	0.064	0.067	0.079	-	-
	Superiority (%)	38.37	8.08	8.07	9.49	-	-

^a the value of C associated with within-family selection from the nested-mating populations at $h^2 = 0.25$ and 0.1 , respectively;

^{b,c} the values of C associated with combined-index selection from the nested-mating populations at $h^2 = 0.25$ and 0.1 , respectively.

tion of individuals that maximizes ΔG at given C is not necessarily a PIS solution. Such a combination is referred to as local optimal solution, in contrast to PIS solutions. Local optimal solutions can be classified into 3 groups. First, PIS solutions with $\beta < 0$ or $\gamma < 0$. Although mathematically maximizing ΔG at given C , PIS1 with $\beta < 0$ or PIS2 with $\gamma < 0$ is biologically inferior to PIS1 with $\beta = 0$ or PIS2 with $\gamma = 0$ (LINDGREN and WEI, 1994; WEI and LINDGREN, 1995). Second, non-PIS solutions with lower ΔG but higher C than one of their two adjacent PIS solutions (Fig. 2). Finally, other non-PIS solutions that include solutions with higher ΔG and C (Fig. 2), this group is probably biologically worse than their neighbouring PIS solutions.

PIS1 vs. PIS2

Two methods differ in the term of the relationship among the individuals involved (eqns 7–10). PIS1 con-

siders only the relationship a concerned individual has with other selections, while PIS2 includes all interactions among involved individuals (LINDGREN and MULLIN, 1997; ZHENG et al., 1997; WEI and YEH, 2000). However, PIS1 and PIS2 are identical or almost equivalent in producing selection solutions (Figs. 2 and 3a). PIS1 (or PIS2) may occasionally yield one or few exclusive solutions, particularly at low h^2 (Fig. 3b, c), which does not influence the equivalence of both, though.

Choice of a PIS Solution in Application

Although PIS provides a series of potential solutions with different ΔG and C (Fig. 4), a single solution usually needs to be decided in practice. When ΔG or C is an overwhelming consideration, combined-index or within-family selection is a sole choice to maximize ΔG (LUSH, 1947; OSBORNE, 1957; HENDERSON, 1984; FALCONER and MACKAY, 1996) or minimize C (ASKEW and BURROWS,

1983; WEI and LINDGREN, 1995). It becomes complex for choice of a solution between the two extremes. Obviously, knowledge about the role of coancestry in breeding is essential for a good choice. Nevertheless, the diagrammed relationship between ΔG and C (Fig. 4) is still very useful to assist balancing ΔG and C in selection. One of the most important features illustrated in Fig. 4 is that at low value of C , much ΔG can be improved by tolerating minor increase in C , while at high value, C may be dramatically reduced by sacrificing little ΔG (also see Fig. 2; WEI and LINDGREN, 1995; WEI et al., 1998).

Effect of Mating Design

A hierarchical breeding population may be associated with different mating designs in practical breeding, such as open-pollinated, single-pair and nested mating. In selection, mating design therefore plays a role in the improvement of breeding value and control of coancestry (e.g. van BUIJTENEN and BURDON, 1990; WEI et al., 2002). For balancing ΔG and C , a simple mating design is superior to a complicated one such as single-pair to nested mating under comparable conditions (Fig. 4). The superiority is even evident when using conventional selection methods: within-family and combined-index selection (Table 1). We should keep in mind, however, that a complicated mating has other irreplaceable features. For example, extra information can be obtained to accurately predict an individual or family's breeding value (ENDERSON, 1984; WHITE and HODGE, 1988; FALCONER and MACKAY, 1996); non-additive genetic variance can be estimated to evaluate the efficiency of use of interspecies crossing; specific combining ability is an important criterion for selecting a hybrid cross for production purpose.

Other Implications

Although this study concerns balanced hierarchical populations of non-inbred individuals (Fig. 1), PIS2 can directly be applied to all other situations (WRAY and GODDARD, 1994; BRISBANE and GIBSON, 1995a, 1995b; LINDGREN and MULLIN, 1997; ZHENG et al., 1997). PIS1 may also be extendable to other types of populations by properly modifying the term rU^T in (5). In practice, the actual requirements for applying PIS1 or PIS2 are the known breeding values and pedigree structure of individuals contained in a population. Some assumptions or parameters in this study, such as balanced hierarchical population, heritability, variance components and random and normal-distributed variables, are thus not necessary.

References

ASKEW, G. R. and P. M. BURROWS (1983): Minimum coancestry selection I. A *Pinus taeda* population and its simulation. *Silvae Genet.* **32**: 125–131.
 BRISBANE, J. R. and J. P. GIBSON (1995a): Balancing selection response and rate of inbreeding by including genetic relationship in selection decisions. *Theor. Appl. Genet.* **91**: 421–431.
 BRISBANE, J. R. and J. P. GIBSON (1995b): Including genetic relationships in selection decisions: alternative methodologies. *Theor. Appl. Genet.* **91**: 769–775.

BULMER, M. G. (1971): The effect of selection on genetic variability. *Am. Nat.* **105**: 201–211.
 BURROWS, P. P. (1984): Inbreeding under selection from related families. *Biometrics* **40**: 895–906.
 CABALLERO, A. (1994): Developments in the prediction of effective population size. *Heredity* **73**: 657–679.
 FALCONER, D. S. and T. F. C. MACKAY (1996): Introduction to quantitative genetics. Longman Group, London, UK.
 FERNÁNDEZ, J. and M. A. TORO (1999): The use of mathematical programming to control inbreeding in selection schemes. *J. Anim. Breed. Genet.* **116**: 447–466.
 FERNÁNDEZ, J., M. A. TORO and A. CABALLERO (2003): Fixed contributions designs vs. minimization of global coancestry to control inbreeding in small populations. *Genetics* **165**: 885–894.
 ENDERSON, C. R. (1984): Applications of Linear Models in Animal Breeding. University of Guelph, Guelph ON, Canada.
 LINDGREN, D. and A. C. MATHESON (1986): An algorithm for increasing the genetic quality of seed from seed orchards by using the better clones in higher proportions. *Silvae Genet.* **35**: 173–177.
 LINDGREN, D. and R.-P. WEI (1994): Gain versus effective number. p. 164–177. *In*: Progeny testing and breeding strategies (edited by LEE, S. J.), Proc. of the Nordic Group for Forest Genetics and Tree Breeding, Edinburgh, Oct. 6–10, 1993, Forestry Authority.
 LINDGREN, D. and T. J. MULLIN (1997): Combining high gain and low coancestry. *Silvae Genet.* **46**: 124–129.
 LINDGREN, D., R.-P. WEI and L. F. BONDESSON (1993): Optimal selection from families. *Heredity* **70**: 619–621.
 LUSH, J. L. (1947): Family merit and individual merit as bases for selection. *Am. Nat.* **81**: 241–261 and 262–379.
 OSBORNE, R. (1957): The use of sire and dam family averages in increasing the efficiency of selective breeding under a hierarchical mating system. *Heredity* **11**: 93–116.
 QUINTON, M., C. SMITH and M. E. GODDARD (1992): Comparison of selection methods at the same level of inbreeding. *J. Anim. Sci.* **70**: 1060–1067.
 ROBERTSON, A. (1961): Inbreeding in artificial selection programmes. *Genet. Res. Camb.* **2**: 189–194.
 ROSVALL, O. and T. J. MULLIN (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.
 SANTIAGO, E. and A. CABALLERO (1995): Effective size of populations under selection. *Genetics* **139**: 1013–1030.
 TORO, M. and M. PEREZ-ENCISO (1990): Optimization of selection response under restricted inbreeding. *Genet. Sel. Evol.* **22**: 93–107.
 VAN BUIJTENEN, J. P. and R. D. BURDON (1990): Expected efficiencies of mating designs for advanced-generation selection. *Can. J. For. Res.* **20**: 1648–1663.
 VERRIER, E., J. J. COLLEAU and J. L. FOULLEY (1993): Long-term effects of selection based on the animal model BLUP in a finite population. *Theor. Appl. Genet.* **87**: 446–454.
 WANG, T., S. N. AITKEN, J. H. WOODS, K. POLSSON and S. MAGNUSSEN (2004): Effects of inbreeding on coastal Douglas fir growth and yield in operational plantations: a model-based approach. *Theor. Appl. Genet.* **108**: 1162–1171.
 WEI, R.-P. (1995): Predicting genetic diversity and optimizing selection in breeding programmes. PhD Thesis, Swedish University of Agricultural Sciences, Umeå, Sweden, ISBN 91-576-4900-1

- WEI, R.-P. and D. LINDGREN (1995): Optimal family contributions and a linear approximation. *Theor. Pop. Biol.* **48**: 318–332.
- WEI, R.-P. and F. C. YEH (2000): Optimal family contributions in selection toward maximum gain under restricted average coancestry. *Silvae Genet.* **49**: 281–290.
- WEI, R.-P., F. C. YEH and N. K. DHIR (2002): Investigation of status number following selection from populations under different mating designs. *Silvae Genet.* **51**: 87–92.
- WEI, R.-P., C. R. HANSEN, N. K. DHIR and F. C. YEH (1998): Genetic gain with desired status number in breeding programs: a study on selection effects. *Can. J. For. Res.* **28**: 1861–1869.
- WHITE, T. L. and G. R. HODGE (1988): Best linear prediction of breeding values in a forest tree improvement program. *Theor. Appl. Genet.* **76**: 719–727.
- WILLIAMS, C. G. and O. SAVOLAINEN (1996): Inbreeding depression in conifers: implications for using selfing as a breeding strategy. *For. Sci.* **42**: 102–117.
- WRAY, N. R. and M. E. GODDARD (1994): Increasing long-term response to selection. *Genet. Sel. Evol.* **26**: 431–451.
- ZHENG, Y. Q., D. LINDGREN, O. ROSVALL and J. WESTIN (1997): Combining genetic gain and diversity by considering average coancestry in clonal selection of Norway spruce. *Theor. Appl. Genet.* **95**: 1312–1319.

Accuracy of Ranking Individuals in Field Tests of Different Designs: A Computer Simulation

By C.-Y. XIE^{1,*}, Y.-B. FU² and A. D. YANCHUK¹

(Received 4th October 2005)

Abstract

A computer simulation was conducted to investigate the accuracy of ranking individual trees in field tests of different designs. A test population that consists of 900 trees from 45 full-sib families generated by three 6-parent disconnected half-diallels was considered. One incomplete block design with single-tree plots and four complete block designs with 1, 2, 4, and 10-tree row plots were examined. Various narrow-sense heritabilities, site variation patterns (patch sizes and gradient slopes), and two levels of dominant to additive genetic variance ratios (0 and 0.30) were evaluated. Results indicate that the accuracy of ranking depends more on the heritability of the trait and less on the magnitude of the dominant genetic variance, site variation patterns, and field designs. With patchy site variation, differences in ranking accuracy were observed for different designs, but became smaller with higher heritabilities. Impact of environmental gradient on the accuracy of individual ranking was negligible. Incomplete block design with single-tree plots (ICB1) provided the most accurate ranking when heritability was low while complete block design with 2-tree plots (RCB2) appeared to be the best when heritability was high. Large row plot designs were among the least effective in all the simulated scenarios. For traits with medium heritabilities, the statistical efficiencies of ICB1 and RCB with 1 and 2-tree plots are comparable.

Key words: breeding value, progeny testing, field design, ranking accuracy, computer simulation.

¹ British Columbia Ministry of Forests, Research Branch, P.O. Box 9519, Victoria, BC V8W 9C2, Canada.

² Plant Gene Resources of Canada, AAFC Saskatoon Research Centre, 107 Science Place, Saskatoon, SK S7N 0X2, Canada.

* Corresponding Author: CHANG-YI XIE, Tel: (250) 387-8911, Fax: (250) 387-0046. E-mail: chang-yi.xie@gov.bc.ca.

Introduction

Progeny testing is one of the most important and expensive phases of a tree-breeding program. To make maximum progress from selection, it is necessary to use a field design that ranks the test materials as closely as possible to the ranking of their true genetic merit, or breeding values. Single-tree plot (STP) designs (with either randomized complete or incomplete blocks) have long been recognized as the best for ranking families, or parents, for backward selection (CONKLE, 1963; WRIGHT, 1976; LIBBY and COCKERHAM, 1980). As they maximize the number of micro-sites sampled with a given number of trees per family, STP designs minimize the errors in estimating family means (or parental breeding values) and thus maximize the accuracy of family ranking. In tree breeding, forward selection is another common practice, which requires accurate ranking of individuals.

The breeding value of an individual (IBV) consists of two components corresponding to family and within-family additive genetic effects. Although STP designs are efficient for evaluating the first component (i.e., parental breeding values), they may not be for the second (i.e., within-family breeding value) since they are unable to provide assessment of plot effects for micro-site adjustments to the observed individual phenotypes. Hence STP designs may not always be superior to multiple-tree plot (MTP) designs in ranking individuals, and their relative superiority may vary with the heritability of the trait and the patterns of environmental variation (VAN BUIJTENEN, 1983).

In this paper, we report the results from a computer simulation on the accuracy of ranking individuals in field tests of different designs. This simulation consid-