M. J. W. and Sherman-Broyles, S. L.: Factors influencing levels of genetic diverse in woody plant species. New Forests **6**, 95–124 (1992). — Kim, Z. S. and Lee, S. W.: Genetic diversity of three native *Pinus* species in Korea. In: Population genetics and genetic conservation of forest trees. Edited by Ph. Baradat, W. T. Adams and G. Müller-Starck. SPB Academic Publishing, Amsterdam, The Netherlands. pp. 211–218 (1995). — Kim, Z. S., Lee, S. W., Lim, J. H., Hwang, J. W. and Kwon, K. W.: Genetic diversity and structure of natural populations of *Pinus koraiensis* (Sieb. et Zucc.) in Korea. Forest Genetics **1**, 41–49 (1994). — Lee, D. H.: Geographic variation in leaf and seed characteristics of natural populations of *Pinus thunbergii* Parl. in Korea. M. A. thesis. Yeungnam University. (in Korean). (1989). — Mirov, N. T.: The genus *Pinus*. The Ronald Press Company, New York (1967). — Miyata, M. and Ubukata, M.: Genetic variation of allozymes in natural stands of Japanese black pine. (in Japanese). J. Jpn. For. Soc. **76**, 445–455 (1994). — Nei, M.: Analysis of gene diversity in subdivided populations. Proc. Nat. Acad. Sci. **70**, 3321-3323 (1973). — Nei, M.: Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics **89**, 583–590 (1978). — Shin, D. H.: Inheritance and linkage of allozymes in *Pinus thunbergii* Parl. M. A. Thesis. Korea University (1987). — Shiraishi, S.: Linkage relationships among allozyme loci in Japanese black pine, *Pinus thunbergii* Parl.. Silvae Genetica **37**, 60–66 (1988a). — Shiraishi, S.: Inheritance of isozyme variations in Japanese black pine, *Pinus thunbergii* Parl.. Silvae Genetica **37**, 93–100 (1988b). — Sneath, P. H. A and Sokal, R. R.: Numerical Taxonomy. W. H. Freeman and Co., San Francisco, CA (1973). — Swofford, D. L. and Selander, R. B.: BIOSYS-1: A computer program for the analysis of allelic variation in population genetics and biochemical systematics. Release 1.7. Illinois Natural History, Survey, IL (1989).

Balancing Gain and Relatedness in Selection

By D. Lindgren[1]) and T. J. Mullin[2])

(Received 23rd January 1997)

Summary

A population merit criterion, B_ω, for a set of genotypes, ω, is formulated as $B_\omega = \bar{g}_\omega - c\Theta_\omega$where, c is a weighting constant, $\bar{g}_w$ is the average of their breeding values, and Θ_ω is the average coancestry of the considered genotypes, which is a measure of their relatedness. The breeding objective studied here is selecting the set ω that maximises B_ω. An iterative search algorithm is proposed for finding this maximum under a given breeding-population size. This algorithm was applied to an example using simulation techniques. Results were presented as graphs where the gain was plotted against the status effective number, which was used to quantify the degree of relatedness as an inverse function of average coancestry. For all except extreme c values the algorithm gave markedly better combinations of gain and average coancestry when compared with a conventional method to control relatedness by restricting contributions from individual parents.

Key words: computer simulation, diversity, effective population size, inbreeding, genetic base, status number, selection response, coancestry, kinship.

FDC: 165.3/.4; 165.6

Introduction

Breeders always face a trade-off between 2 basic desiderata. While they are expected to produce genetic gain, they are also expected to control relatedness so that genetic diversity is conserved. The challenge is first to formulate a breeding objective considering both gain and relatedness, and then to apply a selection procedure that somehow affords an optimal trade-off.

The problem of balancing genetic gain and diversity of the genetic base by including relatedness in selection decisions has been approached by several investigators (e.g., Toro and Pérez-Enciso, 1990; Quinton and Smith, 1995; Wei, 1995; Brisbane and Gibson, 1995). Quinton *et al.* (1992) introduced comparisons of breeding methods at the same level of inbreeding, and it seems to be common to regard inbreeding as the entity that must be compromised in the pursuit of gain (e.g., Caballero *et al.*, 1996). Recently, combining consideration of breeding value and relatedness among selections has been suggested as a way to achieve that goal (Wray and Goddard, 1994). Brisbane and Gibson (1995) developed a selection algorithm amounting to the estimated breeding values of the selected individuals, minus their assumed influence on average coancestry, in order to maximise gain in relation to the genetic base maintained.

Wei (1995) showed how gain and diversity may be pursued for a single breeding cycle, assuming a symmetric population structure with all individuals equally inbred. For deployment of unrelated clones in forestry, either as parents in seed orchards or as mass-propagated clonal mixtures, the problem has a simple solution, requiring only that the frequencies of deployed genetic entries vary in linear relationship to their predicted genetic values (Lindgren, 1986; Lindgren *et al.*, 1989). While this method was also shown to serve as a good approximation for families (Wei and Lindgren, 1995), where a similar optimising method can be used Lindgren *et al.*, 1993), such methods are only applicable in simple situations (which, however, are of major practical importance in forest tree breeding) and are somewhat complicated to use.

The aims of the current study were to formulate an expression for the merit of a given set of genotypes considering their relatedness in conjunction with their breeding values as a composite breeding objective, and to suggest a way to optimise the approach to that objective.

---

[1]) Department of Forest Genetics and Plant Physiology, Swedish University of Agricultural Sciences, 901 83 Umeå, Sweden

[2]) Genesis Forest Science Canada Inc., C. P 64, Succursale Haute-Ville, Québec (QC) G1R 4M8, Canada

Methods

Theoretical development of the breeding objective

To describe the state of relatedness in a population, we must consider relationships *among* individuals. *Coancestry* is the probability that genes sampled from corresponding loci in a pair of individuals are identical by descent. Coancestry of a pair of individuals is equivalent to the inbreeding of their progeny if they are bred. The pairing may be an individual with itself, in which case we speak of *self-coancestry*. *Average coancestry* of a population is the average of the coancestry for all possible pairs of individuals in the population, including self-coancestry. In a population of n individuals, there are n^2 different pairings to consider when calculating its average coancestry. Average coancestry is the expected average inbreeding that would occur after random mating, including reciprocals and selfs, and regardless of sex or other mating barriers, and is independent of any assumptions relating to the mating behaviour of the considered population. Average coancestry is also equivalent to the probability that two genes sampled from the gene pool of the population will be identical by descent.

We suggest formulating the "merit" of a set ("population") of genotypes as

$$B_{\omega} = \bar{g}_{\omega} - c\Theta_{\omega} \qquad (1)$$

where B_{ω} is the "population merit"; ω is the set of genotypes which constitute the considered population; $\bar{g}_{w}$ is the average of the breeding values of ω; c is a weighting constant; and Θ_{ω} is the average coancestry of ω.

An equivalent formulation is

$$B_{\omega} = \bar{g}_{\omega} - \frac{c}{2N_{s_{\omega}}} \qquad (2)$$

where $N_{s_{\omega}}$ is the "status effective number" of ω, defined by LINDGREN *et al.* (1996) as half the inverse of average coancestry. This latter formulation (eq. 2) is preferred in presenting the results, as "effective number" is better understood than the concept of "average coancestry" and makes more clear what the number actually means. This status effective number has some attractive properties, e.g., it is equivalent to the actual census number of a set of non-inbred, unrelated genotypes. The factor 2 in the denominator is applicable when considering diploid organisms.

We formulate *the objective of selection as finding a population ω, which maximises the population merit,* B_{ω}.

A selection algorithm for maximising population merit

There is no known procedure to identify the unique set of individuals that maximises B_{ω}, nor to validate a suggested maximum, other than by an exhaustive consideration of all possible sets. We present here a selection algorithm that seeks to maximise the population merit of selecting N genotypes by means of an iterative procedure. We will call our procedure "population-merit selection". The algorithm evaluates each candidate before each decision by calculating

$$B_{ij\omega} = \hat{g}_{ij} - c\Theta_{\omega} \qquad (3)$$

where $B_{ij\omega}$ is the prediction of population merit that will result if the ith member of the jth family is added to (or subtracted from) the already selected population so that the new selected population becomes ω; $\hat{g}_{ij}$ is the predicted breeding value for this individual; c is a penalty weight describing the relative value of coancestry against additive genetic gain; and Θ_{ω} is the resulting average coancestry.

In this way, a candidate for selection is penalised by the average coancestry that would result if it were included in the population together with a given set of already selected members. The advantage a candidate offers is compared with other candidates for a particular stage of selection. Similarly, one can also compare individuals considered for substitution of already selected genotypes, and thus search for a set of members that maximises the population merit. The penalty weight can be set to give the desired balance between gain and relatedness. A penalty of zero would obviously maximise expected gain, while an infinite penalty would minimise relatedness and an infinitely negative penalty would maximise relatedness.

The applied procedure to select N individuals can be described as follows:

1. Breeding values are predicted for each candidate, by conventional methods such as combined-index, BLUP, etc.
2. A value for the penalty weight c is chosen.
3. The subtraction term $c\Theta_{\omega}$ for each family is calculated, in turn, by considering the resulting average coancestry in the selected population, ω, after the addition of another individual from this family. This term is subtracted from the predicted breeding values for each member of the family to estimate the population merit after the addition of that individual (eq. 3). Note that Θ_{ω} is self-coancestry during the initial pass of the algorithm as it selects the first member of ω.
4. The candidate individuals are ranked for the selection criterion and the highest-ranking candidate is added to the breeding population.
5. Steps 3 and 4 are repeated until N selections have been made, and the population merit of this set is recorded.
6. Steps 3 and 4 are repeated to add an N+1th individual to the selected set.
7. Each of the N+1 selected genotypes is then removed, in turn, from the breeding population and the population merit of the remaining N individuals calculated. The new set of N individuals giving the maximum population merit is identified.
8. If the population merit given by this new set of N individuals is less than before, the previous set of N individuals is retained and the selection algorithm is complete. Otherwise, this new set of N individuals is retained, and Steps 6 through 8 are repeated.

What the algorithm actually does is to first select N individuals consecutively, and then to substitute earlier selections by alternatively adding and discarding individuals until the process fails to find a set of N individuals with higher population merit. Note that a family will always be represented by its top-ranking phenotypes.

The most onerous part of the selection process is the recalculation of the average coancestry for the set ω considered for selection. This requires analysis of the resulting pedigree and recalculation of the average coancestry for many different populations as each of the families is considered for addition to, or removal from, the selected population. Computer algorithms for pedigree analysis are available that can efficiently calculate the inbreeding coefficients of all members of a pedigree (TIER, 1990; MEUWISSEN and LUO, 1992). These algorithms are easily adapted to calculate average coancestry LINDGREN *et al.*, 1997). Average coancestry can be derived from a matrix of the additive relationship. Before these algorithms and fast computers were available, the task of calculating average coancestry was too large for the numbers of individuals in real-life breeding populations. An EXCEL programme (COADEMO.XLS) has been made to illustrate the algorithm and is available on the world wide web at http://linne.genfys.slu.se/breed/breed.htm.

Computer simulation

We illustrate the application of population-merit selection by means of a simple simulation study. A computer simulation was carried out on a closed population, using stochastic modelling techniques described by MULLIN and PARK (1995). The quantitative genetic model assumed infinitesimal additive gene effects, normally distributed with variance of 100, where the initial population was assumed to be in linkage equilibrium (FOULLEY and CHEVALET, 1981). Environmental effects were normally distributed around an arbitrary trait mean 100 with variance of 1900, giving an initial narrow-sense heritability $h^2 = 0.05$.

A breeding population with $N = 40$ individuals was selected in each generation. The initial 40 individuals were non-inbred and unrelated. Individuals in the breeding population were single-pair mated, producing 20 full-sib families (no half-sib families) with a fixed family size of 200, in which the next round of selection was to be carried out. The choice of mating partners in the breeding population was made at random. Unbiased breeding values for the test progeny were predicted using a combined index of individual and family performance, weighted by their respective heritabilities (FALCONER and MACKEY, 1996; BAKER, 1986).

Selection on the selection criterion (eq. 3) was simulated for various penalty factors, c, covering the full range of values. Comparisons were made with a conventional practice to avoid relatedness when selecting for predicted breeding values with restrictions on parental contributions to the selected population, ranging from a limit (m) of 2 selections per parent (fully restricted) to no restrictions ($m = N$).

The simulation was continued through 5 discrete generations and repeated 100 times to produce reliable mean results. The simulation was stopped after 5 generations, as we do not believe that our simulation model is relevant for a much larger number of generations, where it does not address mechanisms that can reduce relatedness, such as mutation. Simulation results were interpreted by plotting gain achieved against the status effective number of the selected population, over the whole range of simulated conditions.

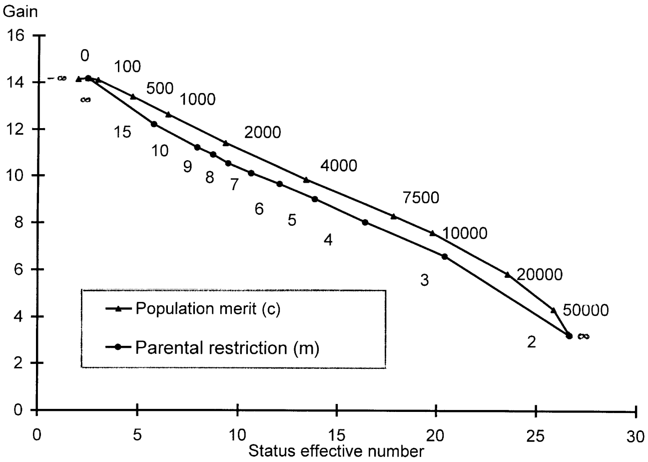

Figure 1. – Gain (%) and effective number (N_S) in a selected population of 40 trees, after one round of selection, when $h^2 = 0.05$. The line for population-merit selection connects results from simulations obtained with c-values chosen arbitrarily to cover the full range, while the parental restriction line connects results generated when the maximum contribution from a given parent was varied ($m = 2$, 3, 4, etc.). The m- and c-values used are shown beside the corresponding results. The limiting cases where the curves coincide correspond to family selection (leftmost), unrestricted combined-index selection (uppermost) and within-family selection (rightmost).

Results

The relationship of genetic gain and status effective number after the first round of selection is illustrated in *figure 1*. When no penalty for relatedness is applied ($c = 0$), population-merit selection becomes identical to conventional selection with no parental restrictions (unrestricted combined-index selection). When the penalty for relatedness is very high, population-merit selection is equivalent to selection with maximum restrictions on parental contributions (where an equal number of selections is made from each cross, i.e., within-family selection). Between these cases, the 2 methods produce markedly different results; population-merit selection *always* produced more gain than restrictions on parental contributions at a given N_S or, put another way, resulted in a higher N_S for any given gain. The relative superiority of population-merit selection was generally greatest for scenarios that maintained higher Ns, but the differences were considerable for all but the extreme cases of maximum and minimum restriction.

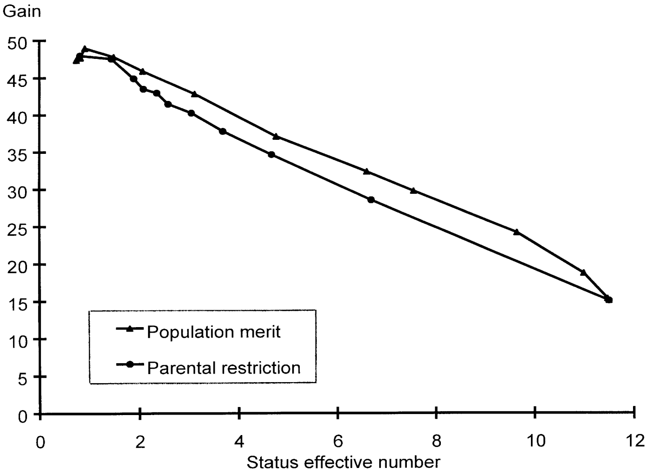

Figure 2. – Gain and effective number in a selected population of 40 trees, after following the simulations from *figure 1* for 5 generations of population-merit selection and selection with parental restriction.

The superiority of the population-merit selection was maintained after 5 generations, as shown in *figure 2*. It can be seen that recurrent use of population-merit selection produces even greater improvement over the parental restriction method after 5 generations. It may be worth noting that N_S becomes extremely low after 5 generations of combined-index selection unless measures are taken to restrict relatedness, like using population-merit selection with a reasonably high c value.

Discussion

Similar work

Our suggested breeding objective has similarities with that proposed by WRAY and GODDARD (1994), but instead of deducting a function of average inbreeding as such from the gain we deduct a function of average coancestry. BRISBANE and GIBSON (1995) demonstrated that an adjustment of estimated breeding values for average coancestry, in a way that is only slightly different from ours, results in an efficient compromise between gain and inbreeding. Their basic idea of what a selection algorithm should do can be said to be identical to ours, but we believe our algorithm is more able to meet its objective. Ballou and Lacy (1995) showed that a selection algorithm, which is similar to our formulation of population merit with a high c, was the most efficient in conserving genetic diversity. Lindgren and Wei (1994) subtracted the inverted effective number times a constant from average gain to get a value index similar to ours, but their effective number was defined in a much more narrow sense than status number.

A balanced situation

Figure 1 illustrates a situation where a given number of selections are made among unrelated, non-inbred full-sib families to replace the breeding population. Classical family, within-family and combined-index selection are special cases. In family selection, where $c \rightarrow -\infty$ (approximated here by $c = -10^6$), both gain and N_S suffer; thus, family selection is never likely to offer a good compromise between gain and effective number (this result is more clear from Lindgren and Wei, 1994, whose effective population size concept is equivalent to status number for their particular case, although this is dependent on their somewhat unrealistic assumptions, such as an infinite number of selections). Within-family selection, is equivalent to population-merit selection where $c \rightarrow +\infty$ (approximated by $c = 10^6$), as is combined-index selection (which here is a BLUP-solution) when $c = 0$ and thus only predicted breeding value is considered.

Different sexes

As presented, our algorithm does not consider sex differences. Forest tree breeders and other crop breeders can usually choose to use a genotype as a seed parent or pollen parent, and an algorithm involving gender would often be regarded as irrelevant or an unnecessary complication. From the standpoint of demonstrating principles, it is also better first to develop a general breeding model, and then to deal with gender groups as a special case.

In situations where *a priori* knowledge exists about fertility differences, such as when the breeding population consists of widely different numbers of males and females, our suggested definition of B_ω is not practical. Such differences can probably be handled by weighting. This weighting must be applied when calculating both average gain and average coancestry. If there are N members of the breeding population, composed of N_f females, and $N_m = N - N_f$ males, numbered with the females first followed by the males, the average coancestry of the selected population, Θ_ω, weighted for sex, is given as a function of individual coancestries (θ_{ij})

$$\Theta_\omega = \frac{N_m^2 \sum_{i=1}^{N_f}\sum_{j=1}^{N_f} \theta_{ij} + N_f^2 \sum_{i=N_f+1}^{N} \sum_{j=N_f+1}^{N} \theta_{ij} + 2N_m N_f \sum_{i=1}^{N_f} \sum_{j=N_f+1}^{N} \theta_{ij}}{4N_f^2 N_m^2} \tag{4}$$

When the sexes are more or less evenly distributed in the breeding population, this weighting is not important.

The diversity measure

The topic of conserving allelic diversity has tended to be addressed in terms of minimising the level of inbreeding that would occur within more or less random mating populations. Most recent investigators have expressed the relationship of gain (on the Y-axis) and average inbreeding on the X-axis (e.g., Quinton and Smith, 1995; Klieve *et al.*, 1994). We suggest that, for many situations, it is more appropriate to base the X-axis on average coancestry. The difference is not so great from a practical point of view for advanced programmes without sublining, as average coancestry becomes inbreeding in the offspring following random mating. For animal breeding, the breeding population is often a segment of the production population and, in that situation, inbreeding may be regarded as a serious consequence of selection, while in forest tree breeding the breeding population is usually not a part of the production population, and in such situations average inbreeding may deviate considerably from average coancestry; it can become very high if selfing is practised, and yet can return to zero after many generations of breeding if individuals from different lines are mated. Forest tree breeders, therefore, have much less reason to relate gain to actual inbreeding than do animal breeders.

N_S is equivalent to the concept of "founder genome equivalents", proposed by Ballou and Lacy, (1995). N_S (and average coancestry) is directly related to loss of gene diversity defined as the heterozygosity expected if the population were in Hardy-Weinberg equilibrium. Therefore, genetic management strategies that maximise N_S also maximise gene diversity Ballou and Lacy, 1995). Genetic diversity is closely related to the proportion of genes that are not identical by descent in the population, and it is likely that populations with similar N_S also have a similar long-term breeding potential and breeding limit, even if they are derived differently.

If the population merit is defined as a function of gain and average coancestry, it seems logical to plot these against each other. A clear demonstration that inbreeding is not an adequate measure is given by the situation illustrated in *figure 1*. Here, the situation in the breeding population after the first round of selection is depicted, when inbreeding will still be zero, as founders are neither related nor selfed. This consideration is of major practical importance for forest tree breeding programmes, most of which are currently managing the first few generations. Forest tree breeders usually still mate unrelated trees, and thus get no inbreeding. The current inbreeding tells little about the undesirable loss of diversity in many practically important situations, but more about historical mating systems that may have little impact on breeding potential.

Instead of using average coancestry explicitly on the X-axis, which conveys a meaningful message to few people, we suggest that it is heuristically preferable to use a coancestry-based effective number, N_S Lindgren *et al.*, 1996, see above). A further advantage is that greater diversity is a desirable thing, and this is more easily associated with a high N_S, rather than a low Θ or F.

It should be noted that the concepts of coancestry and inbreeding are relative to some reference population of individuals assumed to be unrelated and non-inbred, so that no negative values occur. It is also assumed that the studied populations have known coancestry structures that can be derived by pedigree analysis. The suggested algorithm can handle a population with an existing inbreeding and coancestry structure; so there is no need to start from the reference population, as was done in our example.

The average coancestry concept

Average kinship, as used by Ballou and Lacy (1995), is equivalent to our average coancestry. Averages can, however, be formulated differently. Their inclusion of self-coancestry is not evident, nor is it evident that pair-coancestry appears twice but self-coancestry only once. Averages can be taken over the n combinations with individuals in a population (referred to by Ballou and Lacy, 1995, as "mean kinship") or over the n^2 ways that individuals can be paired with each other or just over the pairs that actually mate (e.g. to minimise the "average coancestry" of mating couples). Many people regard it as contraintuitive that a population consisting of unrelated individuals has a positive average coancestry. Coancestry usually refers to pair-wise relations, but here it is applied as a population concept. Therefore, it may be a good idea to call our average coancestry concept "population coancestry" instead. This is the reason we use Θ as a symbol for average coancestry rather than some variant of θ.

Selection algorithms

Theoretically, B_ω could be maximised by an exhaustive search of all ω that can be formed from available candidates. That may be a valid option for some cases, and deserves consideration before trying other ways, but usually there are too many alternatives. We have considered the problem of selecting a set of 40 from 4000 candidates. There are $4000![40!/(4000-40)!]$ possible sets – a huge number! There is no known general procedure to find the exact maximum, nor to validate a suggested maximum, other than by an exhaustive consideration of all possible sets. There is a risk that our iterative procedure may identify a local, rather than a global maximum. However, there are evidently many ways to construct iterative search algorithms that are likely to find the maximum. There are tricks to reduce the problem size (e.g., only top-ranking individuals in a family need to be considered). The most important contribution of this paper is the formulation of the proposed objective that considers breeding value and relatedness jointly. Our effort to develop a selection algorithm may be regarded as an example of its application.

Practical implementation

The formulation of the underlying breeding objective in terms of population merit seems a sound way to consider both gain and relatedness. The advantages of the population-merit selection method seem to be considerable, and there are no evident drawbacks. Graphs, such as those used in *figure. 1* and *2*, that demonstrate the gain achievable over the full range of possible effective numbers, can be valuable as a basis for a breeder's decision.

If the factors related to c could be quantified, a relevant value for c could be established. However, in real-life situations, it is usually difficult to establish *a priori* a particular c, and this is one reason we have studied selection over the full range of c-values. However, diagrams similar to *figure 1* are a valuable tool for making the decisions of a desirable N_S, and for choosing the most appropriate value of c for a given situation. Similar decisions are often made in conventional selection, such as the establishment of limits on parental contributions, which effectively imply a corresponding value for c. In selection for population merit, the requirement to explicitly choose a value of c will probably encourage more thought and better documentation on how these decisions are made. The c will vary with the situation; it is likely that lower values will be more acceptable when selecting for a seed orchard than for long-term breeding.

The superiority of the suggested method over alternatives will vary with the specific conditions. Provided the algorithm succeeds in finding the maximum, there does not seem to be a way to get a better combination of additive genetic gain and status effective number. There are, however, some problems, e.g., predictions rather than known breeding values are used. In the later generations of our example, combined-index selection is not the best predictor of breeding value, but the comparison with the conventional method to control relatedness should still be fair. The genetic parameters change by selection, and in our current formulation these parameters are absorbed by c. This may cause problems if selection is repeated for several generations. One may ask whether generation-by-generation optimisation will assure optimisation after several generations.

The population-merit selection algorithm is neither difficult to automate nor to understand. While it requires substantial computing time to repeat it hundreds of times when applied in simulation models, less time is required to complete selection using data sets from real-life breeding programmes. However, the time required for computations increases exponentially with the size of the selected population, thus faster computers is not enough, but improvement in methodology would be desirable to handle larger populations. For many practical purposes, BLUP-values of predicted gain would probably be used. In forest tree breeding programs, there is often no need to trace back more than a generation, so that the calculation of average coancestry can be simplified.

Other uses

We have constructed an algorithm for a preset number of selections (N), but the algorithm can be extended to identify the N which maximises population merit. The current study focuses on selection, but the suggested breeding objective and population merit concepts can be useful for a wider range of purposes. Suitable mating designs can be defined or the whole breeding system can be analysed.

Acknowledgements

This research was supported by funding from the Jacob Wallenberg foundation and the Kempe foundation. The journal referees made very helpful comments.

Literature

Baker, R. J.: Selection indices in plant breeding. CRC Press, Boca Raton, FL (1986). — Ballou, J. D. and Lacy, R. C.: Identifying genetically important individuals for management of genetic variation in pedigreed populations. Chapter 5. In: Population management for survival and recovery. Edited by J. D. Ballou, M. Gilpin and T. J. Foose. Columbia Univ. Press, New York. pp. 76–111 (1995). — Brisbane, J. R. and Gibson, J. P.: Balancing selection response and rate of inbreeding by including genetic relationships in selection decisions. Theor. Appl. Genet. **91**, 421–431 (1995). — Caballero,A., Santiago, E. and Toro, M. A.: Systems of mating to reduce inbreeding in selected populations. Anim. Sci. **62**, 431–442 (1996). — Falconer, D. S. and Mackey, T. F. C.: Introduction to quantitative genetics. 4th ed.. Longman Group, London, UK. 463 pp. (1996). — Foulley, J.-L. and Chevalet, C.: Méthode de prise en compte de la consanguinité dans un modèle simple de simulation de performances. Ann. Génét. Sél. Anim. **13**, 189–196 (1981). — Klieve, H. M., Kinghorn, B. P. and Barwick, S. A.: The joint regulation of genetic gain and inbreeding under mate selection. J. Anim. Breed. Genet. **111**, 81–88 (1994). — Lindgren, D.: How should breeders respond to breeding values? In: Proc. IUFRO Conf. on Breeding Theory, Progeny Testing and Seed Orchards, 13 to 17 October 1986, Williamsburg, VA. pp. 361–371 (1986). — Lindgren, D., Gea, L. D. and Jefferson, P. A.: Loss of genetic diversity monitored by status number. Silvae Genet. **45**, 52–59 (1996). — Lindgren, D., Gea, L. D. and Jefferson, P. A.: Status number for measuring genetic diversity. Forest Genetics **4**(2), 69–76 (1997). — Lindgren, D., Libby, W. J. and Bondesson, F. L.: Deployment to plantations of numbers and proportions of clones with special emphasis on maximizing gain at a constant diversity. Theor. Appl. Genet. **77**, 825–831 (1989). — Lindgren, D. and Wei, R-P.: Gain versus effective number. Ed. Lee S. Proceedings — Nordic Group for Tree Breeding, Edinburgh, 6 to 10 October 1993. 164–177 (1994). — Lindgren, D., Wei, R.-P. and Bondesson, F. L.: Optimal selection from families. Heredity **70**, 619–621 (1993). — Meuwissen, T. H. E. and Luo, Z.: Computing inbreeding coefficients in large populations. Genet. Sel. Evol. **24**, 305–313 (1992). — Mullin, T. J. and Park, Y. S.: Stochastic simulation of population management strategies for tree breeding: a new decision support tool for personal computers. Silvae Genet. **44**, 132–141 (1995). — Quinton, M. and Smith, C.: Comparison of evaluation-selection systems for maximizing genetic response at the same level of inbreeding. J. Anim. Sci. **73**, 2208–2212 (1995). — Quinton, M., Smith, C. and Goddard, M. E.: Comparison of selection methods at the same level of inbreeding. J. Anim. Sci. **70,** 1060–1067 (1992). — Tier, B.: Computing inbreeding coefficients quickly. Genet. Sel. Evol. **22**, 419–430 (1990). — Toro, M. and Pérez-Enciso, M.: Optimization of selection response under restricted inbreeding. Genet. Sel. Evol. **22**, 93–107 (1990). — Wei, R.-P.: Predicting genetic diversity and optimizing gain for tree breeding programs. Ph. D. thesis, Swedish University of Agricultural Sciences, Department for Forest Genetics and Plant Physiology, Umeå (1995). — Wei, R.-P. and Lindgren, D.: Optimal family contributions and a linear approximation. Theor. Pop. Biol. **48**, 318–332 (1995). — Wray, N. R. and Goddard, M. E.: Increasing long-term selection response. Genet. Sel. Evol. **26**, 431–451 (1994).

Geographical Variation in Seed and Seedling Traits of Neem (*Azadirachta indica* A. Juss.) Among Ten Populations Studied in Growth Chamber

By S. K. Kundu and P. M. A. Tigerstedt(1)

(Received 17th February 1997)

Abstract

Nineteen biometric characters of seed and seedling of 10 populations of *Azadirachta indica* from Myanmar, Bangladesh, India, Pakistan and the Sudan were studied under growth chamber conditions. The results obtained by the use of principal component analysis (PCA) and cluster analysis of data revealed three distinct groups of populations. Provenances from the high rainfall areas were separated from lower rainfall areas in PCA. Correlation analyses between biometric traits and geoclimatic factors indicated the existence of ecoclinal variation in seed weight, shoot:root ratio, leaf number, and leaflet ratio. Leaflets of the northern provenances were narrower than those of the southern provenances, with a tendency for a separation between the northern (Pakistan, Myanmar and Bangladesh) and southern occurrences (Sudan and India). Provenances having lower shoot:root ratio and lesser number of leaves were separated with their corresponding mean annual rainfall. The results suggest that *A. indica* employs an adaptive strategy in response to water deficit at the initial phase of seedling growth and establishment.

Key words: Adaptation, *Azadirachta indica*, drought, ecoclinal variation, geographical variation, neem.

FDC: 165.52; 232.12; 164; 181.31; 181.525; 232.312.31; 232.318; 176.1 *Azadirachta indica*.

Introduction

Neem (*Azadirachta indica* A. Juss.) is a multipurpose tree species native to the Indian subcontinent and Myanmar. It has been transferred to many of the warmer parts of the world. Apart from its value as a timber tree, neem has antiseptic, medicinal, insecticidal properties and sociocultural values (Radwanski and Wickens, 1981; Koul *et al*., 1989; Maramorosch, 1991; Mordue and Blackwell, 1993; Ahmed, 1995; Biswas *et al*., 1995; Ketkar and Ketkar, 1995). The effective exploration, identification, documentation and use of genetic resources of the neem is a prerequisite for its efficient use.

The exact origin of *A. indica* is not known. Some authors suggest it may lie in Myanmar and/or in parts of southern India such as Karnataka (Troups, 1921; Vartak and Ghate, 1990). Nowadays, *A. indica* is widely distributed by introduction, mainly in the drier tropical and subtropical zone of Asia, Africa, the Americas, Australia and the South Pacific islands. The neem tree is well known for its drought resistance (Schmutterer, 1995). The wide ranges of habitats indicate good adaptation to drier conditions. Neem thrives well in low rainfall (130 mm/year) areas. In high rainfall (3,000 mm/year to 4,000 mm/year) areas, it has failed completely (Schmutterer, 1995). Neem performs well on a wide range of soil types. In its native environment, neem is found growing in mixed forests in association with *Acacia* and *Dalbergia spp.* (Benge, 1989).

Knowledge of the distribution of genetic variability, breeding system of trees and of the evolutionary forces that have shaped them, is a prerequisite for tree improvement (Muona, 1990). Such information is also needed for planning gene conservation and establishing breeding populations. Provenance testing is done for very practical reasons, to screen the naturally available genetic variation and to allow selection of the best available types for reforestation or further breeding (Wright, 1976). Provenance research also aims at defining the genetic and environmental components of phenotypic variability between trees from different geographic origins (Callaham, 1964). Several studies (Hamrick *et al*., 1979, 1981; and Hamrick and Godt, 1989) suggest that the extent of the geographic range correlates closely with the within population genetic variation in tropical tree species.

Variation in seed and seedling traits and its significance for seed source studies have been studied in a number of tree species. Various ecotypes of *A. indica* exhibit variation in several characters (Arora, 1993). Growth and architectural analysis of neem showed significant variation in shoot elongation, height growth, branching pattern and internode length (Bisht and Toky, 1993). Significant provenance variation in height growth and survival rates of neem was reported among 39 seed sources from India by Rajawat *et al.* (1994). Geographical variation in seed size, oil content and tree growth related characters has also been reported (Veerendra, 1995; Surendran *et al.,* 1993; Dwivedi, 1993). Ermel (1995) showed distinct ecotypic differences in azadirachtin and oil contents of seed kernels.

---

(1) Department of Plant Biology, University of Helsinki, P.O. Box 27, FIN-00014 Helsinki, Finland.
E-mail: Sunil.Kundu@Helsinki.fi
Fax: 358 9 708 5434