

A Decision Support System Inferring the Genotypes of Seed Trees From Open Pollinated Progenies

By A. C. PAPAGEORGIOU* and L. S. ILIADIS

Department of Forestry, Environmental Management and Natural Resources, Democritus University of Thrace,
P.O. Box 129, Pantazidou 193, GR-68200 Orestiada, Greece

(Received 1st June 2004)

Summary

The knowledge of the genetic background of seed lots used in forestry operations (i.e., afforestations, reforestations) is important for the evaluation of the future potential of forests to survive under changing environments. An approach to infer the genotype of seed trees from their open pollinated progenies is presented. Sampling difficulties of adult tree tissue make the genetic analysis of seed trees often impossible. The present decision support system (TGDSS – Tree Genotype Decision Support System) allows the reconstruction of the genotypes of the parent population (adult trees) through the analysis of only the genotypes of the progeny generation (seeds). The method consists of three steps, each one representing a specific logical test. The system was tested with data obtained from the genetic analysis of tree populations. The genotypes inferred through the TGDSS and the ones observed in the laboratory were identical. The factors defining the effectiveness of the system are discussed.

Key words: Decision Support System, genotypes, seed trees, progeny, inheritance, C++.

Introduction

A specific question that often arises is the knowledge of the genotype of the female parents of a given set of seeds from open pollination. This information is of crucial importance for the performance of genetic studies on the origin of seeds, the mating system of tree species, the spatial dynamics of forest stands and the mode of inheritance of specific genes (ELLSTRAND, 1984; BROWN *et al.*, 1985; CHAKRABORTY *et al.*, 1988; HATTEMER, 1991).

If the knowledge of the seed tree genotype is required, sampling of fresh tissue from that adult tree is necessary. However, transportation and storage of fresh tissue can be very difficult, making the analysis of the female parent in many cases impossible. This is often a major obstacle for the performance of genetic studies of tree populations, since most researchers can easily sample seeds belonging to the same female parent, but not tissue from the tree itself. As a consequence, the data structures of many population genetic studies are characterized by single-plant progenies (GILLET, 1997).

A method of reconstruction of the female parent genotype through the genotypes of its open pollinated progeny is presented here. Previous studies described methods for seed tree reconstruction using likelihood equations and assumptions about the mating system (BROWN and ALLARD, 1970; BROWN *et al.*, 1975; RITLAND, 1986). GILLET and GREGORIUS (2000) developed a system analytical approach for seed tree genotype reconstruction, by identifying heterozygous parents for the performance of inheritance analysis. The Tree Genotype Decision Support System (TGDSS) presented in this study aims at the reconstruction of the genotypes – heterozygous as well as

homozygous – of the parent population (adult trees) by the analysis of the genotypes of the progeny. This method is based on a simpler approach than the previous attempts and combines the use of various mathematical equations with the use of logical and heuristic methods.

A Decision Support System (DSS) couples the intellectual resources of individuals with the capabilities of computers to improve the quality of decisions (KEEN and MORTON, 1978). It is a computer-based system that consists of three interactive components, namely, a language system, a knowledge system and a problem-processing system (BONCZEK *et al.*, 1983). The two most important parts of the TGDSS are the knowledge base and the inference engine. The knowledge base consists of rules and data that are put in by the user and stored in text files for future use. The function of any knowledge representation scheme is to capture the essential features of a problem domain and to make that information accessible to a problem-solving procedure (LUGER and STUBBLEFIELD, 1998). The inference engine is the part of the system that manipulates the rules in order to generate information (DOUKIDIS and WHITLEY, 1988) and was designed to be Forward Chaining (data-driven). The inferencing starts from the known data and the system tests each rule. It fires every rule whose antecedent can be shown to be true. The system performs an iterative process, until no more rules fire.

The TGDSS is written in C++, a conventional object-oriented language, which is suitable for implementing algorithms and repetitive processes (PRATA, 1998). The inference engine of the TGDSS is easily implemented using a set of rules and the reasoning of the system is sequential and quite straightforward.

The use of computerized decision support systems is spreading in various fields and domains as computer applications are moving from transaction processing and monitoring activities to problem analysis and solution applications (TURBAN and ARONSON, 1998). The TGDSS is designed to perform speedy computations and to achieve the combination of the Mendelian laws and their mathematical expression with the human decision capability based on experience.

Methods

Description of the decision process

Consider a sample of seeds collected from each of a number of trees from the same stand. This sample is analysed genetically in the laboratory, by the means of gene markers, in order to describe the genotypes of the seeds (progenies) at a single gene. Assuming that tissue of the female parent (seed tree) of each of these open-pollinated single-tree progenies is not available for analysis, the only possibility for identifying the genotype of the female parent is to infer it through the genotypes of its progenies.

The method presented here comprises of three steps. The first applies a test of qualitative nature, which is based on the

* Corresponding author: Tel: +30 25520 41155, Fax: +30 25520 41192.
Email: apapage@fmenr.duth.gr

presence or absence of specific genotypes. The two other steps are quantitative tests, based on the comparison of specific genotype frequencies in the sample. The second step is applied for those seed trees that could not be identified through the first step and – respectively – the third step to the remaining seed trees of unknown genotype after the second step. Each step is performed after the previous step is completed for all seed tree genotypes that remain unidentified.

The **first step** (qualitative test) is based on the fact, that all progeny genotypes have at least one allele derived from their female parent. Thus, when the genotype of a progeny is homozygous for a specific allele, then the genotype of its female parent definitely contains the same allele. Following this rationale, when two different homozygous genotypes appear under the same progeny sample, then the female parent can be identified as heterozygous at that locus. The qualitative test of the first step controls the existence of two different homozygous genotypes among the progenies of each tree.

The genotypes of the progenies are presented in the following form: a capital letter symbolizes the gene and an index the allele. The array of the genotypes of all progenies of each single female parent is used as input-data for the qualitative test. The test is performed for the progenies of each female parent separately.

The aim of the first step is to control whether a progeny sample contains two homozygous genotypes. If this is the case, we can safely consider the female parent as heterozygous for the alleles, for which their progeny were found to be homozygous. The parent tree genotype is identified successfully and we do not need to perform any other test for the specific tree. If for example a female parent contains two genotypes X_aX_a and X_bX_b among its progenies, then the female tree can be identified as heterozygous and its genotype is X_aX_b , with an accuracy of 100%.

If only one homozygous progeny is detected among the progenies of a female tree, then this tree will be examined in the next step. In the rare and complicated case when no homozygous progeny can be found in the seed sample of a female tree, the current method we present here fails to identify the genotype of the specific female parent and this tree is skipped. Furthermore, if more than two homozygous progenies containing different alleles (X_a , X_b , X_c) are detected, the tree is omitted assuming an error in sampling or in the laboratory procedures, since these results make no biological meaning for diploid organisms, as most trees are.

STEP 1

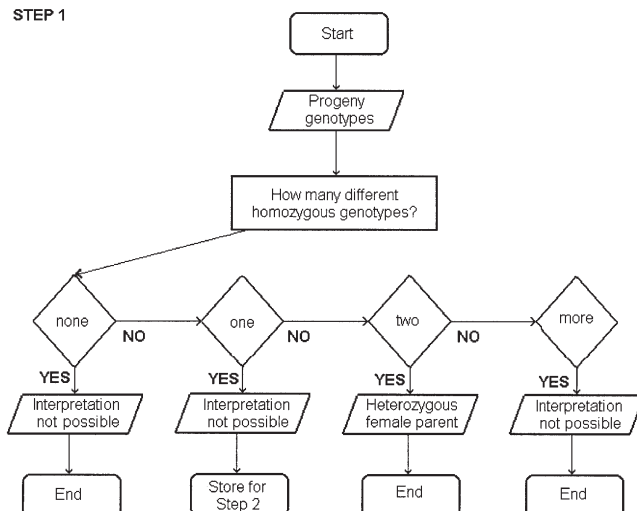


Figure 1. – Flow chart of Step 1.

The same procedure is repeated for all progeny samples. The first step can identify heterozygous trees only. When the first step is finished, the information for the genotypes of a number of heterozygous female trees is thus obtained (Figure 1).

The **second step** is applied to the progenies of the female parents, the genotypes of which remain unidentified after the first step. It examines the frequencies of the genotypes within the sample of progenies of a single tree and compares them with the expected ratios according to the second law of Mendel. If for example a female tree is heterozygous at a gene X (X_aX_b), then the sum of the homozygous progenies (X_aX_a and X_bX_b) of the specific tree must be equal to the frequency of the heterozygous progeny carrying the same alleles as the parent tree (X_aX_b) (GILLET and HATTEMER, 1989). Therefore, if the symbol $N(X_aX_b)$ to describe the amount of progenies carrying the genotype (X_aX_b), the following relation among the progenies of a heterozygous female tree with the genotype X_aX_b is expected:

$$N(X_aX_a) + N(X_bX_b) = N(X_aX_b) \quad (1)$$

An array of the genotypes of all progenies of each unidentified female parent is the input data for this step. The test is performed for the progenies of each female parent separately. Since each progeny sample that is examined here contains only one type of homozygous genotype (i.e., X_aX_a) for a specific allele (X_a), then we know that the genotype of the female tree contains this specific allele (X_a). For this reason, the expected ratio of genotype frequencies that we examine in this test is:

$$N(X_aX_a) = N(X_aX_b), \forall a \neq b \quad (2)$$

Equation (2) is used to assess the frequency of each different heterozygous genotype in the progeny sample that contains the allele of the homozygous progeny genotype. If all comparisons of frequencies show significant deviations from equation (2), the parent tree cannot be heterozygous and is characterized as homozygous for the same allele as its homozygous progeny. If only one heterozygous genotype appears in a frequency that is non-significantly different than the frequency of the homozygous genotype, we proceed to the next step. If more than one heterozygous progeny genotypes appear in non-significantly different frequencies than the homozygous genotype, the case

STEP 2

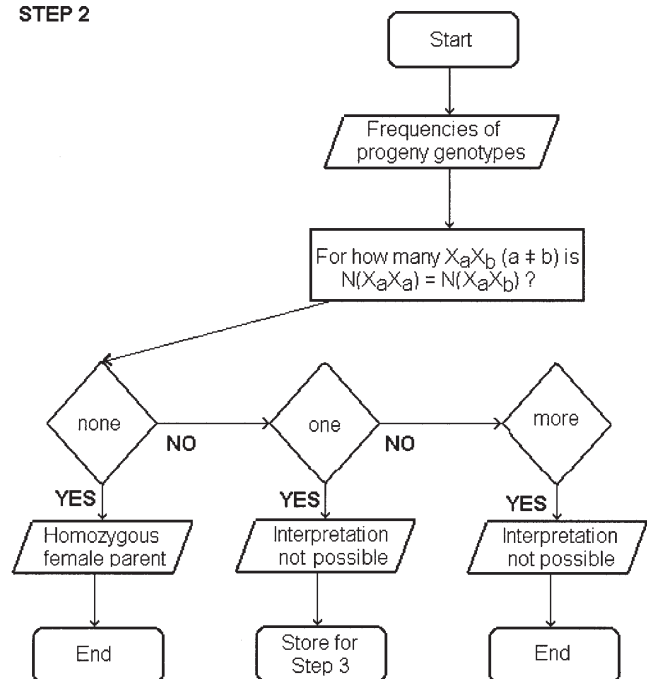


Figure 2. – Flow chart of Step 2.

is too rare and complicated to be solved through this method (Figure 2). Any significance test (i.e., chi-square) can be used here. The second step can identify homozygous trees only.

The qualitative test of the second step is based on the assumption that the second law of Mendel is valid. Random deviations from the Mendelian law decrease the accuracy of the test. This law can be verified through an inheritance analysis (HATTEMER, 1991). If the inheritance analysis has not been carried out for the specific species and the specific gene before, it can be performed through the comparison of the frequencies of the progeny genotypes of the heterozygous female parents that are already identified after the first step, following a procedure suggested by GILLET and HATTEMER (1989). Thus, the results of the first step can be used for the verification of the assumptions used in the second test.

The **third step** includes a quantitative test that is based on the information of the allele frequencies at a specific locus in the population. The results of the first two steps can give a first list of the genotypes inferred and thus provide information about the frequency of certain genotypes and alleles in the population. An assumption of this step is that the genetic structure of the adult tree population is reflected in the pollen cloud that has pollinated the female trees by random mating. Thus, by estimating the allele frequencies in the population, through the results of the previous steps, we can make an estimation about the frequency of these alleles in the pollen cloud that have fertilized the trees that remain unidentified.

The remaining trees examined here have one type of homozygous progeny (i.e., X_aX_a) and one type of heterozygous progeny (i.e., X_aX_b), both appearing under non-significantly different frequencies to each other ($N(X_aX_a) = N(X_aX_b)$). Any other case has been dealt within the previous steps. There are two possible explanations for these genotype frequencies among the progenies:

- either the allele X_b is frequent in the pollen cloud and the female parent is homozygous X_aX_a or,
- the allele X_b is rare in the pollen cloud and the female parent is heterozygous X_aX_b .

Hence, the knowledge of the rarity of a certain allele in the pollen cloud of a female parent can lead us to the identification of the female parent genotype. Following the assumption of random mating, the knowledge of the allele frequencies in the population can provide this information.

The question in this point is how to define a frequency threshold for allele rarity. It is subjective to decide if an allele is rare or not. Several researchers have proposed various rarity thresholds for allele frequencies, according to the aim of their studies (KIMURA, 1983; CONKLE, 1992; CROSSA *et al.*, 1992). For the purpose of the third step of our method, we consider an allele as "rare", if no progeny homozygous for this allele can be found, although the seed tree is heterozygous carrying the rare allele. This means that in case no homozygous genotype X_bX_b appears under the progeny of a heterozygous tree X_aX_b , then the allele X_b is considered as "rare" in the population.

Based on the binomial sampling theory, the threshold in that case should correspond to the sample size (the number of seeds analysed for each seed tree). The allele frequency threshold in the population is then defined as the relative allele frequency above of which the specific allele should be definitely found in a homozygous condition within a given sample of progenies. If we denote the allele threshold as $P(x)$, the size of progeny sample as N and the accuracy level of our decision as p , then the following relation derives from the binomial probability function:

$$P(x) = 2(1 - \sqrt[N]{1-p}) \quad (3)$$

In order to avoid complicated calculations we present an easy way for the estimation of a rarity threshold for a given sample and a specific accuracy level in Table 1. If for example we have a sample of 30 seeds per tree, then any allele in the population having a frequency lower than 0.19 is considered rare for an accuracy level of 0.95.

Table 1. – Rarity threshold for relative allele frequencies in the population of adult trees for two different accuracy levels (95%, 99%).

Sample size (seeds / tree)	p = 95%	p = 99%
10	0.51	0.73
15	0.36	0.52
20	0.27	0.41
25	0.22	0.33
30	0.19	0.28
40	0.14	0.21
50	0.11	0.17
60	0.09	0.14
70	0.08	0.12
80	0.07	0.11
90	0.06	0.09
100	0.05	0.09

The input data for the third test is an array of progeny genotypes for a specific female parent that remains unidentified after the two first steps. The information of the frequencies of the alleles in the population can be derived from the results of the previous steps and is considered as well. First, the rarity threshold $P(x)$ is defined for the given number of progenies. We denote the relative frequency of allele X_b in the population as $P(X_b)$, for which the heterozygous progeny genotype (X_aX_b)

STEP 3

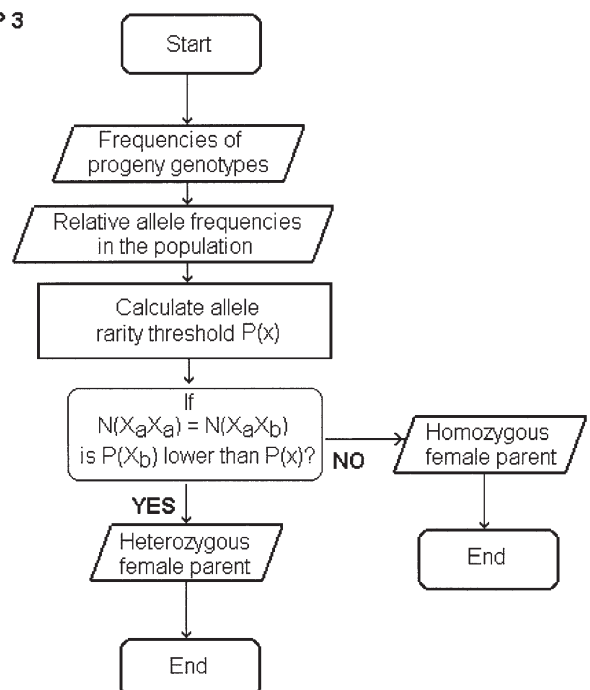


Figure 3. – Flow chart of Step 3.

appears in equal frequencies as the homozygous one (X_aX_a) in the progeny set ($N(X_aX_a) = N(X_aX_b)$). If $P(X_b)$ is lower than $P(x)$, then the allele X_b is considered rare and thus we conclude that the genotype of the female parent is heterozygous X_aX_b . If $P(X_b)$ is higher than $P(x)$, then the allele X_b is considered frequent and the genotype of the female parent is considered to be homozygous X_aX_a (Figure 3). This step leads to a final decision about the genotype of the female parent.

The test of the third step can have a reduced accuracy, when strong deviations from random mating occur and the allele frequencies observed in the population of adult trees do not correspond with the allele frequencies in the pollen clouds of these trees. However, the existence of such non-random mating events can be identified from the results of the previous steps. Pollen clouds can be safely estimated for the homozygous trees identified through the second step and compared to each other (PAPAGEORGIOU, 1995). If there exists large differentiation between the pollen clouds, then random mating should be doubted. Another possible test is the comparison between the genetic structures of all seed trees identified through the first two steps and their progeny. When non-random events are assumed, the user can either modify the rarity threshold and make the test less strict, or skip the results taken from the last step. Information from the first steps can be incorporated in the methodology of the last step to increase the accuracy.

Description of the Decision Support System

The TGDSS is a typical rule-based system, since the existing knowledge in the field of tree genotype determination, including equations, qualitative judgments and heuristics has been systematically coded in C++ language in order to form rules.

The database is used to store the data input by the user, every time the system is executed. The stored data can easily be retrieved and manipulated by the system, on user's request. Data are stored initially in complex data structures. After the consultation of the TGDSS finishes, these data are stored automatically in the form of a text file for future use. The system has been designed and implemented to run for any number of examined trees and for any number of seed genotypes. A dynamic array of structures was designed and implemented for the storage of the data. Each structure consists of two fields. The first field is the genotype found in each seed and the sec-

ond is the quantity of seeds found having the specific genotype.

The system has been designed and constructed to apply the forward chaining reasoning methodology. The TGDSS applies data to the rules that are necessary to give the variable "goal" the value "achieved". The TGDSS uses 12 rules in the main rule set. The overall structure and operation of the system can be seen in Figure 4. The TGDSS is a user-friendly computer programme that is available on the web site of the Democritus University of Thrace.

An Illustrative Example using TGDSS

An accuracy test for the TGDSS is performed, in order to control its ability to describe the genotypes of the female parents (seed trees) in a population. Real experimental data are used. FINESCHI *et al.* (1990) analysed open pollinated progenies from 20 seed trees of sweet chestnut (*Castanea sativa*), using biochemical gene markers. The outcome of this study was the information of the genotypes of 50 progenies per seed tree and the genotypes of all seed trees at several genes. In order to test the ability of the TGDSS to reconstruct the seed tree genotypes using the genotypes of their progenies, we initially ignore the existing information of the seed tree genotypes and we use it only in order to compare them with the outcome of the TGDSS.

An example of the results of the TGDSS is shown in Table 2. The genotypes presented here refer to a gene coding for the enzyme PGI-B. This gene contains three alleles; B_1 , B_2 and B_3 . The genotypes observed among the progenies are symbolized B_1B_1 , B_1B_2 , B_1B_3 and B_3B_3 . The absolute frequency of a genotype B_xB_y in the progeny samples is symbolized N_{xy} .

The progeny arrays of all seed trees are examined through the first step. The genotypes of seed trees 4, 22, 255, 500, 515 and 523 can be identified through the first step, two different homozygous seeds were observed within their progeny sample. The genotype of the seed parent was in all cases B_1B_3 . The rest of the seed trees have only one homozygous genotype among their progenies and are forwarded to the next step. The second step applies the chi-square test in order to control if the fre-

Table 2. – Comparison between the TGDSS output and the results from the genetic analysis in the laboratory, at the gene locus PGI-B. N_{xy} denotes the absolute frequency of the progenies carrying the genotype B_xB_y .

Tree	N_{11}	N_{12}	N_{13}	N_{33}	Step 1	Step 2	Step 3	TGDSS output	Laboratory results
4	14	0	24	12	B_1B_3			B_1B_3	B_1B_3
5	12	0	39	0	Step 2	B_1B_1		B_1B_1	B_1B_1
22	3	0	24	26	B_1B_3			B_1B_3	B_1B_3
84	0	0	5	45	Step 2	B_3B_3		B_3B_3	B_3B_3
93	0	0	18	31	Step 2	B_3B_3		B_3B_3	B_3B_3
95	0	0	11	38	Step 2	B_3B_3		B_3B_3	B_3B_3
100	11	2	39	0	Step 2	B_1B_1		B_1B_1	B_1B_1
172	0	0	23	26	Step 2	Step 3	B_3B_3	B_3B_3	B_3B_3
174	0	0	24	21	Step 2	Step 3	B_3B_3	B_3B_3	B_3B_3
245	0	0	20	29	Step 2	Step 3	B_3B_3	B_3B_3	B_3B_3
255	9	0	27	14	B_1B_3			B_1B_3	B_1B_3
320	0	0	27	16	Step 2	B_3B_3		B_3B_3	B_3B_3
450	13	0	31	0	Step 2	B_1B_1		B_1B_1	B_1B_1
471	0	0	16	30	Step 2	B_3B_3		B_3B_3	B_3B_3
472	0	0	8	33	Step 2	B_3B_3		B_3B_3	B_3B_3
488	0	0	7	45	Step 2	B_3B_3		B_3B_3	B_3B_3
500	7	0	27	16	B_1B_3			B_1B_3	B_1B_3
511	0	0	22	29	Step 2	Step 3	B_3B_3	B_3B_3	B_3B_3
515	6	0	21	19	B_1B_3			B_1B_3	B_1B_3
523	10	0	24	17	B_1B_3			B_1B_3	B_1B_3

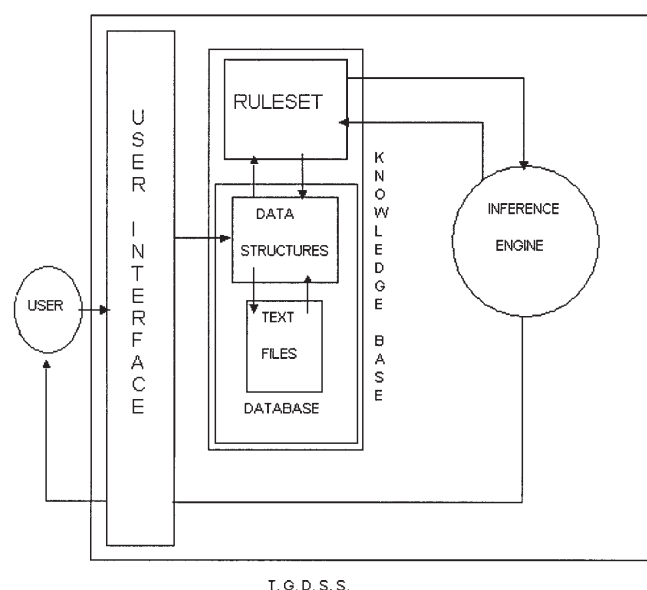


Figure 4. – The overall structure and the operation of the TGDSS.

quency of the homozygous progeny in each progeny sample is equal to the frequency of the heterozygous progenies. Nine seed trees show significant differences between the frequencies of their heterozygous and homozygous progenies and are identified as homozygous. Tree 100 has the genotype B_1B_1 , since N_{11} is significantly higher than N_{12} and significantly lower than N_{13} . Trees 5 and 450 have the genotype B_1B_1 as well, because N_{11} is significantly lower than N_{13} . Trees 84, 93, 95, 320, 472 and 488 can be identified as B_3B_3 , since N_{33} is significantly different than N_{13} . Further, no significant genotype frequencies between B_1B_1 and B_1B_3 exist for trees 172, 174, 245 and 511. The progeny samples of these seed trees are forwarded to the next step.

In order to perform the third step, we need to estimate a rarity threshold for allele frequencies in the population. Since the number of seeds analysed for each seed tree is 50, the rarity threshold $P(x)$ is 0.11 for an accuracy of 0.95. This means that an allele having a frequency in the population lower than 0.11 is considered rare. Next, we need to know the frequency of alleles B_1 and B_3 in the population. Considering the trees identified during the two previous steps, we see that the relative frequency of B_1 is $P(B_1) = 0.375$ and the one of B_3 is $P(B_3) = 0.625$. We realize that both alleles are above the rarity threshold and can be considered as frequent. For this reason, the genotypes of trees 172, 174, 245 and 511 are B_3B_3 .

As shown in Table 2, the genotypes derived as an output of the TGDSS and the real ones obtained through the genetic analysis in the laboratory match without any deviation, indicating a success level of 100%. The TGDSS was tested thoroughly using the data for all genes. In all of the cases the system performed well and gave the expected results.

Discussion and Conclusions

The method and the decision support system presented in this paper can be useful for genetic studies, when parent tree tissue is missing. The success of the results depends mainly on the amount of seeds analysed per tree. Large samples sizes increase the validity of the tests and the probability to identify the female parent genotype in the first two steps already, where no assumptions about the mating system are made. Therefore, the sample of seeds examined should be large enough, in order to allow both alleles of a heterozygous seed parent to appear among the progeny in homozygous condition. This depends on the frequency of the parental alleles in the pollen cloud and the sample size. Furthermore, the consideration of the sample size in the estimation of the rarity threshold for allele frequencies increases the accuracy of the method.

The assumption of the Mendelian laws needed for the performance of the second step represents a usual case in plants. Any deviations from this assumption can be detected using the results of the first step. The assumption used for the third step is less realistic, but it can be still verified through the results of the previous tests. Furthermore, it is an usual assumption made in genetic studies of this kind. The gradual use of data from the first steps into the next ones (as input data or as a control for the assumptions) is a major advantage of the TGDSS and increases its effectiveness.

The TGDSS describes a realistic solution for researchers who work with forest tree breeding and population genetic studies, in particular with aspects of the reproduction system and gene flow. It is simple, based on the classical Mendelian laws and uses assumptions that can be tested. Initial forms of this method have been already applied in studies of *Nothofagus nervosa* (GODOY, 1994), *Cupressus sempervirens* (PAPAGEORGIOU, 1995), *Prosopis flexuosa* and *Prosopis chilensis* (VERGA, 1995).

The TGDSS makes the use of this methodology easier and more effective.

Acknowledgements

The authors wish to thank ANIBAL RAMON VERGA for his stimulation and contribution to the initial discussions leading to this method and S. AVRAMIDIS for comments on the manuscript. Kind assistance was provided by the members of the Institute of Forest Genetics and Forest Tree Breeding, of the University Göttingen in Germany. Valuable advice on the issue of allele frequencies and rarity has been provided by OMAR CORNEJO, HAROLD W. KERSTER, SERGIOS-ORESTIS KOLOKOTRONIS, DMITRI PETROV, PAUL SCHLIEKELMAN, PAUL SUNNUCKS and BRUCE WEIR.

References

- BONCZEK, R. H., HOLSAPPLE, C. W., WHINSTON, A. B. (1983): The evolving role of models in decision support systems. In: HOUSE, W. C. (Ed.). *Decision Support Systems*. Petrocelli, New York, pp. 343–370.
- BROWN, A. H. D., ALLARD, R. W. (1970): Estimation of the mating system in open-pollinated maize populations using isozyme polymorphisms. *Genetics* **66**, 133–145.
- BROWN, A. H. D., MATHESON, A. C., ELDRIDGE, K. G. (1975): Estimation of the mating system of *Eucalyptus obliqua* L'Herit. by using allozyme polymorphisms. *Australian Journal of Botany* **23**, 931–949.
- BROWN, A. H. D., BARRETT, S. C. H., MORAN, G. F. (1985): Mating system estimation in forest trees: Models, Methods and meanings. In: GREGORIUS, H.-R. (Ed.). *Population Genetics in Forestry. Lecture Notes in Biomathematics*, 60. Springer Verlag, Heidelberg, New York, Tokyo, pp. 32–49.
- CHAKRABORTY, R., MEAGHER, T., SMOUSE, P. E. (1988): Parentage analysis with genetic markers in natural populations. I. The expected proportion of offspring with unambiguous paternity. *Genetics* **118**, 527–536.
- CONCLE, M. T. (1992): Genetic diversity – seeing the forest through the trees. *New Forests* **6**, 5–22.
- CROSSA, J., JEWELL, D. C., DEUTSCH, J. A., TABA, S. (1992): Gene action and the bottleneck effect in relation to sample size for maintenance of cross-pollinated populations. *Field Crops Research* **29**, 225–239.
- DOUKIDIS, G., WHITLEY, E. (1988): *Developing Expert Systems*. Chartwell-Bratt Publishing Ltd., Sussex, U.K.
- ELLSTRAND, N. C. (1984): Multiple paternity within the fruits of the wild radish, *Raphanus sativus*. *American Naturalist* **123**, 819–828.
- FINESCHI, S., GILLET, E., MALVOLTI, M. E. (1990): Genetics of sweet chestnut (*Castanea sativa* Mill.) III. Genetic analysis of zymogrammes of single tree offspring. *Silvae Genetica* **39**, 188–194.
- GILLET, E. (1997): Maximum likelihood estimators of the gametic contributions to single – plant progenies. *Biometrics* **53**, 504–523.
- GILLET, E., HATTEMER, H. H. (1989): Genetic analysis of isoenzyme phenotypes using single tree progenies. *Heredity* **63**, 135–141.
- GILLET, E., GREGORIUS, H.-R. (2000): Qualified testing of single-locus codominant inheritance using single tree progenies. *Biometrics* **56**, 801–807.
- GODOY, M. M. (1994): Determinación de isoenimas y análisis genéticos en *Nothofagus nervosa* (Phil) Dim et Milrauli. Msc thesis, Faculty of Forestry, Georg-August Universität Göttingen, Germany.
- HATTEMER, H. H. (1991): Genetic analysis and population genetics. In: FINESCHI, S., MALVOLTI, M. E., CANNATA, F., HATTEMER, H. H. (Eds.). *Biochemical Markers in the Population Genetics of Forest Trees*. SPB Academic Publishing, The Hague, pp. 5–22.
- KEEN, P. G. W., MORTON, M. S. S. (1978): *Decision Support Systems: An Organisational Perspective*. Addison Wesley, Reading, UK.
- KIMURA, M. (1983): Rare variant alleles in the light of the neutral theory. *Molecular Biology and Evolution* **1**, 84–93.

- LUGER, G. F., STUBBLEFIELD, W. A. (1998): Artificial Intelligence. Addison Wesley, USA.
- PAPAGEORGIOU, A. C. (1995): Genetische Untersuchungen zur Züchtung und Generhaltung der Mittelmeerzypresse (*Cupressus sempervirens* L.) in Griechenland. Göttingen Research Notes in Forest Genetics 18, Georg-August Universität Göttingen, Germany.
- PRATA, S. (1998): C++ Primer Plus, 3rd edition. Sams Publishing, Indiana, U.S.A.
- RITLAND, K. (1986): Joint maximum likelihood estimation of genetic and mating structure using open-pollinated progenies. *Biometrics* **42**, 25–43.
- TURBAN, E., ARONSON J. (1998): Decision Support Systems and Intelligent Systems. Prentice Hall, New Jersey, USA.
- VERGA, A. R. (1995): Genetische Untersuchungen an *Prosopis chilensis* und *P. flexuosa* im trockenen Chaco Argentiniens. Göttingen Research Notes in Forest Genetics 19, Georg-August Universität Göttingen, Germany.

Seed Source Variation in Cone, Seed and Seedling Characteristic Across the Natural Distribution of Himalayan Low Level Pine *Pinus roxburghii* Sarg.

By S. MUKHERJEE ROY*, R. C. THAPLIYAL and S. S. PHARTYAL

(Received 5th July 2004)

Abstract

The wide range of climatic condition in the natural distribution of chir pine is expected to result in high genetic variation within different populations of the species. The present study on the provenance variation of chir pine aims to determine the nature and extent of variation in wide range of populations with respect to 23 morphological traits of cone, seed and seedling, (at nursery stage) of 63 provenances of *Pinus roxburghii*. Seed sources exhibited a wide range of variability in terms of mean values for various traits, standard deviation, variance, coefficient of variability, broad sense heritability, genetic advance and genetic gain and offer ample scope for undertaking screening for the desired traits. Genotypic variance (V_g) and genotypic coefficient of variability (GCV) for most of the parameters were found to be higher than corresponding environmental variance (V_e) and environmental coefficient of variability (ECV), except cone fresh weight, survival%, days taken for bud break and collar diameter where V_e and ECV dominated the V_g and GCV, thus indicating influence of environment on the expression of these traits. Moderate to high

percentage of heritability coupled with same intensity of gain, was observed for many of the traits studied under laboratory conditions and at nursery stage e.g. for germination percentage, MGT, GV. This signifies that these traits are under strong genetic control and good amount of heritable additive genetic component can be exploited for further selection and improvement of this species.

Key words: *Pinus roxburghii* Sarg., seed source variation, seed and seedling characters, bud break, genotypic variance (V_g), genotypic coefficient of variability (GCV), environmental variance (V_e) and environmental coefficient of variability (ECV), heritability, genetic gain.

Introduction

Chir (*Pinus roxburghii*) is one of the five indigenous species of pine widely distributed in the outer ranges and principal valleys of Himalaya between 450–2300 m elevation, and from Afghanistan in the North-West to Arunachal Pradesh in the North-Eastern part of India between 26°N–36°N latitude and 71°E–93°E longitude. The species inhabits mostly the subtropical and warm temperature monsoon belt but occasionally can be found in areas where winter snow is a regular feature or the tropical climates where summer temperatures sometimes reach as high as 40°C. This diverse range of climatic condition

* Forest Tree Seed Laboratory, Forest Research Institute, P.O.–New Forest, DehraDun, 248006 Uttaranchal, INDIA.
E-mail: ndroy@indiatimes.com

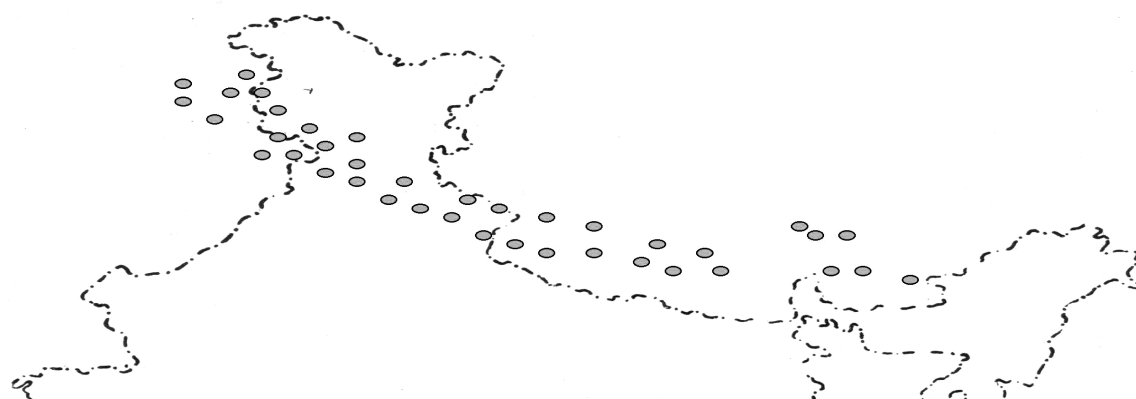


Figure 1. – Geographic distribution of *Pinus roxburghii* in Indian subcontinent.