Theor Appl Genet (1991) 82:313–320

# Effective number of pollen parents in clonal seed orchards

**J. H. Roberds**[1], **S. T. Friedman**[2] **and Y. A. El-Kassaby**[3]

[1] U.S. Department of Agriculture, Forest Service, Genetics Department, Box 7614, North Carolina State University, Raleigh, NC 27695-7614, USA
[2] U.S. Department of Agriculture, Forest Service, National Forest GEL, 2375 Fruitridge Road, Camino, CA 95709, USA
[3] Canadian Pacific Forest Limited, Tahsis Pacific Region, 8067 East Saanich Rd., R.R. 1, Saanichton, BC VOS 1MO, Canada

Received December 13, 1990; Accepted December 18, 1990
Communicated by P.M.A. Tigerstedt

**Summary.** A method for quantifying mating behavior in clonal seed orchards of forest tree species is presented. It involves the estimation of effective numbers of pollen parents from seed samples collected from individual ramets in such orchards. These effective numbers are variance effective numbers for populations of male gametes that are successful in uniting with ovules to produce viable seed. Three such effective numbers are defined for clonal seed orchards: $N_p(a)$ for male gamete populations for ramets within clones, $N_p(b)$ for male gamete populations for clones, and $N_p(c)$ for male gamete populations for entire orchards. Estimators for these effective numbers and for standardized variances of allele frequencies in the male gametic populations are presented. Expressions are also given for the confidence intervals for each of the three effective numbers. Estimates of these parameters and the corresponding confidence intervals for two seed orchards are presented and interpreted.

**Key words:** Variance effective number – Temporal method – Gene frequency variance – Genetic drift – Male gamete populations

## Introduction

Seed orchards are the link between breeding programs and reforestation efforts. Orchard populations most effectively achieve their purpose when they consistently produce adapted, genetically improved seed that meet reforestation needs. The value of seed crops produced by orchard populations is compromised, however, when mating patterns deviate from the panmictic optimum. Genetic gains potentially available through breeding and testing are not realized in such cases. It is now apparent that mating behavior in wind-pollinated seed orchards frequently deviates from panmixia. Several factors are known to contribute to such departures from random mating. Pollen contributions from outside sources can cause such departures and were found to be substantial in two loblolly pine (*Pinus taeda* L.) seed orchards (Friedman and Adams 1985). Selfing can be a factor (Shaw and Allard 1982; Ritland and El-Kassaby 1985) and phenological variation in reproductive traits may also cause uneven contribution of gametes (Griffin 1984; El-Kassaby et al. 1984). Similarly, differences in reproductive output (Griffin 1982; Schmidtling 1983; El-Kassaby et al. 1989) can influence the clonal distribution of gametes in orchard seed crops. Deficient pollen production in young orchards (Daniels 1978) and for specific clones in mature ones (Askew 1988; El-Kassaby and Askew 1990) has been shown to exist and to result in differential gametic output of clones. Male-female complementarities such as have been found in Douglas fir [*Pseudotsuga menziesii* (Mirb) Franco] (Apsit et al. 1989) also may result in disproportionate gametic contributions by particular pollen parents in matings involving some females.

In this paper we use an ideal clonal seed orchard model as the basis for quantifying mating processes in clonal seed orchards. We introduce the concept of effective number of pollen parents ($N_p$) as a measure of mating behavior in this ideal orchard. Our ideal clonal seed orchard is similar to the ideal population described by Wright (1931) for purposes of defining effective population number. Wright defined effective population number as the size of the ideal population that has the same increase in homozygosity and random allele drift as an observed population. In this context the ideal population is considered to be the collection of N diploid individuals that are produced in each generation by random pairing of 2 N gametes randomly sampled from the previous generation.

Our interest is focused on disproportionate contributions by clones to the pool of male gametes that are successful in uniting with ova to produce seed. Such disproportionate contributions may cause frequencies of neutral genes in the gamete pool to differ from their frequencies in the ramet population of the clonal orchard. These differences are equivalent to those produced by random sampling of genes from the ramets in the orchard and therefore are consistent with deviations caused by genetic drift. Variance effective numbers are appropriate for evaluating the amount of genetic drift or variance in allele frequency among samples (Crow and Denniston 1988). Effective number of pollen parents is therefore derived as a variance effective number. The variance effective number and methods to estimate it are described in Crow and Denniston (1988) and Kimura and Crow (1963).

## Effective number of pollen parents

We define the *effective number of pollen parents* ($N_p$) as the number of parents in samples of male gametes from an ideal seed orchard that has the same variance in allele frequency that occurs among populations of male gametes which have united with ovules to form embryos in seed. Such populations of male gametes can be associated with ramets of a clone, clones, or entire orchards. The effective number of pollen parents may be determined by equating the variance of allele frequencies among populations of male gametes to the sampling variance for frequencies given by the binomial distribution, and then solving for the number of individuals that yield the appropriate sample of genes. This binomial distribution has as a parameter the allele frequency for the population of ramets in the seed orchard. Such an effective number will fall in the interval, $1/2 \leq N_p < \infty$, and is independent of the number of clones in the orchard. A large value of $N_p$ indicates that most of the clones are equally contributing genes as male parents to the progeny population. If $N_p$ is small, either a few clones are contributing disproportionately large numbers of male gametes to the pool that is effective in producing seed or there is a large influx of male gametes from outside sources into the orchard.

The effective number of pollen parents is determined from the sampling variance for allele frequencies of the ramet population in seed orchards, as determined from allele frequencies in male gametes that were successful in forming embryos in the progeny generation. Thus, it is based on differences between the allele frequencies in two generations and is a reflection of temporal variation in allele frequencies. Use of temporal variation in allele frequencies as a basis for determining variance effective size was introduced by Krimbas and Tsakas (1971). This approach is based on the drift variance in allele frequency between generations, which is given by $p(1-p)/2\,N_e$ where $p$ is the allele frequency in the initial generation and $N_e$ is the variance effective size. An estimate of the standardized variance for allele frequency change, $\hat{f}=(p_1-p)^2/p(1-p)$, where $p_1$ is the observed allele frequency in the subsequent generation, is obtained and $N_e$ is then estimated by equating the expected and observed values of $f$ and solving for $N_e$. This method was further developed by Lewontin and Krakauer (1972), who investigated the distribution of $f$ and described a method of computing the approximate variance for estimates of $N_e$. Subsequently, Nei and Tajima (1981) identified two sampling methods to estimate $f$, suggested two additional estimators for $f$, and described a method for computing confidence intervals for $N_e$. Pollack (1983) introduced a different estimator for $f$ and derived its sampling distribution. A revised model for gene sampling was described by Waples (1989), who also investigated two gene sampling procedures for the estimation of $f$ and $N_e$. Waples also introduced new methods of estimating confidence intervals for both $f$ and $N_e$. In this report we introduce estimators for $f$ and $N_p$ appropriate for samples of genes taken from male gamete pools in clonal seed orchards, and we give confidence intervals for $N_p$.

## Seed orchard model

We consider a seed orchard for a monoecious, diploid species that is made up of multiple clones, each of which is represented by several ramets. It is assumed that each clone contributes male gametes to the seed orchard's pollen pool. As a basis for defining effective numbers of pollen parents ($N_p$), we hypothesize and ideal seed orchard in which each ramet of each clone makes an equal contribution to the male gamete population that fertilizes the ovules produced by clones within the orchard. We further assume that the pollen pool for the ideal orchard contains only gametes contributed by ramets within the orchard. Allele frequencies of genes in the ideal orchard's pollen pool are then exactly equal to the allele frequencies in the ramet population that makes up the ideal orchard. This pollen pool serves as a reference population from which binomial samples of genes in male gametes are drawn. Such sampling generates the allele frequency drift that forms the basis for defining three effective numbers of pollen parents $N_p(a)$, $N_p(b)$, and $N_p(c)$.

These three effective numbers, considered as a triple, provide a description of mating patterns in clonal seed orchards. $N_p(a)$ is the effective number of pollen parents defined on the basis of binomial sampling at the level of ramets within clones; $N_p(b)$ is the effective number defined for binomial sampling at the clonal level; and $N_p(c)$ is the effective number of pollen parents defined at the orchard level. Both $N_p(a)$ and $N_p(b)$ may vary among clones and can be determined for individual clones as well as for an entire orchard.

## Sampling method

Male gametes that are successful in fertilizing ovules of each ramet in the orchard may be considered to be independent binomial samples withdrawn from the orchard's pollen pool, which has allele frequency $p_{oij}$ for the $j^{\text{th}}$ allele of the $i^{\text{th}}$ locus. As a result of the sampling, allele frequencies in the male gametes for the $k^{\text{th}}$ ramet of the $l^{\text{th}}$ clone, $p_{ijkl}$ can differ from $p_{oij}$, particularly if the gamete sample size is small, i.e., if $N_p(a)$ is not large. From the natural sample of male gametes for each ramet, a subsample of size $m$ is obtained from seed for estimating allele frequencies. Allele frequencies for the subsample, $q_{ijkl}$, can differ substantially from $p_{ijkl}$ because of binomial sampling variation, especially if $m$ is small.

The estimators we use for the standardized variances, ($f$), and the effective numbers of pollen parents, ($N_p$), are appropriate only if sample sizes at the ramet and clone levels are equal. The confidence intervals presented are derived under the same sampling restriction. A sample of $r$ ramets is considered for each of the $n$ clones investigated. We assume a biallelic genetic model for $x$ independently segregating loci. When multiple allelic loci are used, estimates may be obtained for each allele. In these cases, the alternate form for each allele consists of the entire collection of the remaining alleles. The allele frequency for such an alternate form is one minus the frequency of the allele under consideration.

To estimate $N_p(a)$ we require the sampling variance of $q_{ijkl}$, assuming $p_{oij}$ is known. To obtain an expression for the sampling variance of $q_{ijkl}$, Var $(q_{ijkl})$, we first consider the variance of $p_{ijkl}$ given $p_{oij}$. Since $p_{ijkl}$ is a random variable and is subject to allele frequency drift because of binomial sampling, this conditional variance is

$$\mathrm{Var}(p_{ijkl}|p_{oij})=p_{oij}(1-p_{oij})/2N_p(a). \tag{1}$$

Likewise, since $q_{ijkl}$ is the allele frequency for gamete samples of size $m$ produced by binomial sampling,

$$\mathrm{Var}(q_{ijkl}|p_{ijkl})=\{p_{ijkl}(1-p_{ijkl})\}/m. \tag{2}$$

Incorporation of the expression for conditional variance (Rao 1973, p. 97) gives

$$\mathrm{Var}(q_{ijkl})=\mathrm{Var}(q_{ijkl}|p_{oij})=\mathrm{E}[\mathrm{Var}(q_{ijkl}|p_{ijkl})]+\mathrm{Var}[\mathrm{E}(q_{ijkl}|p_{ijkl})], \tag{3}$$

where E indicates the operator for expected value.

Since $\mathrm{E}(q_{ijkl}|p_{ijkl})=p_{ijkl}$, substitution of the right members of the identities in Eqs. 1 and 2 into Eq. 3 yields

$$\mathrm{Var}(q_{ijkl})=\mathrm{E}[p_{ijkl}(1-p_{ijkl})]/m+[p_{oij}(1-p_{oij})]/2N_p(a). \tag{4}$$

Substitution of $p_{oij}$ for $\mathrm{E}(p_{ijkl})$ leads to

$$\mathrm{Var}(q_{ijkl})=p_{oij}(1-p_{oij})\,[m+2N_p(a)-1]/2N_p(a)\,m. \tag{5}$$

$\mathrm{Var}(\bar{q}_{ijt})$ is required for estimation of $N_p(b)$ and $\mathrm{Var}(\bar{q}_{ij})$ is needed for estimation of $N_p(c)$. Here $\bar{q}_{ijt}$ is the allele frequency of the male gamete sample for clone $t$ and $\bar{q}_{ij}$ is the allele frequency in the male gamete pool for all clones. These frequencies can be obtained from

$$\bar{q}_{ijt}=\sum_{k=1}^{r} q_{ijkt}/r \quad \text{and} \quad \bar{q}_{ij}=\sum_{l=1}^{n}\sum_{k=1}^{r} q_{ijkl}/r\,n,$$

where $r$ indicates the number of ramets per clone and $n$ the number of clones. Expressions for the variances $\mathrm{Var}(\bar{q}_{ijt})$ and $\mathrm{Var}(\bar{q}_{ij})$ can be derived in the same way as $\mathrm{Var}(q_{ijkl})$ was derived, yielding

$$\mathrm{Var}(\bar{q}_{ijt})=p_{oij}(1-p_{oij})\,\{[r\,m+2N_{pt}(b)-1]/[2N_p(b)\,m\,r]\} \tag{6}$$

and

$$\mathrm{Var}(\bar{q}_{ij})=p_{oij}(1-p_{oij})\,\{[r\,m\,n+2N_p(c)-1]/[2N_p(c)\,r\,m\,n]\}. \tag{7}$$

## Estimation of the $f$ and $N_p$ parameters

For estimation of the $f$ parameters, we employ estimators of the form suggested by Krimbas and Tsakas (1971) and used by Lewontin and Krakauer (1973) for their temporal method. Since allele frequencies in the ramet population of the orchard are assumed to be known, we replace estimates of the base population allele frequency in Krimbas and Tsaka's (1971) estimator by the known orchard allele frequencies $p_{oij}$.

We begin by giving the estimator for the standardized variance for gamete pools associated with ramets within single clones,

$$\hat{f}_t(a)=\sum_{i=1}^{x}\sum_{k=1}^{r}(q_{ijkt}-p_{oij})^2/x\,r\,p_{oij}(1-p_{oij}), \tag{8}$$

where $t$ indicates the $t^{\text{th}}$ clone, $r$ indicates the number of ramets per clone, and $x$ indicates the number of independently segregating loci. For the multilocus estimators discussed here, only the frequency for the allele with the largest frequency, designated the $j^{\text{th}}$ allele, is used at each locus.

The expected value for this estimator can be expressed as

$$\begin{aligned} E[\hat{f}_t(a)] &= \sum_{i=1}^{x}\sum_{k=1}^{r} E[(q_{ijkt}-p_{oij})^2/x\,r\,p_{oij}(1-p_{oij})] \\ &= \sum_{i=1}^{x}\sum_{k=1}^{r} \mathrm{Var}(q_{ijkt})/x\,r\,p_{oij}(1-p_{oij}). \end{aligned} \tag{9}$$

Since the gamete pool for each ramet within a clone is considered to be an independent sample from the orchard's male gamete pool, each ramet has an identical $\mathrm{Var}(q_{ijkt})$, and substitution of Eq. 5 into Eq. 9 yields

$$E[\hat{f}_t(a)]=\sum_{i=1}^{x}[(m+2N_{pt}(a)-1)/2N_{pt}(a)\,m\,x].$$

Here $N_{pt}(a)$ indicates the effective number of pollen parents for ramets of clone $t$. The alleles at each locus are assumed to be selectively neutral and to segregate independently of the alleles at other loci; hence, they share a common $N_{pt}(a)$ and the expectation is reduced to

$$E[\hat{f}_t(a)]=(m+2N_{pt}(a)-1)/2N_{pt}(a)\,m\,. \tag{10}$$

An estimator for $N_{pt}(a)$ for the $t^{\text{th}}$ clone can be obtained by equating $\hat{f}_t(a)$ to $E[\hat{f}_t(a)]$ and solving for $N_{pt}(a)$. This estimator can be expressed as

$$\hat{N}_{pt}(a)=(m-1)/2(\hat{f}_t(a)\,m-1) \tag{11}$$

For large $m$, the estimator

$$\hat{N}_{pt}(a)=(x\,r-2)/2(x\,r)\,\hat{f}_t(a) \tag{12}$$

is asymptotically unbiased and is appropriate for estimating $N_{pt}(a)$.

An estimator for $f\,(a)$, the standardized variance male gamete pool allele frequencies for ramets within clones for entire orchards, is given by the average $\hat{f}_t(a)$ over clones and can be written as

$$\hat{f}\,(a)=\left[\sum_{l=1}^{n}\sum_{i=1}^{x}\sum_{k=1}^{r}(q_{ijkl}-p_{oij})^2/p_{oij}(1-p_{oij})\right]/n\,x\,r. \tag{13}$$

A corresponding entire seed orchard estimator for $N_p(a)$ is given by

$$\hat{N}_p(a)=(m-1)/2(\hat{f}\,(a)\,m-1). \tag{14}$$

For large $m$, the estimator given in Eq. 15 may be used in place of the estimator shown in Eq. 14:

$$\hat{N}_{p}(a)=(n\,x\,r-2)/2(n\,x\,r)\,\hat{f}\,(a). \tag{15}$$

An estimator for the standardized variance in allele frequency among male gamete pools for a single clone is given by

$$\hat{f}_t(b)=\left[\sum_{i=1}^{x}(\hat{q}_{ijt}-p_{oij})^2/p_{oij}(1-p_{oij})\right]/x, \tag{16}$$

where $t$ again indicates clone $t$. Using the expression for the variance of $\tilde{q}_{ijt}$ given in Eq. 6, the expected value of $\hat{f}_t(b)$ is found to be

$$E[\hat{f}_t(b)]=(r\,m+2\,N_{pt}(b)-1)/2\,N_{pt}(b)\,r\,m. \tag{17}$$

An estimator for number of effective pollen parents in the gamete pool for clone $t$ is then

$$\hat{N}_{pt}(b)=(r\,m-1)/2(\hat{f}_t(b)\,r\,m-1). \tag{18}$$

For large $r\,m$ the estimator

$$\hat{N}_{pt}(b)=(x-2)/2\,x\,\hat{f}_t(b) \tag{19}$$

can be used.

The comparable estimator for the parameter for clonal gamete polls over an entire orchard is given by

$$\hat{f}_t(b)=\left[\sum_{l=1}^{n}\sum_{i=1}^{x}(\bar{q}_{ijl}-p_{oij})^2/p_{oij}(1-p_{oij})\right]/n\,x, \tag{20}$$

and the estimator for the effective number of pollen parents in clonal male gamete pools across all clones is

$$\hat{N}_p(b)=(r\,m-1)/2(r\,m\,\hat{f}(b)-1). \tag{21}$$

Again, for large $r\,m$, the estimator

$$\hat{N}_p(b)=(x\,n-2)/2(x\,n)\,\hat{f}(b) \tag{22}$$

is appropriate.

We now turn to estimators for the parameters describing the male gamete pool for an entire orchard. Estimates of the $f$ parameter for an orchard's gamete pool can be obtained from

$$\hat{f}(c)=\left[\sum_{i=1}^{x}(\bar{q}_{ij}-p_{oij})^2/p_{oij}(1-p_{oij})\right]/x. \tag{23}$$

Taking the expected value of the right-hand side of Eq. 23 and substituting the expression in Eq. 7 for the variance of $\bar{q}_{ij}$ yields

$$E[\hat{f}(c)]=[r\,m\,n+2\,N_p(c)-1]/2\,N_p(c)\,r\,m\,n.$$

The effective number of pollen parents for an orchard can be estimated by

$$\hat{N}_p(c)=(n\,r\,m-1)/2(\hat{f}(c)\,n\,r\,m-1). \tag{24}$$

However, if $n\,r\,m$ is large, the simpler estimator

$$\hat{N}_p(c)=(x-2)/2\,x\,\hat{f}(c) \tag{25}$$

can be used.

## Confidence intervals for the $N_p$

Appealing to the central limit theorem, we find that the random variable

$$X_{ijkl}=(q_{ijkl}-p_{oij})/[p_{oij}(1-p_{oij})\cdot(m+2\,N_p(a)-1)/2\,N_p(a)\,m]^{1/2}$$

is approximately normally distributed as a $N(0,1)$ random variable for large $2\,N_p(a)\,m$. Therefore, $X^2_{ijkl}$ is approximately distributed as $\chi^2$ with one degree of freedom when $2\,N_p(a)\,m$ is large. In our model, the male gamete pool for each ramet is considered to be an independent sample of the gametes from the orchard and the genes at each locus are assumed to segregate independently. Therefore,

$$\sum_{i}^{n}\sum_{k}^{r}X^2_{ijkl}=[(m+2\,N_p(a)-1)/2\,N_p(a)\,m]^{-1}\cdot\sum_{i}^{x}\sum_{k}^{r}(q_{ijkl}-p_{oij})^2/p_{oij}(1-p_{oij}) \tag{26}$$

is distributed approximately as $\chi^2$ with $x\,r$ degrees of freedom. Substituting Eq. 8 into Eq. 26 and rewriting the term in brackets yields $m\,D\,\hat{f}_t(a)/[1+(m-1)/2\,N_{pt}(a)]$ distributed approximately as $\chi^2$ with $D=x\,r$ degrees of freedom. For an observed value of $\hat{f}_t(a)$, we obtain

$$\chi^2_{1-\alpha/2\,(D)}\leq m\,D\,\hat{f}_t(a)/[1+(m-1)/2\,N_{pt}(a)]\leq\chi^2_{\alpha/2\,(D)} \tag{27}$$

with confidence coefficient $(1-\alpha)$. Following revision of Eq. 27, the $(1-\alpha)$ confidence interval (CI) for $N_{pt}(a)$ may be expressed as

$$(1-\alpha)\text{ CI for }N_{pt}(a)=(\chi^2_{1-\alpha/2\,(D)}(m-1)/2[D\,\hat{f}_t(a)\,m-\chi^2_{1-\alpha/2\,(D)}],\ \chi^2_{\alpha/2\,(D)}(m-1)/2[D\,\hat{f}_t(a)\,m-\chi^2_{\alpha/2\,(D)}]. \tag{28}$$

Confidence intervals for the remaining $N_p$ can be derived in an analogous manner and are presented in Table 1. However, for the cases in which number of gametes sampled to determine allele frequencies is large, a method is available to determine confidence intervals for $N_p$ based on the sampling distribution for large sample estimates of $N_p$. We develop the confidence intervals for $N_{pt}(a)$ for these samples here. Confidence intervals of this type can be obtained for the remaining $N_p$ by similar methods.

For sufficiently large $m$, the variance of allele frequency among ramets of clone $t$ can be shown from the expression given in Eq. 5 to be approximately

$$\mathrm{Var}(q_{ijkt})\approx p_{oij}(1-p_{oij})/2\,N_{pt}(a).$$

The expected value of $\hat{f}_t(a)$ is then found from Eq. 9 to be approximately

$$E[\hat{f}_t(a)]\approx 1/2\,N_{pt}(a). \tag{29}$$

Substitution of Eq. 8 into Eq. 26 and replacement of

$(m+2\,N_{pt}(a)-1)/2\,N_{pt}(a)\,m$ by $E[\hat{f}_t(a)]$

**Table 1.** Confidence intervals for effective numbers of pollen parents in different gamete pools

| Gamete pool | Reference unit | Degrees of freedom ($D$) | $(1-\alpha)$ Confidence interval |
|---|---|---|---|
| Small seed sample per ramet | | | |
| Ramet | Entire orchard | $x\,n\,r$ | $\frac{(m-1)\,\chi^2_{1-\alpha/2\,(D)}}{2[D\,m\,\hat{f}\,(a)-\chi^2_{1-\alpha/2\,(D)}]}$, $\frac{(m-1)\,\chi^2_{\alpha/2\,(D)}}{2[D\,m\,\hat{f}\,(a)-\chi^2_{\alpha/2\,(D)}]}$ |
| Clone | Single clone | $x$ | $\frac{(r\,m-1)\,\chi^2_{1-\alpha/2\,(D)}}{2[D\,r\,m\,\hat{f}_t(b)-\chi^2_{1-\alpha/2\,(D)}]}$, $\frac{(r\,m-1)\,\chi^2_{\alpha/2\,(D)}}{2[D\,r\,m\,\hat{f}_t\,(b)-\chi^2_{\alpha/2\,(D)}]}$ |
| Clone | Entire orchard | $x\,n$ | $\frac{(r\,m-1)\,\chi^2_{1-\alpha/2\,(D)}}{2[D\,r\,m\,\hat{f}\,(b)-\chi^2_{1-\alpha/2\,(D)}]}$, $\frac{(r\,m-1)\,\chi^2_{\alpha/2\,(D)}}{2[D\,r\,m\,\hat{f}\,(b)-\chi^2_{1-\alpha/2\,(D)}]}$ |
| Orchard | Entire orchard | $x$ | $\frac{(n\,r\,m-1)\,\chi^2_{1-\alpha/2\,(D)}}{2[D\,n\,r\,m\,\hat{f}\,(c)-\chi^2_{1-\alpha/2\,(D)}]}$, $\frac{(n\,r\,m-1)\,\chi^2_{\alpha/2\,(D)}}{2[D\,n\,r\,m\,\hat{f}\,(c)-\chi^2_{\alpha/2\,(D)}]}$ |
| Large seed sample per ramet | | | |
| Ramet | Entire orchard | $x\,n\,r$ | $\frac{\hat{N}_p(a)\,\chi^2_{1-\alpha/2\,(D)}}{D-2}$, $\frac{\hat{N}_p(a)\,\chi^2_{\alpha/2\,(D)}}{D-2}$ |
| Clone | Single clone | $x>2$ | $\frac{\hat{N}_{pt}(b)\,\chi^2_{1-\alpha/2\,(D)}}{D-2}$, $\frac{\hat{N}_{pt}(b)\,\chi^2_{\alpha/2}}{D-2}$ |
| Clone | Entire orchard | $x\,n$ | $\frac{\hat{N}_p(b)\,\chi^2_{1-\alpha/2\,(D)}}{D-2}$, $\frac{\hat{N}_p(b)\,\chi^2_{\alpha/2}}{D-2}$ |
| Orchard | Entire orchard | $x>2$ | $\frac{\hat{N}_p(c)\,\chi^2_{1-\alpha/2\,(D)}}{D-2}$, $\frac{\hat{N}_p(c)\,\chi^2_{\alpha/2\,(D)}}{D-2}$ |

indicates that

$$D\,\hat{f}_t(a)/E[f_t(a)]$$

is distributed approximately as $\chi^2$ with $D=x\,r$ degrees of freedom. Substitutions involving Eqs. 12 and 29 reveal that

$$(D-2)\,N_{pt}(a)/\hat{N}_{pt}(a)$$

is also approximately distributed as $\chi^2$ with $D$ degrees of freedom. This form suggests

$$\hat{N}_{pt}(a)\,\chi^2_{1-\alpha/2\,(D)}/(D-2)\leq N_{pt}(a)\leq \hat{N}_{pt}(a)\,\chi^2_{\alpha/2\,(D)}/(D-2) \tag{30}$$

as the $1-\alpha$ confidence interval for $N_{pt}(a)$.

## Numerical results from two seed orchards

In coniferous species, the allelic state of the male gamete involved in forming a zygote can be ascertained for a number of allozyme loci. If an allozyme locus is active in both the haploid megagametophytic and the diploid embryonic tissues, the allelic state of the male contribution can be determined by deduction from the allelic state of the megagametophyte and genotype of the zygote. Thus, for these loci, allele frequencies in pollen pools can be estimated from samples of seed.

Data for seven clones in the Canadian Pacific Forest Products Limited (formerly CIP Inc.) combined clonal and seedling high-elevation Douglas fir seed orchard located near Victoria, BC were suitable for use in estimating the effective numbers of pollen parents. This orchard is arranged in 16 blocks, and seed samples were collected from each seed-bearing tree in six randomly selected blocks in 1983. It had been cooled with a water-spray treatment to reduce pollen contamination from outside sources and to promote random mating among individuals within the orchard (Fashler and El-Kassaby 1987). A detailed description of this orchard and the sampling method employed is given in Ritland and El-Kassaby (1985) and El-Kassaby and Ritland (1986). Genotypic scores for six unlinked allozyme loci were available for megagametophytic and embryonic tissues of freshly germinated seed from three ramets of each of the seven clones. The genotype of every cone-producing tree in the six blocks was determined for each of the loci studied in the embryonic tissue. From these data, estimates of pollen allele frequencies were obtained for each ramet, clone, and for the clonal portion of the orchard. Allele frequencies computed for the cone-bearing population in the sampled blocks were assumed to be representative of the allele frequencies of the pollen-producing trees in the entire orchard. Using these data, $f$ parameters, effective numbers of pollen parents, and confidence intervals for

**Table 2.** Estimates of standardized variances and effective numbers of pollen parents, with confidence intervals, in a Douglas fir seed orchard

| Clone number | Standardized variance $\hat{f}$ | Effective number $\hat{N}_p$ | 95% Confidence intervals for $N_p$ |
|---|---|---|---|
| Ramets in clones | | | |
| 25 | 0.0588 | $\infty$[a] | 7.81, $\infty$[b] |
| 87 | 0.0540 | $\infty$ | 9.07, $\infty$ |
| 160 | 0.2923 | 2.07 | 0.82, 4.66 |
| 168 | 0.1126 | 10.15 | 2.60, $\infty$ |
| 344 | 0.1207 | 8.64 | 2.37, 208.21 |
| 410 | 0.0766 | 47.03 | 4.63, $\infty$ |
| 439 | 0.0674 | 645.16 | 5.78, $\infty$ |
| Orchard-wide | 0.1116 | 10.38 | 5.94, 21.39 |
| Clones | | | |
| 25 | 0.0168 | $\infty$ | 8.30, $\infty$ |
| 87 | 0.0246 | 210.02 | 5.06, $\infty$ |
| 160 | 0.1820 | 3.06 | 0.57, 9.12 |
| 168 | 0.0444 | 22.02 | 2.54, $\infty$ |
| 344 | 0.0508 | 17.10 | 2.19, $\infty$ |
| 410 | 0.0292 | 70.47 | 4.11, $\infty$ |
| 439 | 0.0278 | 88.28 | 4.36, $\infty$ |
| Orchard-wide | 0.0536 | 15.56 | 7.58, 34.32 |
| Orchard level | | | |
| | 0.0048 | 310.46[c] | 24.92, $\infty$ |
| | | 69.74[d] | 21.57, 251.91 |

[a] Negative values for $\hat{N}_p$ are interpreted as infinite effective number

[b] Negative values for end points are interpreted as infinite

[c] Computed using the estimator given in Eq. 24

[d] Computed using the estimator given in Eq. 25

the effective numbers were estimated and are presented in Table 2.

$N_p$ varied considerably among the clones investigated both on ramets-in-clones and clonal bases. Large $N_p$ were found for clones 25, 87, and 439. These observations are compatible with a pattern of mating that involves contributions by a large number of individuals in the orchard to the male gamete pool for each clone. The floral phenology for these clones is congruous with such an interpretation, since all three flower in the middle of the orchard's flowering season. In contrast, clone 160 has particularly small $N_p$'s, which indicates that pollen from only a few individuals in the orchard were successful in fertilizing the majority of the ovules produced by this clone. As clone 160 does not exhibit extreme phenological reproductive behavior, its small $N_p$ is as yet unexplained. Small $N_p$ for a clone, however, is a result that is consistent with male-female interaction in mating, which has been found to exist in Douglas fir seed orchards (Apsit et al. 1989). Clones 168 and 344, which have small $N_p$, are receptive to pollination early in the season, and clone 410, which flowers late, was found to have intermediate $N_p$ values.

Primarily because few seeds were sampled per ramet ($m = 15$), confidence intervals for the $N_p$ were large except for the cases with small $N_p$. It is thus evident that seed samples of much larger size are necessary for each ramet in situations for which moderate and large $N_p$ are possible.

The large $\hat{N}_p$ values obtained at the orchard level suggest that there was little effective contamination by pollen from outside the orchard in the progeny of the seven clones investigated. El-Kassaby and Ritland (1986), who studied the same seed crop in this orchard, found there was almost no contamination from outside sources. The interpretation of our effective number results therefore agrees with the conclusion of El-Kassaby and Ritland (1986).

The second source of data for estimates of $N_p$ was the Champion International low-specific-gravity loblolly pine seed orchard located near Newberry/SC. Electrophoretic assays of embryos and megagametophytes for seven unlinked gene loci were obtained from seed collected in 1976 from two ramets for each of nine clones. Samples for each ramet consisted of 56 seeds. A description of the seed orchard and the allozyme loci investigated can be found in Friedman and Adams (1985). Allele frequencies in male gamete pools were computed from the results of the isozyme assays and used to estimate $f$, $N_p$, and confidence intervals for $N_p$, for ramets in clones, clones, and the entire orchard. Estimates of the $f$ and $N_p$ parameters and confidence intervals for the $N_p$ are presented in Table 3.

Values for $\hat{N}_p$ did not vary greatly among clones and almost always fell in a range ($10 \leq N_p \leq 30$) that can be considered moderate for effective numbers. While the extreme $N_p$ values observed for some clones in the Douglas fir seed orchard were not found in this loblolly pine orchard, the $N_p$ values did fall within a range that indicates departure from panmictic mating behavior. While little information concerning flowering phenology is available for clones in this orchard, it is known that clones 7, 17, and 20, which have $N_p$ estimates below 20, tend to flower late in the orchard's flowering season. The relative uniformity of $\hat{N}_p$ among clones and the intermediate magnitude of values could have resulted from a combination of substantial pollen contamination from outside sources and out-of-phase flowering phenology for some clones. Friedman and Adams (1985) found that considerable pollen contamination occurred in the 1976 seed crop of this orchard.

Confidence intervals for the $N_p$ are much narrower than intervals computed for comparable $\hat{N}_p$ for the Douglas fir orchard. This reduction probably is a reflection of the larger per-ramet seed samples for the loblolly pine orchard ($m = 56$ for loblolly pine versus $m = 15$ for the Douglas fir). The intervals estimated for values of $\hat{N}_p$ for individual clones are still large for intermediate $N_p$ values.

**Table 3.** Estimates of standardized variances and effective numbers of pollen parents, with confidence intervals, in a loblolly pine seed orchard

| Clone number | Standardized variance $\hat{f}$ | Effective number $\hat{N}_p$ | 95% Confidence intervals for $N_p$ |
|---|---|---|---|
| Ramets in clones | | | |
| 3 | 0.0515 | 14.60 | 4.45, 50.38 |
| 6 | 0.0479 | 16.35 | 4.85, 62.90 |
| 7 | 0.0497 | 15.43 | 4.65, 55.98 |
| 10 | 0.0480 | 16.29 | 4.83, 62.35 |
| 12 | 0.0308 | 38.03 | 8.37, $\infty$ [a] |
| 13 | 0.0396 | 22.61 | 6.09, 146.24 |
| 14 | 0.0360 | 27.00 | 6.84, 336.21 |
| 17 | 0.0615 | 11.27 | 3.64, 32.55 |
| 20 | 0.0564 | 12.75 | 4.01, 39.74 |
| Orchard-wide | 0.0468 | 16.96 | 11.40, 25.52 |
| Clones | | | |
| 3 | 0.0409 | 15.50 | 3.09, 55.33 |
| 6 | 0.0442 | 14.06 | 2.85, 47.72 |
| 7 | 0.0468 | 13.10 | 2.68, 43.06 |
| 10 | 0.0397 | 16.09 | 3.18, 58.75 |
| 12 | 0.0269 | 27.54 | 4.83, 174.51 |
| 13 | 0.0248 | 31.28 | 5.29, 260.86 |
| 14 | 0.0303 | 23.19 | 4.25, 114.78 |
| 17 | 0.0512 | 11.73 | 2.44, 36.86 |
| 20 | 0.0502 | 12.00 | 2.49, 38.02 |
| Orchard-wide | 0.0394 | 16.24 | 10.10, 25.17 |
| Orchard level | | | |
| | 0.0313 | 16.47 [b] | 3.88, 39.33 |
| | | 11.40 [c] | 3.85, 36.52 |

[a] Negative value for an end point is interpreted as being infinitely large

[b] Computed using the estimator given in Eq. 24

[c] Computed using the estimator given in Eq. 25

It is evident that somewhat larger seed samples per ramet ($m > 56$) are necessary to produce desirable confidence intervals for $N_p$'s for individual clones when $N_p$ values are in the moderate range.

## Discussion

Values of the parameters in the triple $[N_p(a), N_p(b), N_p(c)]$ provide insight into mating patterns in clonal seed orchards. Here reference is to $N_p(a)$'s and $N_p(b)$'s determined on an orchard-wide basis. It can be shown that the three parameters have values that occur in the order $N_p(a) \leq N_p(b) \leq N_p(c)$. The following partition of the $N_p$ parameter space is useful for interpreting mating behavior.

### *Case I*

*$N_p(a)$ small, $N_p(b)$ and $N_p(c)$ large.* This indicates behavior that results in allele frequency variation among male gamete populations for ramets within clones, which is equivalent to the sampling variation produced by genetic drift, but which yields negligible allele frequency sampling variation among the gamete populations for clones and for the entire orchard. This pattern occurs because different ramets within clones have different gamete samples. It will be produced when most matings occur among neighboring plants and ramets are randomly located with respect to the clonal composition.

### *Case II*

*$N_p(a) = N_p(b)$ small, $N_p(c)$ large.* Behavior results in allele frequency variation that is equivalent to that produced by genetic drift among male gamete populations for clones. In addition, male gamete populations for the ramets of each clone are identical genetic replicate populations. This pattern of variation could result from clonal variation in flowering phenology. The Douglas fir seed orchard we investigated, with its large $N_p(c)$ and its moderate but similar $N_p(a)$ and $N_p(b)$, most closely fits this case.

### *Case III*

*$N_p(a) = N_p(b) = N_p(c)$ small.* This indicates a pattern that is consistent with genetic drift at the level of the male gamete pool for the entire orchard, but one in which the male gamete populations for all ramets, regardless of clone, are identical genetic replicate populations. The loblolly orchard studied, which had moderate but almost equal $N_p$ values, most closely fits this case.

### *Case IV*

*$N_p(a)$ and $N_p(b)$ small, but $N_p(a) < N_p(b) \leq N_p(c)$.* The pattern here is consistent with genetic drift variation among male gamete pools for clones, with additional drift effects occurring among male gamete pools for ramets in clones. If $N_p(c)$ is moderate in value, behavior is equivalent to genetic drift at the level of the orchard male gamete pool, with additional drift variation occurring among clonal male gamete pools. If $N_p(c)$ is large, allele frequency sampling variation at the level of the orchard gamete pool is negligible.

### *Case V*

*$N_p(a)$, $N_p(b)$, and $N_p(c)$ large.* When this case occurs, the seed orchard behaves as a panmictic unit. There is little allele frequency variation among the male gamete pools at the ramets in clones, clones, or orchard levels.

Pollen migration from outside sources into seed orchards probably causes allele frequencies in orchard gamete pools to differ from the allele frequency of the ramet population of the orchard. The result is a reduction in $N_p$ that is indistinguishable from one produced by other factors. It is therefore difficult to determine whether pollen contamination is a contributing factor. Even when the pollen contamination rate is high, the interpretation of the reduced $N_p$ is the same as if other factors are responsible, i.e., the number of effective pollen parents from the orchard itself is small. Consequences of the factors producing the small effective numbers, however, are different. In the case of pollen contamination, orchard seed pools suffer from an infusion of genes from undesirable sources. If migration of alien pollen is not a factor, orchard seed gene pools are depauperate because of the disproportionate large contributions of genes from a few clones.

Waples (1989) has pointed out that the temporal method of estimating effective population size, which is the method we propose for determining $N_p$, is less precise for populations with large effective number than for those with a small value. However, such a method is exactly suited for use in the situation of interest in seed orchards. There, concern is focused on detecting clones and orchard populations that have small $N_p$, and thus the greatest estimation precision is desired for populations with small values for $N_p$, while less precision is necessary for those with high $N_p$ values.

Although in this paper we address only estimation of $N_p$ in clonal seed orchards, it is clear that the methods described can be adapted for use in natural populations. Similarly, these methods can be employed to determine numbers of effective males in seed crops produced by various natural regeneration methods, such as seed tree and shelterwood methods as well as in seed production areas.

## References

Apsit VJ, Nakamura RR, Wheeler NC (1989) Differential male reproductive success in Douglas fir. Theor Appl Genet 77: 681–684

Askew GR (1988) Estimation of gamete pool compositions in clone seed orchards. Silvae Genet 37:227–232

Crow JF, Denniston C (1988) Inbreeding and variance effective population numbers. Evolution 42:482–495

Daniels JD (1978) Efficacy of supplemental mass pollination in Douglas fir seed orchards. Silvae Genet 27:52–58

El-Kassaby YA, Askew GR (1991) The relation between reproductive phenology and reproductive output in determining the potential gametic pool profile in a Douglas fir seed orchard. For Sci (in press)

El-Kassaby YA, Ritland K (1986) Low levels of pollen contamination in a Douglas fir seed orchard as detected by allozyme markers. Silvae Genet 35:224–229

El-Kassaby AY, Fashler AMK, Sziklai O (1984) Reproductive phenology and its impact on genetically improved seed production in a Douglas fir seed orchard. Silvae Genet 33:120–125

El-Kassaby YA, Fashler AMK, Crown M (1989) Variation in fruitfulness in a Douglas fir seed orchard and its effect on crop management decisions. Silvae Genet 38:113–120

Fashler AMK, El-Kassaby YA (1987) The effect of water spray cooling treatment on reproductive phenology in a Douglas fir seed orchard. Silvae Genet 36:245–249

Friedman ST, Adams WT (1985) Estimation of gene flow into two seed orchards of loblolly pine (*Pinus taeda* L.). Theor Appl Genet 69:609–615

Griffin AR (1982) Clonal variation in radiata pine seed orchards. I. Some flowering, cone, and seed production traits. Aust For Res 12:295–302

Griffin AR (1984) Clonal variation in radiata pine seed orchards. II. Flowering phenology. Aus For Res 14:271–281

Kimura J, Crow JF (1963) The measurement of effective population number. Evolution 17:279–288

Krimbas CB, Tsakas S (1971) The genetics of *Dacus oleae*. V. Changes of esterase polymorphism in a natural population following insecticide control – selection or drift? Evolution 25:454–460

Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. Genetics 74:175–195

Nei M, Tajima F (1981) Genetic drift and estimation of effective population size. Genetics 98:625–640

Pollak E (1983) A new method for estimating the effective population size from allele frequency changes. Genetics 104: 531–548

Rao CR (1973) linear statistical inference and its applications, 2nd edn. John Wiley, New York

Ritland K, El-Kassaby YA (1985) The nature of inbreeding in a seed orchard of Douglas fir as shown by an efficient multilocus model. Theor Appl Genet 71:375–384

Schmidtling RC (1983) Genetic variation in fruitfulness in a loblolly pine (*Pinus taeda* L.) seed orchard. Silvae Genet 32:76–80

Shaw DV, Allard RW (1982) Estimation of outcrossing rates in Douglas fir using isozyme markers. Theor Appl Genet 62:113–120

Waples RS (1989) A generalized approach for estimating effective population size from temporal changes in allele frequency. Genetics 121:379–391

Wright S (1931) Evolution in Mendelian populations. Genetics 16:97–159