Theor Appl Genet (1994) 89:936–942

J. Crossa · R. Vencovsky

Implications of the variance effective population size on the genetic conservation of monoecious species

Received: 9 September 1993 / Accepted: 10 June 1994

Abstract The concept of variance effective population size [$Ne_{(v)}$] and other expressions are reviewed and described for specific sampling steps in germplasm collection and regeneration of monoecious species. Special attention is given to procedures for computing the variance of the number of contributed gametes [V(k)] to the next generation. Drift, as it occurs between generations, was considered to contain a component due to the sampling of parents and a subsequent component due to the sampling of gametes. This demonstrates that drift, caused by reduction of seed viability, damages the genetic integrity of accessions stored in germplasm banks. The study shows how mating designs, such as plant-to-plant or chain crossings with additional female gametic control, can partially alleviate this problem. Optimal procedures for increasing $Ne_{(v)}$ when collecting germplasm in the field are also discussed. The effect of different female and male gametic control strategies on $Ne_{(v)}$ is considered under several situations. Practical examples illustrating the use of V(k) and $Ne_{(v)}$ expressions are given.

Key words Effective population size · Variance of the number of contributed gametes · Genetic resources preservation · Monoecious species

Introduction

Germplasm conservation aims to preserve the genetic variability in a crop species by avoiding, as much as possible, genetic erosion. The effectiveness of a germplasm bank in preserving, regenerating, and collecting germplasm depends on sampling procedures, random genetic drift, and seed viability.

The effective population size, taken as a measure of the genetic representativeness of a seed sample, can be adapted to specific aspects of genetic resources preservation such as seed regeneration and collection. Crossa (1989) suggests that, in accession regeneration, germplasm banks should establish procedures aimed at controlling the number of progeny per individual or family. This author reinforces the known fact that equalizing the reproductive output of each family doubles the effective size of the population.

The concept of effective population size was first introduced by Wright (1931) when studying breeding structure and genetic representativeness of an actual population as related to an idealized population. In an ideal population all individuals have equal expectation of being the parents of any individual progeny. However, in an arbitrary breeding population not all individuals produce progeny and, therefore, may not influence the genetic structure and composition of the succeeding generation. Only progenitors that produce progeny influence the genetic constitution of the next generation and consequently affect the effective population size.

Several factors affect the effective size of the population: the number of viable offspring of a given generation, the number of gametes contributed per individual in the parental generation, variation in the number of individuals per generation, unequal numbers of individuals of each sex, or the species' natural rate of self-fertilization. All these factors influence the genetic contribution through successive generations and should therefore be considered when effectively measuring the size of a breeding population.

Two different approaches have historically been used to quantify the effective population size (Crow and Kimura 1970). The first relates a breeding population's inbreeding coefficient to the inbreeding of an ideal population. The second relates the sampling variance of allelic frequency in a breeding population to the sampling variance of an idealized population (Kimura and Crow 1963; Wright

---

Communicated by A. R. Hallauer

J. Crossa (✉)
Biometrics and Statistics Unit, International
Maize and Wheat Improvement Center (CIMMYT).
Lisboa 27, Apdo. Postal 6-641, Mexico, D. F., Mexico

R. Vencovsky
Instituto de Genetica, ESALQ, USP, C. P. 83,13.400-970,
Piracicaba, Sao Paulo, Brazil

1969; Crow and Kimura 1970). The effective population size of a population, whether actual or hypothetical, is the size of a theoretical population that has the same inbreeding coefficient, or the same allelic frequency variance, as that of the actual or hypothetical population under study. However, researchers, in general, lack direct formulae for calculating the effective population size either (1) for populations used in artificial selection schemes (inbred line selection, recurrent selection, etc.) or (2) for specific populations used in genetic resource conservation, such as germplasm collection and regeneration.

Vencovsky (1978) outlined procedures for calculating the variance effective population size for monoecious populations subjected to artificial selection. He later (Vencovsky 1987) adapted some of these procedures to specific aspects of germplasm collection and preservation.

In the present study, we review and describe methods for calculating: (1) the variance of the number of male and female gametes contributed by individuals of monoecious species, (2) the variance effective population size relative to germplasm collection and accession regeneration and (3) the variance of the number of gametes as related to the mating system and the manner in which gametes (male and female) are sampled for germplasm collection and regeneration.

The variance effective population size [$Ne_{(v)}$]

Kimura and Crow (1963) and Crow and Kimura (1970) proposed two different methods of describing effective population size. One is the inbreeding effective size [$Ne_{(f)}$] which is naturally dependent upon the number of individuals in the parental generation and the resulting probability that alleles of individuals are identical by descent. The second is the variance effective size [$Ne_{(v)}$], which is a function of the number of gametes (offspring) contributed by individuals in a population to the next generation. The inbreeding effective number is related to the number of parents, while the variance effective number is related to the number of offspring.

The concept of variance effective population size, which is particularly relevant to germplasm conservation, is defined by Crow and Kimura (1970) as

$$Ne_{(v)}=2N_t/[\{s^2(k)/(\bar{k}\}(1+\alpha_{t-1})+(1-\alpha_{t-1})] \quad (1)$$

where N_t is the number of offspring in generation t, α_{t-1} is a measure of departure from Hardy-Weinberg equilibrium in the parental generation, $s^2(k)$ and $\bar{k}$ are the variance and mean of the number of gametes contributed by the parents, respectively. Under random mating the expected value of α_{t-1} is $-1/(2N_{t-1}-1)$ (Crow and Kimura 1970). In most plant conservation studies, however, $2N_{t-1}$ is large enough such that taking $\alpha_{t-1}=0$ for panmictic species will not cause a serious bias in $Ne_{(v)}$. Therefore, based on this assumption, the variance effective population size reduces to

$$Ne_{(v)}=2N_t/[s^2(k)/\bar{k}+1] \quad (2)$$

Wright (1969) gives a similar expression for $Ne_{(v)}$, assuming Hardy-Weinberg proportions in the parental generation ($F_{IS}=0$), which includes the ratio $V(k)/\bar{k}$ instead of $s^2(k)/\bar{k}$. In our applications of $Ne_{(v)}$, we will use the latter ratio with the Gaussian correction $s^2(k)=[N_{t-1}/(N_{t-1}-1)]\ V(k)$, whenever necessary.

In the special case of random sampling of gametes, assuming that each parent has an equal probability ($1/N_{t-1}$) of contributing to the next generation, k follows a binomial distribution with variance $V_{(k)}=2N_t(1/N_{t-1})(1-1/N_{t-1})$. Since $\Sigma k=2N_t$ is the total number of gametes contributed such that $\bar{k}=(\Sigma k)/N_{t-1}=(2N_t)/N_{t-1}$ we have $V(k)=\bar{k}(1-1/N_{t-1})$ and $s^2(k)=\bar{k}$, which is a Poisson variance for k. Consequently we get $Ne_{(v)}=N_t$. With large population sizes, $s^2(k)$ approximates $V(k)$.

Extended model for variance of the contributed gametes [V(k)] in monoecious species

In applying $Ne_{(v)}$, we consider the model proposed by Vencovsky (1978) with the simplified notation $N_{t-1}=N$ and $N_t=n$ for the potential number of parents and number of offspring, respectively.

This model assumes that from an initial set of N monoecious diploid plants, F are sampled for seed collection ($0<F\leq N$). These F parents contribute both female and male gametes. From the remaining N–F plants, R are additionally sampled to contribute only male gametes ($0<R\leq N-F$). We then have the following proportions: $u=F/N$ and $v=(F+R)/N=M/N$ for seed parents and pollen parents, respectively ($0<u\leq 1$; $0<v\leq 1$). The model can accommodate the derivation of $Ne_{(v)}$ for recurrent selection schemes or germplasm regeneration and collection procedures.

When M=F=N, there is no selection or sampling and all individuals (N) from generation t–1 potentially contribute male and female gametes to the next generation (t) and u=v=1. This is the case when an accession from a germplasm bank is regenerated, all plants are used for pollination, and seed is harvested from all plants. However, loss of seeds due to germination problems may result in only a proportion uN=F of individuals in the initial generation producing progeny in the next generation. This model could also be used to derive the effective population size for germplasm collection, where F female plants are collected in the field, and a large (unknown) number of plants (M=N) contribute male gametes.

Vencovsky (1978) derived the variance of the number of contributed gametes [V(k)] under the proposed model, assuming sampling of plants without replacement and taking into consideration that N-M plants contribute no gametes. In subsequent sampling of gametes, a locus with two alleles was assumed, and the sampling variance in gene frequency was obtained following the reasoning given by Crow and Kimura (1970) for deriving $Ne_{(v)}$. Variance of the number of contributed gametes is given by

$$V(k)=uV(k_{fs})+u(1-u)(\bar{k}_{fs})^2+vV(k_{ms})+v(1-v)(\bar{k}_{ms})^2+ 2u(1-v)\bar{k}_{fs}\bar{k}_{ms} \text{ (for } u\leq v\text{) (Vencovsky 1978).} \quad (3)$$

The first two terms on the right hand side of equation 3 relate to a female component, the third and fourth to a male component, and the last is a covariance term between male and female gametes. With k_f and k_m standing for the number of female and male gametes, we have $V(k_{fs})$ as the variance of k_f within F, while $V(k_{ms})$ denotes the variance of k_m within M=F+R. With n offspring and a total of 2n gametes, $\bar{k}_{fs}=n/F$ (mean of k_f within F) and $k_{ms}=n/M$ (mean of k_m within M).

Relative to female gametes there are two groups of plants: (1) F, contributing, on average, $\bar{k}_{fs}$ gametes and (2) N–F, with zero contribution. The term $u(1-u)(\bar{k}_{fs}-0)^2=u(1-u)(\bar{k}_{fs})^2$ is the female between-groups component of $V(k_f)$, whereas $v(1-v)(\bar{k}_{ms}-0)^2$ is the male between-groups component of $V(k_m)$. The last term of equation 3 is a between-groups covariance that arises from the two pairs of means $(\bar{k}_{fs},\bar{k}_{ms};\ 0,\ 0)$ and is present only when a set of plants is totally discarded as parents of the next generation.

One basic assumption underlying V(k) (equation 3) is that there is no correlation between the number of female and male gametes contributed per individual within the set of F monoecious plants. Therefore, the model excludes any self-fertilization (natural or artificial).

The overall means of k_f and k_m relative to the initial set of N parents are $\bar{k}_f=n/N$, $\bar{k}_m=n/N$ and $\bar{k}=(2n/N)=\bar{k}_f+\bar{k}_m$. Variance V(k) reduces to $V(k_f)+V(k_m)$ when u=v=1.

For practical purposes, we can visualize V(k) and $Ne_{(v)}$ as being dependent upon the manner in which female and male gametes are sampled, the mating scheme, and the total number of gametes, 2n.

Another factor affecting V(k) and $Ne_{(v)}$ is the proportion of parental individuals. In accession regeneration, this factor is the result of seed loss due to poor germination. We will examine how different ways of sampling female and male gametes affect the magnitude of V(k) (equation 3) and the resulting $Ne_{(v)}$ (equation 2).

$Ne_{(v)}$ expressions for large parental population size

Our objective is to develop manageable $Ne_{(v)}$ expressions which are not overly dependent on underlying assumptions and, consequently, are useful in practical situations. As will be seen, considerable simplification is attained if the number of plants (N) of the parental population is large enough that $N/(N-1)\cong 1$, which also implies that $V(k)\approx s^2(k)$ and $\alpha\cong 0$. Assuming a panmictic population of N individuals with n offspring stemming from F seed parents and M pollen parents, according to the model, $\bar{k}=2n/N$, u=F/N, and v=M/N; then, with random sampling of n female and n male gametes, $V(k_{fs})$ and $V(k_{ms})$ (terms of equation 3) are binomial variances

$V(k_{fs})=n(1/F)(1-1/F)$ and

$V(k_{ms})=n(1/M)(1-1/M)$.

We will describe the female, male, and covariance components of V(k) (equation 3) and the final form of $Ne_{(v)}$ (equation 2) for four different alternative samplings of female and male gametes. Case 1: no control of the number of male and female gametes; Case 2: control only of the number of female gametes; Case 3: control only of the number of male gametes; Case 4: control of the number of female and female gametes. Case 3 is not feasible in practice and is included here only for theoretical comparison with its counterpart (Case 2).

Case 1

Pollination is at random and the number of seeds per plant is not controlled; that is, there is no hand pollination, and unequal numbers of seeds are randomly taken from F parental plants. In this case equation 2 can be written as

$$Ne_{(v)}=n/\{1+[n(1-u)-1]/(4F)+[3n(1-v)-1]/(4\,M)\} \qquad (4)$$

(see Appendix).

Case 2

In this case the number of female gametes is controlled by taking the same amount of seeds per plant, but pollination is at random. Then equation 2 becomes

$$Ne_{(v)}=n/\{3/4+[n(1-u)]/(4F)+[3n(1-v)-1]/(4\,M)\} \qquad (5)$$

(see Appendix).

Case 3

Control of the number of male gametes through hand pollination, but unequal numbers of seeds are randomly taken from each plant. Then equation 2 is

$$Ne_{(v)}=n/\{3/4+[n(1-u)-1]/(4F)+[3n(1-v)]/(4\,M)\} \qquad (6)$$

(see Appendix).

Case 4

Using hand pollination and counting equal numbers of seeds per plant, we achieve male and female gametic control. Then equation 2 can be written as

$$Ne_{(v)}=n/\{1/2+[n(1-u)]/(4F)+[3n(1-v)]/(4\,M)\} \qquad (7)$$

(see Appendix).

$Ne_{(v)}$ in germplasm regeneration with constant population size

The following are applications of equation 4 to 7 in specific situations in which population size is kept constant (n=N). First, consider accession regeneration with N individuals planted in the field; all of them provide female and male gametes. Therefore, M=F=N and u=v=1. Thus, for Case 1 equation 4 reduces to

$Ne_{(v)}=N/[1-(1/2N)]\cong N$,

for Cases 2 and 3 (equations 5 and 6)

$Ne_{(v)}=N/[(3/4)-(1/4N)]\cong N(4/3)$,

and for Case 4 (equation 7) $Ne_{(v)}=2N$.

Although equalizing the number of male and female gametes among individuals is not likely to occur in nature, with hand pollination, and equal numbers of seeds from each pollinated plant, the size of $Ne_{(v)}$ is twice the magnitude of N, the census population.

Another factor affecting effective population size in germplasm regeneration is loss of seeds due to poor germination. Now, with M=F, u=F/N=v=M/N are less than one but n=N. This is similar to mass selection prior to pollination (selection on both sexes) and mating among F selected plants. For Case 1, with n=N, equation 4 reduces to

$$Ne_{(v)}=Nu/(1-1/2N)\cong Nu. \tag{8}$$

For Cases 2 and 3 equations 5 and 6 reduce to

$$Ne_{(v)}=N[4u/(4-u-1/N)]\cong N[4u/(4-u)], \tag{9}$$

and when u=1, $Ne_{(v)}\cong N(4/3)$.

For Case 4, equation 7 becomes

$$Ne_{(v)}=N[2u/(2-u)], \text{ and} \tag{10}$$

when u=1, $Ne_{(v)}=2N$.

For example, assume that 1000 (N) seeds are collected and stored in a germplasm bank and that, after a certain number of years, the accession is regenerated. All 1000 seeds are planted in an isolated block. Assume also that only 33% (u=1/3) of seeds germinate. If pollination is at random and unequal numbers of seeds are taken from each pollinated plant (Case 1), all 1000 offspring seeds will have $Ne_{(v)}=333$ (equation 8). When only female or male gametes (but not both) are controlled (Cases 2 and 3), $Ne_{(v)}$ in the next generation is 364 (equation 9). When pollination is controlled (hand pollination) and equal numbers of seeds are taken from each pollinated plant (Case 4), $Ne_{(v)}$ in the following generation is 400 (equation 10).

These results demonstrate that reduction of seed viability can seriously affect the genetic representativeness of the sample and produce genetic erosion. However, appropriate pollination techniques for male and female gametic control [i.e., plant-to-plant crossing (for male gametic control) and retaining equal numbers of seeds per pollinated plant (for female gametic control)] can partially alleviate this problem.

Table 1 summarizes $Ne_{(v)}$ expressions related to germplasm regeneration. Formulae in the last column of this table are applicable when there is no reduction in germination rate, such that M=N plants are available as pollen parents (v=1), but seeds are taken from a sub-set (F≤N) of plants. This situation resembles half-mass selection (Vencovsky 1976).

Table 1 $Ne_{(v)}$ for germplasm regeneration alternatives with constant population size (n=N) and monoecious species[a]

Gametic control		Number of parents	
Female	Male	Seed → F=Nu Pollen → M=F=Nu	Seed → F=Nu Pollen → M=N
		$Ne_{(v)}$	
Yes	Yes	N[2u/(2–u)]	N[4u/(1+u)]
Yes	No	N[4u/(4–u)]	N[4u/(1+2u)]
No	No	Nu	N[4u/(1+3u)]

[a] $0<u\leq 1$; $1-(1/N)\cong 1$

$Ne_{(v)}$ in germplasm collection

When germplasm is collected, the total number of pollinator plants (M) in the field is unknown but potentially large. No exact statement can be made about the proportion u=M/N and an extreme situation is M≅N and u≅1. If the number F of seed parents is a very small fraction of the entire population [such that u=(F/N)≅0] from which n seeds are randomly taken, from equation 4 (Case 1)

$$Ne_{(v)}=n/[(n-1)/(4F)-1/(4N)+1]\cong n/[(n-1)/(4F)+1]. \tag{11}$$

With large n this approaches $Ne_{(v)}=4F$, which is the effective number of F half-sib families of infinite size. For Case 2, equation 5 now has the reduced form

$$Ne_{(v)}=n/[n/(4F)-1/(4N)+3/4]\cong n/[n/(4F)+3/4]. \tag{12}$$

Since pollen control is not possible, Cases 3 and 4 (equations 6 and 7) do not apply here.

Consider that n=100 seeds are randomly collected from F=20 plants; we have $Ne_{(v)}=45$ (equation 11), but $Ne_{(v)}=50$ with equal numbers of seeds taken per plant (equation 12) (for Cases 1 and 2, respectively). In this case, the limiting factor is the number of plants (F) from which n seeds are taken. Obviously, as F increases, $Ne_{(v)}$ increases. For example, consider collecting a total of n=10 000 seeds using three sampling strategies: (1) F=100 plants (100 seeds per plant), (2) F=500 plants (20 seeds per plant) and (3) F=1000 plants (10 seeds per plant). In the first sampling strategy, (F=100) $Ne_{(v)}=388$, in the second sampling scheme (F=500) $Ne_{(v)}=1739$ and in the third sampling procedure (F=1000) $Ne_{(v)}=3077$.

When collecting germplasm in the field, the recommended strategy is taking, at random, equal numbers of seeds from the largest possible number of parental plants.

Direct method for calculating V(k) and $Ne_{(v)}$

The following examples are specific mating schemes such that the number of contributed female and male gametes are known in advance. We also assume a parental population derived from random mating and comprising a small number (N) of plants. Under such circumstances $s^2(k)$ and k can be computed directly from the data, with $Ne_{(v)}$ necessarily incorporating the Gaussian correction and the expected value of α.

With n offspring and N parents, $\bar{k}=(2n)/N$. Since $k=k_f+k_m$ per individual we have

$$V(k)=V(k_f)+V(k_m)+2Cov(k_f,k_m)$$

where $V(k_f)=(1/N)\,\bar{2}_f(k_f-\bar{k}_f)^2$, $V(k_m)=(1/N)\,\bar{2}_m(k_m-\bar{k}_m)^2$ and $Cov(k_m,k_f)=(1/N)\,\bar{2}(k_f-\bar{k}_f)(k_m-\bar{k}_m)$.
Also $s^2(k)=[N/(N-1)]V(k)$.

As already pointed out, the covariance term is a consequence of parental seed loss due to germination problems or any other situation in which a sub-set of plants is totally discarded prior to reproduction.

Only when gametes are randomly drawn can we consider $s^2(k)$ as a Poisson variance, and specifically when $u=v=1$ we have $s^2(k)=s^2(k_f)+s^2(k_m)=\bar{k}_f+\bar{k}_m=n/N+n/N=2n/N=\bar{k}$ with $Cov(k_f,k_m)=0$.

Example

The objectives of this example are to show: (1) how to obtain exact values of $Ne_{(v)}$ under specific circumstances, without any approximation or initial assumption and 2) the meaning and validity of equation 3 (variance of the number of contributed gametes).

Assume a parental population of N=6, where only F=2 plants (u=0.33) germinate and survive for reproduction (N–F=4 plants do not contribute gametes). Assume that the two surviving plants are crossed reciprocally (hand pollination) and three seeds are taken from each pollination to recover the original six plants (Table 2). With n=N=6, population size remains constant and $\bar{k}=2n/N=2$; also $\bar{k}_f=\bar{k}_m=n/N=1$. For variances we have:

$$V(k_f)=V(k_m)=(1/6)[9+9+0+0+0+0-(36/6)]=2.$$

However, the variance of the total number of gametes ($k=k_f+k_m$) per individual is

$$V(k)=(1/6)[36+36+0+0+0+0-(144/6)]=8.$$

Then $V(k)>V(k_f)+V(k_m)$, because

$$2Cov(k_f,k_m)=2(1/6)\,\{(3\times3)+(3\times3)+0+0+0+0-[(6\times6)/6]\}=4.$$

Using equation 1 and taking $\alpha=-1/(12-1)$ and $s^2(k)=[N/(N-1)]V(k)=(6/5)8=9.6$, we obtain $Ne_{(v)}=(2)(6)/[(9.6/2)(1-1/11)+(1+1/11)]=2.2$. For this case equation 10 gives the approximate value $Ne_{(v)}=6[2(0.33)/(2-0.33)]=2.4$, which is slightly biased because neither $N/(N-1)=6/5$ nor $\alpha=-1/11$ are negligible.

Applying equation 3 for comparison, we get $V(k_f)=uV(k_{fs})+u(1-u)(\bar{k}_{fs})^2=0+(2/6)(1-2/6)(6/2)^2=2$, $V(k_m)=vV(k_{ms})+v(1-v)(\bar{k}_{ms})^2=0+(2/6)(1-2/6)(6/2)^2=2$, $2Cov(k_f,k_m)=2u(1-v)\bar{k}_{fs}\bar{k}_{ms}=2(2/6)(1-2/6)(6/2)(6/2)=4$, and $V(k)=2+2+4=8$, as before.

If alternatively we assume: (1) an initial accession size of N=6 individuals, (2) that pollination is not random (control of male gametes), (3) equal numbers of seeds are taken from each pollination (control of female gametes), and (4) 100% germination (u=1), then $s^2(k)=0$, and based on equation 1 with $\alpha=-1/11$ $Ne_{(v)}=(2)(6)/[0+(1+1/11)]=11$. Using equation 10, we obtain the approximation $Ne_{(v)}=(6)[2/(2-1)]=12$, bias being due to the $\alpha=0$ assumption used in deriving equation 10.

Table 2 Number of male and female contributed gametes with N=6 and F=2 (u=0.33) and reciprocal crosses. Three seeds taken per plant[a]

Parent number	Contributed gametes		
	k_f	k_m	k
1	3	3	6
2	3	3	6
3	0	0	0
4	0	0	0
5	0	0	0
6	0	0	0
Totals	6	6	12

[a] Constant population size (n=N)

Table 3 Number of male and female contributed gametes with N=6, u=1 and chain crosses. One seed taken per plant[a]

Parent number	Contributed gametes		
	kf	km	k
1	1	1	2
2	1	1	2
3	1	1	2
4	1	1	2
5	1	1	2
6	1	1	2
Totals	6	6	12

[a] Constant population size (n=N)

$Ne_{(v)}$ under some mating systems for seed regeneration

In germplasm regeneration we tend to believe that different mating schemes will result in different effective population sizes. However, the following alternatives show that this is not necessarily true. In fact, when n and k are kept the same, $Ne_{(v)}$ will depend on how a given mating scheme affects the variance of the number of contributed gametes [V(k)].

Pollination systems most commonly used in accession regeneration are chain crosses and plant-to-plant crosses (with or without reciprocals). We present simple examples of these mating systems as related to male and female gametic control.

(1) Chain crosses. In this scheme, every plant is used as male and female (monoecious). It involves crossing plant 1 with plant 2, plant 2 with plant 3, and so on up to plant L crossed back with plant 1 (Table 3). This generates a chain of n half-sib related families; that is, ear j and j+1 are half-sibs, as are ears j+1 and j+2, and so on. From Ta-

Table 4 Number of male and female contributed gametes with N=6, u=1, and plant to plant crosses with reciprocals. One seed taken per plant[a]

Parent number	Contributed gametes		
	k_f	k_m	k
1	1	1	2
2	1	1	2
3	1	1	2
4	1	1	2
5	1	1	2
6	1	1	2
Totals	6	6	12

[a] Constant population size (n=N)

Table 5 Number of male and female contributed gametes with N=6, u=1, and plant to plant crosses without reciprocals. Two seeds taken per plant[a]

Parent number	Contributed gametes		
	k_f	k_m	k
♀ 1	2	0	2
♂ 2	0	2	2
♀ 3	2	0	2
♂ 4	0	2	2
♀ 5	2	0	2
♂ 6	0	2	2
Totals	6	6	12

[a] Constant population size (n=N)

ble 3, $k=k_f+k_m=6+6=12$, $\bar{k}_f=\bar{k}_m=1$, $\bar{k}=2$ and from the direct method for calculating the variance of contributed gametes, V(k)=0 [because $V(k_f)=V(k_m)=Cov(k_m,k_f)=0$]. Therefore, equation 1 becomes $Ne_{(v)}=(2)(6)/\ [0+(1+1/11)]=11$.

(2) Plant-to-plant crosses with reciprocals. This involves taking plants both as male and female (monoecious) (Table 4). From Table 4, $k=k_f+k_m=6+6=12$, $\bar{k}_f=\bar{k}_m=1$, $\bar{k}=2$ and V(k)=0 because $V(k_f)=V(k_m)=Cov(k_m,k_f)=0$. Thus, equation 1 becomes again $Ne_{(v)}=(2)(6)/[0+(1+1/11)]=11$.

(3) Plant-to-plant crosses without reciprocals. This involves using plants as females or as males but not both (Table 5). In this scheme, the accession to be regenerated is considered as being dioecious. Manipulating monoecious plants as dioecious plants requires appropriate expressions for $Ne_{(v)}$ for separate sexes (Crow and Denniston 1983). In this example, the initial population consists of $N_f=3$ and $N_m=3$ female and male parental plants, respectively, and with control of the number of gametes contributed, $Ne_{(v)}=8N_mN_f/(N_f+N_m)$ (Hallauer and Miranda 1981), such that $Ne_{(v)}=[8(3\times3)]/(3+3)]=12$, very close to $Ne_{(v)}$ values obtained for monoecious schemes (Tables 3 and 4).

For seed regeneration, all three pollination schemes provide a similar $Ne_{(v)}$, doubling the actual size of the initial population; therefore, only practical reasons are relevant for choosing a specific one.

Conclusions

This study presents procedures for computing the variance of the number of contributed gametes and equations for calculating the variance effective population size as related to genetic resources preservation. In seed regeneration and collection, drift occurs at two stages, namely: when sampling (random selection or random loss) parents from an initial base population and, subsequently, when sampling gametes contributing to the next generation. These are the basic assumptions underlying equation 3, which allow examining how the proportion of sampled parents, relative to the base population, affects the final $Ne_{(v)}$ value. The model describes situations in which the proportion of sampled parents is small, such as when sampling is done in a species' natural habitat or when the proportion of surviving seeds (plants) after loss of seed viability in an accession is small.

In germplasm collection, when the total number of seeds (n) is not too small, $Ne_{(v)}$ is dominated by the number of seed parents (F) (equations 11 and 12). Upon comparing those two expressions, we also observed that female gametic control only has a sizeable effect on increasing $Ne_{(v)}$ when n is small relative to F. For example, with n=F, taking equal numbers of seeds per parent (one single seed) will tend to increase $Ne_{(v)}$ by 25%, relative to random sampling of seeds. The appropriate strategy is to collect equal numbers of seeds from the largest possible number of plants (F).

In accession regeneration, when gametic control keeps population size constant (n=N), its positive effect is more evident when there is no great loss in germination rate (u not too small). For instance, if we compare N[(2u)/(2–u)] with Nu (Table 1) when u=1, we have 2N vs N; when u is small (e.g., u=0.1), we have (0.105) N vs (0.100) N, respectively. In this case (small u), random parental loss has an overwhelming effect on drift.

Field pollination procedures, such as plant-to-plant crosses (with or without reciprocals) and chain crossing, provide an $Ne_{(v)}$ that doubles the actual size of the initial population.

Appendix

Case 1

The female component of V(k) (equation 3) can be expressed as $V(k_{fs})=u[n(1/F)(1-1/F)+(1-u)(n^2/F^2)]$. After division by $\bar{k}=2n/N$ and simplifying, we obtain the term $(1/2)+[n(1-u)-1]/(2F)$. Similarly, the male component of V(k) can be reduced to $(1/2)+[n(1-v)-1]/(2\ M)$. The covariance component of equation 3, divided by $\bar{k}$, is [n(1–v)]/M. Therefore equation 2 can be written as

$$Ne_{(v)}=2n/\{1+(1/2)+[n(1-u)-1]/(2F)+(1/2)+[n(1-v)-1]/(2\ M)+[n(1-v)]/(M)\}.$$

After collecting similar terms and simplifying,

$$Ne_{(v)}=n/\{1+[n(1-u)-1]/(4F)+[3n(1-v)-1]/(4\ M)\}.$$

Case 2

Since the number of female gametes is controlled, $V(k_{fs})=0$ but $V(k_{ms})=n(1/M)(1-1/M)$. Since F=uN, the female component of equation 3, divided by $\bar{k}$, is [n(1-u)]/(2F). The male and covariance components of equation 3 are the same as in Case 1. Thus equation 2 can be written as

$$Ne_{(v)}=2n/\{1+[n(1-u)]/(2F)+(1/2)+[n(1-v)-1]/(2\ M)+[n(1-v)]/(M)\}.$$

After collecting similar terms and simplifying,

$$Ne_{(v)}=n/\{3/4+[n(1-u)]/(4F)+[3n(1-v)-1]/(4\ M)\}.$$

Case 3

In this case $V(k_{fs})=n(1/F)(1-1/F)$, but $V(k_{ms})=0$. The female and covariance components of equation 3 are the same as in Case 1, but the male component is $v(1-v)(n^2/M^2)=(1-v)(n^2/MN)$ (because M=vN), which divided by $\bar{k}$ reduces to [n(1–v)]/(2 M). Thus equation 2 is

$$Ne_{(v)}=2n/\{1+(1/2)+[n(1-u)-1]/(2F)+[n(1-v)]/(2\ M)+n(1-v)/(M)\}, \text{ or}$$

$$Ne_{(v)}=n/\{3/4+[n(1-u)-1]/(4F)+[3n(1-v)]/(4\ M)\}.$$

Case 4

Taking an equal number of seeds per plant and doing hand pollination produces $V(k_{fs})=0$ and $V(k_{ms})=0$. After division by $\bar{k}$, the female component is the same as in Case 2, [n(1–u)]/(2F), and the male component the same as in Case 3 [n(1–v)]/(2 M). There is no change in the covariance term and equation 2 can be written as

$$Ne_{(v)}=2n/\{1+[n(1-u)]/(2F)+[n(1-v)]/(2\ M)+[n(1-v)]/(M)\}, \text{ or}$$

$$Ne_{(v)}=n/\{1/2+[n(1-u)]/(4F)+[3n(1-v)]/(4\ M)\}.$$

References

Crossa J (1989) Methodologies for estimating the sample size required for genetic conservation of outbreeding crops. Theor Appl Genet 77:153–161

Crow JF, Kimura M (1970) An introduction to population genetic theory. Burgess Publishing, Minnesota

Crow JF, Denniston C (1983) Inbreeding and variance effective numbers. Evolution 42:482–495

Hallauer AR, Miranda Fo JB (1981) Quantitative genetics in maize breeding. Iowa State University Press, Ames, Iowa

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

Vencovsky R (1976) Extension of the concept of effective population size to populations under selection. Instituto de Genetica, ESALQ/USP, Piracicaba, SP, Brasil, Relatorio Cientifico 10: 223–228

Vencovsky R (1978) Effective size of monoecious populations submitted to artificial selection. Brasil J Genet I:181–191

Vencovsky R (1987) Tamanho efectivo populacional na coleta e preservacao de germoplasma de especies alogamas. ESALQ, USP, Piracicaba, Sao Paulo, Brasil, Instituto de Pesquisas e Estudos Florestais 35:79–84

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

Wright S (1969) Evolution and the genetics of populations, vol 2. The University of Chicago Press. Chicago