Statistical Genetics and Simulation Models in Genetic Resource Conservation and Regeneration

Jiankang Wang,* Jose Crossa, Maarten van Ginkel, and Suketoshi Taba

ABSTRACT

A major objective when regenerating germplasm collections is to maintain as many genes (or alleles) as possible. Two stages, sampling and propagating, can be identified in any regeneration cycle. A rare allele can be lost at either stage, when a cross-pollinated population is being regenerated. The objectives of this study were (i) to derive the theoretical probability that one or a few rare alleles will be retained after one regeneration cycle and (ii) to determine, through computer simulation, the average number of alleles lost after a certain number of regeneration cycles. The effects of four mating methods commonly used on cross-pollinated crops (i.e., random pollen pollination, chain cross, paired cross, and self-pollination) were compared. If more than five seeds are harvested from each regenerated plant, the probability that an allele will be lost during Stage 2 can be ignored. However, harvesting and maintaining more seeds than are required for subsequent regeneration cycles will have some negative effects on the retention of genetic diversity. Self-pollination is the best mating strategy for retaining rare alleles only if the germination rate is 100% and all the reserved seeds are regenerated. It can be used for species where inbreeding depression is not serious and the seeds of individual plants from an accession can be stored separately. Paired cross without reciprocal is recommended for regenerating cross-pollinated species, as it results in the same genetic consequence as paired cross with reciprocal or chain cross, but requires only half the number of labor-intensive crosses.

Genetic resource conservation programs and genebanks store thousands of accessions of different cultivated species and their wild and weedy relatives. Sackville Hamilton et al. (2002) addressed many issues related to accession management, including cost efficiency. It is difficult to provide simple, efficient sampling schemes and optimal sample sizes for the maintenance of all species because of the great complexity of the genetic structure of most plant populations (i.e., species may be cross-pollinated or self-pollinated, or may have mixed self- and random-mating reproductive systems, or different levels of inbreeding) and the numerous ways genetic resources may be used (Engels and Ramanatha Rao, 1998; Crossa and Vencovsky, 1999). Random changes in allele frequency caused by sampling error in small populations (random genetic drift) lead to fixation and loss of alleles, and reduce the proportion of heterozygous individuals in the population (Wright, 1931; Crow and Kimura, 1970). When random changes in allele frequency occur in finite populations, they affect the genetic diversity of the population.

One of the main objectives of genebanks is to retain alleles of landraces and other populations that may have been collected or obtained at great expense and effort. It is important that both globally- and regionally-rare alleles be retained and made available for crop improvement efforts. Genebanks have a large responsibility in this regard: if only the most common alleles or genes are present in samples distributed to cooperators, genetic advances in crop improvement based on these supposed sources of genetic diversity may come to a halt. With climate change upon us, genebank collections increasingly need to be fully exploited to maintain production in the face of a diminishing natural resource base.

The genetic constitution of a collection changes during collecting, storing, regenerating, distributing, and restocking processes. These changes affect the genetic integrity of the reference population in a number of ways. First, when collecting germplasm in the field, sampling strategies influence the size of the population and can result in bottlenecks when the sample is much smaller than the original population. Second, when an accession is stored, different seed survival rates and the accumulation of mutations may affect its genetic integrity. Third, when seed of an accession is regenerated, changes in the genetic constitution of the collection can occur because of sampling or other factors such as differential germination or fecundity. Fourth, when accessions are regenerated in the field, insects, diseases, and other environmental factors can affect plant stand, thereby limiting the plants' gametic contribution (offspring) to the next generation. Fifth, the seed lot that is finally sent to cooperators is itself a sample of either the seed lot collected originally or a regenerated seed lot, with reduced probabilities of rare alleles being present.

The sample size for genetic resource conservation can be studied from different perspectives. From a statistical genetic perspective, a realistic model should take into account drift caused by sampling parents from the original collection and drift caused by sampling gametes from those parents (Crossa and Vencovsky, 1994). Increasing the number of randomly selected parents may reduce drift due to sampling. Drift caused by gamete sampling may be reduced by controlling male gametes through hand-pollination and by controlling female gametes by taking an equal number of seeds per pollinated parent (Crossa and Vencovsky, 1994, 1997, and 1999; Vencovsky and Crossa, 1999).

From a probabilistic perspective, once a collection is stored in the genebank, it becomes the reference population, which can be considered finite or infinite, depending on its actual size. At this point, two main sam-

International Maize and Wheat Improvement center (CIMMYT), Apdo. Postal 6-641, 06600 Mexico, D.F., Mexico. Received 9 Jan. 2004. *Corresponding author (jkwang@cgiar.org).

Published in Crop Sci. 44:2246–2253 (2004).

677 S. Segoe Rd., Madison, WI 53711 USA

Abbreviations: CIMMYT, International Maize and Wheat Improvement Center; GRSimulator, germplasm regeneration simulator.

pling stages that may change the genetic constitution of a collection can be identified in the regeneration process. Stage 1 accounts for changes in genetic constitution during seed storage in the genebank that are mainly due to different seed survival and mutation rates, reduced germination, and the actual sample size taken from the original collection to be planted in the field. Stage 2 accounts for changes due to insects, diseases, and environmental factors that may affect plant stand in the field, and the mating design used during pollination. A rare allele present in the original collection to be regenerated could be lost at either stage. Crossa and Vencovsky (1999) concluded that general recommendations and/or guidelines concerning sample size should be given within ranges (or intervals), depending on model assumptions and biological considerations such as number of loci and number of alleles.

The main objective of Stage 1 is to find a sample size at which the probability of detecting at least one allele of each allele class is greater than a specific quantity. Gregorius (1980) developed a general formula for determining the probability that not all alleles present in a population will be represented in a sample of diploid individuals. Crossa (1989) gave the probability that each of the various allele classes at a locus will be represented at least once in a sample, and Crossa et al. (1993) extended the expression to include several independent loci. Allele frequencies are unknown in real situations. Therefore, Crossa et al. (1993) made the assumption that several alleles occur at an identical low frequency and are inherited independently. However, in reality, the assumption of locus independence is improbable, and changes in genetic constitution may cause nonrandom associations between linked loci, given the finite number of chromosomes within a genome. For a large population under random mating, linkage equilibrium can be assumed for all possible pairs of loci, with the result that a similar number of coupling and repulsion combinations occur. Evidently this does not hold for self-pollinated species or species with mixed self- and random-mating systems of reproduction. It should be pointed out that computations for determining the sample size at which there is a certain probability of retaining at least one allele from each allele class at each locus offer approximate guidelines on the range of sample sizes that, as previously mentioned, are greatly dependent on the frequency of rare alleles.

The range of sample sizes discussed by Crossa et al. (1993) is based solely on probability models and does not consider the genetic structure of a population or specify how well a particular sample represents the reference population in terms of genetic parameters such as variance of allele frequency, inbreeding, random genetic drift due to sampling error, and genetic linkage. Furthermore, probability models, by themselves, do not designate appropriate mating and reproductive systems such as cross-pollination (panmixia), mixed self and random mating, full-sibs (chain crosses and paired plant crosses), half-sibs (random pollen pollination), and self-fertilization, that could, under specific circumstances, maximize the genetic representativeness of a sample. Furthermore, the mathematical equations developed by Crossa et al. (1993) only considered the changes in genetic constitution occurring in the first stage of the regeneration process (Stage 1). However, a complete regeneration cycle comprises sampling Stage 1 and propagating Stage 2. Furthermore, the mathematical model developed by Crossa et al. (1993) is not able to consider more than one regeneration cycle.

The first objective of this research was to extend the statistical and genetic equations developed by Crossa et al. (1993) to account for factors affecting the genetic constitution of a collection during the two stages of the regeneration process, so that the resulting formula can be used to calculate the sample size required to achieve a certain probability that one or more rare alleles will be retained following a regeneration cycle. The second objective of this study was to use computer simulation to investigate the probability that rare alleles will be retained and to determine the average number of alleles lost after several regeneration cycles because of different numbers of seeds being taken from each pollinated plant and the use of different pollination methods [e.g., random pollen, chain cross, paired cross (with or without reciprocal cross), and self-pollination].

MATERIALS AND METHODS

Theory

As noted earlier, two stages can be identified in the regeneration process of a population during germplasm conservation, sampling and propagation. An allele with a frequency of no more than 0.05 is normally viewed as a rare allele and can be lost easily during regeneration. Assume an initial population in Hardy–Weinberg equilibrium, with two alleles, A and a, at one locus, where a is a rare allele with a low frequency q (Table 1). Allele a is considered to be retained at Stage 1 if at least one individual with genotype Aa or aa exists in a sample consisting of n individuals. For a cross-pollinated species, the probability that allele a will be retained at Stage 1 is $P\{\text{allele } a \text{ retained at Stage 1}\} = 1 - (1 - q)^{2n}$, and for a self-pollinated species, $P\{\text{allele } a \text{ retained at Stage 1}\} = 1 - (1 - q)^{n}$. For a cross-pollinated species and for k rare allele classes (which can be located at the same locus or at different loci), the probability that all alleles will be retained at Stage 1 is

$$P\{k \text{ alleles retained at Stage 1}\} = \prod_{i=1}^{k}[1 - (1 - q_i)^{2n}], \quad [1]$$

where q_i is the frequency of the ith allele and n is the sample size. Equation [1] was used by Crossa (1989) and Crossa et al. (1993) to determine the probability of retaining rare alleles.

For a self-pollinated species, Eq. [1] becomes $P\{k \text{ alleles retained at Stage 1}\} = \prod_{i=1}^{k}[1 - (1 - q_i)^{n}]$, and an allele will not be lost if it was retained at the sampling stage. Clearly, the

Table 1. Frequency and sample size for the three genotypes in a sample from a cross-pollinated population in Hardy–Weinberg equilibrium.

Genotype	AA	Aa	aa	Sum
Frequency	$(1 - q)^2$	$2(1 - q)q$	q^2	1
Sample size	n_1	n_2	n_3	n

Table 2. Frequency of each genotype in the half-sib progeny for a sample from an infinite cross-pollinated population where there is no genotype *aa*.

Genotype of parental plant	Sample size	Genotype of the half-sib progeny		
		AA	Aa	*aa*
AA	$n - n_2$	$(2n - n_2)/2n$	$n_2/2n$	0
Aa	n_2	$(2n - n_2)/4n$	1/2	$n_2/4n$

situation is much more complex and urgent for cross-pollinated species, and therefore the discussion that follows centers mainly on these crops, which include major food crops such as maize.

Suppose n_1, n_2, and n_3 are the numbers of individuals in a sample that have genotypes *AA*, *Aa*, and *aa*, respectively (Table 1), that random pollen is used for pollination, and m seeds are harvested from each individual plant at Stage 2. The event that allele *a* is retained at Stage 1 is the union of three exclusive events, i.e., {allele *a* retained at Stage 1} = {$n_2 > 0$ or $n_3 > 0$} = {$n_2 = 0$ and $n_3 > 0$} $\cup$ {$n_2 > 0$ and $n_3 = 0$} $\cup$ {$n_2 > 0$ and $n_3 > 0$}. In an event {$n_2 = 0$ and $n_3 > 0$} or {$n_2 > 0$ and $n_3 > 0$}, where at least one individual with genotype *aa* is retained, the probability that allele *a* will be lost at Stage 2 is 0, no matter how few the seeds (minimum 1) harvested from each pollinated plant. Thus, only for the event {$n_2 > 0$ and $n_3 = 0$} is there some possibility of losing allele *a* at Stage 2. In this case, there are $n - n_2$ *AA* families and n_2 *Aa* families at Stage 2 (Table 2). If all m individuals harvested from each of the n families have genotype *AA*, then allele *a* is lost. Therefore, the probability that allele *a* will be lost at Stage 2 for any possible event among {$n_2 > 0$ and $n_3 = 0$} is

$$P\{\text{allele } a \text{ lost at Stage 2} \mid n_1 = n - n_2 - n_3, n_2 > 0, n_3 = 0\} = \left(\frac{1}{2}\right)^{mn_2}\left(1 - \frac{n_2}{2n}\right)^{mn}.$$

Therefore, the probability that allele *a* will be retained at Stage 1 but lost at Stage 2 is

$$P\{\text{allele } a \text{ retained at Stage 1 but lost at Stage 2}\} = \sum_{n_2=1}^{n}\left(\frac{1}{2}\right)^{mn_2}\left(1 - \frac{n_2}{2n}\right)^{mn} P\{n_1 = n - n_2 - n_3, n_2, n_3 = 0\},$$

where $P\{n_1, n_2, n_3\} = \frac{n!}{n_1!n_2!n_3!}[(1 - q)^2]^{n_1}[2(1 - q)q]^{n_2}[q^2]^{n_3}$. The probability that allele *a* will be retained after the two stages is equal to the probability that allele *a* will be retained at Stage 1 but lost at Stage 2; therefore,

$$P\{\text{allele } a \text{ retained after one regeneration cycle}\} = 1 - (1 - q)^{2n} - \sum_{n_2=1}^{n}\left(\frac{1}{2}\right)^{(m-1)n_2}\left(1 - \frac{n_2}{2n}\right)^{mn}\frac{n!}{(n - n_2)!n_2!}(1 - q)^{2n-n_2}q^{n_2}.$$

When k alleles are included, the probability that all alleles will be retained after regeneration is

$$P\{k \text{ alleles retained after one regeneration cycle}\} = \prod_{i=1}^{k}\left[1 - (1 - q_i)^{2n} - \sum_{n_2=1}^{n}\left(\frac{1}{2}\right)^{(m-1)n_2}\left(1 - \frac{n_2}{2n}\right)^{mn}\frac{n!}{(n - n_2)!n_2!}(1 - q_i)^{2n-n_2}q_i^{n_2}\right] \quad [2]$$

where q_i is the frequency of the ith allele and n is the sample size.

In practice, it would also be useful to know whether a rare allele has been lost after more than one regeneration cycle. This is important for genebanks around the world, particularly those that do not have the appropriate equipment for long-term seed storage and thus need to regenerate their collections at a regular frequency. However, to expand Eq. [2] to include more than one regeneration cycle is mathematically intractable and can only be done by computer simulation.

Computer Simulation Program (GRSimulator)

A computer program called GRSimulator (Germplasm Regeneration Simulator, freely available from http://www.cimmyt.org/tools/grsimulator.htm; verified 3 June 2004) was developed to determine the probability of retaining alleles of interest and calculate the average number of alleles lost after several regeneration cycles. The parameters included in a regeneration strategy are sample size (number of plants to be regenerated) at Stage 1, mating method, and the number of seeds taken from each plant at Stage 2. The mating methods available with GRSimulator are panmixia, random pollen pollination (bulk of pollen from all n regenerated plants applied to stigmas of the same plants), chain cross (Plant 1 crossed to Plant 2, Plant 2 crossed to Plant 3, Plant 3 crossed to Plant 4, etc.), paired without reciprocal cross (Plant 1 crossed to Plant 2, Plant 3 crossed to Plant 4, etc.), paired with reciprocal cross (Plant 1 crossed to Plant 2, Plant 2 crossed to Plant 1, Plant 3 crossed to Plant 4, Plant 4 crossed to Plant 3, etc.), and self-pollination (Plant 1 crossed to Plant 1, Plant 2 crossed to Plant 2, etc.). The parameters of a regeneration strategy may have different values in different cycles. The computer simulation program was written in Fortran 90/95 and can run under Microsoft Windows.

RESULTS AND DISCUSSION

Sample Size Required to Retain Rare Alleles after One Regeneration Cycle

Equation [1] can be used to find the sample size required for achieving a certain probability that one or several rare alleles will be retained at sampling Stage 1 (Table 3). It can also be used to calculate the probability that some rare alleles will be included in a sample of seeds taken from regenerated accessions with the purpose of studying and distributing genetic diversity. Numbers in Table 3 are half of those presented in Crossa et al. (1993) because a cross-pollinated species was considered in this instance. For an allele with a frequency of 0.03 or higher, a sample size of 38 has a 0.90 probability of retaining that allele in the sample; a sample size of 49 individuals has a 0.95 probability; and a sample size of 86 individuals has a 0.99 probability. For an allele with a frequency in the population of only 0.01, a sample size of 115 is required to retain the allele with a 0.90 probability; a sample size of 149 is required to achieve a 0.95 probability; and a sample size of 229 will achieve a 0.99 probability. The lower the frequency of rare alleles, the larger the sample size required for achieving a certain probability of retaining the rare alleles. Lawrence et al. (1995) suggested that a sample size appropriate for regeneration should comprise 172 individuals. However, results of Table 3 concerning sample Stage 1 indicate that 172 individuals will retain no more than two alleles at a frequency of 0.01 with 0.90 and 0.95 probabilities.

Table 3. Sample size required to achieve a 0.90, 0.95, or 0.99 probability that one or several rare alleles will be retained at the sampling stage.

Number of alleles†	$P = 0.90$			$P = 0.95$			$P = 0.99$		
	$q = 0.05$	$q = 0.03$	$q = 0.01$	$q = 0.05$	$q = 0.03$	$q = 0.01$	$q = 0.05$	$q = 0.03$	$q = 0.01$
1	22	38	115	29	49	149	45	86	229
2	29	49	148	36	60	183	52	102	263
5	38	64	193	45	75	228	61	113	309
10	44	75	227	51	87	262	67	120	343
15	48	81	247	55	93	286	71	125	364
20	51	86	261	58	98	297	74	229	378
100	67	113	341	74	124	377	90	151	458

† When more alleles than one are included, all of them have the same frequency.

Equation [2] can be used to determine the probability that one or more rare alleles will be retained in the sample when one full regeneration cycle is completed. For comparison, Table 4 shows the results from both equations and the GRSimulator for three sample sizes (50, 100, and 200), six numbers of seeds harvested from each plant (1, 2, 3, 4, 5, and 10), and five rare alleles at equal or different frequencies. There is almost the same probability of retaining all five rare alleles when more than five seeds are taken from each pollinated plant. Therefore, when the number of seeds harvested from each individual plant exceeds five, the probability that an allele will be retained at Stage 1 but lost at Stage 2 can be ignored (Table 4). In cases where only one or a few seeds are taken from each plant, simulation results are closer to those of Eq. [2] than to those of Eq. [1], which indicates that the probability that rare alleles will be lost at Stage 2 cannot be ignored when only a few seeds are harvested from each pollinated plant.

As expected, when several rare alleles are considered simultaneously, the probability that all alleles will be retained after one regeneration cycle depends mainly on the frequency of each allele in the original collection. Whether these alleles are located at the same locus (Case 1: one locus with six alleles, five of which are rare, and one common) or at different loci (Case 2: five loci, two alleles per locus, and there is one rare and one common allele at each locus) does not affect the final probability of retaining the alleles (Table 4). And, as long as the initial germplasm collection is in Hardy–Weinberg equilibrium, the presence or absence of linkage does not affect the probability either. Additional simulations confirm these results (results not shown).

Regardless of the number of seeds taken from each pollinated plant (except when only one seed is taken), Table 4 clearly shows that a sample size of around 200 at Stage 1 will retain the five alleles present at a low frequency (0.01) with a probability of more than 0.90. When the sample size is 100, the probability of retaining all rare alleles drops to between 0.40 and 0.50 for any number of seeds taken from each pollinated plant (except when one seed is taken). A sample size of 50 had a very low probability (around 0.10) of retaining all five alleles, but that probability increases as allele frequencies increase, as is to be expected.

Probability That Rare Alleles Will Be Retained after Several Regeneration Cycles

From Eq. [2], it can be concluded that the higher the number of seeds taken from each plant, the lower the probability of losing an allele at Stage 2. However, if more seeds are taken during an earlier regeneration cycle than can be grown in subsequent cycles, this will

Table 4. Probabilities that five rare alleles will be retained after one regeneration cycle using Eq. [1] and [2], and 1000 simulation runs.

Sample size	Number of seeds per plant harvested	Five rare alleles with the same frequency 0.01				Five rare alleles with frequencies 0.01, 0.02, 0.03, 0.04, and 0.05			
		Theoretical probability		Probability from GRSimulator†		Theoretical probability		Probability from GRSimulator†	
		Eq. [1]	Eq. [2]	Case 1‡	Case 2§	Eq. [1]	Eq. [2]	Case 1‡	Case 2§
50	1	0.102	0.032	0.024	0.036	0.512	0.306	0.294	0.288
	2		0.077	0.076	0.082		0.455	0.450	0.469
	3		0.094	0.099	0.095		0.495	0.486	0.484
	4		0.100	0.085	0.091		0.507	0.492	0.490
	5		0.102	0.085	0.107		0.510	0.544	0.524
	10		0.102	0.096	0.104		0.512	0.520	0.508
100	1	0.487	0.243	0.256	0.248	0.849	0.695	0.717	0.707
	2		0.415	0.393	0.419		0.814	0.824	0.829
	3		0.466	0.433	0.461		0.839	0.840	0.841
	4		0.481	0.459	0.481		0.846	0.847	0.838
	5		0.485	0.500	0.483		0.848	0.850	0.837
	10		0.487	0.490	0.520		0.849	0.852	0.852
200	1	0.913	0.731	0.738	0.736	0.982	0.936	0.943	0.938
	2		0.877	0.886	0.878		0.973	0.970	0.981
	3		0.904	0.892	0.905		0.980	0.976	0.980
	4		0.911	0.923	0.901		0.981	0.981	0.978
	5		0.913	0.927	0.924		0.982	0.981	0.986
	10		0.913	0.915	0.921		0.982	0.984	0.984

† The initial population used in simulation is of an infinite size.
‡ Case 1: one locus with six alleles, of which five are rare.
§ Case 2: five loci, each with two alleles: one rare and one common.

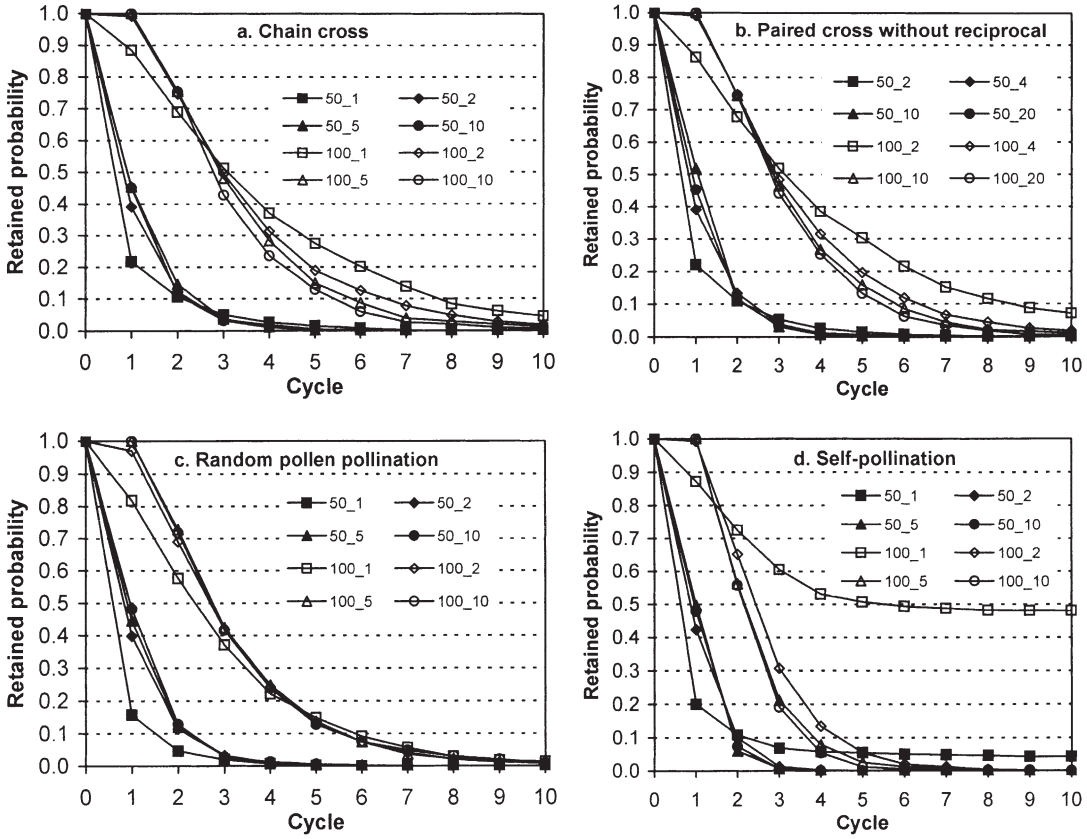

Fig. 1. Probability that 10 rare alleles will be retained after 10 regeneration cycles using four mating systems: chain cross (a), paired cross without reciprocal (b), random pollen pollination (c), and self-pollination (d). The initial population consists of 100 individuals, in which two alleles have frequency 0.01, two have frequency 0.02, two have frequency 0.03, two have frequency 0.04, and two have frequency 0.05. The strategies are denoted by the sample size (50 or 100 plants) and the number of seeds taken from each plant (1, 2, 5, or 10).

increase the probability of losing alleles in the long run. In Fig. 1, two sample sizes (50 and 100) and four different numbers of seeds (1, 2, 5, and 10) taken from each pollinated plant are considered. It was assumed the initial population consisted of 100 individual seeds with 10 rare alleles (at 10 different loci; two alleles with frequency 0.01, two with frequency 0.02, two with frequency 0.03, two with frequency 0.04, and two with frequency 0.05) and four mating systems (chain cross, paired cross, random pollination, and self-pollination) were applied.

For chain cross, paired cross, and random pollen pollination, the probability of retaining all 10 rare alleles declined to below 0.10 after three regeneration cycles when the sample size was 50, and close to 0.50 when sample size was 100 (Fig. 1). For sample size 100 and all mating systems except random pollination, only in the first two to three regeneration cycles did selecting just one seed per pollinated plant have the lowest probability of retaining the rare alleles. However, this trend was completely reversed in subsequent regeneration cycles. The number of seeds taken from each plant had an intermediate effect in paired crossing, chain crossing, and random pollen pollination (Fig. 1a, 1b, and 1c). It had a greater effect in self-pollination, where the practice of taking only one seed per pollinated plant seems to significantly increase the probability of retaining rare alleles throughout the 10 regeneration cycles for both samples sizes (Fig. 1d).

The effect of seed number can also be seen from the average number of alleles lost after several regeneration cycles (Fig. 2). After 10 regeneration cycles, self-pollination with one seed taken from each pollinated plant loses, on average, 2.37 alleles when the sample size is 50 and less than one allele when the sample size is 100 (Fig. 2.1a and 2.2a, respectively). But self-pollination becomes the worst mating system if more than one seed per pollinated plant is taken and only a portion of these can be planted during the next regeneration cycle, because population size needs to be kept nearly constant (Fig. 2.1b, 2.2b, 2.1c, and 2.2c). However, self-pollination with one seed harvested from each regenerated plant is the only strategy that can keep more alleles from being lost after several cycles (Fig. 2). This can be viewed as an extreme case of a strategy that splits one heterogeneous accession to create several more homogeneous accessions (Sackville Hamilton et al., 2002). Self-pollination could be adopted where inbreeding depression is not serious and seeds of individual plants of an accession can be stored separately. By using this strategy, more than one seed should be harvested from each regenerated plant because of incomplete germination. The number of seeds required to form at least one plant in the next regeneration cycle can be estimated

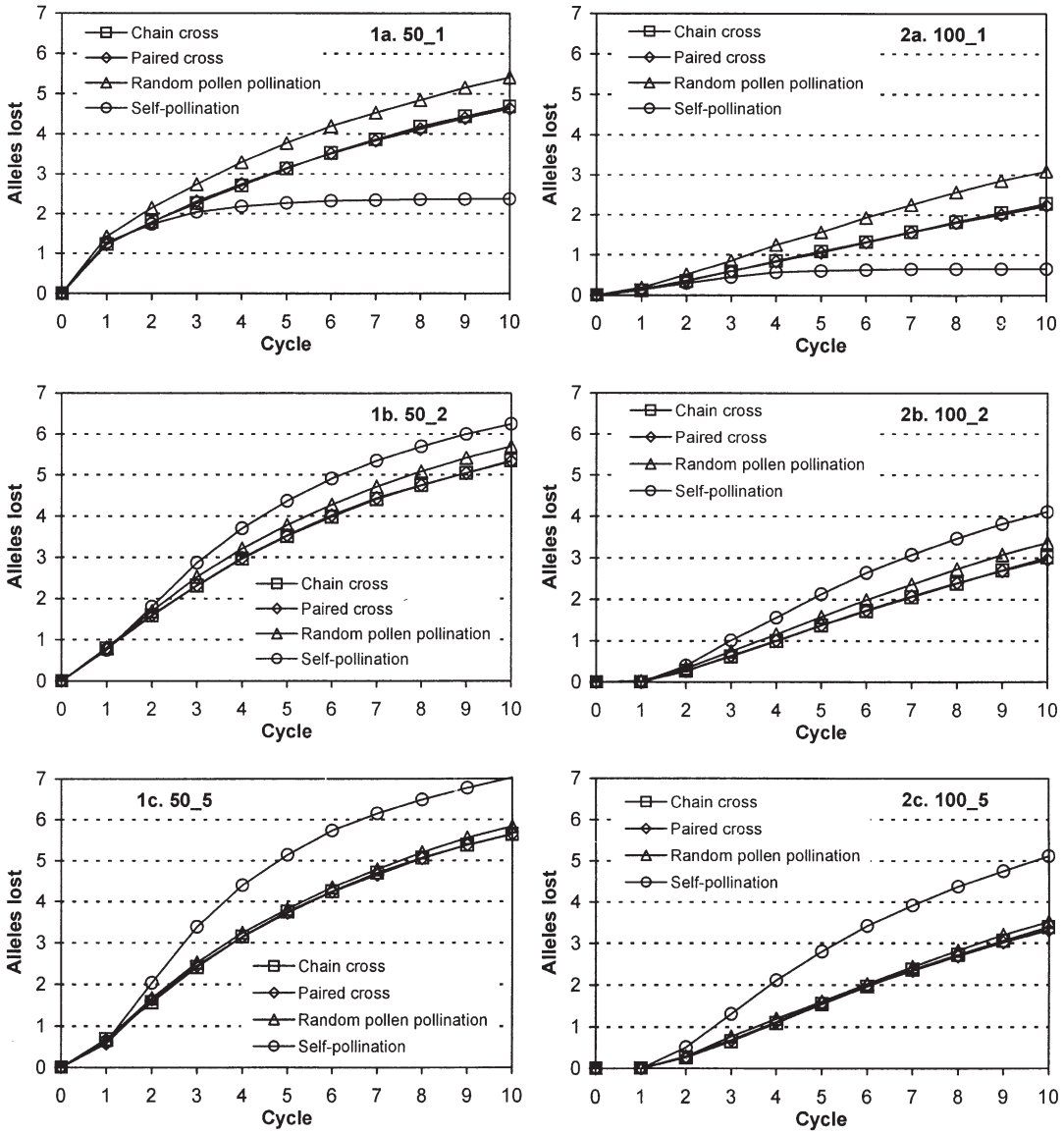

Fig. 2. Number of alleles lost after 10 regeneration cycles. Ten rare alleles were included in the initial population consisting of 100 individuals, in which two alleles have frequency 0.01, two have frequency 0.02, two have frequency 0.03, two have frequency 0.04, and two have frequency 0.05. The four mating systems are random pollination, chain cross, paired cross (with reciprocal cross), and self-pollination. The two sample sizes are 50 and 100, and the three numbers of seeds taken from each plant are 1, 2, and 5, which are represented by 50_1 (1a), 50_2 (1b), 50_5 (1c), 100_1 (2a), 100_2 (2b), 100_5 (2c).

from the formula $m = \frac{\log(1 - P)}{\log(1 - r)}$, where P is the probability, and r is the germination rate. Values for different germination rates and probabilities are presented in Table 5. In the next regeneration cycle, all the stored seeds from each plant should be sown together. No matter how many plants germinate, only one will be kept. This is similar to the strategy where one seed is harvested and germination rate is 100% (assuming that germination rate has a neutral effect on the rare alleles to be retained), and therefore more alleles will not be lost after four to five regeneration cycles (Fig. 1 and 2).

Chain cross, paired cross, and random pollen pollination lose similar numbers of alleles during regeneration cycles in which five or more seeds are harvested from each regenerated plant. Random pollen pollination loses more alleles than chain cross and paired cross when one or two seeds are harvested (Fig. 2.1a, 2.2a, 2.1b, and 2.2b). Chain cross and paired cross achieve the same result for any number of harvested seeds (Fig. 2). The paired cross without reciprocal has the same genetic consequence as the paired cross with recip-

Table 5. Number of seeds that needs to be harvested from a regenerated plant for at least one seed to germinate in the next regeneration cycle.

	Probability		
Germination rate	0.999	0.99	0.95
0.9	3	2	1
0.8	4	3	2
0.7	6	4	2
0.6	8	5	3
0.5	10	7	4

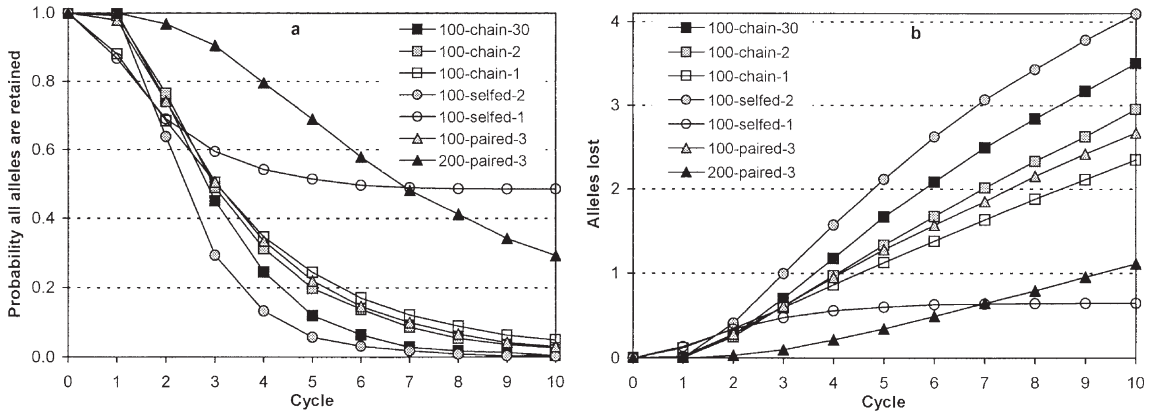

Fig. 3. Probability of retaining all 10 rare alleles, two each with frequencies 0.01, 0.02, 0.03, 0.04 and 0.05 (a), and the number of alleles lost (b). The initial population has 100 individuals. For paired cross, there is no reciprocal cross. Each strategy is denoted by the sample size (100 or 200), mating method (chain cross, self-pollination, and paired cross without reciprocal cross), and the number of seeds to be harvested from each cross (1, 2, 3, or 30).

rocal, where twice as many seeds need to be harvested. However, for the paired plant cross without reciprocal cross, the number of crosses to be made is just half that needed for the paired with reciprocal cross or chain cross.

For a fixed, small collection size (n) produced by the bottleneck effect of a cross-pollinated population species, all n plants should be regenerated in the first cycle (Crossa et al., 1994). If only the same number of plants can be regenerated because of cost constraints in the following regeneration cycle, and germination is not a problem, the best long-term strategy may be to harvest only one seed from each plant during the regeneration process. For a larger collection, the regeneration strategy would be to take two or more seeds from each plant in the first regeneration cycle. Then all the harvested seeds should be regenerated, and only one seed should be taken from each plant in the following cycles. This will greatly decrease the probability of losing alleles. Obviously, the fewer regeneration cycles that are performed, the lower the probability of losing rare alleles. Results obtained from this simulation study confirm results previously reported by Crossa et al. (1994), Crossa and Vencovsky (1997), and Vencovsky and Crossa (1999).

Simulating Maize Germplasm Regeneration

In maize genebanks collections, accessions of this cross-pollinated species are usually regenerated by taking 100 individual seeds and growing them in the field. The chain cross is used, and all seeds from each ear are taken. To represent this regeneration strategy, we used the abbreviation 100-chain-30, where the number 100 indicates the original number of seeds planted and 30 the number of seeds harvested from each cross. Other options include 100-chain-2, 100-chain-1, 100-selfed-2, 100-selfed-1, 100-paired-3, and 200-paired-3. Here, *selfed* means self-pollination was applied, and *paired* means paired without reciprocal cross. In the case of 200-paired-3, because of the initial population size of 100, six seeds are harvested from each of the 50 paired crosses in the first regeneration cycle, so that 300 seeds will be harvested and kept in reserve to establish 200 plants for the second cycle.

Suppose there are 10 rare alleles in the initial collection of 100 individuals, two alleles with frequency 0.01, two with 0.02, two with 0.03, two with 0.04, and two with 0.05. For the strategy 100-chain-30, the probability of retaining all alleles is close to 1.00 after the first regeneration cycle, 0.12 after five regeneration cycles, and almost zero after 10 regeneration cycles (Fig. 3a). The average number of alleles lost is 0.00 after the first regeneration cycle, 1.67 after five regeneration cycles, and 3.50 after 10 regeneration cycles (Fig. 3b). For the common regeneration strategy of planting 100 plants and taking 30 seeds from each plant, simulation results indicated that by harvesting two seeds (instead of 30) from each plant, the long-term probability of retaining all five rare alleles (after 10 regeneration cycles) will increase from 0.01 to 0.03 (Fig. 3a), and the alleles lost will decrease from to 3.50 to 2.95 (Fig. 3b).

When comparing strategy 100-selfed-1 with 100-chain-1, assuming no germination problem, the strategy 100-selfed-1 can dramatically increase the probability of retaining all 10 alleles after 10 regeneration cycles from 0.05 (100-chain-1) to 0.49 (100-selfed-1), and decrease the loss of alleles from 2.35 (100-chain-1) to 0.65 (100-selfed-1). If there is a germination problem, at least two seeds must be harvested to maintain a constant population. In this case, self-pollination becomes the worst regeneration strategy for retaining alleles.

The strategy 100-paired-3 (without reciprocal cross) results in a genetic consequence similar to that of strategy 100-chain-2. However, the number of crosses from paired cross is only half the number from chain cross. In this case, a larger population (i.e., strategy 200-paired-3) can be regenerated without excessively increasing expenses. This strategy (200-paired-3) can increase the retained probability to 0.29 and decreases the number of lost alleles to 1.11 after 10 cycles of regeneration.

In this study, we have assessed the probability of losing rare alleles in the regeneration process. If additional sampling occurs after Stage 2 of any regeneration

cycle, that sampling should be considered Stage 1 of the next regeneration cycle. As previously mentioned, one of the objectives of genebanks is to distribute accessions and collections to the international community that will use this genetic diversity in breeding; the sample should thus represent, as much as possible, the genetic diversity present in the basic reference population. Table 3 provides an assessment of the probability of losing alleles for sampling while distributing accessions.

Recommendations

In summary, if the collection to be regenerated has size n, we recommend that $0.5n$ paired crosses be made, and three or four (depending on germination rate) seeds from each cross be harvested in bulk as the collection for the next regeneration cycle. The remaining seeds can be used for distribution. In the first cycle, more seeds should be harvested (e.g., four or five) to quickly eliminate the bottleneck effect, provided a larger regeneration population is possible the next cycle. All the reserved seeds should be regenerated in the following cycles to minimize resampling random drift. The regenerated population should be as large as possible, and the number of regeneration cycles should be minimized. Self-pollination should be applied where inbreeding depression is not serious, and the seeds of individual plants of an accession can be stored separately. This is the only mating method that can keep more alleles from being lost after several regeneration cycles.

ACKNOWLEDGMENTS

This research was funded in part by the Grains Research and Development Corporation (GRDC), Australia under the grant UQ123 (CIM8).

REFERENCES

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

Crossa, J., C.M. Hernandez, P. Bretting, S.A. Eberhart, and S. Taba. 1993. Statistical genetic considerations for maintaining germplasm collections. Theor. Appl. Genet. 86:673–678.

Crossa, J., S. Taba, S.A. Eberhart, P. Bretting, and R. Vencovsky. 1994. Practical considerations for maintaining germplasm in maize. Theor. Appl. Genet. 89:89–95.

Crossa, J., and R. Vencovsky. 1994. Implication of the variance effective population size on the genetic conservation of monoecious species. Theor. Appl. Genet. 89:936–942.

Crossa, J., and R. Vencovsky. 1997. Variance effective population size for two-stage sampling of monoecious species. Crop Sci. 37:14–26.

Crossa, J., and R. Vencovsky. 1999. Sample size and variance effective population size for genetic resources conservation. Plant Genet. Resour. Newsl. 119:15–25.

Crow, J.F., and M. Kimura. 1970. An introduction to population genetics theory. Burgess Publishing, Minneapolis, MN.

Engels, J.M.M., and R. Ramanatha Rao (ed.) 1998. Regeneration of seed crops and their wild relatives. Proceedings of a consultation meeting, ICRISAT, Hydrabad, India. 4–7 Dec. 1995. International Plant Genetic Resources Institute, Rome, Italy.

Gregorius, H.-R. 1980. The probability of losing an allele when diploid genotypes are sampled. Biometrics 36:643–652.

Lawrence, M.J., D.F. Marshall, and P. Davies. 1995. Genetics of genetics conservation. I. Sample size when collecting germplasm. Euphytica 84:89–99.

Sackville Hamilton, N.R., J.M.M. Engels. Th.J.L. van Hintum, B. Koo, and M. Smale. 2002. Accession management. Combining or splitting accessions as a tool to improve germplasm management efficiency. IPGRI Tech. Bull. 5. International Plant Genetic Resources Institute, Rome, Italy.

Vencovsky, R., and J. Crossa. 1999. Variance effective population size under mixed self and random mating with applications to genetic conservation of species. Crop Sci. 39:1282–1294.

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