PLANT GENETIC RESOURCES

Effective Population Size during Grass Germplasm Seed Regeneration

R. C. Johnson,* V. L. Bradley, and M. A. Evans

ABSTRACT

Effective population size \( (N_e) \) is the key parameter for predicting genetic drift associated with germplasm regeneration. A major factor reducing \( N_e \) below the census population size \( (N_c) \) is variation in seed production among plants in a given population. The objectives of this study were to estimate \( N_e/N_c \) associated with variation in seed production in three model wind pollinated, perennial grass species [Lolium perenne L., Festuca pratensis Huds., and Pseudoroegneria spicata (Pursh) Á. Löve] and to recommend cost effective sampling methodology to maximize \( N_e/N_c \) during seed regeneration. Three accessions of each species were grown at two field locations and variation in seed number among plants and mean seed production per plant used to estimate \( N_e/N_c \). Mean seeds per whole plant, standard deviations, and \( N_e/N_c \) differed among species, and accessions within species \( (P < 0.05) \). For whole plant samples, average \( N_e/N_c \) for each species differed with values of 0.42, 0.51, and 0.63 for \( L. \) perenne, \( F. \) pratensis, and \( P. \) spicata, respectively. However, average \( N_e/N_c \) based on two inflorescences per plant was 0.69, 0.88, and 0.86 for \( L. \) perenne, \( F. \) pratensis, and \( P. \) spicata, respectively, which was higher than that of whole plant samples. This higher \( N_e/N_c \) resulted from the elimination of the variation in inflorescence number per plant, a major source of variation in seed number among plants. The results showed the high potential for genetic drift in small regeneration populations. Increased plant populations and harvesting a constant number of inflorescences per plant are recommended as cost-effective methods to minimize genetic drift during regeneration of outcrossing grass germplasm.

More than 17,000 forage and turf grass accessions are maintained by the Western Regional Plant Introduction Station (WRPIS). The majority of these are self-incompatible, wind-pollinated species with high levels of heterogeneity. Germplasm accessions received at the WRPIS usually require an initial seed increase before the quantity and quality of seed is adequate for storage and distribution to users for research purposes. A regeneration sample from the initial seed increase should be assembled and placed in long-term storage (−18°C or lower). In this way, accessions are preserved as long as possible between regeneration cycles. After the initial stock of regenerated seed is depleted or has low germination, and if viable original seed is no longer available, the regeneration sample must be used to grow plants to replenish seed stocks.

Among the factors that have a strong bearing on the genetic quality of accessions maintained in gene banks are the initial germplasm collecting process, the need for initial and periodic seed regeneration of accessions, and consideration of differences in mating systems (Crossa and Vencovsky, 1994, 1997; Vencovsky and Crossa, 1999). The potential for random genetic drift is a major concern in the relatively small populations associated with regeneration. The effective population size \( (N_e) \) rather than the census population size \( (N_c) \) is the key parameter in genetic drift, and is defined as the size of an ideal population that would have the same amount of random genetic drift as the actual census population (Wright, 1931; Crow and Kimura, 1970). There are a number of factors relating to germplasm conservation that affect \( N_e \) (Frankel et al., 1995). Normally, \( N_e \) is lower than \( N_c \). A major cause of this is the differential production of gametes among parents; that is, variation in potential fecundity (Heywood, 1986). However, if mating is controlled so that every individual in a population contributes two gametes to the next generation, then \( N_e \) is essentially doubled relative to \( N_c \) (Crow and Kimura, 1970). Other regeneration schemes to maximize \( N_e \) include various types of controlled polycrosses (Breese, 1989). When male and female gametes are uncontrolled and seeds are bulk harvested, variation in fecundity can substantially reduce \( N_e \) compared with \( N_c \).

Thus, whenever possible, germplasm managers should control mating to maximize \( N_e \). But because of the need to regenerate large numbers of accessions with limited resources, controlled mating is often difficult. In any case, it is important for germplasm managers to understand the degree that variation in fecundity among plants can lower \( N_e \) during regeneration. With this information, the extent that \( N_e \) is reduced relative to \( N_c \) can be estimated and different methods of sampling considered that would minimize potential genetic drift. There are, however, very few estimates of \( N_e \) available, especially generic to germplasm regeneration. In addition, there is a need to develop and implement cost effective sampling methods that maximize \( N_e \). The objectives of this study were to estimate \( N_e/N_c \) in three model wind pollinated perennial grass species and to recommend cost effective sampling methodology to maximize \( N_e/N_c \) during seed regeneration.

MATERIAL AND METHODS

Calculation of \( N_e/N_c \)

Heywood (1986) showed that seed production per plant is correlated with the number of gametes individuals contribute to the population gamete pool, which in turn affects genetic drift through its association with variation in fecundity. Heywood (1986) derived the following equation to quantify the


proportional reduction in the effective population size associated with variation in potential fecundity:

\[ N_e/N_i = 1/(1 + F) (\sigma^2/\mu^2) + 1, \quad [1] \]

where \( F \) is the fixation index and \( \sigma \) and \( \mu \) is the standard deviation among plants and the mean family size. Thus, when each parent contributes equally to the gamete pool, \( \sigma^2 \) is zero and \( N_e = N_i \). Under these conditions, genetic drift is equal to the traditional binomial sampling model (Heywood, 1986).

Variation in seed production per plant normally reflects variation in pollen and ovule output. When the variance and mean number of seeds sampled per plant is unknown then an estimate of \( \sigma^2/\mu^2 \) can be obtained as follows:

\[ \sigma^2/\mu^2 = s^2/z^2 - 1/z, \quad [2] \]

where \( s^2 \) is the sample variance in seed number among plants and \( z \) is the sample mean seed number per plant in a given population (Heywood, 1986). Seed production of most plants sampled is usually high enough that the correction term, \( 1/z \), has little effect.

### Statistical Model

Consider the following definitions:

\[ Y_{ij} = \text{Number of seeds on the } i\text{th inflorescence of the } j\text{th plant} \quad (j = 1, 2, \ldots, m_i; i = 1, 2, \ldots, n), \quad [3] \]

\[ \bar{Y}_i = \frac{\sum_{j=1}^{m_i} Y_{ij}}{m_i} \quad \text{is the mean number of seeds per inflorescence on the } i\text{th plant}, \quad [4] \]

\[ \bar{Y} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{m_i} Y_{ij}}{\sum_{i=1}^{n} m_i} = \text{the mean number of seeds per inflorescence across all plants}. \quad [5] \]

Using the above definitions, we developed an ANOVA (Table 1) to examine how sampling can be used to minimize variation in fecundity and thereby minimize reductions in \( N_e/N_i \). The between- or among plant component (Table 1) is equivalent to \( \sigma^2 \) in Eq. [2] above, which is the mean of the among and within plant variance for seed number, and \( \bar{Y} \) is the mean number of seeds per inflorescence times the mean number of inflorescences per plant.

Both of these factors are included in the within-plant component for variance in seed number per plant. Thus, if every plant naturally produced the same number of inflorescences per plant and the same number of seeds per inflorescence, seed number among plants would not vary and \( N_e = N_i \). This would not be expected to occur naturally, but artificial sampling can be completed so that the number of seeds harvested per plant is equal for each plant. Equalizing the number of seeds harvested per plant eliminates the variation in seed number among plants. Although assembling such a regeneration sample would not control the variation in pollen production occurring before seed development, it would eliminate the maternal variation in gamete production, and \( N_e \) would be increased compared with harvesting all seeds from whole plants. Instead of sampling an equal number of seeds per plant, an equal number of inflorescences could be sampled from each plant. Even though seeds per plant sampled would still vary among plants as seeds per inflorescence varies, a major source of variation; that is, inflorescence number per plant, is eliminated. As a result, the variation in seed number among plants would be reduced along with the maternal variation to the gamete pool. Although \( N_e \) would not be as high as when seeds per plant are fully equalized, \( N_e \) would nonetheless increase compared with whole plant sampling.

### Field Testing

Estimates of \( N_e/N_i \) were made on the basis of seed number per plant from three accessions of three model perennial grass species maintained at the WRPIIS. The three species were *Festuca pratensis*, *Lolium perenne*, and *Pseudoroegneria spicata*, and each species was represented by three accessions. These species are outcrossing, wind-pollinated, and self-incompatible, so a value of \( F = 0 \) was assumed for the calculations of \( N_e/N_i \) (Brown, 1979; Johnson, 1998). All accessions were diploid. *Festuca pratensis* and *L. perenne* are known diploid species (Berg et al., 1979; Terrell, 1968), and diploid *P. spicata* accessions were selected on the basis of information in the Germplasm Resources Information Network (GRIN).

The experiment consisted of a randomized complete block with two replications established at both Pullman and Central Ferry, WA. Each plot consisted of 30 plants, and there were 18 plots at each location, for a total of 1080 plants. The Pullman location (46°43′55″N, 117°9′25″W) is about 800 m in elevation and plants were grown under dryland conditions. Central Ferry (46°40′13″N, 117°45′8″W) is located approximately 50 km from Pullman in the Snake River Canyon at

### Table 1: Decomposition of the sums of squares, associated degrees of freedom, and mean squares associated with sampling among and within plants where \( Y_{ij} = \text{number of seeds on the } j\text{th inflorescence of the } i\text{th plant} \quad (j = 1, 2, \ldots, m_i; i = 1, 2, \ldots, n), \quad \bar{Y} \text{ is the mean number of seeds per inflorescence on the } i\text{th plant}, \quad \bar{Y} \text{ the mean number of seeds per inflorescence across all plants.}

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of squares</th>
<th>Degrees of freedom</th>
<th>Mean squares</th>
<th>Expected mean squares†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among plants</td>
<td>[ \sum_{i=1}^{n} \sum_{j=1}^{m_i} (Y_{ij} - \bar{Y})^2 ]</td>
<td>( n - 1 )</td>
<td>( \bar{S}<em>{\text{among}} = \frac{\sum</em>{i=1}^{n} \sum_{j=1}^{m_i} (Y_{ij} - \bar{Y})^2}{n - 1} )</td>
<td>( \sigma^2_{\text{total}} + c\sigma^2_{\text{among}} )</td>
</tr>
<tr>
<td>Within plants</td>
<td>[ \sum_{i=1}^{n} \sum_{j=1}^{m_i} (Y_{ij} - \bar{Y})^2 ]</td>
<td>( m_i - 1 )</td>
<td>( \bar{S}<em>{\text{within}} = \frac{\sum</em>{i=1}^{n} \sum_{j=1}^{m_i} (Y_{ij} - \bar{Y})^2}{m_i - 1} )</td>
<td>( \sigma^2_{\text{within}} )</td>
</tr>
<tr>
<td>Total</td>
<td>[ \sum_{i=1}^{n} \sum_{j=1}^{m_i} (Y_{ij} - \bar{Y})^2 ]</td>
<td>( \sum_{i=1}^{n} m_i - \sum_{i=1}^{n} m_i^2/n - 1 )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

† \( \sigma^2_{\text{within}} \) and \( \sigma^2_{\text{among}} \) represent the true within and between variance, respectively, and the constant \( c = \frac{\sum_{i=1}^{n} m_i - \sum_{i=1}^{n} m_i^2/n}{n - 1} \).

If the number of spikes per plant were constant, and \( m_i = m \) for all \( i = 1, 2, \ldots, n \), then the constant \( c = m \).
RESULTS AND DISCUSSION

Nc/Nc on Whole Plants

An analysis of variance showed that the interactions with location were not significant for mean seeds per whole plant, standard deviations for seed number among plants, and Nc/Nc. Values for standard deviations at Central Ferry were 24% lower than at Pullman, but mean seeds per plant were 50% lower. This combination resulted in generally lower s²/σ² and a higher average Nc/Nc at Pullman (X = 0.59) than at Central Ferry (X = 0.45). Although other factors may have contributed to location differences, there was a high incidence of rust (Puccinia spp.) on L. perenne, and to a lesser extent on F. pratensis observed at the irrigated Central Ferry site. At the dryland Pullman site, rust was less severe. Plants of P. spicata did not show disease symptoms at either location, and location means for P. spicata did not differ between Central Ferry and Pullman for treatment factors (P < 0.05). There were no obvious differences in rust incidence among accessions within Lolium and Festuca.

There were significant differences among species, and accessions within species, for mean seeds per plant, standard deviations for seed number among plants, and Nc/Nc (P < 0.05). Since the interactions with location were not significant, the accessions were averaged over locations (Table 2). Although seed number per plant was highly variable, it differed significantly within F. pratensis accessions, but not among L. perenne or P. spicata accessions. Differences in standard deviations were detected for F. pratensis and L. perenne accessions, but not P. spicata accessions (Table 2). For Nc/Nc, differences among accessions within species were only observed for L. perenne. For the L. perenne accessions W6 9344 and W6 9359, Nc/Nc values were less than 0.40, showing strong reductions in Nc associated with high variation in seed number among plants. Overall, F. pratensis had significantly higher seed numbers per plant, and standard deviations among plants, than L. perenne or P. spicata, but L. perenne and P. spicata did not differ (Table 2). The differences among species for Nc/Nc were always significant (P < 0.05) with P. spicata having the highest and L. perenne the lowest values (Table 2). As far as we know, these are the first estimates of Nc in regeneration populations of these species. The relatively low Nc/Nc emphasizes the strong potential for genetic drift in small regeneration populations of outcrossing grass germplasm.

Nc/Nc with Inflorescence Sampling

For Nc/Nc, based on seed number per inflorescence, the location mean at Pullman was 0.77 and at Central Ferry it was 0.84. The species × location interaction was not significant, so the species data for the inflorescence samples were averaged over locations (Table 3). Mean seed number per inflorescence differed for each species, with F. pratensis having the highest and P. spicata the lowest values (Table 3). The mean value for Nc/Nc based on inflorescence samples was 0.81, which was significantly larger (P < 0.01) than the corresponding mean
value for total seeds per plant for the same three accessions ($N_s/N_c = 0.51$). Thus inflorescence samples always had a higher $N_s/N_c$ than the whole plant samples as $s^2/z^2$ was reduced. Although substantial variation in seed number per inflorescence still exists, a major source of variation in seed number among plants; that is, inflorescence number per plant, was eliminated. Since inflorescence sampling does not control variation in pollen production among plants before seed development, the values from this method represent improvements in $N_s/N_c$ only associated with the contribution of maternal gametes in the population. For whole plant seed sampling, however, the variation in seed number among plants represents the variation in both ovule and pollen production on $N_s/N_c$.

Heywood’s (1986) estimates of $N_s/N_c$ for $F = 0$ on the basis of seed number among plants from a wide array of species ranged from 0.14 to 0.68. Johnson (1998) found $N_s/N_c$ based on seed per whole plant averaged 0.68 for three Lolium multiflorum Lam. accessions, and averaged 0.77 for seeds per inflorescence. Thus the inflorescence samples were on average only 13% greater than the seeds per whole plant sample in that study, compared with an average of 59% in the current study. Nevertheless, sampling a constant number of inflorescences per plant would always result in higher $N_s/N_c$ than whole plant samples as inflorescence number per plant is eliminated as a source of variation in seed number among plants (Table 1).

We assumed a fixation index $F$ of zero, which is expected for the species in this study (Brown, 1979; John-son, 1998). In some cases, a degree of selfing does occur in normally self incompatible species resulting in an $F > 0$. The effect of this is to increase the denominator in Eq. [1] leading to lower $N_s/N_c$. For example, if $F = 0.25$ then $N_s/N_c$ would have declined to 0.47 compared with the experimental wide average of 0.52 when $F = 0$. Thus, a 25% increase in $F$ would have led to just a 10% decreases in $N_s/N_c$.

Greater variation in seed number among plants and fewer seeds per plant would be expected when plants are under either biotic or abiotic stress. In this study the prevalence of rust in L. perenne and to a lesser extent F. pratensis appeared to increase variation in seed number among plants and reduce seed production, especially at Central Ferry. The extent of rust in 1998 was unexpected on the basis of previous experience. But as a result, careful monitoring for stem rust and control through fungicides has been implemented to promote plant health during regeneration. Differences in seed production could also arise from naturally occurring genetic variation within and among accessions in response to environment, vernalization requirements, and different stresses.

How may germplasm managers offset limitations in resources to provide the best regeneration sample possible for the maintenance of diversity? It should be remembered that as long as seed held in long-term storage

### Table 2. Estimates of mean seed number per plant, standard deviations for seed number among plants, and the proportional reduction in effective population size ($N_s/N_c$) associated with variation in seed number among plants for three grass accessions of Festuca pratensis, Lolium perenne, and Pseudoroegneria spicata.

<table>
<thead>
<tr>
<th>Species</th>
<th>Accession</th>
<th>Mean seed number</th>
<th>Standard deviation</th>
<th>$N_s/N_c$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Festuca pratensis</strong></td>
<td>W6 11519</td>
<td>4.102b</td>
<td>3.281b</td>
<td>0.55a</td>
</tr>
<tr>
<td></td>
<td>W6 17784</td>
<td>11.592a</td>
<td>9.821a</td>
<td>0.56a</td>
</tr>
<tr>
<td></td>
<td>W6 17806</td>
<td>10.857a</td>
<td>12.262a</td>
<td>0.42a</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>8.851†</td>
<td>8.455†</td>
<td>0.51†</td>
</tr>
<tr>
<td><strong>Lolium perenne</strong></td>
<td>W6 9344</td>
<td>5.916a</td>
<td>6.353a</td>
<td>0.39b</td>
</tr>
<tr>
<td></td>
<td>W6 9359</td>
<td>1.741a</td>
<td>1.537b</td>
<td>0.31b</td>
</tr>
<tr>
<td></td>
<td>W6 9362</td>
<td>3.564a</td>
<td>2.958ab</td>
<td>0.56a</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>3.647</td>
<td>3.613</td>
<td>0.42</td>
</tr>
<tr>
<td><strong>Pseudoroegneria spicata</strong></td>
<td>PI 232131</td>
<td>3.144a</td>
<td>2.419a</td>
<td>0.63a</td>
</tr>
<tr>
<td></td>
<td>PI 232139</td>
<td>1.608a</td>
<td>1.122a</td>
<td>0.66a</td>
</tr>
<tr>
<td></td>
<td>PI 236681</td>
<td>1.848a</td>
<td>1.576a</td>
<td>0.59a</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>2.200</td>
<td>1.706</td>
<td>0.63</td>
</tr>
</tbody>
</table>

* Within a species and column, numbers with different letters are different at $P < 0.05$ on the basis of the LSD.  
† Differences between F. pratensis and other species were significant at $P < 0.05$ on the basis of the LSD; L. perenne and P. spicata did not differ.  
‡ Differences among species were significant at $P < 0.05$ using the LSD.  
§ Calculated as $N_s/N_c = \left[1 + (1 + F)\sigma_2/\mu_2^2 \right]^{-1}$, where $N_s$ and $N_c$ are the effective and census population sizes, respectively, $F$ is the fixation index (assumed to be zero), and $\sigma_2/\mu_2^2$ was estimated as $s^2/z^2-1/z$ for each plot. Using mean $z$ and $s$, above will give somewhat different values.

### Table 3. Estimates of mean seed number per inflorescence sample, standard deviations for seed number per inflorescence, and the proportional reduction in effective populations size ($N_s/N_c$) associated with variation in seeds per inflorescence for Festuca pratensis, Lolium perenne, and Pseudoroegneria spicata.

<table>
<thead>
<tr>
<th>Species</th>
<th>Accession</th>
<th>Mean seed number</th>
<th>Standard deviation</th>
<th>$N_s/N_c$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Festuca pratensis</strong></td>
<td>W6 17784</td>
<td>368a*</td>
<td>136a</td>
<td>0.88a</td>
</tr>
<tr>
<td><strong>Lolium perenne</strong></td>
<td>W6 9344</td>
<td>134b</td>
<td>105ab</td>
<td>0.69a</td>
</tr>
<tr>
<td><strong>Pseudoroegneria spicata</strong></td>
<td>PI 236681</td>
<td>71c</td>
<td>29b</td>
<td>0.86a</td>
</tr>
</tbody>
</table>

* Species with different letters are different at $P < 0.05$ on the basis of the LSD.  
† Calculated as $N_s/N_c = \left[1 + (1 + F)\sigma_2/\mu_2^2 \right]^{-1}$, where $N_s$ and $N_c$ are the effective and census population sizes, respectively, $F$ is the fixation index (assumed to be zero), and $\sigma_2/\mu_2^2$ was estimated as $s^2/z^2-1/z$ for each plot. Using mean $z$ and $s$, above will give somewhat different values.
can be drawn from an original, or close to original sample with high viability, there would be no need for regeneration and no genetic drift would occur. But at some point regeneration is inevitable, and a decision must be made concerning the basic plant population number per accession for a regeneration program. This should be in terms of $N_e$ rather than $N_c$, recognizing that in a random mating diploid population, theory indicates that $1/(2N_e)$ heterozygotes will be lost per generation (Crow and Kimura, 1970). Once the target $N_e$ is established the manager may control mating to maximize $N_c$ or simply increase $N_c$ to compensate for the estimated reductions in $N_e$ associated with variation in fecundity.

To maximize $N_e$ with the fewest plants possible, mating can be controlled so that each individual in the population contributes two gametes to the next generation. With this sampling system $N_e = 2N_c - 1$ so $N_e$ is essentially doubled compared with $N_c$ (Crow and Kimura, 1970). To accomplish this in self-incompatible grasses, paired inflorescences from different plants could be bagged together to obtain a reciprocal cross. If the progeny from the crosses are kept separate, a viable seed from each parent can be used for future regeneration. But the expense of bagging, separate plant harvests, seed cleaning, storage, and the potential for differential poor seed set and seed quality of bagged plants often makes this approach impractical.

At the WRPI, about 600 accessions are regenerated each year and this has barely kept pace with the average number of incoming accessions. If the target $N_e$ is 50, corresponding to an expected erosion in heterozygosity of 1.0% per regeneration, 30,000 separate harvests, seed cleaning, and seed counts would be needed before equal numbers of seeds per plant could be assembled for the 600 accessions. Paired crossing or controlled polycrosses described by Breese (1989) that require individual plant harvests are not possible with current resources. Bulk sampling with uncontrolled mating, but with increased $N_c$ to approximate the desired $N_e$, may be the most cost effective option. Considerable mechanization for planting and care of nurseries is available making increases in plant populations per accession a feasible alternative to controlled crossing. On the basis of the $N_c/N_e$ average of 0.52 (Table 2), it would take 96 plants to obtain a $N_e$ of 50. Sampling a constant inflorescence number per plant would further reduce the $N_e$ needed to attain the desired $N_e$.

Research is needed concerning the optimum number of inflorescences to be sampled per plant, and this is underway at the WRPI for wide range of species and accessions. Because of the known relationship between the variance among individuals and among means of individuals (mean $s^2 = s^2/m$) (Steel and Torrie, 1980), increasing the number of inflorescences $m$ sampled per plant would decrease mean $s^2$ within plants. Since $s^2$ within plants is a component of total $s^2$ among (Table 1), a reduction in $s^2$ among would be expected with increased $m$. As a result, $s^2/z^2$ in Eq. [2] would decrease and $N_c/N_e$ for the maternal gamete variation would increase. When seeds per plant are equal, the maternal gamete variation is zero and $N_c$ is increased by about one-third compared with whole plant sampling (Breese, 1989). In other words, without controlling the variation in pollen production among plants this would be the upper limit of the expected gain in $N_e$ associated with methods that equalize seeds per plant.

On the basis of the average $N_c/N_e$ of 0.52, the WRPI has set a minimum regeneration target population of 100 plants per accession. In addition, if a constant number of inflorescences per plant is sampled, $N_c/N_e$ will consistently increase compared with bulking seeds from whole plants. When available resources prevent the application of controlled crosses to maximize $N_e$, selecting an appropriate population size based on $N_c/N_e$ estimates, along with sampling a constant number of inflorescences, is recommended as a cost effective approach to grass seed regeneration.

ACKNOWLEDGMENTS

Thanks to Connie LeClaire Foiles for her fine technical assistance.

REFERENCES


