Theor Appl Genet (1994) 88: 885–890

R. J. Cross · A. R. Wallace

Loss of genetic diversity from heterogeneous self-pollinating genebank accessions

Received: 6 July 1993 / Accepted: 25 November 1993

Abstract Genebank seed accessions of predominantly self-pollinating species may be stored either as bulked (mixed) seed lines or as pure line cultivars. If seed lines are bulked in storage then when considered over several regeneration cycles, loss of genetic diversity within heterogeneous self pollinating genebank accessions is shown to be severe. This within-accession loss of diversity represents opportunities foregone through the random loss of individual genotypes. Amongst working collections, the utility and repeatability of genebank accessions is paramount in the justification of the germ plasm resource. Therefore, the only practical solution to the management of predominantly self-pollinating species is to preserve individual accessions as pure lines.

Key words Genetic diversity · Genetic truncation
Genetic resources · Single seed descent

Introduction

The principles of ex situ conservation of orthodox seeds are well established, as are the methods and procedures of monitoring the viability and timing of rejuvenation (e.g. Hanson 1985; Ellis et al. 1985a, b). However, there remains conjecture and uncertainty over rejuvenation procedures and a lack of guidance from the International Board for Plant Genetic Resources (IBPGR) on the most appropriate rejuvenation method for each species. This uncertainty has been expressed by the IBPGR (McCusker 1989) and is the likely reason for the construction of practical rejuvenation experiments to investigate the loss of genetic diversity in genetically heterogeneous genebank *Phaseolus* accessions (IBPGR 1988, 1989).

Communicated by A. R. Hallauer

R. J. Cross (✉) · A. R. Wallace
New Zealand Institute for Crop & Food Research Limited,
Private Bag 4704,
Christchurch, NZ

The consequences of different pollination mechanisms and their genetic contribution to the next generation are discussed in an excellent overview by Breese (1989). However, the generalities discussed by Breese cannot be transformed into specific crop recommendations until the genetic integrity of an individual genebank accession has been defined. Although this paper discusses the problem of genetic truncation with reference to the regeneration of heterogeneous genebank accessions, the arguments and principles developed here apply equally to the necessity of strict pure-line maintenance of individual genotypes under the single seed descent plant breeding procedure.

Should individual genebank accessions be genetically heterogeneous or pure? This paper considers the consequences of either option with reference to the maintenance of self-pollinating (autogamous) species.

Materials and methods

Three general areas of investigation are developed. These comprise a computer simulation study, the development of a mathematical model and worked example, and a graphical representation of the genetic truncation over several regeneration cycles. Precise details of the theory may be obtained from one of the authors (ARW).

Results

Sampling theory

If it is assumed that each and every individual genotype within an autogamous heterogeneous genebank accession has an equal opportunity of being selected for subsequent plantings, and is of equal reproductive fitness when compared to all other genotypes within the selected regeneration cycle, then shifts in genotypic composition can be predicted from probability theory. However, for a realistic rejuvenation nursery size of say 1000 lines, the calculations required to sum the expected number of genotypes lost (and

Table 1 Simulation of a bulked seed method of maintaining a seed collection, showing mean and standard deviation of the percentage of genotypes remaining after one, two and three regeneration cycles, for selected reproductive and sampling rates and two population sizes.

Reproductive rate and sampling intensity at each generation	Population size (initial number of genotypes)					
	1000			100		
	% of genotypes remaining	Standard deviation	Number of simulations	% of genotypes remaining	Standard deviation	Number of simulations
	(a) One regeneration					
10	65.2	0.92	200	65.2	3.1	1000
	(b) Two regenerations					
10	49.0	0.95	200	49.3	3.0	1000
	(c) Three regenerations					
2	51.7	0.94	500	51.7	2.8	1000
5	–	–	–	41.9	3.0	1000
8	–	–	–	40.2	3.0	1000
10	39.4	0.88	200	39.9	3.0	1000
12	–	–	–	39.3	2.9	1000
15	–	–	–	38.9	3.0	1000

associated standard deviations) comprise an extremely tedious task when considered over several regeneration cycles. Accordingly, the regeneration process was computer simulated to evaluate the effect of bulking genebank material on the composition of lines within the genebank. This simultation is very simple to program, although expensive in computer CPU time. For example, simulation of three regeneration cycles of 1000 seedlines with a 10-fold reproductive rate required 8.6 h of CPU time per 100 simulations on a VAX 11/780. The model is described below, and results of these simulations are given in Table 1.

Simulation model

The loss which is expected to occur with the bulking of seedbank material after each regeneration cycle was investigated by a series of runs of a computer simulation model, where:

1) The original collection consisted of 1000 seeds, and each seed was "genetically" distinct.
2) Each of the 1000 seeds in the collection produced 10 seeds (equal reproductive fitness), producing a total of 10 000 rejuvenated seeds that were "bulked".
3) One thousand seeds were selected at random and returned to the genebank.
4) This regeneration cycle was repeated to give results for two and three regeneration cycles.

The computer simulation of the conditions described above showed a very large reduction in genetic diversity. Individual traces of the truncation rate for five of the simulations, continued for eight regeneration cycles, are shown in Fig. 1, and the average truncation rates with associated standard deviations at each regeneration cycle are shown in Table 1. For example, given the constraints detailed above, after three regeneration cycles only 39.4% of the original genotypes remained with a very small variation about this mean. Tests confirmed that the distributions in each case were very close to Normal.

Branching process model

An approximation to the true plant regeneration process can be made by considering the following probabilistic branching process model which has characteristics similar to the regeneration process.

Let X_k be the size of the population, comprising 1 or more seeds of each of U_k different genotypes after k regeneration cycles. Initially there is 1 seed of each of n genotypes in the genebank, i.e. $X_0 = U_0 = n$. At each regeneration cycle, individuals are assumed independently to produce s offspring, where s is between 0 and r with the Binomial probability distribution $B\left(r, s; \frac{1}{r}\right)$, and then die. Thus:

$$P(X_k = s \mid X_{k-1} = 1) = \binom{r}{s}\left(\frac{1}{r}\right)^s \left(\frac{r-1}{r}\right)^{r-s} \quad (1)$$

Let Z_{kj} be the number of offspring of the jth genotype present after k regenerations. Then the size of the population after k regenerations (X_k) is:

$$X_k = \sum_{j=1}^{n} Z_{kj}$$

From standard branching process theory (e.g. Feller 1968) the expected value of X_k is n and the variance of X_k is given by:

$$Var(X_k) = \frac{nk(r-1)}{r}$$

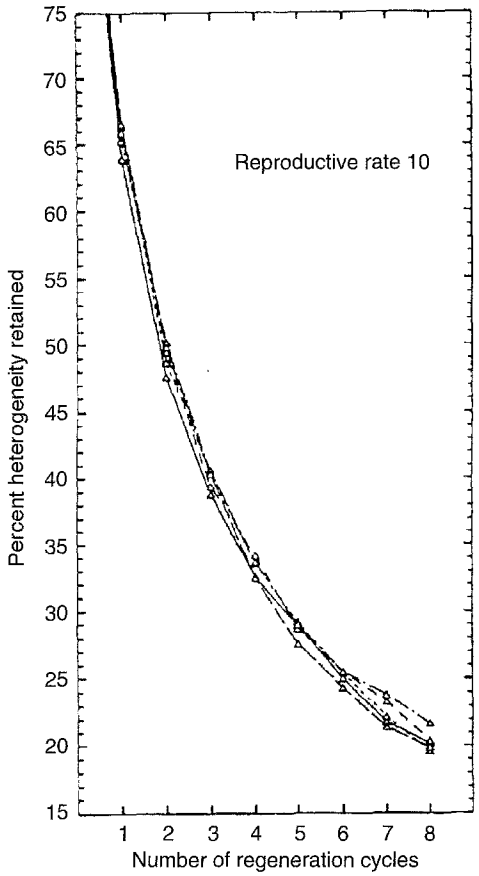

Fig. 1 Percent heterogeneity retained in five simulations of a population of 1000 unique individuals with a reproductive rate of 10 seeds per individual and selection intensity of 10%

Thus, this branching process model maintains the genebank close to its initial size with fairly high probability for small values of k regeneration cycles.

Individual genotypes are lost from the genebank when the number of offspring of the jth genotype present after k regenerations is zero i.e. $Z_{kj}=0$. If we regard the Binomial distribution in Eq. 1 as a set of r Bernouilli trials and use a recurrence relation then:

$$P(Z_{k+1,j}=0)=\left[\frac{r-1}{r}+\frac{1}{r}P(Z_{kj}=0)\right]^r, \quad k\geq 1$$

with $P(Z_{1j}=0)=\left(\frac{r-1}{r}\right)^r$

Let P_m^k be the probability that m or more genotypes are lost after k regeneration cycles. Then:

$$P_m^k=\sum_{s=m}^{n}\binom{n}{s}\left[P(Z_{kj}=0)\right]^s\left[1-P(Z_{kj}=0)\right]^{n-s}$$

This is the upper tail of the Binomial distribution and, since n is large, the distribution may be approximated by an equivalent Normal distribution.

Worked example

Consider the case of 1000 seeds with a 10-fold reproductive rate after the first regeneration cycle, i.e. $n=1000$, $r=10$, $k=1$.

Thus $P(Z_{1j}=0)=0.3487=p$, say.

Mean number of genotypes lost $(np)=348.7$

Standard deviation of number of genotypes lost $\left(\sqrt{np(1-p)}\right)=15.07$

i.e. after adding a continuity adjustment of 0.5 to the approximating Normal distribution we have:

$$P_m^1=1-\Phi\left(\frac{m-349.2}{15.07}\right)$$

(where $\Phi(x)$ is the cumulative Normal distribution function).

Thus, the mean percentage of genotypes remaining after the first regeneration cycle of 1000 genotypes is 65.1% with a standard deviation 1.51%.

Similarly,

$$P_m^2=1-\Phi\left(\frac{m-510.4}{15.81}\right)$$

$$P_m^3=1-\Phi\left(\frac{m-605.5}{15.46}\right)$$

The means and standard deviations for these cases may be directly compared with the simulation results for 1000 genotypes shown in Table 1, where the means are expressed as the percentage of genotypes remaining. The branching process approximation provided excellent estimates of the mean percentage of genotypes remaining but over-estimated the standard deviation by about 70%. The latter overestimation was expected because the number of genotypes remaining are not independent, as the branching process has assumed. This lack of independence arose because as more than the mean number become extinct in a given generation there must be more than the expected number of the remaining genotypes present since the total is constrained to remain constant; and vice versa. Thus, in the next generation there will be a tendency to regress toward the expected number of genotypes, reducing the dispersion to a level less than the Binomial dispersion of the branching process.

Influence of differing reproductive fitness (sensitivity analysis)

The sensitivity of the loss of genetic diversity to changes in reproductive fitness was investigated by simulation, for three regeneration cycles only, on a population of 100 genotypes. This smaller population of genotypes was chosen because of the very high computer processing times required for simulation. Reproductive rates of 2, 5, 8, 10, 12 and 15 were assumed, and the selection intensity was adjusted in each case to maintain the mean size of the collec-

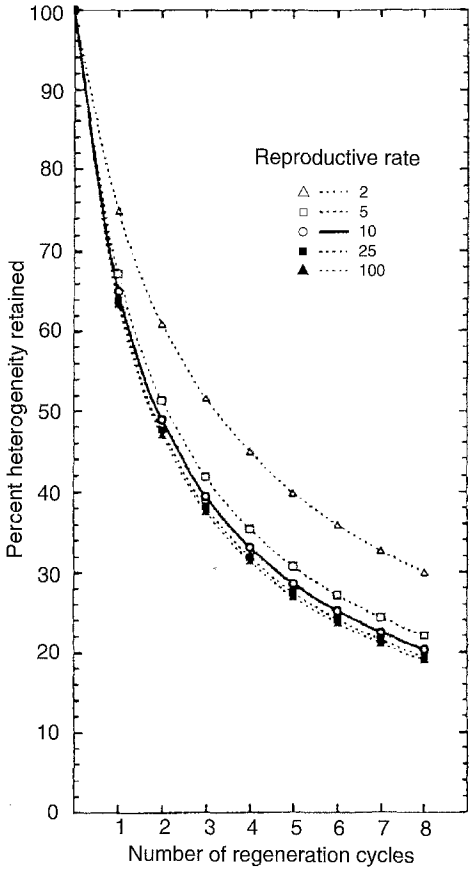

Fig. 2 Average heterogeneity retained in a population of 1000 unique individuals for a range of reproductive rates and selection intensities evaluated with the branching process approximation

tion at 100 seeds after each regeneration cycle. Results for the 2-fold and 10-fold reproductive fitness agreed closely with the corresponding results for 1000 seeds, confirming that the approximation by the smaller population does not distort the estimated losses of genetic diversity. Results for the tested reproductive rates are shown in Table 1. They indicate very good stability of the outcome over the range of fitness from 5- to 15-fold. Table 1 also shows expected numbers of genotypes remaining after one and two regeneration cycles for the case of 10-fold reproductive fitness, to further demonstrate the validity of the 100 seed line approximation.

The branching process model was also evaluated for a range of reproductive rates from 2 to 100. The average number of genotypes remaining after successive regeneration cycles is shown in Fig. 2. The genetic truncation rates are shown to be insensitive to reproductive fitness and are in very close agreement with the simulation results of Table 1. For each regeneration cycle, the mean number of remaining genotypes approaches an asymptopic value as the reproductive rate is increased. The curve for the 10-fold reproductive rate lies approximately midway between the asymptotic curve for very high (100-fold) and low (5-fold) reproductive rates of fitness. This range of reproductive rates should cover most autogamous species in a genebank, and Fig. 2 shows that there is only a 3–4% variation in mean performance of the genebank regeneration procedure over this range. The simulation results quoted earlier for a 10-fold reproductive fitness are, therefore, likely to be correct for the more general situation encountered in practice where the reproductive fitness of an individual autogamous seed may vary between 5- and 100-fold.

Discussion

Whether identified by the branching process model or by computer simulation it is clear that over several regeneration cycles, the rapid truncation of genotypic variation occurs within originally heterogeneous genebank accessions. In reality, the original genebank accession would neither be 100% heterogeneous, nor would each seed within the sample have equal reproductive contribution to the next generation. However, the sowing of a similar plot size at each regeneration cycle and genetic heterogeneity of individual accessions within most genebanks, including those centres with IBPGR world or regional storage responsibilities, is a common feature. Heterogeneous landrace varieties sampled in the early part of this century would have been subjected to several regeneration cycles, and significant genetic truncation of the original sample would be expected. Therefore, present day genetic heterogeneity within individual genebank accessions would likely bear little similarity to the original sample. If the overall purpose of genebanking is to preserve samples of seeds as a reliable and repeatable resource to plant breeders and other interested researchers, then the storage of genetically heterogeneous accessions represents a major obstacle to fulfilling that objective.

The inspection of differing reproductive fitness rates assumed each seed had an equal reproductive rate. This obviously is seldom met in practice. Further, different environments will favour some individuals over others. Therefore, directional selection pressures will mean that some individuals will contribute more to the next generation, and the rate of genetic truncation will be even faster than reported here. Outcrossing between different genotypes would be expected to place some restriction on the rate of truncation reported here. Self-pollinating species are seldom 100% autogamous as demonstrated, for example, with modest outcrossing rates of 1–2% in cultivated barley (Doll 1987). However, such a low outcrossing rate would likely not achieve a balanced polymorphic population structure when subject to the regeneration procedure described in this paper.

The current study, although based on a mathematical model, demonstrates the potential for significant and substantial reduction in genetic diversity within an originally heterogeneous genebank accession. Because the genotypic content of a heterogeneous accession is changing with successive regeneration cycles, problems are expected with the accuracy of genotypic merit and of varietal description.

Therefore, the only way to preserve the genotypic integrity of an individual genebank sample is to maintain accessions as individual pure lines. Practical field trials based on the regeneration strategies described in this paper are in progress to test this contention.

Other studies

Although there are no formal regeneration studies of genebank accessions reported in the literature, two major field studies into the changes of genetic variability over subsequent generations have relevance. These studies are the "Composite Cross" programme involving bulked hybrid populations of cultivated barley at Davis, California (e.g. Suneson 1956), and the ex situ conservation study of wild tetraploid wheats (the "Ammiad" experiment) in Eastern Galilee (Anikster and Noy-Meir 1991). In the composite cross (CC) experiment, both agronomic and biochemical data demonstrate a rapid reduction in genetic variation, especially in the early generations. For example, in CCV the average variance for heading time and plant height within individual families was reduced by 11% and 53%, and 32% and 88% respectively, between generations 4 and 14 (Allard and Jain 1962). Similarly, the elimination of or a rapid reduction in "poor producing lines" (yield per 80 m^2 plot) was noted between generations 3 and 7 in four reported CC populations, and a substantial skew of CCII in the F25 generation where a severe infestation of yellow dwarf virus had occurred (Suneson 1956). These examples of truncation of variance are likely related to reproductive fitness and therefore would increase the probability of more numerous seed types being present in the next sown generation. Furthermore, this "dilution" effect of "selectively neutral" characters such as grain colour and leaf esterases may explain the loss of variability of black seeds in CCV from 19.5% in F_1 to 5.9% in F_8 and 0.7% in F_{14} (Suneson 1956), and a marked convergence to one pair of four-locus complementary leaf esterase gametic types in populations CCII and CCV (Clegg et al. 1972).

In the Ammiad experiment, similar observations of rapid truncation of variability were observed among 250 differing micro-environments. Noy-Meir et al. (1991) identified rainfall, rockiness and cattle grazing as major determinants in the rates of extinction and subsequent recolonisation of wild tetraploid wheats. Although these authors noted that the number and density of plants varied greatly from year to year, the phenological morphological traits within most micro-environments (Anikster et al. 1991) remained similar over the 6-year-long study period. Diversity appeared to parallel habitat heterogeneity where, for example, Nevo et al. (1991) observed the highest level of allozyme polymorphism in the "Karst" area, which was also the most variable for soil moisture. However, this simple explanation has an apparent contradiction, as the "Karst" area also contained the least diversity for high molecular weight glutenin polymorphism (Felsenburg et al. 1991). Whatever the cause of the underlying patterns of diversity, these results demonstrate that the observed genetic differentiation is not random and that plants contributing to the next generation do vary greatly from year to year. These assumptions underlie classical population genetic theory and, therefore, may invalidate model predictions. These inconsistencies are most likely to occur where subsequent generations are reliant upon high levels of subsampling within the diversity of the parent population.

Practical considerations

With regard to genebank management of originally heterogeneous seed samples, a consequence of maintaining genetically pure accessions would be an obvious manifold increase in the number of pure line accessions as compared to bulked seed lots. In favour of pure line maintenance, the high level of confidence in the integrity of the seed source, due to its accuracy and repeatability, would have obvious benefits for curator and researcher alike. The improved accuracy of the resource would mean that the information base of evaluation descriptors would become more relevant to multiple users.

The decision whether to maintain autogamous seed collections as pure lines or as bulks should be the subject of a cost-benefit study. These options must be addressed as an integral part of corporate strategic plan objectives of germ plasm conservation. Until the problem is fully addressed and resolved, indecision on appropriate rejuvenation procedures will persist, and the unconscious loss of genetic diversity within genebanks will continue.

Acknowledgements The authors thank Dr. S. F. L. Gallot (DSIR Physical Sciences, Wellington, New Zealand) for assistance and useful comment on the statistical derivations.

References

Allard RW, Jain SK (1962) Population studies in predominately self pollinated species. II. Analysis of quantitative genetic changes in a bulk hybrid population of barley. Evolution 16: 90–101

Anikster Y, Noy-Meir I (1991) The wild field laboratory at Ammiad. Isr J Bot 40:351–362

Anikster Y, Eshel A, Azrati W, Horovitz A (1991) Patterns of phenotypic variation in wild tetraploid wheat at Ammiad. Isr J Bot 40:397–418

Anon (1989) In: Geneflow. International Board for Plant Genetic Resources, Rome

Breese EL (1889) Regeneration and multiplication of germ plasm resources in seed genebanks: the scientific background. Int Board Plant Genet Res, Rome

Clegg MT, Allard RW, Kahler AL (1972) Is the gene the unit of selection? Evidence from two experimental plant populations. Proc Natl Acad Sci USA 69:2474–2478

Doll H (1987) Outcrossing rates of autumn and spring-sown barley. Plant Breeding 98:339-341

Ellis RH, Hong TD, Roberts EH (1985a) Handbook of seed technology for genebanks, volume 1. Principles and methodology. AGPG:IBPGR/85/84. Int Board Plant Genet Res, Rome

Ellis RH, Hong TD, Roberts EH (1985b) Handbook of seed technology for genebanks, vol 2. Compendium of specific germination information and test recommendations. AGPG:IBPGR/85/85. Int Board Plant Genet Res, Rome

Feller W (1968) In: An introduction to probability theory and its applications, vol. 1 (3rd edn). Wiley, New York

Felsenburg T, Levy AA, Galili G, Feldman M (1991) Polymorphism of high molecular weight glutenins in wild tetraploid wheat: spatial and temporal variation in a native site. Isr J Bot 40: 451–479

Hanson J (1985) Procedures for handling seeds in genebanks. AGPG:IBPGR/85/86. Inter Board Plant Genet Res, Rome

IBPGR (1988) IBPGR annual report. Int Board Plant Genet Res, Rome

IBPGR (1989) IBPGR annual report. Int Board Plant Genet Res, Rome

McCusker A (1989) In: Breese EL (ed) Regeneration and multiplication of germ plasm resources in seed genebanks: the scientific background. Int Board Plant Genet Res, Rome, p 3

Nevo E, Noy-Meir I, Beiles A, Krugman T, Agami M (1991) Natural selection of allozyme polymorphisms: micro-geographical spatial and temporal ecological differentiations in wild emmer wheat. Isr J Bot 40:419–449

Noy-Meir I, Agami M, Anikster Y (1991) Changes in the population density of wild emmer wheat (*Triticum turgidum* var dicoccoides) in a Mediterranean grassland. Isr J Bot 40:385–395

Suneson CA (1956) An evolutionary plant breeding method. Agron J 48:188–191