INTERNATIONAL BIOLOGICAL PROGRAMME 2

Crop genetic resources for today and tomorrow

EDITED BY

O. H. Frankel

Senior Research Fellow
Division of Plant Industry, CSIRO, Canberra, Australia

AND

J. G. Hawkes

Professor of Botany, University of Birmingham, UK

CAMBRIDGE UNIVERSITY PRESS

CAMBRIDGE
LONDON · NEW YORK · MELBOURNE

THE INTERNATIONAL BIOLOGICAL PROGRAMME

The International Biological Programme was established by the International Council of Scientific Unions in 1964 as a counterpart of the International Geophysical Year. The subject of the IBP was defined as 'The Biological Basis of Productivity and Human Welfare', and the reason for its establishment was recognition that the rapidly increasing human population called for a better understanding of the environment as a basis for the rational management of natural resources. This could be achieved only on the basis of scientific knowledge, which in many fields of biology and in many parts of the world was felt to be inadequate. At the same time it was recognised that human activities were creating rapid and comprehensive changes in the environment. Thus, in terms of human welfare, the reason for the IBP lay in its promotion of basic knowledge relevant to the needs of man.

The IBP provided the first occasion on which biologists throughout the world were challenged to work together for a common cause. It involved an integrated and concerted examination of a wide range of problems. The Programme was co-ordinated through a series of seven sections representing the major subject areas of research. Four of these sections were concerned with the study of biological productivity on land, in freshwater, and in the seas, together with the processes of photosynthesis and nitrogen-fixation. Three sections were concerned with adaptability of human populations, conservation of ecosystems and the use of biological resources.

After a decade of work, the Programme terminated in June 1974 and this series of volumes brings together, in the form of syntheses, the results of national and international activities.

4. Optimum sampling strategies in genetic conservation

D. R. MARSHALL & A. H. D. BROWN

There are definite limits to the numbers of samples which can be handled effectively in programmes for the conservation and utilisation of crop genetic resources. These limits are imposed by the financial and personnel resources available to carry out each stage in the process:

In most species there is one *major limiting factor* in the above process which determines what the upper sample limit will be. This major limiting factor varies from species to species. For example, in the case of the primitive land races of the temperate cereals in the Mediterranean basin which are rapidly approaching extinction, the major limiting factor is our capacity to collect the material before it is lost for ever. By contrast, in the case of vegetatively reproduced crops such as yams, sweet potatoes and manioc, where considerable variation still exists in the field but is difficult to conserve, the major limiting factor is conservation. Finally, in the case of the hexaploid weed relatives of cultivated oats (e.g. *Avena fatua* and *A. sterilis*) where ample material still exists in the field and there are no particular conservation problems, the major limiting factor is the breeder's capacity to evaluate and utilise the collected materials.

Obviously, by committing more resources to conservation and breeding programmes, it will be possible to raise these limits substantially. Further, by the judicious allocation of resources it will also be possible to change the major limiting factor in any species from, say, exploration to conservation. However, limits to the growth of collections will always exist, because the total resources which can be devoted to the exploration, conservation and utilisation of crop germplasm will always be

finite. Consequently in practice we should aim to conserve sufficient stocks of each species to saturate the plant breeder's capacity to evaluate and utilise the conserved germplasm both now and in the foreseeable future. That is, we should aim to make evaluation and utilisation the major limiting factors in all species.

The existence of crop specific limits and major limiting factors to the growth of germplasm collections highlights the need for co-ordinated and systematic planning of conservation programmes. There is no point in collecting material if it cannot be adequately conserved, nor is there any point in conserving material which cannot be evaluated and utilised. Moreover, unnecessary effort in any one species further reduces the resources which can be devoted to other species. The need for careful planning is nowhere more evident than in the formulation of sampling procedures for species where exploration is the major limiting factor to the growth of collections. These include species which are severely threatened by extinction as well as those which occur in remote areas or difficult terrain such that they are likely to be sampled only once. In these cases the plant explorer carries an immense responsibility, for it is his decisions which determine to what extent the gene resources of such species will be available for the use of future generations.

The development of efficient exploration programmes for a particular species requires decisions, and hence information, at two levels. At the first level we have decisions concerning the regions or areas of the world to be explored. Objective decisions at this level require information on (i) where significant gene pools of the species occur and where they are most threatened by extinction and (ii) the areas covered by previous exploration missions and their relative effectiveness as measured by the material held in existing collections. The need for such information has been repeatedly discussed in the literature in recent years (Bennett, 1965; Frankel, 1967; 1970*a*; Frankel & Bennett, 1970). Moreover, an extensive survey of crop genetic resources in their centres of diversity has recently been completed (Frankel, 1973, see also Chapter 37) and the Crop Ecology and Genetic Resources Unit of FAO has initiated a survey of existing collections, and this work is continuing. Consequently, these points will not be considered further here.

At the second level we have decisions concerning sampling procedures within the selected areas. These are:

(*a*) the number of plants to sample per site;

(*b*) the total number of sites to sample;

(*c*) the distribution of sampling sites within each area.

Objective decisions at this level, and hence the full definition of optimum sampling strategies, require a knowledge of the kinds and amounts of genetic variation in the target species and the apportionment of this variation among plants within populations, among populations within regions, and among regions within the selected areas. Unfortunately, quantitative information on the distribution of genic variation within and between populations of cultivated plants and their weed relatives is not available. Nor is it likely to become available. There is insufficient time to undertake population surveys of all the species and regions of interest prior to the preparation of exploration plans, particularly when the target populations are in jeopardy. Therefore in formulating sampling procedures for genetic conservation there is no practical alternative to using the information available from species which have been studied in some detail and extrapolating from those to the species where basic information is lacking. Much of the information on distribution of genic variation in nature in plant species has been summarised by Jain (Chapter 2) and Bradshaw (Chapter 3). Our purpose here is to formulate, as far as possible with current information, a quantitative sampling theory for genetic conservation which permits the collection of the maximum amount of genetically useful variability in the target species while keeping the number of samples within the practical limits discussed above. However, before we can consider the problem of optimal strategies in detail we need to define (i) an appropriate measure of genetic diversity – the parameter we wish to maximise, and (ii) what we regard as genetically 'useful' variability.

Measures of genetic diversity

There are a variety of measures which may be used to characterise the level of genetic diversity in a species and the apportionment of this variation within and between individuals, populations and regions. These measures fall into two classes.

(i) Measures based on genetic variance in quantitative characters. These measures are commonly used in population biology and plant breeding and have the advantage that they are familiar to all scientists in these fields.

(ii) Measures based on allelic diversity at loci governing qualitative characters. Such measures have seen increasing application in recent years due to the development of gel electrophoretic techniques to study allelic frequencies at single loci. The most commonly used measures,

which are analogous to those used by ecologists to measure species diversity (Hurlbert, 1971; Lewontin, 1972) are:

(*a*) Total number of alleles in the population.

(*b*) The proportion of heterozygotes that would be produced if the population were random mated. If the frequency on the *i*th allele at a locus in p_i, the heterozygosity (H) is given by:

$$H = \sum_{i,j=1}^{k} p_i p_j \quad (i \neq j)$$

$$= 1 - \sum_{i=1}^{k} p_i^2$$

where k is the number of alleles.

(*c*) The Shannon–Weaver information function

$$H' = -\sum_{i=1}^{k} p_i \log_2 p_i$$

p_i and k were defined previously.

Obviously, measure (*a*) depends only on number of alleles in the population (allelic richness) while measures (*b*) and (*c*) are functions of the frequencies (allelic evenness) as well as the number of the alleles in the populations.

We reasoned that any measure of genetic diversity chosen for the present purposes should meet two criteria. First, it should be a direct measure of genic diversity. This criterion precludes the use of all measures of diversity based on variance analyses of quantitative characters (class i above). Such parameters measure only that portion of the genic variability which is expressed phenotypically. The proportion of expressed variability varies markedly with the character under consideration and the genetic background and environment in which it is expressed. Consequently, measures of genetic diversity based on variance in quantitative characters may be unreliable indicators of the diversity in a population at the level of the individual gene.

Second, it should be a function predominantly of the number of different alleles in the population. Genetic conservationists are not primarily interested in preserving a representative sample of the target species, where representative is defined in as many senses as possible. Rather, they are interested in preserving at least one copy of each of the different alleles in the target species (Bennett, 1970*b*). An accurate representation in the sample of the allelic *frequencies* in the population is of much less interest and requires sampling on a far more extensive scale. Consequently, in the context of sampling for genetic conservation,

the property of allelic richness is considerably more important than the property of allelic evenness. This criterion, therefore, precludes all measures of diversity which are functions of both the number and frequency of alleles in the population. Thus, the average number of alleles per locus provides the simplest and least ambiguous measure of genetic diversity for the purposes of exploration and conservation, consistent with the above criteria.

It may be argued that the number of alleles per locus (n_a) has the disadvantage that it focuses attention on single genes, whereas the aim of conservation is to collect and conserve adaptive gene complexes. As emphasised by Brock (1971), in many circumstances induced mutations may represent a more efficient source of single gene variability than gene pools conserved from nature. Yet it is extremely unlikely that populations treated with artificial mutagens can replace natural gene pools as a source of co-adapted gene complexes which are of fundamental importance in the adaptation of populations to their environment (Dobzhansky, 1970). However, this disadvantage is more apparent than real. Firstly, co-adapted gene complexes can be regarded as 'alleles' at a 'super gene' and in the context of sampling strategies are no different in principle from alleles at a locus. Secondly, in population studies, it is impossible to study single loci independently of closely linked loci. It follows that where we use experimental data from single loci to draw inferences about collecting strategies, these inferences refer to the linked segments marked by the individual loci rather than the marker gene themselves. Thus, while we frame our discussion in terms of alleles at single loci for simplicity, the same conclusions hold for co-adapted gene complexes, which are the real target of genetic conservation.

'Useful' genetic variability

Since each gene consists of hundreds of nucleotides, each capable of base substitutions and with additional permutations possible through sequence rearrangements, additions and deletions, the potential number of allelic states at a single locus is virtually infinite. The number of different allelic combinations which may exist at several loci is even greater. Obviously, only a small fraction of these potential alleles or allelic combinations exist in a species at any one time. However, while the actual numbers of variants found in populations are small in relation to the total which may exist, they are nevertheless extremely large. In

fact, as emphasised by Allard (1970*a*), most species probably contain millions, if not hundreds of millions, of different genotypes. It might be argued that we should collect and preserve all existing variants. Yet, this is clearly impractical. Consequently, we need to define the class or classes of variants which we regard as potentially most useful and to which we will assign greater priority. As a first step towards this end we will consider the distribution of alleles or allelic combinations in populations in both theory and practice.

Theoretical distribution of alleles in populations

Neutral alleles. Kimura & Crow (1964; 1970) have formulated the expected number of neutral alleles (n_a), with allelic frequencies lying between p and q ($0 < p < q < 1$), in an equilibrium population of effective size, N_e, as:

$$n_a = \int_p^q \phi(x)\,dx = \theta \int_p^q (1-x)^{\theta-1} x^{-1} dx$$

where $\phi(x)dx$ is the expected number of alleles with frequencies lying between x and $x+dx$ ($0 < x < 1$); u is the mutation rate to a novel new allele, and $\theta = 4N_e u$. In this model, the shape of the allelic profile is determined by the value of θ. Fig. 4.1 depicts the allelic profiles for θ equal to 0.5, 1.0 and 2.0. When θ is small (i.e. in relatively small populations) most alleles are either very common or very rare. As the population size or mutation rate increases, a greater proportion of alleles occur at intermediate (0.05 to 0.30) frequencies. The curves also emphasise that in large populations a virtually infinite number of alleles occur at very low frequencies.

To find the expected number of alleles with frequencies greater than 0.50, 0.10, 0.05 and 0.01, $\phi(x)$ was integrated for each value of θ. The calculated values of n_a are given in Table 4.1. The important point of note is that populations usually contain only two to four alleles per locus in intermediate to high frequency (> 0.05); the other alleles are relatively rare.

It may be objected that the assumption of neutrality is unrealistic. Consequently, it is also desirable to consider deleterious variants on the one hand and alleles favoured by balancing selection on the other. The distributions of deleterious and neutral alleles will differ in that selection will reduce the number of common and intermediate variants and increase the number of relatively rare alleles. The distribution of

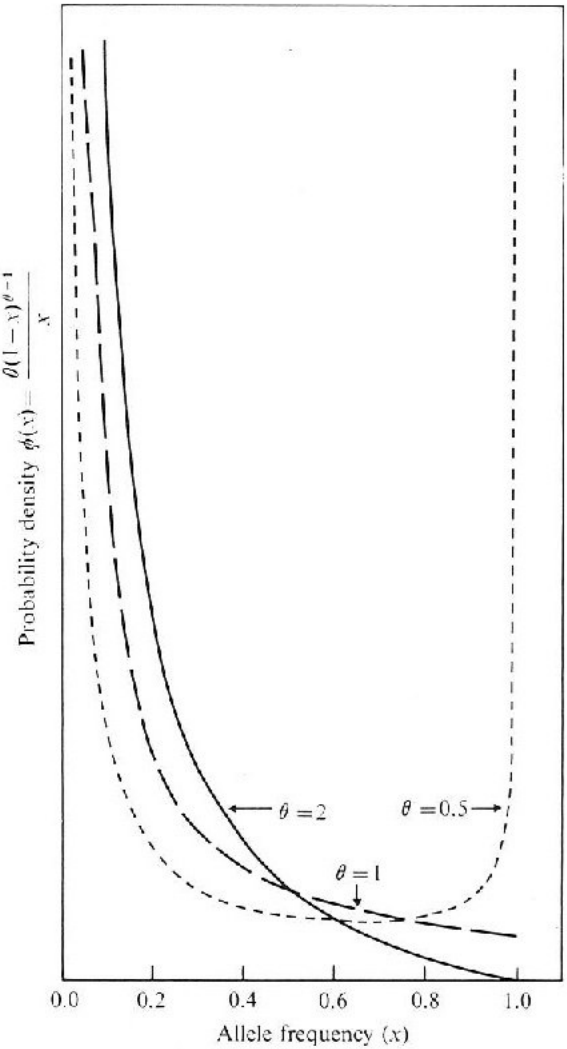

Fig. 4.1. Theoretical probability distributions of selectively neutral isoallele frequencies for various values of $\theta = 4 N_e u$. See text for further details.

neutral and selectively maintained alleles will also differ. As a case in point we will consider alleles maintained by over-dominance.

Over-dominant alleles. To study the effect of over-dominance on the allelic profile of a population we will use Kimura & Crow's (1970) model in which all heterozygotes have unit fitness and all homozygotes have relative fitness of $(1-s)$. Under this model

$$\phi(x) = Cx^{-1} \exp\left[-\sigma(x-f)^2 - \theta x\right]$$

where $\sigma = 2N_e s$, $1-f$ is the probability that an individual chosen at random is heterozygous at the locus in question and C is such that

$$\int_0^1 x\phi(x)\,dx = 1.$$

Table 4.1. *Expected number of neutral alleles (n_a) in various frequency classes*

Model	$\theta = 0.5$	$\theta = 1.0$	$\theta = 2.0$
$\phi(x)$	$(2x\sqrt{(1-x)})^{-1}$	x^{-1}	$2(1-x).x^{-1}$
$\int\phi(x)$	$\log_e[(1-\sqrt{(1-x)})/\sqrt{x}]$	$\log_e x$	$2[\log_e x - x]$
Frequency class	Expected number of alleles (n_a)		
> 0.50	0.85	0.69	0.39
> 0.10	1.83	2.30	2.81
> 0.05	2.19	3.00	4.09
> 0.01	3.00	4.61	7.23

Table 4.2. *Effective number of over-dominant alleles (n_e) and expected number of alleles (n_a) with frequency > 0.05, assuming constant selection differential s against homozygotes*

Selection differential	$\theta = 0.5$	$\theta = 1.0$	$\theta = 2.0$
	Effective number of alleles (n_e)		
$s = 0$	1.5	2.0	3.0
$s = 0.01$	9.0	14.0	21.0
$s = 0.10$	25.0	36.0	54.0
	Actual number of alleles (n_a)		
$s = 0$	2.2	3.0	4.1
$s = 0.01$	9.0	8.0	7.0
$s = 0.10$	3.0	2.0	0

Kimura & Crow (1970) have tabulated values of f for various values of s, N_e and u. Let us assume $u = 10^{-5}$, $s = 0.01$ or 0.1 and that θ takes values 0.5, 1.0 and 2.0 as before. The values of f^{-1}, the 'effective' number of alleles (n_e) or the number of equally frequent alleles required to give the observed level of heterozygosity, are shown in Table 4.2. It will be noted that over-dominance ($s > 0$) substantially increases the effective number of alleles over neutrality ($s = 0$).

Table 4.3. *Number of alleles per variable locus in populations of* Avena barbata, Drosophila willistoni, *and* Homo sapiens

Allelic frequencies	*Avena barbata*[1]	*Drosophila willistoni*[2]		*Homo sapiens*[3]
		Island	Continental	
> 0.50	1.00	0.97	1.00	0.93
> 0.10	1.84	1.39	1.43	1.91
> 0.05	1.71	1.71	1.98	2.21
> 0.01	2.00	2.80	2.92	2.61
Number of loci	5	16	16	14

[1] Clegg, M. T. & Allard, R. W. (1972). *Proc. natn. Acad. Sci. USA*, **69**, 1820–4.
[2] Ayala, F. J., Powell, J. R. & Dobzhansky, T. (1971). *Proc. natn. Acad. Sci. USA*, **68**, 2480–3.
[3] Lewontin, R. C. (1972). *Evol. Biol.*, **7**, 381–98.

Approximate values of C can be obtained as $C = \theta \exp[\sigma f^2]$, $\phi(x)$ plotted and the actual number (n_a) of common and intermediate alleles ($0.05 < P < 1$) estimated graphically as the area under the curve. The estimates obtained in this way are given in Table 4.2. It is clear that weak over-dominance leads to a marked increase in the number of intermediate and common alleles in the population. However, under this model strong over-dominance leads to maintenance of large numbers of alleles all at relatively low frequencies.

Distribution of alleles in natural populations

In Table 4.3, the above theoretical distributions can be compared with some published experimental data. It will be noted that on the average there are approximately two detectable alleles at each variable locus with frequency greater than 0.05.

The impact of a breeding system of predominant self-pollination is displayed by the *A. barbata* results. Under close inbreeding, it is more difficult to maintain large numbers of alleles in populations at low frequencies. Consequently, in *A. barbata* there were usually only two alleles per polymorphic locus and those occurred at intermediate frequencies. By contrast, populations of *D. willistoni* have a greater number of alleles in the class $0.01 < P < 0.05$. These presumably represent deleterious recessives and constitute the mutational load. This load seems to be less in human populations. However, many very rare variants have been found in human populations when the sample size

has been sufficiently large. For example, Hopkinson & Harris (1971) have detected 2.6 rare variants per locus for 12 genes specifying soluble enzymes in samples of 3000–13000 gametes. This represents a total average frequency of rare heterozygotes of 0.0017 or average frequency for rare alleles of 0.0008.

Assignment of priorities

From the above considerations it is obvious that the alleles in a population can be *arbitrarily* divided into those which are common ($P > 0.05$; usually less than four alleles) and those which are rare ($p < 0.05$; many alleles). In addition, any particular allele can be categorised as to whether it is *widespread* and occurs in many of the populations in the target area or *local* and restricted to one, or a few adjacent populations. These subdivisions yield four possible classes of alleles based on frequency and distribution. As it is possible to collect only a fraction of the many variants in a species, the critical question is: should we give priority to any particular class of alleles?

Considering first the common alleles or variants of widest occurrence; this class is the easiest to sample (indeed they are probably often present in modern agricultural varieties) and will be inevitably included in the sample regardless of strategy. Second, there is the class of widely occurring but locally rare alleles. From probability considerations, it follows that the sampling of this class depends only on the *total* number of plants taken from the target area, rather than how these are distributed between and within sampling sites. Consequently, the number of such alleles recovered is also largely independent of sampling strategy.

The third class of alleles, those which are *locally common*, are of greatest importance as regards sampling strategy. By contrast, the final class of alleles, which are both rare and restricted to a few populations, do not merit the same attention. We reach this conclusion from two arguments. First, the great majority of common alleles or allelic combinations (whether widespread or local) presumably represent adaptive variants maintained in populations by some form of balancing selection (Dobzhansky, 1970). Consequently, common variants are likely to be of far greater interest to plant breeders than rare variants which presumably represent either newly arising mutants or recombinations, or deleterious genes or gene combinations maintained in the population by a balance between mutation, migration or recombination and selection. Second, rare variants are obviously much more difficult

to collect than their common counterparts. Indeed, artificially induced mutation or in the case of gene blocks, artificially enhanced recombination, coupled with an effective selection screen, could be a far more efficient means of obtaining rare variants than randomly searching for them in natural populations. Consequently, the aim of plant exploration can be defined as *the collection of at least one copy of each variant occurring in the target populations with frequency greater than 0.05.* It follows that the most appropriate measure of genetic conservation is not simply the average number of alleles per locus, but the average number with population frequency greater than 0.05.

Definition of optimum sampling strategies

In discussing sampling strategies we will first consider a basic procedure for use in crops where we have no specific information on the distribution of variation in nature. We will then consider more sophisticated procedures for use where such information is available.

The basic strategy

Number of plants per site

By the term 'site' we wish to designate that area from which one bulked sample will be drawn and one set of ecological recordings kept in a single collecting instance.

The delineation of the most appropriate sampling area is usually relatively simple in annual crop plants. Populations of these species are harvested in bulk and a portion is used to sow the following crop. The high degree of mixing which occurs during harvesting and sowing each year ensures that all fields planted from a single seed source will not vary in genetic structure. Obviously, some genetic differentiation may take place by selective plant losses during the growth of the crop. However, in most circumstances, this will be minimal. Consequently, in annual crops the sampling target will be the individual field or farm if each farmer uses different seed stocks or a group of fields or farms if farmers use a common seed stock.

In wild or weed species the definition of the appropriate sampling unit is much more difficult. First, such species are often continuously distributed in nature and there are no artificial boundaries defining discrete, relatively homogeneous populations as there are in crop plants. Thus, the plant explorer must decide where in the continuum he will

collect and the total size of the area to be sampled (e.g. 1000, 10000 or 100000 m^2). Secondly, as emphasised by Bradshaw (Chapter 3), gene flow through pollen and seed dispersal is extremely limited in wild plants, particularly in comparison with cultivated annuals, and as a result, natural populations of wild species often show marked genetic differentiation over distances as small as a few metres. In these circumstances, the explorer must also decide whether to collect a single random sample from the chosen area or to sample the differentiated subpopulations separately. A number of populational and ecological factors will influence the explorer's decisions on these matters. The most important populational factor is the size of the interbreeding unit which is a function of the number and density of plants, the mating system and the level of pollen and seed dispersal. The most important ecological factor is the degree of environmental heterogeneity for such variables as soil type, aspect, slope, moisture regime, and associated flora. Consequently, the delineation of the appropriate sampling area is best left to the explorer at the time of collection since he is in a position to evaluate the factors involved.

Once an appropriate target population has been defined we are interested in obtaining as representative a sample of the common alleles in the population as possible. However, since excessive sampling of any one site limits the explorer's opportunities to discover and sample other, perhaps more interesting, sites the samples should be kept as small as possible. Therefore, it seems logical to define the optimum sample size per site to be the number of plants required to obtain, with 95 per cent certainty, all the alleles at a random locus occurring in the target population with frequency greater than 0.05. This definition is very similar to that of Oka (1969) based on '95 per cent of the genes distributed in the population with a frequency of 5 per cent'.

If complete information were available on the distribution of alleles in the target species, it would be possible to define an optimum sample size for each population. However, in species where such information is lacking and which are of greatest interest here, this is obviously impossible. The simplest and safest practical alternative for such species is to define, using the theoretical and experimental data on the distribution of alleles in populations discussed earlier (Fig. 4.1, Tables 4.1–4.3), a sample size which ensures that we meet our objective in the most common circumstances, and to apply this sample size to all populations sampled. This we will now proceed to do.

Sampling theory of allelic profiles. Consider a population in which two alleles (A_1, A_2) occur with frequencies p_1 and p_2, respectively. The probability that a random sample of n gametes contains at least one copy of each allele ($P[A_1^+, A_2^+]$) is given by:

$$P[A_1^+, A_2^+] = 1-(1-p_1)^n-(1-p_2)^n+(1-p_1-p_2)^n.$$

If $p_1 = 0.95$ and $p_2 = 0.05$ then a sample of 59 gametes is required to obtain at least one copy of each allele with 95 per cent certainty.

For a larger number of alleles, the exact probability expression becomes more cumbersome (Moran, 1968). For example, for four alleles:

$$P[A_1^+, A_2^+, A_3^+, A_4^+] = 1-\sum_{i=1}^{4}(1-p_i)^n+\sum_{i,j=1}^{4}(1-p_i-p_j)^n$$
$$-\sum_{i,j,k=1}^{4}(1-p_i-p_j-p_k)^n+(1-\sum_{i=1}^{4}p_i)^n;\ (k>j>i).$$

The behaviour of this and analogous expressions for increasing values of n was studied for five types of allelic profiles (Table 4.4). At one extreme, the first profile simulates the case of four alleles maintained by strong over-dominance. The other extreme is the fifth profile, obtained when three alleles are held in low but significant frequency under a mutation-selection balance. The intermediate cases are derived from the three profiles drawn in Fig. 4.1.

Table 4.4 shows the number of gametes (n) which must be sampled to be 95 per cent certain of obtaining at least one copy of each allele, and the probability, P, of achieving this objective given that $n = 100$. The results show that a surprisingly small random sample could be claimed as the 'minimum representative sample' under our criterion. It is clear from Table 4.4 that the value of n for which P achieves 0.95 is heavily dependent on the actual frequency of the rarest allele of interest to the collector. In general, alleles which are actively maintained in the population by balancing selection at intermediate frequencies require smaller samples than rarer deleterious recessives. Nevertheless, even in the limiting case of twenty alleles with frequency 0.05, a random sample of about 120 gametes will include with 95 per cent certainty one copy of each allele. Consequently, it seems safe to conclude that a random sample of 50–100 would be more than adequate under most circumstances.

This conclusion is further strengthened by the fact that in practice we do not sample individual gametes or even seeds, rather we sample single heads, panicles, other fruiting bodies or vegetative propagules. Individual heads or panicles will represent a number of different

Table 4.4. *Sample sizes (n) required to be 95 per cent certain of obtaining at least one copy of each common allele (frequency > 0.05) and probabilities (P) of achieving this objective if n = 100 for five contrasting models of allelic profiles*

Model	Balanced polymorphism 1	Neutral $\theta = 0.5$ 2	Neutral $\theta = 1.0$ 3	Neutral $\theta = 2.0$ 4	Mutation/ selection 5
Allele					
A_1	0.25	0.76	0.63	0.49	0.80
A_2	0.25	0.20	0.23	0.22	0.05
A_3	0.25	—	0.09	0.12	0.05
A_4	0.25	—	—	0.07	0.05
Remainder	—	0.04	0.05	0.10	0.05
Sample size (n)	19	15	37	43	80
Probability (P)	1.00	1.00	1.00	0.99	0.98

gametes in most crop species, particularly if they are outbreeding. Vegetative propagules carry at least the diploid complement.

It would also be instructive to know how many different alleles per locus we might expect to include on average in a random sample of say 50 or 100 gametes. Unfortunately, a general answer to this question is only available for the case of neutral loci. Ewens (1972) has shown that in a random sample of n gametes one would expect to capture

$$E(\mathrm{k}) \simeq \frac{\theta}{\theta} + \frac{\theta}{\theta+1} + \ldots + \frac{\theta}{\theta+n-1}$$

neutral alleles where $\theta = 4N_e u$. The sampling variance of k is

$$\mathrm{var}\,(\mathrm{k}) = E(\mathrm{k}) - \left[\frac{\theta^2}{\theta^2} + \frac{\theta^2}{(\theta+1)^2} + \ldots + \frac{\theta^2}{(\theta+n-1)^2}\right].$$

He has tabulated these moments for many values of θ and n. For illustrative purposes we will consider only two cases. When $\theta = 1$, the expected number of alleles in a random sample of 50 gametes is 4.5, with variance 2.9; whereas in a sample of 100 gametes, the expected number is 5.2 and the variance is 3.6. Thus on average one expects to collect some rare alleles additional to the 3.0 common alleles (Table 4.1) we define as the conservation target.

Biased sampling procedures. Bennett (1970*a*) has argued in favour of enriching the random sample with a biased sample of rare phenotypic variants if these occur in the target population. Such a procedure has

the obvious advantage that it allows the explorer to collect more of the observable variants in the population. Nevertheless, it suffers from a number of serious disadvantages. First, biased sampling will be far more time consuming than simple random sampling because of the need to search through large numbers of plants for the rare off-types. Consequently, by devoting time to the collection of rare types at one site, the explorer reduces his chances of sampling the common variants at other sites. Further, as we have argued before, despite the few spectacular cases of rare phenotypically distinguishable variants which later proved to be of great value (e.g. the opaque mutant in maize), common, currently adaptive alleles are more likely to be valuable to the plant breeder than rare variants. Second, there is the danger of collecting unrecognised disease specimens which could cause the whole collection to be destroyed by quarantine authorities. Third, if the random and biased samples are bulked, this will upset the correlation between that sample and its location and thus severely reduce any value the sample possesses for population research. On the other hand, if the samples are kept separate, this will greatly increase the effort which must be devoted to recording, multiplication and screening of the material. We conclude therefore that sampling of deviates should play only a minor role, if any, in genetic conservation.

In some instances an explorer may wish to sample only a small number of populations and can afford to spend more than the minimum amount of time at each site. In such situations, we would suggest that he should increase the size of the random sample at those sites which are highly polymorphic for morphological marker loci or highly variable for quantitative characters. Recent studies in *Avena fatua* and *A. barbata* indicate that populations which are highly polymorphic at morphological marker loci also tend to be highly polymorphic for those loci without observable phenotypic effects (Marshall & Allard, 1970). If this finding holds in general, it means that taking a larger random sample at sites highly polymorphic for observable variants offers a more efficient means of increasing the total variation in a collection than just collecting rare morphological variants at each site.

Number of sites

Since each site offers the prospect of sampling a different set of alleles, where we have no information on the distribution of variation in nature, the optimum number of sites to sample is the maximum possible. The

number of populations which can be sampled will be determined by such factors as the length of the collecting season, relative abundance of the target species, roughness of the terrain, etc. These factors place a strict upper limit on the number of samples which can be collected in a single season. Under ideal conditions this limit may be as high as 1000 samples, although in most circumstances it will be considerably less (Bennett, 1970*a*; Allard, 1970*a*).

Distribution of sites within the target area

As emphasised by Bennett (1965; 1970*b*) and Bradshaw (Chapter 3) it has been found repeatedly that the pattern of genetic differentiation within species is strongly correlated with environmental heterogeneity. Consequently, in species where there is no reliable information on the distribution of variation in nature, the explorer should aim to sample as broad a range of environments as possible.

There are many ways of achieving this objective. We will discuss two contrasting sampling strategies to illustrate the issues involved. In the first case the sampling sites are over-dispersed, that is, distributed more or less evenly over the target area (Fig. 4.2*a*). This procedure will capture the maximum amount of genetic variation associated with broad geographic differences in edaphic and climatic factors. In the second case the sampling sites are clustered in, say, groups of five, and the groups are more or less evenly distributed over the target area (Fig. 4.2*b*). This procedure will capture a significant amount of the variability associated with both geographic and *microgeographic* differences in environmental factors.

Each of these strategies has its advantages and disadvantages when considered in relation to a particular species or group of species. The first is the preferred procedure when sampling annual crop species. As noted earlier, such species are harvested in bulk and a portion is used to sow the following crop, often in a different field. The high degree of mixing associated with the harvesting and sowing procedures and the constant migration from field to field, virtually precludes the development of differentiated populations on a microgeographic scale. Thus, there is little point in collecting a number of populations within a region in such species.

The second strategy is likely to be the preferred procedure in wild or weed species. These species often grow on non-arable as well as arable land and as a result, occupy a greater range of habitats than

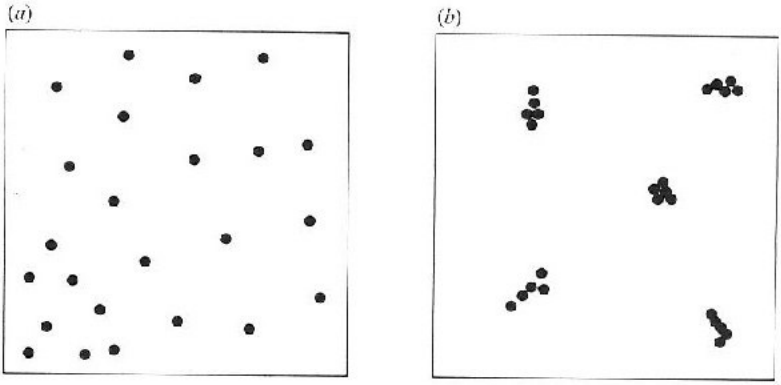

Fig. 4.2. Diagramatic representations of (*a*) evenly dispersed and (*b*) clustered sampling patterns.

crop plants. On the other hand, gene flow through pollen and seed dispersal is usually relatively small in wild species. As a result, a considerable portion of the total variability in such species has developed in response to, and is maintained by, heterogeneity in local environments. The clustering of samples has a number of other benefits in wild species. First, it reduces time spent in travelling between sites and permits the collection of a greater total number of samples in a given time. Second, it forces the explorer to search consciously for markedly different habitats within a region and avoids the possibility, which exists when only one sample is taken per region, that the explorer will unconsciously select, say, all lush or all arid sites. Third, it increases the value of the collections for studies of population structures and the adaptive significance of observable polymorphisms.

The procedures illustrated in Fig. 4.2 represent extremes and many intermediate strategies are possible. In many species it is conceivable that one strategy may be preferred in one area and an alternative strategy preferred in other areas. However, whatever strategy is used it is obviously important that the plant explorer assemble the fullest information possible, on the climate, soil and vegetation and their local variations, in the area to be explored before deciding on the distribution of sampling sites.

Modifications to the basic procedure

In a number of species which are targets for conservation we have at least some information on the distribution of variation in nature (Frankel & Bennett, 1970). This information can allow the explorer to

develop sampling procedures which are more efficient than the basic exploration strategy discussed above. For example, if it is known that the target species shows little, or no, inter-populational differentiation (e.g. *Phalaris tuberosa* in south-eastern Australia; McWilliam, Schroeder, Marshall & Oram, 1971) then the optimum strategy would be to collect a large number of individuals from a few populations. Alternatively, if it is known that the target species is highly differentiated and the majority of populations contain one, or a few, homozygous genotypes (e.g. *Trifolium subterraneum* in Western Australia; Gladstones, 1966), then the optimum strategy would be to collect a few individuals from a large number of populations. Consequently, we will now consider what modifications can be made to the basic strategy when something is known of the distribution of genetic variation in the target species.

Number of plants per site and number of sites

The critical problem here is to determine, given information on levels of variation within and between populations of the target species, the optimal allocation of sampling resources within and between sites so that the explorer recovers the maximum amount of variability. Oka (1969) has presented a useful formulation of this problem. He estimated the fraction of the total genetic variation in the target area captured by a particular sampling procedure, G, as

$$G = 1-\{(1-P)+P(1-p)^n\}^N,$$

where 1 = total genetic variation in the target area,
P = proportion of total variation represented by population or sampling site,
p = proportion of genetic variation per population or site represented by an individual plant (or seed),
N = number of populations or sites sampled,
n = number of plants (or seeds) sampled per site.

Oka described the effect of increasing n and N on G, assuming values of P and p suggested by experimental data on various kinds of target species (modern and primitive cultivars of rice and its wild relatives), in order to determine the values required for G to exceed 0.95. That is, he calculated the values of N and n required to ensure that the explorer captured more than 95 per cent of the total genetic variability in the target area.

However, this treatment does not immediately answer the problem of

Table 4.5. *Theoretical values of (i) optimal number of plants to sample per site and (ii) optimal number of sites to sample per day for a range of genetic models*

Population		Modern cultivars (1)	Primitive cultivars (2)	Wild relatives (3)	Outbreeding species (4)	Outbreeding species (5)
P		0.01	0.05	0.10	0.50	0.75
p		0.95	0.20	0.05	0.05	0.05
a/b ratio			(i) Number of plants per site (n)			
25		1	10	15	30	36
100		2	15	39	50	55
a	b		(ii) Number of sites per day (N)			
25	1	18	14	12	9	8
50	0.5	10	8	7	6	6

the *optimal* allocation of sampling resources. There are in fact many combinations of N and n which give $G > 0.95$. The critical point is which of these combinations is the most efficient. To answer this question we must assume a relationship between N and n consequent upon the limited availability of sampling resources. The simplest kind of relationship is:

$$E = N(a + bn),$$

where E represents the total effort available for sampling,
a represents that amount required for each site visited (i.e. transport, ecological records, site inspection, etc.), and
b represents that amount expended to sample one plant.

This relationship assumes that as bn is reduced, there is more time to travel to and sample other sites, allowing a greater density of sampling within the target area.

The problem now becomes to choose from values of N and n subject to the constraint E, those which maximise G, for particular values of p and P. The solutions to five cases are presented in Table 4.5. Three of these cases (examples 1–3) employ values of p and P given by Oka (1969) for modern and primitive cultivars of rice and their wild relatives. The remaining two use a much higher value of P as implied by recent studies of genic variability in cross-fertilising organisms (e.g. Lewontin, 1972; Selander & Johnson, 1972). Table 4.5(i) gives the values of n, the number of plants per site, for maximum efficiency. These estimates depend on the ratio of a/b and not on their actual values. However, the estimates of N, the number of sites sampled, does depend on ascribing

Table 4.6. *Number of days required to collect 95 per cent of variation ($G > 0.95$) using maximally efficient sampling procedure under a range of genetic models*

Population		Modern cultivars (1)	Primitive cultivars (2)	Wild relatives (3)	Outbreeding species (4)	Outbreeding species (5)
P		0.01	0.05	0.10	0.50	0.75
p		0.95	0.20	0.05	0.05	0.05
a	b	Days to collect 95% of variation				
25	1	16	5	5	1	1
50	0.5	30	8	5	1	1

actual values to a and b. For simplicity, we have used units of time, since this is the major factor limiting the explorer's activities once he is in the field. The values used are $E = 500$ minutes (i.e. it is assumed that the explorer spends 8.33 hours/day in sampling), $a = 25$ or 50 minutes (average time travelling between sites etc.) and $b = 1$ or 0.5 minutes (average time spent in collecting a single plant). Table 4.5(ii) gives the number of populations optimally sampled in a day.

It is clear from these results that partial genetic information may cause the explorer to alter markedly his sampling pattern from that previously discussed. In particular, these calculations reinforce the point that too much attention to one site sacrifices considerably the efficiency of sampling. Even when the amount of effort required per site compared with the amount to sample one plant is very high, we still find it most efficient to sample more sites per day and fewer plants per site.

Table 4.6 gives the number of days required to collect 95 per cent of the variation in a particular region, given that the population structure of the species is uniform in that area. It will be noted that in outbreeders where each population may contain 50–75 per cent of the total variation in the region as a whole, one day in the field will yield 95 per cent of the common alleles in the area. On the other hand, in plants where each population may contain as low as one per cent of the total variation in a region, the explorer must spend 30 days in the field to achieve the same objective. These calculations serve to emphasise that the greater the genetic differentiation in a species, the harder it is to sample effectively.

The above model suffers from a number of deficiencies. First, it assumes that each population contains a constant proportion of the total variation in the target species. This assumption is, of course, highly unrealistic. The level of variation generally varies quite markedly

from population to population. However, deviations from this assumption do not detract from the conclusions that the optimal strategy is to take as few plants as possible per site and sample a maximum number of sites. Obviously, if the level of within-population variation varies markedly from population to population the more populations sampled the greater is the probability the explorer will sample the more variable populations. Second, it is implicit in this model that equal numbers of plants will be collected at each site. In many instances it is neither practical nor wise to do this. For example, if the genetic structure of populations varied markedly between sites, a more efficient procedure would be to collect the minimum number of plants at uniform sites and increase sample size at highly variable sites. Third, the model ignores important practical realities. In particular, it takes no account of the fact that samples need to be distributed as well as conserved. In most species the practical minimum sample size is at least 25 plants per population unless the explorer is prepared to spend a great deal of time multiplying samples prior to distribution.

Despite these deficiencies, the above analysis does provide an insight into the relative efficiency of different sampling strategies. In particular it emphasises that collecting 200–300 plants per population as recommended by Bennett (1970*a*) and Allard (1970*a*) is an extremely wasteful procedure. In most circumstances, sample size should not exceed 50 plants per population and in no circumstances is it desirable to collect more than 100 plants per population.

Distribution of sites within the target area

Information on the population structure of the target species can also be important in determining the most effective distribution of sampling sites within the area to be explored. Recent studies of the patterns of population structure and differentiation for a variety of morphological and enzymatic marker loci in the slender wild oat (*Avena barbata*) in central California (Jain, 1969; Marshall & Allard, 1970; Clegg & Allard, 1972; Hamrick & Allard, 1972) illustrates this point well. Over large areas of California, populations of this species are of two types for the loci surveyed. Either they are predominantly homozygous and monotypic for one array of alleles, or they are homozygous and monotypic for an alternate array. Further, there is a strong correlation between the occurrence of the two genotypes and the major climatic regions of California. One type occurs primarily in the arid interior of

the state (Mediterranean warm summer region), while the other is limited to the more mesic coastal areas (Mediterranean cool summer region). However, highly polymorphic populations of *A. barbata*, displaying most of the allelic combinations of the loci surveyed, occur in the area of overlap between the major climatic zones. Obviously, if the aim of an exploration mission was to obtain as representative a sample of allelic combinations of these genes as possible in central Californian populations of *A. barbata*, the most efficient strategy would be to take all samples from the zone of overlap between the two major climatic regions and ignore the rest of the state.

Conclusions

It is obvious that information on the population structure of the target species can greatly improve the efficiency of sampling procedures by indicating the optimal numbers of plants per site, numbers of sites per region and the distribution of sampling sites within the target area. For estimating optimum sample sizes, only a minimum amount of information is required (e.g. estimates of the average number of alleles per locus or within and between family variances for five to ten populations). However, for determining the optimal distribution of sites within a region, substantial information is required before the optimum strategy can be defined with any degree of confidence. Nevertheless, some information is better than none, no matter how meagre. Consequently, the explorer should endeavour to collect all available information on the population structure of the target species prior to the planning of the collecting mission.

Maintenance of collected samples

There is little point in striving to collect the maximum amount of genetic variability in a species unless the material is adequately conserved to prevent the subsequent loss of the variation during the maintenance of the collection. We do not intend to consider in detail all aspects of the maintenance of germplasm collections. Rather, our purpose is to focus on one central issue which has been the subject of considerable debate in recent years – the relative merits of the maintenance of sexually propagated crops as stored seed compared with that by propagation in bulk populations or mass reservoirs (Simmonds, 1962; Frankel, 1970*b*; Allard, 1970*b*).

There are two main issues involved here:

(i) how best to conserve collections so that the maximum amount of variability is maintained for future generations?

(ii) how best to maintain collected variability for use by current plant breeders?

These represent distinct but interrelated issues. Obviously, if a particular collection is maintained in the long term as a bulk population it must be used in the short term in that form. However, if a collection is maintained in long-term storage as separate items, then these can be compounded into a bulk population or used individually as the situation demands.

*Base collections**

The question we must answer here is: will maintenance as seed or living collections preserve the most variation for future generations?

Simmonds (1962) originally proposed the use of mass reservoirs for long-term conservation because he regarded 'museum collections', as he termed them, as a wasting resource, often with very high rates of attrition (see also Bennett, 1965; 1968). At the time Simmonds wrote his article this was true in many instances. However, over the last decade there have been considerable advances in the technology of seed storage and it is now possible to maintain seed of many species for periods of 20–25 years or more before regeneration is necessary, as compared to three to five years in the past. At the same time, there has also been a greater appreciation of the need for genetic conservation, and first-class seed storage facilities have been established in several countries including the USA, USSR, Japan, Turkey, Philippine Islands and Mexico, and more are planned or being developed in Europe, Asia and Latin America. The increased availability of long-term storage facilities and the greater appreciation of their function means that attrition of samples in storage is no longer the problem it was in the past. We do not wish to imply that there will be no losses during storage. As noted by Allard (1970*b*) completely 'static' preservation is impossible and loss of genetic variation can occur through differential survival of genotypes in storage and selection, hybridisation and genetic drift during the rejuvenation process. However, we emphasise that these losses can be controlled by the judicial management of the collection and reduced to acceptable levels such that the stored seed even after 50 years and two

* For definition of the different types of collections see Chapter 37.

cycles of regeneration would contain a significant fraction of the original variation. The same cannot be said for 'mass reservoirs'. One of the most notable features of the composite cross populations was the rapid reduction in genetic variation in the early generations (Jain, 1961; Allard & Jain, 1962). For example, in Composite Cross v (CCv), the total genetic variation for height and heading date was reduced by 30–40 per cent between generations 4 and 14. By generation 19, the depletion in genetic variation for these characters was from 50 to 70 per cent. The conclusion seems inescapable that the bulk populations, or 'mass reservoirs' as they have come to be called, are of little value in *preserving* variation, potential or expressed, for future use. It is often argued that this tremendous loss of variability inherent in the use of bulk populations can be avoided by growing the bulk populations at a number of sites. It is envisaged that a different fraction of the variation would be preserved at each of the sites. However, the need to grow the populations at different sites negates to a large degree one of their main potential advantages, that is, there are few entries to grow and maintain. To take this argument to its ultimate conclusion, we might envisage growing CCv at 20 sites to ensure maintenance of as much variability as possible. Yet, it was originally derived by intercrossing 30, apparently pure-line, parental varieties. We would expect that the original 30 lines would be far easier to maintain than 20 bulk hybrid populations.

A second line of evidence which argues against the use of bulk populations for the long-term maintenance of variability comes from a study by Clegg, Allard & Kahler (1972) of allelic frequencies at four esterase loci in CCII and CCv. It might be assumed that complex interallelic and intergenic interactions would maintain different alleles in different CC populations, as these populations do not have all parents in common. However, this study indicated that the same pair of four locus-complementary, gametic types became predominant in both populations at Davis, California. In other words, both populations were approaching the same end point despite their different initial compositions. If it should prove to be at all general, this finding has important implications in the present context. In particular it means that bulk populations, although they may be developed from lines from different regions of the world, could after 20–40 generations become very similar. In other words, bulk populations when grown in a common environment not only retain a small portion of the potential variability expressed on crossing a set of parents, but they also tend to retain the same spectrum of variation.

Consequently, we would conclude that seed stored as separate items offers a far more effective means of preserving variation collected from nature than bulk populations. Further, as emphasised by Frankel (1970*b*), collections make it possible to study individual gene pools and to relate their characteristics to the environment from which they originated and this potential is lost if most or all the populations are bulked into mass reservoirs for maintenance.

Active or working collections

The question we must answer here is: should working collections, which represent the seed of lines held in storage for immediate distribution, be in the form of individual entries or bulk populations? Again we would argue for separate entries. However, in this case our arguments are based largely on the question of flexibility of use. If seed is supplied as individual entries, the breeder can decide how many or how few he requires and he uses them as he wishes.

It should not be assumed that bulk populations have no place in plant breeding. They are obviously a valuable adjunct to breeding for the same reason as they are unsuitable for use in the long-term maintenance of germplasm. They offer an effective means of selecting a range of locally adapted genotypes from the vast numbers generated by intercrossing parental lines or populations. However, it is important to realise that the requirements of each breeder will differ along with his local environment. Therefore he should be able to determine the parental composition of a bulk population, rather than receive a preformed pool from a conservation centre, since such a pool may have lost much variation potentially useful to him.

Summary and overall recommendations

1. There are definite limits to the numbers of samples which can be handled effectively in programmes for the conservation and utilisation of crop genetic resources. Consequently, there is a need for the coordinated and systematic planning of conservation programmes to ensure the preservation of the maximum amount of useful genetic variability while keeping the total number of samples within these limits.

2. The number of alleles per locus is the simplest measure of genetic diversity for the purposes of exploration and conservation. Linked,

co-adaptive gene complexes can be regarded as 'alleles' of a 'supergene' and in the present context are no different in principle from alleles at a single locus. Consideration of allelic profiles encountered in populations in theory and practice on the one hand, and the virtually infinite potential for genes to vary on the other, led us to focus on the class of alleles which are *locally common*, as critical in determining optimal sampling strategies. With this in mind, the aim of plant exploration was defined as the collection of at least one copy of each variant occurring in the target populations with a frequency greater than 0.05.

3. Where little, or no, information is available on the distribution of variation in nature, the optimal strategy is (i) to collect 50–100 individuals per site, (ii) to sample as many sites as possible within the time available, and (iii) to ensure that sampling sites represent as broad a range of environments as possible.

4. When partial information on the population structure of the target species is available it is possible to modify this basic procedure. A detailed analysis of alternative strategies indicates that, in many circumstances, even fewer than 50 plants per site, but more sites, should be sampled in order to maximise the genetic diversity in the collected sample.

5. There is no point in striving to collect the maximum amount of genetic variability in a species unless the material is adequately conserved. An important issue here is the relative merits of maintenance as stored seed compared with that by propagation in bulk populations. This question is resolved in favour of maintenance as stored seed on the basis of increased availability of first-class, long-term storage facilities and the now clearer insight into the goals of gene conservation.

References

Allard, R. W. (1970*a*). Population structure and sampling methods. In *Genetic Resources in Plants – their Exploration and Conservation* (eds O. H. Frankel and E. Bennett), *IBP Handbook*, no. **11**, pp. 97–107. Blackwell, Oxford and Edinburgh.

Allard, R. W. (1970*b*). Problems of maintenance. In *Genetic Resources in Plants – their Exploration and Conservation* (eds O. H. Frankel and E. Bennett), *IBP Handbook*, no. **11**, pp. 491–4. Blackwell, Oxford and Edinburgh.

Allard, R. W. & Jain, S. K. (1962). Population studies in predominantly self-pollinated species. II. Analysis of quantitative changes in a bulk hybrid population of barley. *Evolution*, **16**, 90–101.

Bennett, E. (1965). Plant introduction and genetic conservation: genecological aspects of an urgent world problem. *Scott. Pl. Breed. Stn Rec.*, pp. 27–113.

Bennett, E. (1968). *Record of the FAO/IBP Technical Conference on the Exploration, Utilization and Conservation of Plant Genetic Resources, 1967.* FAO, Rome.

Bennett, E. (1970*a*). Adaptation in wild and cultivated plant populations. In *Genetic Resources in Plants – their Exploration and Conservation* (eds O. H. Frankel and E. Bennett), *IBP Handbook*, no. **11**, pp. 115–29. Blackwell, Oxford and Edinburgh.

Bennett, E. (1970*b*). Tactics in plant exploration. *Genetic Resources in Plants – their Exploration and Conservation* (eds O. H. Frankel and E. Bennett), *IBP Handbook*, no. **11**, pp. 157–79. Blackwell, Oxford and Edinburgh.

Brock, R. D. (1971). The role of induced mutations in plant improvement. *Radiation Botany*, **11**, 181–96.

Clegg, M. T. & Allard, R. W. (1972). Patterns of genetic differentiation in the slender wild oat species *Avena barbata*. *Proc. natn. Acad. Sci. USA*, **69**, 1820–4.

Clegg, M. T., Allard, R. W. & Kahler, A. L. (1972). Is the gene the unit of selection? Evidence from two experimental plant populations. *Proc. natn. Acad. Sci. USA*, **69**, 2474–8.

Dobzhansky, T. (1970). *Genetics of the Evolutionary Process.* Columbia Univ. Press, New York.

Ewens, W. J. (1972). The sampling theory of selectively neutral alleles. *Theor. Pop. Biol.*, **3**, 87–112.

Frankel, O. H. (1967). Guarding the plant-breeder's treasury. *New Scientist*, **14**, 538–40.

Frankel, O. H. (1970*a*). Genetic conservation of plants useful to man. *Biological Conservation*, **2**, 162–9.

Frankel, O. H. (1970*b*). Genetic conservation in perspective. In *Genetic Resources in Plants – their Exploration and Conservation* (eds O. H. Frankel and E. Bennett), *IBP Handbook*, no. **11**, pp. 469–89. Blackwell, Oxford and Edinburgh.

Frankel, O. H. (1973). *Survey of Crop Genetic Resources in their Centres of Diversity. First Report.* FAO/IBP, Rome.

Frankel, O. H. & Bennett, E. (eds) (1970). *Genetic Resources in Plants – their Exploration and Conservation, IBP Handbook*, no. **11**, 554 pp. Blackwell, Oxford and Edinburgh.

Gladstones, J. S. (1966). Naturalised subterranean clover (*Trifolium subterraneum* L.) in Western Australia: the strains, their distributions, characteristics and possible origins. *Aust. J. Bot.*, **14**, 329–54.

Hamrick, J. L. & Allard, R. W. (1972). Microgeographical variation in allozyme frequencies in *Avena barbata*. *Proc. natn. Acad. Sci. USA*, **69**, 2100–4.

Hopkinson, D. A. & Harris, H. (1971). Recent work on isozymes in man. *Ann. Rev. Genet.*, **5**, 5–32.

Hurlbert, S. A. (1971). The non-concept of species diversity. A critique and alternative parameters. *Ecology*, **52**, 577–86.

Jain, S. K. (1961). Studies on the breeding of self-pollinating cereals. The composite cross bulk population method. *Euphytica*, **10**, 315–24.

Jain, S. K. (1969). Comparative ecogenetics of *Avena fatua* and *A. barbata* occurring in central California. *Evol. Biol.*, **3**, 73–118.

Kimura, M. & Crow, J. F. (1964). The number of alleles that can be maintained in a finite population. *Genetics*, **49**, 725–38.

Kimura, M. & Crow, J. F. (1970). *An Introduction to Population Genetics Theory*. Harper and Row, New York, 591 pp.

Lewontin, R. C. (1972). The apportionment of human diversity. *Evol. Biol.*, 7, 381–98.

Marshall, D. R. & Allard, R. W. (1970). Isozyme polymorphisms in natural populations of *Avena fatua* and *A. barbata*. *Heredity*, **25**, 373–82.

McWilliam, J. R., Schroeder, H. E., Marshall, D. R. & Oram, R. N. (1971). Genetic stability of Australian phalaris (*Phalaris tuberosa* L.) under domestication. *Aust. J. agric. Res.*, **22**, 895–908.

Moran, P. A. P. (1968). *An Introduction to Probability Theory*. Oxford Univ. Press, London.

Oka, Hiko-ichi (1969). A note on the design of germplasm presentation work in grain crops. *SABRAO Newsletter*, **1**, 127–34.

Selander, R. K. & Johnson, W. C. (1972). Genetic variation among vertebrate species. *Proc. XVII Int. Congr. Zool.*, pp. 1–31.

Simmonds, N. W. (1962). Variability in crop plants, its use and conservation. *Biol. Rev.*, **37**, 442–65.