JA1513

The sustainable preservation of biodiversity in self-pollinating plant species: sample size and collection methodology

1097

IMEP - PROGRAM OFFICE
29-7-94
RECEIVED ON

H.A. van Rheenen[1], S. Bhatti[2] and K.V.S. Rao[1]

Summary

Collectors of seed for *ex situ* conservation have adopted a common concept about the number of plants and seeds per plant that should go into a sample. The literature is reviewed and the conclusion drawn, that for self-pollinating species the number of plants per sample should be increased under certain conditions and that time can be saved, if in fields of uniform plant stands the sampling is spot wise rather than by transection.

Introduction

Biodiversity is a word much used, not only in common conversation, but also in professional and popular publications. A literature search using "AGRICOLA" database, and "biodiversity", "preservation" and "*ex situ*" as key words, yielded reference numbers as shown in Fig. 1. World-wide, dedicated efforts are being made to preserve the globally available botanical biodiversity (FAO, 1992). Examples of organizations being involved in such activities are the International Board for Plant Genetic Resources (IBPGR), Rome, Italy; National Bureau for Plant Genetic Resources (NBPGR), New Delhi, India; International Centres for Agricultural Research such as the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, India; and National Centres for Agricultural Research such as the National Agricultural Research Centre (NARC), Islamabad, Pakistan. The Genetic Resources Division (GRD) of ICRISAT has collected and preserves up-to-date 16 443 accessions of chickpea, 13 234 of groundnut, 12 024 of pigeonpea, 23 862 of millet, and 33 108 of sorghum respectively, while NARC maintains and evaluates a total of 15 991 accessions of 35 different plant species. Crucial for the proper preservation of the world's biodiversity of plant species is the collecting strategy utilised, and in particular the sample size and sampling method applied. Also of importance are the composition of the sample taken and the method of *ex situ* preservation.

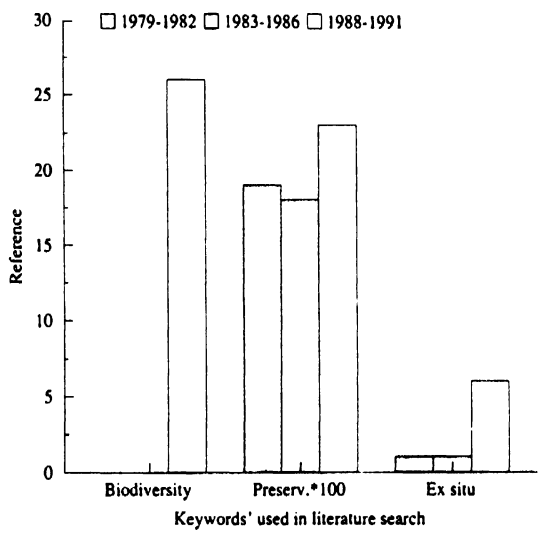

Fig. 1. Literature search on key words: biodiversity, preservation, *ex situ*

Source: Agricola Database

During an exploration trip in 1992 to collect seed samples of wild *Cicer* species in the mountain areas of North West Frontier Province of Pakistan in July 1992, the first two authors had the opportunity not only to collect seeds of the relatives of the chickpea or garbanzo bean: *C. microphyllum*, *C. nuristanicum* and *C. macranthum* (Bhatti *et al.*, 1993), but also to discuss the all important issue of collection strategy. During the collecting mission, direct competition in maintaining biodiversity was encountered, (Fig. 2), namely overgrazing. This prompted the authors to share their view upon collecting strategy with others, and the third author was called upon to check on the statistical aspects involved.

[1] International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) Patancheru 502 324, A.P., India
[2] Pakistan Agricultural Research Council, National Agricultural Research Centre, Islamabad, Pakistan

Fig. 2. Competition in collection

Sample size

A classic work on biodiversity of plant species and its conservation is Frankel and Bennet's "Genetic Resources in plants", published in 1970. In it, Allard (1970) addresses in a transparent manner the problems of sampling, caused partly by unknown factors such as variation between and within populations at different locations. His suggestion was to sample 200-300 plants per population, and a population, for practical purposes, meant a farmer's field for cultivated species and an area of occupied land, for wild species. Similarly, Bennett (1970) recommended sample sizes of 200-500 plants.

At a later date, Marshall and Brown (1975) [M/B] provided the theoretical basis for estimating the optimal sample size, and their conclusion was, that in most circumstances sample sizes should not exceed 50 plants per population and in no circumstance should more than 100 plants be collected per population. Many germplasm collectors follow a procedure of taking about 10-50 seeds from about 50-100 plants at each collection site (Hawkes 1991; Abebe Demissie 1991; Pundir and Mengesha 1991), and it appears M/B's theory is usually referred to, when the adopted methodology is described (Chapman, 1989). The approach M/B followed has been useful, but their sampling theory can be expanded somewhat.

They discuss several cases, simple and more complex, with allelic differences at one locus only, and present formulas to calculate how many gametes are required to obtain at least one copy of each allele with a definite probability say, for instance, 95%. If we assume, that more than one locus show allelic differences, their formulas need a raise to the power l (l = number of loci). This is of practical significance, as in a plant population differences may not be restricted to alleles at a single locus, but can also extend to different loci simultaneously. It will give the following expansion of their cases:

- The simplest case:

A population has two alleles A_1, A_2, with frequencies $p_1 = 0.95$ and $p_2 = 0.05$ respectively. Then the answer to the question how many gametes (n) are required to obtain at least one copy of each allele with 95% certainty is obtained from the modified formula of M/B.

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

Answer: 1 locus (M/B; l = 1): 59; 2 loci (l = 2): 72; 4 loci (l = 4): 86; 8 loci (l = 8): 99.

- A more complex case:

The population has 4 main alleles with frequencies $p_1 = 0.80$, $p_2 = 0.05$, $p_3 = 0.05$, $p_4 = 0.05$. Now the answer to the question how many gametes (n) are required to obtain at least one copy of each allele with 95% certainty is obtained from the modified extended formula of M/B:

$$P[A_1^+, A_2^+, A_3^+, A_4^+] = [1 - \sum_{i=1}^{4}(1-p_i)^n + \sum_{i,j=1}^{4}(1-p_i-p_j)^n - \sum_{i,j,k=1}^{4}(1-p_i-p_j-p_k)^n + (1-\sum_{i=1}^{4}p_i)^n]^l; (k>j>i) \quad [2]$$

Answer: 1 locus (M/B; l = 1): 80; 2 loci (l = 2): 94; 4 loci (l = 4): 107; 8 loci (l = 8): 120.

An important observation is to be made here: The above formulas used for determining sample size are applicable to Mendelian populations, and thus to cross pollinating crops. It appears that the number of plants required per collection site, as calculated by M/B for one locus, will increase, if more loci are considered. However, they rightly observe that in practice the sampling is not on gametes but on single heads, panicles, other fruiting bodies or vegetative propagules. In the case of cross-pollinating crops, single heads and panicles will represent indeed a number of different gametes, *but* for vegetative propagules and fruits of self-pollinating crops that is not the case. For self-pollinating crops simplified formulas can be used. Consider there are a alleles at each locus, and there are l loci. Then the number of variants will be

$$v = a^l \quad [3]$$

The frequency of the variants can be obtained by expanding the expression:

$$\prod_{j=1}^{l}\sum_{i=1}^{a} p_{ij} \quad [4]$$

Where p_{ij} is the frequency of the ith allele at the jth locus of a variant.

From the number and frequency of the variants we can calculate how many plants have to be taken in a sample to obtain at least one copy of each variant by using the formula

$$P[V_1, V_2, \ldots V_v] = 1 - \sum_{i=1}^{v}(1-f_i)^n \quad [5]$$

Where v is the number of variants (V), f_i is the frequency of the ith variant, and n is the number of plants required for a given probability.

For example, with 3 loci ($l = 3$), each having 2 alleles (a = 2) with different frequency (p_{ij}), the possible variants and their frequencies are as shown in Table 1.

Table 1. Variants and their frequency (f_i) in a population of plants with 2 different alleles (a = 2) of different frequency (p_i) at 3 different loci ($l = 3$, L_3). $p_{11} = 0.5$, $p_{21} = 0.5$, $p_{12} = 0.4$, $p_{22} = 0.6$, $p_{13} = 0.3$, $p_{23} = 0.7$, v = 8

Variant	Frequency (f_i)
L_{11}, L_{12}, L_{13}	$p_{11} \times p_{12} \times p_{13} = 0.06$
L_{11}, L_{12}, L_{23}	$p_{11} \times p_{12} \times p_{23} = 0.14$
L_{11}, L_{22}, L_{13}	$p_{11} \times p_{22} \times p_{13} = 0.09$
L_{11}, L_{22}, L_{23}	$p_{11} \times p_{22} \times p_{23} = 0.21$
L_{21}, L_{12}, L_{13}	$p_{21} \times p_{12} \times p_{13} = 0.06$
L_{21}, L_{12}, L_{23}	$p_{21} \times p_{12} \times p_{23} = 0.14$
L_{21}, L_{22}, L_{13}	$p_{21} \times p_{22} \times p_{13} = 0.09$
L_{21}, L_{22}, L_{23}	$p_{21} \times p_{22} \times p_{23} = 0.21$

The number of plants needed, to obtain at least one representative of each of the v variants with 95% certainty if the allelic frequencies are equal for all loci is shown in Table 2.

With unequal frequencies, the rare alleles determine the sample size. For example, if one of the alleles has a frequency of 0.05 and the remaining alleles have an equal distribution, the sample sizes required, corresponding with those shown in Table 2, are 59 for 1 locus, 1195 for 2 loci and 23510 for 3 loci, whatever the number of different alleles would be.

Table 2. Sample size required to obtain with 95% certainty at least one representative of each variant for plant populations with 1-3 different independent loci each with 2-4 different alleles at equal frequencies.
In brackets is the number of variants.

	Loci		
Alleles	1	2	3
2	5 (2)	15 (4)	38 (8)
3	10 (3)	44 (9)	167 (27)
4	15 (4)	89 (16)	454 (64)

The experienced collector will try to guess the existing variation - as also advocated by M/B - and then decide on sample size, guided by calculations as presented. M/B arrived at the conclusion that a random sample of 50-100 plants would be more than adequate under most circumstances. For cross-pollinating crops this may be adequate, but for self-pollinating species we suggest that the sample size may be small (<50) if the collector expects little variation, but may be large (up to 500 as Bennett 1970 suggested) if much variation is expected or observed.

Sampling method

Finally we wish to address one more important aspect of preserving biodiversity: the time factor in collecting samples. We present Fig. 2 to show that often the time for collection may run out rather rapidly. In fact, the collection sites for the wild *Cicer* species we visited were national disaster areas one month after our visit due to rainfall of extreme intensity that caused land slides.

Usually collectors will take their plant samples at regular intervals while walking through a collection site (Chapman, 1989; Hawkes, 1991; Abebe Demissie, 1991; Pundir and Mengesha, 1991). However, Marshall and Brown (1975) correctly realized, that populations of annual crops are harvested in bulk, that a portion is used to sow the following crop, and that the high degree of mixing during harvesting and sowing each year assures, that all fields planted from a single seed source will not vary in genetic structure. Therefore it may be concluded, that to collect seed samples by walking through a collection site is not required, if the plant stand of a field is uniform. The sampling time per plant can be reduced effectively by taking samples from all plants in one small spot. For crops like wheat, tef and chickpea the farmer will also appreciate, that his field is not traversed too much.

Conclusion

Marshall and Brown's (1975) suggestion that seed samples should not be taken from more than 100 plants, may be correct for cross-pollinating species but for self-pollinating species we propose that the number may range from 25-500 depending on the expected variation. We further suggest that time can be saved by sampling from small crop areas rather than from random transects through a farm if plant stands are regular. However, if *selective* sampling is the objective, the strategy will be different and the sampler may effectively search for desired types.

References

Abebe Demissie 1991. A decade of germplasm exploration and collecting activities by the Plant Genetic Resources Centre/Ethiopia. *In:* Engels, J.M.M., Hawkes, J.G. and Melaku Worede (eds.). *Plant genetic resources of Ethiopia.* pp. 202-217. Cambridge University Press, Cambridge, U.K.

Allard, R.W. 1970. Problems of maintenance. *In:* Frankel, O.H. and Bennett, E. (eds.). *Genetic resources in plants - their exploration and conservation*, IBP Handbook No. 11, pp. 491-444.

tion and conservation, IBP Handbook No. 11, pp. 491-444. Blackwell Scientific Publications, Oxford, U.K.

Bennett, E. 1970. Facties in plant exploration. *In:* Frankel O.H. and Bennett, E. (eds.). *Genetic resources in plants - their exploration and conservation*, IBP hand book No. 11, pp. 157-179. Blackwell, Oxford.

Bhatti, M.S., Anwar, R. and van Rheenen, H.A. 1993. Collection of wild *Cicer* species in Pakistan. International Chickpea Newsletter 28: 4-5

Chapman, C.G.D. 1989. Collection strategies for the wild relatives of field crops. *In:* Brown, A.H.D., Frankel, O.H., Marshall, D.R. and Williams, J.T. (eds.). *The use of plant genetic resources*. pp. 136-156. Cambridge University Press, Cambridge, U.K.

FAO (Food and Agriculture Organization of the United Nations) 1992. Sustainable Development and the Environment. FAO policies and actions - Stockholm 1972 - Rep 1992. 89pp.

Frankel, O.H. and Bennett, E. (eds.). 1970. *Genetic resources in plants - their exploration and conservation*. IBP Handbook No. 11, pp. 554. Blackwell Scientific Publications, Oxford, UK.

Hawkes, J.G. 1991. Theory and practice of collecting germplasm in a centre of diversity. *In:* Engels, J.M.M., Hawkes, J.G. and Melaku Worede (eds.). *Plant genetic resources of Ethiopia*. pp. 189-201. Cambridge University Press, Cambridge, U.K.

Marshall, D.R. and Brown, A.H.D. 1975. Optimum sampling strategies in genetic conservation. *In:* Frankel, O.H. and Hawkes, J.G. (eds.). *Crop Genetic Resources for today and tomorrow*. pp. 53-80. Cambridge University Press, Cambridge, U.K.

Pundir, R.P.S. and Mengesha, M.H. 1991. Collection strategies for food legumes. *In:* Smith, A. and Robertson, L. (eds.). *Legume genetic resources for semi-arid temperate environments*. pp.187-203. Proceedings of an international workshop, 19-24 June 1987, Cairo, Egypt: ICARDA.

Résumé

Conservation durable de la biodiversité chez les espèces végétales autopollinisatrices: taille des échantillons et méthodologie de collecte

Les collecteurs de semences pour la conservation *ex situ* ont adopté un concept commun au sujet du nombre de plantes et de semences par plante qui devrait composer un échantillon. On passe en revue la littérature et on en tire la conclusion que pour les espèces autopollinisatrices, le nombre de plantes par échantillon devrait être augmenté dans certaines conditions et qu'il est possible de gagner du temps, s'il s'agit de champs de populations végétales homogènes, en prélevant les échantillons au hasard et non transversalement.

Resumen

La conservación sostenible de la biodiversidad en las especies vegetales autopolinizadas: tamaño de la muestra y metodología de recolección

Los recolectores de semillas para la conservación *ex situ* han adoptado un criterio común acerca del número de plantas y semillas por planta que debe integrar una muestra. En el artículo se examina la bibliografía y se concluye que, en el caso de las especies autopolinizadas, el número de plantas por muestra se debe aumentar bajo ciertas condiciones, y que se puede ahorrar tiempo si en los campos de poblaciones uniformes de plantas se realiza un acertado muestreo al azar y no uno transversal.