Theor Appl Genet (1998) 97: 187–194 

**I. Zoro Bi · A. Maquet · J. Degreef · B. Wathelet**
**J. P. Baudoin**

# Sample size for collecting seeds in germplasm conservation: the case of the Lima bean (*Phaseolus lunatus* L.)

Received: 10 March 1998 / Accepted: 1 April 1998

**Abstract** The design of optimum sampling strategies integrating criteria of efficiency relevant to multilocus models and many target populations has been investigated with respect to the number of plants and the number of seeds per plant to be sampled for a Lima bean (*Phaseolus lunatus* L.) gene pool. This study, using five populations and six polymorphic enzyme loci, shows that the number of plants rather than the number of seeds collected per plant primarily determines the success of seed sampling, suggesting that plant number plays an essential part in maintaining the allelic multiplicity of predominantly selfing species like Lima bean. According to the results, it appears that among Lima bean populations an efficient sampling procedure is achieved by collecting 1–4 seeds from 200 to 300 plants. These sample sizes will retain 8–10 alleles, regardless of their frequencies. When we consider polymorphism at the 5% level, it is expected that sampling 10–80 plants will collect combinations of 4–8 alleles. Based on data from genetic and demographic studies, we suggest an efficient sampling scheme for Lima bean germplasm at both population and geographical levels.

Communicated by P. M. A. Tigerstedt

I. Zoro Bi (✉) · A. Maquet · J. Degreef · J. P. Baudoin
Faculté Universitaire des Sciences Agronomiques,
UER Phytotechnie des Régions Intertropicales,
2 Passage des Déportés,
B-5030 Gembloux,
Belgium
e-mail: Zoro.i@fsagx.ac.be
Fax: + 32 81 614 544

B. Wathelet
Faculté Universitaire des Sciences Agronomiques,
UER Chimie Biologique et Industrielle,
2 Passage des Déportés,
B-5030 Gembloux,
Belgium

**Key words** *Phaseolus lunatus* L. · Genetic diversity · Sampling strategies · Sample size · Field collection

## Introduction

Current threats to plant genetic resources and the biological and economic consequences of their decline have been well-documented and publicized (Falk 1990; Guillaumet and Morat 1990; Langaney et al. 1990; Guerrant 1992; Ellstrand and Elam 1993). The threats appear as diverse as the plants themselves (Falk 1990) and the genetic losses have become greater than the genetic gains, due to either the destruction of natural habitats or the widespread cultivation of very few genotypes (Brown 1983; Wilcox and Murphy 1985; Lande 1988; Harris and Silva-Lopez 1992). Consequently, the conservation of plant genetic resources by both ex situ and in situ methods attracts a growing public and scientific interest as well as support.

Although ex situ methods have contributed to the conservation of the genetic resources of several important crops, and consequently to their improvement (Frankel 1977), many problems related to these methods have been reported, such as inadequate sampling procedures during field collection and lack of representation of the total gene pool for a crop in gene banks (Altieri and Merrick 1987; Brown et al. 1997).

Sampling methods for field exploration have been investigated in many studies using probability models and population genetics theory (Ewens 1972; Marshall and Brown 1975; Oka 1975; Crossa 1989; Yonezawa and Ichihashi 1989; Falk 1991; Crossa et al. 1993; Lawrence et al. 1995 a). The results can provide rough predictions for unstudied plant species, suggesting basic management approaches. However, because there is considerable variation among plant species for ecological traits that influence the distribution of genetic variation (e.g. population size, degree of isolation,

pollination mechanism, seed dispersal mechanism), a genetically effective management strategy for one species may not be effective for another (Hamrick et al. 1991). Therefore, methodological and detailed studies of the ecology, population biology, genetics, and reproductive biology of the target plant species are essential to define efficient sampling strategies for germplasm collection from natural populations (Hawkes 1971; Lande 1988; Yonezawa and Ichihashi 1989; Falk 1990). Such studies, however, are lacking in the literature (Brown and Munday 1982; Moran et al. 1988; Watson et al. 1994; Padulosi et al. 1995; Lamboy et al. 1996).

As a model to develop a strategy for in situ conservation, a study has been conducted on the genetic diversity structure and the dynamics of wild Lima bean (*Phaseolus lunatus* L.) populations. The Central Valley of Costa Rica was selected as the geographical site due to the presence of about 400 natural *P. lunatus* populations and to their risk of extinction as a result of growing urbanization and intensive agriculture (Debouck 1987). This species has been chosen due to its alternative outbreeder-inbreeder behaviour: according to Baudoin (1991), the Lima bean is a self-compatible species with a mixed mating system, e.g. predominantly self-pollinating but with a fair amount of outcrossing (0–48%). Since in situ and ex situ conservations are complementary, and should not be viewed as alternatives (Falk 1987; Given 1987; Iwanaga 1995), investigations have also been carried out with the aim of designing the optimal sample size for field collections of *P. lunatus* germplasm. To achieve this objective, we used six allozyme markers and the theoretical model of Yonezawa and Ichihashi (1989) integrating both gene frequencies and mating-system parameters, e.g. the selfing rate and the inbreeding coefficient (Wright 1978). The main advantage of this model is its multilocus approach, since the aim of sampling for genetic conservation is both to obtain genetic variation at as many loci as possible and to retain alleles conferring specific adaptation to biotic and abiotic constraints (Brown 1978).

In the present paper we attempt to integrate current results with genetic and demographic data obtained from previous studies (Maquet et al. 1996; Degreef et al. 1997; Maquet et al. 1997) in order to design an optimal sampling strategy for Lima bean germplasm. Our analyses were limited to wild populations because of the genetic uniformity of many cultivated materials and the great influence of cropping practices, such as the deliberate selection of seeds from few plants for sowing during the next cultivation periods (Debouck 1988). In addition, with respect to the assayed loci, wild forms are generally more variable than their corresponding crops and have a larger and/or different spectra of alleles due to the founder effect of domestication and other genetic bottlenecks (Chapman 1989).

## Material and methods

### Plant materials

Five wild Lima bean populations located in the Central Valley of Costa Rica were sampled during the first trimestrial period of 1995, corresponding to the time of physiological maturity of the plant. These populations are representative of the four ecological zones in which the majority of the populations are located (Degreef et al. 1997) and were identified by several alpha-numeric codes. We sampled 10–60 plants per population and 4–6 racemes were collected per sampled plant according to the number of pod-bearing plants and their seed production. Reproductive individuals can bear several racemes (around 400) with 1–20 pods per raceme, each pod containing 1–5 seeds. One seed was randomly chosen per raceme for electrophoretic analysis, resulting in sample sizes ranging from 58 to 334 seeds per population.

### Electrophoretic analysis

To estimate the gene frequencies and inbreeding parameters, we selected six polymorphic loci resolved from six enzyme systems: alcohol dehydrogenase (Adh, E.C. 1.1.1.1), diaphorase (DIA, E.C. 1.8.1.4), fluorimetric esterase (fEST, E.C 3.1.1.-), malate dehydrogenase (Mdh, E.C. 1.1.1.37), phosphoglucomutase (Pgm, E.C. 5.4.2.2), and shikimate dehydrogenase (SKDH, E.C. 1.1.1.25). Enzyme extraction was done by grinding 5-day-old cotyledon tissues in a potassium phosphate buffer, pH 7.0, containing 20% sucrose (Sigma #S-8501), 5% PVP-40, 0.05% triton X-100 (Sigma #T-8532), 14 mM 2-mercaptoethanol (Sigma #M-6250), and 0.1 M $KH_2PO_4$. The pH value was adjusted to 7.0 with a solution of 5 N NaOH. Electrophoresis was performed using a horizontal 10% starch-gel (Sigma #S-4501) containing 3% sucrose. Two buffer systems were employed: continuous histidine-citrate, pH 6.1 (Kazan et al. 1993) for ADH and MDH, and discontinuous lithium-borate, pH 8.1/Tris-citrate, pH 8.4 (Murphy et al. 1990) for DIA, fEST, PGM, and SKDH. The techniques for gel electrophoresis and histochemical staining procedures are those reported elsewhere (Zoro Bi et al. 1996).

Loci were labelled sequentially, with those migrating closest to the anodal end designated as number 1 (Koenig and Gepts 1989). The first and the last stacks correspond to accession G25221, a Mexican wild form, considered as a standard for our analyses. The allozyme from this genotype is designated as 100 and all other allozymes are assessed according to their relative distance from this standard. The genetic control and the quaternary structure of the analysed enzyme systems were discussed previously (Maquet 1995; Zoro Bi et al. 1997 b). These studies showed that at least three loci, *Adh-2*, *Mdh-2* and *Pgm-2*, are independently inherited.

### Data analysis

For this study we used the theoretical model developed by Yonezawa and Ichihashi (1989). The sample size in this model is defined by m, the number of plants to be sampled, and n, the number of seeds to be taken from a single plant. It is assumed that the m plants are randomly sampled from the target population, the n seeds being collected from each plant.

If there are two alleles, $A_1$ and $A_2$, at a locus of interest, three genotypes, $A_1A_1$, $A_1A_2$ and $A_2A_2$, are expected to segregate in the target population. Representing the frequencies of these genotypes by $G_1$, $G_2$ and $G_3$ and assuming that plants in this population are self-pollinated with a selfing rate s, the genotypic array of the seeds borne on single plants can be formulated as shown in Table 1. If the two alleles are selectively neutral and this population is at genetic

**Table 1** Genotypic arrays in a target population and seed embryos on a single plant in the case of two alleles (adapted from Yonezawa and Ichihashi 1989)

| Plant in population | | Embryos on single plant | | | |
|---|---|---|---|---|---|
| Genotype | Frequency | Genotype | Frequency in selfed seeds | Frequency in outcrossed seeds | Total |
| $A_1A_1$ | $G_1$ | $A_1A_1$ | 1 | $p_1$[a] | $g_{11}$[c] $= s$[d] $+ (1 - s)p_1$ |
| | | $A_1A_2$ | 0 | $p_2$[b] | $g_{12} = (1 - s)p_2$ |
| | | $A_2A_2$ | 0 | 0 | 0 |
| $A_1A_2$ | $G_2$ | $A_1A_1$ | 1/4 | $p_1/2$ | $g_{21} = s/4 + (1 - s)p_1/2$ |
| | | $A_1A_2$ | 1/2 | 1/2 | $g_{22} = s/2 + (1 - s)/2$ |
| | | $A_2A_2$ | 1/4 | $p_2/2$ | $g_{23} = s/4 + (1 - s)p_2/2$ |
| $A_2A_2$ | $G_3$ | $A_1A_1$ | 0 | 0 | 0 |
| | | $A_1A_2$ | 0 | $p_1$ | $g_{32} = (1 - s)p_1$ |
| | | $A_2A_2$ | 1 | $p_2$ | $g_{33} = s + (1 - s)p_2$ |

[a] Frequency of allele $A_1$ in target population ( $= G_1 + G_2/2$)
[b] Frequency of allele $A_2$ in target population ( $= G_2/2 + G_3$)
[c] Genotypic frequency of selfed and outcrossed seed progenies according to the mother plant genotype ($g_{i\cdot}$) and their respective genotype ($g_{\cdot j}$)
[d] Selfing rate

equilibrium, the following relations hold between the frequencies of genotypes and alleles (Allard et al. 1968):

$$G_1 = p_1^2 + p_1p_2F;\ G_2 = 2p_1p_2(1 - F);\ G_3 = p_2^2 + p_1p_2F,$$

$$\text{with } F = s/(2 - s),$$

where F represents Wright's inbreeding coefficient.

The sampling-scheme efficiency is measured by the probability that all alleles included in the target population are collected, or in other words that none of the alleles are missed in the sampled seeds. For the two-allele case this probability is:

$$P = 1 - (G_1g_{11}^n + G_2g_{21}^n)^m - (G_2g_{23}^n + G_3g_{33}^n)^m,$$

where genotype frequencies $G_1$, $g_{11}$ etc. follow the definitions presented in Table 1, the sample size, $N$, being the product of m and n. The probability for the three-allele case is formulated in Yonezawa and Ichihashi (1989).

The probability ($P$) that all alleles at some independent loci are collected in a seed sample of size $N$ is given by the multiplication of the probabilities calculated for each of the loci using the above formula.

We used the multilocus mixed-mating model and estimation procedure of Ritland and Jain (1981), implemented by Ritland (1990), to estimate the population multilocus outcrossing rates, $t_m$, from which the selfing rates were deduced as $s = 1 - t_m$. The goodness-of-fit of the observed to the expected frequencies of progeny genotypes was tested for each locus using a $\chi^2$ test generated by MLT. Only data for loci showing a good fit to the mixed-mating model were used for $t_m$ estimation. This method provides not only greater statistical power, but is also free of the inbreeding equilibrium assumption (Ritland 1983).

## Results

All analysed loci showed two alleles (Table 2), which enabled the use of the formula suggested for the diallelic model (Yonezawa and Ichihashi 1989). A locus was considered polymorphic if more than one allele was observed, regardless of their frequencies. From the estimates of Table 2 we computed the probability $P$ for all combinations of allelic frequencies. Results based on nine fixed values of the total seed number showed that the highest probability in each of the five populations corresponds to the combination of $m = N$ and $n = 1$ (Table 3). The probability decreases steadily with decreasing values of m. This indicates that the number of plants per population, rather than the number of seeds per plant, is a primary determinant of the efficiency in the sampling scheme at least for the wild Lima bean.

Considering each population separately, we notice that the decrease in the number of plants (m) affects markedly the reduction of $P$ values especially when loci with unbalanced allelic frequencies are targeted (e.g. J48). By contrast, when populations possessing loci with well-balanced allelic frequencies are targeted (e.g. E88), a 0.90 probability is reached with lower fixed values of the total seed number ($N = 20$). These results indicate that the outcome of the sampling scheme depends not only on the number of collected plants but also on the target allelic frequencies.

From Table 3, we selected the number of seeds to be sampled in order to collect all alleles observed at the assayed loci in the target population with a 0.90 probability. The results are presented in Table 4. This table indicates that a larger number of plants per population (200–300) and few seeds per selected plant (1–4) are required to collect all alleles even with frequencies as low as 0.008 (e.g. J48, *Pgm-2*, $p_2$). Such a result could be explained by the fact that in a predominantly selfing species, seeds from a single plant are highly homozygous and homogeneous (Yonezawa and Ichihashi 1989; Lawrence et al. 1995 a).

**Table 2** Allele and genotype frequencies, average selfing rate (s) and Wright's inbreeding coefficient (F) for five wild Lima bean populations

| Population | $N^a$ | Locus | Allele frequency | | Genotype frequency | | | $s^b$ | F |
|---|---|---|---|---|---|---|---|---|---|
| | | | $p_1$ | $p_2$ | $G_1$ | $G_2$ | $G_3$ | | |
| E76 (10[c]) | 58 | *Adh-2*[d] | 0.888 | 0.112 | 0.862 | 0.051 | 0.086 | 0.926 | 0.742 |
| | 58 | *Dia-1* | 0.733 | 0.267 | 0.724 | 0.017 | 0.259 | | 0.956 |
| | 58 | *Mdh-2*[d] | 0.103 | 0.897 | 0.069 | 0.068 | 0.862 | | 0.631 |
| | 58 | *Pgm-2* | 0.776 | 0.224 | 0.759 | 0.034 | 0.207 | | 0.902 |
| | 58 | *Skdh* | 0.034 | 0.966 | 0.034 | 0 | 0.966 | | 1 |
| E84 (18) | 109 | *Adh-2* | 0.427 | 0.573 | 0.395 | 0.064 | 0.541 | 0.822 | 0.869 |
| | 105 | *Dia-1* | 0.933 | 0.067 | 0.905 | 0.057 | 0.038 | | 0.543 |
| | 109 | *Mdh-2*[d] | 0.784 | 0.216 | 0.758 | 0.052 | 0.189 | | 0.845 |
| | 89 | *Pgm-2* | 0.213 | 0.787 | 0.181 | 0.064 | 0.754 | | 0.808 |
| | 109 | *Skdh* | 0.042 | 0.958 | 0.037 | 0.009 | 0.954 | | 0.885 |
| E88 (21) | 151 | *Adh-2* | 0.430 | 0.570 | 0.338 | 0.185 | 0.477 | 0.613 | 0.623 |
| | 107 | *fEst-2* | 0.164 | 0.836 | 0.150 | 0.028 | 0.822 | | 0.898 |
| | 151 | *Mdh-2*[d] | 0.503 | 0.497 | 0.385 | 0.237 | 0.378 | | 0.525 |
| | 125 | *Pgm-2*[d] | 0.604 | 0.396 | 0.481 | 0.247 | 0.273 | | 0.484 |
| E100 (60) | 333 | *Adh-2* | 0.836 | 0.164 | 0.820 | 0.033 | 0.147 | 0.916 | 0.880 |
| | 334 | *Dia-1*[d] | 0.988 | 0.012 | 0.988 | 0 | 0.012 | | 1 |
| | 332 | *Mdh-2* | 0.452 | 0.548 | 0.380 | 0.144 | 0.476 | | 0.710 |
| | 244 | *Pgm-2* | 0.191 | 0.809 | 0.152 | 0.078 | 0.771 | | 0.748 |
| J48(51) | 198 | *Adh-2*[d] | 0.985 | 0.015 | 0.975 | 0.020 | 0.005 | 0.559 | 0.325 |
| | 198 | *Mdh-2* | 0.268 | 0.732 | 0.227 | 0.081 | 0.692 | | 0.794 |
| | 197 | *Pgm-2*[d] | 0.992 | 0.008 | 0.990 | 0.005 | 0.005 | | 0.665 |
| | 196 | *Skdh* | 0.054 | 0.946 | 0.051 | 0.005 | 0.944 | | 0.950 |

[a] Number of analysed seeds for the corresponding locus
[b] s values are deduced from $t_m$ estimated by the procedure of Ritland (1990): $s = 1 - t_m$
[c] Number of sampled plants
[d] Locus showing a good fit to a mixed-mating model and used to compute the selfing rate for a corresponding population

## Discussion

The results obtained from this study reveal that the plant number plays an essential role in maintaining the allelic multiplicity of predominantly selfing species such as the Lima bean. Indeed, among wild *P. lunatus* populations, an efficient sampling procedure is achieved by collecting 1–4 seeds from 200 to 300 plants. These sample sizes will retain eight (e.g. J48, Table 3) to ten (e.g. E76, Table 3) alleles from the analysed loci, even with some of them occurring at frequencies less than 1%. Similar sample sizes were reported by Brown (1978), Crossa (1989), Yonezawa and Ichihashi (1989) and Lawrence et al. (1995 a, b) using theoretical models. However, when we consider polymorphism at the 5% level, it is expected that sampling 1–2 seeds from 10 to 80 plants, according to tested population and allele frequencies, will collect all allelic combinations (data not shown). These results fit with those reported by Marshall and Brown (1975), Oka (1975), Brown and Briggs (1991), and Crossa et al. (1993).

Results indicate that for population E88, when $N \geq 40$, all combinations give a 0.90 probability of collecting all alleles, even from a single plant. Such results are not realistic since, in the wild Lima bean, the low gene flow (Hardy et al. 1997) and the predominantly selfing mating system result in populations being spatially structured (Zoro Bi et al. 1997 a). Consequently, for the Lima bean, a reliable genetic sampling should include in each selected population a high number of plants regularly distributed over the site.

The number of populations to sample can be inferred from recent studies carried out by Maquet et al. (1996, 1997) on of the genetic structure of *P. lunatus*. These studies revealed that Nei's gene diversity (Ht) did not vary markedly either between the two recognized gene pools (Mesoamerican: $Ht = 0.160$ and Andean: $Ht = 0.240$) or between the two botanical forms (wild: $Ht = 0.353$, and cultivated $Ht = 0.331$). The same studies reported a high gene differentiation between populations ($Gst = 0.775$). On the basis of these results, we recommend the collection of populations of wild and cultivated forms from as many distinctive ecological sites as possible. However, more accurate information concerning the relationships between genetic and ecological variations in Lima bean is necessary to confirm this recommendation.

Because within-site environmental heterogeneity should exert selective pressures on a population

**Table 3** Probability (*P*) of collecting all alleles from the analysed enzyme loci with different combinations of the number of plants (m) and seeds per plant (n) in five wild Lima bean populations

| Combination plant/seed number | | | Probability (*P*) for each population | | | | |
|---|---|---|---|---|---|---|---|
| m | n | $N^a$ | E76 | E84 | E88 | E100 | J48 |
| 20 | 1 | 20 | 0.470 | 0.574 | 0.993 | 0.225 | 0.089 |
| 10 | 2 | 20 | 0.197 | 0.307 | 0.960 | 0.111 | 0.032 |
| 5 | 4 | 20 | 0.051 | 0.130 | 0.896 | 0.044 | 0.013 |
| 2 | 10 | 20 | 0.005 | 0.038 | 0.801 | 0.010 | 0.006 |
| 1 | 20 | 20 | 0.001 | 0.018 | 0.744 | 0.003 | 0.004 |
| 40 | 1 | 40 | 0.773 | 0.865 | 1 | 0.407 | 0.288 |
| 20 | 2 | 40 | 0.508 | 0.647 | 0.998 | 0.242 | 0.134 |
| 10 | 4 | 40 | 0.241 | 0.404 | 0.990 | 0.131 | 0.065 |
| 8 | 5 | 40 | 0.176 | 0.339 | 0.985 | 0.105 | 0.054 |
| 5 | 8 | 40 | 0.082 | 0.232 | 0.972 | 0.064 | 0.038 |
| 4 | 10 | 40 | 0.056 | 0.195 | 0.966 | 0.050 | 0.033 |
| 2 | 20 | 40 | 0.018 | 0.122 | 0.950 | 0.024 | 0.024 |
| 1 | 40 | 40 | 0.007 | 0.088 | 0.939 | 0.013 | 0.021 |
| 80 | 1 | 80 | 0.951 | 0.984 | 1 | 0.648 | 0.608 |
| 20 | 4 | 80 | 0.563 | 0.745 | 1 | 0.273 | 0.229 |
| 10 | 8 | 80 | 0.322 | 0.561 | 0.999 | 0.170 | 0.151 |
| 8 | 10 | 80 | 0.259 | 0.510 | 0.999 | 0.146 | 0.136 |
| 4 | 20 | 80 | 0.126 | 0.392 | 0.998 | 0.093 | 0.106 |
| 2 | 40 | 80 | 0.067 | 0.324 | 0.996 | 0.064 | 0.092 |
| 1 | 80 | 80 | 0.043 | 0.288 | 0.996 | 0.049 | 0.086 |
| 100 | 1 | 100 | 0.977 | 0.994 | 1 | 0.729 | 0.710 |
| 50 | 2 | 100 | 0.865 | 0.952 | 1 | 0.505 | 0.483 |
| 25 | 4 | 100 | 0.663 | 0.833 | 1 | 0.330 | 0.313 |
| 20 | 5 | 100 | 0.587 | 0.782 | 1 | 0.288 | 0.275 |
| 10 | 10 | 100 | 0.360 | 0.624 | 1 | 0.189 | 0.197 |
| 5 | 20 | 100 | 0.200 | 0.505 | 0.999 | 0.127 | 0.158 |
| 4 | 25 | 100 | 0.167 | 0.477 | 0.999 | 0.114 | 0.151 |
| 2 | 50 | 100 | 0.102 | 0.417 | 0.999 | 0.085 | 0.135 |
| 1 | 100 | 100 | 0.073 | 0.385 | 0.999 | 0.071 | 0.128 |
| 140 | 1 | 140 | 0.995 | 0.999 | 1 | 0.839 | 0.840 |
| 70 | 2 | 140 | 0.940 | 0.986 | 1 | 0.626 | 0.645 |
| 35 | 4 | 140 | 0.792 | 0.927 | 1 | 0.430 | 0.465 |
| 28 | 5 | 140 | 0.731 | 0.896 | 1 | 0.380 | 0.419 |
| 14 | 10 | 140 | 0.525 | 0.780 | 1 | 0.263 | 0.320 |
| 7 | 20 | 140 | 0.347 | 0.676 | 1 | 0.191 | 0.266 |
| 2 | 70 | 140 | 0.183 | 0.572 | 1 | 0.129 | 0.227 |
| 1 | 140 | 140 | 0.149 | 0.547 | 1 | 0.115 | 0.219 |
| 200 | 1 | 200 | 0.999 | 1 | 1 | 0.927 | 0.932 |
| 100 | 2 | 200 | 0.982 | 0.998 | 1 | 0.755 | 0.798 |
| 50 | 4 | 200 | 0.896 | 0.978 | 1 | 0.552 | 0.638 |
| 40 | 5 | 200 | 0.852 | 0.964 | 1 | 0.495 | 0.592 |
| 20 | 10 | 200 | 0.688 | 0.901 | 1 | 0.357 | 0.482 |
| 10 | 20 | 200 | 0.524 | 0.830 | 1 | 0.272 | 0.417 |
| 8 | 25 | 200 | 0.479 | 0.811 | 1 | 0.253 | 0.404 |
| 4 | 50 | 200 | 0.371 | 0.763 | 1 | 0.212 | 0.376 |
| 2 | 100 | 200 | 0.309 | 0.735 | 1 | 0.189 | 0.361 |
| 1 | 200 | 200 | 0.275 | 0.719 | 1 | 0.177 | 0.354 |
| 300 | 1 | 300 | 1 | 1 | 1 | 0.980 | 0.983 |
| 150 | 2 | 300 | 0.998 | 1 | 1 | 0.879 | 0.918 |
| 75 | 4 | 300 | 0.967 | 0.997 | 1 | 0.700 | 0.812 |
| 60 | 5 | 300 | 0.943 | 0.994 | 1 | 0.641 | 0.776 |
| 15 | 20 | 300 | 0.711 | 0.941 | 1 | 0.383 | 0.615 |
| 10 | 30 | 300 | 0.645 | 0.924 | 1 | 0.344 | 0.591 |
| 5 | 60 | 300 | 0.557 | 0.900 | 1 | 0.301 | 0.565 |
| 3 | 100 | 300 | 0.514 | 0.888 | 1 | 0.283 | 0.554 |
| 2 | 150 | 300 | 0.490 | 0.882 | 1 | 0.274 | 0.549 |
| 1 | 300 | 300 | 0.465 | 0.875 | 1 | 0.264 | 0.543 |
| 340 | 1 | 340 | 1 | 1 | 1 | 0.988 | 0.990 |
| 170 | 2 | 340 | 0.999 | 1 | 1 | 0.908 | 0.942 |
| 85 | 4 | 340 | 0.979 | 0.999 | 1 | 0.745 | 0.854 |
| 68 | 5 | 340 | 0.961 | 0.997 | 1 | 0.687 | 0.823 |
| 17 | 20 | 340 | 0.761 | 0.961 | 1 | 0.421 | 0.675 |
| 10 | 34 | 340 | 0.683 | 0.944 | 1 | 0.370 | 0.646 |
| 5 | 68 | 340 | 0.607 | 0.927 | 1 | 0.330 | 0.624 |
| 4 | 85 | 340 | 0.589 | 0.923 | 1 | 0.321 | 0.619 |
| 2 | 170 | 340 | 0.549 | 0.914 | 1 | 0.304 | 0.609 |
| 1 | 340 | 340 | 0.527 | 0.910 | 1 | 0.295 | 0.605 |
| 400 | 1 | 400 | 1 | 1 | 1 | 0.995 | 0.996 |
| 200 | 2 | 400 | 1 | 1 | 1 | 0.940 | 0.966 |
| 100 | 4 | 400 | 0.989 | 1 | 1 | 0.800 | 0.900 |
| 80 | 5 | 400 | 0.978 | 0.999 | 1 | 0.745 | 0.875 |
| 20 | 20 | 400 | 0.818 | 0.979 | 1 | 0.475 | 0.749 |
| 10 | 40 | 400 | 0.733 | 0.965 | 1 | 0.407 | 0.716 |
| 5 | 80 | 400 | 0.671 | 0.955 | 1 | 0.370 | 0.698 |
| 4 | 100 | 400 | 0.656 | 0.952 | 1 | 0.362 | 0.694 |
| 2 | 200 | 400 | 0.624 | 0.947 | 1 | 0.346 | 0.687 |
| 1 | 400 | 400 | 0.606 | 0.944 | 1 | 0.338 | 0.683 |

[a] Total number of seeds arbitrarily fixed ($N = mn$)

(Huenneke 1991), sampling should include representatives from any distinctive microhabitats in a site (e.g. topographical features or soil heterogeneity).

In order to sample more genetic diversity among Costa Rican populations, the collecting missions in the field should be carried out two or three times at least, since the populations are characterized by a demographic instability in human disturbed habitats, the number of reproductive individuals in a target population fluctuating greatly (from 0 to > 50 plants) over the years (Degreef et al. 1997). In addition, according to these authors, after their dispersal, the seed germination rate equals 84% per year, the 16% remaining seeds staying in the soil seed bank. About 56% of seeds from the soil seed bank germinate and reach juvenile individuals stage the next year. Considering that the germination rate remains constant the following year, at least 97% of the initial set of seeds will germinate after 3 years. Alleles from individuals born from these latest germinating seeds should be sampled through repeated collection trips.

Our objective was to estimate the minimum sample size required to conserve at a high probability all the genetic variation revealed from six enzyme loci. In practice, the conservationist may well need to take samples that are larger than the values we suggested so as to provide, for example, enough material for distribution to other investigators, for evaluation and regeneration of material, or for testing the seed germination.

Further investigations should focus on an increased number of enzyme loci or more polymorphic genetic

**Table 4** Number of sampled seeds corresponding to a 0.90 probability of collecting all alleles from the analysed loci in a target population

| Population | $N^a = 20$ | | $N = 40$ | | $N = 80$ | | $N = 100$ | | $N = 140$ | | $N = 200$ | | $N = 300$ | | $N = 340$ | | $N = 400$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | m[b] | n[c] | m | n | m | n | m | n | m | n | m | n | m | n | m | n | m | n |
| E76 | –[d] | – | – | – | 80 | 1 | 100 | 1 | 140 | 1 | 200 | 1 | 300 | 1 | 340 | 1 | 400 | 1 |
| | | | | | | | | | 70 | 2 | 100 | 2 | 150 | 2 | 170 | 2 | 200 | 2 |
| | | | | | | | | | | | 50 | 4 | 75 | 4 | 85 | 4 | 100 | 4 |
| | | | | | | | | | | | | | 60 | 5 | 68 | 5 | 80 | 5 |
| E84 | – | – | – | – | 80 | 1 | 100 | 1 | 140 | 1 | 200 | 1 | 300 | 1 | 340[e] | 1 | | |
| | | | | | | | 50 | 2 | 70 | 2 | 100 | 2 | 150 | 2 | 170 | 2 | | |
| | | | | | | | | | 35 | 4 | 50 | 4 | 75 | 4 | 85 | 4 | | |
| | | | | | | | | | 28 | 5 | 40 | 5 | 60 | 5 | 68 | 5 | | |
| | | | | | | | | | | | 20 | 10 | 15 | 20 | 17 | 20 | | |
| | | | | | | | | | | | | | 10 | 30 | 10 | 34 | | |
| | | | | | | | | | | | | | 5 | 60 | 5 | 68 | | |
| | | | | | | | | | | | | | | | 4 | 85 | | |
| | | | | | | | | | | | | | | | 2 | 170 | | |
| | | | | | | | | | | | | | | | 1 | 340 | | |
| E88 | 20 | 1 | 40[e] | 1 | | | | | | | | | | | | | | |
| | 10 | 2 | 20 | 2 | | | | | | | | | | | | | | |
| | 5 | 4 | 10 | 4 | | | | | | | | | | | | | | |
| | | | 8 | 5 | | | | | | | | | | | | | | |
| | | | 5 | 8 | | | | | | | | | | | | | | |
| | | | 4 | 10 | | | | | | | | | | | | | | |
| | | | 2 | 20 | | | | | | | | | | | | | | |
| | | | 1 | 40 | | | | | | | | | | | | | | |
| E100 | – | – | – | – | – | – | – | – | – | – | 200 | 1 | 300 | 1 | 340 | 1 | 400 | 1 |
| | | | | | | | | | | | | | | | 170 | 2 | 200 | 2 |
| J48 | – | – | – | – | – | – | – | – | – | – | 200 | 1 | 300 | 1 | 340 | 1 | 400 | 1 |
| | | | | | | | | | | | | | 150 | 2 | 170 | 2 | 200 | 2 |
| | | | | | | | | | | | | | | | | | 100 | 4 |

[a] Total arbitrarily fixed number of seeds ($N = mn$)
[b] Number of plants per population
[c] Number of sampled seeds per plant
[d] Sample sizes for which $P < 0.9$
[e] Values of $N$ for which all combinations of m and n give $P > 0.9$ for the concerned populations. Such values indicate the optimal sample size

markers like microsatellites (Schaal et al. 1991). Such a study could set limits on either sample size, time, labour or cost. Indeed, our study does not include a hypothesis concerning the benefit-cost analysis, namely the allocation of resources associated to the sampling of multiple individuals or multiple populations. Note that such investigations can be only partially guided by biological considerations since judging how to allocate resources invokes other questions of societal values and priorities as well (Falk 1991).

The model of Yonezawa and Ichihashi (1989) is designed for populations at genetic equilibrium. The study of Maquet et al. (1996) indicated that some Costa Rican wild Lima bean populations did not fit with the Hardy-Weinberg law. To correct bias induced by this factor, we used the mixed mating model of Ritland (1983) in estimating the autogamy rate, but even this should not be sufficient to optimize our results. As far as we know, no model based on the assumption of genetic disequilibrium has been reported in the literature.

There are currently many discussions about the way to maintain a representative genetic diversity while reducing the amount of material stored in gene banks, and therefore to constitute the core collection (Crossa et al. 1993; Balfourier et al. 1994; Yonezawa et al. 1996; Brown et al. 1997; Crossa and Vencovsky 1997). The results reported in the present paper could partly assist in this task.

**Acknowledgements** This study was financed by the Belgian Administration of Cooperation for Development (A.G.C.D., Brussels, Belgium) and supervised by the International Plant Genetic Resources Institute (I.P.G.R.I., Roma, Italy). We thank Dr. Xavier Vekemans (Laboratoire de Génétique et d'Écologie Végétales, Université Libre de Bruxelles, Belgium) for helpful comments on the manuscript and Mrs. Moreman (Bibliothèque Centrale, FUSAGx, Gembloux, Belgium) for checking the English style. A scholarship to the first author was provided by the Scientific Research Ministry of the Ivory Coast.

## References

Allard RW, Jain SK, Workman PL (1968) The genetics of inbreeding populations. Adv Genet 14:55–131

Altieri MA, Merrick LC (1987) In situ conservation of crop genetic resources through maintenance of traditional farming systems. Econ Bot 41:86–96

Balfourier F, Charmet G, Grand-Ravel C (1994) Conservation of allelic multiplicity and genotypic frequency by pooling wild populations of perennial ryegrass. Heredity 73 : 386–396

Baudoin JP (1991) La culture et l'amélioration de la légumineuse alimentaire *Phaseolus lunatus* L. en zones tropicales. Fac Sci Agron and CTA, Gembloux, Belgium and Wageningen, The Netherlands

Brown AHD (1978) Isozymes, plant population genetic structure and genetic conservation. Theor Appl Genet 52 : 145–157

Brown AHD, Briggs JD (1991) Sampling strategies for genetic variation in ex situ collections of endangered plant species. In: Falk DA, Holsinger KE (eds) Genetics and conservation of rare plants. Oxford University Press, Oxford, UK, pp 99–119

Brown AHD, Munday J (1982) Population-genetic structure and optimal sampling of land races of barley from Iran. Genetica 58 : 85–96

Brown AHD, Brubaker CL, Grace JP (1997) Regeneration of germplasm samples: wild versus cultivated plant species. Crop Sci 37 : 7–13

Brown WL (1983) Genetic diversity and genetic vulnerability—an appraisal. Econ Bot 37 : 4–12

Chapman CGD (1989) Collection strategies for the wild relatives of field crops. In: Brown AHD, Frankel OH, Marshall DR, Williams JT (eds) The use of plant genetic resources. Cambridge University Press, Cambridge, USA, pp 263–279

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

Crossa J, Hernandez CM, Bretting P, Eberhart SA, Taba S (1993) Statistical genetic considerations for maintaining germplasm collections. Theor Appl Genet 86 : 673–678

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

Debouck DG (1987) *Phaseolus* germplasm collection in Central Costa Rica. Report No AGPG/IBPGR: 87/52, International Board for Plant Genetic Resources (IBPGR), Roma, Italy

Debouck DG (1988) *Phaseolus* germplasm exploration. In: Gepts P (ed) Genetic resources of *Phaseolus* beans. Kluwer Academic Publishers, London, UK, pp 3–29

Degreef J, Baudoin JP, Rocha OJ (1997) Case studies on breeding systems and its consequences for germplasm conservation. 2. Demography of wild Lima bean populations in the Central Valley of Costa Rica. Genet Res Crop Evol 44 : 429–438

Ellstrand NC, Elam DR (1993) Population genetic consequences of small population size: implications for plant conservation. Annu Rev Ecol Syst 24 : 217–242

Ewens WJ (1972) The sampling theory of selectively neutral alleles. Theor Pop Biol 3 : 87–112

Falk DA (1987) Integrated conservation strategies for endangered plants. Nat Areas J 7 : 118–123

Falk DA (1990) Integrated strategies for conserving plant genetic diversity. Ann Missouri Bot Gard 77 : 38–47

Falk DA (1991) Joining biological and economic models for conserving plant genetic diversity. In: Falk DA, Holsinger KE (eds) Genetics and conservation of rare plants. Oxford University Press, Oxford, UK, pp 209–223

Frankel OH (1977) Genetic resources. Ann NY Acad Sci 287 : 332–344

Given DR (1987) What the conservationist requires of ex situ collections. In: Branwell D, Hamann O, Heywood V, Synge H (eds) Botanic gardens and the world conservation strategy. Academic Press, London, UK, pp 103–116

Guerrant EO, Jr (1992) Genetic and demographic considerations in the sampling and reintroduction of rare plants. In: Fiedler PL, Jain SK (eds) Conservation biology: the theory and practice of nature conservation, preservation, and management. Chapman and Hall, London, UK, pp 321–344

Guillaumet J-L, Morat P (1990) Menaces sur la flore. Cahiers d'Outre-Mer 43 : 343–362

Hamrick JL, Godt MJW, Murawski DA, Loveless MD (1991) Correlations between species traits and allozyme diversity: implications for conservation biology. In: Falk DA, Holsinger KE (eds) Genetics and conservation of rare plants. Oxford University Press, Oxford, UK, pp 75–86

Hardy O, Dubois S, Zoro Bi I, Baudoin JP (1997) Gene dispersal and its consequences on the genetic structure of wild populations of Lima bean (*Phaseolus lunatus*) in Costa Rica. Plant Genet Res Newslett 109 : 1–6

Harris LD, Silva-Lopez G (1992) Forest fragmentation and the conservation of biological diversity. In: Fiedler PL, Jain SK (eds) Conservation biology: the theory and practice of nature conservation, preservation, and management. Chapman and Hall, London, UK, pp 197–237

Hawkes JG (1971) Conservation of plant genetic resources. Outlook Agric 6 : 248–253

Huenneke LF (1991) Ecological implications of genetic variation in plant populations. In: Falk DA, Holsinger KE (eds) Genetics and conservation of rare plants. Oxford University Press, Oxford, UK, pp 31–44

Iwanaga M (1995) IPGRI strategy for in situ conservation of agricultural biodiversity. In: Engels JMM (ed) In situ conservation and sustainable use of plant genetic resources for food and agriculture in developing countries. IPGRI/DSE, Bonn-Röttgen, Germany, pp 13–26

Kazan K, Muehlbauer FJ, Weeden NF, Ladizinsky G (1993) Inheritance and linkage relationships of morphological and isozyme loci in chickpea (*Cicer arietinum* L.). Theor Appl Genet 86 : 417–426

Koenig R, Gepts P (1989) Allozyme diversity in wild *Phaseolus vulgaris*: further evidence for two major centers of genetic diversity. Theor Appl Genet 78 : 809–817

Lamboy WF, Yu J, Forsline PL, Weeden NF (1996) Partitioning of allozyme diversity in wild populations of *Malus sieversii* L. and implications for germplasm collection. J Am Soc Hort Sci 121 : 982–987

Lande R (1988) Genetics and demography in biological conservation. Science 241 : 1455–1460

Langaney A, Nadot R, van Blijenburgh H (1990) S.O.S. génomes! Connaître, gérer et sauver les patrimoines génétiques. Cahiers d'Outre-Mer 42 : 534–546

Lawrence MJ, Marshall DF, Davies P (1995 a) Genetics of genetic conservation. I. Sample size when collecting germplasm. Euphytica 84 : 89–99

Lawrence MJ, Marshall DF, Davies P (1995 b) Genetics of genetic conservation. II. Sample size when collecting seed of cross-pollinating species and the information that can be obtained from the evaluation of material held in gene banks. Euphytica 84 : 101–107

Maquet A (1995) Etude de la diversité génétique de la légumineuse alimentaire *Phaseolus lunatus* L. par l'analyse de caractères morphophysiologiques et de marqueurs protéiques. PhD thesis, Fac Univ Sci Agron, Gembloux, Belgium

Maquet A, Zoro Bi I, Rocha OJ, Baudoin JP (1996) Case studies on breeding systems and its consequences for germplasm conservation. 1. Isoenzyme diversity in wild Lima bean populations in central Costa Rica. Genet Res Crop Evol 43 : 309–318

Maquet A, Zoro Bi I, Delvaux M, Wathelet B, Baudoin J-P (1997) Genetic structure of a Lima bean base collection using allozyme markers. Theor Appl Genet 95 : 980–991

Marshall DR, Brown AHD (1975) Optimum sampling strategies in genetic conservation. In: Frankel OH, Hawkes JG (eds) Crop genetic resources for today and tomorrow. Cambridge University Press, Cambridge, UK, pp 53–80

Moran GF, Bell JC, Eldridge KG (1988) The genetic structure and the conservation of five natural populations of *Pinus radiata*. Can J For Res 18 : 506–514

Murphy RW, Sites JW, Jr, Buth DG, Haufler CH (1990) Proteins. I. Isozyme electrophoresis. In: Hillis DM, Moritz C (eds) Molecular systematics. Sinauer Associates Inc, Sunderland, Massachusetts, USA, pp 45–126

Oka HI (1975) Consideration on the population size necessary for conservation of crop germplams. In: Matsuo T (ed) JIBP Synthesis, pp 57–63

Padulosi S, Caruso T, Barone E (1996) Taxonomy, distribution, conservation and uses of *Pistacia* genetic resources. IPGRI Rome, Italy

Ritland K (1983) Estimation of mating systems. In: Tanksley SD, Orton TJ (eds) Isozymes in plant genetics and breeding, Part A. Elsevier, Amsterdam, The Netherlands, pp 289–302

Ritland K (1990) A series of FORTRAN computer programs for estimating plant mating systems. J Hered 81 : 235–237

Ritland K, Jain S (1981) A model for the estimation of outcrossing rate and gene frequencies using *n* independent loci. Heredity 47 : 35–52

Schaal BA, Leverich WJ, Rogstad SH (1991) A comparison of methods for assessing genetic variation in plant conservation biology. In: Falk DA, Holsinger KE (eds) Genetics and conservation of rare plants. Oxford University Press, Oxford, UK, pp 123–134

Watson LE, Uno GE, McCarty NA, Kornkven AB (1994) Conservation biology of a rare plant species, *Eriocaulon kornickianum* (Eriocaulaceae). Am J Bot 81 : 980–986

Wilcox BA, Murphy DD (1985) Conservation strategy: the effects of fragmentation on extinction. Am Nat 125 : 879–887

Wright S (1978) Variability within and among natural populations. University of Chicago Press, Chicago, USA

Yonezawa K, Ichihashi H (1989) Sample size for collecting germplasms from natural populations in view of the genotypic multiplicity of seed embryos borne on a single plant. Euphytica 41 : 91–97

Yonezawa K, Ishii T, Nomura T, Morishima H (1996) Effectiveness of some management procedures for seed regeneration of plant genetic resource accessions. Genetica 43 : 517–524

Zoro Bi I, Maquet A, Wathelet B, Baudoin JP (1996) New isozyme markers for the study of the gene pool of *Phaseolus lunatus* L. Annu Rep Bean Improv Coop 39 : 247–248

Zoro Bi I, Maquet A, Baudoin JP (1997 a) Spatial patterns of allozyme variants within three wild populations of *Phaseolus lunatus* L. from the central valley of Costa Rica. Belg J Bot 129 : 149–155

Zoro Bi I, Maquet A, Wathelet B, Baudoin JP (1997 b) Genetic control of alcohol dehydrogenase, malate dehydrogenase, and phosphoglucomutase isozymes in Lima bean (*Phaseolus lunatus* L.). Plant Breed 116 : 181–185