PROC. V INT. SYMP. BIOL. CONTR. WEEDS, BRISBANE, AUSTRALIA, 1980: 103–11

OPTIMAL SAMPLING STRATEGIES IN THE BIOLOGICAL CONTROL OF WEEDS

D.R. Marshall, J.J. Burdon and A.H.D. Brown[1]

ABSTRACT

The objective of exploration in the target plant in biological programs for weed control is the collection of the maximum amount of genetic variation at loci controlling host/parasite or host/predator interactions within a limited number of samples. The sampling strategy which is optimal in attaining this objective is: (i) to collect 50 to 100 individuals/site; (ii) to sample as many sites as possible within the time available; and (iii) to ensure that the sample sites represent as broad a range of environments as possible. Modifications to this strategy which may be necessary in the field are also discussed.

INTRODUCTION

The need for comprehensive genetic sampling of potential control agents has long been recognized and stressed in the biological control literature (e.g., Clausen 1936, Wilson 1965, DeBach 1974). As a result, considerable emphasis has been given to developing efficient sampling procedures which meet this need (Goeden *et al.* 1974, Wapshere 1975, Winder and Harley 1976). Such sampling is necessary to ensure the discovery and collection of the most effective control agents. This entails the discovery of not only the most effective species, but the most effective biotypes of those species. As emphasized by Winder and Harley (1978) if the target weed consists of a number of ecotypes or races which differ in their susceptibility to the chosen predators or parasites, then biotypes matching each of these ecotypes or races must be used.

In contrast, the sampling of the range of genetic variation in the target weed has received much less attention. Yet, adequate sampling of the weed is as important as adequate sampling of control agents if the aim is to ensure the best possible matching of the two. Substantial evidence from crop and pasture plants indicates that marked genetic differences in susceptibility to both fungal pathogens (Day 1974) and animal parasites and predators (Maxwell *et al.* 1972) are common in plant species. The best possible matching of a control agent and its host therefore requires the collection and testing of the full range of genetic variation in both species.

As a result, one major requirement of collections of the target weed made for evaluation with potential control agents is that they contain as much relevant genetic variation as possible. However, scientists involved in such programs invariably have limited resources at their disposal and can collect, and more particularly, evaluate, only a limited number of samples. A second major requirement, therefore, is that the total number of samples in such collections should be kept within this practical limit. The problem facing the collector is to satisfy these conflicting requirements. That is, to define a sampling procedure which permits the collection of the maximum amount of genetically relevant variation within a specified and limited number of samples.

The definition of efficient exploration programs requires decisions at two levels. At the first level are decisions concerning the regions or geographical areas

[1] C.S.I.R.O. Division of Plant Industry, P.O. Box 1600, Canberra City, A.C.T. 2601, Australia.

to be explored. In the case of the target weed decisions at this level pose no problem. The area to be explored is the total area where the weed is to be controlled.

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

(i) number of individuals to sample per site;
(ii) the total number of sites to sample; and
(iii) the distribution of sampling sites within the exploration region.

Our aim here is to develop optimum sampling strategies in terms of the above parameters which yield the maximum relevant genetic variability in a specified total number of samples. We developed such a procedure previously for use in genetic conservation of crop and forage plants (Marshall and Brown 1975, 1980). We use the same approach here taking into account differences in the nature of weed versus cultivated plant populations and the end use of the samples collected.

However, in order to determine how to collect it is first necessary to specify what to collect, that is, to define what is relevant genetic variability in this context.

BASIC SAMPLING STRATEGY

What to collect

From the viewpoint of biocontrol, the only genetic variation of relevance in the target weed is that concerned with differences in resistance or tolerance to control agents. In particular, the explorer is interested in collecting all variants in the target weed which differ significantly in their susceptibility to particular potential control agents. Such variants may represent alleles at a single locus or allelic combinations at two or more loci.

In framing optimum sampling strategies, such variants can be arbitrarily divided into four classes depending on their population frequency and distribution in the target weed (Marshall and Brown 1975, 1980, Brown 1978). First, the variants in any one population can be classed into those which are *common* (frequency greater than, say, five per cent) and those which are *rare* (frequency less than five per cent). Next, each variant is classed as to whether it is *widespread* and occurs in many populations or *local* and restricted to one or a few adjacent populations.

Clearly, the widely occurring common variants will be easiest to collect and will be included in the sample at high frequency regardless of strategy. Similarly, the probability of including the widely occurring but locally rare variants in a sample will depend largely on the total number of individuals collected rather than on how those individuals are distributed between and within sample sites. Therefore, the number of this class of variants recovered will also be largely independent of sampling strategy.

The third class of variants, those which are common within populations but localized in their distribution, will obviously be more difficult to collect than their counterparts with a widespread distribution. Further, locally common variants with resistance or tolerance to control agents are likely to be of importance in biological control programs because they could act as foci for the rapid spread of resistant strains of the weed. Because such variants require a special collecting effort and because they are likely to be of practical import,

they warrant special attention from collectors.

The final class of variants, those which are rare within populations and restricted in their distribution among populations, will also be difficult to collect. They are also likely to be of considerable interest and importance because they could, like the more common variants, act as foci for the spread of resistant strains of the weed. However, in this case, the buildup would presumably be less rapid allowing time for control agents to evolve to meet the new challenge or the introduction of additional agents to control the resistant strain.

On these bases we would argue that the variants which merit priority in the framing of sampling strategies are those specifying resistance to control agents which are *locally common*; i.e., those which occur in a restricted number of populations at high frequency. As yet there is only limited evidence of the existence of such locally common variants in target weeds in biological control programs. One of the best of these examples is the differential susceptibility of the broad intermediate and narrow leaf width races of skeleton weed (*Chondrilla juncea*; Compositae) in Australia to two introduced control agents—a rust *Puccinia chondrillina* Bubak and Syd. (Uredinales:Pucciniaecae) (Hasan 1972) and a gall mite, *Aceria chondrillae* Can. (Acari:Eriphyidae) (Caresche and Wapshere 1974).

However, in recent years, electrophoretic studies of variation in soluble proteins indicate that locally common alleles can make up a substantial fraction (of the order of 35 per cent) of variants at a locus in a range of plant species (Brown 1978). To emphasize this point we present in Tables 1 and 2 data for allele frequencies at two enzyme loci (*alcohol dehydrogenase*, ADH, and *glutamate-oxalate transaminase*, GOT, loci) in 13 populations of Paterson's curse, *Echium plantagineum* (Boraginaceae), a common weed occurring throughout much of temperate Australia, which is the target weed in a biological control program. Clearly, two of the six detectable alleles at these loci (*Adh* S_s and *Got* S) fall into the locally common class and occur in appreciable frequencies (> 5 per cent) in one or a few populations. Further, it is unlikely

Table 1. Incidence and frequency of the four ADH alleles in 13 *Echium plantagineum* populations.

Site	No. of individuals sampled	Per cent allele incidence and frequency			
		F	I	S	Ss
Ariah Park		88.5	4.5	8.0	0
Corowa	50	88.0	0	13.0	0
Deniliquin	43	100.0	0	0	0
Gundagai	50	96.0	1.0	3.0	0
Hall	49	78.6	4.6	16.8	0
Holbrook	50	89.0	1.0	10.0	0
Mildura	26	99.4	0.6	0	0
Northam	49	92.9	0	0	7.1
Three Springs	51	99.5	0.5	0	0
Thurgoona	50	99.0	0	1.0	0
Toowoomba	42	100.0	0	0	0
Wombat	50	70.5	7.5	22.0	0
Young	50	72.5	4.0	23.5	0

Table 2. Incidence and frequency of the two GOT alleles in 13 *Echium plantagineum* populations.

Site	No. of individuals sampled	Per cent allele incidence and frequency	
		F	S
Ariah Park	50	100.0	0
Corowa	50	100.0	0
Deniliquin	43	100.0	0
Gundagai	50	100.0	0
Hall	49	100.0	0
Holbrook	50	100.0	0
Mildura	26	100.0	0
Northam	49	90.3	9.7
Three Springs	51	100.0	0
Thurgoona	50	100.0	0
Toowoomba	42	100.0	0
Wombat	50	100.0	0
Young	50	98.0	2.0

that the patterns of allelic frequencies at loci of interest to collectors concerned with biological control would differ from patterns at other loci. Indeed, evidence from agricultural studies suggest that locally common alleles are prevalent at loci specifying disease and pest resistance in crop plants and their wild relatives (e.g., Qualset 1975, Khush 1977).

How to collect

(i) Number of plants per site

Since excessive sampling at any one site limits an explorer's opportunities to sample other sites and hence his capacity to collect locally common variants, the sample at each site should be kept as small as practical. On this basis the optimum sample size is defined (Marshall and Brown 1975, 1978) to be the number of individuals required to collect, with 95 per cent certainty, at least one copy of all the common variants in a population. The use of the 95 per cent probability level is arbitrary, chosen simply because it is a widely accepted limit among biologists.

When the distribution of allelic frequencies in the target species is unknown, no calculation of the optimum sample size is possible. The only alternative is to define, using the available theoretical and experimental data on population structure in plants a sample size which ensures that the above objective is met under most circumstances, and to collect a sample of that size from all populations sampled.

Consider a population in which two variants, say A_1 and A_2, occur with frequencies p_1 and p_2, respectively. The probability that random sample of n gametes contains at least one copy of both variants symbolized $P[A_1^+, A_2^+]$, is

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

If $p_1 = 0.95$ and $p_2 = 0.05$ then 59 gametes are required to obtain at least one copy of each variant with 95 per cent probability. If $p_1 = 0.90$ and $p_2 = 0.10$ then only 29 gametes are required to achieve the same objective. As the number

of variants in the population increases, the exact expression for the probability of obtaining at least one copy of each rapidly becomes more complex (Moran 1968). However, a lower bound for this probability, of sufficient accuracy for our proposes is,

$$P[A_1^+, a_2^+, \text{----}, Ak^+] \geq 1 - \sum_{i=1}^{k} (1-p_i)^n$$

To illustrate the behaviour of this expression the number of gametes, n, required to be 95 per cent sure of obtaining one copy of each common variant ($p_i > 0.05$) are given in Table 3 for three populations which differ in their composition. Obviously, these examples are contrived. Nevertheless, they cover the spectrum of gene frequency profiles encountered in practice, at least for electrophoretic variants (Brown 1978). Table 3 also gives for each population the probability, P, of collecting at least one copy of each common variant given a sample size of 100 gametes.

Table 3. Sample sizes (n) required to be 95 per cent certain of obtaining at least one copy of each common variant (frequency $p_i > 0.05$) and the probability (P) of achieving this objective given a sample size of 100 gametes.

Example	1	2	3
Variant			
A_1	0.25	0.63	0.80
A_2	0.25	0.23	0.05
A_3	0.25	0.09	0.05
A_4	0.25	0	0.05
Remainder	0	0.05	0.05
Sample sizes			
n	16	34	80
Probabilities			
P	1.00	1.00	0.98

It is clear from these calculations that the sample size necessary to achieve our defined objective is heavily dependent on the frequency of the rarest variant of interest to the collector. Variants which are maintained at intermediate or high frequencies in populations require smaller samples than their rarer counterparts. Nevertheless, these data reinforce the conclusion that a surprisingly small random sample is generally required to satisfy our criterion. Indeed, in the limiting case of twenty variants with frequency of 0.05, a random sample of 117 gametes will include with 95 per cent certainty one copy of each. We therefore conclude, particularly since explorers collect panicles or pods and other fruiting structures from individual plants and *not gametes*, that a random sample from 50 individuals would be sufficient in most cases.

(ii) The number of sites

Since the aim is to collect as many variants as possible in the target species, and since each new site offers the prospect of sampling a new set of locally common variants then the optimum number of sites to sample is the maximum

possible. In practice, the number of populations which can be sampled will be determined by the length of the collecting season, relative abundance of the target species and roughness of the terrain. These factors place a strict upper limit on the number of samples which can be collected in a single mission and it is important therefore, that they be distributed to maximum advantage.

(iii) Distribution of sites within the target area

It is increasingly evident in plant species that the patterns of genetic differentiation result from, and are strongly correlated with, environmental heterogeneity (Bennett 1965, Allard 1970, Antonovics 1971, Bradshaw 1975, Snaydon 1978). To maximize the number of locally common variants collected the explorer should, therefore, sample as many different environments as possible.

As we noted previously (Marshall and Brown 1975, 1980) there are many ways of achieving this objective. In practice, the key decision which has to be made is the relative emphasis to be given to macrogeographic versus microgeographic variation. If the collector's aim is to capture the maximum amount of genetic variation associated with broad geographic differences in environmental factors then the sampling sites should be over-dispersed; that is, more or less evenly distributed over the target area. Alternatively, if the collector aims to capture variability associated with both geographic and microgeographic differences in, say, soil and climatic factors, then the sampling sites should be clustered in small groups (say three to five sites/group) and the groups distributed according to the obvious geographic heterogeneity in the environment.

The latter procedure will in most circumstances be the preferred method in species involved in biological control programs. Weedy species often occupy a wide range of habitats and gene flow between and within populations is usually limited. As a result, a considerable portion of the total variability in such species is associated with heterogeneity in local environments. The clustering of samples also has other obvious benefits. First, it reduces the 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, of unconsciously collecting samples from all similar, say, all lush or all arid, sites.

MODIFICATIONS TO THE BASIC STRATEGY

The above analysis indicates that the optimal sampling strategy is: (i) to collect 50 to 199 individuals per site; (ii) to sample as many sites as possible; and (iii) to cluster sampling sites in obviously different geographic areas so as to sample both macrogeographic and microgeographic components of variation. This strategy is based on a number of assumptions, either implicitly or explicitly, which may not always be met in practice. Here we examine ways to modify the basic procedure when one of these important assumptions is known to be invalid.

Breeding systems

A basic assumption underlying the above procedure is that the explorer lacks any knowledge of the breeding system of the target species. To take account of this, the number of individuals to sample per site was set equal to the number of independent gametes required. That is, the requirement of a sample of 100 independent gametes was formulated as a recommendation to collect 50 to 100

plants. However, this recommendation is conservative and would only be necessary in apomictic, parthenogenetic or clonally reproducing species where progeny are generally identical with their material parent or inbreeding species where the bulk of the population is homozygous of most loci. In outbreeding species a smaller sample of individuals may be adequate. In fact, if a plant species was truly random mating then, a sample of 100 seeds from any plant would suffice since, in theory, those seeds would represent on the male side 100 randomly chosen gametes. The point is that if an explorer knows that the species he is sampling is outbreeding then he may be able to improve his sampling efficiency by collecting more seeds off fewer individuals. However, we emphasize two points. First, the number of individuals sampled in known outbreeders should be kept at a reasonable level of, say, 20 per population. Very few species in reality practise random mating and too drastic a reduction in sample size could conceivably lead to marked reduction in the number of alleles collected. Second, this procedure is only valid in known outbreeding species – in asexual and inbreeding species the only way to ensure the collection of 100 random gametes is to sample 100 random individuals.

Physical constraints limiting the number of individuals sampled per site

Another assumption implicit in the framing of the basic sampling strategy was that the effort required to sample an additional individual at a site is small in comparison to the effort required to sample an entirely new site. In other words, we assumed it would be relatively easy for the explorer to collect 50 to 100 individuals of each site. Indeed, it may be extraordinarily difficult or impossible to collect 50 to 100 individuals per site if the target species is sparsely distributed, exists as small disjunct populations, or rapidly sheds its seed.

If the collection of 50 to 100 individuals at each stop would seriously restrict the number of sites which could be sampled, we would recommend a reduction in the number of individuals collected per site. For example, if only 1000 individuals could be collected it would be better to take 10 individuals from 100 populations than 100 individuals from 10 populations. The justification is simply that in these circumstances the collection of fewer plants from more sites allows greater opportunities to collect locally common alleles.

Genetic structure of the target species

A third, and possibly the most important, assumption underlying the basic strategy is that the explorer lacks any knowledge of the population structure of the target species. Information on the kinds and amounts of genetic variation in the target species and how this variation is distributed within and between populations can be used to develop more sophisticated and efficient sampling procedures. These are described in detail by Marshall and Brown (1975, 1980). We will not repeat them here because in the case of species involved in biological control programs the assumption that the explorer has little or no knowledge of the population structure of the species he is collecting is usually valid.

CONCLUSIONS

1. The objective of exploration in the target plant in biological programs for weed control is the collection of the maximum amount of relevant genetic variation within a strictly limited number of samples to permit optimum matching of host and parasite biotypes.
2. We define relevant genetic variation in biological weed control programs as

all variants specifying resistance or tolerance in the host with population frequency greater than five per cent.

3. The basic sampling strategy which is optimal in attaining the objective is: (i) to collect 50 to 100 individuals per site; (ii) to sample as many sites as possible within the time available; and (iii) to ensure that the sample sites represent as broad a range of environments as possible. This basic strategy is recommended when information on the population structure and breeding system of collected species is lacking.

4. Samples sizes of 50 or more individuals per site are generally conservative. Where there is insufficient easily procurable material to collect 50 individuals per site, this number should be reduced and more sites sampled.

REFERENCES

Allard, R.W. (1970). 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. (Blackwells:Oxford and Edinburgh.)

Antonovics, J. (1971). The effects of a heterogeneous environment of the genetics of natural populations. *Am. Sci.* 59:593—9.

Bennet, E. (1965). Plant introduction and genetic conservation: geneocological aspects of an urgent world problem. *Scotl. Pl. Breed. Stn.Rec.* pp.27—113.

Bradshaw, A.D. (1975). Population structure and the effects of isolation and selection. In 'Crop Genetic Resources for Today and Tomorrow'. (Eds. O.H. Frankel and J.G. Hawkes.) (Cambridge Univ. Press:Cambridge.), pp.37—51.

Brown, A.H.D. (1978). Isozymes, plant population genetic structure and genetic conservation. *Theor. Appl. Genet.* 52:145—57.

Clausen, C.P. (1936). Insect parasitism and biological control. *Ann. Ent. Soc. Am.* 29:201—33.

Caresche, L.A., and Wapshere, A.J. (1974). Biology and host specificity of the *Chondrilla* gall mite *Aceria chondrillae* (C. Can.) (Acarina, Eriophyidae). *Bull. Ent. Res.* 64:183—92.

Day, P.R. (1974). 'Genetics of Host-Parasite Interaction.' (Freeman:California.)

DeBach, P. (1974). 'Biological Control by Natural Enemies.' (Cambridge Univ. Press:Cambridge.)

Goeden, R.D., Kovalev, O.V., and Ricker, D.W. (1974). Arthropods exported from California to the U.S.S.R. for ragweed control. *Weed Science* 22:156—8.

Hasan, S. (1972). Specificity and host specialization of *Puccinia chondrillina*. *Ann. Appl. Biol.* 72:257—63.

Khush, G.S. (1977). Disease and insect resistance in rice. *Adv. Agron.* 29:265—341.

Marshall, D.R., and Brown, A.H.D. (1975). Optimum sampling strategies in genetic conservation. In 'Crop Genetic Resources for Today and Tomorrow'.

(Eds. O.H. Frankel and J.G. Hawkes.) (Cambridge Univ. Press:Cambridge.), pp.53–70.

Marshall, D.R., and Brown, A.H.D. (1980). Theory of forage plant collection. In 'Genetic Resources of Forage Plants'. (Eds. J.G. McIvor and R.A. Bray.) (In Press.)

Maxwell, F.G., Jenkins, J.N., and Parrott, W.L. (1972). Resistance of plants to insects. *Adv. Agron.* 24:187–265.

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

Qualset, C.O. (1975). Sampling germplasm in a center of diversity: an example of disease resistance in Ethiopian barley. In 'Crop Genetic Resources for Today and Tomorrow'. (Eds. O.H. Frankel and J.G. Hawkes.) (Cambridge Univ. Press: Cambridge.), pp.81–96.

Snaydon, R.W. (1978). Genetic changes in pasture populations. In 'Plant Relations in Pastures'. (Ed. J.R. Wilson.) (CSIRO:Melbourne.), pp.253–69.

Wapshere, A.J. (1975). A protocol for programmes for biological control of weeds. *P.A.N.S.* **21**:295–303.

Wilson, F. (1965). Biological control and the genetics of colonising species. In 'The Genetics of Colonising Species'. (Eds. H.G. Baker and G.L. Stebbins.) (Academic Press:New York and London.)

Winder, J.A., and Harley, K.L.S. (1978). Exploration for organisms for biological control of weeds. Proc. IV Int. Symp. Biol. Contr. Weeds, Gainesville, Florida, 1976, pp.35–8.