

Multilocus Diversity in an Outbreeding Weed, *Echium plantagineum* L.

A. H. D. Brown and J. J. Burdon

Division of Plant Industry, CSIRO, G.P.O. Box 1600, Canberra, A.C.T. 2601.

Abstract

A population of *E. plantagineum* was surveyed for its genetic structure at 23 isozyme loci. More than half (13) of these loci were polymorphic, with an average number of three alleles per locus, a gene diversity of 38% and heterozygosity of 35%. More importantly, the distribution of multilocus heterozygosity over individuals was found to approximate that assuming independence among loci or no linkage disequilibrium. The population consisted of a vast array of multilocus genotypes. This pattern indicates that the outbreeding system encourages recombination sufficient to outweigh the effects on multilocus structure of bottlenecks in population size. Genic and genotypic variation presumably allows high levels of biochemical flexibility in populations of *E. plantagineum*. Such flexibility could hamper attempts at biological control.

Introduction

Weed infestations often originate through the accidental or deliberate movement of plant species from one part of the world to another. When such migration involves few individuals, only a limited sample of genetic variability from the source population is likely to arrive [see Brown and Marshall (1981) and Barrett (1982) for review]. Examples of depauperate gene pools in Australian weeds are *Xanthium strumarium* (Moran and Marshall 1978), *Emex spinosa* (Marshall and Weiss 1982) and *Opuntia* spp. (Murray 1982).

Even when appreciable variability is introduced, it is still possible for populations to be composed of a restricted number of multilocus genotypes. This happens when the different alleles at each polymorphic locus are found in a restricted number of combinations. For example, *Chondrilla juncea*, an apomictic weed, occurs in Australia as three distinct races. These races differ for several isozyme loci (Burdon *et al.* 1980) so that the species has a high level of genetic diversity on the basis of single loci. Yet its multilocus diversity is severely restricted to the three predominant genotypes. Likewise, Californian populations of *Avena barbata* (Allard *et al.* 1972), and Israeli populations of *Hordeum spontaneum* (Brown *et al.* 1980) exhibit high levels of linkage disequilibrium. In these populations, certain multilocus genotypes are more frequent than would be expected if alleles at different loci occurred at random with respect to one another. Both species are predominantly self pollinated, colonizing species.

Echium plantagineum L. (Paterson's curse) is an insect-pollinated, self-compatible, annual diploid species. It was introduced into Australia as a garden species before 1850, and escaped to become a serious weed of pastures of south-eastern and south-western Australia (Piggin 1982). Although several introductions were made, they may have included only a limited sample of genetic variation. Furthermore, the rapid spread and colonizing

habit of the species is likely to reduce still further the variation found at any one site. Bottlenecks in population size might also be expected to generate substantial levels of linkage disequilibrium among the variants that remain. On the other hand, the outbreeding system of *E. plantagineum* could outweigh these historic effects and break down association among loci. Some modes of selection (e.g. heterotic selection) indeed accelerate this decay of disequilibrium under random mating (Clegg 1978). The aim of this study was to determine which of these forces, namely restricted population sizes or the open-mating system, predominated in shaping the genetic structure of a representative population of this weed.

Several statistical methods for studying multilocus data are available (see Brown and Weir 1983 for review and literature). Brown *et al.* (1980) developed a composite measure based on the distribution of the number of heterozygous loci in two randomly chosen gametes. In order to apply this measure to predominantly outbreeding species it is assumed that the observed distribution of heterozygous loci among individuals in the population provides a satisfactory approximation to the gametic distribution. The major advantage of the summarizing measure based on this distribution is that its sampling behaviour is known under the null hypothesis of absence of association between genes. Indeed, given the values of heterozygosity at each locus, two extreme distributions, that assuming *zero* associations on the one hand, and that assuming *complete* association of heterozygosity on the other, are readily calculated.

Materials and Methods

Plants

The population of *E. plantagineum* used was that previously studied on the river flats of the Murrumbidgee river at Gundagai, N.S.W., in 1980 (Burdon *et al.* 1983). The area sampled was the same as the area used in that transect study. In June 1982, some 200 seedlings at the rosette stage were transplanted from the field into single pots and grown in a heated glasshouse (25/20°C, day/night) with natural lighting. After 3 months, the plants flowered, and various tissues were assayed electrophoretically for a range of enzymes. By using transplanted individuals each plant could be repeatedly sampled to obtain sufficient loci and a complete data set. The final sample size was 193 plants.

Electrophoresis

Single, immature flower buds were prepared for horizontal, starch-gel electrophoresis as previously described for leaf samples (Burdon *et al.* 1983). The enzymes assayed were as follows: alcohol dehydrogenase (ADH; EC 1.1.1.1) or NAD(P)H dehydrogenase (synonym menadiene reductase MDR; EC 1.6.99.2) and esterase (EC 3.1.1.?) on 15 mm Tris–3 mm citrate gels with 0.1 M sodium hydroxide, 0.3 M boric acid tray buffer; acid phosphatase (ACP, EC 3.1.3.2), shikimate dehydrogenase (SDH, EC 1.1.1.25), glucosephosphate isomerase (PGI, EC 5.3.1.9), peroxidase (PRX, EC 1.11.1.7), aconitate hydratase (ACO, EC 4.2.1.3), phosphoglucosyltransferase (PGMT, EC 2.7.5.1) on 5 mm histidine (pH 8.0) gels with 0.4 M sodium citrate (pH 8.0) tray buffer. In addition, aspartate aminotransferase (EC 2.6.1.1), dipeptidase (EC 3.4.13.11), malate dehydrogenase (EC 1.1.1.37), phosphoglucuronate dehydrogenase (EC 1.1.1.43) and isocitrate dehydrogenase (EC 1.1.1.42) were assayed in a subsample of 40 plants and found to be uniform in this population. Assuming that each band on these five zymograms represents a locus, the estimated number of invariant loci in the total sample of stains is 10. Enzyme assays were as previously reported, or followed procedures in Siciliano and Shaw (1976).

Statistical Methods

The first step is to compute estimates of the usual single-locus parameters of genetic variation, namely allele frequencies, heterozygosity and gene diversity. The gene diversity at the *j*th locus (*h_j*) is defined as

$$h_j = 1 - \sum_i p_{ij}^2,$$

where p_{ij} is the frequency of the i th allele at the j th locus. The observed frequency of heterozygotes at the j th locus is H_j .

The multilocus isozyme genotype for each plant is now screened and reduced to the number of loci (out of 13 polymorphic) at which the plant was heterozygous. The empirical frequency distribution is then constructed and its variance (s_k^2) computed.

The theoretical distribution of heterozygous loci *assuming independence* among the loci is obtained by convolution of the generating functions for heterozygosity at each single locus. In the present case of 13 polymorphic loci, the generating function is:

$$G(X) = \prod_{j=1}^{13} [(1-H_j) + H_j X]. \quad (1)$$

This function is expanded numerically and the coefficient of the term X^i is the probability that the number of loci out of 13, for which an individual is heterozygous is i . From this distribution the expected variance is readily calculated. This variance, assuming independence of loci, is also obtainable directly using the values of H_j (Brown *et al.* 1980):

$$\sigma_k^2 = \sum H_j - \sum H_j^2.$$

In applying the multilocus methods of Brown *et al.* (1980) to outbreeding populations, it is necessary to substitute H_j for h_j in formulae 3, 4, 5, 20, and 22 of that paper. Thus the variance of the estimate of s_k^2 , assuming the null hypothesis of independence among the loci is true, is:

$$\text{var}(s_k^2 | \text{independence}) = [\sum H_j - 7\sum H_j^2 + 12\sum H_j^3 - 6\sum H_j^4 + 2(\sum H_j - \sum H_j^2)^2] / n,$$

where n is the number of plants scored and the vertical line indicates the conditional hypothesis of independence. Using this variance and assuming the sampling distribution of σ_k^2 approximates normality, the upper 95% confidence limit for s_k^2 is:

$$L \cong \sum H_j - \sum H_j^2 + 2 [\text{var}(s_k^2 | \text{independence})]^{1/2}.$$

Thus, if the observed s_k^2 exceeds L , the null hypothesis of independence at the level of locus pairs is rejected.

At the other extreme, it is of interest to compute a distribution which displays 'maximum association' in some specific sense, for given levels of heterozygosity at each locus. We suggest that a suitable yardstick would be a distribution with maximum dispersion or variance of heterozygosity. Such a distribution can be constructed by *ranking* the observed (H_j ; $j = 1, \dots, m$) such that

$$H_1 > H_2 > H_3 \dots > H_m,$$

where m is the number of polymorphic loci. Assume now that the genotypic scores for each locus can be allocated in any manner to individuals. A set of hypothetical multilocus genotypes is constructed such that $n(H_{m-1} - H_m)$ are heterozygous at $m-1$ loci, etc., and $n(1 - H_1)$ are heterozygous at no loci. The variance of this hypothetical distribution is:

$$s_k^2(\text{max}) = 2\sum jH_j - \sum H_j - (\sum H_j)^2.$$

Results

Table 1 lists for each polymorphic locus the number of alleles and the allele frequencies. The alleles are designated a, b, c, d in decreasing order of electrophoretic mobility. The loci *Mdr*, *Pgi*, *Pgm* 1 and *Sdh* were the same as in the 1980 transect study, and their allelic frequencies were remarkably similar to that study (Burdon *et al.* 1983). The *Est* and *Prx* loci were difficult to score. For the *Est* locus, the isozymes are close to the borate front and differed among themselves only slightly in migration rate. The peroxidase isozymes are weakly expressed. In both cases, repeated sampling was used to check classification.

The values for expected heterozygosity or gene diversity (h) were compared with observed heterozygosity (H). As in 1980, observed heterozygosity for *Pgi* exceeded its expected value, although not significantly. As with the early seedling data of the previous study, there was no evidence of heterozygote excess early in the generation. The overall

average heterozygosity for the 13 loci was 0.357 compared with the expected value under random mating of 0.380. The estimated fixation index is $F[= 1 - \sum(H_j/h_j)/13] = 0.08$, indicative of a predominantly random mating population.

The average number of alleles per polymorphic locus was 3.0, and if the 10 presumed monomorphic loci are included, the value was 2.13. This represents a remarkably high level of genic polymorphism.

Table 1. Single-locus polymorphism in *E. plantagineum* at Gundagai

Locus ^a	No of alleles	Allele frequencies				Heterozygosity ^b	
		a	b	c	d	<i>h</i>	<i>H</i>
<i>Aco 1</i>	2	0.801	0.199			0.319	0.264
<i>Aco 2</i>	3	0.104	0.886	0.010		0.204	0.124
<i>Aco 3</i>	3	0.106	0.884	0.010		0.208	0.176
<i>Aco 4</i>	2	0.990	0.010			0.021	0.021
<i>Acp</i>	3	0.010	0.575	0.415		0.497	0.544
<i>Adh</i>	2	0.956	0.044			0.084	0.078
<i>Est</i>	4	0.104	0.355	0.350	0.192	0.704	0.606
<i>Mdr</i>	4	0.021	0.655	0.192	0.132	0.516	0.487
<i>Pgi</i>	4	0.015	0.526	0.231	0.228	0.618	0.668
<i>Pgm 1</i>	3	0.233	0.731	0.036		0.411	0.373
<i>Pgm 2</i>	2	0.249	0.751			0.374	0.425
<i>Prx</i>	4	0.070	0.067	0.171	0.692	0.493	0.415
<i>Sdh</i>	3	0.370	0.609	0.021		0.492	0.456
Average	3.0					0.380	0.357

^a*Acp*, *Adh* and *Pgi* show a single additional heteroallelic, multimeric isozyme in heterozygotes; whereas *Mdr* shows three such 'hybrid bands'; and the isozymes are presumably tetrameric.

^bThe observed heterozygosity is denoted *H*, and the expected heterozygosity *h*.

The observed distribution of heterozygosity among individuals is depicted in Fig. 1*a*, compared with the distribution expected under random association of loci. It is clear that they are very similar. The observed variance (s_K^2) was 2.837, whereas the expected variance (σ_K^2) was 2.468. These values lead to a multilocus index [$X(2) = s_K^2/\sigma_K^2 - 1$] of 0.150 which is not significantly different from zero. [The upper 95% confidence limit for the variance, from Brown *et al.* (1980), is $2.468 + 0.486 = 2.954$.] The distribution which gives maximum variance given the single-locus values of heterozygosity is also shown in Fig. 1*b*. Its variance is 19.63, so that the maximum observable value for $X(2)$ is 6.95. Hence on this basis, the population had only about 2% of the potential multilocus organization for polymorphism at these loci. Of course, this procedure gives an overall picture of the association among the marker loci in this population. When the overall measure approximates the expectation assuming independence, it cannot be concluded that all alleles at all these loci are in linkage equilibrium. The specification of the two distributions (as in Figs 1*a* and 1*b*) is a further development of the procedures of Brown *et al.* (1980).

The above treatment is based on whether or not a single locus is heterozygous, as seen in function (1). This treatment of multilocus structure can readily be extended to summarizing structure over loci for other two-state, single-locus variables. For example Gottlieb (1981) has proposed a measure of genetic variation, called the 'homogeneity value' based on whether or not the diploid genotype at each locus in each individual is the commonest homozygote at that locus. The multilocus extension of this variable uses the above formulae but with H_j redefined as the population frequency of the commonest homozygote for the *j*th locus. The summary variances have the subscript *Y* replacing the *K* to indicate this change.

In the present data, the observed variance in the number of loci with the most frequent homozygous genotype (s_y^2) was 2.110. The expected variance under independence (σ_y^2) was 2.392, indicating a slight deficiency of extreme types. In other words genetic variation was spread over individuals at random, and not confined to a few rare 'off types' in the midst of a common optimum genotype.

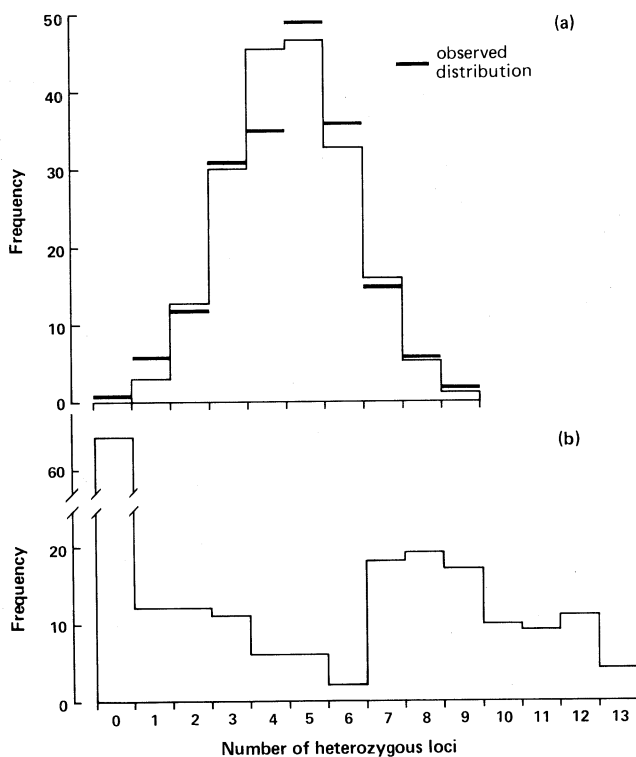


Fig. 1. (a) Histogram showing the expected distribution of the number of heterozygous loci assuming independence (equation 1). The observed distribution is shown as thick bars. (b) Hypothetical distribution of heterozygosity which has maximum variance.

Discussion

Despite the fact that *E. plantagineum* is a weedy, annual herb, deliberately introduced from Europe in limited, if not severely restricted, numbers its Australian populations can be remarkably variable. The level of isozyme polymorphism found in the Gundagai population exceeds the average level observed in a range of other plant species (Hamrick *et al.* 1979). It is typical of the levels found in several other *E. plantagineum* populations, which have been surveyed for isozyme loci that express in seeds (Marshall *et al.* 1981; Burdon, unpublished data). Furthermore, the breeding system of *E. plantagineum* is evidently one which leads to free recombination so that alleles at different loci are organized into genotypes essentially at random. The high level of polymorphism and its generally random organization implies that at least in populations like the one studied, each individual in the population is probably genetically unique.

These results are relevant to current proposals to control this weedy species using insects introduced from the plant's native range in the Mediterranean region (Delfosse and Cullen 1981). Previous experience with the biological control of weedy plant species around the world has shown that those species which are presumed to be genetically uniform or possess low levels of genetic diversity have been controlled more effectively and more often than those species which are highly variable (Burdon and Marshall 1981). Thus in Australia most species which have been at least reasonably well controlled by the use of natural enemies are either apomictic weeds such as *Opuntia* spp. and *Chondrilla juncea* or are species that reproduce clonally such as *Alternanthera philoxeroides* and *Salvinia* spp. Their population genetic structure stands in marked contrast to the present findings for *E. plantagineum*. With its propensity for high genic and multilocus genotypic diversity, *E. plantagineum* is likely to be more difficult to control by biological means than a genetically uniform weed. It is therefore likely to require a greater input of resources to achieve the same level of control as a uniform weed.

In this study isozyme variants have been exploited only as markers, or indicators of the population genetic structure of the genome including those genes responsible for resistance to specific insect pests and fungal pathogens. Yet the relationship between isozyme variation and host resistance variation is unknown (Barrett 1982). The biological significance of isozyme variation itself remains a matter of controversy. However, Clarke (1979) suggested that biochemical variation may have a direct role in withstanding pest or pathogen pressure. One possible example of this has been found in peas, where an esterase isozyme is closely correlated with wilt resistance (Hunt and Barnes 1982). If this proposal is correct, then it is clear that *E. plantagineum* in Australia possesses a high level of biochemical flexibility which any biological control program must endeavour to overcome.

References

- Allard, R. W., Babbel, G. R., Clegg, M. T., and Kahler, A. L. (1972). Evidence for coadaptation in *Avena barbata*. *Proc. Natl Acad. Sci. U.S.A.* **69**, 3043–8.
- Barrett, S. C. H. (1982). Genetic variation in weeds. In 'Biological Control of Weeds with Plant Pathogens'. (Eds R. Charudattan and H. L. Walker.) pp. 73–98. (Wiley: New York.)
- Brown, A. H. D., Feldman, M. W., and Nevo, E. (1980). Multilocus structure of natural populations of *Hordeum spontaneum*. *Genetics* **96**, 523–36.
- Brown, A. H. D., and Marshall, D. R. (1981). Evolutionary changes accompanying colonization in plants. In 'Evolution Today'. (Eds G. G. E. Scudder and J. L. Reveal.) *Proc. Second Int. Congr. Syst. Evol. Biol.* pp. 351–63. (Hunt Institute for Botanical Documentation, Carnegie-Mellon University: Pittsburgh.)
- Brown, A. H. D., and Weir, B. S. (1983). Measuring genetic variability in plant populations. In 'Isozymes in Plant Genetics and Breeding'. (Eds S. D. Tanksley and T. J. Orton.) pp. 219–39. (Elsevier: Amsterdam.)
- Burdon, J. J., Marshall, D. R., and Groves, R. H. (1980). Isozyme variation in *Chondrilla juncea* L. in Australia. *Aust. J. Bot.* **28**, 193–8.
- Burdon, J. J., and Marshall, D. R. (1981). Biological control and the reproductive mode of weeds. *J. Appl. Ecol.* **18**, 649–58.
- Burdon, J. J., Marshall, D. R., and Brown, A. H. D. (1983). Demographic and genetic changes in populations of *Echium plantagineum* L. *J. Ecol.* **71**, in press.
- Clarke, B. C. (1979). The evolution of genetic diversity. *Proc. R. Soc. Lond.* **B205**, 434–74.
- Clegg, M. T. (1978). Dynamics of correlated genetic systems. II. Simulation studies of chromosomal segments under selection. *Theoret. Pop. Biol.* **13**, 1–23.
- Delfosse, E. S., and Cullen, J. M. (1981). New activities in biological control of weeds in Australia. II. *Echium plantagineum*: Curse or Salvation? *Proc. Fifth Int. Symp. Biol. Contr. Weeds*, Brisbane, Australia, 1980. pp. 563–74. (CSIRO: Melbourne.)
- Felsenstein, J. (1974). The evolutionary advantage of recombination. *Genetics* **78**, 737–56.
- Gottlieb, L. D. (1981). Electrophoretic evidence and plant populations. *Progr. Phytochem.* **7**, 1–46.

- Hamrick, J. L., Linhart, Y. B., and Mitton, J. B. (1979). Relationship between life history characteristics and electrophoretically-detectable genetic variation in plants. *Ann. Rev. Ecol. Syst.* **10**, 173–200.
- Hunt, J. S., and Barnes, M. F. (1982). Molecular diversity and plant disease resistance: An electrophoretic comparison of near-isogenic lines of wilt-resistant or -susceptible *Pisum sativum* L. cv. William Massey. *Euphytica* **31**, 341–8.
- Marshall, D. R., Burdon, J. J., and Brown, A. H. D. (1981). Optimal sampling strategies in the biological control of weeds. Proc. Fifth Int. Symp. Biol. Contr. Weeds, Brisbane, Australia, 1980. pp. 103–11. (CSIRO: Melbourne.)
- Marshall, D. R., and Weiss, P. W. (1982). Isozyme variation within and among Australian populations of *Emex spinosa* (L.) Campd. *Aust. J. Biol. Sci.* **35**, 327–32.
- Moran, G. F., and Marshall, D. R. (1978). Allozyme uniformity within and variation between races of the colonizing species *Xanthium strumarium* L. (Noogora Burr). *Aust. J. Biol. Sci.* **31**, 283–91.
- Murray, N. D. (1982). Ecology and evolution of the *Opuntia*—*Cactoblastis* ecosystem in Australia. In 'Ecological Genetics and Evolution: The Cactus: Yeast: *Drosophila* Model System'. (Eds J. S. F. Barker and W. T. Stamer.) pp. 17–30. (Academic Press: Sydney.)
- Piggin, C. M. (1982). The biology of Australian weeds 8. *Echium plantagineum* L. *J. Aust. Inst. Agric. Sci.* **48**, 3–16.
- Siciliano, M. J., and Shaw, C. R. (1976). Separation and visualization of enzymes on gels. In 'Chromatographic and Electrophoretic Techniques'. Vol. 2. Zone electrophoresis. 4th Edn. (Ed. I. Smith.) pp. 185–209. (Heinemann: London.)

