RARE ELECTROPHORETIC VARIANTS IN A HYBRID ZONE

N. H. BARTON*, R. B. HALLIDAY† AND G. M. HEWITT

School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, England

Received 29.ix.82

SUMMARY

Populations of the grasshopper *Podisma pedestris* were collected from two ends of a zone of hybridization between two chromosome races, at Seyne and Tende in southern France. 21 enzyme and protein loci were detected by gel electrophoresis. Six of these loci showed widespread polymorphism, and a further eleven had very little or no variation. Two loci (*Idh, 6Pgd*) had rare alleles in different frequencies in the two areas surveyed. The remaining two loci (*Mdh-1, Mdh-2*) showed a marked increase in the frequency of rare variants, from 1 per cent outside the hybrid zone, up to 5 per cent at its centre. This region of increased electrophoretic variation coincided with the chromosomal cline between the two races, and with a region of decreased viability. It was spread over about the same width as the chromosomal cline. Possible explanations for this extra variation include intragenic recombination and elevated mutation rates.

1. INTRODUCTION

Two chromosome races have been observed in the grasshopper *Podisma pedestris*. Many populations contain the typical Acridid karyotypes of 22 + X0 males and 22 + XX females. In an area of south-eastern France, an X-autosome fusion has occurred, resulting in the appearance of a neo-XY race in which males appear as 20 + XY and females as 20 + XX. These two races meet and hybridize along a hybrid zone which runs for 100 km across the Alpes Maritimes (Hewitt, 1975; Barton and Hewitt, 1981). Crosses between these races reveal substantially decreased viability of the F1. This inviability is not associated with the chromosome fusion, which is subject to only weak selective pressures (Barton, 1980; Barton and Hewitt, 1981). An extensive electrophoretic survey of both pure and hybridizing populations has revealed no allozymic differentiation at all between the two races for 21 loci, including six which exhibit widespread polymorphism (Halliday *et al.*, 1982; 1983).

The genetic relationships between hybridizing taxa may be studied not only by surveying the differences between them, but also by examining changes which occur within the region of hybridization. In *P. pedestris*, we have observed that animals from hybrid populations have lower viability (Barton and Hewitt, 1982), higher levels of chromosomal aberations (Barton, 1980), and more variable Fuelgen staining of spermatids (Westermann *et al.*, in prep.), than those in pure populations. The present paper describes a similar anomaly in the frequency of rare variant alleles at electrophoretically detected loci within the hybrid zone.

* Present address: Department of Genetics, University College, London NW1 2HE, England

† Present address: CSIRO Division of Entomology, P.O. Box 1700, Canberra City A.C.T. 2601, Australia

2. Materials and Methods

Samples of approximately 20 adult male *P. pedestris* were taken from 98 local populations, each from an area of 50×50 metres or smaller. Eleven populations were from near Seyne, at the western end of the hybrid zone (Halliday *et al.*, 1983), and 87 were from near Tende, at the eastern end (Barton and Hewitt, 1981; Halliday *et al.*, 1982). These two areas are separated by a distance of approximately 100 km and in each case collections spanned an area of approximately 3 km either side of the hybrid zone. Smaller collections were also made from Switzerland (Offenpass) and Austria (Wiener Neustadt).

Techniques for the handling of animals and the preparation and electrophoresis of extracts are described by Halliday *et al.* (1982). The main aim of the isozyme survey was to examine the differences between the two chromosome races. It therefore concentrated on the most highly polymorphic isozyme loci, namely *Amylase*-1, *Amylase*-2, *Esterase*-1, *Esterase*-2, and *Malic enzyme*, for which the mean total number of alleles examined exceeds 1800. Between five and 15 animals from each local population were scored for the other loci.

3. Results

A total of 21 loci was examined in a general isozyme survey of *P. pedestris* populations. Six of these loci (*Esterase*-1, *Esterase*-2, *Amylase*-1, *Amylase*-2, *Phosphoglucomutase*-3, *Malic enzyme*) exhibited widespread polymorphisms which are described elsewhere (Halliday *et al.*, 1982; 1983). These loci will not be discussed further in the present paper. A further five loci showed no variation at all in the animals surveyed—*Alcohol dehydrogenase* (total sample size 792 alleles), *Hexokinase* (992), *Phosphoglucomutase*-1 (1222), *Phosphoglucomutase*-2 (868) and *Glutamate dehydrogenase* (678). Six loci had a common allele whose overall frequency in the populations examined was greater than 99 per cent. *Protein*-2 had two variant alleles out of the 1050 sampled, *Superoxide dismutase* had a single variant in a sample of 1020, and *α-Glycerophosphate dehydrogenase* had one in 1242. For *Protein*-1 (sample size 1048 alleles), the common allele was $P\text{-}1^a$. Six examples of the variant $P\text{-}1^b$ were found in a single population of 19 individuals (14aa: 4ab: 1aa), and another variant $P\text{-}1^c$ was represented by a single heterozygote in each of two other populations. *P*-1, *P*-2, *Sod* and *α-Gpd* were previously thought to be completely monomorphic in both of the areas surveyed, but we now recognize the existence of variant alleles at these very low frequencies. For *Glyceraldehyde*-3-*phosphate dehydrogenase* (sample size 858 alleles, common allele $G3pd^b$), three examples of the variant $G3pd^a$ were found. Two of these were contributed by heterozygotes in the same sample of five individuals. Seven variants (Ao^a) were found for the *Aldehyde oxidase* locus, out of a total of 1296 examined; two of these were from heterozygotes in the same sample of ten individuals. The remaining four loci showed intermediate levels of variation, as indicated by the figures in table 1. These loci are *Isocitrate dehydrogenase* (*Idh*), 6-*Phosphogluconate dehydrogenase* (*6Pgd*), *Malate dehydrogenase*-1 (*Mdh*-1), and *Malate dehydrogenase*-2 (*Mdh*-2). Since variant alleles at these loci are uncommon, we have not

TABLE 1

Frequency of rare variants at the loci Idh, 6Pgd, Mdh-1, and Mdh-2, broken down into data from Seyne and from Tende, and into data from karyotypically pure and mixed samples.

	Tende		Seyne	
Idh	Pure	Mixed	Pure	Mixed
Alleles scored	442	458	62	138
No. of rare alleles $(a+c)$	3	6	0	0
Populations with 0 rare alleles	33	32	4	7
Populations with 1 rare alleles	3	3	—	—
Populations with 2 rare alleles	—	0	—	—
Populations with $\geqq$3 rare alleles	—	1	—	—
6Pgd				
Alleles scored	532	526	80	116
No. of rare alleles $(a+x)$	11	11	7	8
Populations with rare alleles	34	31	1	3
Populations with 1 rare alleles	7	9	1	2
Populations with 2 rare alleles	2	1	1	1
Populations with $\geqq$3 rare alleles	—	—	1	1
Mdh-1				
Alleles scored	688	634	80	140
No. of rare alleles $(a+c)$	6	19	10	8
Populations with 0 rare alleles	37	31	0	3
Populations with 1 rare alleles	6	4	2	1
Populations with 2 rare alleles	—	4	—	2
Populations with $\geqq$3 rare alleles	—	2	2	1
Mdh-2				
Alleles scored	540	660	80	140
No. of rare alleles (a)	4	31	0	1
Populations with 0 rare alleles	35	33	4	6
Populations with 1 rare alleles	4	4	—	1
Populations with 2 rare alleles	—	2	—	—
Populations with $\geqq$3 rare alleles	—	5	—	—

been able to perform crosses to confirm their Mendelian inheritance. It is possible that the phenotypes for these enzymes could be subject to non-genetic variation. However, if variation of this kind was to exist, we would expect it to be random with respect to both geographic location and karyotype.

The locus *Idh* has three alleles, the commonest of which is Idh^b. The rare alleles Idh^a and Idh^c produce isozymes which migrate faster and slower respectively, than the Idh^b isozyme. These rare alleles may be more common at Tende than at Seyne (table 1), but they are nowhere common enough to allow meaningful population comparisons.

The *6Pgd* locus has common allele $6Pgd^b$ in most areas, with $6Pgd^a$ occurring at low frequency in a number of populations from Tende. The highest local frequency recorded there for $6Pgd^a$ is 0·25 (population 136), but it usually occurs at frequencies below 0·1. $6Pgd^a$ also occurs in low frequency at Seyne, where it is joined by the allele $6Pgd^x$, which does not

occur at Tende. Karyotypically pure and mixed populations seem to contain *6Pgd* variant alleles at similar frequencies. The only exception is the pure X0 population from Switzerland, in which $6Pgd^a$ reached a gene frequency of 0·64 (18/28).

Rare alleles at the *Mdh*-1 locus occur in higher frequency at Seyne than they do at Tende. For *Mdh*-2, rare alleles occur at similar frequencies in the two areas. At Tende, rare alleles for both *Mdh*-1 and *Mdh*-2 occur at higher frequencies in mixed populations than they do in pure populations. Also, the proportion of populations containing several rare alleles is higher in mixed than in pure populations at Tende, for both loci. These differences between pure and mixed populations are not apparent at Seyne, perhaps due to the reduced sample sizes and the generally lower frequencies of rare alleles in that area. The increase in the frequency of rare alleles in mixed populations at Tende is demonstrated more clearly by plotting the frequency of variants against the position of each population in the chromosomal cline (fig. 1). This frequency rises from ~1 per cent outside the cline
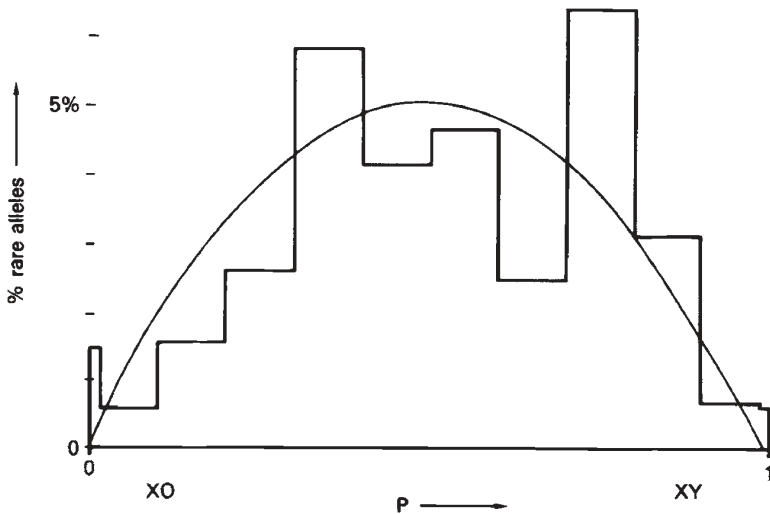

FIG. 1. A histogram showing the combined frequency of rare alleles at *Mdh*-1 and *Mdh*-2, and a smooth arc indicating the expected chromosomal heterozygosity across the hybrid zone. The latter frequency was calculated by fitting a smooth cline to observed chromosomal data by the method of maximum likelihood. P = frequency of the neo-*XY* karyotype.

up to ~5 per cent at its centre. This region of increased electrophoretic variation coincides with the chromosome cline, and with the region of decreased viability, and occupies a comparable distance. The increase is statistically significant for both loci. Regressions against chromosomal heterozygosity (h) give significant increases of log likelihood over the hypothesis of uniform frequencies ($2\Delta L \sim \chi_1^2 = 5\cdot84$ for $p = 0\cdot0136 + 0\cdot04\ h$, as against $p = 0\cdot0252$ at *Mdh*-1; and $2\Delta L \sim \chi_1^2 = 28\cdot66$ for $p = 0\cdot0051 + 0\cdot10\ h$, as against $p = 0\cdot0298$ at *Mdh*-2).

For three loci (*Idh, Mdh*-1, *Mdh*-2), as well as for two of the less variable loci described above (*G3pdh, P*-1), the data suggest that there

may be an excess of populations carrying two or more rare alleles, or in other words, that the frequency of the rare alleles varies from sample to sample. There are also significant local fluctuations in allele frequencies at the more polymorphic loci (Halliday *et al.*, 1982), and one would like to know not only whether the fluctuations in rare allele frequency are statistically significant, but also whether they are of similar magnitude to those of common alleles. We will therefore estimate Wright's F_{ST} ($= \text{var}(p)/pq$) by finding the likelihood with which the data could have arisen from spatial patterns characterised by any given F_{ST} coefficient. This will give a support curve whose maximum is the most likely estimate, $\hat{F}_{ST}$, and whose width gives the reliability of this estimate. In large samples, the difference in log likelihood between $F = \hat{F}_{ST}$ and $F = 0$ is distributed as $\frac{1}{2}\chi_1^2$, and gives a statistical test for the significance of the fluctuations (fig. 2; Edwards, 1972).
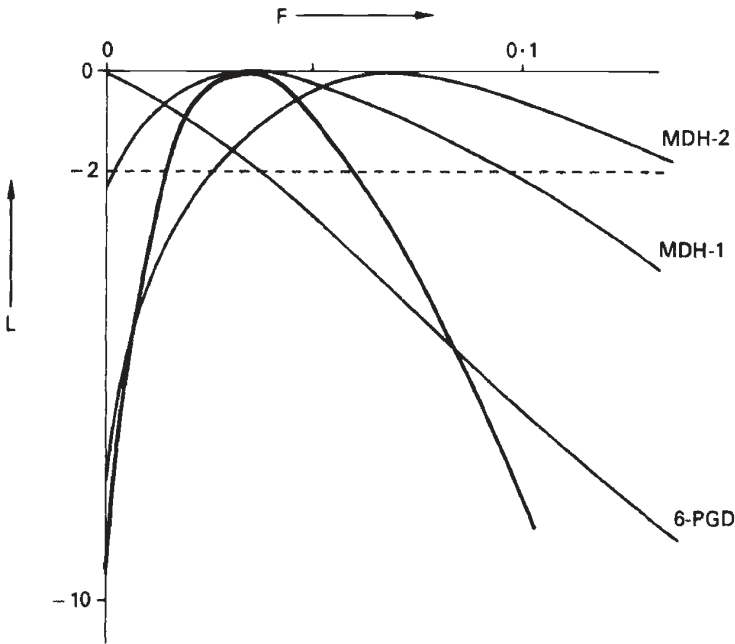

FIG. 2. The log likelihood of the observed data from Tende plotted against the theoretical F coefficient (variance of allele frequency/mean allele frequency) for three loci *6Pgd*, *Mdh*-1, *Mdh*-2. The heavy curve shows the total log-likelihood for the above data combined with that from Seyne.

(Slatkin (1981) has proposed an alternative, qualitative, method of analysing spatial patterns of rare variants; however, this method is not directly applicable to these data, since it assumes a uniform mutation rate.)

To calculate the likelihoods, we must make some assumption about the form of the distribution of frequencies. In a single population of size N_e, sampling drift, mutation (μ), and selection against deleterious alleles (s) will combine to give an equilibrium distribution of the form $(p/\alpha)^{\beta-1} e^{-\alpha p}/\alpha \Gamma(\beta)$ (where, if $\beta = 4 N_e \mu$, $\alpha = 4 N_e s$, $p \ll 1$, the mean frequency is (β/α), and $F_{ST} = 1/\alpha$: Wright (1937)). It would be more realistic to assume that *P. pedestris* is distributed more or less continuously,

but the gene frequency distribution then becomes more difficult to calculate. Differences in allele frequency were therefore analysed by assuming that the above single population distribution holds in each site sampled, and that the sites are far enough apart for the fluctuations to be uncorrelated with each other. For $6Pgd$, the expected allele frequency (β/α) was taken as the average frequency. However, since rare allele frequency at Mdh-1, 2 clearly increases in chromosomally heterozygous populations, for these loci the regression of allele frequency against chromosomally heterozygosity was used. Because of the complexity of this analysis, only data from the three most polymorphic loci (Mdh-1, Mdh-2, $6Pgd$) have been used.

The log likelihood of the data is plotted against Wright's F_{ST} in fig. 2. There are significant fluctuations at Mdh-2 at Tende ($\hat{F} \sim 0.069$, $2\Delta L \sim \chi_1^2 = 15.66$), marginally significant fluctuations at Mdh-1 at Tende ($\hat{F} = 0.038$, $2\Delta L \sim \chi_1^2 = 4.78$), and no significant fluctuations at $6Pgd$ at Tende ($\hat{F} = 0$). These estimates do not differ significantly from each other, or from the much less accurate estimates derived from the Seyne data on $6Pgd$ and Mdh-1 ($\chi_4^2 = 5.02$, n.s.; overall, $\hat{F} = 0.032$). This degree of variation is of the same order as the variation in frequency at the more polymorphic loci, and for the chromosomal polymorphism. Possible causes for these fluctuations will be discussed in a later paper.

4. DISCUSSION

At two different loci, and in one of two sampling areas, there is a five-fold increase in the frequency of rare alleles in hybridizing populations, over their background frequency in pure populations. This phenomenon was first found in a hybrid zone in the land snail, *Partula* (Clarke, 1968), and has since been reported in hybrid zones in *Mus musculus* (Hunt and Selander, 1973), between *Uroderma bilobatum davisii* and *U.b. convexum* (Greenbaum, 1981), between *Rana belandieri* and *R. utricularia* (Sage and Selander, 1979), and, most strikingly, between *Cerion* spp. (Woodruff, 1981).

Before discussing possible explanations for these results, we note that they do not occur indiscriminately with respect to either genes or to alleles. Increases by a substantial factor at some loci, and in *Mdh*-2 at Seyne, may have been missed simply because of the very low levels of variation in that area. Nevertheless, it does seem that the effect is restricted to certain genes and certain places. It also seems to be restricted to certain electromorphs, in that only one or two distinct alleles are involved. It seems unlikely that each "allele" could be traced back to a single ancestral variant, since such a variant would be very old by the time it had achieved a noticeable frequency. By that time it would also have spread out geographically, blurring the distinction between pure and mixed populations. Such selectivity is typical of the other cases cited above, but gives little information as to the processes responsible.

Since they are so rare, and are distributed so sporadically, it seems plausible that the observed variants are deleterious, and are maintained in a balance between their production by intragenic recombination or by

mutation, and their elimination by selection. Their increased frequencies could therefore be due either to increased rates of production in hybrids, or to a relaxation of selection pressures within the hybrid zone. Mechanisms for the latter are difficult to imagine, but several mechanisms exist for the production of new variation in hybrids, and these may be difficult to distinguish. Intragenic recombination between alleles from two taxa which had diverged while retaining the same electrophoretic phenotype, could readily account for the limited numbers of loci involved, and the limited number of alleles produced.

Alternatively, the rate of mutation in hybrids might be increased by a mechanism analogous to hybrid dysgenesis in *Drosophila* (*e.g.*, Woodruff *et al.*, 1982, Engels, 1981). There is a greater proportion of chromosomal aberrations in grasshopper embryos from hybrid populations (Barton, 1980) and an increased variability in the staining of spermatids by Fuelgen (Westermann, *et al.*, in prep.), suggesting some general breakdown of control over the genetic system in hybrid *P. pedestris*.

Although we cannot see any way to distinguish among the above possibilities without more direct information, we can estimate the strengths of the processes involved by examining the spatial distribution of the variants. These are found within a region about the same width as the chromosomal cline, about 800 metres across. This suggests that they may be acted upon by selective pressures similar in strength to those which operate on the chromosomal difference. Comparison of the cline width with the measured dispersal distance of ~20 metres in a generation, gives an estimate of $s_0 \sim 0 \cdot 5$ per cent (Barton and Hewitt, 1981), which is a rough lower limit for the selection (s) acting against the electrophoretic variants. Since these reach a frequency of ~5 per cent in the centre of the hybrid zone, the mutation rate (or, for that matter, the intragenic recombination rate), must be at least $2 \cdot 5 \times 10^{-4}$.

The phenomenon described here seems to occur widely, both in other species, and for other characters in *Podisma*. However, we find it hard to assess its evolutionary significance. It probably does not contribute very greatly to the isolation between the races, since the estimated loss of fitness caused by the effect ($\mu \geq 2 \cdot 5 \times 10^{-4}$ per locus) is much less than the average selection pressure of $\sim 3 \cdot 5 \times 10^{-2}$ which was estimated to be acting on each inviability gene (by comparison of the width of the region of hybrid inviability with the dispersal rate; Barton and Hewitt, 1982). Whether the effect contributes significantly to the evolution of new types from within the hybrid zone, as suggested by Woodruff *et al.*, (1982), depends on the difficult question of whether or not the rate of evolution is limited by selective constraints or by the availability of useful variation, and on whether hybrid zones occupy a large enough proportion of a species' range to provide a significant source of new genotypes. Still, there are cases where new species have originated *via* hybridisation between two ancestral types (Mayr, 1963; Grant, 1971), and the availability of increased variation may have aided their adaptation or even their initiation. In sexually reproducing species originating from two quite distinct parental genotypes, simple intergenic recombination in hybrids would probably release more variability than could be provided by new mutation or intragenic recombination. However, the phenomenon described here may be of some importance in establishing new asexual or selfing species derived from hybrid populations.

5. REFERENCES

BARTON, N. H. 1980. The fitness of hybrids between two chromosomal races of the grasshopper *Podisma pedestris*. *Heredity 45*, 47–59.

BARTON, N. H. AND HEWITT, G. M. 1981. A chromosomal cline in the grasshopper *Podisma pedestris*. *Evolution 35*, 1008–1018.

BARTON, N. H. AND HEWITT, G. M. 1982. The genetic basis of hybrid inviability between two chromosomal races of the grasshopper *Podisma pedestris*. *Heredity 47*, 367–383.

CLARKE, B. C. 1968. Balanced polymorphism and regional differentiation in land snails. In Drake, E. T. (ed.) *Evolution and Environment*, Yale University Press, pp. 351–368.

EDWARDS, A. W. F. 1972. *Likelihood*. Cambridge University Press.

ENGELS, W. R. 1981. Hybrid dysgenesis in *Drosophila* and the stochastic loss hypothesis. *Cold Spring Harbor Symp. Quant. Biol. 45*, 561–565.

GRANT, V. 1971. *Plant Speciation*. Columbia University Press, New York.

GREENBAUM, I. F. 1981. Genetic interactions between hybridizing cytotypes of the tent-making bat (*Uroderma bilobatum*). *Evolution 35*, 306–321.

HALLIDAY, R. B., BARTON, N. H. AND HEWITT, G. M. 1982. Electrophoretic analysis of a chromosomal hybrid zone in the grasshopper *Podisma pedestris*. *Biol. J. Linn. Soc.* (in press).

HALLIDAY, R. B., WEBB, S. F. AND HEWITT, G. M. 1983. Genetic and chromosomal polymorphism in hybridizing populations of the grasshopper *Podisma pedestris*. (MS submitted to *Biol. J. Linn. Soc.*).

HEWITT, G. M. 1975. A sex-chromosome hybrid zone in the grasshopper *Posidma pedestris* (Orthoptera: Acrididae) *Heredity 35*, 375–387.

HUNT, W. G. AND SELANDER, R. K. 1973. Biochemical genetics of hybridization in European house mice *Heredity 31*, 11–13.

MAYR, E. 1963. *Animal Species and Evolution*. Belknap, Harvard.

SAGE, R. D. AND SELANDER, R. K. 1979. Hybridization between species of the *Rana pipiens* complex in central Texas, USA. *Evolution 33*, 1069–1088.

SLATKIN, M. 1981. Estimating levels of gene flow in natural populations. *Genetics 99*, 323–335.

WOODRUFF, D. S. 1981. Towards a genodynamics of hybrid zones: studies of Australian frogs and West Indian land snails. In Atchley, W. R. and Woodruff, D. S. (eds.) *Evolution and Speciation*, Cambridge University Press, Cambridge, U.K.

WOODRUFF, R. C., SLATKO, B. E. AND THOMPSON, J. R. 1982. Factors affecting mutation rate in natural populations. In Ashburner, M., Carson, H. L. and Thompson, J. N. (eds.) *Genetics and Biology of Drosophila, Vol.* 3c, Academic Press, New York.

WRIGHT, S. 1937. The distribution of gene frequencies in populations. *Proc. Nat. Acad. Sci. 23*, 307–320.