

# Environmental correlates of inversion frequencies in natural populations of seaweed flies (*Coelopa frigida*)

R. K. Butlin\* and  
T. H. Day†

\* Department of Zoology, University College,  
P.O. Box 78, Cardiff, U.K.

† Department of Genetics, Queen's Medical Centre,  
University of Nottingham, Nottingham, U.K.

Two populations of seaweed flies (*Coelopa frigida*) from the north-east coast of England were sampled at approximately monthly intervals. Single samples were also collected from populations on the south coast. The allele and genotype frequencies at the alcohol dehydrogenase locus, known to be associated with the large  $\alpha/\beta$  chromosomal inversion, were determined for each of the 34 samples.

Aspects of the habitats were quantified at the time of sampling, as were 12 variables extracted from meteorological data. The genetic, ecological and climatic data are used to test predictions deriving from laboratory experiments and from previous, smaller scale, sampling programmes. There was an excess of heterokaryotypes over Hardy–Weinberg expectations which was stable over time. While the frequencies were close to the predicted equilibrium values, the rate of return following departures from the equilibrium was more rapid than predicted from the operation of heterokaryotypic advantage alone. Estimates of the effective population size were very much lower than census estimates. Karyotype frequencies cycled seasonally and were associated with both temperature and the incidence of gales. The interactions of ecological and life-historical factors with the observed genetical variation are discussed.

## INTRODUCTION

One of the principal aims of ecological genetics is to explain the frequencies of genetic variants in natural populations in terms of environmental variables. Despite many years of endeavour there are still rather few examples for which a clear link has been established between specific environmental variables and allele frequencies. Among the best known cases are the visible polymorphisms in *Biston betularia* (Brakefield, 1987) and *Cepaea nemoralis* (Cain and Sheppard, 1954), allozyme polymorphisms in *Mytilus edulis* (Koehn, 1978) and *Drosophila melanogaster* (van Delden, 1982), and of course, sickle cell disease in man (Allison, 1954). Even in these species a complete description of the selection pressures involved is not yet available. Unlike enzyme polymorphisms, there is generally no question about the existence of selection maintaining chromosomal inversion polymorphisms, and the selection pressures may be substantial. Inversion systems should, therefore, be good candidates for studying the ecological factors responsible for selection but this approach has not been exploited because of the difficulty of defining the natural habitat of species such as *D. pseudo-*

*obscura* on which the classical studies of inversion polymorphism were conducted (Wright and Dobzhansky, 1946).

The identification of selective agents influencing a polymorphism must involve at least two phases: correlation of allele or genotype frequencies in natural populations with environmental variables, and manipulation of environmental variables in laboratory populations. Only a combination of these two types of observation can establish a link between an environmental variable and the polymorphism, and even then it may be necessary to demonstrate the mechanistic relationship between the gene and the selective agent to establish that the link is causal (Clarke, 1979).

The habitat of the seaweed fly, *Coelopa frigida*, is clearly defined; it lives only in piles of decaying seaweed cast up on beaches by tides and winds (Dobson, 1974a, 1976). So far all populations studied in northern Europe are polymorphic for a large inversion on chromosome I (Butlin *et al.*, 1982a; Day *et al.*, 1983; and unpublished data). Inversion frequencies in natural populations strongly suggest that the polymorphism is maintained by selection, and laboratory experiments indicate that at least three types of selection

may operate: heterokaryotypic advantage in egg to adult viability (Butlin *et al.*, 1984), variation in development time (Day *et al.*, 1980), and sexual selection (Butlin *et al.*, 1982*b*, Day *et al.*, 1987). This is therefore a promising system in which to study the connection between environmental parameters and inversion frequencies. In this paper we report on a series of samples from natural populations and examine the extent to which selection pressures identified in the laboratory can explain variations in the inversion frequencies observed.

## METHODS

Samples of seaweed flies were collected as larvae from ten sites on the north-east and south coasts of England from March 1981 to March 1982 (see Appendix for details of sites). Two sites were sampled repeatedly (St. Mary's Island, SM, and Whitburn, WH) to give a total of 34 samples (12 from SM, 10 from WH and 12 others). Larvae were kept in the seaweed in which they were collected and emerging adults were removed and stored at  $-25^{\circ}\text{C}$ . A sub-sample from the total pool of adults emerging from a collection (often several thousand individuals) was used for genetic analysis. (This is the "larval" method of Butlin *et al.*, 1982*a*).

Chromosome I in *C. frigida* has two arrangements, designated  $\alpha$  and  $\beta$ , related by a complex pattern of overlapping inversions (Aziz, 1975). To score flies for the inversion we took advantage of a close association between the inversion and the *Adh* locus (Day *et al.*, 1982). The *Adh-B* allele is always found with the  $\alpha$  arrangement and the *Adh-D* allele with the  $\beta$ . Three other *Adh* alleles are present; the *Adh-C* allele which occurs with both arrangements, and the rare *Adh-A* and *-E* alleles whose inversion associations are unknown. Methods for electrophoresis have been described by Butlin *et al.*, (1982*a*). For reasons outlined below only data on females were used in the analysis although electrophoresis was carried out on both sexes. A total of more than 3700 individual females were scored from the 34 collections; sample sizes are given in the Appendix.

In addition to these genetic data, information was also collected at the time of sampling on a range of environmental variables as follows:

- (1) date of collection
- (2) composition of the seaweed bed (mainly *Laminaria*, *Laminaria* and *Fucus*, or mainly *Fucus*)
- (3) area of seaweed bed (large, medium, or small)

- (4) depth of seaweed bed ( $>1\text{m}$ ,  $0.2\text{--}1.0\text{m}$ ,  $<0.2\text{m}$ , or "strings" among pebbles)
- (5) extent of decomposition (fresh, moderate, very decomposed, extremely decomposed, extremely decomposed loose and dry, or extremely decomposed wet and foul smelling)
- (6) temperature of seaweed bed relative to air temperature (cold, cool, warm, hot, or very hot)
- (7) wetness of seaweed bed (dry, moderate, or wet)
- (8) size of *Coelopa* larvae (small, medium, or large)
- (9) larval density (sparse, moderate, dense, or very dense)
- (10) presence or absence of *Coelopa* adults
- (11) proportion of *Coelopa pilipes* in the emerging adults (scored after collection of adults in the laboratory).

Meteorological data were obtained from published sources (Meteorological Office, Bracknell, Berks.) for Tynemouth, Tyne and Wear (north east coast) and Twist, Devon (south coast). The following parameters have been considered:

- (12) monthly mean temperatures ( $T$ )
- (13) total monthly rainfall ( $R$ )
- (14) number of days per month on which a gale was blowing, i.e. wind speed exceeded 34 knots ( $G$ )
- (15) total hours of sunshine per month ( $S$ )

Each of these values was cumulated over one, two or three months to test for delayed or averaged effects, giving 12 environmental parameters  $T_1$ ,  $T_2$ ,  $T_3$ ,  $R_1$ ,  $R_2$ ,  $R_3$ ,  $G_1$ ,  $G_2$ ,  $G_3$  and  $S_1$ ,  $S_2$ ,  $S_3$ .

## PREDICTIONS

The total data set consists of 34 samples for each of which the frequencies of six common genotypes and two rare alleles are available, plus up to 23 environmental parameters. It is clearly not productive to test for associations of each genetic parameter with each environmental parameter because many such tests would give significant results due to chance alone. We have, therefore, made specific predictions about the expected associations based on information already available concerning the  $\alpha/\beta$  inversion system. This places us in a much stronger position when testing for associations in the field data.

Before describing these predictions it will be helpful to consider a caveat about the sampling

procedure. Seaweed beds were sampled at random times in their life span, which is principally determined by the cycle of spring tides and by gales (Dobson, 1974*a*). The chromosome I inversion is known to influence development time (Day *et al.*, 1980) and so there is a very real risk of biased sampling. We attempted to reduce this bias by collecting from beds containing only larvae—that is before pupation had started. However, this was not always easy to ascertain and so it is possible that some samples were deficient in the fast developing  $\beta\beta$  karyotype. On the other hand, when seaweed was collected with very young larvae there may have been some truncation of the emergence period due to sub-optimal laboratory conditions, such as the build up of waste products or predator populations. This could cause under-representation of the slow developing  $\alpha\alpha$  karyotype.

Development time variation is much greater in males than females (Day *et al.*, 1980) and so sampling errors of this kind could be expected to be greater in males. Consequently we have restricted our analysis entirely to females. This causes a substantial reduction in sample sizes but, nevertheless, reduces the variability of samples considerably.

#### Prediction 1

In laboratory tests  $\alpha\beta$  heterokaryotypes have a higher egg to adult viability than  $\alpha\alpha$  or  $\beta\beta$  homokaryotypes. The  $\alpha\alpha$  karyotype generally has the lowest viability (Butlin *et al.*, 1984). This predicts that in field samples there should be an excess of  $\alpha\beta$  (*Adh-BD*) over Hardy-Weinberg expectations with corresponding deficiencies of  $\alpha\alpha$  (*BB*) and  $\beta\beta$  (*DD*), the greater deficiency being for *BB*. Assuming no direct effect of the *Adh* locus, the genotypes *BC*, *CD* and *CC* should show little departure from Hardy-Weinberg expectations. These predictions have usually been borne out in previous samples (Butlin *et al.*, 1982*a*, 1984) but their generality can be tested with this much larger data set.

#### Prediction 2

Viability differences increased with larval density in laboratory experiments as well as a small series of field samples (Butlin *et al.*, 1984). This predicts that there will be a relationship between larval density and deviations from Hardy-Weinberg expectations.

#### Prediction 3

The viabilities observed in laboratory trials predict an equilibrium frequency of  $\alpha$  in the region 0.4–0.5 in the absence of other selection pressures. This should be reflected in a similar frequency of the *Adh-B* allele amongst *Adh-B* and *-D* alleles, although these allele frequencies can vary to some extent independently of inversion frequencies because of the presence of the *Adh-C* allele on both arrangements.

Viability selection also predicts that in sequences of samples from a single location the change in allele frequency from one generation to the next  $\Delta q$ , should be related to the starting allele frequency. As all samples are likely to be at frequencies close to equilibrium, there should be a negative linear regression of  $\Delta q$  on  $q$  with an equilibrium point  $\hat{q}$  in the region 0.4–0.5 (where  $q$  is the frequency of  $\alpha$ , estimated as the frequency of *Adh-B* amongst *Adh-B* and *-D* alleles).

#### Prediction 4

Census population sizes of *C. frigida* are very large, often in the region of  $10^9$ – $10^{10}$  larvae in a seaweed bed of 1000 m<sup>3</sup> (Butlin *et al.*, 1984 and unpublished data). If the effective population size is also large then the regression in Prediction 3 should explain most of the variation in allele frequencies from generation to generation for *Adh-B* and *Adh-D*, while the rarer alleles *Adh-A*, *-C* and *-E*, which may be much less affected by selection on the inversion, should show little variation in frequency with time. However, the extent of inter-colony migration in *C. frigida* is unknown (but see Dobson, 1974*a*), so that no firm prediction can be made about geographical variation for these alleles. *C. frigida* is quite a strong flier and can on occasions cover large distances (Oldroyd, 1954) so that one might expect little variation between sites on one coast (either south or north-east), but potentially substantial variation between coasts.

#### Prediction 5

The variation in development time of the karyotypes ( $\beta\beta$  developing fastest and  $\alpha\alpha$  slowest, Day *et al.*, 1980) is expected to interact with the life span of seaweed beds to influence inversion frequencies. Seaweed beds are typically deposited by high spring tides and removed by the next cycle of spring tides 28 days later. On this basic cycle is superimposed the effect of storms, with more frequent storms producing shorter lifetimes on

average. High temperatures also affect the effective longevity of seaweed beds by speeding decomposition such that seaweed fly eggs are laid sooner, and by accelerating fly development.

We therefore predict that the frequency of the  $\alpha$  inversion (*Adh-B*) should increase when the ambient temperature is high and when storm frequencies are low, and should decrease in the opposite conditions. This may generate an annual cycle in inversion frequencies with  $\alpha$  increasing in summer and decreasing during winter. It may also produce a coastal difference if, for example, the south coast is more sheltered or warmer.

#### Prediction 6

*C. pilipes* develops more slowly than *C. frigida* under similar conditions (Dobson, 1974b; Butlin, 1983). This predicts a correlation between the frequency of *C. pilipes* and the frequency of the  $\alpha$  inversion.

These predictions have been tested with the data presented in the Appendix. In addition, other environmental variables have been scanned for effects on gene and genotype frequencies but, as noted above, results from this approach carry much less weight than the tests based on *a priori* predictions.

## RESULTS

In the following analyses the individual samples have been treated as independent data points. In reality this is clearly not always true since the gene frequencies in a population at any one time must depend in part on the frequencies in preceding generations. In practice this is only likely to be a problem in the two populations sampled at frequent intervals—St. Mary's Island and Whitburn. The extent of the interdependence of samples can be estimated by autocorrelation. The correlation between frequencies of *Adh-B* in samples taken from the same population less than 40 days apart (and probably from consecutive generations) using samples from both SM and WH is  $r = 0.503$  ( $n = 18$ ,  $P < 0.05$ ). Hence there is significant autocorrelation which must be taken into consideration in some of the following analyses. However, it only accounts for one quarter of the variation in allele frequencies leaving substantial variation which may be correlated with the environmental parameters of interest.

The data have been analysed without transformation except with *C. pilipes* frequencies which

span a very wide range. In this case arcsine transformation of  $pilipes/(pilipes + frigida)$  was used. Otherwise transformation of the gene and genotype frequencies had little effect because sample sizes were large and homogeneous, and frequencies were generally in the range 5–60 per cent.

#### Prediction 1

Deviations from Hardy-Weinberg expectations have been measured using the proportional departure from expectation [(observed number – expected number)/expected number] for each genotype. With the exception of the *CC* genotypes for which sample sizes were small, this parameter was found to follow an approximately normal distribution. The observed mean deviations are presented in table 1.

**Table 1** Departures from Hardy-Weinberg expectations. The departures are expressed as the mean ( $\pm$ standard error) of the parameter (observed – expected)/expected number for each genotype. Means were calculated for samples from St. Mary's Island (SM), Whitburn (WH) and all samples

<i>Adh</i> genotype	All samples	SM	WH
<i>BB</i>	-0.205 ( $\pm 0.033$ )	-0.264 ( $\pm 0.049$ )	-0.147 ( $\pm 0.061$ )
<i>BC</i>	0.017 ( $\pm 0.048$ )	0.022 ( $\pm 0.065$ )	0.129 ( $\pm 0.111$ )
<i>BD</i>	0.188 ( $\pm 0.024$ )	0.208 ( $\pm 0.028$ )	0.144 ( $\pm 0.035$ )
<i>CD</i>	0.067 ( $\pm 0.038$ )	0.061 ( $\pm 0.047$ )	0.048 ( $\pm 0.093$ )
<i>DD</i>	-0.129 ( $\pm 0.022$ )	-0.138 ( $\pm 0.028$ )	-0.071 ( $\pm 0.030$ )

The data are clearly in full agreement with the predictions. There is a highly consistent excess of *BD*, and deficiencies of both *BB* and *DD* genotypes. The deficiency is greater for the *BB* genotype. *BC* and *CD* genotypes do not differ significantly from expectation.

#### Prediction 2

Deviations from Hardy-Weinberg expectations have been compared among the four larval density classes for the genotypes *BB*, *BD* and *DD* (table 2). No significant variation was observed among classes although the smallest departures from expectation were in the lowest density class in accordance with the prediction. This was also true of the SM samples alone, the only population represented in all density classes.

A significant effect of larval density in natural seaweed beds has been found by Butlin *et al.*, (1984). They used a more objective method of estimation of larval density that allowed for patch-

**Table 2** Mean departures from Hardy-Weinberg expectations in all samples at different levels of larval density

Adh genotype	Larval density				$F_{3,11}$
	Sparse	Moderate	Dense	Very dense	
BB	-0.139	-0.289	-0.192	-0.269	0.42
BD	0.144	0.238	0.232	0.244	0.30
DD	-0.071	-0.134	-0.160	-0.127	0.23

ness in the distribution of larvae within seaweed beds. The present method may not have been accurate enough to distinguish variation in density, especially when densities were more than moderate.

**Prediction 3**

Allele frequencies at the *Adh* locus are given in table 3(a) together with the frequency of *B* alleles amongst *B* and *D* alleles alone ( $B^*$ ). ( $B^*$  is calculated ignoring the *Adh-A*, *-C* and *-E* alleles whose associations with the  $\alpha$  and  $\beta$  inversions remain unknown). This is, therefore, the best available estimate of the  $\alpha$  arrangement frequency.  $B^*$  is both very stable, and lies within the ranges predicted both by laboratory viability estimates and by the observed departures from Hardy-Weinberg expectations.

**Table 3** Variation in *Adh* allele frequencies

(a) Mean frequencies ( $\pm$  standard error)

Allele	All samples	SM	WH
<i>B</i>	0.391 ( $\pm 0.0082$ )	0.380 ( $\pm 0.0141$ )	0.396 ( $\pm 0.0163$ )
<i>C</i>	0.099 ( $\pm 0.0039$ )	0.095 ( $\pm 0.0056$ )	0.103 ( $\pm 0.0237$ )
<i>D</i>	0.493 ( $\pm 0.0088$ )	0.511 ( $\pm 0.0164$ )	0.473 ( $\pm 0.0177$ )
$B^*$	0.442 ( $\pm 0.0091$ )	0.427 ( $\pm 0.0167$ )	0.456 ( $\pm 0.0192$ )

(b) Regressions of  $\Delta q$  on  $q$  where  $q$  = frequency of *B*/(*B* + *D*) in generation  $n$  and  $\Delta q = q_{(n+1)} - q_n$

SM	$\Delta q = 0.279 - 0.669q, t = -1.68, n = 8, P > 0.10$
WH	$\Delta q = 0.527 - 1.11q, t = -6.68, n = 6, P < 0.01$
SM and WH	$\Delta q = 0.295 - 0.658q, t = -2.61, n = 14, P < 0.05$

(c) Estimates of selection coefficient (see text for explanation)

	$w_\alpha \cdot w_\beta$ (95% confidence limits)	$s_\alpha$	$s_\beta$	$q$
SM	0.488 (0.486-0.489)	0.35	0.25	0.417
WH	0.480 (0.477-0.482)	0.32	0.29	0.476

Sufficient data were available to test the regression of  $\Delta q$  on  $q$  for both SM and WH (table 3(b)). Neither the slopes ( $t = 1.02, P > 0.20$ ) nor the intercepts ( $t = 0.08, P > 0.20$ ) differ significantly between populations. These equations predict an equilibrium frequency ( $\hat{q}$ ) for  $\alpha$  of 0.417 for SM, 0.475 for WH and 0.448 for the combined regression. Again this is fully consistent with other predictions and with the observed frequencies, including the higher frequency for WH (table 3(a)).

However, a surprising feature of the regressions is the steepness of the slopes which are much greater than expected, and much greater than in comparable systems (e.g., Wright and Dobzhansky, 1946). The expected slope can be estimated from the deviations from Hardy-Weinberg expectations and the observed  $\hat{q}$  on the assumption that selection was complete when the samples were taken. If  $w_\alpha (= 1 - s_\alpha)$  is the fitness of  $\alpha\alpha$ , and  $w_\beta (= 1 - s_\beta)$  the fitness of  $\beta\beta$  relative to  $\alpha\beta$ , and  $P, H,$  and  $Q$  are the genotype frequencies of  $\alpha\alpha, \alpha\beta$  and  $\beta\beta$  respectively after selection, then  $PQ/(H/2)^2 = w_\alpha w_\beta$  (Falconer 1981).  $w_\alpha \cdot w_\beta$  has been estimated from the data for SM and WH and can be used with the estimates of  $\hat{q}$  [ $= s_\beta/(s_\alpha + s_\beta)$ ] for heterozygote advantage] to calculate the values of  $s_\alpha$  and  $s_\beta$  (table 3(c)). These estimates, which appear reasonable in relation to laboratory studies of viability, predict a slope of the regression of  $\Delta q$  on  $q$  of about  $-0.15$  to  $-0.20$  in the region of  $\hat{q}$ . This is much less than the observed slope.

The slope of the regression may be inflated by errors in the estimation of allele frequencies in a few samples since aberrant points will be followed by an apparent return towards the equilibrium frequency. However, we consider that with the precautions outlined above, we have achieved a reliable measure of gene frequency. Another possibility is that more than one generation might have elapsed between consecutive samples; this is most unlikely since only samples collected less than 40 days apart were used. An alternative explanation of the regressions is that additional selection pressures are operating. Extremely strong heterozygote advantage would be needed to produce the observed slope; another possibility is that some types of frequency-dependent selection may be operating.

**Prediction 4**

Effective population sizes can be estimated from temporal sequences of allele frequencies using the

method of Pollak (1983). This method was designed for use with data over several generations and so is particularly appropriate here. Pollak's sampling method 2 corresponds closely to our sampling. Using data on all five *Adh* alleles and all samples, the estimated values of  $N_e$  are  $985 \pm 3375$  for SM and  $497 \pm 1339$  for WH. The errors on these values are large but nevertheless the values are strikingly low by comparison with the census estimates in excess of  $10^6$ .

Pollak's method assumes that at most weak selection operates on the loci studied. This is clearly not the case for the *B* and *D* alleles whose frequencies are influenced by selection on the inversion, but the assumption may be more valid for the other three alleles. If the method is adapted by allowing for predicted changes in the frequencies of *B* and *D* from the values of  $s_\alpha$  and  $s_\beta$  above, the new values of  $N_e$  are  $4738 \pm 71,460$  for SM, and  $859 \pm 3757$  for WH. Since selection accounts for some of the changes in allele frequencies, these estimates are higher. However, they remain much smaller than the estimates derived from census studies. This may reflect bottlenecks of *Coelopa* populations when there are long intervals between suitable seaweed deposits. A similar contrast between census and effective population sizes in *Coelopa* was found by Burnet (1961) using entirely different methods and is typical of such studies in general (Pollak, 1983).

Geographical variation in frequencies of the *A*, *C* and *E* alleles which were assumed to be little affected by selection on the inversion, indicates substantial isolation between the north-east and south coasts (table 4). This is expected because they are separated by several hundred kilometres of unsuitable coastline. There was also a difference in the frequency of the *A* allele between SM and WH. This is much more surprising as the two sites are separated by less than 20 km, and intervening colonies do exist. It may reflect the fact that inter-colony migration in *Coelopa* is sporadic—flies normally move very short distances (Dobson, 1974a) although occasional movement over long distances can occur (Oldroyd, 1954).

#### Prediction 5

The possibility of seasonal fluctuations in inversion frequencies was examined by including the month of the year as a factor in a stepwise multiple regression analysis (discussed further below). "Month" was in all cases the first term to be added and it explained about 40 per cent of the variance in *Adh-B* and *-D* allele frequencies (table 5). There

Table 4

(a) Frequencies of rare *Adh* alleles

	No. of alleles		Total sample
	<i>A</i>	<i>E</i>	
SM	17	7	3646
WH	16	8	1750
Other NE coast sites	3	0	342
S coast sites	0	6	1664
Total	36	21	7402
$\chi^2_1$ (NE coast vs. S coast)	10.45**	0.46	
$\chi^2_1$ (SM vs. WH)	3.91*	3.00	

$\chi^2$  values have been calculated by comparing the number of rare alleles (*A* or *E*) with the number of remaining alleles. \*  $P < 0.05$ , \*\*  $P < 0.01$ .

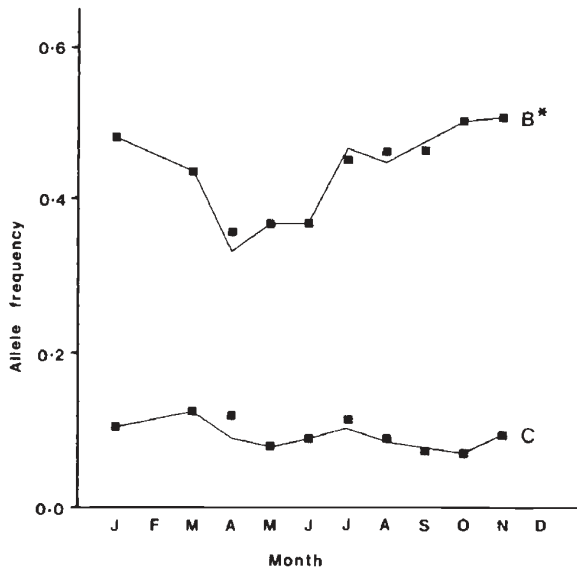
(b) Frequencies of *Adh-C*

	Mean <i>Adh-C</i> frequency ( $\pm$ standard error)
SM	0.0954 ( $\pm 0.0056$ )
WH	0.1029 ( $\pm 0.0075$ )
All NE coast sites	0.1019 ( $\pm 0.0046$ )
All S coast sites	0.0905 ( $\pm 0.0065$ )
All sites	0.0992 ( $\pm 0.0039$ )

Table 5 Results of the stepwise multiple regression analysis

	Parameters fitted in order	Total variance explained (per cent)	<i>F</i> (df)
All samples			
<i>B</i>	Month	40.1	4.13* (9, 12)
	Area	60.7	4.67* (2, 12)
<i>C</i>	Month	17.9	1.56 (9, 14)
	Area	44.4	5.25** (9, 12)
<i>D</i>	Month	67.8	6.09* (2, 12)
	Area	40.7	4.96** (9, 12)
<i>B/(B + D)</i>	Month	67.1	6.61* (2, 12)
	Area		
SM and WH			
<i>B</i>	Month	45.6	6.80* (8, 7)
	Area	74.3	7.33* (2, 7)
	G3	76.9	1.91 (1, 7)
<i>C</i>	Month	23.5	1.69 (8, 10)
	Area	48.4	8.11** (8, 7)
<i>D</i>	Month	80.1	9.01* (2, 7)
	Area	80.2	1.04 (1, 7)
	G3	80.2	1.04 (1, 7)
<i>B/(B + D)</i>	Month	44.3	8.94** (8, 7)
	Area	78.0	9.73** (2, 7)
	G3	80.4	1.99 (1, 7)

\*  $P < 0.05$ , \*\*  $P < 0.01$ .



**Figure 1** Mean monthly *Adh* allele frequencies for St. Mary's Island and Whitburn pooled (continuous line), and for all populations pooled (■). B\* is the frequency of  $B/(B+D)$ .

was a clear seasonal cycle in frequencies with an amplitude of about 0.1. A sharp fall is apparent from December to April, and steady rise for the remainder of the year (fig. 1). This is precisely the pattern predicted from the data on development times of the inversion types.

The prediction that the south coast should have a higher mean  $\alpha$  frequency is uncertain since there is much variation on both coasts in temperature and shelter from gales. In addition, the test is somewhat insensitive since fewer samples are available from the south coast than from the north east coast. The overall mean proportion of *Adh-B* alleles amongst *B* and *D* alleles was  $0.452 \pm 0.016$  for the south coast, and  $0.439 \pm 0.011$  for the north east coast. This difference is in the expected direction but is small, especially by comparison with the difference between SM and WH (table 3(a)).

Regression of *Adh-B*, *-D* and  $B/(B+D)$  frequencies on each of the twelve weather variables were calculated for both the full data set and for SM and WH alone. Few significant values and no patterns were obtained for the sunshine or rainfall parameters. However, patterns were evident for both gales and temperatures. Each of these parameters when summed over the three months preceding a sample gave significant regressions ( $P < 0.05$  in all six comparisons for *T3*, and in five out of six comparisons for *G3*). Two regressions for *G2* and one for *G1* were also significant but none

for *T1* nor *T2*. In all cases *Adh-B* frequency increased at higher temperatures and with less frequent gales.

Clearly both gales and temperatures are correlated with month of the year and so their regressions do not provide evidence for a causal association. In the stepwise multiple regression analysis (see below) the effects of *G3* and *T3* were never significant after the term for the month had been added. Only *G3* explained any additional variation and in this case the improvements were very slight. On this basis there is no way to distinguish the effects of gales or temperatures in inducing the seasonal cycle.

#### Prediction 6

Regression of allele frequencies on the proportion of *C. pilipes* in samples were in the directions expected from the longer development time of *C. pilipes*, but were not statistically significant (table 6). No annual cycle was detected in the data for proportions of the two species although this has been suggested in the past (Dobson, 1974a), and might also be expected using the same argument as predicted a cycle in inversion frequencies. Presumably other factors influence the relative numbers of the species.

**Table 6** Regressions of *Adh* allele frequencies (in *C. frigidula*) on frequencies of *Coelopa pilipes*

<i>Adh</i> allele	Regression coefficients ( $\pm$ standard error)	
	All samples	SM and WH
<i>B</i>	0.0391 ( $\pm 0.0245$ )	0.0545 ( $\pm 0.0276$ )*
<i>C</i>	-0.0122 ( $\pm 0.1021$ )	-0.0140 ( $\pm 0.0124$ )
<i>D</i>	-0.0424 ( $\pm 0.0266$ )	-0.0612 ( $\pm 0.0327$ )
$B/(B+D)$	0.0477 ( $\pm 0.0275$ )	0.0666 ( $\pm 0.0330$ )*

\*  $P < 0.10$

#### Other environmental parameters

All possible relationships between environmental parameters and both allele frequencies and deviation from Hardy-Weinberg expectations were assessed using the full data set. It is to be expected that this large number of tests would generate a number of spuriously significant results. The significant results were, therefore, scanned for consistent patterns. No such patterns were detected for the great majority of variables, including: algal composition, depth of deposit, state of decomposition, the temperature and wetness of seaweed, larval size, presence of adults, extent of rainfall

and sunshine. However, potentially meaningful patterns were revealed for two variables: seaweed bed area and larval density.

Seaweed bed area showed significant effects ( $P < 0.05$ ) for *Adh-B* frequency,  $B/(B+D)$ , *BB* excess, and *BD* excess over Hardy-Weinberg expectations. Smaller beds were associated with lower *B* frequencies, greater *BB* deficiencies and greater *BD* excesses than were observed with larger beds. This is comparable to the effect predicted for increased larval density and it may be true that small seaweed beds tend to have higher larval densities, or that they are more "stressful" in other aspects.

*Adh-B* frequency varied significantly among larval density classes, with the major differences existing between the lowest density and the other three density classes. This is similar to the relationship between density and deviations from Hardy-Weinberg excess. However, it is surprising that the higher densities were associated with higher *Adh-B* frequencies:

density 1	$B/(B+D) = 0.383 \pm 0.018,$
density 2	$B/(B+D) = 0.468 \pm 0.026,$
density 3	$B/(B+D) = 0.463 \pm 0.028,$
density 4	$B/(B+D) = 0.428 \pm 0.026.$

#### Stepwise multiple regression

Five parameters suspected of showing effects on allele frequencies were included in a stepwise multiple regression analysis for the allele frequencies *Adh-B*, *-C*, *-D* and  $B/(B+D)$  and for either the full data set or for SM and WH samples separately. At each step the parameter was added that caused the greatest reduction in the residual variance. The parameters included were; month, area, *C. pilipes* frequency, temperature (*T*<sub>3</sub>) and gales (*G*<sub>3</sub>). The results are summarised in table 5.

*Adh-C* frequencies showed no significant associations, whilst *Adh-B* and *-D* frequencies showed reciprocal effects as expected if selection acts on the inversion rather than the *Adh* locus. There was a highly consistent effect of month reflecting an annual cycle in inversion frequencies as described above. The only other term to enter significantly was seaweed bed area, reinforcing the conclusion from single parameter analyses, and demonstrating the independence of this effect from the seasonal cycle. After removing variation due to time of year, the effect of area on  $B/(B+D)$  (relative to "large" seaweed beds) was: "medium"  $-0.11 \pm 0.030$ , "small"  $-0.11 \pm 0.037$ .

#### DISCUSSION

A substantial measure of agreement exists between, on the one hand, the predictions based on laboratory experiments with *C. frigida* and, on the other the patterns of variation in allele and genotype frequencies in natural populations. Heterokaryotype advantage, which can be very strong in laboratory conditions, is clearly a major force in maintaining the  $\alpha\beta$  inversion polymorphism. This is reflected in deviations from Hardy-Weinberg expectations, the stability of inversion frequencies, and a strong tendency for populations to return towards equilibrium frequencies. However, while heterozygote advantage may maintain the polymorphism, it cannot fully explain the patterns of inversion frequencies. Other selective forces must clearly be involved.

Laboratory studies have demonstrated variation in egg-to-adult development times amongst karyotypes and these differences appear to provide the simplest explanation for seasonal variation in inversion frequencies. In winter, when seaweed beds have on average short lifetimes, rapid development will be favoured leading to an increase in the  $\beta$  inversion. The increase in  $\alpha$  during the summer is rather more difficult to explain. Longer seaweed bed lifetimes and higher temperatures allow more flies to complete development but this, in itself, should not select against  $\beta$ . One possibility is that the early-emerging  $\beta\beta$  females lay some eggs in the seaweed bed in which they developed and these have a low probability of completing development. However, development time differences amongst female genotypes are small relative to differences amongst males, especially under natural conditions (Butlin, 1983). Another possibility is that slow developing, large males have a mating advantage as has been observed in some, though not all, laboratory experiments (Butlin *et al.*, 1982b; Day *et al.*, 1987). When seaweed beds are short lived,  $\beta\beta$  males would obtain more matings because of their fast development time, but in long-lived seaweed beds,  $\alpha\beta$  and  $\alpha\alpha$  males would gain an advantage due to their larger size, mediated either by male-male competition or by female preference (Day *et al.*, in preparation). Karyotypes also differ in longevity and fecundity both of which are also associated with adult size (Butlin and Day, 1985). These components of selection may also favour the  $\alpha$  inversion when sufficient time is available for development, or if there are times when no food at all is available.

Butlin (1983) was able to generate a seasonal cycle in inversion frequencies in a simulation



Appendix

Pop. (4)	Date	Adh genotypes (1)										Rare alleles (2)		Sample Sizes		Environment variables (3)								
		BB	BC	BD	CC	CD	DD	A	E	Females	Total	No. of C. pilipes		1	2	3	4	5	6	7	8	9		
SM	18.3.81	21	11	62	0	23	41	2	3	162	288	0	2	2	*	3	3	4	2	*	*	*		
WH	18.3.81	26	12	39	0	18	14	2	2	112	287	586	2	2	1	2	3	4	2	*	*	*		
SM	15.4.81	9	5	71	3	18	50	5	0	159	285	3	2	2	2	2	3	*	*	*	5	*		
RU	20.4.81	14	11	61	2	19	42	0	1	150	287	0	1	3	4	4	1	1	*	*	*	*		
SM	14.5.81	10	9	63	1	18	36	4	1	139	273	0	2	2	2	3	3	2	2	*	*	5		
SS	14.5.81	8	6	71	0	11	42	2	2	140	280	0	2	3	3	3	2	2	2	*	*	1		
SM	18.6.81	7	4	53	0	11	27	4	1	104	217	37	2	3	3	3	4	2	2	*	*	*		
WH	18.6.81	3	9	30	0	6	23	2	0	72	144	34	2	3	3	1	2	3	*	*	*	*		
LU	24.6.81	11	7	30	0	8	11	0	2	67	138	0	*	*	*	*	*	*	*	*	*	*		
PO	26.6.81	24	13	68	1	14	35	0	0	155	309	0	*	*	*	*	*	*	*	*	*	*		
SM	8.7.81	24	16	58	1	13	32	7	2	149	277	5	2	2	3	2	2	2	*	*	3	*		
WH	8.7.81	12	9	25	0	7	14	3	1	69	144	28	2	3	4	4	1	1	3	2	1	*		
RH	8.7.81	6	4	27	1	14	8	3	0	63	126	0	2	3	4	3	3	*	*	2	4	*		
RU	13.7.81	5	3	36	1	9	15	0	2	71	141	0	1	3	4	4	3	1	*	*	4	*		
SM	30.7.81	22	11	60	1	13	34	4	4	144	319	20	2	3	3	3	1	3	3	1	2	1		
WH	30.7.81	11	8	47	4	7	20	1	1	99	144	1	2	3	3	3	1	3	3	1	2	1		
PO	19.8.81	9	5	40	2	5	11	0	1	73	141	1	2	3	3	3	2	2	2	2	3	1		
SM	27.8.81	13	9	72	0	8	31	3	0	135	283	368	3	2	2	2	2	3	3	*	4	0		
WH	27.8.81	10	8	46	2	8	16	4	1	92	143	27	2	2	3	2	4	2	*	*	5	1		
BE	9.9.81	16	4	27	1	5	21	0	2	75	141	16	2	1	2	3	4	2	3	4	2	1		
SM	25.9.81	21	12	72	1	19	20	1	1	146	287	1633	*	*	*	*	*	*	*	*	*	*		
WH	25.9.81	11	6	28	0	5	13	4	0	64	146	175	*	*	*	*	*	*	*	*	*	*		
WH	11.10.81	12	6	29	0	4	13	9	1	70	189	1531	2	2	2	2	3	2	1	*	*	1		
FL	25.10.81	9	5	34	2	9	12	1	0	71	141	75	*	*	*	*	*	*	*	*	*	*		
SM	4.11.81	27	15	68	1	13	20	3	1	146	288	7	2	1	1	6	3	2	*	*	*	*		
WM	4.11.81	12	5	36	0	9	13	1	0	76	135	0	2	2	3	1	2	3	1	*	*	*		
SM	31.12.81	25	15	87	2	19	29	0	0	177	368	0	2	2	2	2	6	3	2	3	2	*		
WH	21.1.82	15	5	47	0	9	27	1	1	105	259	727	*	*	*	*	*	*	*	*	*	*		
SM	21.1.82	18	16	83	2	17	47	0	0	183	366	0	*	*	*	*	*	*	*	*	*	0		
FL	5.2.82	6	3	13	0	4	11	0	0	37	63	140	*	*	*	*	*	*	*	*	*	*		
BE	28.2.82	13	4	51	0	7	12	0	2	88	271	17	*	*	*	*	*	*	*	*	*	*		
LR	28.2.82	12	13	89	0	10	29	0	0	153	302	20	*	*	*	*	*	*	*	*	*	*		
WH	18.3.82	10	8	42	2	19	31	5	3	116	235	25	2	2	2	3	4	3	4	3	5	1		
SM	18.3.82	13	23	86	2	18	37	0	0	179	328	0	2	2	2	2	6	5	3	2	3	1		

(1) Number of individuals, females only.

(2) Number of alleles, both sexes.

(3) Environmental variables: 1, composition of seaweed bed; 2, area of seaweed bed; 3, depth of seaweed bed; 4, extent of decomposition; 5, temperature; 6, wetness; 7, size of *Coelopa* larvae; 8, larval density; 9, presence of *Coelopa* adults. Numbers correspond to the categories described in the Methods section, in the order given. \* indicates missing value.

(4) Populations: SM, St. Mary's Island, Tyne & Wear, NZ350753; SS, St. Mary's Island (South bay), Tyne & Wear, NZ349747; WH, Whitburn, Tyne & Wear, NZ408614; RH, Robin Hood's Bay, North Yorkshire, NZ954048; FL, Flamborough, Humberside, IA 255708; RU, Ruxington, Sussex, TV049014; LU, Lulworth, Dorset, SY825799; PO, Portland, Dorset, SY683687; BE, Beer, Devon, SY229890.

model that incorporated some of these differences—development time, longevity, fecundity and mating success—and in which the probability of stochastic removal of seaweed beds varied with season. The role of temperature was not considered in this simulation. The data available do not provide a distinction between the potential influences of gales and temperature but their effects should be separable given sufficiently detailed data on the sequence of seaweed beds at a site. Continuous observation of a single locality would appear to be the best approach for further field sampling.

An alternative method of separating influences on inversion frequencies may be to apply the present conclusions to a completely separate data set from an area of very different environmental influences. An extensive programme of sampling has been carried out in Sweden which meets these requirements (Day *et al.*, 1983, and unpublished data). In the Baltic region, seaweed beds are more stable than in the North Sea or English Channel probably due to differences in tidal patterns. As the tidal influence declines the frequency of the  $\beta$  inversion initially falls, but then rises again. This can be interpreted in the same way as the annual cycle in British samples: as the longevity of seaweed beds increases, selection initially favours  $\alpha$  due to the advantage of larger size. However, when lifetimes are so long that  $\beta\beta$  flies can regularly complete two generations in a single bed, selection would switch to favouring  $\beta$ . The pattern of inversion frequencies could also be explained by other environmental parameters that vary in the entrance to the Baltic, such as temperature, salinity, and algal composition. Further sampling will clearly be necessary before we can distinguish between these alternatives.

*Acknowledgements* This work was supported by an N.E.R.C. grant to T. H. Day, and by an S.E.R.C. studentship to R.K. Butlin. This support is greatly appreciated.

## REFERENCES

- ALLISON, A. C. 1954. Notes on sickle-cell polymorphism. *Ann. Hum. Genet.*, **19**, 39–57.
- AZIZ, J. B. 1975. *Investigations into chromosomes 1, 2 and 3 of Coelopa frigida*. Ph.D. Thesis, University of Newcastle upon Tyne.
- BRAKEFIELD, P. M. 1987. Industrial melanism: do we have the answers? *Trends Ecol. Evol.*, **2**, 117–122.
- BURNET, B. 1961. On the distribution of recessive embryonic lethals in a natural population of *Coelopa frigida* (Fab.). *Genet. Res., Camb.*, **2**, 249–271.
- BUTLIN, R. K. 1983. *The maintenance of an inversion polymorphism in Coelopa frigida*. Ph.D. Thesis, University of Nottingham.
- BUTLIN, R. K., COLLINS, P. M., SKEVINGTON, S. J. AND DAY, T. H. 1982a. Genetic variation at the alcohol dehydrogenase locus in natural populations of the seaweed fly, *Coelopa frigida*. *Heredity*, **48**, 45–55.
- BUTLIN, R. K. AND DAY, T. H. 1985. Adult size, longevity and fecundity in the seaweed fly, *Coelopa frigida*. *Heredity*, **54**, 107–110.
- BUTLIN, R. K., READ, I. L. AND DAY, T. H. 1982b. The effects of a chromosomal inversion on adult size and male mating success in the seaweed fly, *Coelopa frigida*. *Heredity*, **49**, 51–62.
- BUTLIN, R. K., COLLINS, P. M. AND DAY, T. H. 1984. The effect of larval density on an inversion polymorphism in the seaweed fly, *Coelopa frigida*. *Heredity*, **52**, 415–423.
- CAIN, A. J. AND SHEPPARD, P. M. 1954. Natural selection in *Cepaea*. *Genetics*, **39**, 89–116.
- CLARKE, B. C. 1979. The evolution of genetic diversity. *Proc. R. Soc. Lond. B.*, **205**, 453–474.
- DAY, T. H., DOBSON, T., HILLIER, P. C., PARKIN, D. T. AND CLARKE, B. C. 1980. Different rates of development associated with the alcohol dehydrogenase locus in the seaweed fly, *Coelopa frigida*. *Heredity*, **44**, 321–326.
- DAY, T. H., DOBSON, T., HILLIER, P. C., PARKIN, D. T. AND CLARKE, B. 1982. Associations of enzymic and chromosomal polymorphisms in the seaweed fly, *Coelopa frigida*. *Heredity*, **48**, 35–44.
- DAY, T. H., DAWE, C., DOBSON, T. AND HILLIER, P. C. 1983. A chromosomal inversion polymorphism in Scandinavian populations of the seaweed fly, *Coelopa frigida*. *Hereditas*, **99**, 135–145.
- DAY, T. H., MILES, S., PILKINGTON, M. D. AND BUTLIN, R. K. 1987. Differential mating success in populations of seaweed flies (*Coelopa frigida*). *Heredity*, **58**, 203–212.
- DOBSON, T. 1974a. Studies on the biology of the kelp fly *Coelopa* in Great Britain. *J. Nat. Hist.*, **8**, 155–177.
- DOBSON, T. 1974b. Mechanisms controlling species composition in natural populations of the seaweed fly *Coelopa*. *J. Nat. Hist.*, **8**, 653–673.
- DOBSON, T. 1976. Seaweed flies (Diptera: Coelopidae, etc.). In Cheng, L. (ed.) *Marine Insects*, North-Holland Publishing Company, Amsterdam. pp. 447–463.
- FALCONER, D. S. 1981. *Introduction to Quantitative Genetics 2nd edn*. Longman, London.
- KOEHN, R. K. 1978. Physiology and biochemistry of enzyme variation: the interface of ecology and population genetics. In Brussard, P. and Solbrig, O. (eds) *The Interface of Ecology and Genetics*, Springer-Verlag, Berlin, pp. 51–72.
- OLDROYD, H. 1954. The seaweed fly nuisance. *Discovery*, **15**, 198–202.
- POLLAK, E. 1983. A new method for estimating the effective population size from allele frequency changes. *Genetics*, **104**, 531–548.
- VAN DELDEN, W. 1982. The alcohol dehydrogenase polymorphism in *Drosophila melanogaster*. Selection at an enzyme locus. *Evol. Biol.*, **15**, 187–222.
- WRIGHT, S. AND DOBZHANSKY, T. 1946. Genetics of natural populations. XII. Experimental reproduction of some of the changes caused by natural selection in certain populations of *Drosophila pseudoobscura*. *Genetics*, **31**, 125–156.