

# Population structure of *Rhagoletis pomonella*, the apple maggot fly

STEWART H. BERLOCHER\* & BRUCE A. McPHERON†

Department of Entomology/320 Morrill Hall, University of Illinois at Urbana-Champaign, 505 S. Goodwin Avenue, Urbana, IL 61801 and †Department of Entomology/501 Agricultural Sciences & Industries Building, Pennsylvania State University, University Park, PA 16802–3088, U.S.A.

*Rhagoletis pomonella*, the apple maggot fly, is the best-known member of the *R. pomonella* species complex. *Rhagoletis pomonella* has attracted interest because of its economic importance and because it has been a central focus of debate over sympatric speciation via host race formation. Based on electrophoretic data for 17 allozyme loci, the population structure of the ancestral hawthorn race of *R. pomonella* is characterized by: (i) strong geographical differentiation across eastern North America ( $F_{ST} = 0.148$ ), (ii) pronounced latitudinal clines in allele frequencies at three loci, *Aat-2*, *Dia-2* and *Had*, and (iii) very strong linkage disequilibrium involving two of these loci, *Dia-2* and *Aat-2*, which are closely linked. The 10-fold greater level of population differentiation in *R. pomonella* compared with its close relative *R. mendax* may be related to its greater flexibility in life history timing.

**Keywords:** allozymes, population genetics, population structure, *Rhagoletis*, sympatric speciation, Tephritidae.

## Introduction

The tephritid fruit flies of the *Rhagoletis pomonella* species group (henceforth the *pomonella* group) have for almost three decades been the touchstone of debate over sympatric speciation via host race formation, a mode of speciation in which reproductive isolation evolves not through geographical isolation, but as a result of the adaptation of parasites to new hosts (Bush, 1966, 1992, 1994; Mayr, 1963; Futuyma & Mayer, 1980). The Bush model of sympatric speciation via host races hypothesizes that genotypes that seek out and survive on new hosts can produce a new population — a host race — that is at least partially reproductively isolated from the original population on the ancestral host. One host race formed in historical time, the apple race of *R. pomonella* (Walsh), has been thoroughly examined and found to be distinct from the ancestral hawthorn race in behavioural (Prokopy *et al.*, 1987), phenological (Smith, 1988; Feder *et al.*, 1993, 1994) and allozyme (Feder *et al.*, 1988, 1990a,b; McPheron *et al.*, 1988b) characteristics. Recent work (Feder *et al.*, 1994) has shown by mark–release–recapture that only about 5 per cent of matings are between apple and hawthorn populations.

*Rhagoletis pomonella* is apparently the only species in the *pomonella* group (Bush, 1966) that has formed host races on domesticated plants in historical time (Bush, 1966; Prokopy & Berlocher, 1980). One point of reference in attempting to correlate the fly's 'host race potential' with other aspects of its biology and genetics is its close relative, the blueberry maggot *R. mendax* Curran. *Rhagoletis mendax* has very little population structure (at least for allozymes) at any level — among individual host plants, between different host species, or across eastern North America. This lack of structure has been suggested to be the result of inflexible diapause characteristics, which may constrain it to a single host species over much of its geographical range (Berlocher, 1995; Payne & Berlocher, 1995b). *Rhagoletis pomonella*, on the other hand, has pronounced allozyme population structure, from the level of individual conspecific host plants (McPheron *et al.*, 1988a; Feder *et al.*, 1990b) to latitudinal allele frequency clines in the northern part of its range (Feder & Bush, 1989; Feder *et al.*, 1990a), to continental-scale geographical differentiation over eastern North America (initial estimate of  $F_{ST} = 0.220$ , McPheron, 1987). Feder *et al.* (1993) have argued that *R. pomonella* has a flexible diapause strategy, and have shown that different allozyme genotypes are associated with different

\*Correspondence.

diapause and eclosion phenologies. This implies that allozyme loci are linked to, and are in linkage disequilibrium with, genes involved in diapause control, so that phenological adaptation indirectly produces population structure at allozyme loci (although some direct role for selection on the allozymes themselves is not ruled out). For brevity we henceforth refer to this explanation for allozyme differentiation in *R. pomonella* as the 'climate-linkage hypothesis'. The climate-linkage hypothesis is compatible with the observation that *R. pomonella* infests many species of its native host hawthorn (Rosaceae: *Crataegus*; Bush, 1966; Berlocher & Enquist, 1993), and many of these species have different fruiting phenologies (Berlocher & Enquist, 1993).

In this paper we examine geographical population structure in *R. pomonella* over a much greater geographical range than has previously been studied, and include partial data from central Mexico. The most striking finding is that the clines discovered by Feder and colleagues (Feder & Bush, 1989; Feder *et al.*, 1990a) in the northern part of the range of *R. pomonella* continue linearly to the southern end of the range in eastern North America, where alleles which are almost fixed in the north are totally absent. The presence of 'northern' alleles at relatively high frequencies was found to correlate with the range limits of the apple host race.

#### *Biology and systematics of R. pomonella*

The life cycle of *R. pomonella* is typical of most *Rhagoletis* species (Boller & Prokopy, 1975). The insect is univoltine, with eclosion of adults in the summer as host fruits ripen, mating on host plants, oviposition into fruit, larval development in a single fruit and pupal diapause and overwintering in the soil. The native range of *R. pomonella* is eastern North America and Mexico (Bush, 1966). *Rhagoletis pomonella* is a serious agricultural pest because of the existence of the apple race. All species in the *pomonella* group (Bush, 1966; Berlocher *et al.*, 1993; Payne & Berlocher, 1995a) are morphologically very similar 'sibling' species (Bush, 1966).

### Materials and methods

#### *Data sets*

Data from five sources, referred to henceforth as data sets 1–5, are used in this paper. All samples are from the native host plant, hawthorn. The primary data, on which almost all statistical analysis was carried out, are in data sets 1 and 2. Data set 1 (populations 1–16 in Table 1 and the Appendix)

consists of 14 new population samples and two previously unpublished ones (McPheron, 1987). Data set 2 consists of published data (McPheron, 1990) on six additional populations (Amherst, MA; Ithaca, NY; East Lansing, MI; Hardin County, OH; Hancock County, IN; and Wayne County, IL). The same 17 loci were studied in the same laboratory for all populations of both data sets 1 and 2, using the same buffers for all but one locus. The enzyme that was not electrophoresed identically is the protein encoded by *Mpi*, run on buffer 1 for the new samples instead of buffer 2 of McPheron (1990). These buffers have very similar pHs (pH 8.6, buffer 1; pH 8.4, buffer 2), and although the mobilities were different the same alleles were observed (for brevity, we equate the terms allele and electromorph). The *Mpi* mobilities (and those of all other loci) used in the present paper are the same as in all papers of the current series (Berlocher, 1995; Payne & Berlocher, 1995a). The equivalent *Mpi* alleles are 63, 70, 85, 100 and 112 for the new mobilities and 31, 37, 70, 100 and 125 for those of McPheron (1990). *Dia-1* was monomorphic and is not included in the Appendix.

Data sets 3, 4 and 5 contain either partial data or were obtained for a different set of loci under a different set of electrophoretic conditions, and are therefore used only to supplement and extend the primary statistical analysis carried out on sets 1 and 2. Data set 3 (population 17 in Table 1 and the Appendix) consists of unpublished data from a population from central Mexico, obtained under the same electrophoretic conditions as data sets 1 and 2, but including only seven loci (Berlocher, 1976). Data set 4 is published frequency data by Feder & Bush (1989): Cadillac, MI; Ephraim, WI; Hatley, WI and Waukegan, IL. These sites were added to increase the number and maximize the geographical dispersion of northern sites in studying latitudinal clinal structure. For the clinally varying *Aat-2*, *Dia-2* and *Had*, and the highly structured *Acon-2*, the same common alleles were observed using both the Feder/Bush and Berlocher/McPheron buffers (Feder & McPheron, unpublished observations). Data set 5 consists of unpublished frequencies for the same four loci listed above for a site in Menasha, MN (Feder, unpublished observations):  $Aat-2^F$  (see Results) = 0.849;  $Dia-2^{100}$  = 0.884;  $Had^{125}$  = 0.888;  $Acon-2^{89}$  = 0.027,  $Acon-2^{95}$  = 0.690.

#### *Sampling*

Sampling for data set 1 stressed previously under-sampled south-eastern North America. Most insects

**Table 1** Sample information for *Rhagoletis pomonella* collection sites not described elsewhere (see text). 'Run as' indicates life stage that was electrophoresed; 'Code' is used in Appendix to abbreviate sites. All sites except 17 are in U.S.A.

Code	Site		Collection date	Host species	Latitude	Run as
	State	Locality (county)				
1	Alabama	US 82 & Autauga County 10 (Autauga)	9/9/89	<i>Crataegus flava</i>	32°32'	A
2	Georgia	US 80 & I-75 (Bibb)	13/9/89	<i>C. flava</i>	32°48'	A
3		Musella (Bibb)	16/10/89	<i>C. viridis</i>		A
4	Illinois	Lake of the Woods Park (Champaign)	14/9/89	<i>C. mollis</i>	40°5'	A
5	Kentucky	Reelfoot Lake (Fulton)	9/9/94	<i>C. mollis</i>	36°29'	L
6	Maryland	Beltville Natl. Agric. Res. Center (Prince Georges)	17/9/89	<i>C. mollis</i>	39°4'	A
7		US 40/48 at eastern continental divide (Garrett)	18/9/89	<i>C. pruniosa</i> †‡	39°35'	A
8	Pennsylvania	Meyersdale (Somerset)	18/9/89	<i>C. pruniosa</i> †‡	39°49'	A
9	S. Carolina	Clarks Hill Reservoir (McCormick)	8/9/85	<i>C. flabellata</i> †	33°42'	L
10		Clemson University Sandhills Res. Sta. (Kershaw)	30/8/89	<i>C. flava</i>	34°5'	A
11	Texas	Dinosaur Road (Gillespie)	8/10/89	<i>C. greggiana</i>	30°14'	A
12		Nacogdoches (Nacogdoches)	2/9/85	<i>C. brachyacantha</i>	31°32'	A
13		Urbana (San Jacinto)	29/10/89	<i>C. viridis</i>	31°32'	A
14		Waxahachie (Ellis)	6/10/89	<i>C. invisat</i> †	31°25'	A
15	Tennessee	Savage Gulf Natural Area (Grundy)	21/10/90	<i>C. flabellata</i> †‡	35°27'	A
16	Virginia	Mt. Rogers (Grayson)	27/9/91	<i>C. pruniosa</i> †‡	36°46'	A
17	Hidalgo (Mexico)	Zoquizoquipan	15/9/74	<i>C. mexicana</i> ‡	21°40'	A

A, adult; L, larva.

†Tentative identification; see text.

‡Sites at conspicuously higher altitude than remainder; see text.

were reared from infested fruit collected in the field and frozen at  $-70^{\circ}\text{C}$  for later analysis, as described by Berlocher & Enquist (1993), but at two sites larvae were frozen. Fruit was collected from as many individual plants as feasible at each site, and pooled there. Most sites were at altitudes of less than 300 m a.s.l. Exceptions are US 40/48 at the continental divide, MD (c. 848 m), Meyersdale, PA (c. 650 m), Mt. Rogers, VA (c. 920 m), Savage Gulf Natural Area, TN (c. 520 m), and Zoquizoquipan, Hidalgo, Mexico (1850 m). For insect and plant vouchering procedures see Berlocher & Enquist (1993). Although *Crataegus* is a difficult genus taxonomically (see discussion by Berlocher & Enquist, 1993), all but six host plants could be confidently identified to species (Table 1). Sites are mapped in Fig. 1; for exact site locations contact S. H. B.

### Electrophoresis

Standard single-condition horizontal starch gel electrophoresis of 17 loci expressed in both larvae and adults was performed for samples in data sets 1 and 2 as described previously (McPheron, 1990; Berlocher, 1995). [Berlocher (1995) states that only 16 loci were studied in *R. mendax*, but the totally invariant *Acon-1* was inadvertently omitted from discussion; it was studied on the same buffer with *Acon-2*, and in essentially identical numbers]. Electrophoresis for data sets 1 and 2 was carried out on adult flies with the exception of the two larval samples (Table 1). All allozyme loci in *R. pomonella* have been at least partially mapped (Feder *et al.*, 1989).

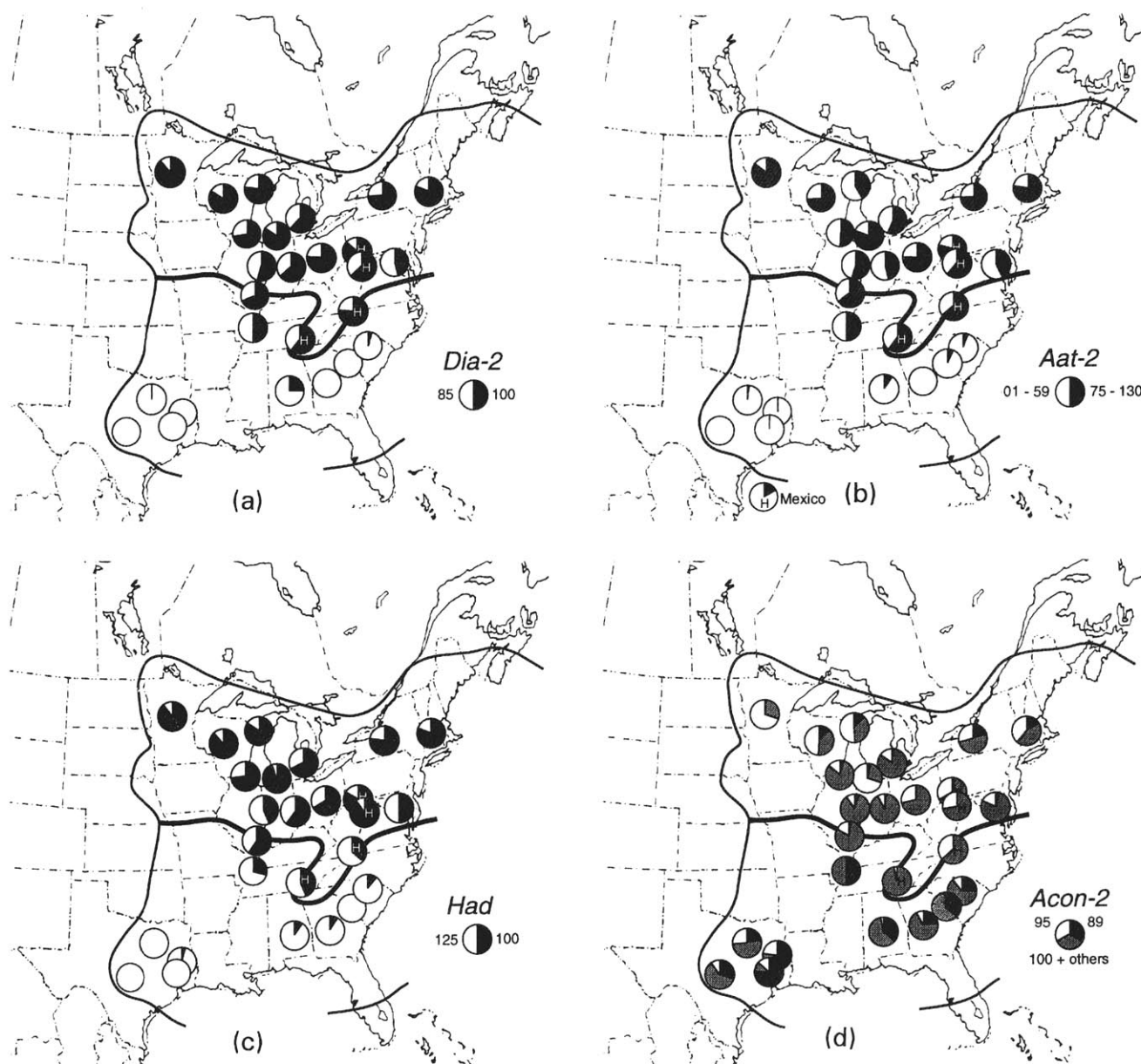
### Data analysis

As with the electrophoresis, data analysis was very similar to that in the previous paper on *R. mendax* (Berlocher, 1995). The reader is referred to that paper for details of some analyses. The BIOSYS-1 data file for data sets 1 and 2 may be obtained from S. H. B. via the following e-mail address: [stewart\\_berlocher@qms1.life.uiuc.edu](mailto:stewart_berlocher@qms1.life.uiuc.edu).

The BIOSYS-1 (Swofford & Selander, 1981) exact probability test option was used to determine whether populations in data sets 1 and 2 were in Hardy-Weinberg Equilibrium (HWE). To study geographical population differentiation in data sets 1 and 2 under a fixed population model (Weir, 1990), we used contingency analysis (using a Monte Carlo simulation of an exact  $R$  populations  $\times$   $C$  alleles test on allele counts; MONTE CARLO  $R \times C$  2.2, Macintosh program provided by W. Engels, Department of

Genetics, University of Wisconsin, Madison, WI; Fisher's exact test used on  $2 \times 2$  tables). To ensure that rejection of the null hypothesis was even possible, only tests in which the total count of the second most abundant allele at each locus was  $\geq 5$  were performed (this value was determined by numerical experimentation using a range of sample sizes typical of our data). See the discussion in Berlocher (1995) of the number of significant decimal places for  $P$ -values from Monte Carlo simulation. Geographical structure in the 22 samples of data sets 1 and 2 under a random population model was studied using  $F$ -statistics (Weir, 1990). As in the previous paper, for comparison we also calculated the original  $F$ -statistics of Wright (1978) using BIOSYS-1.

As in the *R. mendax* paper, the first and critical analysis of potential latitudinal clines was carried out by nonparametric regression of the frequency of each allele against latitude using Kendall's  $\tau$ , because neither allele frequencies nor latitude are necessarily normally distributed, even after transformation. Eighteen samples from data sets 1 and 2 were analysed for latitudinal effects. Four sites at high elevations in the Appalachian Mountains and Cumberland Plateau (Fig. 1, Table 1) were excluded after observing that they possessed allele frequencies characteristic of high latitudes (Results). For the remaining 18 samples, alleles were chosen for regression against latitude as follows. First, alleles for which the highest frequency among the samples was  $< 0.02$  were eliminated. Secondly, to eliminate dependencies in the frequency data, for each locus the allele with the highest mean frequency among the samples was eliminated. Alleles with significant nonparametric regressions (see discussion of multiple tests) were further subjected to conventional regression analysis for descriptive purposes, with data from sets 4 and 5 included; high altitude sites from data sets 1 and 2 were graphed but not included in the analyses. We initially used arcsine transformed frequencies and heterozygosities in the parametric regressions, but the results were so similar to those obtained with untransformed data that only the latter are shown here. Latitudinal analysis of HWE expected, average heterozygosity,  $\bar{H}$ , and average number of alleles,  $\bar{a}$ , was carried out in a similar fashion, except that average number of alleles was also analysed using data from only eight flies chosen haphazardly (the first eight individuals complete for all loci in the order electrophoresed) from each sample (eight being the smallest  $n$  for a locus among the 18 included samples), to eliminate the effect of sample size on the number of alleles



**Fig. 1** Geographical allele frequency variation at four loci of *Rhagoletis pomonella* in eastern North America. The thin line represents the range of *R. pomonella*, and the southern limit of the apple race is marked with a thick line. Mapped are data from 21 of the 22 uniformly electrophoresed populations of data sets 1 and 2 (the Musella, GA population is omitted as it is both geographically and genetically very close to the Lizella, GA population). Also mapped are the five populations of data sets 4 and 5 (see text), and for *Aat-2* the Mexican population comprising data set 3 (not mapped to site). H indicates a high elevation site; see text. (a) *Dia-2*; (b) *Aat-2* — see text for discussion of pooling of alleles; (c) *Had*; (d) *Acon-2*.

observed.

As in the *R. mendax* study (Berlocher, 1995), the analysis of linkage disequilibrium had two stages. First, contingency tests on two-locus genotype numbers were carried out for all pairwise combinations of loci (because linkage relationships are

incompletely known) for all samples in data sets 1 and 2. To avoid the laborious cross-tabulation analysis by spreadsheet program used by Berlocher (1995), a BASIC program (for modified BIOSYS-1 files, written by S. H. B.) that cross-tabulates genotype numbers and carries out a Monte Carlo simula-

tion (1000 replications) of an exact probability  $R \times C$  contingency test was used as an initial screen of the data. Comparisons with  $P < 0.10$  were then subjected to more thorough contingency analysis using MONTE CARLO  $R \times C$  2,2, with cases with  $n < 5$  for the second most common allele eliminated.

In general (see Results) only samples with probabilities below the Bonferoni corrected critical value in the contingency tests were carried to the second stage of measuring linkage disequilibrium,  $D$ .  $D$  was computed for data pooled to two alleles per locus (most common and all others; see Results) using the maximum likelihood method of Hill (1974). Because the *R. pomonella* populations were in HWE (Results), the relatively simple method of Hill (1974) could be used instead of that of Weir (1990).  $\chi^2_I$  was calculated from  $D$  as suggested by Hill (1974) for comparison with the contingency test results.  $D'$ , which scales  $D$  between  $-1$  and  $1$  (Hedrick *et al.*, 1978), was also calculated.

When multiple tests are carried out the significance level must be lowered appropriately. The sequential Bonferoni procedure (Holm, 1979; cited by Rice, 1989) was used for tests of HWE, population differentiation and  $D$ . [Re-analysis of the *R. mendax* data (Berlocher, 1995) using the sequential rather than the nonsequential procedure employed there does not change any conclusions.] See Results for experiment limits.

## Results

Allele frequencies for data sets 1 and 3, and average heterozygosities and allele numbers for data set 1, are shown in the Appendix. One low frequency null allele was observed at the dimeric locus *Pgi*.

$\bar{H}$  averaged across all 22 samples in data sets 1 and 2 is 0.198, and  $\bar{a}$  averaged across samples is 2.436. Although selection of loci for this paper was biased towards polymorphic loci,  $\bar{H}$  based on the 29 locus data set of Berlocher *et al.* (1993), in which loci were chosen simply on the basis of acceptable electrophoretic resolution, regardless of whether polymorphism occurred, was almost identical at two sites at or near Urbana, IL ( $\bar{H} = 0.221$ , 29 loci, at Urbana, IL (Berlocher *et al.*, 1993);  $\bar{H} = 0.229$ , 17 loci (including *Dia-1*), at the Lake of the Woods site of the present work; sites *c.* 19 km apart).

However, unlike the case in *R. mendax*,  $\bar{H}$  in *R. pomonella* showed a significant clinal pattern ( $\tau = 0.682$ ,  $P < 0.0001$ ) of greater heterozygosity at high latitudes (18 samples from data sets 1 and 2). The parametric regression of  $\bar{H}$  on latitude is shown in Fig. 2a.  $\bar{a}$  showed a similar latitudinal cline

[ $\tau = 0.489$ ,  $P = 0.0046$  with  $n = 8$  in the 18 samples, indicating that the generally smaller sample sizes in the south (see Appendix) do not solely account for the reduction in  $\bar{a}$ ]; Fig. 2b shows parametric regression with complete samples.

Tests of HWE indicate that *R. pomonella* populations are panmictic, as described previously (Feder & Bush, 1989 for data set 4 of this paper; Feder *et al.*, 1988, 1990a,b, 1993; McPheron, 1988a,b, 1990). After application of the Bonferoni procedure with

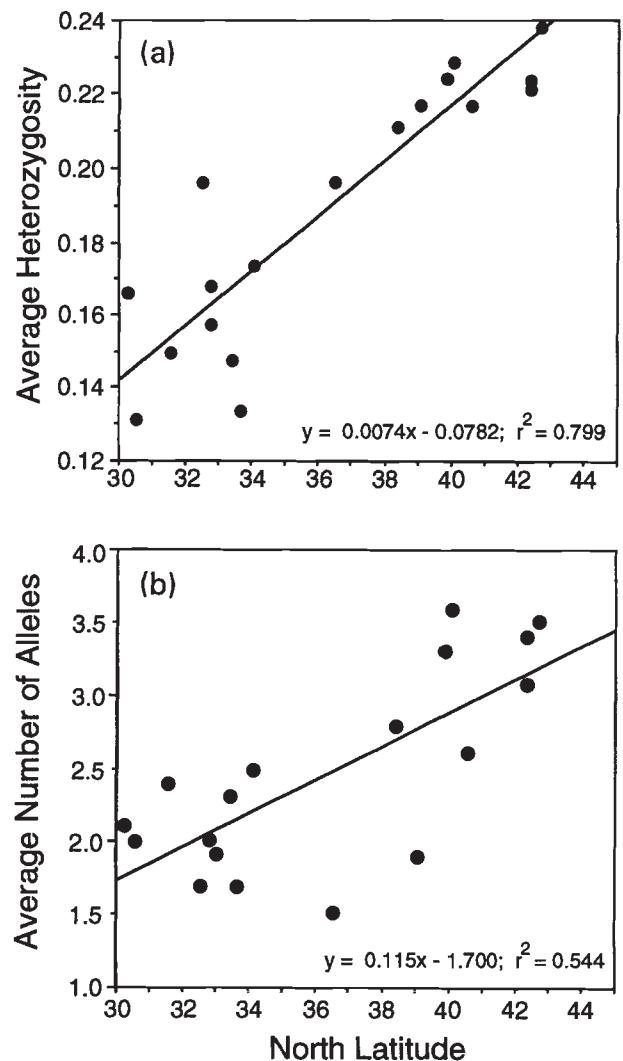


Fig. 2 Parametric regression of (a) average heterozygosity and (b) average number of alleles on latitude including 18 populations of *Rhagoletis pomonella* from the 22 populations of data sets 1 and 2 (four high elevation sites removed). Regression statistics included for comparative purposes; primary significance determined by nonparametric regression (see text). Average heterozygosity and number of alleles calculated for entire samples (see text).

an experiment wise 0.05 probability level, no tests gave significant results, regardless of whether the experiment was considered to consist of data sets 1 and 2 considered together (240 total tests), or whether each sample was treated as a separate experiment, as in Berlocher (1995). Mean  $f$  over all loci by the Cockerham and Weir method (Weir, 1990) is  $-0.003$ , not significantly different from zero by bootstrapping.

Mapping and visual inspection of allele frequencies reveals conspicuous patterns of differentiation at several loci, the most striking of which, for *Dia-2*, *Aat-2* (pooling of alleles discussed below), *Had* and *Acon-2*, are mapped in Fig. 1 (all data sets included). By far the most conspicuous pattern is a tendency for latitudinal clines. Elevation effects are evident as well; comparison of sites 6 with 7, and 10 with 15 (each pair being on a similar latitude) reveals that the higher altitude site of each pair has a higher allele frequency, but more extensive sampling is needed for adequate statistical analysis. At one locus, *Adh-1* (not mapped), the populations in central Texas (Gillespie Co., TX) and Mexico are quite differentiated from the remainder (see Appendix).

Large-scale geographical differentiation is overwhelmingly significant, as shown by both the fixed population contingency tests and the random population  $F$ -statistics (Table 2). Using contingency tests,

**Table 2**  $F_{ST}$  and probability values from contingency tests for 22 geographical samples representing the geographical range of *Rhagoletis pomonella* (17 loci, data sets 1 and 2; see text). See text for explanation of zero  $P$ -values

Locus	$F_{ST}$	$P$
<i>Aat-2</i>	0.117	0
<i>Acon-2</i>	0.132	0
<i>Adh-1</i>	0.105	0
<i>Ak-3</i>	0.003	0.11
<i>Dia-2</i>	0.389	0
<i>Fum</i>	0.007	0.02
<i>Had</i>	0.342	0
<i>Idh</i>	0.011	0
<i>Mpi</i>	0.035	0
<i>Pgi</i>	0.024	0
<i>Pgm</i>	0.014	0
<i>Tpi</i>	$-0.003$	NT
<i>Acon-1</i>	$-0.003$	NT
<i>Ald</i>	$-0.002$	NT
<i>Dia-1</i>	M	M
<i>Mdh-1</i>	0.006	0.03
<i>Mdh-2</i>	0.042	0.0001

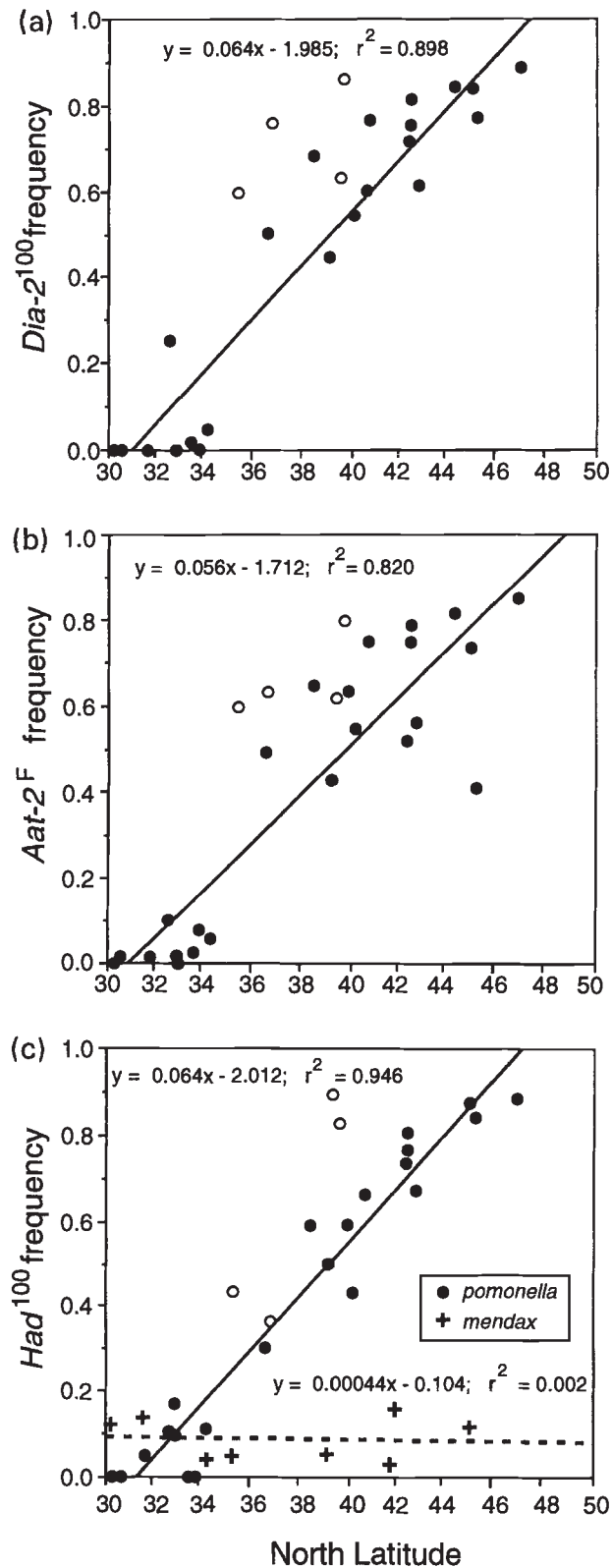
NT, not testable (see Methods); M, monomorphic.

10 of the 13 testable loci show significant differentiation at the 0.01 level (after Bonferoni correction), and the level of significance is probably much higher at nine of the loci ( $P = 0$  using at least  $10^7$  replications of the Monte Carlo procedure). Single-locus  $F_{ST}$  values are also quite large, with the largest being 0.389 for *Dia-2*.  $F_{ST}$  over all 16 variable loci is 0.148 (bootstrap 95 per cent confidence limits are 0.046 and 0.277).  $F_{ST}$  calculated using Wright's (1978) method produces a value of  $F_{ST}$  identical to 3 decimal places. Because of strong disequilibrium between *Aat-2* and *Dia-2* (see below), we also recalculated the Cockerham and Weir  $F_{ST}$  with each of these two loci removed from the data. The effect was relatively minor ( $F_{ST} = 0.156$  with *Aat-2* removed, and 0.113 with *Dia-2* removed), so in the Discussion we use the value based on all loci.

Analysis of the apparent clines revealed that five out of 35 testable alleles were significantly associated with latitude at the 0.05 level: *Aat-2*<sup>21</sup> ( $\tau = -0.544$ ,  $P = 0.0023$ ); *Aat-2*<sup>75</sup> ( $\tau = 0.706$ ,  $P < 0.0001$ ); *Aat-2*<sup>100</sup> ( $\tau = 0.618$ ,  $P = 0.0003$ ); *Dia-2*<sup>100</sup> ( $\tau = 0.691$ ,  $P < 0.0001$ ); *Had*<sup>100</sup> ( $\tau = 0.765$ ,  $P < 0.0001$ ). *Aat-2*<sup>75</sup>, *Dia-2*<sup>100</sup> and *Had*<sup>100</sup> are also significant at the  $P = 0.01$  level. The frequency clines for *Aat-2*, *Dia-2* and *Had* are plotted in Fig. 3, with data sets 4 and 5 included. The four high altitude sites (open circles), not included in the statistical analysis, are all above the regression lines for *Aat-2*, *Dia-2* and *Had*, supporting the altitude effect suggested by Fig. 1.

Inspection of the data revealed a very strong, widespread and consistent pattern of linkage disequilibrium involving *Aat-2* and *Dia-2*. Table 3 shows the complete analysis for the largest sample, Lake of the Woods, IL (130 genotypes complete for both loci).  $D$  was computed after partitioning *Aat-2* alleles into a test allele and a pool of all others. Alleles *Aat-2*<sup>18</sup> to *Aat-2*<sup>59</sup> all show negative disequilibrium with *Dia-2*<sup>100</sup>, whereas *Aat-2*<sup>75</sup> to *Aat-2*<sup>123</sup> all show positive disequilibrium with *Dia-2*<sup>100</sup>, with highly significant  $P$ -values (from both the contingency tests and the  $\chi^2$  approximation from  $D$ ) for four of the alleles. Because of the consistency of this pattern in almost all populations with sufficient polymorphism to analyse (data not shown), *Aat-2* alleles were pooled into two groups referred to henceforth as the 'slow' (S) allele (*Aat-2*<sup>01</sup> to *Aat-2*<sup>59</sup>) and the 'fast' (F) allele (*Aat-2*<sup>75</sup> to *Aat-2*<sup>123</sup>) for geographical analysis of  $D$  and for mapping and regression of geographical patterns of *Aat-2* (Figs 1b and 3b).

In the search for  $D$  among all locus  $\times$  locus comparisons, little besides the *Aat-2*–*Dia-2* association was discovered. If the hypothesis to be tested is



that no disequilibrium exists anywhere in *R. pomonella* data sets 1 and 2, then the experimentwise Bonferroni  $P$ -value for a nominal 0.05 test value is 0.00008 (658 tests); only 10 *Aat-2-Dia-2* comparisons (Table 4) and the *Dia-2-Idh* and *Aat-2-Idh* comparisons (both uncorrected  $P = 0.00002$ ) at Reelfoot Lake, KY are significant. Considering each sample as a separate experiment substantially increases the corrected critical  $P$ -values (for a typical population with about nine testable loci, 0.05 critical value, corrected  $P \approx 0.001$ ), but fails to add additional cases of significant disequilibrium. Given the strength of the *Aat-2-Dia-2* relationship in some populations, we included in Table 4 all populations, with significant  $D$  or not, in order to discern any geographical trends.

The *Fum*<sup>158</sup> allele, previously reported to be restricted to *R. mendax* (Berlocher, 1995 and references therein), was observed in nine different heterozygotes from four populations, out of a total of 1130 *R. pomonella* examined at this locus. This raises the possibility of hybridization between *R. pomonella* and *R. mendax*, to be discussed in detail elsewhere.

## Discussion

The population structure of *R. pomonella* is clearly very different from that of *R. mendax* (Berlocher, 1995), with the most striking differences concerning large-scale geographical differentiation. *Rhagoletis mendax* is almost uniform across its range, with little differentiation ( $F_{ST} = 0.015$ ) and no clinal structure (Berlocher, 1995), whereas *R. pomonella* shows very significant differentiation ( $F_{ST} = 0.148$ , significantly greater than in *R. mendax* as the 95 per cent confidence limits do not overlap) and strong clinal structure. The differences with respect to clines are especially striking at the *Had* locus (Fig. 3c), where the two species share the two common alleles. No

Fig. 3 Parametric regression of *Rhagoletis pomonella* allele frequencies on latitude in eastern North America. Plotted are data from 27 populations: 22 populations from the uniformly electrophoresed data sets 1 and 2 (four high elevation sites, not used in any statistical analysis, are open circles), plus the five additional samples of data sets 4 and 5. Regression statistics included for comparative purposes; primary significance determined by nonparametric regression of the 18 low altitude populations from data sets 1 and 2 (see text). (a) *Dia-2*<sup>100</sup>; (b) *Aat-2*<sup>F</sup> — see text for discussion of pooling of alleles; (c) *Had*<sup>100</sup> (including data for *R. mendax* — see text).

**Table 3** Linkage disequilibrium between *Dia-2*<sup>100</sup> and the eight alleles of *Aat-2* in the Lake of the Woods, IL sample of *Rhagoletis pomonella*. To test for disequilibrium, *Aat-2* alleles were pooled into one single allele ('allele' column) and all others. *P* is calculated from contingency tests,  $\chi^2$  from *D* (*P*-values for 1 d.f.); neither corrected by Bonferroni procedure

<i>Aat-2</i> allele	<i>P</i>	<i>D</i>	<i>D'</i>	$\chi^2$
18	0.30	-0.006	-1.000	1.826
21	0	-0.083	-0.882	25.452****
25	0.18	-0.002	-1.000	0.604
50	0	-0.118	-0.832	37.494****
59	0.15	-0.006	-1.000	1.826
75	0.0002	0.068	0.703	16.605****
100	0	0.138	0.831	47.150****
123	0.09	0.012	0.368	1.296

No. of complete genotypes = 130.

\*\*\*\**P* ≤ 0.001.

**Table 4** Linkage disequilibrium between *Aat-2*<sup>F</sup> and *Dia-2*<sup>100</sup> in samples of *Rhagoletis pomonella* in which both loci were polymorphic (i.e. more than one allele at both loci; see text)

Sample	<i>n</i>	<i>P</i>	<i>D</i>	<i>D'</i>	$\chi^2$
US 82 & Autauga County 10, AL	10	0.22	0.075	1.000	3.333
Wayne County, IL	54	0	0.154	0.755	26.046****
Lake of the Woods Park, IL	130	0	0.225	0.919	106.394****
Hancock County, IN	70	0	0.180	0.772	41.758****
Reelfoot Lake, KY	15	0.00002	0.250	1.000	15.000****
Amherst, MA	99	0	0.127	0.858	64.086****
Beltsville Natl. Agric. Res. Center, MD	20	0.00002	0.207	0.884	14.104****
US 40/48 at eastern continental divide, MD	30	0	0.209	0.924	23.868****
East Lansing, MI	68	0	0.199	0.865	47.828****
Ithaca, NY	99	0	0.142	0.792	55.777****
Hardin County, OH	53	0	0.149	0.836	35.200****
Meyersdale, PA	15	0.0043	0.107	1.000	9.231**
Clemson University Sandhills Res. Sta., SC	23	0.0119†	0.041	1.000	14.985****
Savage Gulf Natural Area, TN	15	0.0029	0.166	0.692	7.189**
Waxahachie, TX	34	0.94	-0.000	-0.030	0.015
Mount Rogers, VA	15	0.0092	0.148	1.000	7.885**

*n*, no. of genotypes complete for both loci.

\*\**P* < 0.01, \*\*\*\**P* ≤ 0.001. *P*-values uncorrected.

†*P* from Fisher's exact 2 × 2 test.

trivial reason such as greatly reduced variation in *R. mendax* can explain these differences, as levels of allozyme variation are similar in the two species.  $\bar{H}$  averaged over all populations is 0.145 in *R. mendax* and 0.198 in *R. pomonella*, and  $\bar{a}$  averaged over all populations is 2.176 in *R. mendax* and 2.436 in *R. pomonella* [*R. mendax* values recalculated from Berlocher (1995) based on 17 instead of 16 loci], and in fact the two species share some or all alleles at all loci (Berlocher *et al.*, 1993).

Our data provide additional support for the hypothesis that the life history phenology of *R. mendax* is less flexible than that of *R. pomonella*. Dates of fruit collections yielding flies provide an approximation of when adult flies and larvae are active during the year, although only a very rough guide, as different fruit collections contain life stages ranging from newly laid eggs to mature third instar larvae. Nonetheless, the observation that the variance in collection dates is significantly greater

( $P = 0.0042$ ) in *R. pomonella* (data from Table 1) than *R. mendax* (data from Payne & Berlocher, 1995b; variance comparison using Levene test on the middle 80 per cent of individual coefficients of variation as suggested by Van Valen, 1978) strongly suggests more constraint of emergence time in *R. mendax* than *R. pomonella*.

Our results in part confirm and amplify earlier allozyme studies of *R. pomonella* based on less widespread geographical sampling. The large  $F_{ST}$  was presaged by McPheron (1987), who obtained  $F_{ST} = 0.220$  (including only two southern populations, Nacogdoches and Clarks Hill Reservoir). In a series of papers Feder and colleagues (Feder & Bush, 1989; Feder *et al.*, 1990a, 1993) have described strong allozyme clines in *R. pomonella* populations (of both hawthorn and apple races) in the northern part (roughly, from central Illinois north) of the range of the species. Our results demonstrate that these clines continue linearly all the way to the southern extent of the range of *R. pomonella* in the U.S.A. (Fig. 3).

Feder *et al.* (1990a) have previously described strong linkage disequilibrium between *Aat-2* and *Dia-2*, and noted that this is not surprising in light of their close linkage. Our data suggest that the disequilibrium may be of roughly equal intensity from north to south, as two of the three nonsignificant  $D'$  values in the south are nonetheless maximal and positive (Table 4), despite the gradual decrease of both  $Dia-2^{100}$  and  $Aat-2^F$  towards zero in the south. Much larger sample sizes for southern populations will be needed to confirm this finding.

However, in several regards our results are either at odds with prior work or further complicate our knowledge of the already complex biology of *R. pomonella*. In three respects the pattern of population differentiation we have observed departs from the simple clinal pattern described previously (Feder & Bush, 1989; Feder *et al.*, 1990a). First, the pattern at *Acon-2* is complex, with  $Acon-2^{95}$  having the highest frequency in the north,  $Acon-2^{89}$  having the highest frequency in the south, and  $Acon-2^{100}$  having the highest frequency in the middle of the range (Fig. 1d). Secondly, *Adh-1* shows potentially nonclinal variation, with the  $Adh-1^{35}$  allele absent in most northern samples, reaching a maximum frequency of 0.387 in the Gillespie Co., TX population. If only the  $Adh-1^{-100}$  allele is considered, the data suggest yet another cline, as this allele occurs at very high frequency in northern populations: at 0.548 at Gillespie County, TX, and at only 0.254 in central Mexico. However, we note that the common allele in Mexico, with a frequency of 0.729, is  $Adh-1^{01}$ ,

which occurs sporadically throughout the rest of the range (see Appendix). If variation at *Adh-1* is indeed clinal, the pattern will be more like that of *Acon-2* than the simpler pattern of the other loci. Thirdly, we did not find the  $Mpi^{70}$  allele (=  $Mpi^{37}$  of Feder *et al.*, 1990a) to be clinally distributed. However, our data indicate that this allele becomes common only in the northern part of the range of *R. pomonella*, where Feder *et al.*'s (1990a) sampling was denser than ours, so our data do not contradict the possibility of a cline at *Mpi*.

Our results also do not entirely agree with Feder *et al.* (1990a) with respect to linkage disequilibrium, but the differences may have little significance. In the case of the *Mpi-Acon-2* disequilibrium observed by Feder *et al.* (1990a) but not us, only one sample is unambiguously significant after Bonferroni correction, and in the case of the three-locus disequilibrium involving the *Aat-2-Dia-2-Idh* linkage group, observed by us but not by Feder *et al.* (1990a), the Reelfoot, KY sample was apparently from a small population (only 16 larvae were found); drift in a tiny population is thus a possible explanation.

Under the climate-linkage hypothesis, adaptation to local climatic regimes indirectly affects allozyme frequencies owing to linkage disequilibrium between an unknown number of diapause-controlling genes and linked groups of allozyme loci; three groups of loci [*Aat-2-Dia-2*, *Me-Acon-2-Mpi* and *Had* (we examined all except *Me*)] do appear to be associated with diapause characteristics (Feder *et al.*, 1993). The climate-linkage hypothesis thus explains the allozyme clines as the by-product of adaptation to a longer pre-eclosion period in the southern part of the range. This hypothesis also explains the involvement of the same loci in host race formation, as apple and most hawthorn species have different fruiting phenologies (Feder *et al.*, 1988, 1990a,b; McPheron *et al.*, 1988b), and may potentially explain differentiation related to individual conspecific hawthorn trees (McPheron *et al.*, 1988a). Our results are consistent with the climate-linkage hypothesis.

Our data also suggest at least two other factors that need to be considered before the climate-linkage hypothesis can be completely accepted. First, the loss of some alleles in the south, as reflected in the reduction in  $\bar{H}$  and  $\bar{a}$  (Fig. 2a,b) may be at least partially caused by small population size, i.e. by drift. Based on our field experience (McPheron, 1990; Berlocher & Enquist, 1993; Berlocher & McPheron, unpublished observations) we have the strong impression that *R. pomonella* populations in the southern part of the range are on average smaller

and more isolated than northern populations, although quantitative data sufficient for a statistical test are lacking. The second factor is that the data do not completely rule out a role for secondary contact as an explanation for the clines, although this explanation was not favoured by Feder and colleagues (Feder & Bush, 1989; Feder *et al.*, 1990a). Selection and secondary contact are not, of course, mutually exclusive.

Comparison of the range of the apple race of *R. pomonella* with the electrophoretic data (Fig. 1) shows that the apple race is found only in areas where the frequencies of 'northern' alleles of the clinally varying loci *Aat-2*, *Dia-2* and *Had* are well above 0.5 in the hawthorn race, even though apples are grown in the south. This indicates two things. First, as the frequencies of these alleles in the apple race are also high (Feder *et al.*, 1993 and references therein), the apple race clearly originated in the northern part of the range, as also indicated by the historical data (Bush, 1992 and references therein). Secondly, if the climate-linkage hypothesis is correct, then allele frequencies in a given area may be of predictive value in measuring the likelihood of future host shifts.

The partial data from the single site in Mexico indicate that the eastern North American and Mexican populations have diverged substantially. In addition to the differentiation at *Adh-1* already discussed, frequencies are quite different at *Aat-2*, with *Aat-2*<sup>21</sup> absent, and a very high frequency of *Aat-2*<sup>50</sup>; the Mexican population resembles neither northern nor southern populations very closely at this locus. The allozyme differences together with the wing band difference discussed by Bush (1966) support Bush's conclusion that the Mexican population is not a recent introduction.

Given that a detailed understanding of the population structure of *R. pomonella* is essential for a complete understanding of both the formation of host races in historical time and speciation of these flies in evolutionary time, much remains to be done. The elevational effect we observed needs to be studied in detail. Phylogenetic analysis of DNA sequences of the allozymes and other genes is needed to determine whether secondary contact has played a role in the origin of the clines. Another gap in our knowledge concerns the population structure and diapause characteristics of other species of the *R. pomonella* group. We have focused primarily on a comparison of *R. pomonella* and *R. mendax* because both have been studied in terms of diapause characteristics, behaviour, host plant range and population structure. Comparison of the population structure of

*R. pomonella* with that of a third species, the 'sparkleberry fly' (Payne & Berlocher, 1995a), suggests that the clines and large  $F_{ST}$  of *R. pomonella* may be the exception rather than the rule (sparkleberry fly  $F_{ST} = 0.018$ , no clinal structure).

Perhaps the most critical need is for study of the effect of host plant species on population structure. Although we sampled many species of *Crataegus*, we have yet to analyse fly samples reared from different species of *Crataegus* growing side-by-side, analogous to the apple-hawthorn comparisons critical to establishing the existence of the apple race (Feder *et al.*, 1988; McPheron *et al.*, 1988a). Yet such work is necessary, as the fruiting times of some hawthorn species differ as much as or more than do apples and *C. mollis*, the common northern hawthorn of apple race studies (Berlocher & Enquist, 1993). For example, the difference in collection dates between the *C. brachyacantha* collection at Nacogdoches, TX and the *C. viridis* collection at Urbana, TX (c. 132 km apart) is 58 days — substantially greater than the typical difference in collection dates between apple and hawthorn flies at typical northern sites of 20–30 days (McPheron, unpublished observation). These differences in hawthorn fruiting times do not necessarily mean that host races on each hawthorn species are expected — other factors such as oviposition behaviour are involved in differentiating the apple and hawthorn races (Prokopy *et al.*, 1987), and the mere existence of early and late ripening hawthorns does not necessarily produce temporally distinct fly populations — but if the climate-linkage hypothesis is true, then flies reared from early and late ripening hawthorn species should possess different allele frequency arrays at loci such as *Aat-2*, *Acon-2*, *Dia-2* and *Had*.

### Acknowledgements

This project would never have been possible without the assistance of S. Lyons, J. A. Payne, R. Payne, A. Amis, W. S. Sheppard, C. Anelli, Th. Anderson, M. G. Anderson, S. H. Berlocher, N. N. Berlocher, G. P. Berlocher and G. L. Bush (Cmdr, Ret.), and in particular J. L. Feder for the use of unpublished data. NSF grant 89-06217 provided funding.

### References

- BERLOCHER, S. H. 1976. *Genetics of Speciation in Rhagoletis (Diptera: Tephritidae)*. Ph.D. Thesis, University of Texas at Austin.
- BERLOCHER, S. H. 1995. Population structure of the blueberry maggot, *Rhagoletis mendax*. *Heredity*, **74**, 542–555.

- BERLOCHER, S. H. AND ENQUIST, M. 1993. Distribution and host plants of the apple maggot fly, *Rhagoletis pomonella* (Diptera: Tephritidae) in Texas. *J. Kans. Entomol. Soc.*, **66**, 145–166.
- BERLOCHER, S. H., McPHERON, B. A., FEDER, J. L. AND BUSH, G. L. 1993. Genetic differentiation at allozyme loci in the *Rhagoletis pomonella* (Diptera: Tephritidae) species complex. *Ann. Entomol. Soc. Am.*, **86**, 716–727.
- BOLLER, E. F. AND PROKOPY, R. J. 1975. Bionomics and management of *Rhagoletis*. *Ann. Rev. Ent.*, **21**, 223–246.
- BUSH, G. L. 1966. The taxonomy, cytology, and evolution of the genus *Rhagoletis* in North America (Diptera: Tephritidae). *Bull. Mus. Comp. Zool.*, **134**, 431–562.
- BUSH, G. L. 1992. Host race formation and sympatric speciation in *Rhagoletis* fruit flies (Diptera: Tephritidae). *Psyche*, **99**, 335–357.
- BUSH, G. L. 1994. Sympatric speciation in animals: new wine in old bottles. *Trends Ecol. Evol.*, **9**, 285–288.
- FEDER, J. L. AND BUSH, G. L. 1989. Gene frequency clines in host races of *Rhagoletis pomonella* (Diptera: Tephritidae) in the midwestern United States. *Heredity*, **63**, 245–266.
- FEDER, J. L., CHILCOTE, C. A. AND BUSH, G. L. 1988. Genetic differentiation between sympatric host races of *Rhagoletis pomonella*. *Nature*, **336**, 61–64.
- FEDER, J. L., CHILCOTE, C. A. AND BUSH, G. L. 1989. Inheritance and linkage relationships of allozymes in the apple maggot fly. *J. Hered.*, **80**, 277–283.
- FEDER, J. L., CHILCOTE, C. A. AND BUSH, G. L. 1990a. The geographic pattern of genetic differentiation between host-associated populations of *Rhagoletis pomonella* (Diptera: Tephritidae) in the eastern United States. *Evolution*, **44**, 570–594.
- FEDER, J. L., CHILCOTE, C. A. AND BUSH, G. L. 1990b. Regional, local, and microgeographic allele frequency variation between apple and hawthorn populations of *Rhagoletis pomonella* in western Michigan. *Evolution*, **44**, 596–608.
- FEDER, J. L., HUNT, T. A. AND BUSH, G. L. 1993. The effect of climate, host plant phenology, and host fidelity on the genetics of apple and hawthorn-infesting races of *Rhagoletis pomonella*. *Entomologia exp. appl.*, **69**, 117–135.
- FEDER, J. L., OPP, S. B., WLAZLO, B., REYNOLDS, K., GO, W., AND SPIZAK, S. 1994. Host fidelity as an effective premating barrier between sympatric races of the apple maggot fly. *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 7990–7994.
- FUTUYMA, D. J. AND MAYER, G. C. 1980. Non-allopatric speciation in animals. *Syst. Zool.*, **29**, 254–271.
- HEDRICK, P., JAIN, S. AND HOLDEN, L. 1978. Multilocus systems in evolution. *Evol. Biol.*, **11**, 104–154.
- HILL, W. G. 1974. Estimation of linkage disequilibrium in randomly mating populations. *Heredity*, **33**, 229–239.
- MAYR, E. 1963. *Animal Species and Evolution*. Belknap Press, Cambridge, MA.
- McPHERON, B. A. 1987. *The Population Genetics of the Colonization of the Western United States by the Apple Maggot, Rhagoletis pomonella (Walsh) (Diptera: Tephritidae)*. Ph.D. thesis, University of Illinois, Urbana.
- McPHERON, B. A. 1990. Genetic structure of apple maggot fly (Diptera: Tephritidae) populations. *Ann. Entomol. Soc. Am.*, **83**, 568–577.
- McPHERON, B. A., SMITH, D. C. AND BERLOCHER, S. H. 1988a. Microgeographic genetic variation in the apple maggot, *Rhagoletis pomonella*. *Genetics*, **119**, 445–451.
- McPHERON, B. A., SMITH, D. C. AND BERLOCHER, S. H. 1988b. Genetic differences between host races of the apple maggot fly. *Nature*, **336**, 64–66.
- PAYNE, J. A., AND BERLOCHER, S. H. 1995a. Phenological and electrophoretic evidence for a new blueberry-infesting species in the *Rhagoletis pomonella* (Diptera: Tephritidae) sibling species complex. *Entomologia exp. appl.*, **75**, 183–187.
- PAYNE, J. A. AND BERLOCHER, S. H. 1995b. Distribution and host plants of the blueberry maggot, *Rhagoletis mendax* (Diptera: Tephritidae) in southeastern North America. *J. Kans. Entomol. Soc.*, **68**, 133–142.
- PROKOPY, R. J. AND BERLOCHER, S. H. 1980. Establishment of *Rhagoletis pomonella* (Diptera: Tephritidae) on rose hips in southern New England. *Can. Ent.*, **51**, 1319–1320.
- PROKOPY, R. J., DIEHL, S. R. AND COOLEY, S. S. 1987. Behavioral evidence for host races in *Rhagoletis pomonella* flies. *Oecologia*, **76**, 138–147.
- RICE, W. R. 1989. Analyzing tables of statistical tests. *Evolution*, **43**, 223–225.
- SMITH, D. C. 1988. Heritable divergence of *Rhagoletis pomonella* host races by seasonal asynchrony. *Nature*, **336**, 66–68.
- SWOFFORD, D. L. AND SELANDER, R. B. 1981. BIOSYS-1: A FORTRAN program for the comprehensive analysis of electrophoretic data in population genetics and systematics. *J. Hered.*, **72**, 281–283.
- VAN VALEN, L. 1978. The statistics of variation. *Evol. Theory*, **4**, 33–43.
- WEIR, B. S. 1990. *Genetic Data Analysis*. Sinauer Associates, Sunderland, MA.
- WRIGHT, S. 1978. *Evolution and the Genetics of Populations*, vol. 4, *Variability in and among Natural Populations*. University of Chicago Press, Chicago.

Appendix Allele frequencies in *Rhagoletis pomonella*

Locus	Population code								
	1	2	3	4	5	6	7	8	9
<i>Aat-2</i>									
(n)	10	35	30	131	15	20	30	15	20
01	0.000	0.000	0.050	0.000	0.000	0.000	0.000	0.000	0.000
18	0.000	0.000	0.000	0.011	0.000	0.000	0.000	0.000	0.000
21	0.200	0.343	0.500	0.172	0.067	0.225	0.167	0.067	0.275
25	0.000	0.000	0.000	0.004	0.000	0.000	0.000	0.000	0.000
50	0.700	0.629	0.383	0.263	0.433	0.325	0.217	0.133	0.625
59	0.000	0.029	0.050	0.011	0.000	0.025	0.000	0.000	0.025
75	0.000	0.000	0.017	0.176	0.000	0.200	0.183	0.300	0.050
100	0.100	0.000	0.000	0.305	0.267	0.225	0.400	0.400	0.025
123	0.000	0.000	0.000	0.057	0.233	0.000	0.033	0.100	0.000
<i>Acon-2</i>									
(n)	10	35	30	137	15	17	30	15	20
75	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
80	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
89	0.000	0.243	0.367	0.062	0.500	0.059	0.017	0.033	0.375
95	0.200	0.086	0.033	0.095	0.000	0.176	0.283	0.333	0.000
100	0.650	0.414	0.533	0.675	0.500	0.500	0.483	0.433	0.575
106	0.150	0.257	0.067	0.146	0.000	0.265	0.200	0.200	0.050
114	0.000	0.000	0.000	0.022	0.000	0.000	0.017	0.000	0.000
<i>Adh-1</i>									
(n)	10	36	30	143	15	20	30	15	20
83	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
45	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
40	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
35	0.000	0.000	0.000	0.000	0.000	0.000	0.100	0.033	0.000
01	0.000	0.000	0.000	0.003	0.000	0.000	0.017	0.000	0.000
-38	0.350	0.000	0.000	0.056	0.000	0.000	0.050	0.000	0.000
-46	0.000	0.000	0.000	0.003	0.000	0.000	0.000	0.000	0.000
-100	0.650	1.000	1.000	0.937	1.000	0.975	0.817	0.767	1.000
-140	0.000	0.000	0.000	0.000	0.000	0.025	0.017	0.200	0.000
<i>Ak-3</i>									
(n)	10	36	30	144	15	20	30	15	20
68	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
83	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
87	0.250	0.028	0.150	0.132	0.000	0.125	0.117	0.133	0.075
100	0.750	0.958	0.850	0.861	1.000	0.875	0.867	0.867	0.925
111	0.000	0.014	0.000	0.007	0.000	0.000	0.017	0.000	0.000
<i>Dia-2</i>									
(n)	10	36	30	143	15	20	30	15	20
70	0.000	0.028	0.000	0.000	0.000	0.000	0.000	0.000	0.000
85	0.750	0.972	1.000	0.455	0.500	0.550	0.367	0.133	1.000
100	0.250	0.000	0.000	0.545	0.500	0.450	0.633	0.867	0.000
<i>Fum</i>									
(n)	10	36	30	139	15	20	30	15	20
100	1.000	0.972	1.000	0.996	1.000	1.000	1.000	1.000	1.000
158	0.000	0.028	0.000	0.004	0.000	0.000	0.000	0.000	0.000

## Appendix Continued

Locus	Population code								
	1	2	3	4	5	6	7	8	9
<i>Had</i>									
(n)	10	36	30	140	15	20	30	15	20
72	0.000	0.000	0.000	0.014	0.000	0.000	0.000	0.000	0.000
98	0.000	0.000	0.000	0.004	0.000	0.000	0.000	0.000	0.000
100	0.100	0.097	0.167	0.432	0.300	0.500	0.900	0.833	0.000
125	0.900	0.903	0.833	0.550	0.700	0.500	0.100	0.167	1.000
<i>Idh</i>									
(n)	10	36	30	144	15	20	30	15	20
82	0.000	0.000	0.000	0.003	0.000	0.000	0.000	0.000	0.000
87	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
93	0.000	0.000	0.000	0.014	0.000	0.000	0.017	0.000	0.000
100	0.800	0.639	0.583	0.583	0.500	0.775	0.633	0.567	0.600
113	0.000	0.000	0.017	0.003	0.000	0.000	0.000	0.000	0.000
118	0.200	0.361	0.350	0.385	0.500	0.225	0.333	0.433	0.400
128	0.000	0.000	0.050	0.010	0.000	0.000	0.017	0.000	0.000
135	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
<i>Mpi</i>									
(n)	10	36	30	144	15	20	30	15	20
63	0.050	0.014	0.000	0.024	0.000	0.000	0.000	0.067	0.050
70	0.000	0.000	0.083	0.021	0.000	0.000	0.083	0.033	0.025
85	0.000	0.000	0.017	0.028	0.000	0.125	0.017	0.067	0.100
100	0.950	0.986	0.900	0.927	0.667	0.875	0.883	0.800	0.825
112	0.000	0.000	0.000	0.000	0.333	0.000	0.017	0.033	0.000
<i>Pgi</i>									
(n)	10	36	30	144	15	20	30	15	20
43	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
53	0.000	0.000	0.000	0.003	0.000	0.000	0.000	0.000	0.000
n76	0.000	0.000	0.000	0.000	0.000	0.000	0.017	0.000	0.000
80	0.000	0.000	0.000	0.007	0.000	0.000	0.017	0.000	0.000
100	1.000	0.819	0.900	0.865	1.000	0.925	0.917	0.900	1.000
130	0.000	0.042	0.000	0.024	0.000	0.000	0.033	0.000	0.000
145	0.000	0.139	0.100	0.101	0.000	0.075	0.017	0.100	0.000
<i>Pgm</i>									
(n)	10	36	30	144	15	20	30	15	20
80	0.000	0.000	0.000	0.003	0.000	0.000	0.000	0.000	0.000
92	0.000	0.153	0.000	0.076	0.000	0.075	0.083	0.067	0.000
95	0.200	0.000	0.000	0.010	0.000	0.000	0.000	0.033	0.000
100	0.750	0.847	0.967	0.854	0.900	0.875	0.817	0.833	0.875
104	0.000	0.000	0.000	0.000	0.000	0.000	0.017	0.000	0.000
111	0.000	0.000	0.033	0.038	0.000	0.050	0.067	0.033	0.125
118	0.050	0.000	0.000	0.017	0.100	0.000	0.017	0.033	0.000
125	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
<i>Tpi</i>									
(n)	10	36	30	120	15	8	30	15	20
68	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
85	0.000	0.000	0.000	0.004	0.000	0.000	0.000	0.000	0.000
90	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
100	1.000	1.000	1.000	0.992	1.000	1.000	1.000	1.000	1.000
120	0.000	0.000	0.000	0.004	0.000	0.000	0.000	0.000	0.000

## Appendix Continued

Locus	Population code								
	1	2	3	4	5	6	7	8	9
<i>Acon-1</i>									
(n)	10	36	30	144	15	20	30	15	20
-100	1.000	1.000	1.000	0.990	1.000	1.000	1.000	1.000	1.000
-120	0.000	0.000	0.000	0.003	0.000	0.000	0.000	0.000	0.000
-300	0.000	0.000	0.000	0.007	0.000	0.000	0.000	0.000	0.000
<i>Ald</i>									
(n)	10	36	30	144	15	20	30	15	19
100	1.000	1.000	0.933	1.000	1.000	1.000	1.000	1.000	1.000
162	0.000	0.000	0.067	0.000	0.000	0.000	0.000	0.000	0.000
<i>Mdh-1</i>									
(n)	10	36	30	142	15	20	30	15	20
-73	0.000	0.000	0.000	0.011	0.000	0.000	0.000	0.000	0.000
-100	1.000	1.000	1.000	0.989	1.000	1.000	0.983	1.000	1.000
-108	0.000	0.000	0.000	0.000	0.000	0.000	0.017	0.000	0.000
<i>Mdh-2</i>									
(n)	10	36	30	142	15	19	30	15	20
43	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
100	1.000	0.986	1.000	0.993	1.000	1.000	1.000	1.000	1.000
147	0.000	0.014	0.000	0.007	0.000	0.000	0.000	0.000	0.000
$\bar{H}$	0.196	0.157	0.167	0.229	0.196	0.217	0.227	0.232	0.133
$\bar{a}$	1.7	1.9	2.0	3.6	1.5	1.9	2.8	2.3	1.7
Locus	Population code								
	10	11	12	13	14	15	16	17	
<i>Aat-2</i>									
(n)	23	28	60	34	34	15	15	51	
01	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
18	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
21	0.413	0.482	0.550	0.397	0.324	0.100	0.100	0.000	
25	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
50	0.522	0.482	0.333	0.515	0.559	0.267	0.267	0.833	
59	0.000	0.036	0.100	0.074	0.088	0.033	0.000	0.000	
75	0.043	0.000	0.017	0.015	0.029	0.067	0.200	0.157	
100	0.022	0.000	0.000	0.000	0.000	0.533	0.367	0.010	
123	0.000	0.000	0.000	0.000	0.000	0.000	0.067	0.000	
<i>Acon-2</i>									
(n)	32	26	60	30	31	15	15		
75	0.000	0.019	0.017	0.000	0.016	0.000	0.000		
80	0.000	0.000	0.025	0.000	0.000	0.000	0.000		
89	0.266	0.308	0.642	0.767	0.290	0.000	0.000		
95	0.109	0.096	0.225	0.133	0.242	0.033	0.367		
100	0.609	0.577	0.092	0.100	0.435	0.733	0.467		
106	0.016	0.000	0.000	0.000	0.016	0.233	0.167		
114	0.000	0.000	0.000	0.000	0.000	0.000	0.000		

## Appendix Continued

Locus	Population code							
	10	11	12	13	14	15	16	17
<i>Adh-1</i>								
(n)	32	31	57	35	35	15	15	59
83	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.017
45	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.000
40	0.000	0.065	0.000	0.000	0.000	0.000	0.000	0.000
35	0.000	0.387	0.000	0.043	0.029	0.000	0.067	0.000
01	0.000	0.000	0.009	0.000	0.000	0.000	0.000	0.729
-38	0.000	0.000	0.000	0.000	0.000	0.033	0.100	0.000
-46	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-100	1.000	0.548	0.991	0.957	0.971	0.967	0.700	0.254
-140	0.000	0.000	0.000	0.000	0.000	0.000	0.100	0.000
<i>Ak-3</i>								
(n)	32	23	60	35	35	15	15	
68	0.000	0.000	0.017	0.000	0.000	0.000	0.000	
83	0.016	0.000	0.008	0.000	0.000	0.000	0.000	
87	0.047	0.043	0.108	0.100	0.086	0.133	0.100	
100	0.938	0.913	0.825	0.900	0.914	0.867	0.900	
111	0.000	0.043	0.042	0.000	0.000	0.000	0.000	
<i>Dia-2</i>								
(n)	32	31	60	35	35	15	15	
70	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
85	0.953	1.000	1.000	1.000	0.986	0.400	0.233	
100	0.047	0.000	0.000	0.000	0.014	0.600	0.767	
<i>Fum</i>								
(n)	32	30	60	35	35	15	15	59
100	0.969	1.000	1.000	1.000	1.000	1.000	1.000	0.966
158	0.031	0.000	0.000	0.000	0.000	0.000	0.000	0.034
<i>Had</i>								
(n)	32	31	60	35	35	15	15	
72	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
98	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
100	0.109	0.000	0.050	0.000	0.000	0.433	0.367	
125	0.891	1.000	0.950	1.000	1.000	0.567	0.633	
<i>Idh</i>								
(n)	32	31	60	35	35	15	15	40
82	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.013
87	0.000	0.000	0.025	0.000	0.000	0.000	0.000	0.000
93	0.000	0.032	0.000	0.000	0.000	0.000	0.000	0.000
100	0.672	0.613	0.658	0.514	0.657	0.600	0.467	0.700
113	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
118	0.328	0.226	0.317	0.471	0.300	0.367	0.533	0.287
128	0.000	0.129	0.000	0.014	0.043	0.000	0.000	0.000
135	0.000	0.000	0.000	0.000	0.000	0.033	0.000	0.000
<i>Mpi</i>								
(n)	24	31	60	35	35	15	15	
63	0.000	0.000	0.017	0.000	0.000	0.000	0.000	
70	0.042	0.000	0.008	0.029	0.014	0.000	0.100	
85	0.042	0.016	0.008	0.000	0.029	0.000	0.000	
100	0.875	0.984	0.967	0.943	0.957	1.000	0.667	
112	0.042	0.000	0.000	0.029	0.000	0.000	0.233	

## Appendix Continued

Locus	Population code							
	10	11	12	13	14	15	16	17
<i>Pgi</i>								
(n)	32	31	60	35	35	15	15	59
43	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.060
53	0.000	0.016	0.000	0.000	0.000	0.000	0.000	0.000
n76	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
80	0.016	0.032	0.000	0.014	0.000	0.033	0.033	0.000
100	0.891	0.952	0.967	0.943	0.957	0.967	0.900	0.720
130	0.047	0.000	0.033	0.029	0.014	0.000	0.000	0.220
145	0.047	0.000	0.000	0.014	0.029	0.000	0.067	0.000
<i>Pgm</i>								
(n)	32	31	60	35	35	15	15	59
80	0.000	0.000	0.017	0.000	0.014	0.000	0.000	0.000
92	0.063	0.032	0.050	0.086	0.086	0.200	0.167	0.085
95	0.000	0.000	0.000	0.014	0.029	0.000	0.000	0.000
100	0.813	0.887	0.783	0.900	0.829	0.667	0.700	0.915
104	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
111	0.094	0.081	0.150	0.000	0.029	0.100	0.067	0.000
118	0.000	0.000	0.000	0.000	0.014	0.033	0.067	0.000
125	0.031	0.000	0.000	0.000	0.000	0.000	0.000	0.000
<i>Tpi</i>								
(n)	32	31	60	35	35	15	15	
68	0.016	0.000	0.000	0.000	0.000	0.000	0.000	
85	0.000	0.000	0.000	0.000	0.014	0.000	0.000	
90	0.016	0.000	0.008	0.000	0.000	0.000	0.000	
100	0.969	1.000	0.992	1.000	0.986	1.000	1.000	
120	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
<i>Acon-1</i>								
(n)	32	31	60	35	35	15	15	
-100	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
-120	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
-300	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
<i>Ald</i>								
(n)	32	31	60	35	35	15	15	58
100	1.000	1.000	0.992	1.000	1.000	1.000	1.000	1.000
162	0.000	0.000	0.008	0.000	0.000	0.000	0.000	0.000
<i>Mdh-1</i>								
(n)	32	23	60	31	35	15	15	
-73	0.016	0.000	0.000	0.032	0.000	0.000	0.000	
-100	0.984	1.000	1.000	0.968	1.000	1.000	1.000	
-108	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
<i>Mdh-2</i>								
(n)	32	23	60	31	35	15	15	
43	0.000	0.043	0.000	0.000	0.000	0.000	0.000	
100	0.984	0.957	1.000	1.000	1.000	0.833	1.000	
147	0.016	0.000	0.000	0.000	0.000	0.167	0.000	
$\bar{H}$	0.173	0.166	0.149	0.131	0.147	0.222	0.273	
$\bar{a}$	2.5	2.1	2.4	2.0	2.3	2.0	2.2	