

Host race or species? Allozyme characterization of the 'flowering dogwood fly', a member of the *Rhagoletis pomonella* complex

STEWART H. BERLOCHER

University of Illinois at Urbana-Champaign, Department of Entomology/320 Morrill Hall, 505 S. Goodwin Avenue, Urbana, IL 61801, U.S.A.

The term 'flowering dogwood fly' has been used in the literature for a poorly understood member of the *Rhagoletis pomonella* sibling species complex infesting the fruits of flowering dogwood (*Cornus florida*). Electrophoretic analysis of 17 allozyme loci in 21 populations reveals significant frequency differences between the flowering dogwood fly and its closest relative the apple maggot fly, *R. pomonella*, and between it and the somewhat more distant 'sparkleberry fly'. Frequency differences between the flowering dogwood fly and *R. pomonella* are as great as 0.817 in the north, but are less in the south, with a maximum difference at one site of only 0.328. No fixed allozyme differences distinguish the flowering dogwood fly anywhere; its only consistent, unique feature is the highest frequency of *Aat-2*⁵⁹ in the *pomonella* species group. Population structure of the flowering dogwood fly is moderate with $F_{ST} = 0.084$ and fewer latitudinal clines than *R. pomonella*. The conclusion from the allozyme and life history data is that the flowering dogwood fly is a species, although some interspecific gene flow may be occurring. Additional issues discussed include how to estimate interspecific gene flow when genetic markers are under divergent selection, the appropriate species concept when there is gene flow, and the future of the flowering dogwood fly in the face of the dogwood anthracnose epidemic. The possible utility of a new species concept for phytophagous insects, using as a criterion the capacity of a host race to regenerate the ancestral population, is also discussed.

Keywords: allozymes, gene flow, host race, *Rhagoletis*, species concepts, sympatric speciation.

Introduction

The *Rhagoletis pomonella* species group (Bush, 1966), a complex of very close sibling species, remains the primary focus of debate on sympatric speciation by ecological specialization. The initial step of Bush's sympatric speciation model for *Rhagoletis* and other phytophagous and parasitic animals (Diehl & Bush, 1989; Bush, 1994) is the evolution of a host race (a partially reproductively isolated population on a new host; Feder, 1998). The model requires both larval host specialization and adult mating on the host, or host

fidelity as it has been recently termed (Feder *et al.*, 1994). The best documented example of host race formation is the ~140-year-old (= 140 generations of this univoltine insect) apple race of *R. pomonella* (Walsh), which arose from the ancestral hawthorn-infesting race (review in Feder, 1998).

The second theoretical step in sympatric speciation is improvement of host fidelity, leading eventually to the production of species (Bush, 1994; Berlocher, 1998). Study of this second step is more difficult than study of host race formation, for several reasons. One is that the appropriate species concept (reviewed in de Queiroz, 1998; Harrison, 1998) is unclear. The biological species concept (BSC) has been the universal choice of speciation workers, as the evolution of reproductive isolation is what allows subsequent evolutionary divergence. Yet in principle a host race may become evolutionarily independent long before gene flow completely ceases

Correspondence: Stewart H. Berlocher, University of Illinois at Urbana-Champaign, Department of Entomology/320 Morrill Hall, 505 S. Goodwin Avenue, Urbana, IL 61801, U.S.A.
Tel.: (217) 333-2929. Fax: (217) 244-3499. E-mail: stewartb@uiuc.edu
*E-mail: stewartb@uiuc.edu

(Bush, 1994). Thus under the BSC *sensu stricto* (permitting absolutely no gene flow) such an independent lineage could not be recognized as a species. This problem can be solved by using a 'loose' or nonstrict version of the BSC, which is the approach taken here, but then a second problem surfaces: some sort of critical value or species threshold for gene flow is now needed. I am unaware of any published guidelines on a gene flow threshold value below which populations are considered species. Any attempt to define such a threshold would need to consider standing allele frequency differences as well as direct gene flow measurements, because allele frequency data are much easier to obtain than direct gene flow measurements, and can be used to estimate gene flow indirectly (Neigel, 1997).

Any threshold gene flow value will differ among taxa, as it would depend on the strength of divergent selection imposed by the hosts. The apple and hawthorn host races of *R. pomonella* provide an example of one particular balance between gene flow and selection. Feder *et al.* (1994) determined gene flow to be ~6% by direct, mark-release-recapture methods, but frequency differences between the races at six allozyme loci are generated by the different temperature regimes experienced by the races (apple race flies experience earlier season, warmer conditions than hawthorn flies; Feder *et al.*, 1997). In the absence of selection for genotypes adapted to different temperatures, the ~6% interracial gene flow would rapidly homogenize allele frequencies — but with selection added, the races reach equilibria with allele frequency differences of ~0.15. A parallel balance between selection and gene flow is likely in very close species. Thus, the common practice of treating all loci as neutral 'markers' when determining species status from indirectly estimated gene flow may result in erroneous conclusions.

The 'flowering dogwood fly' (Berlocher *et al.*, 1993) is a promising candidate for the study of step two of sympatric speciation, the transition from host race to species. As indicated by the name 'flowering dogwood fly' (henceforth dogwood fly), it infests only the fruits of *Cornus florida* L. The dogwood fly is very close to both *R. pomonella* (Berlocher *et al.*, 1993) and another undescribed *pomonella* group species, known at present as the 'sparkleberry fly' (Payne & Berlocher, 1995a; sparkleberry is unique in being a fall-fruiting blueberry). *Rhagoletis pomonella*, the dogwood fly, and the sparkleberry fly are all broadly sympatric in eastern North America. Prior allozyme analysis (Berlocher *et al.*, 1993) revealed only frequency differences between the dogwood fly and *R. pomonella*, and morphologically the dogwood fly is indistinguishable from *R. pomonella* (Bush, 1966). No fixed differences between the dogwood fly and *R. pomonella* were observed at the mitochondrial COII locus (Smith & Bush, 1996).

This paper analyses allozyme divergence of the dogwood fly from *R. pomonella* and the sparkleberry fly. However, I concentrate on the dogwood fly and *R. pomonella*, as these taxa are closer to each other than either is to the sparkleberry fly, based on a 29-locus allozyme analysis (S. H. Berlocher, unpubl. data). I examine in detail the geographical population structure of the dogwood fly, as the allele frequency clines and other structures of *R. pomonella* (Berlocher & McPherson, 1996) complicate efforts to understand species divergence in the *pomonella* species group. The central issue, however, is where the dogwood fly should be placed on the continuum from host race to species. In attempting to determine this point, I discuss whether alternatives to the BSC are useful in this particular case, and for phytophagous insects in general.

Materials and methods

Sampling of material and analysis of data was performed as described in previous papers of this series, so only a brief overview is given here. Henceforth 'this series' refers to the host plant papers of Berlocher & Enquist (1993) and Payne & Berlocher (1995a,b), and the population structure papers of Payne & Berlocher (1995a), Berlocher (1995), and Berlocher & McPherson (1996). Insects were reared to adulthood from larvae in infested fruit collected in the field, and frozen at -80°C for later analysis. Where possible, fruit collections from sets of hawthorn, flowering dogwood and sparkleberry trees in close proximity were made to avoid confounding host-related differences with ordinary geographical variation. Samples chosen for electrophoretic analysis are given in Table 1; for exact site locations contact the author. Standard single-condition horizontal starch gel electrophoresis was employed. As with all papers of this series, a standard 17-locus subset of the 29 studied by Berlocher *et al.* (1993) was studied.

As with the electrophoresis, population genetic analysis was identical to that in the previous papers in this series. The zero probabilities recorded in some tables indicate that 10^7 random samples of a Monte Carlo exact contingency test deviated less from expectations than did the actual sample, although of course the probability must be some very small nonzero value. The complete allozyme data file may be obtained from the author.

Results

Ecology of the dogwood fly

I collected a total of ~56 000 flowering dogwood fruit at 33 geographical sites. Flies were successfully reared at 31

Table 1 Sample information for flowering dogwood fly (FDF) collection sites. All collections were made in the U.S.A. from *Cornus florida*

Site locality (County/Parish), state	Collection		Infestation rate (%)	Latitude
	Date	No. fruit		
Tuskegee National Forest (Macon), AL	1/11/89	1880	10.4	32°30'
Gainesville (Alachua), FL	5/11/89	1255	40.8	29°48'
Morgan State Nursery, nr. Byron (Peach), GA	2/11/89	10 566	11.4	32°39'
Folkston (Charleston), GA	5/11/89	1602	34.4	30°50'
Oak Park (Emanuel), GA	6/11/89	2318	5.4	32°22'
3 miles north-west of Byron (Peach), GA	2/11/89	2806	7.7	32°40'
Carbondale (Jackson), IL	9/10/89			37°44'
Fairfield (Wayne), IL	Autumn/81			38°23'
I-57 rest stop (Union), IL	28/10/89	191	46.6	37°49'
Baton Rouge (E. Baton Rouge), LA	31/10/89	742	38.7	30°27'
Bogalusa (Washington), LA	31/10/89	1336	54.5	30°48'
Beltsville Natl. Agri. Res. Center (Prince Georges), ML	10/89	2524	2.1	39°04'
Bill Meyer Wildlife Manag. Area (Allegany), ML	18/9/89			39°33'
Route 8 and Co. Y (Crawford), MO	14/10/91	1328	48.6	37°48'
Fayetteville (Cumberland), NC	7/11/89	1947	3.8	35°00'
Rocky Mt. (Edgecombe), NC	7/11/89	4724	2.6	36°02'
Princeton (Mercer), NJ	Autumn/88			40°20'
Clarks Hill Reservoir (McCormick), SC	6/11/89	1384	3.8	33°42'
Nacogdoches (Nacogdoches), TX	29/10/89	427	19.7	31°32'
Wright Patman Lake (Bowie), TX	29/10/89	1390	9.6	33°23'
Bowling Green (Caroline), VA	8/11/89	1575	7.2	38°03'

(94%) of these. Mean infestation rate was 17.7% (SD = ±20.6%, range 0–54.5%), the highest observed in this series; the dogwood fly is a common insect where its host occurs. Based upon collection dates for infested fruit (from this study, other papers in this series, and J. L. Feder, B. A. McPheron & S. H. Berlocher, unpubl. data), the dogwood fly occurs later in the year than most *R. pomonella* populations, but concurrently with the sparkleberry fly (Fig. 1; see Discussion). For all three taxa infested fruit was found later at lower latitudes (Fig. 1).

Population structure of the dogwood fly

Samples from 18 of my collection sites were chosen for electrophoretic analysis (Table 1), augmented by three collections from colleagues (Fairfield, IL, D. C. Smith; Beltsville National Agricultural Research Center, MA, G. Steck; Princeton, NJ, J. L. Feder.) The mean and standard deviation across populations of average heterozygosity is 0.134 ± 0.013 , which places it below *R. pomonella* (0.191; Berlocher & McPheron, 1996) and *R. mendax* (0.154; Berlocher, 1995), but above the sparkleberry fly (0.115; Payne & Berlocher, 1995a). The fact that the two species with the lowest heterozygosities are also both specialists on single species of host plants

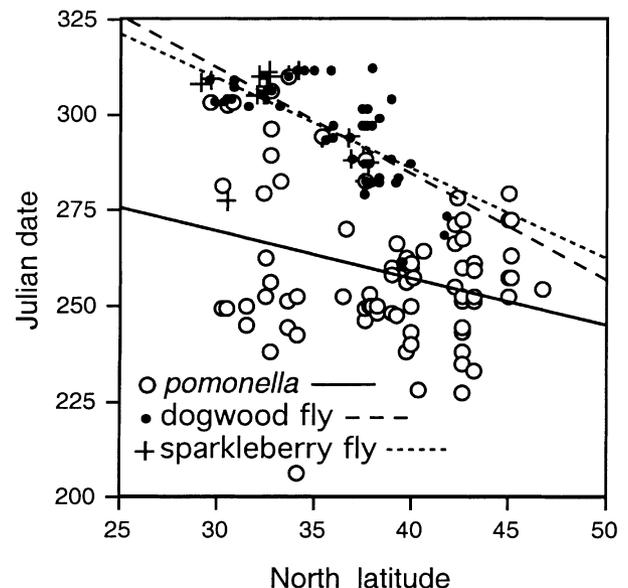


Fig. 1 Date of collection of infested fruit vs. latitude of collection site for *Rhagoletis pomonella*, the dogwood fly (FDF), and the sparkleberry fly (SF). Data include many collections not used for electrophoresis (see text). Slopes are ($x = ^\circ\text{N}$ latitude, $y = \text{Julian date}$): *R. pomonella*, $y = -1.21x + 305.55$; FDF, $y = -2.78x + 395.65$; SF, $y = -2.36x + 379.82$. All $P < 0.01$.

raises the possibility that multiple-niche selection may be operating to maintain greater amounts of variation in the species using more host species.

All populations were in HWE, as in other *pomonella* group species reported on in this series. In the geographical contingency tests, 11 of the 13 testable (Berlocher & McPheron, 1996) loci showed significant interpopulation heterogeneity (at the 5% level after sequential Bonferroni correction), and at eight loci the heterogeneity is highly significant (Table 2). Geographical differentiation is also significantly greater than zero under a random model approach, as F_{ST} calculated as Cockerham & Weir θ (Weir, 1990) is 0.084, with 95% confidence limits calculated by bootstrapping over loci of 0.011 and 0.215. The reason for the very large bootstrapped confidence limits is that one locus, *Acon-2*, displays far more geographical differentiation than any other (Table 2). As *Acon-2* (or some locus near it) is known to be strongly affected by selection related to larval developmental temperature (Feder *et al.*, 1997), the use of mean F_{ST} to infer intraspecific, geographical gene flow is probably invalid. This does not obviate the use of F_{ST} as a descriptive statistic.

Significant latitudinal clines exist at two loci, *Acon-2* and *Had* (no longitudinal effects were seen). As with *R. pomonella* (Berlocher & McPheron, 1996), clines first determined to be significant by nonparametric statistics ($P < 0.05$ after correction for multiple tests) were then plotted and regressed with untransformed frequencies

Table 2 θ values and contingency test P -values at 17 loci for 21 geographical samples representing the geographical range of the flowering dogwood fly. See text for explanation of zero P -values. NT, too little variation to test (Berlocher & McPheron, 1996); M, monomorphic

Locus	θ	P
<i>Aat-2</i>	0.011	0
<i>Acon-2</i>	0.346	0
<i>Adh-1</i>	0.011	0.0013
<i>Ak-3</i>	0.019	0
<i>Dia-2</i>	0.017	0
<i>Fum</i>	0.009	NT
<i>Had</i>	0.033	0
<i>Idh</i>	0.027	0
<i>Mpi</i>	0.012	0
<i>Pgi</i>	0.011	0
<i>Pgm</i>	0.008	0.0003
<i>Tpi</i>	-0.004	0.95
<i>Acon-1</i>	0.004	NT
<i>Ald</i>	-0.003	NT
<i>Dia-1</i>	M	M
<i>Mdh-1</i>	0.012	0.0090
<i>Mdh-2</i>	0.011	0.0478

for display (Fig. 2). Both *Acon-2*⁸⁹ and *Acon-2*⁹⁵ showed a decrease towards the north, and *Had*¹⁰⁰ an increase towards the north. The clines in *Acon-2* are the cause of the very large θ at this locus (Table 2). Unlike *R. pomonella* (Berlocher & McPheron, 1996), no significant latitudinal clines in average heterozygosity or average number of alleles were found. Also unlike *R. pomonella* (Berlocher & McPheron, 1996) and *R. mendax* (Berlocher, 1995), no linkage disequilibrium was detected.

Genetic differences between the dogwood fly and close relatives

As reported previously based on a 29-locus comparison of a single pair of Illinois samples (Berlocher *et al.*, 1993), only allozyme frequency differences occur between *R. pomonella* and the dogwood fly. Similarly, only frequencies differentiate the sparkleberry fly [Payne & Berlocher (1995a) contains frequencies for six representative populations; the remaining six also analysed here are available from the author]. With the larger set of dogwood fly samples now available, a complicated pattern of differentiation is observed, in particular between *R. pomonella* and the dogwood fly. For northern populations results are in accord with Berlocher *et al.* (1993). In this study three northern paired *R. pomonella*-dogwood fly samples, in which both members of a pair were from the same year, were available (*R. pomonella* data from Berlocher & McPheron, 1996): Beltsville Agricultural Research Center, MA samples ~0.5 km apart; Fairfield, IL and Wayne Co., IL samples ~2.2 km apart; Bill Meyer Wildlife Area, MA and U.S. 40/48 at eastern Continental divide, samples ~70 km apart. At all three sites, large, highly significant, geographically consistent differences occur at four loci: *Aat-2*, *Acon-2*, *Had* and *Dia-2*. Data for these loci for the Fairfield and Wayne Co., IL, pair of sites are shown in Table 3. Note that in the Tables and Discussion the expression Δp is used for the largest allele frequency difference in a pairwise comparison, and only loci at which differentiation is statistically significant are discussed. Table 3 also includes a comparison of the Fairfield, IL dogwood fly sample with a Champaign, IL *R. pomonella* sample (sites ~60 km apart) for three loci not considered in this study. The Δp for *Aat-1* is the highest observed thus far, 0.817.

However, in the south the Δp s between the dogwood fly and *R. pomonella*, and also the sparkleberry fly, are much smaller. Fewer of the 17 loci show significant differentiation, and the two that do, *Aat-2* and *Acon-2*, show smaller Δp s. Table 4 presents frequency data for the three closest paired dogwood fly and *R. pomonella* samples: dogwood fly and *R. pomonella* sampled at

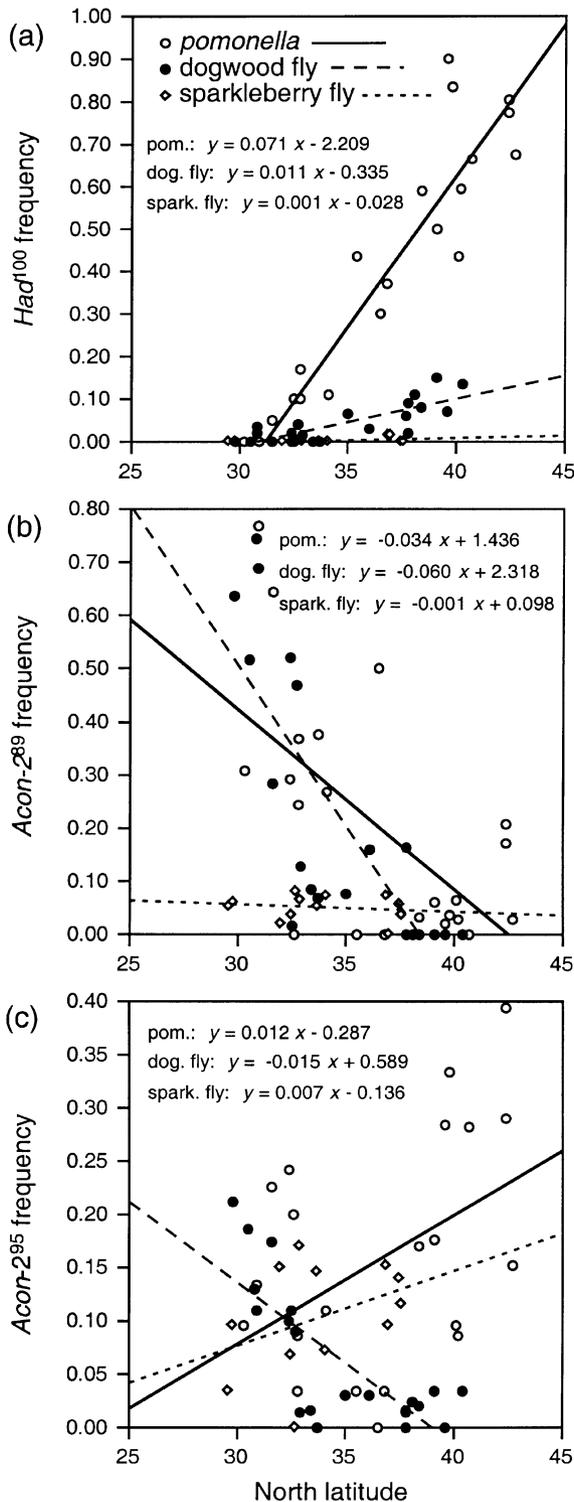


Fig. 2 Allele frequency differences between the sibling species for loci that are generally clinal. The parametric regressions against latitude are used descriptively here; nonparametric regression was used to test significance (see text). (a) Had ; (b) $Acon-2^{89}$; (c) $Acon-2^{95}$.

Nacogdoches, TX (~10 m apart) in 1985 and 1989, respectively; near Byron, GA (~16 km apart) in 1989; and Clarks Hill Reservoir, SC (~3 km apart) in 1985 and 1989. In addition, at the SC site, a sparkleberry fly sample ~3 km from each of the other two host collections (Payne & Berlocher, 1995a) was also available for comparison. Despite the genetic similarity of the dogwood fly, sparkleberry fly and *R. pomonella* in the south, Table 4 shows that consistent allozyme frequency differences do occur. In comparisons of *R. pomonella* and the dogwood fly, for example, $Aat-2^{59}$ occurs at all three sites at significantly higher frequency in the dogwood fly, and the opposite is seen for $Acon-2^{89}$.

These patterns suggested that analysis of allozyme frequencies across the geographical ranges of the three taxa might reveal consistent patterns. Such patterns do occur, and can be divided into those that involve primarily nonclinally varying loci, and those that involve clinal loci.

$Aat-2^{59}$, $Aat-2^{100}$ and Pgi^{130} were involved in non-clinal patterns (Fig. 3). The dogwood fly is characterized by the highest frequency of $Aat-2^{59}$ among the three taxa considered here, and in fact it possesses the highest frequency of this allele anywhere in the *pomonella* group. Frequencies of $Aat-2^{59}$ in the sparkleberry fly and dogwood fly do not overlap ($P < 0.001$ by nonparametric Mann-Whitney U -test), and overlap between *R. pomonella* and the dogwood fly is slight ($P < 0.001$). For $Aat-2^{100}$ the frequencies of the dogwood fly and sparkleberry fly do not overlap ($P < 0.001$), but the dogwood fly and *R. pomonella* overlap almost entirely [although $Aat-2^{100}$ is clinal in *R. pomonella* (Berlocher & McPherson, 1996) but not in the dogwood fly]. For Pgi^{130} the sparkleberry fly has the highest frequency among the three taxa (and in the *pomonella* group as a whole), and does not overlap with the dogwood fly (but does overlap slightly with *R. pomonella*, with $P < 0.001$ in both cases); *R. pomonella* and the dogwood fly overlap broadly.

Patterns of frequency difference for the clinal alleles (Fig. 2) are particularly complicated. At latitudes of 29–34°N, all three taxa display Had^{100} frequencies at or very close to 0.0. As one moves north, Had^{100} remains at very low frequency in sparkleberry fly populations (Payne & Berlocher, 1995a). The dogwood fly shows a shallow but highly significant clinal pattern ($P = 0.0003$), whereas *R. pomonella* shows a very steep cline, with northern populations approaching fixation for Had^{100} (Berlocher & McPherson, 1996). Thus all three taxa become increasingly different from one another as one moves northwards.

At $Acon-2^{89}$, all three taxa possess relatively low frequencies at around 40°N, but differ as one moves southwards. The sparkleberry fly is again uniformly characterized by very low frequencies of $Acon-2^{89}$, but

Table 3 Allele frequency comparison of northern dogwood fly and *Rhagoletis pomonella* samples in close geographical proximity. Differences significant at these seven loci at $P = 0$ (see text). Δp = largest allele frequency difference. N = sample size

(a) Comparison of 1981 Fairfield, IL dogwood fly and Wayne Co., IL *R. pomonella* samples

Allele	<i>Aat-2</i> , $\Delta p = 0.459$		<i>Acon-2</i> , $\Delta p = 0.273$		<i>Had</i> , $\Delta p = 0.526$		<i>Dia-2</i> , $\Delta p = 0.651$				
	dogwood fly ($N = 76$)	<i>pomonella</i> ($N = 59$)	dogwood fly ($N = 76$)	<i>pomonella</i> ($N = 62$)	dogwood fly ($N = 76$)	<i>pomonella</i> ($N = 62$)	dogwood fly ($N = 76$)	<i>pomonella</i> ($N = 59$)			
21	0.072	0.203	89	0.000	0.032	98	0.007	0.008	85	0.967	0.316
32	0.000	0.008	95	0.020	0.169	100	0.079	0.605	100	0.033	0.684
50	0.612	0.153	100	0.934	0.661	111	0.007	0.000			
59	0.257	0.000	106	0.039	0.129	125	0.901	0.387			
75	0.033	0.127	114	0.007	0.008	135	0.007	0.000			
100	0.026	0.492									
123	0.000	0.008									
130	0.000	0.008									

(b) Comparison of 1981 Fairfield Co., IL dogwood fly sample with 1985 Champaign Co., IL *R. pomonella* sample

	<i>Aat-1</i> , $\Delta p = 0.817$		<i>Acy</i> , $\Delta p = 0.625$		<i>Pep</i> , $\Delta p = 0.700$			
	dogwood fly ($N = 30$)	<i>pomonella</i> ($N = 60$)	dogwood fly ($N = 30$)	<i>pomonella</i> ($N = 60$)	dogwood fly ($N = 30$)	<i>pomonella</i> ($N = 60$)		
-48	0.083	0.000	89	0.033	0.142	94	0.000	0.017
-64	0.917	0.158	100	0.217	0.733	100	0.000	0.542
-88	0.000	0.008	110	0.750	0.125	106	0.967	0.267
-100	0.000	0.817				110	0.033	0.175
-110	0.000	0.017						

Table 4 Allele frequency comparison of southern dogwood fly with southern *Rhagoletis pomonella* and sparkleberry fly samples, sampled in close geographical proximity. Numbers after abbreviations are years. See text for locales. *P*-values for contingency tests of differences between the dogwood fly and *R. pomonella* (see text) are given below each pair of allele frequency columns tested. Δp = largest allele frequency difference. *N* = sample size
(a) *Aat-2*

	Texas, $\Delta P = 0.507$		Georgia, $\Delta P = 0.403$		South Carolina, $\Delta P = 0.201$		
	dogwood fly	<i>pomonella</i>	dogwood fly	<i>pomonella</i>	dogwood fly	<i>pomonella</i>	sparkleberry fly
	1989 (<i>N</i> = 23)	1985 (<i>N</i> = 60)	1989 (<i>N</i> = 36)	1989 (<i>N</i> = 30)	1989 (<i>N</i> = 31)	1985 (<i>N</i> = 20)	1989 (<i>N</i> = 41)
1	0.000	0.000	0.000	0.050	0.000	0.000	0.000
21	0.043	0.550	0.097	0.500	0.129	0.275	0.085
50	0.609	0.333	0.722	0.383	0.645	0.625	0.805
59	0.348	0.100	0.181	0.050	0.226	0.025	0.000
75	0.000	0.017	0.000	0.017	0.000	0.050	0.000
100	0.000	0.000	0.000	0.000	0.000	0.025	0.110
	<i>P</i> = 0		<i>P</i> = 0		<i>P</i> = 0.001		

(b) *Acon-2*

	Texas, $\Delta P = 0.451$		Georgia, $\Delta P = 0.300$		South Carolina, $\Delta P = 0.328$		
	dogwood fly	<i>pomonella</i>	dogwood fly	<i>pomonella</i>	dogwood fly	<i>pomonella</i>	sparkleberry fly
	1989 (<i>N</i> = 23)	1985 (<i>N</i> = 60)	1989 (<i>N</i> = 36)	1989 (<i>N</i> = 30)	1989 (<i>N</i> = 31)	1985 (<i>N</i> = 20)	1989 (<i>N</i> = 41)
75	0.000	0.017	0.000	0.000	0.000	0.000	0.000
80	0.000	0.025	0.000	0.000	0.000	0.000	0.000
89	0.283	0.642	0.125	0.367	0.065	0.375	0.049
95	0.174	0.225	0.014	0.033	0.000	0.000	0.146
100	0.543	0.092	0.833	0.533	0.903	0.575	0.805
106	0.000	0.000	0.028	0.067	0.016	0.050	0.000
114	0.000	0.000	0.000	0.000	0.016	0.000	0.000
	<i>P</i> = 0		<i>P</i> = 0.001		<i>P</i> = 0.001		

(c) *Pgi*

	Texas		Georgia		South Carolina		
	dogwood fly	<i>pomonella</i>	dogwood fly	<i>pomonella</i>	dogwood fly	<i>pomonella</i>	sparkleberry fly
	1989 (<i>N</i> = 23)	1985 (<i>N</i> = 60)	1989 (<i>N</i> = 36)	1989 (<i>N</i> = 30)	1989 (<i>N</i> = 31)	1985 (<i>N</i> = 20)	1989 (<i>N</i> = 41)
53	0.000	0.000	0.000	0.000	0.016	0.000	0.000
80	0.109	0.000	0.000	0.000	0.032	0.000	0.000
100	0.891	0.967	0.931	0.900	0.887	1.000	0.780
123	0.000	0.000	0.000	0.000	0.016	0.000	0.000
130	0.000	0.033	0.014	0.000	0.000	0.000	0.220
145	0.000	0.000	0.056	0.100	0.048	0.000	0.000

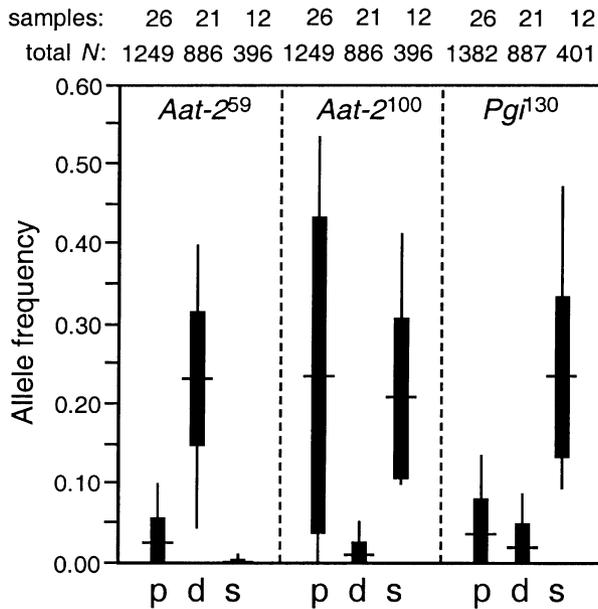


Fig. 3 Allele frequency differences (mean, SD, range) between the sibling species for loci that are generally nonclinal. p, *Rhagoletis pomonella*; d, flowering dogwood fly; s, sparkleberry fly. Samples, number of samples; N, sum of individuals across all samples.

both *R. pomonella* ($P=0.021$) and the dogwood fly ($P < 0.0001$) show a clinal increase in frequency towards the south, in this case being very similar in the two taxa. The *R. pomonella* regression for this allele was not reported in Berlocher & McPherson (1996) because the P -value did not fall below the 0.05 critical value after Bonferroni correction, but in light of the similarity of the regression line to that of the dogwood fly it is likely to be significant with further sampling.

The pattern at *Acon-2*⁸⁹ is statistically the weakest, but potentially of great interest because the direction of the cline may be different in the dogwood fly and *R. pomonella*. In the dogwood fly, allele frequencies decline significantly ($P=0.001$) as one moves northwards, the opposite of the nonsignificant ($P=0.225$, $P=0.455$) trends in the other two taxa.

Discussion

Reduced to the barest essence, the results of this allozyme study are: frequency differences as high as 0.817 occur between the dogwood fly and *R. pomonella* in north-eastern North America, with significant differences at seven of 29 loci (at four of the 17 studied in detail here), but in south-eastern North America the maximum frequency difference at a site was as low as 0.328, with significant differences at only two of 17 loci.

Lack of fixed allozyme differences greatly complicates decisions on species status. With fixed differences a straightforward decision under the BSC is possible because no gene flow can be occurring (Avise, 1994). Even without fixation, most workers have accepted species status for taxa distinguished by very large frequency differences; for example, *R. pomonella* and its fellow *pomonella* group taxon *R. mendax* (blueberry maggot) are universally regarded as species, with allozyme frequency differences at several loci between ~ 0.80 and ~ 0.90 (Feder & Bush, 1989). But given smaller frequency differences, such as those between the dogwood fly and *R. pomonella*, one can say only that gene flow must be within a range from zero to some value short of panmixia.

The allozyme data could be consistent with zero gene flow between the dogwood fly and *R. pomonella* in two ways. First, fixed interspecific differences might occur somewhere in the genome, and simply not have been sampled in this study. Secondly, fixed allele differences might be lacking, but gene flow prevented by essentially fixed *phenotypic* differences in prezygotic isolation/host fidelity characters. This situation is possible if characters contributing to host fidelity are polygenically controlled by many loci; in this case, the allozymes may not be misleading about the true level of allelic differentiation. I note that two characters integral to prezygotic isolation in these flies do appear to be polygenically inherited. These are postdiapause emergence time [dogwood-fly flies are active later than most *R. pomonella* (Fig. 1)], and fruit size choice (dogwood-fly flies prefer smaller artificial fruit than do *R. pomonella*; Smith, 1986).

Turning to the possibility that gene flow is *not* zero between the dogwood fly and *R. pomonella*, it is obvious that natural selection would be required to prevent genetic homogenization. As discussed in the Introduction, such selection maintains the apple and hawthorn host races of *R. pomonella*. Three aspects of the population genetics of the dogwood fly and *R. pomonella* suggest that similar selection–migration equilibria are occurring. First, the same loci and in many cases the same allozymes are involved in differentiating both the host races, and the dogwood fly and *R. pomonella*. Secondly, the dogwood fly and *R. pomonella* develop under different temperature regimes (the dogwood fly is active later in the year, and is exposed to cooler temperatures in the adult, larval and early pupal stages), and are thus likely to be exposed to the same kind of divergent climate-related selection as are the host races. Thirdly, the dogwood fly shows clines (Fig. 2) at some of the same loci showing climate-related clines in *R. pomonella* (Feder *et al.*, 1993).

However, allele frequency differences between *R. pomonella*, the dogwood fly, and the sparkleberry

fly are clearly larger than those between the *R. pomonella* host races. Averaged across several years, localities and loci in the Midwest, $\Delta p \sim 0.15$ for the host races (Feder *et al.*, 1993 and references therein). By comparison, mean Δp for the dogwood fly and *R. pomonella* in the south is somewhat more than twice as large, at 0.36 (Table 4). Values for Δp are greater in the north. In the Wayne Co., IL comparison, mean Δp across *Aat-2*, *Acon-2*, *Had* and *Dia-2* is 0.474 (Table 3). With the addition of the three loci from Berlocher *et al.* (1993), mean Δp in Illinois over seven loci is 0.577. These larger Δp s imply either less gene flow or stronger divergent selection. By making some reasonable assumptions about the strength of divergent selection, quantitative estimates of gene flow can be obtained.

For the Grant, MI site Feder *et al.* (1997) calculated the strength of selection needed to maintain the observed interracial frequency differences in the face of ongoing gene flow, using a model based on the classic single-gene, diallelic equations for selection–migration equilibrium of Wright (1931). Alleles are exchanged at rate m between two populations, each specializing on a different host. Constant viability selection after migration acts against one allele in one host and the other allele in the other host in an additive manner with selection coefficient s . From the model, Feder *et al.* (1997) calculated that $s=0.079$ would maintain the observed frequency differences between the races in the face of 6% gene flow. In the laboratory selection experiment, Feder *et al.* (1997) measured selection coefficients of the order of 0.28 (mean over four loci), a value much greater than that actually needed to maintain equilibria.

To estimate potential gene flow between the dogwood fly and *R. pomonella*, I employed the same model as Feder *et al.* (1997). I calculated, assuming that both populations start at $p=0.5$, values of s and m needed to produce Δp s of 0.15 (representative of the host races of *R. pomonella*), 0.35 (representative of *R. pomonella*–dogwood fly in the south), 0.60 (representative of *R. pomonella*–dogwood fly in the north) and 0.80 (typical *R. pomonella*–*R. mendax* value).

With $m=0.06$ (Feder *et al.*, 1994), the model predicts that $s=0.072$ will maintain a host race Δp of 0.15. [This s -value is slightly different from that of Feder *et al.* (1997) because they calculated using frequencies from only one site.] However, with this amount of gene flow, much greater values of s would be needed to maintain the larger Δp s observed between the dogwood fly and *R. pomonella*. The s -value required to maintain a ‘southern’ $\Delta p=0.35$ would be 0.174, and the value for a ‘northern’ $\Delta p=0.6$ would be 0.365; the s -value for $\Delta p=0.8$ is a very large 0.682. These values imply large selectional loads, at least in the north; for a single locus

at which $\Delta p=0.6$, the model predicts that 0.086 of both populations will die in each generation. If viability selection of this magnitude is occurring at all seven allozyme loci, then only 0.532 of the population is surviving selection at these loci each generation, which is clearly unrealistic. However, with smaller Δp s in the south, and with viability selection at only the two loci with significant differences, the selectional load is much less. Feder *et al.* (1997) did measure large s -values in their selection experiment, including one of 0.552 for *Had*, but the constant laboratory temperature regimes were more severe than flies would be exposed to in nature, with no cooling period at night. Overall, I conclude that at least in the north, m between the dogwood fly and *R. pomonella* must be smaller than the 0.06 value for host races, in order to reduce s to realistic values.

The question then becomes, how much smaller? One way to answer this question is to assume that s acting in the dogwood fly and *R. pomonella* is the same as that estimated for the host races in the field, or $s=0.072$. Under this assumption, the value of m that will give a Δp of 0.8 is 0.004, which is in accord with the observed absence of matings between *R. mendax* and *R. pomonella* in field studies (Feder & Bush, 1989). For $\Delta p=0.60$, representative of differences between the dogwood fly and *R. pomonella* in the north, the value of m is 0.010. However, m -values could be much smaller than the 1% I calculated. If one uses the largest observed Δp rather than the mean, one of course obtains a smaller m estimate. Using the largest observed northern Δp , 0.817 (at *Aat-1*, not sampled in the south), m is much smaller at 0.004 — the same as for *R. pomonella* and *R. mendax*. For $\Delta p=0.35$, however, representative of the dogwood fly and *R. pomonella* in the south, the m -value needed to produce an equilibrium is 0.023, of the same order of magnitude as the 0.06 measured for *R. pomonella* host races. [The range of Δp s in the south is smaller than in the north (Table 4), and results using the largest Δp are similar to those using the mean value.]

We now come to the crux of the problem: is an m of ~ 0.02 diagnostic of a species or a host race? Mayr (1963; p. 26) was probably not envisaging such a large value when he admitted that some gene exchange between biological species may occur. However, recent work has led to the conclusion that interspecific hybridization is common in some animals (Avisé, 1994; Dowling & Secor, 1997). In Darwin’s finches, for example, 1.8% of matings of *Geospiza fuliginosa* are with non-conspecifics (Grant, 1993).

Why is *G. fuliginosa* universally accepted as a species, despite the potential for $\sim 2\%$ gene flow? The likely reason is that morphological differences are maintained

between it and its congeners — unlike the case for the dogwood fly and *R. pomonella*. However, for several nonmorphological features the dogwood fly is distinct from *R. pomonella*. A small but significant amount of postzygotic isolation between the dogwood fly and *R. pomonella* occurs, consisting of a ~10% reduction in egg hatch in backcrosses (Smith, 1986). As already mentioned, differences in fruit size choice exist, as do allochronic differences (Fig. 1). Although some collection dates for the dogwood fly and *R. pomonella* overlap, in both the north and south, the vast majority differ substantially. The later emergence of the dogwood fly has been shown to have a genetic basis (Smith, 1988). Yet another biological difference between the dogwood fly and *R. pomonella* is that oviposition preferences contributing to host fidelity appear to be quite strong, at least in *R. pomonella*. Urbana, IL is just west of the northwesternmost edge of the natural range of *C. florida*, but has many ornamental plantings of the species. *Rhagoletis pomonella* on hawthorns is abundant in Urbana. But in 20 years of observing and collecting (S. H. Berlocher, unpubl. data), I have found not a single *R. pomonella* (nor any *Rhagoletis*) in *C. florida* fruit in Urbana, even though both native *C. florida* and the dogwood fly are common at sites starting just 50 km east of Urbana. Finally, recent wind tunnel studies by W. Roelofs and C. Linn have demonstrated that the flies very consistently discriminate between the fruit volatiles of apples and flowering dogwood (Roelofs and Linn, unpubl. data).

Overall, the allozyme data indicate that gene flow between the dogwood fly and *R. pomonella* must be relatively low, at least in the north, and the behavioural and ecological evidence indicate that the dogwood fly is maintaining itself as a unique population, even if some gene flow is occurring. Significantly, almost all of the traits discussed in the preceding paragraph can be viewed not only as 'characters' providing evidence for species status, but also as traits that increase host fidelity, and thus restrict gene flow. I conclude that species status under a 'nonstrict' version of the BSC is as justifiable for the dogwood fly as for *Geospiza* species.

The need to invoke a rather ramshackle, loose version of the BSC suggests that some other species concept might be more appropriate for the dogwood fly, and for other similar phytophagous insects. A search among the many current alternative species concepts or definitions (de Queiroz, 1998; Harrison, 1998) for any that explicitly permit ongoing gene flow reveals three possibilities. These are the cohesion concept of Templeton (1981), the genotypic cluster definition of Mallet (1995), and the lineage concept of de Queiroz (1998). But none of these provides a guide as to how much gene flow indicates a species rather than a host race; for example, application

of the genotypic cluster definition to the *pomonella* group reveals clusters for the apple race–hawthorn race comparison, as well as for comparison of unambiguous species like *R. pomonella* and *R. mendax*, with a continuum of decreasing degree of cluster overlap as level of genetic divergence increases from host race to distinct species (Feder, 1998). No species threshold is apparent — nor is one specified by the method. And in one sense this is acceptable, even desirable, as the logic of the sympatric speciation hypothesis absolutely requires the existence of sympatric transitional populations which may not be classifiable as either a host race or species.

However, for practical reasons the number of unclassifiable cases must be kept to an absolute minimum; all biological information is organized around the concept of species and species names. One idea that has not been seriously explored as a basis for a species concept for specialized phytophagous insects is that species should not rapidly revert back to their ancestral host plant, given the opportunity, whereas host races are almost certainly capable of such a reverse shift. The lack of perfect host fidelity in the apple race of *R. pomonella* (Feder, 1998) strongly suggests that if only apple race flies were introduced to an isolated experimental planting of hawthorns and apples, a hawthorn-infesting population would be 'reconstituted'. The experiment has not been carried out, but is feasible for both the host races and for more divergent taxa like the dogwood fly. For many organisms a species concept based on the idea of evolutionary irreversibility would not increase our understanding of nature, as it could not be readily tested, but for *Rhagoletis* and some other specialized phytophagous insects it may be worth investigating.

A final note is that, systematic decisions aside, the dogwood fly may be destined to have a short evolutionary existence. As a result of the dogwood anthracnose epidemic in North America, dogwood fly populations may soon face severely reduced fruit availability (Sherald *et al.*, 1996); the baseline data presented here may ultimately be of use in understanding the evolutionary genetics of population size reduction, or perhaps even extinction.

Acknowledgements

As with previous papers in this series, outstanding technical support was provided by S. Lyons. Additional invaluable support was provided by A. Amis, M. G. Anderson, Th. Anderson, C. Anelli, G. P. Berlocher, N. N. Berlocher, S. Harry Berlocher, G. L. Bush, J. L. Feder, J. A. Payne, R. Payne, W. S. Sheppard and D. C. Smith. Two anonymous reviewers improved the manuscript. NSF 89-06217 provided funding.

References

- AVISE, J. C. 1994. *Molecular Markers, Natural History, and Evolution*. Chapman & Hall, New York, NY.
- BERLOCHER, S. H. 1995. Population structure of the blueberry maggot, *Rhagoletis mendax*. *Heredity*, **74**, 542–555.
- BERLOCHER, S. H. 1998. Can sympatric speciation be proven from phylogenetic and biogeographic evidence?. In: Howard, D. J. and Berlocher, S. H. (eds) *Endless Forms: Species and Speciation*, pp. 99–113. Oxford University Press, New York, NY.
- 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. AND MCPHERON, B. A. 1996. Population structure of *Rhagoletis pomonella*, the apple maggot fly. *Heredity*, **77**, 83–99.
- 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.
- 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. 1994. Sympatric speciation in animals: new wine in old bottles. *Trends Ecol. Evol.*, **9**, 285–288.
- DE QUEIROZ, K. 1998. The general lineage concept of species, species criteria, and the process of speciation: A conceptual unification and terminological recommendations. In: Howard, D. J. and Berlocher, S. H. (eds) *Endless Forms: Species and Speciation*, pp. 57–78. Oxford University Press, New York, NY.
- DIEHL, S. R. AND BUSH, G. L. 1989. The role of habitat preference in adaptation and speciation. In: Otte, D. and Endler, J. A. (eds) *Speciation and its Consequences*, pp. 527–553. Sinauer Associates, Sunderland, MA.
- DOWLING, T. E. AND SECOR, C. L. 1997. The role of hybridization and introgression in the diversification of animals. *Ann. Rev. Ecol. Syst.*, **28**, 593–620.
- FEDER, J. L. 1998. The apple maggot fly, *Rhagoletis pomonella*: flies in the face of conventional wisdom about speciation? In: Howard, D. J. and Berlocher, S. H. (eds) *Endless Forms: Species and Speciation*, pp. 130–144. Oxford University Press, New York, NY.
- FEDER, J. L. AND BUSH, G. L. 1989. A field test of differential host usage between two sibling species of *Rhagoletis* fruit flies (Diptera: Tephritidae) and its consequences for sympatric models of speciation. *Evolution*, **43**, 1813–1819.
- 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 pre-mating barrier between sympatric races of the apple maggot fly. *Proc. Natl. Acad. Sci. U. S. A.*, **91**, 7990–7994.
- FEDER, J. L., ROETHELE, J. B., WLAZLO, B. AND BERLOCHER, S. H. 1997. Selective maintenance of allozyme differences among sympatric host races of the apple maggot fly. *Proc. Natl. Acad. Sci. U. S. A.*, **94**, 11417–11421.
- GRANT, P. R. 1993. Hybridization of Darwin's finches on Isla Daphne Major. *Phil. Trans. R. Soc. B.*, **340**, 127–139.
- HARRISON, R. G. 1998. Linking evolutionary pattern and process: The relevance of species concepts for the study of speciation. In: Howard, D. J. and Berlocher, S. H. (eds) *Endless Forms: Species and Speciation*, pp. 19–31. Oxford University Press, New York, NY.
- MALLET, J. 1995. A species definition for the Modern Synthesis. *Trends Ecol. Evol.*, **10**, 294–299.
- MAYR, E. 1963. *Animal Species and Evolution*. Belknap Press, Cambridge, MA.
- NEIGEL, J. E. 1997. A comparison of alternative strategies for estimating gene flow from genetic markers. *Ann. Rev. Ecol. Syst.*, **28**, 105–128.
- 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.
- SHERALD, J. L., STIDHAM, T. M., HADIDIAN, J. M. AND HOELDTKE, J. E. 1996. Progression of the dogwood anthracnose epidemic and the status of flowering dogwood in Catoctin Mountain Park. *Plant Dis.*, **80**, 310–312.
- SMITH, D. C. 1986. *Genetics and Reproductive Isolation of Rhagoletis Flies*. PhD. Thesis, University of Illinois, Urbana, IL.
- SMITH, D. C. 1988. Heritable divergence of *Rhagoletis pomonella* host races by seasonal asynchrony. *Nature*, **336**, 66–68.
- SMITH, J. J. AND BUSH, G. L. 1996. Phylogeny of the genus *Rhagoletis* (Diptera: Tephritidae) inferred from DNA sequences of mitochondrial cytochrome oxidase II. *Mol. Phylogen. Evol.*, **7**, 33–43.
- TEMPLETON, A. R. 1981. Mechanisms of speciation—a population genetic approach. *Ann. Rev. Ecol. Syst.*, **12**, 23–48.
- WEIR, B. S. 1990. *Genetic Data Analysis*. Sinauer Associates, Sunderland, MA.
- WRIGHT, S. 1931. Evolution in Mendelian populations. *Genetics*, **16**, 97–159.