*Entomol. exp. appl.* **57**: 191–200, 1990.



# Genetic differentiation and speciation in leaf-mining flies of the genus *Phytomyza*

Jürg E. Frey[1], Thierry Latscha & Doyle B. McKey[2]
*Department of Zoology, University of Basel, CH-4051 Basel, Switzerland;* [1] *Present address: Federal Research Station, Schloss, Lab. 4, CH-8820 Wädenswil, Switzerland;* [2] *Present address: Department of Biology, University of Miami, Coral Gables, FL 33124, USA*

Accepted: June 22, 1990

*Key words: Phytomyza chaerophylli*, Diptera, Agromyzidae, population genetics, F-statistics, gene flow, extinction, bottleneck, rapid speciation

## Abstract

Population structure is one of the most important factors determining speciation potential. We analyzed the spatial variation in allozyme frequencies of *Phytomyza chaerophylli* Kalt. (Diptera: Agromyzidae) using Wright's F-statistics. The degree of interpopulation variation is substantial ($F_{ST}$ = 0.092). Variation between local populations separated by only a few kilometers is as high as between populations as far as 1000 kilometers apart, indicating low interpopulation gene flow. The results are supported by ecological studies which revealed that host plant populations of *P. chaerophylli*, as well as of other, closely related *Phytomyza* species, are patchily distributed and have a short average life span. Populations of many *Phytomyza* species are thus mostly small and partly isolated and are prone to frequent extinctions and bottlenecks. Such conditions greatly increase the chance of rare mutations being fixed in new populations which may eventually lead to speciation. It is argued that the ecology and population structure of *P. chaerophylli* and congeners favor rapid speciation.

## Introduction

Patterns and processes of speciation within a taxonomic group depend on the ecology and population structure of the organisms concerned. Differences in these traits may cause intergroup differences in speciation rates. For example, a parasitic species that is highly host specific may initiate speciation if it shifts to a new host (Bush, 1969, 1975a, b; Zwölfer & Bush, 1984). This process may be greatly accelerated if the species lives in small demes with limited gene flow and frequent founder events (Wright, 1931, 1940; Mayr, 1970; Bush, 1975a; Templeton, 1981). In general, interactions of ecological and genetic factors can promote reproductive isolation between demes or effectively prevent its occurrence. Information on both types of factors is important to understand the potential for speciation. Leaf-mining flies of the family Agromyzidae seem to be in a period of fast adaptive radiation on several plant families (Zwölfer, 1978). To identify some of the main factors promoting rapid speciation in this family, we studied the population structure and ecological parameters of *Phytomyza chaerophylli*, which we present here in com-

parison with available data from related species.

The Agromyzidae is one of the phylogenetically youngest families of Diptera, consisting of a relatively small number of typically large genera (Zwölfer, 1978). The genus *Phytomyza* consists of 273 different species in Europe indicating a high speciation over extinction ratio (Spencer, 1976; Zwölfer, 1978). *P. chaerophylli* is an oligophagous leaf-miner feeding on several host plants of the family Umbelliferae, in northern Switzerland mostly on *Anthriscus sylvestris* Hoffm. and *Chaerophyllum temulum* L. Its distribution covers all of Europe and the western parts of Asia with the center apparently in Central Europe. The host plant species are closely related and show little variation in secondary compounds (Frey, 1986). Their distribution is distinctly patchy. Different host plant species occur in similar habitats, although they are very rarely found in sympatry. Many host plant populations of these flies have a short life span. Thus, bottlenecks, extinctions and recolonizations, sometimes by only few individuals, seem to be rather common in *P. chaerophylli*. Both sexes meet and mate on the larval host plants (Frey, 1986; Zwölfer & Bush, 1984), and the pupal period is spent in the soil below the host. Thus, many habitat elements of *P. chaerophylli* are locally aggregated. The available ecology and behavior data suggest that these flies usually do not migrate actively, and therefore gene flow between populations may be low (Frey, 1986 and in prep.). For means of comparison, we included a second species in our study, *Phytomyza aurei* Her. This species is closely related to *P. chaerophylli* and shows similar habitats and mating behaviors (Frey, 1986 and in prep.). It feeds exclusively on *Chaerophyllum aureum* L., a host plant which is sometimes also used by *P. chaerophylli*.

## Materials and methods

Larvae of *P. chaerophylli* were collected from eight sites around Basel and from five sites on a transect of 1000 km through Southwestern Europe (Fig. 1). In two populations (LES, LEF), the host

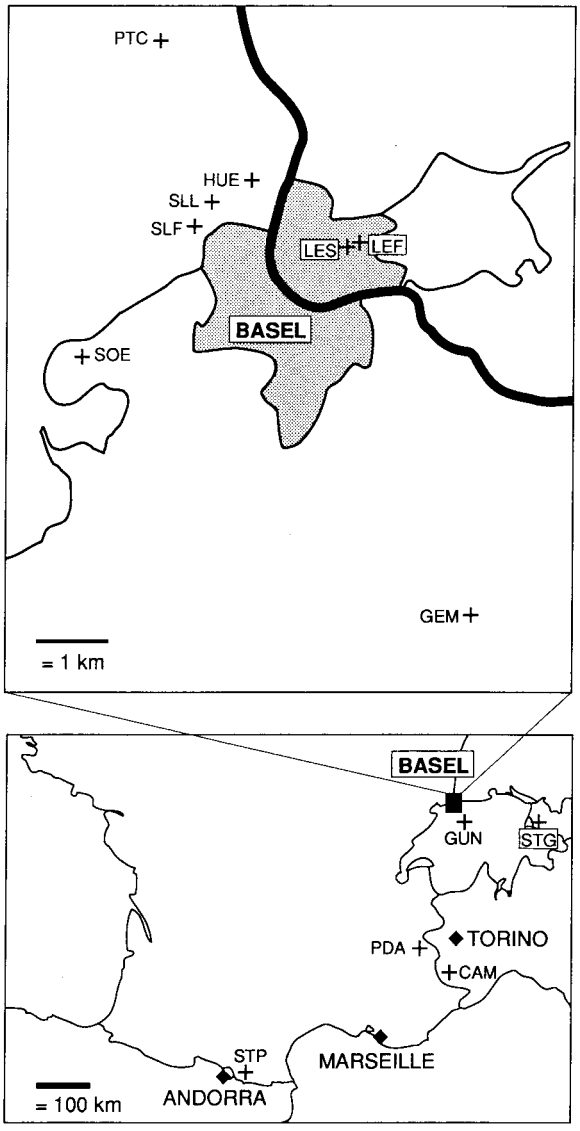


*Fig. 1.* Location of populations.

plant was exclusively *C. temulum*, in one population (SLL) both *A. sylvestris* and *C. temulum*, and in the remaining 10 populations exclusively *A. sylvestris*. *P. aurei* was collected from three populations with *Chaerophyllum aureum* as host plant. GEMCA is situated 500 m north-east of the GEM population, CHICA 2 km north-west of CHI in the Italian Alps, and VOLCA 25 km south-west of STP in the Spanish Pyrenees. All sites have been collected in 1985.

Collected larvae and/or pupae were stored in

*Table 1.* Genetic variation in 13 populations of *P. chaerophylli*. Indicated are allele frequencies, number of individuals, expected heterozygosity ($H_{exp}$), and the mean number of alleles per locus

| Locus | Allele | Populations | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | LES | LEF | SLL | SLF | PTC | HUE | SOE | GEM | GUN | STG | CAM | PDA | STP | Mean |
| *Idh* | N | 49 | 30 | 60 | 20 | 20 | 19 | 20 | 27 | 23 | 29 | 12 | 20 | 35 | |
| | 97 | 0.051 | 0.017 | 0.000 | 0.000 | 0.000 | 0.263 | 0.000 | 0.056 | 0.043 | 0.069 | 0.083 | 0.050 | 0.086 | 0.055 |
| | 100 | 0.949 | 0.983 | 1.000 | 1.000 | 1.000 | 0.737 | 1.000 | 0.944 | 0.957 | 0.931 | 0.917 | 0.950 | 0.914 | 0.945 |
| | H (exp) | 0.097 | 0.033 | 0.000 | 0.000 | 0.000 | 0.388 | 0.000 | 0.105 | 0.083 | 0.128 | 0.153 | 0.095 | 0.157 | 0.095 |
| *G6pdh* | N | 59 | 30 | 60 | 20 | 20 | 27 | 20 | 35 | 26 | 28 | 15 | 20 | 40 | |
| | 100 | 1.000 | 1.000 | 0.967 | 1.000 | 0.975 | 0.981 | 1.000 | 0.958 | 0.962 | 0.982 | 1.000 | 1.000 | 0.950 | 0.983 |
| | 104 | 0.000 | 0.000 | 0.033 | 0.000 | 0.025 | 0.019 | 0.000 | 0.042 | 0.038 | 0.018 | 0.000 | 0.000 | 0.050 | 0.017 |
| | H (exp) | 0.000 | 0.000 | 0.064 | 0.000 | 0.049 | 0.036 | 0.000 | 0.080 | 0.074 | 0.035 | 0.000 | 0.000 | 0.095 | 0.033 |
| *6Pgdh* | N | 57 | 30 | 60 | 20 | 20 | 28 | 20 | 37 | 27 | 28 | 12 | 20 | 40 | |
| | 100 | 0.955 | 0.983 | 0.692 | 0.950 | 0.975 | 0.857 | 0.925 | 0.908 | 1.000 | 0.786 | 1.000 | 0.950 | 1.000 | 0.922 |
| | 104 | 0.045 | 0.017 | 0.308 | 0.050 | 0.025 | 0.143 | 0.075 | 0.092 | 0.000 | 0.214 | 0.000 | 0.050 | 0.000 | 0.078 |
| | H (exp) | 0.085 | 0.033 | 0.427 | 0.095 | 0.049 | 0.245 | 0.139 | 0.167 | 0.000 | 0.337 | 0.000 | 0.095 | 0.000 | 0.129 |
| *Me* | N | 60 | 30 | 60 | 20 | 20 | 28 | 20 | 38 | 27 | 29 | 13 | 20 | 40 | |
| | 95 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.077 | 0.000 | 0.000 | 0.006 |
| | 98 | 0.000 | 0.000 | 0.008 | 0.000 | 0.000 | 0.036 | 0.000 | 0.092 | 0.130 | 0.000 | 0.000 | 0.050 | 0.050 | 0.028 |
| | 100 | 0.983 | 0.967 | 0.992 | 0.950 | 1.000 | 0.929 | 1.000 | 0.763 | 0.759 | 1.000 | 0.923 | 0.950 | 0.950 | 0.936 |
| | 102 | 0.017 | 0.033 | 0.000 | 0.050 | 0.000 | 0.036 | 0.000 | 0.145 | 0.111 | 0.000 | 0.000 | 0.000 | 0.000 | 0.030 |
| | H (exp) | 0.033 | 0.064 | 0.017 | 0.095 | 0.000 | 0.133 | 0.000 | 0.361 | 0.366 | 0.000 | 0.142 | 0.095 | 0.095 | 0.108 |
| *Pgi* | N | 60 | 30 | 60 | 20 | 20 | 28 | 20 | 40 | 28 | 30 | 20 | 20 | 40 | |
| | 89 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.150 | 0.000 | 0.000 | 0.012 |
| | 92 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.150 | 0.000 | 0.000 | 0.012 |
| | 98 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.025 | 0.036 | 0.033 | 0.000 | 0.000 | 0.000 | 0.007 |
| | 100 | 1.000 | 1.000 | 0.967 | 1.000 | 0.975 | 1.000 | 1.000 | 0.938 | 0.875 | 0.967 | 0.650 | 1.000 | 0.913 | 0.945 |
| | 104 | 0.000 | 0.000 | 0.033 | 0.000 | 0.025 | 0.000 | 0.000 | 0.038 | 0.089 | 0.000 | 0.050 | 0.000 | 0.087 | 0.025 |
| | H (exp) | 0.000 | 0.000 | 0.064 | 0.000 | 0.049 | 0.000 | 0.000 | 0.117 | 0.219 | 0.064 | 0.455 | 0.000 | 0.160 | 0.087 |
| *Pgm* | N | 38 | 18 | 60 | 10 | 20 | 10 | 20 | 29 | 19 | 29 | 6 | 20 | 38 | |
| | 92 | 0.000 | 0.056 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.167 | 0.000 | 0.000 | 0.000 | 0.000 | 0.118 | 0.026 |
| | 95 | 0.013 | 0.000 | 0.058 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.025 | 0.000 | 0.007 |
| | 98 | 0.000 | 0.000 | 0.017 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.001 |
| | 100 | 0.487 | 0.389 | 0.625 | 0.850 | 0.700 | 0.800 | 0.850 | 0.717 | 0.711 | 0.690 | 0.833 | 0.800 | 0.776 | 0.710 |
| | 102 | 0.211 | 0.222 | 0.092 | 0.150 | 0.100 | 0.100 | 0.100 | 0.050 | 0.132 | 0.000 | 0.000 | 0.050 | 0.026 | 0.095 |
| | 104 | 0.289 | 0.333 | 0.208 | 0.000 | 0.200 | 0.100 | 0.050 | 0.067 | 0.158 | 0.310 | 0.167 | 0.125 | 0.079 | 0.161 |
| | H (exp) | 0.500 | 0.475 | 0.469 | 0.255 | 0.420 | 0.320 | 0.255 | 0.406 | 0.411 | 0.428 | 0.278 | 0.320 | 0.347 | 0.376 |
| *Nadh-1* | N | 38 | 10 | 49 | 10 | 20 | 10 | 20 | 26 | 10 | 20 | 15 | 20 | 40 | |
| | 97 | 0.064 | 0.000 | 0.020 | 0.000 | 0.000 | 0.000 | 0.000 | 0.038 | 0.000 | 0.000 | 0.067 | 0.000 | 0.000 | 0.015 |
| | 100 | 0.936 | 1.000 | 0.980 | 1.000 | 1.000 | 1.000 | 1.000 | 0.962 | 1.000 | 1.000 | 0.533 | 1.000 | 1.000 | 0.955 |
| | 102 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.400 | 0.000 | 0.000 | 0.031 |
| | H (exp) | 0.120 | 0.000 | 0.040 | 0.000 | 0.000 | 0.000 | 0.000 | 0.074 | 0.000 | 0.000 | 0.498 | 0.000 | 0.000 | 0.056 |
| *Lap* | N | 41 | 18 | 48 | 10 | 20 | 19 | 20 | 30 | 12 | 29 | 13 | 20 | 20 | |
| | 96 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.077 | 0.000 | 0.000 | 0.006 |
| | 98 | 0.573 | 0.278 | 0.208 | 0.700 | 0.475 | 0.579 | 0.375 | 0.433 | 0.250 | 0.345 | 0.308 | 0.225 | 0.225 | 0.383 |
| | 100 | 0.427 | 0.722 | 0.792 | 0.300 | 0.525 | 0.421 | 0.625 | 0.567 | 0.750 | 0.655 | 0.615 | 0.775 | 0.775 | 0.611 |
| | H (exp) | 0.489 | 0.401 | 0.330 | 0.420 | 0.499 | 0.488 | 0.469 | 0.491 | 0.375 | 0.452 | 0.473 | 0.349 | 0.349 | 0.430 |
| *Est-1* | N | 39 | 29 | 55 | 18 | 10 | 19 | 10 | 28 | 15 | 17 | 11 | 19 | 34 | |
| | 100 | 0.641 | 0.879 | 0.891 | 0.750 | 0.750 | 0.421 | 0.650 | 0.821 | 0.800 | 0.706 | 0.091 | 0.553 | 0.735 | 0.668 |
| | 105 | 0.000 | 0.000 | 0.009 | 0.083 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.029 | 0.000 | 0.000 | 0.000 | 0.009 |
| | 107 | 0.064 | 0.000 | 0.036 | 0.000 | 0.000 | 0.000 | 0.350 | 0.161 | 0.000 | 0.000 | 0.545 | 0.053 | 0.074 | 0.099 |
| | 110 | 0.295 | 0.121 | 0.064 | 0.167 | 0.250 | 0.579 | 0.000 | 0.018 | 0.200 | 0.265 | 0.364 | 0.395 | 0.191 | 0.224 |
| | H (exp) | 0.460 | 0.212 | 0.194 | 0.375 | 0.375 | 0.488 | 0.455 | 0.293 | 0.320 | 0.415 | 0.496 | 0.494 | 0.398 | 0.382 |

*Table 1.* continued

| Locus | Allele | Populations | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 LES | 2 LEF | 3 SLL | 4 SLF | 5 PTC | 6 HUE | 7 SOE | 8 GEM | 9 GUN | 10 STG | 11 CAM | 12 PDA | 13 STP | Mean |
| *α-Gpdh* | N | 10 | 0 | 30 | 0 | 10 | 10 | 0 | 0 | 0 | 10 | 10 | 20 | 20 | |
| | 100 | 1.000 | 0.000 | 1.000 | 0.000 | 1.000 | 1.000 | 0.000 | 0.000 | 0.000 | 1.000 | 1.000 | 1.000 | 1.000 | |
| | 102 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | |
| | H (exp) | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | |
| *Nadh-2* | N | 25 | 5 | 40 | 0 | 20 | 7 | 20 | 8 | 15 | 25 | 13 | 15 | 15 | |
| | 98 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.040 | 0.000 | 0.000 | 0.000 | |
| | 100 | 1.000 | 1.000 | 1.000 | 0.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.960 | 1.000 | 1.000 | 1.000 | |
| | H (exp) | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.077 | 0.000 | 0.000 | 0.000 | 0.006 |
| Mean | H (exp) | 0.162 | 0.111 | 0.146 | 0.113 | 0.131 | 0.191 | 0.120 | 0.190 | 0.168 | 0.176 | 0.227 | 0.132 | 0.145 | 0.155 |
| Mean no. of alleles | | 1.910 | 1.800 | 2.182 | 1.667 | 1.636 | 1.818 | 1.500 | 2.400 | 2.000 | 1.818 | 2.091 | 1.818 | 1.910 | 1.888 |

liquid nitrogen until they were used for starch gel electrophoresis. The methods used for electrophoresis were modified from Brewer (1970), Ayala *et al.* (1972) and Menken (1980) and are described in detail in Frey (1986). We screened 24 enzyme systems, 11 of which proved to be useful for routine assays (locus abbreviation in parentheses): Isocitrate dehydrogenase-NADP (*Idh*), Glucose-6-phosphate dehydrogenase (*G6pdh*), 6-Phosphogluconate dehydrogenase (*6Pgdh*), Phosphoglucose isomerase-2 (*Pgi*), Phosphoglucomutase (*Pgm*), Malic enzyme (*Me*), Nadh dehydrogenase-1 (*Nadh-1*), Nadh dehydrogenase-2 (*Nadh-2*), α-Glycerophosphate dehydrogenase-2 (*α-Gpdh*), Leucine aminopeptidase (*Lap*), and Esterase-1 (*Est-1*). Single larvae or pupae were ground in 18 μl distilled water and pipetted onto filter paper (Whatman No. 6, 5 × 11 mm). 50 to 70 individuals were put side by side in a starch gel (15 × 30 cm, 12%) and run for 5 to 8 h. Samples from various sites were occasionally run side by side to standardise allelic nomenclature. Alleles were coded by their mobility relative to the allele that shows the highest mean frequency at this locus in the total data set (allele 100). Loci were considered monomorphic if the frequency of the most common allele was higher than 99%. Samples of *P. aurei* were run together with *P. chaerophylli* in order to establish the degree of interspecific genetic differentiation between these two *Phytomyza* species.

Gene frequency variation was analyzed by means of Nei's genetic distance (Nei, 1972) and UPGMA clustering method (Sneath & Sokal, 1973). The structuring of the populations was analyzed using F-statistics as described in Weir and Cockerham (1984). Allele frequency heterogeneity was tested applying G-contingency statistics on the allele numbers (Sokal and Rohlf, 1981). Significant deviations from Hardy-Weinberg genotypic expectations were determined by G-test using the William's correction for low sample sizes. Standard errors of the means were calculated according to the Jackknife procedure (Sokal & Rohlf, 1981).

## Results

Of 11 loci, two were monomorphic (>99% frequency), *α-Gpdh* and *Nadh-2* (Table 1). Of the polymorphic loci, two groups can be distinguished:

1) Moderate allelic variation between populations, with the most common allele being the same in all populations (*Idh*, *G6pdh*, *6Pgdh*, *Me*, *Pgi*, and *Nadh-1*).
2) Considerable allelic variation between populations, with allele 100 being not the most common in all populations (*Pgm*, *Lap*, and *Est-1*).

Several populations show unique allele frequency patterns at one or more loci. The popula-

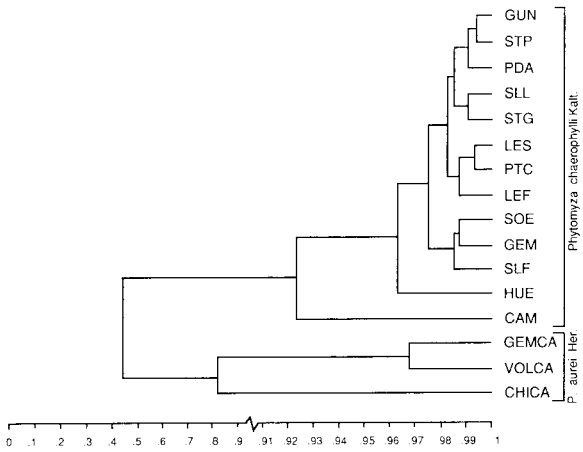


*Fig. 2.* Dendrogram of genetic identities between 13 populations of *P. chaerophylli* and three populations of *P. aurei*, based on nine polymorphic loci.

tion CAM, for example, has five unique alleles and the frequency of allele 100 is distinctly different from the other populations at *Pgi*, *Nadh-1*, and *Est-1*. Similar deviations were also found in the HUE population at *Idh* and *Est-1*. Allele frequencies at the Pgm locus seem to be host specific. In populations on *C. temulum* (LES, LEF), the frequency of allele 100 is low ($p = 0.438 \pm 0.049$), whereas in the populations on *A. sylvestris* (GEM, HUE, PTC, SLF, SOE, GUN, STG, CAM, PDA, STP) these frequencies are high throughout the sample area ($p = 0.773 \pm 0.020$). In the SLL population, where both host plants occur side by side, but only *C. temulum* could be sampled, the frequency of allele 100 is intermediate ($p = 0.625$).

Mean Nei's genetic identity and genetic distance between all pairs of populations are $I = 0.967$ and $D = 0.034$. The genetic identify I ranges from 0.994 (between GUN and STP) to 0.885 (between SLL and CAM). From these data, we constructed a dendrogram using the UPGMA method (Fig. 2). The two species, *P. chaerophylli* and *P. aurei*, form two distinct clusters. Within the *P. chaerophylli* cluster, two populations are separated from the rest (HUE and CAM) and show distinctly lower values of I than are measured between the other populations.

There is no obvious relationship between genetic identity and geographic distance between the populations, e.g. GUN, situated north of the Swiss Alps, and STP, situated north of the French Pyrenees, are very similar genetically ($I = 0.994$, Fig. 2) although they are separated by a distance of 720 km. In contrast, HUE and SLL, which are separated geographically by only 900 m, are genetically less similar ($I = 0.938$).

The mean $F_{ST}$, a measure of among population variance in allele frequencies, for all variable loci is $F_{ST} = 0.092$. The heterogeneity G-test shows

*Table 2.* Nei's genetic distance (above diagonal), identity (below diagonal) and average heterozygosity (on diagonal) between populations of *P. charophylli*

| | LES 1 | LEF 2 | SLL 3 | SLF 4 | PTC 5 | HUE 6 | SOE 7 | GEM 8 | GUN 9 | STG 10 | CAM 11 | PDA 12 | STP 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | **0.731** | 0.017 | 0.036 | 0.018 | 0.007 | 0.026 | 0.027 | 0.026 | 0.027 | 0.017 | 0.090 | 0.026 | 0.029 |
| 2 | 0.983 | **0.779** | 0.017 | 0.042 | 0.015 | 0.062 | 0.031 | 0.024 | 0.015 | 0.017 | 0.120 | 0.026 | 0.019 |
| 3 | 0.965 | 0.983 | **0.755** | 0.046 | 0.022 | 0.064 | 0.024 | 0.023 | 0.021 | 0.010 | 0.122 | 0.025 | 0.019 |
| 4 | 0.983 | 0.959 | 0.955 | **0.794** | 0.011 | 0.028 | 0.023 | 0.019 | 0.033 | 0.028 | 0.106 | 0.035 | 0.033 |
| 5 | 0.993 | 0.985 | 0.978 | 0.989 | **0.775** | 0.026 | 0.016 | 0.014 | 0.013 | 0.008 | 0.088 | 0.013 | 0.012 |
| 6 | 0.975 | 0.940 | 0.938 | 0.972 | 0.974 | **0.716** | 0.045 | 0.047 | 0.046 | 0.028 | 0.085 | 0.025 | 0.040 |
| 7 | 0.973 | 0.969 | 0.976 | 0.978 | 0.984 | 0.956 | **0.788** | 0.012 | 0.022 | 0.021 | 0.068 | 0.019 | 0.014 |
| 8 | 0.975 | 0.977 | 0.978 | 0.981 | 0.986 | 0.954 | 0.988 | **0.711** | 0.012 | 0.021 | 0.099 | 0.026 | 0.013 |
| 9 | 0.973 | 0.985 | 0.980 | 0.968 | 0.987 | 0.955 | 0.978 | 0.988 | **0.734** | 0.017 | 0.097 | 0.012 | 0.006 |
| 10 | 0.983 | 0.983 | 0.990 | 0.972 | 0.992 | 0.972 | 0.980 | 0.980 | 0.983 | **0.739** | 0.092 | 0.011 | 0.014 |
| 11 | 0.914 | 0.887 | 0.885 | 0.900 | 0.916 | 0.918 | 0.934 | 0.906 | 0.908 | 0.912 | **0.660** | 0.067 | 0.081 |
| 12 | 0.975 | 0.974 | 0.976 | 0.966 | 0.987 | 0.975 | 0.981 | 0.975 | 0.988 | 0.989 | 0.935 | **0.772** | 0.007 |
| 13 | 0.971 | 0.981 | 0.981 | 0.968 | 0.988 | 0.960 | 0.986 | 0.987 | 0.994 | 0.986 | 0.923 | 0.993 | **0.759** |

*Table 3.* F-statistics (weighted) from all 13 populations, calculated by the method described in Weir and Cockerham (1984). N = number of individuals; $p_i$ = mean frequency of allele 100; G = G-test for heterogeneity. SE = Standard error

| Locus | N | $p_i$ | $F_{is}$ | $F_{it}$ | $F_{st}$ | G | df | P |
|---|---|---|---|---|---|---|---|---|
| *Idh* | 364 | 0.945 | 0.529 | 0.553 | 0.050 | 48 | 12 | *** |
| *G6pdh* | 401 | 0.983 | 0.109 | 0.108 | 0.000 | 19 | 12 | n.s. |
| *6Pgdh* | 399 | 0.922 | 0.084 | 0.185 | 0.110 | 98 | 12 | *** |
| *Me* | 404 | 0.936 | 0.844 | 0.855 | 0.068 | 112 | 36 | *** |
| *Pgi* | 415 | 0.945 | 0.500 | 0.539 | 0.078 | 134 | 48 | *** |
| *Pgm* | 318 | 0.710 | 0.433 | 0.459 | 0.046 | 174 | 60 | *** |
| *Nadh-1* | 276 | 0.955 | 0.758 | 0.818 | 0.246 | 97 | 24 | *** |
| *Lap* | 300 | 0.611 | 0.038 | 0.113 | 0.078 | 68 | 24 | *** |
| *Est-1* | 304 | 0.668 | 0.309 | 0.409 | 0.145 | 198 | 36 | *** |
| Mean | | 0.853 | 0.400 | 0.449 | 0.091 | 948 | 264 | *** |
| SE | | 0.048 | 0.097 | 0.093 | 0.024 | | | |

P: *** < 0.001.
df: degrees of freedom.

significant substructuring of the total population at all but the *G6pdh* locus (Table 3). $F_{ST}$ is strongly affected by including one population, CAM, which is considerably different genetically from the others (I = 0.913; see Fig. 2). There is strong evidence that geographic isolation may have been the main factor accounting for differentiation in this specific case. CAM, and also CHICA, the extremely differentiated population of the second species, *P. aurei* (I = 0.816; see Fig. 2), were collected in the Val Grana, Italy. This small valley is U-shaped and is surrounded by high mountains which could provide a strong geographic barrier to fly dispersion. After the colonization of this valley, the populations may have been genetically undisturbed by immigrants for a long period of time. Also, since the host plants are in high abundance, the populations may have built up large sizes, which could, at least partly, explain the high heterozygosity as well as the relatively large number of unique alleles found in this population (see Table 1) (Slatkin, 1977). The HUE population of *P. chaerophylli* is also genetically distinct from the others (I = 0.964; see Fig. 2) although it is only 2 km distant from at least four other populations without any obvious geographic barriers. The data are insufficient to pinpoint a specific cause for this differentiation (e.g., recent founder event, selection, etc.). Excluding these two populations, the value of $F_{ST}$ decreases to $F_{ST} = 0.061$. In a recent literature survey of 27 species of winged insects of the orders Diptera, Lepidoptera and Coleoptera, $F_{ST}$ averaged to $0.054 \pm 0.008$ with a range from 0.003 to 0.156 (Table 6 in McCauley & Eanes, 1987). According to these values and considering the relatively small interpopulation distances, *P. chaerophylli* shows a rather high degree of between population differentiation.

The degree of differentiation between populations of *P. chaerophylli* in the same region is as high as between populations from locations as far as 1000 km apart. The eight populations of the Basel region (Fig. 1) show a differentiation of $F_{ST} = 0.084$ as compared to the value of $F_{ST} = 0.080$ for the five remaining populations distributed from Austria through Spain.

$F_{IS}$, measuring the inbreeding coefficient due to nonrandom mating within subpopulations, is high ($F_{IS} = 0.327$) indicating significant deviations from Hardy-Weinberg expectations in the form of heterozygote deficiencies. $F_{IT}$, a measure of deviation from random mating within the total population, is also high ($F_{IT} = 0.389$) as can be expected with $F_{IS}$ and $F_{ST}$ both being positive and large. The log likelyhood G-test for deviation of genotype frequencies from Hardy-Weinberg equilibrium for all polymorphic loci and all populations indicates a highly significant heterozygote deficit ($G = 805$, $P < 0.001$). The deficit was found in ten of the 13 populations (LES, LEF, SLL, GEM, GUN, CAM, STP: $P < 0.001$; HUE: $P < 0.01$; STG, PDA: $P < 0.05$) and at five of the nine loci analyzed (*Idh*, *Me*, *Pgm*, *Est-1*: $P < 0.001$; *Nadh-1*: $P < 0.05$).

## Discussion

Population differentiation in *P. chaerophylli* can be explained to some extent by the biology of this species. From ecological studies we know that the average life time of a host plant population is

rather short. Over 30% of the populations studied were eliminated within a period of three years (Frey, 1986). This is mostly due to anthropogenic factors, e.g., farming, but also to natural changes in habitats, e.g., succession to a different vegetation stage. Therefore, extinctions of local populations are frequent in this species. Furthermore, because even very small, isolated host plant populations often were found to be infested, colonization of newly founded host plant populations, probably by very few individuals, seems also to be common.

Interpopulation differentiation is not dependent on the geographic distance between populations. The similar degree of differentiation over small and large areas may be explained by low levels of gene flow over most of the distribution area of *P. chaerophylli* which decreases gene frequency differences caused by gene drift and/or microhabitat selection. Agromyzidae have been found in the aerial plankton as high as 3300 m above the pacific ocean (Gressit *et al.*, 1961). Thus, wind drift is a likely factor for long range gene flow in this fly family. However, no data are available as for the amount of gene flow provided by wind drift.

The host-specific gene frequency differences found at the *Pgm* locus are not likely to be caused by sampling error. Although the two populations with *C. temulum* as host plant (LES, LEF) are separated by a distance of only 300 m, they are still genetically different ($F_{ST} = 0.054$, $P < 0.01$). This difference is mainly due to *Est-1* ($G = 13$, $P < 0.001$) and to *Pgm* ($G = 6$, $P < 0.05$), the other loci showing no significant allele frequency differences. Furthermore, *Pgm* gene frequencies are intermediate in the SLL population where both host plants occur sympatrically. There are many factors that may cause gene frequency differences related to the host plant species, e.g., differences in predator and parasitoid fauna between host specific habitats or differences in phytochemistry between different host plants. Contents in secondary compounds can vary considerably both quantitatively and qualitatively between conspecific host plant populations in Umbelliferae (Gonnet, 1983) as well as in other plant families (Frohne & Jensen, 1985). Furthermore, mortality rates due to hymenopterous parasites, the single most significant mortality factor, can also vary dramatically between populations, ranging from 29% to 72% (Frey, 1986). These differences provide ample opportunity for habitat and/or host specific selection which may eventually increase the degree of isolation between populations of a species living on different host plants.

Heterozygote deficiency can be accounted for by inbreeding, by the pooling of equilibrium populations that differ in gene frequencies (Wahlund, 1928), by selection or by positive assortative mating. Furthermore, an apparent heterozygote deficit may be observed as a result of sex-linkage even if gene frequencies are in Hardy-Weinberg equilibrium (Ferguson, 1980), or as a result of null alleles.

The effect of sex-linkage at one locus should be apparent in all populations. However, only at *Est-1* (11 populations) and *Pgm* (10 populations), we found more than six populations with a significant heterozygote deficit, suggesting that sex-linkage is not involved in the nine polymorphic loci used in this study. We never observed null alleles. Heterozygote deficiency caused either by inbreeding due to nonrandom mating or by pooling of equilibrium populations is expected to affect all loci in the population (Hartl, 1980). However, a heterozygote deficit was only found in five out of nine loci, indicating that inbreeding as well as pooling effects may only marginally, if at all, affect heterozygote frequencies. Because of the narrow host range of *P. chaerophylli*, only moderate habitat differences in abiotic factors other than temperature occur between different populations (Frey, 1986). Large interpopulation differences in temperature may occur, but they are not correlated to population differences in gene frequencies at any of the studied loci. However, as outlined above, large interpopulation differences exist in mortality rates caused by the parasitoid fauna, providing a strong potential for selection which may partly be responsible for the heterozygote deficit observed. Positive assortative mating is another mechanism that may contribute to this heterozygote deficit. Males searching for

mates move frequently between leaves and plants, thus being exposed to predators such as spiders, and to fatal accidents, e.g., drowning under wet weather conditions (Frey, 1986). The sex ratio, which in laboratory experiments was found to be close to 1 : 1 at emergence, could therefore be changed dramatically towards a male deficit. Because males are capable of inseminating several females (Frey, 1986), this could cause assortative mating effects if population densities are low. Although we were not able to assess population sizes quantitatively in nature, the fact that host plant populations often are small, i.e., less than 100 plant individuals, and our experiences from field observations both suggest that most populations are rather small. Positive assortative mating and selection are therefore the most likely factors responsible for the heterozygote deficit found in this species.

Gene frequency data of *P. chaerophylli* support the hypothesis emerging from ecological studies that this species lives in small, partly isolated demes. We know that the area of the GEM population has been under extensive agricultural usage for at least the last 12 years, providing the host, *A. sylvestris*, in abundance and predictably. Mean heterozygosity and mean number of alleles are both relatively high indicating that GEM is a big population with a large number of individuals. The same seems to be true for the CAM population, another population where agricultural practice guarantee an abundant and predictable supply of host plant leaves, probably since decades. The other populations, in contrast, show a much lower mean number of alleles and a reduced mean heterozygosity suggesting that these populations underwent bottlenecks in their past. Because, as mentioned above, the average life span of a typical *P. chaerophylli* population seems to be rather low, these values indicate that the rate of population increase must be high after bottlenecks. This may explain why the heterozygosity values are not as low as could be expected after a bottleneck (Nei *et al.*, 1975; Chakraborty & Nei, 1976).

High degrees of local differentiation between conspecific populations were found in several other *Phytomyza* species. The relatively high genetic distances between the three *P. aurei* populations indicate low levels of gene flow between these populations (Fig. 2). A very similar situation was found between conspecific populations of *P. sphondylii* (Nei's genetic identity I = 0.827; Saner, 1986). In a thorough population structure analysis, Latscha (1986; Latscha *et al.*, 1987) found high $F_{ST}$ values for *P. angelicae* ($F_{ST}$ = 0.148) and for *P. laserpitii* ($F_{ST}$ = 0.088), again demonstrating a high degree of local differentiation between conspecific *Phytomyza* populations, often over small geographic distances.

Similar population structures as those described here for *P. chaerophylli* have also been reported for other plant feeding insects, e.g. milkweed beetle *Tetraopes tetraophthalmus* (Coleoptera: Cerambycidae) (McCauley & Eanes, 1987) and the parthenogenetic moth *Ectoedemia argyropeza* (Lepidoptera: Nepticulidae) (Menken & Wiebosch-Steeman, 1988). As in many *Phytomyza* species, gene flow seems to be provided in both species by passive dispersal mechanisms.

In oligophagous and monophagous parasites, the establishment of a new host race involves genetic changes in at least two components of the parasite's genome. First, the adult insect must acquire the ability to preferentially accept the new host for oviposition, and second, the larvae must be able to survive on the new host (Bush, 1975b). In some phytophagous insect species, host acceptance is monogenetically inherited (Gallun & Khush, 1980). Genetic change in host acceptance genes may thus cause a shift to a new host plant (Bush, 1975b). However, sibling species pairs in the genus *Phytomyza* usually occur on host plants with distinct phytochemical differences. For example, *P. angelicae* lives on *Angelica sylvestris* L., a host plant with high contents in linear and angular furanocoumarins and flavonols and low contents in flavonons whereas its closely related sibling species, *P. laserpitii*, lives on *Laserpitium latifolium* L. which shows the exact opposite pattern in contents of secondary compounds (Latscha, 1986). Such phytochemical differences between host plant species may be too great to allow for larval survival. Larvae of *P. angelicae* survive and develop on *Laserpitium latifolium* L.,

the host plant of its sibling species, *P. laserpitii*, but not vice versa (Latscha, 1986). Successful colonization under such conditions is not possible without additional changes in larval survival genes. Larvae of females with altered host preference are strongly selected for genes enabling them to survive on the new host. Thus, changes in two genes may eventually lead to successful colonization of new, formerly unaccessible host plants and, because individuals carrying both mutations meet and mate on this new host, to speciation.

Instantaneous shifts to extremely different host plants may be possible via chromosome rearrangements that affect both traits simultaneously. Chromosomal inversions are very common in the genus *Phytomyza* and can be used to distinguish between host races and incipient species (Block, 1969, 1974). Chromosomal mutations are favored by population structures as described for *P. chaerophylli* even without extrinsic barriers (Wright, 1941; White, 1978; Lande, 1979; Templeton, 1981), and their chances may be greatly increased by founder events (Mayr, 1970). In the evolution of the genus *Phytomyza*, rapid speciation via chromosomal mutations is therefore likely to be an important mechanism.

**Acknowledgements**

We thank S.M. Williams and S.B.J. Menken for their helpful comments on an earlier draft of this manuscript. This study was supported in part by the Swiss National Science Foundation and the 'Werenfels Fonds der Freiwilligen Akademischen Gesellschaft Basel'.

**Zusammenfassung**

*Genetische Differenzierung und Speziation bei blattminierenden Fliegen der Gattung* Phytomyza

Die Populationsstruktur ist einer der wichtigsten Faktoren, die das Speziationspotential einer Species bestimmen. Wir analysierten die räumliche Variation der Allozymfrequenzen von *Phytomyza chaerophylli* Kalt. (Diptera: Agromyzidae) nach den Methoden von Wrights F-Statistik. Der Grad der genetischen Variation zwischen den Populationen ist beträchtlich ($F_{ST}$ = 0.092). Die Variation zwischen lokalen, lediglich durch wenige Kilometer voneinander getrennten Populationen ist gleich gross wie zwischen Populationen, die bis zu 1000 Kilometer auseinander liegen. Eine solche Differenzierung deutet auf einen niedrigen Genfluss zwischen den Populationen. Die Ergebnisse der genetischen Analyse werden gestützt durch ökologische Studien, welche zeigten, dass Wirtspflanzenpopulationen von *P. chaerophylli* und anderen, nah verwandten *Phytomyza*-Arten eine inselartige Verbreitung aufweisen und eine verhältnismässig kurze mittlere Lebensdauer haben. Die Populationen von vielen *Phytomyza*-Arten sind daher meist klein und teilweise isoliert und unterliegen häufig Ausrottungen und genetischen Flaschenhälsen ('Bottlenecks'). Solche Bedingungen vergrössern beträchtlich die Chance, seltene Mutationen in neuen Populationen zu fixieren, was unter Umständen zu Speziation führen kann. Wir stellen fest, dass Ökologie und Populationsstruktur von *P. chaerophylli* und anderen Arten dieses Genus günstige Voraussetzungen für eine schnelle Speziation darstellen.

**References**

Ayala, F. J., J. R. Powell, M. L. Tracey, C. A. Mourao & S. Perez-Salas, 1972. Enzyme variability in the *Drosophila willistoni* group. IV. Genetic variation in natural populations of *Drosophila willistoni*. Genetics 70: 113–139.

Block, K., 1969. Chromosomal variation in Agromyzidae. I. *Phytomyza abdominalis* Zett. – two incipient species and their natural hybrid. Hereditas 62: 131–152.

Block, K., 1974. Chromosomal variation in Agromyzidae. III. *Cerodonta* (*Butomomyza*) *eucaricis* Nowakowski – two semispecies or sibling species? Hereditas 78: 125–140.

Brewer, G. J., 1970. An introduction to isozyme techniques. Academic Press, NY.

Bush, G. L., 1969. Sympatric host race formation and speciation in frugivorous flies of the genus *Rhagoletis* (Diptera, Tephritidae). Evolution 23: 237–251.

Bush, G. L., 1975a. Modes of animal speciation. Ann. Rev. Ecol. Syst. 6: 339–364.

Bush, G. L., 1975b. Sympatric speciation in phytophagous parasitic insects. pp. 187–204. In: Evolutionary Strategies

of Parasitic Insects and Mites (P. W. Price, ed.). Plenum Press, New York.

Chakraborty, R. & M. Nei, 1976. Bottleneck effects on average heterozygosity and genetic distance with the stepwise mutation model. Evolution 31: 347–356.

Ferguson, A., 1980. Biochemical systematics and evolution.- Blackie, London, 194 pp.

Frey, J. E., 1986. Biologie und Wirtspflanzenbeziehung der blattminierenden Fliegen *Phytomyza chaerophylli* Kalt. und *P. aurei* Her. (Diptera, Agromyzidae). Dissertation Universität Basel.

Frohne, D. & U. Jensen, 1985. Systematik des Pflanzenreiches. Third ed. Gustav Fischer Verlag, Stuttgart, 355 pp.

Gallun, R. L. & G. S. Khush, 1980. Genetic factors affecting expression and stability of resistance. In: F. G. Maxwell and P. R. Jennings (eds.) Breeding Plants Resistant to Insects. John Wiley & Sons, New York: pp. 64–85.

Gonnet, J.-F., 1983. Phytosociological and geographical variation of flavonoid glycosides in *Chaerophyllum aureum*. Phytochemistry 22: 1421–1423.

Gressit, J. L., R. E. Leech, J. Sedlacek & K. A. J. Wise, 1961. Trapping of air-borne insects in the Antarctic area (part 2). Pacific Insects 3: 559–562.

Hartl, D. L., 1980. Principles of population genetics. Sinauer Ass. Inc., Sunderland, Mass, 488 pp.

Lande, R., 1979. Effective deme sizes during long-term evolution estimated from rates of chromosomal rearrangement. Evolution 33: 234–251.

Latscha, T., 1986. Biologie und Wirtspflanzenbeziehungen blattminierender Agromyziden (Diptera) an Umbelliferen: Der '*Phytomyza-angelicae*'-Komplex, biochemische Taxonomie, Populationsstrukturen und wirtsspezifische Divergenz. Dissertation Universität Basel.

Latscha, T., J. E. Frey, P. Ruggle, M. Saner & D. B. McKey, 1987. Host plant relationships and speciation in leafmining Agromyzid flies on Umbelliferae. pp. 261–266. In: V. Labeyrie, G. Fabres, D. Lachaise (eds.), Insects-Plants. Dr. W. Junk Publ., Dordrecht, The Netherlands.

McCauley, D. E. & W. F. Eanes, 1987. Hierarchical population structure analysis of the milkweed beetle *Tetraopes tetraophthalmus* (Forster). Heredity 58: 193–203.

Mayr, E., 1970. Populations, Species and Evolution. Harvard Univ. Press, Cambridge, MA, 453 pp.

Menken, S. B. J., 1980. Allozyme polymorphism and the speciation process in Small Ermine Moths (Lepidoptera, Yponomeutidae). Diss. Univ. Leiden, The Netherlands.

Menken, S. B. J. & M. Wiebosch-Steeman, 1988. Clonal diversity, population structure, and dispersal in the parthenogenetic moth *Ectoedemia argyropeza*. Entomol. exp. appl. 49: 141–152.

Nei, M., 1972. Genetic distance between populations. Am. Nat. 106: 283–292.

Nei, M., 1976. Mathematical models of speciation and genetic distance. pp. 723–765. In: Population Genetics and Ecology, S. Karlin, E. Nevo (eds.). Academic Press.

Nei, M., T. Maruyama & R. Chakraborty, 1975. The bottleneck effect and genetic variability in populations. Evolution 29: 1–10.

Saner, M., 1986. Populationsbiologie und Wirtsbeziehungen der blattminierenden Agromyziden (Diptera) auf *Heracleum sphondylium* L. und *Pastinaca sativa* L. (Umbelliferae). Masters thesis, University of Basel, Switzerland.

Slatkin, M., 1977. Gene flow and genetic drift in a species subject to frequent local extinctions. Theor. Pop. Biol. 12: 253–262.

Sneath, P. H. A. & R. R. Sokal, 1973. Numerical taxonomy. San Francisco, Freeman, 573 pp.

Sokal, R. R. & F. J. Rohlf, 1981. Biometry. Freeman, San Francisco, 859 pp.

Spencer, K. A., 1976. The Agromyzidae (Diptera) of Fennoscandia and Denmark. F. Entomol. Scand. 5, I + II: 1–606.

Templeton, A. L., 1981. Mechanisms of evolution – a population genetic approach. Ann. Rev. Ecol. Syst. 12: 23–48.

Wahlund, S., 1928. Zusammensetzung von Populationen und Korrelationserscheinungen vom Standpunkt der Vererbungslehre aus betrachtet. Hereditas 11: 65–106.

Weir, B. S. & C. C. Cockerham, 1984. Estimating F-Statistics for the Analysis of Population Structure. Evolution 38: 1358–1370.

White, M. J. D., 1978. Modes of speciation. Freeman, San Francisco, 455 pp.

Wright, S., 1931. Evolution in mendelian populations. Genetics 16: 97–159.

Wright, S., 1940. Breeding structure of populations in relation to speciation. Am. Nat. 74: 232–248.

Wright, S., 1941. On the probability of fixation of reciprocal translocations. Amer. Natur. 75: 513–522.

Zwölfer, H., 1978. Mechanismen und Ergebnisse der Co-Evolution von phytophagen und entomophagen Insekten und höheren Pflanzen. Sonderbd. naturwiss. Ver., Hamburg, 2: 7–50.

Zwölfer, H. & G. L. Bush, 1984. Sympatrische und parapatrische Artbildung. Z. zool. Syst. Evolut.-forsch. 22, 3: 211–233.