

The role of geographic setting on the diversification process among *Tephritis conura* (Tephritidae) host races

T Diegisser, J Johannesen and A Seitz

Institut für Zoologie, Abteilung für Ökologie, Universität Mainz, Saarstrasse 21, D-55099 Mainz, Germany

We address the controversy over the processes causing divergence during speciation. Host races of the fruit fly *Tephritis conura* attack the thistles *Cirsium oleraceum* and *Cirsium heterophyllum*. By studying the genetic divergence of *T. conura* in areas where host plants are sympatric, parapatric and allopatric, we assessed the contribution of geography in driving host-race divergence. We also evaluated the relative importance of genetic drift and selection in the diversification process, by analysis of the geographic distribution of genetic variation. Host races were significantly diverged at five out of 13 polymorphic allozyme loci. Variance at two loci, *Hex* and *Pep D*, was almost exclusively attributable to host-plant affiliation in all geographic settings. However, *Hex* was significantly more differentiated between host races in sympatry/parapatry than in allopatry. This result

might be explained by selection against hybridisation or against incorrect host choice in contact areas. Linkage disequilibrium tests suggest the latter: gene flow in contact areas may occur from males of the host-race *C. heterophyllum* to females of the host-race *C. oleraceum*, whereas incorrect oviposition events were never observed. The distinctive patterns of genetic differentiation at the two highly differentiated loci implicate the action of selection (acting directly or on linked loci) rather than genetic drift. Despite their restricted interactions in sympatry, we conclude that host races are stable and that the major diversification process took place before species arrived in today's geographical settings.

Heredity (2006) **96**, 410–418. doi:10.1038/sj.hdy.6800821; published online 29 March 2006

Keywords: allopatry; allozymes; disruptive selection; divergent selection; genetic drift; sympatric speciation

Introduction

Allopatric speciation is traditionally assumed to be the predominant mode of speciation (Mayr, 1963), but the evidence for sympatric speciation has become increasingly strong in the last decade (for reviews see: Via, 2001; Dres and Mallet, 2002). The debate today is not whether sympatric speciation is possible; rather it is a question of its relative importance. The number of well-documented cases is still small (Dres and Mallet, 2002) and further case studies are required to evaluate the processes causing divergence (Jiggins and Mallet, 2000; Schluter, 2001; Via, 2001). In fact, these underlying mechanisms rather than the geographical modes *per se* are the point of controversy. In sympatric speciation models, gene flow becomes restricted due to disruptive selection (eg Kawecki, 1997; Dieckmann and Doebeli, 1999; Kondrashov and Kondrashov, 1999) whereas, in allopatric speciation models, reproductive isolation is a passive consequence of different selection regimes and/or genetic drift (Turelli *et al.*, 2001; Presgraves *et al.*, 2003). However, the geographical mode of speciation does not necessarily implicate a particular mechanism. Sympatric populations might also differentiate due to differential

gene flow (Stam, 1983). In phytophagous insects, for example, allochronic isolation due to differences in host-plant phenology can lead to restricted gene flow in sympatry even in the absence of selection (Butlin, 1990). Furthermore, speciation can be a combination of sympatric or allopatric conditions. For instance, sympatric host-race formation in *Rhagoletis pomonella* is likely to have been facilitated by genetic polymorphisms that evolved during an allopatric stage (Feder *et al.*, 2003). Allopatric speciation in turn might be completed in sympatry via reinforcement processes, that is selection promoting prezygotic isolation after secondary contact (eg Howard, 1993; Noor, 1999). During the speciation process, therefore, different geographic settings might often be involved, with several mechanisms contributing to differentiation.

In the present paper, we address the contribution geography plays in driving genetic divergence in the tephritid fly *Tephritis conura* and, based on geographic distributions, the relative importance of genetic drift and selection in the diversification process. *Tephritis conura* infest at least nine thistle species of the genus *Cirsium* (Cardueae) (Zwölfer, 1988; Romstöck-Völkl, 1997). These univoltine flies mate and oviposit exclusively on host plants in late spring, synchronised with flower head phenology (Romstöck and Arnold, 1987). Larvae feed on the seeds until pupating and adults emerge in July/August, hibernating in places which are not yet known. Seitz and Komma (1984) found that flies emerging from *Cirsium oleraceum* (L.) SCOP. (cabbage thistle) differed from flies emerging from *Cirsium heterophyllum* (L.) HILL.

Correspondence: T Diegisser, Institut für Zoologie, Abteilung für Ökologie, Universität Mainz, Saarstrasse 21, D-55099 Mainz, Germany. E-mail: thorsten.diegisser@uni-mainz.de

Received 13 June 2005; accepted 6 March 2006; published online 29 March 2006

(melancholy thistle) in frequencies at the allozyme locus hexokinase, indicating the existence of host races (we will henceforth refer to flies emerging from *C. oleraceum* as oleraceum flies and those from *C. heterophyllum* as heterophyllum flies). Mitochondrial DNA sequence analysis suggests recent diversification. Both host races share the most common mtDNA haplotype, which is central in a star-like genealogy (Diegisser et al, 2006).

The study area of Seitz and Komma (1984) was situated in the Bavarian Fichtelgebirge where *C. oleraceum* and *C. heterophyllum* occur parapatrically. Outside the Fichtelgebirge the host plants are largely allopatrically distributed, but there also exist areas of true sympatry, for example in southern Scandinavia (Figure 1).

The combination of a young system and multiple geographical distributions makes *T. conura* a promising target for studies to evaluate the contribution that geography makes in driving reproductive isolation. In this study, we analyse genetic divergence of oleraceum and heterophyllum flies from sympatric, parapatric and allopatric regions using allozyme electrophoresis. Basically, we expect to find one of three possible patterns, which differ in their implications for the speciation process. Firstly, host races might differ more in allopatry than in sympatry/parapatry. Differentiation among host

races in contact areas might decrease due to oviposition into the 'wrong' host plant and/or hybrid matings. Such a pattern would indicate that the original diversification among heterophyllum and oleraceum flies was a passive consequence of either genetic drift or selection in isolation but that diversification is unfinished. The reduced differentiation in contact areas (compared to that in allopatric regions) would provide information about the degree of ongoing gene flow and the stability of the system in sympatry.

Under the second scenario, differentiation among host races would be higher in contact areas than in allopatry. This pattern would be coherent with selection against hybridisation and/or oviposition into the alternative host. In an experiment, we forced oleraceum and heterophyllum females to oviposit into flower heads of *C. oleraceum*. The heterophyllum females produced significantly fewer progeny than the oleraceum flies, suggesting that selection against wrong host plant choice exists (T Diegisser, unpublished data). As a consequence, selection should strengthen traits, which reduce maladaptive events in contact zones, for example host choice and/or mating preferences. If allozyme loci are linked to such traits, they should exhibit higher differentiation in contact areas than in allopatric regions.

Finally, the third conceivable pattern is a uniform level of host-related differentiation in all geographic settings. In that case, the pattern could be interpreted in two ways. One possibility (scenario 3a) is that both the process of ongoing gene flow (scenario 1) and selection against maladaptive hybridisation/wrong host choice (scenario 2) occur in contact areas but they counterbalance each other, thereby producing differentiation similar to that in allopatric regions. Alternatively, gene flow might be so strongly restricted that allele frequencies are not significantly affected by hybridisation events (if these occur at all), hence suggesting geographically stable, reproductively isolated host races (scenario 3b).

Materials and methods

Host-plant distribution and sampling

C. oleraceum occurs throughout a large part of Europe (Figure 1), from eastern France to Russia. Its southern limit of distribution is from northern Italy to the northern Balkans; in the North it can be found up to southern Scandinavia and the West Siberian Lowlands. *C. heterophyllum* is distributed from Scotland, Scandinavia and the Baltic States to Siberia, occurring sympatrically with *C. oleraceum* in southern Sweden and parts of Russia. In addition, *C. heterophyllum* can be found in mountain ranges of Europe, for example Alps, Carpathians, Ore Mountains and the Bavarian Fichtelgebirge (Figure 1). In these regions, the host plants occur parapatrically; *C. heterophyllum* at high altitudes, *C. oleraceum* at lower altitudes.

In the present study, a total of 40 *T. conura* populations from all geographic host-plant settings were analysed (19 *C. heterophyllum* sites; 21 *C. oleraceum* sites). We sampled flies from southern Sweden, where host plants occur in true sympatry and often syntopically (Figure 1; site codes 9–13 = heterophyllum populations, 14–19 = oleraceum populations). Parapatric sites were collected in the Bavarian Fichtelgebirge (site codes 29–34 = oleraceum

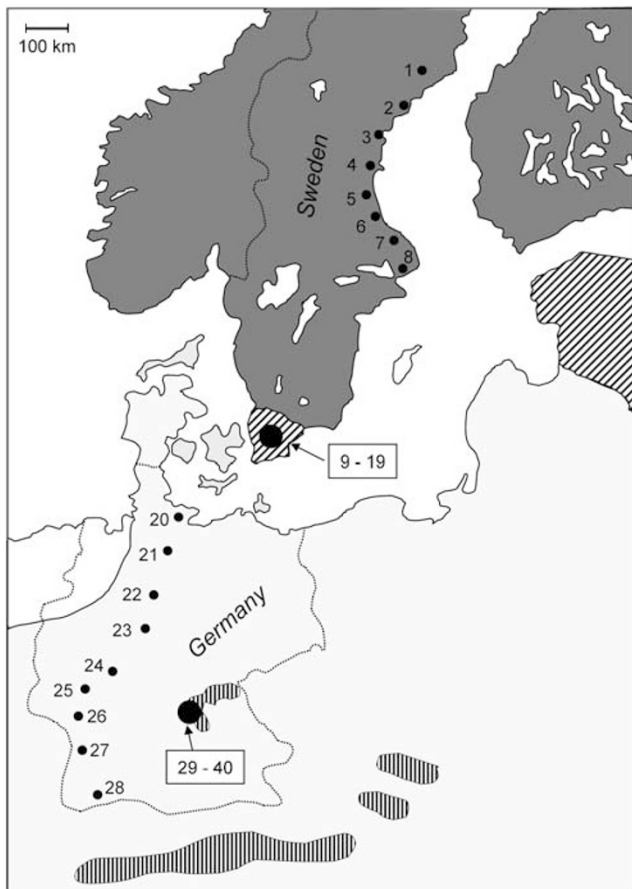


Figure 1 Sampling scheme and host-plant distribution. Numbers refer to sample sites (details given in Appendix A). Pale-grey shaded = allopatric *C. oleraceum* distribution; dark-grey shaded = allopatric *C. heterophyllum* distribution; diagonally hatched = sympatric region; vertically hatched = parapatric region.

populations, 35–40 = heterophyllum populations). From here, three sites were included from the area of contact where both host plants occurred syntopically (sites 32–34 and 35–37, respectively). Flies from allopatric regions were sampled in transects to allow testing for geographic variation. The *C. heterophyllum* transect (sites 1–8) was located in Middle/North Sweden, the *C. oleraceum* transect (sites 20–28) was laid through Germany (Figure 1). The minimum distance between parapatric/allopatric sites was 180 km, between sympatric/allopatric sites 250 km. Details about sample localities are given in Appendix A.

Sampling took place in summer 2002 and 2003. At each site, flower heads from several plants were collected and placed in darkened plastic bottles with a hole in the lid, which was covered with a glass bottle. Emerging flies were attracted by the light and trapped in the bottle. Adults were collected on a daily basis and stored at -80°C until genetic analyses were performed.

Allozyme electrophoresis

Allozyme analysis was conducted using cellulose acetate electrophoresis (Hebert and Beaton, 1989). Whenever possible we analysed 24 individuals per site (12 males and females, respectively), resulting in a total of 950 flies. Thirteen loci stained consistently in all individuals: aconitate hydratase (*Acon*, EC 4.2.1.3), arginine kinase (*Ark*, EC 2.7.3.3), fumarate hydratase (*Fum*, EC 4.2.1.2), glyceraldehyde-3-phosphate dehydrogenase (*G3pdh*, EC 1.2.1.12), glutamate-oxalacetate transferase (*Got*, EC 2.6.1.1), β -hydroxybutyrate dehydrogenase (*Hbdh*, EC 1.1.1.30), hexokinase (*Hex*, EC 2.7.1.1), isocitrate dehydrogenase (*Idh*, EC 1.1.1.42), 6-phosphogluconate dehydrogenase (*6Pgdh*, EC 1.1.1.44), peptidase A (Gly-Leu) (*PepA*, EC 3.4.11.11), peptidase D (Phe-Pro) (*PepD*, EC 3.4.13.9), phosphoglucomutase (*Pgm*, EC 5.4.2.2) and trehalase (*Tre*, EC 3.2.1.28). All enzymes were run at 250 V for 30 min using the buffer systems Tris-Citrate pH 8.2 (Richardson *et al*, 1986) for *G3pdh*, *Hex*, *Idh*, *6Pgdh*, *Tre* and Tris-Glycine pH 8.5 (Hebert and Beaton, 1989) for *Acon*, *Ark*, *Fum*, *Got*, *Hbdh*, *PepA*, *PepD* and *Pgm*. The most common allele at each locus was given the arbitrary score 100; the other alleles were named according to their relative electrophoretic mobility. As the most common allele at the *Hbdh* locus stained exactly at the point of application, we assigned the score 100 to the next most frequent allele.

Data analysis

Allele frequencies were calculated for each population with G-stat (Siegismund, 1993), and used to build a maximum likelihood phenogram with the PHYLIP (Felsenstein, 1993) subprogram CONTML (parameter settings 'global rearrangements' and 'randomise input order of species'). This phenogram visualises the genetic relationships among populations, thereby providing evidence for the pattern of differentiation: under scenario 1, we expect heterophyllum and oleraceum populations from contact areas to be less separated from each other than those of allopatric regions, while the opposite should apply to scenario 2. If populations in the phenogram cluster due to host-plant affiliation and unrelated to geography, this points to scenario 3. For quantification of population differentiation we applied

hierarchical F-statistics, using the method of Weir and Cockerham (1984) as implemented in Arlequin 2.0 (Schneider *et al*, 2000). The total interpopulation variance, F_{ST} , was partitioned into variance, which can be assigned to host affiliation (F_{HT}) and host-independent variance (F_{SH}). This quantifies to which extent the observed population structure is caused by host-race differentiation. The absolute (F_{HT}) and relative (F_{HT}/F_{ST}) degree of host-related differentiation was calculated for all populations and for each geographic setting separately. We used genetic identities (Nei, 1972) to estimate whether host-race differentiation is lower (higher) among heterophyllum and oleraceum populations in contact (sympatry/parapatry) than among heterophyllum and oleraceum populations that are geographically isolated from each other.

To exclude the possibility that geographic separation of populations may bias the genetic identity between allopatric *vs* parapatric/sympatric host-race populations, tests for isolation by distance (IBD) were performed in the two allopatric transects by comparing pairwise F_{ST} estimates between populations (Slatkin, 1993). Significant IBD within transects may indicate that the level of host-related differentiation depends on the geographic locality of a given population. In contrast, in the absence of IBD populations can be pooled within each geographic setting to create an average between-host-race population identity estimate. Significance for IBD was tested with a Mantel test implemented in Genepop (Raymond and Rousset, 2000).

Although the above analyses might allow conclusions about the likelihood of scenarios 1–3, they cannot discriminate between scenario 3a (ongoing gene flow counterbalanced by selection against hybridisation/wrong host-plant choice) and scenario 3b (cessation of gene flow independent of the geographic setting). To discriminate between 3a and b, we applied tests for linkage-disequilibria. If gene flow in contact areas is ongoing (ie hybridisation and/or oviposition into the wrong host plant), this should result in linkage disequilibria at those loci showing strong host-related differentiation. Linkage disequilibrium was tested with exact tests supplied in GDA (Lewis and Zaykin, 2001), pooling sympatric/parapatric populations of each host plant for stronger statistical support (only alleles with frequencies ≥ 0.05 were included).

Results

Nine of the 13 analysed loci were polymorphic at the 95% level (*Acon*, *Got*, *Hbdh*, *Hex*, *6Pgdh*, *PepA*, *PepD*, *Pgm*, *Tre*); allele frequencies tables are available on request). Five loci (*Acon*, *Hex*, *6Pgdh*, *PepA*, *PepD*) revealed (highly) significant allele frequency differences among heterophyllum and oleraceum populations (Table 1). Particularly, hexokinase (*Hex*) and peptidase D (*PepD*) showed extremely divergent allele frequencies: the most common heterophyllum alleles *Hex*₁₀₀ and *PepD*₁₀₀ were rarely found in oleraceum populations where the alleles *Hex*₁₀₇ and *PepD*₉₂ were predominant. By contrast, the allele frequency differences among the different geographic settings within the *same* host were relatively small. Only two alleles, and in heterophyllum flies only, differed significantly in frequency among geographic

Table 1 Significant allele frequency differences between oleraceum (olera) and heterophyllum (hetero) populations over all sites and for each geographic setting (*t*-tests)

Allele	All populations (<i>df</i> = 38)		Allopatry (<i>df</i> = 15)		Sympatry (<i>df</i> = 9)		Parapatry (<i>df</i> = 10)					
	Olera	Hetero	Olera	Hetero	Olera	Hetero	Olera	Hetero				
<i>Acon_125</i>	0.06	*	0.01	0.05	0.01	0.06	0.00	0.07	0.01			
<i>Hex_100</i>	0.19	***	0.93	0.23	***	0.91	0.14	***	0.92	0.17	***	0.97
<i>Hex_107</i>	0.80	***	0.07	0.76	***	0.08	0.82	***	0.08	0.83	***	0.03
<i>6Pgdh_100</i>	0.90	*	0.97	0.92		0.95	0.84		0.97	0.93		1.00
<i>6Pgdh_110</i>	0.07	***	0.01	0.06		0.01	0.11		0.01	0.06	*	0.00
<i>PepA_85</i>	0.24	*	0.13	0.22		0.09	0.29		0.11	0.24		0.20
<i>PepA_100</i>	0.52	*	0.65	0.50	*	0.73	0.50		0.63	0.55		0.54
<i>PepD_80</i>	0.07	***	0.01	0.06		0.01	0.07	**	0.00	0.09	***	0.00
<i>PepD_92</i>	0.88	***	0.06	0.89	***	0.04	0.91	***	0.10	0.83	***	0.05
<i>PepD_100</i>	0.03	***	0.84	0.02	***	0.87	0.02	***	0.73	0.04	***	0.89

P* < 0.05; *P* < 0.01; ****P* < 0.001 (after Bonferroni correction).

settings (*6Pgdh_100* and *PepA_100*; Anova: *F* = 7.87, *P* = 0.004 and *F* = 5.82, *P* = 0.013, respectively).

The degree of host-related allele frequency difference was quantified by hierarchical *F*-statistics. Over all loci, 91.97% of the interpopulation variance was assigned to host-plant affiliation (*F*_{HT}/*F*_{ST}). Only five loci contributed to significant *F*_{HT}-values (Table 2). Particularly important for population structure were the loci *Hex* and *PepD* (*F*_{ST} = 0.707 and *F*_{ST} = 0.735, respectively) where host affiliation explained over 98% of the interpopulation variance. Host-associated differentiation is reflected in the maximum likelihood phenogram (Figure 2). Populations clustered exclusively due to host-plant affiliation, whereas the geographic origin of populations was little important. The only visible geographical effect in the phenogram was Bavarian heterophyllum populations (site codes 35–40) that were separated from most of the Swedish heterophyllum populations.

IBD tests along Swedish heterophyllum and the German oleraceum transects revealed no significant correlations between genetic and geographic distances (28 contrasts *P* = 0.884 and 36 contrasts *P* = 0.212, respectively). Hence, geographic separation of populations within allopatric areas did not affect the comparisons of genetic identity between allopatric and parapatric/sympatric populations.

To assess whether and/or how geographic settings affect the degree of host-related differentiation we compared absolute *F*-values and *F*_{HT}/*F*_{ST}-ratios of allopatric, sympatric and parapatric populations. Over all loci, there was virtually no difference between sympatric and allopatric regions, neither in the amount of differentiation (*F*_{ST}) nor in the percentage of host-affiliated variance (*F*_{ST}/*F*_{HT}) (Table 3). However, the latter was clearly higher in the parapatric Bavarian populations (98.44 %), while *F*_{ST}-values were similar. Analysing the loci separately revealed that this held for almost all loci, which showed significant *F*_{HT}-values (Table 3). In addition to hierarchical *F*-statistics, we used genetic identities (Nei, 1972) as a measure of host-related population differentiation, which could be tested for effects of geographic setting. Figure 3a shows the mean genetic identities between heterophyllum and oleraceum sites of different geographic settings, for example the Bavarian heterophyllum populations had a similar mean genetic identity to parapatric and allopatric oleraceum populations (\bar{I} = 0.847 and \bar{I} = 0.845, respectively). The

Table 2 Hierarchical *F*-statistics over all loci and for single loci

Locus	<i>F</i> _{ST}	<i>F</i> _{SH}	<i>F</i> _{HT}	<i>F</i> _{HT} / <i>F</i> _{ST} (%)
<i>Acon</i>	0.040***	0.037***	0.004	
<i>Ark</i>	0.027***	0.028***	−0.001	
<i>Fum</i>	0.032***	0.034***	−0.001	
<i>G3pdh</i>	0.021*	0.021	−0.001	
<i>Got</i>	0.040***	0.032***	0.009**	21.93
<i>Hbdh</i>	0.071***	0.075***	−0.004	
<i>Hex</i>	0.707***	0.037***	0.696***	98.42
<i>ldh</i>	0.011**	0.010**	0.001	
<i>6Pgdh</i>	0.075***	0.041***	0.036***	47.50
<i>Pgm</i>	0.039***	0.038***	0.001	
<i>PepA</i>	0.049***	0.027***	0.023***	46.06
<i>PepD</i>	0.735***	0.051***	0.721***	98.06
<i>Tre</i>	0.008*	0.012**	−0.001	
All loci	0.343***	0.040***	0.315***	91.97

P* < 0.05; *P* < 0.01; ****P* < 0.001.

F-values are explained in the text.

mean genetic identities between sympatric/parapatric populations and populations of the same region affiliated to the alternative host did not differ significantly from those between sympatric/parapatric populations and allopatric populations of the alternative host, that is host-related differentiation was not higher in sympatric/parapatry than in allopatry (\bar{I} = 0.851 vs \bar{I} = 0.854; *Z*₃₉ = 0.725, *P* = 0.47). Corresponding tests were also performed for single loci that had showed significant *F*_{HT}-values. There were no significant differences at *6Pgdh* (\bar{I} = 0.966 vs \bar{I} = 0.976; *Z*₃₉ = 0.930, *P* = 0.35), *PepA* (\bar{I} = 0.912 vs \bar{I} = 0.893; *Z*₃₉ = −1.683, *P* = 0.09), *PepD* (\bar{I} = 0.126 vs \bar{I} = 0.102; *Z*₃₉ = 1.703, *P* = 0.09) and *Got* (\bar{I} = 0.970 vs \bar{I} = 0.975; *Z*₃₉ = 1.477, *P* = 0.14). However, the host-related differentiation at *Hex* was significantly higher in sympatry/parapatry than between sympatric/parapatric and allopatric regions (\bar{I} = 0.242 vs \bar{I} = 0.307; *Z*₃₉ = −2.647, *P* = 0.008). This difference was found in the parapatric Bavarian region as well as the sympatric Swedish region (Figure 3b).

In addition to the level of differentiation between geographic settings we assessed linkage disequilibria between significant host-race loci in contact areas. Linkage at these loci might be interpreted in terms of oviposition into wrong host plants and/or hybridisation among host races. After Bonferroni correction, no locus combination revealed significant deviations from

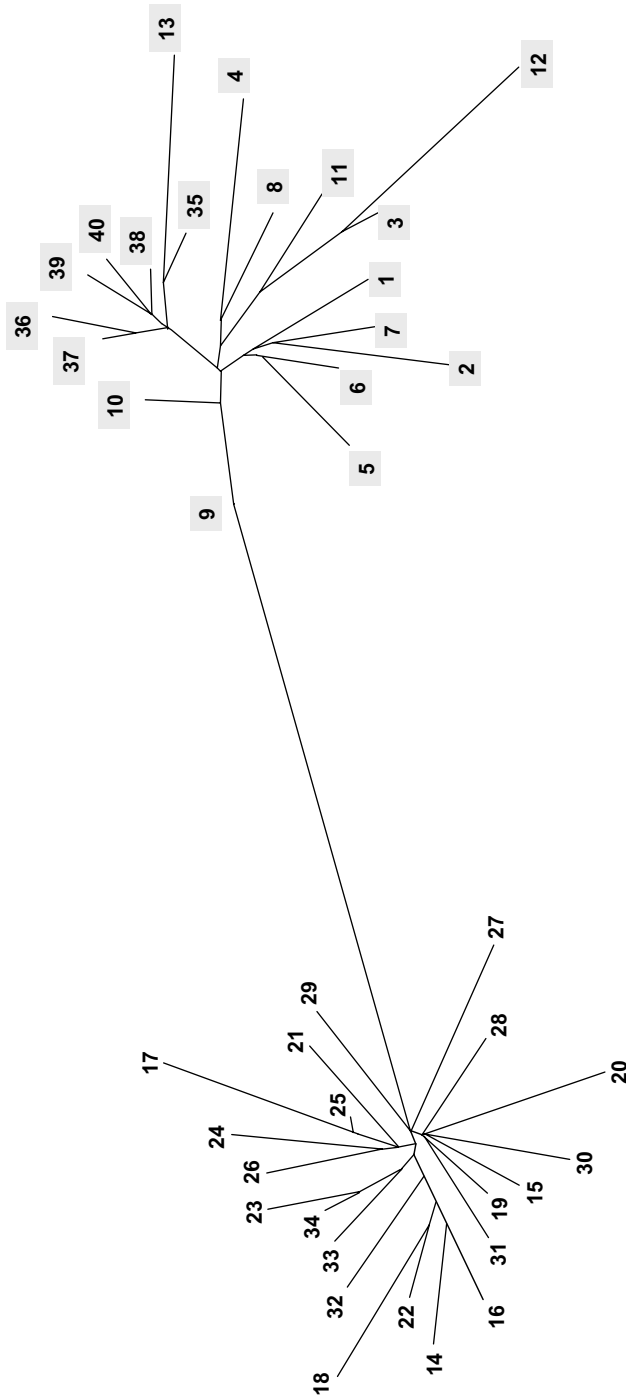


Figure 2 Phenogram of population relatedness. The phenogram builds two clearly distinct groups due to host-plant affiliation (heterophyllum populations are shaded). Site numbers refer to Figure 1.

expected multigenotype distributions. However, for oleraceum flies from contact areas the uncorrected P -value of the locus combination *Hex*/*PepD* was significant ($P=0.01$). By contrast, there was no evidence for linkage among *Hex* and *PepD* in allopatric oleraceum populations ($P=0.82$), nor in any geographic setting in heterophyllum flies (contact area: $P=1.00$; allopatric region: $P=0.70$). The reason for this pattern can be inferred from Table 4: while there is virtual no difference

Table 3 Interbiotype variance estimates for different geographic settings

	<i>Hetero</i> _{allo} vs <i>olera</i> _{allo}	<i>Hetero</i> _{sym} vs <i>olera</i> _{sym}	<i>Hetero</i> _{para} vs <i>olera</i> _{para}
<i>All loci</i>			
F_{ST}	0.341***	0.331***	0.345***
F_{HT}	0.315***	0.306***	0.340***
F_{HT}/F_{ST}	92.35%	92.42%	98.44%
<i>Hex</i>			
F_{ST}	0.641***	0.711***	0.792***
F_{HT}	0.623***	0.703***	0.791***
F_{HT}/F_{ST}	97.27%	98.74%	99.88%
<i>PepD</i>			
F_{ST}	0.776***	0.678***	0.723***
F_{HT}	0.761***	0.660***	0.722***
F_{HT}/F_{ST}	98.07%	97.38%	99.90%
<i>6Pgdh</i>			
F_{ST}	0.026**	0.128***	0.048**
F_{HT}	0.014*	0.054*	0.057**
F_{HT}/F_{ST}	53.79%	41.90%	100.00%
<i>PepA</i>			
F_{ST}	0.080***	0.060***	0.005
F_{HT}	0.058***	0.028*	0.001
F_{HT}/F_{ST}	71.85%	46.51%	—
<i>Got</i>			
F_{ST}	0.025***	0.065***	0.049***
F_{HT}	0.003	0.012	0.032***
F_{HT}/F_{ST}	—	—	65.27%

* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

Hetero = *C. heterophyllum* flies; *olera* = *C. oleraceum* flies; *allo* = allopatric; *sym* = sympatric; *para* = parapatric.

F_{HT}/F_{ST} was not calculated when F_{HT} was not significantly different from zero.

between observed and expected multilocus genotypes in heterophyllum flies, there is a significant excess of the multilocus genotype *Hex*_{100/107-*PepD*_{92/100} (ie F1 equivalents) in the oleraceum flies from contact areas.}

Discussion

Allozyme differentiation among heterophyllum and oleraceum populations was very strong, suggesting the existence of host races or even sibling species. Populations clustered exclusively by host-plant affiliation. The results correspond largely to scenario 3 (see Introduction) with geographic uniform estimates of differentiation between host races. This holds true for both host races in sympatric, parapatric and allopatric regions. These findings indicate that gene flow between heterophyllum and oleraceum flies is significantly restricted and that the system was established before the present geographic distribution arose. Hence, geography in its present state plays, at most, only a fine-tuning role in diversification (see below).

Unfinished lineage sorting where both host races share a central haplotype suggests a rapid diversification process in *T. conura* (Diegisser et al, 2006). Was the primary force promoting allozyme divergence among host races genetic drift (in allopatry) or some kind of divergent selection (independent of geography)? If genetic drift was a major cause for population differ-

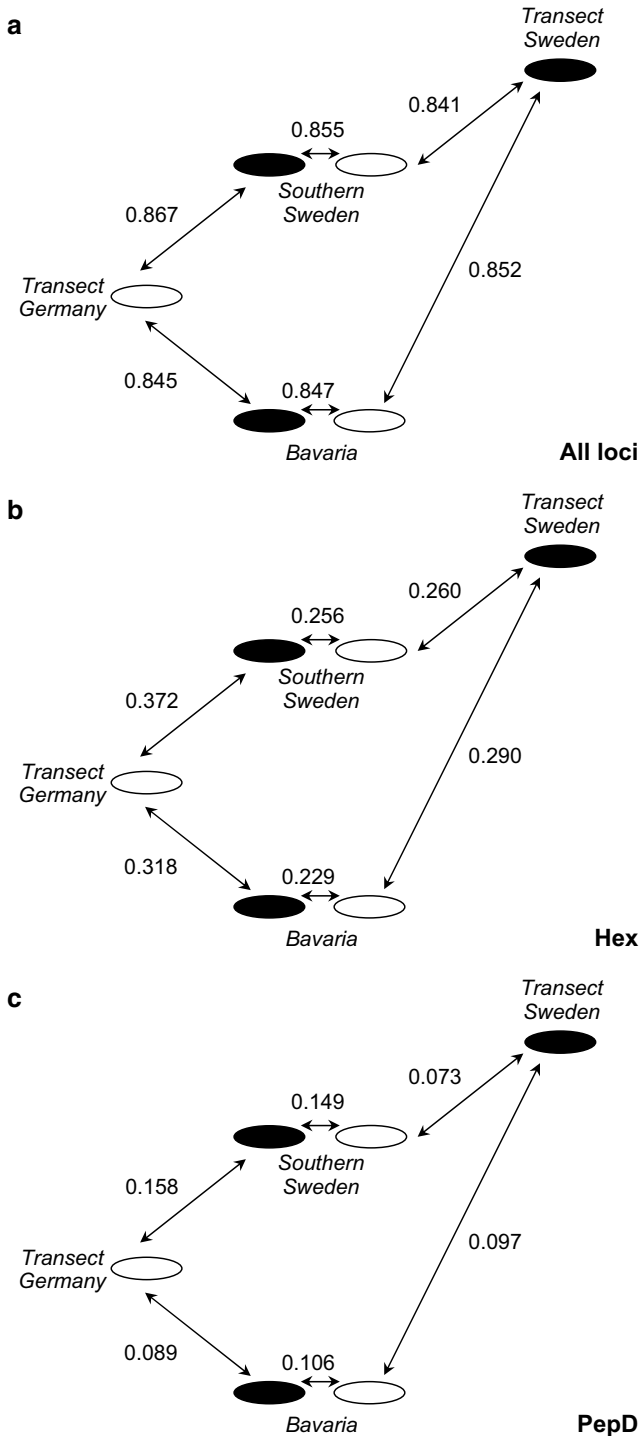


Figure 3 Mean genetic identities between heterophyllum and oleraceum flies of different geographic regions. Heterophyllum and oleraceum populations are represented by black and open circles, respectively. At the locus *Hex*, for example, the Bavarian oleraceum populations have a mean genetic identity of 0.229 with Bavarian heterophyllum populations, but of 0.290 with heterophyllum populations from the Swedish transect. (a) all loci; (b) Hexokinase; (c) Peptidase D.

entiation, all loci should be affected to a similar extent (McPherson *et al*, 1988). Thus, uniform differences at all analysed loci would indicate that populations have become differentiated by genetic drift (Itami *et al*, 1998).

Table 4 Observed and expected (in parenthesis) multigenotype distributions at the host-race characteristic loci *Hex* and *PepD* in contact areas (shown are only alleles with frequencies ≥ 0.05)

	<i>PepD</i> _92/92	<i>PepD</i> _92/100	<i>PepD</i> _100/100
<i>(a) Heterophyllum populations</i>			
<i>Hex</i> _100/100	1 (0.89)	31 (30.20)	159 (159.91)
<i>Hex</i> _100/107	0 (0.10)	3 (3.48)	19 (18.42)
<i>Hex</i> _107/107	0 (0.01)	0 (0.32)	2 (1.67)
<i>(b) Oleraceum populations</i>			
<i>Hex</i> _100/100	2 (1.88)	0 (0.12)	0 (0)
<i>Hex</i> _100/107	50 (54.44)	8 (3.56)	0 (0)
<i>Hex</i> _107/107	147 (142.68)	5 (9.32)	0 (0)

Note the excess of *Hex*_100/107-*PepD*_92/100 in the oleraceum population.

By contrast, if the differences among loci are highly heterogeneous, this would suggest that selection is acting on those loci which are highly differentiated (Slatkin, 1987). Heterophyllum and oleraceum populations revealed significant F_{HT} values at five of 13 loci, and only two of these 13 loci were highly differentiated, making the degree of host-related differentiation highly heterogeneous. Of the five host-related loci, the interpopulation variances at three loci *Got*, *6-Pgdh* and *PepA* were low ($F < 0.036$) and only significant in some but not all geographic settings. By contrast, the two loci *Hexokinase* and *Peptidase D* showed highly significant F_{HT} values in all geographic regions and, what was particularly remarkable, host-related differentiation at these loci was about 20 times higher than at any other locus. This pattern clearly suggests that selection is acting on *Hex* and *PepD*, or that the allozyme loci themselves are selectively neutral but linked to loci experiencing selection.

Feder *et al* (1993, 1997) could show that allele frequencies previously found to differ between *R. pomonella* host races correlated with eclosion time. The latter affects host choice directly and therefore contributes to prezygotic isolation between host races. There exist analogous conditions in *T. conura*. In Bavaria, *C. heterophyllum* provides flower heads suitable for oviposition about 3 weeks earlier than *C. oleraceum*, and Komma (1990) found evidence that the spring reappearance of flies on their respective host is synchronised with host plant phenology. Theoretically, differentiated allozyme loci could be linked to loci determining timing of oviposition. Alternatively, there could be linkage to loci affecting host choice. In experiments, *T. conura* showed significant host preferences (Romst ock-V okl, 1997). This can be explained in part by genetically determined colour preferences (Komma, 1990), but olfactory traits might in general also be important for host-plant recognition (Linn *et al*, 2003). Each of the above-mentioned traits facilitates prezygotic isolation because assortative mating arises as a by-product (Dres and Mallet, 2002). However, differentiated allozymes could also be linked to a second class of loci, namely those affecting performance. Host plants provide habitat and food for larvae. Adaptation to one host might result in lower fitness on the alternative host (Diehl and Bush, 1989). Egg-laying experiments suggest that such negative genetic correlations for larval performance exist in *T. conura*, as heterophyllum flies developed only very

weakly in *C. oleraceum* flower heads (T Diegisser, unpublished data).

Although we cannot infer yet whether *Hexokinase* and *Peptidase D* are under direct selection or linked to loci under selection, and which trait selection is acting on, the highly heterogeneous allozyme differentiation indicates that reproductive isolation between host races might primarily result from adaptive divergence. The question is whether reproductive isolation has evolved merely as a by-product of divergent selection (either in allopatry or in sympatry) or whether it has been driven by disruptive selection in contact areas, that is selection against intermediates. In the present study, we addressed this issue by comparing host differentiation in sympatry, parapatry and allopatry. Less differentiation in contact areas would have indicated that gene flow partly homogenises host-related divergence (scenario 1), while higher differentiation in contact areas (scenario 2) could have been interpreted in terms of selection against maladaptive events (ie hybridisation and/or wrong host choice). The latter process would be analogous to reinforcement ('natural selection strengthening sexual isolation in response to maladaptive hybridisation following secondary contact of two taxa' (Noor, 1999)), except that selection might act not only against hybridisation but also against oviposition into wrong hosts.

Neither of the above scenarios was clearly supported by the allozyme differentiation, which was mostly uniform among geographic settings. However, a uniform pattern might arise in contact areas in the presence of gene flow if it is counterbalanced by selection against maladaptive hybridisation/wrong host choice (scenario 3a). An indication for at least some ongoing gene flow was provided by linkage disequilibria among *Hexokinase* and *Peptidase D* in oleraceum flies. In contact areas, oleraceum flies had a small but significant excess of the heterozygote genotype *Hex_100/107-PepD_92/100*, which can be best explained by the existence of hybrid flies. Although double heterozygotes could result from intrahost-race matings, not a single fly (neither in heterophyllum nor in oleraceum populations) showed the most common double homozygote *Hex/PepD*-genotype of the alternative host race (see Table 4), hence the pattern suggest hybrids. There was no linkage disequilibrium in heterophyllum populations. Thus, gene flow among host races seems only to occur due mating between heterophyllum males and oleraceum females (which oviposit on *C. oleraceum*), while there is evidence neither for matings between oleraceum males and heterophyllum females, nor for oviposition into the wrong host plant.

The allozyme data also provide some, albeit weak, evidence for selection against maladaptive hybridisation/wrong host choice. The host-race divergent enzyme *Hex* was more divergent in contact areas but, generally, this differentiation was small compared to the uniform host-related divergence. The slight but significant difference may indicate that selection on *Hex* is in the final process of fine tuning host and/or mate preference. Although this process may counterbalance the effects of gene flow, the process itself does not depend on gene flow. Selection may continue against maladaptive traits even if there is no gene flow at all (Butlin, 1987). For example, even though there seems to be no gene

flow via oviposition into the wrong host, this does not mean that such oviposition does not occur: if postzygotic isolation already has developed so strongly that oviposition into the wrong host plant is fatal to the larvae, no gene flow occurs but selection might still strengthening oviposition preference because it is maladaptive. If the (weak) reinforcement pattern is true for *Hex*, it may well be selection against habitat choice that is important.

The geographic pattern of differentiation suggests that *T. conura* host races are stable and that the major diversification process took place before today's geographical settings were established. However, although the allozyme data provide some evidence for ongoing gene flow and selection against maladaptive events, today both play only a fine-tuning role in diversification between the two host races that are largely reproductively isolated.

Acknowledgements

We are grateful to two anonymous reviewers for their valuable comments on the manuscript. This study was supported by the Deutsche Forschungsgemeinschaft (DFG Se506/10-1).

References

- Butlin RK (1987). Speciation by reinforcement. *Trends Ecol Evol* 2: 8–13.
- Butlin RK (1990). Divergence in emergence time of host-races due to differential gene flow. *Heredity* 65: 47–50.
- Dieckmann U, Doebeli M (1999). On the origin of species by sympatric speciation. *Nature* 400: 354–357.
- Diegisser T, Seitz A, Johannesen J (2006). Phylogeographic patterns of host-race evolution in *Tephritis conura* (Diptera: Tephritidae). *Mol Ecol* 15: 681–694.
- Diehl SR, Bush GL (1989). The role of habitat preference in adaptation and speciation. In: Otte D, Endler JA (eds) *Speciation and its Consequences*. Sinauer Associates: Sunderland, Massachusetts, USA. pp 356–365.
- Dres M, Mallet J (2002). Host-races in plant-feeding insects and their importance in sympatric speciation. *Philos Trans R Soc Lond B Biol Sci* 357: 471–492.
- Feder JL, Berlocher SH, Roethele JB, Dambroski H, Smith JJ, Perry WL et al (2003). Allopatric genetic origins for sympatric host-plant shifts and race formation in *Rhagoletis*. *Proc Natl Acad Sci USA* 100: 10314–10319.
- Feder JL, Hunt TA, Bush GL (1993). The effects of climate, host plant phenology and host fidelity on the genetics of apple and hawthorn infesting races of *Rhagoletis pomonella*. *Entomol Exp Appl* 69: 117–135.
- Feder JL, Stolz U, Lewis KM, Perry W, Roethele JB, Rogers A (1997). The effects of winter length on the genetics of apple and hawthorn races of *Rhagoletis pomonella* (Diptera: Tephritidae). *Evolution* 51: 1862–1876.
- Felsenstein J (1993). *PHYLIP (Phylogeny Inference Package)* Version 3.5c. Computer program distributed by the author. Department of Genetics, University of Washington, Seattle.
- Hebert PDN, Beaton MJ (1989). *Methodologies for Allozyme Analysis Using Cellulose Acetate Electrophoresis. A Practical Handbook*. Educational Service of Helena Laboratories: Beaumont, Texas, USA.
- Howard DJ (1993). Reinforcement: Origin, dynamics, and fate of an evolutionary hypothesis. In: Harrison RG (ed) *Hybrid Zones and the Evolutionary Process*. Oxford University Press: Oxford. pp 46–69.

- Itami JK, Craig TP, Horner JD (1998). Factors affecting gene flow between the host-races of *Eurosta solidaginis*. In: Mopper S, Strauss SY (eds) *Genetic Structure and local Adaptation in Natural Insect Populations. Effects of Ecology, Life History, and Behavior*. Chapman & Hall: New York. pp 375–407.
- Jiggins CD, Mallet J (2000). Bimodal hybrid zones and speciation. *Trends Ecol Evol* **15**: 250–255.
- Kawecki TJ (1997). Sympatric speciation by habitat specialization driven by deleterious mutations. *Evolution* **51**: 1751–1763.
- Komma M (1990). *Der Pflanzenparasit Tephritis conura und die Wirtsgattung Cirsium*. Dissertation, University of Bayreuth, Germany.
- Kondrashov AS, Kondrashov FA (1999). Interactions among quantitative traits in the course of sympatric speciation. *Nature* **400**: 351–354.
- Lewis PO, Zaykin D (2001). *Genetic data analysis: Computer program for the analysis of allelic data*. Version 1.0 [d16.c]. Distributed by the authors over the internet from <http://lewis.eeb.uconn.edu/lewishome/software.html>.
- Linn C, Feder JL, Nojima S, Dambroski HR, Berlocher SH, Roelofs W (2003). Fruit odor discrimination and sympatric host-race formation in *Rhagoletis*. *Proc Natl Acad Sci USA* **100**: 11490–11493.
- Mayr E (1963). *Animal Species and Evolution*. Belknap Press: Cambridge, MA.
- McPherson BA, Smith DC, Berlocher SH (1988). Microgeographic genetic variation in the apple maggot *Rhagoletis pomonella*. *Genetics* **119**: 445–451.
- Nei M (1972). Genetic distance between populations. *Am Nat* **106**: 283–292.
- Noor M (1999). Reinforcement and other consequences of sympatry. *Heredity* **83**: 503–508.
- Presgraves DC, Balagopalan L, Abmayr SM, Orr HA (2003). Adaptive evolution drives divergence of a hybrid in viability gene between two species of *Drosophila*. *Nature* **423**: 715–719.
- Raymond M, Rousset F (2000). *GENEPOP*. Version 3.2. Distributed by the authors. Institut des Sciences de l'Evolution, Université de Montpellier II, France.
- Richardson BJ, Baverstock PR, Adams M (1986). *Allozyme electrophoresis. A handbook for animal systematics and population studies*. Academic Press: San Diego.
- Romstöck M, Arnold H (1987). Populationsökologie und Wirtswahl bei *Tephritis conura* Loew-Biotypen (Dipt.: Tephritidae). *Zool Anz* **219**: 83–120.
- Romstöck-Völkl M (1997). Host-race formation in *Tephritis conura*: determinants from three trophic levels. *Ecol Stud* **130**: 21–38.
- Schluter D (2001). Ecology and the origin of species. *Trends Ecol Evol* **16**: 372–380.
- Schneider S, Roessli D, Excoffier L (2000). *Arlequin: A software for population genetics data analysis*. Version 2.0. Genetics and Biometry Lab, Department of Anthropology, University of Geneva.
- Seitz A, Komma M (1984). Genetic polymorphism and its ecological background in Tephritid populations (Diptera: Tephritidae). In: Wöhrmann K, Loeschke V (eds) *Population Biology and Evolution*. Springer: Heidelberg. pp 143–158.
- Siegismund HR (1993). *G-Stat. Genetical statistical programs for the analysis of population data*. Version 3. The Arboretum, Royal Veterinary and Agricultural University, Horsholm, Denmark.
- Slatkin M (1987). Gene flow and the geographic structure of natural populations. *Science* **236**: 787–792.
- Slatkin M (1993). Isolation by distance in equilibrium and non-equilibrium populations. *Evolution* **47**: 264–279.
- Stam P (1983). The evolution of reproductive isolation in closely adjacent plant populations through differential flowering time. *Heredity* **50**: 105–118.
- Turelli M, Barton NH, Coyne JA (2001). Theory and speciation. *Trends Ecol Evol* **16**: 330–343.
- Via S (2001). Sympatric speciation in animals: the ugly duckling grows up. *Trends Ecol Evol* **16**: 381–390.
- Weir BS, Cockerham CC (1984). Estimating F-statistics for the analysis of population structure. *Evolution* **38**: 1358–1370.
- Zwölfer H (1988). Evolutionary and ecological relationships of the insect fauna of thistles. *Annu Rev Entomol* **33**: 103–122.

Appendix A

Sample site data
See Table A1.

Table A1

No.	Site code	Site location		Host	Sampled	N	H _c	H _o
<i>Allopatric</i>								
1	HST10	N63°57.722'	E018°51.181'	<i>Hetero</i>	2003	24	0.169	0.171
2	HST9	N63°28.916'	E018°34.759'	<i>Hetero</i>	2003	24	0.140	0.154
3	HST8	N62°53.812'	E018°01.915'	<i>Hetero</i>	2003	24	0.220	0.208
4	HST7	N62°13.975'	E017°22.636'	<i>Hetero</i>	2003	24	0.140	0.170
5	HST6	N61°39.197'	E016°55.620'	<i>Hetero</i>	2003	24	0.158	0.157
6	HST5	N61°02.050'	E016°41.015'	<i>Hetero</i>	2003	24	0.172	0.178
7	HST4	N60°21.146'	E017°28.027'	<i>Hetero</i>	2003	24	0.149	0.147
8	HST3	N59°29.212'	E018°10.773'	<i>Hetero</i>	2003	24	0.202	0.180
<i>Sympatric</i>								
9	HS2	N55°51.103'	E013°52.274'	<i>Hetero</i>	2003	24	0.189	0.192
10	HS4	N55°46.541'	E013°55.570'	<i>Hetero</i>	2003	24	0.142	0.154
11	HS6	N55°51.857'	E013°44.779'	<i>Hetero</i>	2003	24	0.217	0.208
12	HS8	N55°55.645'	E013°34.628'	<i>Hetero</i>	2003	24	0.191	0.184
13	HS13	N56°03.852'	E013°34.242'	<i>Hetero</i>	2003	24	0.169	0.183
14	OS2	N55°50.854'	E013°36.033'	<i>Olera</i>	2002	24	0.196	0.181
15	OS3	N55°47.556'	E013°29.973'	<i>Olera</i>	2002	24	0.226	0.221
16	OS6	N56°17.509'	E013°42.424'	<i>Olera</i>	2002	24	0.209	0.237
17	OS7	N56°16.772'	E012°50.849'	<i>Olera</i>	2002	23	0.271	0.281
18	OS8	N56°04.982'	E012°49.773'	<i>Olera</i>	2002	24	0.174	0.183
19	OS10	N56°05.363'	E013°44.346'	<i>Olera</i>	2002	24	0.208	0.231

Table A1 Continued

No.	Site code	Site location		Host	Sampled	N	H _e	H _o
<i>Allopatric</i>								
20	OGT2	N54°16.906'	E010°34.350'	<i>Olera</i>	2002	24	0.226	0.218
21	OGT4	N53°19.509'	E010°11.112'	<i>Olera</i>	2002	21	0.193	0.165
22	OGT5	N52°08.669'	E009°26.315'	<i>Olera</i>	2002	24	0.167	0.157
23	OGT7	N51°06.768'	E009°35.827'	<i>Olera</i>	2002	24	0.223	0.212
24	OGT8	N50°32.323'	E009°11.950'	<i>Olera</i>	2002	24	0.275	0.279
25	OGT9	N50°03.124'	E008°05.860'	<i>Olera</i>	2002	24	0.222	0.205
26	OGT10	N49°31.781'	E008°06.512'	<i>Olera</i>	2002	24	0.235	0.215
27	OGT12	N48°31.119'	E007°51.059'	<i>Olera</i>	2002	24	0.203	0.192
28	OGT14	N47°43.000'	E008°20.000'	<i>Olera</i>	2003	24	0.207	0.183
<i>Parapatric</i>								
29	OG2	N49°53.092'	E011°36.975'	<i>Olera</i>	2002	24	0.218	0.192
30	OG7	N49°58.683'	E011°31.389'	<i>Olera</i>	2002	24	0.256	0.271
31	OG11	N49°56.578'	E011°40.419'	<i>Olera</i>	2002	24	0.236	0.249
32	OHG1	N50°02.722'	E011°44.509'	<i>Olera</i>	2002	18	0.202	0.222
33	OHG2	N50°00.145'	E011°43.309'	<i>Olera</i>	2002	24	0.195	0.185
34	OHG3	N49°58.159'	E011°45.985'	<i>Olera</i>	2002	24	0.195	0.186
35	HOG1	N50°02.722'	E011°44.509'	<i>Hetero</i>	2002	24	0.146	0.157
36	HOG2	N50°00.145'	E011°43.309'	<i>Hetero</i>	2002	24	0.178	0.199
37	HOG3	N49°58.159'	E011°45.985'	<i>Hetero</i>	2002	24	0.183	0.196
38	HG4	N50°00.978'	E011°48.504'	<i>Hetero</i>	2002	24	0.194	0.193
39	HG9	N50°03.905'	E011°50.411'	<i>Hetero</i>	2002	24	0.179	0.176
40	HG13	N49°58.794'	E011°52.112'	<i>Hetero</i>	2002	24	0.184	0.159

Numbers of first column refer to Figure 1. hetero = *C. heterophyllum*; olera = *C. oleraceum*; H_e = expected heterozygosity; H_o = observed heterozygosity.