Genetic differentiation in *Zeiraphera diniana* (Lepidoptera: Tortricidae, the larch budmoth): polymorphism, host races or sibling species?

IGOR EMELIANOV*, JAMES MALLET & WERNER BALTENSWEILER†
Galton Laboratory, Department of Genetics and Biometry, University College London, 4 Stephenson Way, London NW1 2HE, U.K. and †Blumenbergstrasse 9, CH-8634 Hombrechtikon, Switzerland.

Eleven larch- and pine-feeding populations of the larch budmoth were studied in Western Europe using 24 allozyme loci. Hierarchical *F*-statistics between larch and pine forms ($F_{ST} \approx 0.065$) were much greater than those between localities within larch ($F_{ST} \approx 0.002$) and within pine ($F_{ST} \approx 0.026$), even when sympatric host-associated populations were considered. Analysis of Nei's genetic distance produced similar results, and a UPGMA tree of all populations consistently clustered samples from the same hosts together. Three loci, autosomal *Pgm* and *Mdh-s* and sex-linked *Idh-s*, were the most important loci involved in differences between the two forms. Previously considered to be conspecific, it is now clear that the larch and pine biotypes of *Z. diniana* are either host races (*sensu* Diehl & Bush, 1984) or good sympatric species which hybridize rarely. Regardless of taxonomic status, the distribution of larch and pine in the Pleistocene, phenological differences between the two hosts, together with the great vagility of *Z. diniana*, an outbreak pest, suggest that sympatric or parapatric differentiation is at least as likely as allopatric divergence.

Keywords: allozymes, host races, population structure, sex linkage, speciation, *Zeiraphera diniana*

Introduction

Whether sympatric speciation is common is one of the more controversial issues in biology (Mayr, 1970; Grant & Grant, 1979; Futuyma, 1983; Diehl & Bush, 1984; Bush, 1994). By sympatric speciation we mean a process of genetic differentiation into species occurring between populations that are within normal dispersal distance of each other. The widespread belief in the alternative, allopatric model is so prevalent perhaps because it is simpler to think of divergence in the absence of gene flow, rather than because of any theoretical implausibility of divergence in sympatry (Barton & Charlesworth, 1984). 'Host races' (*sensu* Diehl & Bush, 1984), where host-associated differentiation of phytophagous or parasitic insects is maintained in the face of gene flow, provide an important intermediate link between host adaptation within panmictic populations and 'good'

host-associated species which rarely, if ever, hybridize. Whether or not these host races initially evolved in allopatry, the existence of hybridizing host races implies that there can be a stable route for sympatric speciation because gene flow is not able to swamp genetic differentiation.

Host-associated forms have been found in many insect species (Diehl & Bush, 1984); however, the best-studied examples often either lack significant genetic differentiation (Mitter & Futuyma, 1979; Jaenike & Selander, 1980; Menken, 1981) or are sibling species with no known gene exchange (Guttman *et al.*, 1981; Pashley, 1988). One of the few cases that stands up to close examination is that of the apple versus hawthorn races of *Rhagoletis pomonella* (Feder *et al.*, 1988, 1994; McPheron *et al.*, 1988; Smith, 1988; Bush *et al.*, 1989). Overall, the scarcity of 'true' host races, as defined above, casts doubt on the idea that sympatric speciation is common. The lack of examples of true host races could, however, be because few likely cases have been studied in much detail. This is the first study to use

*Correspondence.

genetic markers to investigate the host-associated populations of *Zeiraphera diniana*.

The larch budmoth, *Zeiraphera diniana* Guenée is a classic outbreak pest of coniferous forests across the Palaearctic (Varley *et al.*, 1973; Berryman, 1986). The species consists of host-associated forms feeding on larch (*Larix*), pine (*Pinus*) and spruce (*Picea*). The population biology of the larch budmoth has been studied in great detail for over 30 years by Baltensweiler *et al.* (reviewed by Baltensweiler *et al.*, 1977; Baltensweiler, 1993). This moth has highly cyclical population dynamics above 1000 m in larch forests of the Alps: every 8–10 years there are outbreaks during which populations peak at 100,000-fold greater densities than during the subsequent crash. Associated with this population fluctuation is an apparent genetic cycle involving 5th instar larval colour morphs; dark morphs predominate during population build-up, whereas pale morphs become commoner during the decline phase. The differences between the dark and pale forms of larvae on larch are similar to the differences between larvae on larch and pine; on larch, larvae are predominately dark, on pine they are pale. This suggests that host-associated differentiation contributes to the cycles of genetic variation on larch, and may even be involved in population regulation (Baltensweiler, 1993).

There is a variety of differences between the host-associated populations of *Z. diniana*, apart from larval colour pattern. (i) Size: larch and pine form adults are not known to differ in structure, but larch moths are on average larger (W. Baltensweiler, unpublished data). (ii) Phenology: egg hatch of the larch form is synchronized with the flush of larch foliage, about 2–4 weeks earlier than the pine form and its host shoots (Day, 1984). (iii) Host preference: larch and pine forms tend to lay eggs on their own hosts (Maksymov, 1959; Bovey & Maksymov, 1959). (iv) Host-associated larval survival: the larch form does not normally survive either on pine or spruce, whereas the pine form does well on all three hosts (Day, 1984; W. Baltensweiler, unpublished

data). (v) Pheromones: larch and pine forms differ in female pheromone blend (Guerin *et al.*, 1984).

In spite of these differences, the two forms have always been considered members of the same species because: (i) they interbreed freely in the laboratory, with no distortion of sex ratio in the progeny (Table 1; Baltensweiler, 1977); (ii) studies of variation in female pheromone blend, variable response of males to these pheromones and the intermediacy of F_1 hybrids suggests that introgression should be common in the wild (Priesner, 1979; Day, 1984; Guerin *et al.*, 1984; Baltensweiler & Priesner, 1988); (iii) their populations fluctuate in synchrony; and (iv) probable hybrid forms are often found on larch (Baltensweiler *et al.*, 1977; Baltensweiler, 1993).

We are currently engaged in an investigation of genetic differences between larch and pine forms of *Z. diniana* and the potential relationship between genetics and population dynamics. In this paper, we use allozymes to investigate genetic differences between host-associated biotypes. Only one previous study on the allozymes of *Z. diniana* exists, in which only three loci were investigated and the differences between host races were not studied (Verdon, 1978).

Materials and methods

A total of 300 larvae were collected during summer 1994 in 11 larch- or pine-feeding populations at nine sites (Fig. 1 and Table 2). About 80 per cent of these larvae were frozen in liquid N_2. Others were transported to the laboratory and reared to adulthood. These adults were used to test for the possible effects of host plants, parasitoids or pathogens, which might have affected larval electrophoretic banding patterns. Adults could be sexed in the usual way. We also found that males of later instar pine form larvae could be detected because their dark testes were visible through the semi-transparent cuticle. Sexing was more difficult for last instar larch form larvae, nearly all of which have a heavily pigmented cuticle.

Table 1 Sex ratio in larch budmoth hybrids†

Type of cross	No. of pairs	F_1 females	F_1 males	Females (%)	P (test: no distortion)
L × L	23	621	628	49.7	>0.5
P × P	11	117	128	47.8	>0.5
L × P	5	211	198	51.6	>0.5
P × L	4	98	82	54.4	>0.1

†Broods are as mentioned in Baltensweiler (1993).
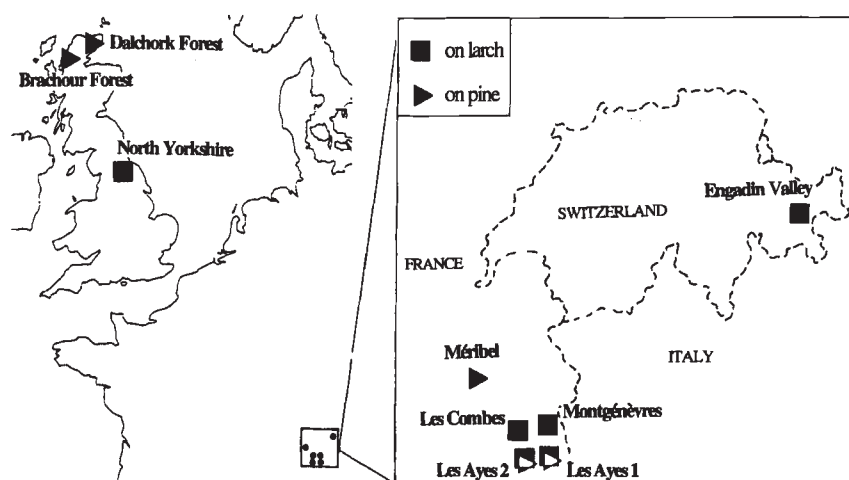L: larch form; P: pine form.

Fig. 1 Geographical locations of 11 samples of *Zeiraphera diniana*. Abbreviated geographical names used further in this paper (Table 2, Fig. 2) are given in bold. BRITAIN: 1: North Yorkshire (**NY**), two individuals from two sites, Harwood Dale Forest (grid ref. NZ960988) and Cropton Forest (NZ810937), larvae on Japanese larch (*Larix leptolepis*); 2: Brachour Forest (**Brac**, DN090500), on *Pinus contorta*; 3: North Dalchork Forest (**Dal**, NC550200), on *P. contorta*. FRANCE: 4: Méribel (**Mér**), 45°23′N 6°35′E, on *P. cembra*; 5: Bois Les Combes (**LC**), near Briançon, 44°54′N 6°34′E, on *L. decidua*; 6, 7: Bois les Ayes sites 1 and 2 (**LA1** and **LA2**), near Briançon, 44°50′N 6°39′E and 44°51′N 6°41′E, respectively, larvae on both *P. cembra* and *L. decidua* at each site; 8: Mont-génèvres (**Mont**), near Briançon, 44°56′ N 6°43′E, larvae on *L. decidua*. SWITZERLAND: 9: Engadin Valley (**Eng**), 46°26′ N 9°50′E, on *L. decidua*.

Larvae were cut in half and adults were also split between the thorax and abdomen, after trimming legs and wings. Posterior larval segments and thoraces of adults were kept for DNA extraction; anterior larval sections and adult abdomens were homogenized on ice in 25–35 μL of extraction buffer containing 0.01 per cent NADP, 0.1 per cent β-mercaptoethanol and 0.1 per cent Triton X-100 detergent (Richardson *et al.*, 1986) and centrifuged. Chloroform (7–12 μL) was added to the homogenate before centrifugation to remove excessive lipids and lipoproteins. Reference aliquots (10 μL) were taken from 20–30 μL of resulting supernatants. Electrophoresis was performed on Helena cellulose acetate plates supplied by Helena Laboratories. Two buffer systems were used: TGB (0.3 per cent Trizma, 1.4 per cent glycine, pH 8.6) and PB (0.36 per cent NaH$_2$PO$_4$, 0.22 per cent Na$_2$HPO$_4$, pH 6.3). At least two reference aliquots with known banding patterns were loaded on each plate along with the new samples. Each electrophoretic run was planned so that wherever possible a variety of samples was represented on each plate to avoid artefactual differences between populations generated by run-to-run variation in resolution and staining. The plates were run for 30 minutes in a cold room (+2°C) with 250 V at 4 mA. After completing the run, plates were stained for particular enzymes as described else-

where (Richardson *et al.*, 1986; Mallet *et al.*, 1993) with minor modifications.

A total of 19 enzyme systems of 36 tested were sufficiently scorable and active, and were selected for further analysis (abbreviated name of enzyme, EC no., and buffer system are shown in parentheses): glutamate-oxaloacetate transaminase (*Got* 2.6.1.1, TGB), phosphoglucose isomerase (*Pgi* 5.3.1.9, TGB), phosphoglucomutase (*Pgm* 2.7.5.1, TGB), malic enzyme (*Me* 1.1.1.40, TGB), phe-pro and leu-ala peptidases (*Pp* and *La* 3.4.11 or 13, TGB), aconitate hydratase (*Acon* 4.2.1.3, PB), esterase (*Est* 3.1.1.1, TGB), isocitrate dehydrogenase (*Idh* 1.1.1.42, TGB), fumarate hydratase (*Fum* 4.2.1.2, PB), adenylate kinase (*Ak* 2.7.4.3, PB), enolase (*Enol* 4.2.1.11, PB), malate dehydrogenase (*Mdh* 1.1.1.37, PB), hexokinase (*Hk* 2.7.1.1, PB), glyceraldehyde-3-phosphate dehydrogenase (*Gapdh* 1.2.1.12, PB), glycerate dehydrogenase (*Gldh* 1.1.1.29, TGB), aldolase (*Ald* 4.1.2.13, PB), glutathione reductase (*Gr* 1.6.4.2, PB), arginine kinase (*Argk* 2.7.3.3, PB).

Allele and genotypic frequencies were calculated separately for each sample. Genotypic frequencies for each polymorphic locus in each sample were tested for fit to Hardy–Weinberg using χ^2 goodness-of-fit tests. Expected heterozygosity (H_e, unbiased) was used as a measure of genetic variability and was

calculated using Nei's formula (1978). The genetic divergence between populations was estimated by making two separate analyses: (i) Wright's F_{ST} (Wright, 1978) was calculated and contingency χ^2-tests were used to test whether interpopulation genetic variation was different from that expected by chance; and (ii) unbiased genetic distance (D_N) was computed using Nei's (1978) formula. Subsequent UPGMA was used in a cluster analysis of D_N matrices. All calculations were performed manually or using BIOSYS-1 for the PC (Swofford & Selander, 1989).

Results

A total of 24 loci from 19 enzyme systems were scored. Electrophoretic variation was detected at 13 of these 24 loci. *Me*, *Acon-s* and *La-s* all have frequencies of their most common allele >0.95 and did not differ significantly between pine and larch larvae. These three loci were therefore considered monomorphic along with *Got-f*, *Fum*, *Ak*, *Enol*, *Mdh-f*, *Hk*, *Gldh*, *Gapdh*, *Ald*, *Gr* and *Argk* which showed no variation. Overall allele frequencies for

Table 2 Allele frequencies at polymorphic loci in *Zeiraphera diniana*

Loci and alleles	Larch						Overall larch	Pine					Overall pine
	Mont	LA1	LA2	LC	Eng	NY		LA1	LA2	Mér	Dal	Brac	
Pgm													
N	37	33	49	42	8	2	*171*	34	3	64	18	7	*126*
0.75	0.000	0.000	0.000	0.000	0.000	0.000	*0.000*	0.015	0.000	0.000	0.000	0.000	*0.005*
0.87	0.000	0.000	0.031	0.000	0.000	0.000	*0.021*	0.000	0.000	0.008	0.056	0.000	*0.021*
1.00	0.743	0.848	0.734	0.702	0.750	1.000	*0.750*	0.647	0.667	0.578	0.556	0.571	*0.594*
1.12	0.257	0.152	0.235	0.286	0.250	0.000	*0.228*	0.338	0.333	0.414	0.388	0.429	*0.380*
1.28	0.000	0.000	0.000	0.012	0.000	0.000	*0.002*	0.000	0.000	0.000	0.000	0.000	*0.000*
Idh-s (sexed males; assumes heterozygous 'females' are actually males)													
N	1	NA	9	5	NA	NA	*15*	NA	NA	10	9	2	*21*
0.65	0.000		0.000	0.000			*0.000*			0.000	0.000	0.000	*0.000*
1.00	1.000		0.500	0.600			*0.567*			0.400	0.277	0.500	*0.357*
2.40	0.000		0.500	0.400			*0.433*			0.600	0.723	0.500	*0.643*
Idh-s (sexed females)													
N	1	NA	9	3	NA	NA	*13*	NA	NA	40	3	2	*45*
0.65	0.000		0.000	0.000			*0.000*			0.000	0.000	0.000	*0.000*
1.00	1.000		0.667	0.333			*0.615*			0.150	1.000	1.000	*0.245*
2.40	0.000		0.333	0.667			*0.385*			0.850	0.000	0.000	*0.755*
Idh-s (total, treated as though autosomal)													
N	37	32	48	43	8	2	*170*	34	3	65	18	7	*127*
0.65	0.014	0.000	0.000	0.000	0.000	0.000	*0.003*	0.000	0.000	0.000	0.000	0.643	*0.000*
1.00	0.621	0.438	0.583	0.477	0.687	0.250	*0.538*	0.191	0.000	0.208	0.500	0.000	*0.264*
2.40	0.365	0.562	0.417	0.523	0.313	0.750	*0.459*	0.809	1.000	0.792	0.500	0.357	*0.736*
Mdh-S													
N	37	31	48	42	8	2	*168*	34	3	65	18	7	*127*
0.70	0.000	0.000	0.031	0.012	0.000	0.000	*0.012*	0.000	0.000	0.000	0.000	0.000	*0.000*
1.00	0.865	0.806	0.740	0.809	0.812	1.000	*0.803*	0.103	0.333	0.015	0.056	0.000	*0.051*
3.00	0.000	0.000	0.000	0.012	0.000	0.000	*0.003*	0.000	0.000	0.008	0.000	0.000	*0.004*
5.50	0.135	0.194	0.229	0.167	0.188	0.000	*0.182*	0.897	0.667	0.977	0.944	1.000	*0.945*
H_e‡	0.169	0.162	0.178	0.184	0.178	0.087	*0.183*	0.162	0.197	0.152	0.170	0.172	*0.173*
	(0.052)	(0.051)	(0.051)	(0.054)	(0.049)	(0.060)	*(0.054)*	(0.049)	(0.060)	(0.048)	(0.054)	(0.057)	*(0.053)*

N: number of tested individuals; NA: not available.
†See Fig. 1 for the key to abbreviated names of populations.
‡Averaged over all 24 loci. Standard error in parentheses.

three polymorphic loci showing differences between host-associated biotypes are given in Table 2. Thirteen cases of significant ($P<0.05$) deviation from Hardy-Weinberg were found in 74 χ^2-tests across all populations both of larch and pine form. This significant deviation persisted when χ^2 values were summed over populations (summed $\chi^2_{366} = 548.7$, $P<0.05$). The deviation mainly occurred at three loci: *Idh-s*, *Est* and *Pp-s* (Table 3). The highly significant deviation at *Idh-s* appears to be caused by strong heterozygous deficit in females and suggests sex-linkage: 18 of 34 sexed males were homozygous at *Idh-s* whereas 58 of 60 sexed females were scored as 'homozygotes' (i.e. hemizygotes, $\chi^2_1 = 31.89$, $P<0.001$; errors in sexing probably account for the two exceptions). Sex linkage is, however, unlikely to be a reason for Hardy-Weinberg deviations at *Est* and *Pp-s* because there were no significant differences in heterozygosity between the sexes at these loci. These distortions might be an artefact from a breakdown of the χ^2 continuity assumption with the large numbers of genotypic classes (up to 28 at *Pp-s* and 21 at *Est*) and too few individuals in each. We grouped *Pp-s* and *Est* alleles into three classes: faster than the common allele, common allele and slower than the common allele; the test for the *Pp-s* locus then became nonsignificant in all populations but one (Méribel on pine) and significant deviation at

Est persisted only at Les Ayes site 2 on larch and pine. Excluding the *Idh-s* locus from the calculations, and grouping *Est* and *Pp-s* alleles reduced the overall deviation from Hardy–Weinberg to a non-significant level (Table 3).

Differentiation between seven large samples of *Z. diniana*, expressed in terms of Wright's F_{ST}, is shown in Table 4. As the overall F_{ST} (0.083) on its own does not show what factor (association with a particular host or geographical variation between localities) contributes most to genetic divergence, we calculated gene frequency variance at three hierarchical levels: within host, between hosts and overall. F_{ST} within pine together with overall F_{ST} were calculated in two different ways (with and without British populations) to estimate the contribution of the more than 1000 km separating the British and Continental sites to the overall differentiation.

Within-host differentiation between populations on larch is about 5 per cent of total and is about 30 per cent of total on pine (F_{ST} values: larch, 0.002; pine, 0.026; overall, 0.083). Between-host variance is clearly larger ($F_{ST} = 0.065$) than that within hosts and accounts for 78 per cent of the among-population variance. Excluding British data (Dalchork) from the pine populations causes a reduction of F_{ST} within pine to 0.009, an increase in between-host F_{ST} to 0.070 and a reduction in overall F_{ST} to 0.078;

Table 3 Test of fit to Hardy–Weinberg equilibrium for polymorphic loci in *Zeiraphera diniana*

Loci	Larch			Pine			Sum		
	χ^2	d.f.	P	χ^2	d.f.	P	χ^2	d.f.	P
Got-s	15.526	12	NS	1.866	3	NS	17.392	15	NS
Pgi	0.155	6	NS	0.167	2	NS	0.322	8	NS
Pgm	6.270	8	NS	3.594	9	NS	9.864	17	NS
Pp-f	6.575	10	NS	2.967	5	NS	9.542	15	NS
Pp-s†	13.226	12	NS	26.139	9	**	39.365	21	*
La-f	20.151	24	NS	20.595	18	NS	40.746	42	NS
Acon-f	21.658	18	NS	13.251	18	NS	34.909	36	NS
Est†	29.170	12	**	28.614	9	***	57.784	21	***
Idh-s	31.700	6	***	33.442	3	***	65.142	9	***
Mdh-s	2.977	16	NS	1.509	5	NS	4.486	16	NS
Total	147.408	124	NS	132.144	81	***	279.552	205	***
Total excluding *Idh-s*	115.708	118	NS	98.702	78	NS	214.410	196	NS

†Calculations for *Pp-s* and *Est* were made after grouping allele frequencies (see Results).
*$P \leqslant 0.01$, **$P \leqslant 0.005$, ***$P \leqslant 0.001$.

Table 4 F_{ST} values for *Zeiraphera diniana*

| Loci | F_{ST} at various hierarchical levels of differentiation | | | |
	Within larch†	Within pine†	Between hosts	Overall
Got-s	0.000	0.001	−0.003	−0.003
Pgi	0.000	0.000	−0.004	−0.003
Pgm	0.003	0.000	0.023***	0.024**
Me	—	0.000	−0.001	−0.001
Pp-f	0.002	0.001	0.000	0.001
Pp-s	0.000	0.016***	−0.003	0.007***
La-f	0.000	0.022**	−0.004	0.009*
La-s	0.000	—	0.000	0.001
Acon-f	0.015**	0.016*	−0.002	0.025***
Acon-s	0.000	0.000	−0.003	−0.003
Est	0.000	0.044***	−0.006	0.019***
Idh-s‡	0.003	0.085**	0.023***	0.078***
Mdh-s	0.000	0.003	0.543***	0.545***
Average overall	0.002	0.026***	0.065***	0.083***

†Performed separately, without hierarchy.
‡As a conservative measure, all homozygotes at the sex-linked locus *Idh-s* were treated as though female (hemizygous) for the purposes of this analysis.
Probability (null hypothesis is no differentiation): *$P \leqslant 0.05$, **$P \leqslant 0.01$, ***$P \leqslant 0.001$.

between-host variance then accounts for 90 per cent of total variance; thus the between-host variance is not an artefact of including a British pine population. Association with a particular host is the major factor contributing to the overall genetic variance.

Significant gene frequency variation at different hierarchical levels occurs at different sets of loci (Table 4). Populations were significantly differentiated at *Acon-f* on larch and at *Pp-s, La-f, Acon-f, Est* and *Idh-s* on pine. In contrast, there is significant variance at *Pgm, Mdh-s* and *Idh-s* between hosts; it is notable that two of the loci differentiated between hosts (*Pgm, Mdh-s*) do not show strong geographical variation within hosts.

UPGMA analysis of the Nei's D_N matrix, based on the complete set of 24 loci in 11 populations, confirms the hierarchical analysis of F_{ST}. The resultant tree divides the samples into two separate clusters, a larch cluster and a pine cluster; the average Nei's D_N between them is 0.039 (Fig. 2a). Even populations with small sample sizes (Eng, NY, LA1 pine and Brac) clustered with other samples from the same host plant. As already pointed out, the key loci in detecting interhost differentiation are *Pgm, Mdh-s* and *Idh-s*. Figure 2b represents seven populations with sample sizes more than 30 genomes analysed for 24 loci; average D_N between host biotypes is 0.036. When the *Mdh-s* and *Idh-s* loci are excluded

from the calculations (Fig. 2c), average D_N drops sharply to 0.005, and the Dalchork pine population joins the larch cluster; the other two pine populations remain separated. Finally, exclusion of *Pgm* causes the D_N of the longest branch to drop to 0.004; between-host differentiation then disappears (Fig. 2d).

Discussion

The larch and pine forms of *Z. diniana* could be: (i) subpopulations of a panmictic species; (ii) host races which hybridize; or (iii) good species which do not hybridize. Our results confirm that host-associated biotypes of *Z. diniana* are genetically differentiated (Fig. 2 and Table 4). The strong differences at *Mdh-s, Idh-s* and *Pgm* throughout Western Europe indicate that gene flow between larch and pine forms is not sufficient to homogenize gene frequencies. These results, coupled with the known biological differences between host biotypes (larval colour, host-related performance, mating behaviour, pheromones, phenology; see Introduction) firmly exclude panmixia but do not distinguish whether the two forms are good species or whether there is still gene flow between them.

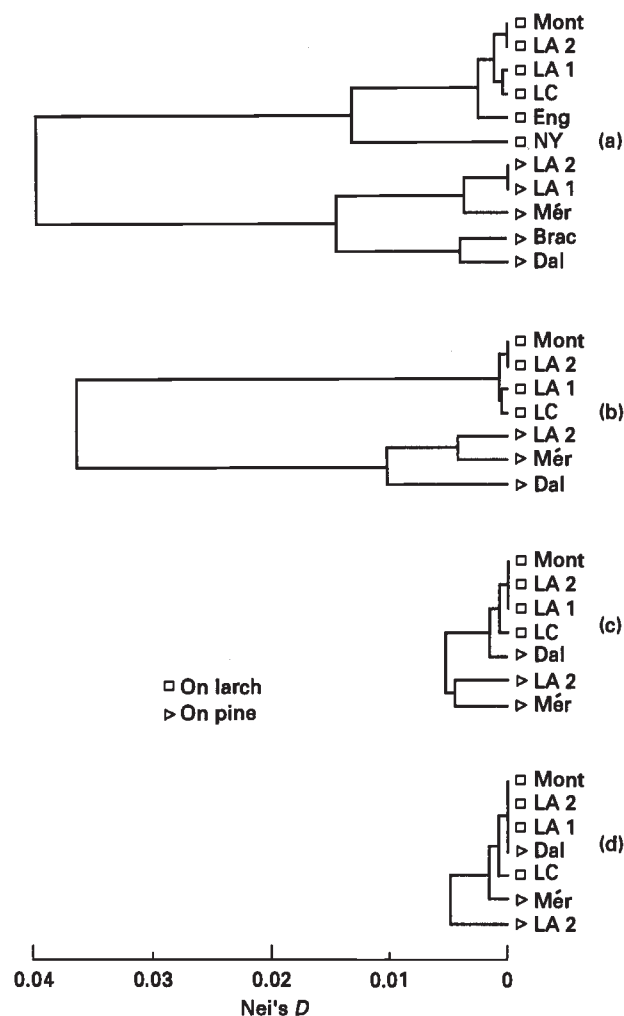On the one hand, the absence of significant genetic differences at most loci examined suggests

Fig. 2 UPGMA clustering based on Nei's (1978) unbiased genetic distance. (a) 11 samples, using all 24 loci. (b) 7 large samples with sample sizes more than 30 genomes, 24 loci. (c) 7 large samples, 22 loci (*Mdh-s*, *Idh-s* excluded); (d) 7 large samples, 21 loci (*Mdh-s*, *Idh-s*, and *Pgm* excluded). The removal of the three differentiated loci between B and D destroys the coherence of the larch- and pine-associated clusters.

the possibility of strong gene flow which homogenizes genetic variation between hosts. Most of the genetic differentiation results from *Mdh-s*, coupled with small differences at other loci. If the two host-associated populations were separated, and are now drifting apart, we can get a rough idea of how likely a strongly divergent allele frequency like this would be. The overall $F_{ST} \approx 0.065$ between hosts (Table 4) gives for an allele with frequency 0.803 a standard deviation of allele frequency of about 0.103. The actual allele frequencies of the *1.00* allele at *Mdh-s* are 0.803 on larch and 0.051 on pine, so its frequency has changed 0.752 in total, about 7.4 standard deviations. Assuming that, for low levels of

genetic drift without fixation, allele frequencies are approximately normally distributed around their starting value, this has an extremely low probability of $\ll 10^{-10}$. It is more likely (but still highly improbable) that both populations drifted from some intermediate frequency, say $[0.051+0.803]/2$, giving a standard deviation of 0.126 and drift of 3.0 standard deviations in each direction, which has probability $(0.0027)^2 \approx 7 \times 10^{-6}$. Even though there are a number of alleles at other loci free to drift, it is therefore virtually impossible to generate this much drift at any locus from the observed overall F_{ST}. We can therefore reject the idea that all of the allozyme loci are drifting apart at an equal rate; it is much more likely that *Mdh-s* is under strong disruptive selection whereas other loci are maintained at similar frequencies by gene flow.

Whereas this explanation is possible, the abundance of biological traits associated with the genetic differentiation (see Introduction) seems to preclude panmictic gene flow over much of the genome. Another reasonable possibility is that the host biotypes are good species between which there is no gene flow. The genetic similarities of the majority of loci would then result from recency of separation, with the strong differences at *Mdh-s*, *Idh-s* and in biological traits being from stronger disruptive selection on some parts of the genome. While we do not believe species should be defined on the basis of genetic distance, it is worth noting that genetic distance between host biotypes ($D_N = 0.04$) falls on the low end of typical genetic distances between sympatric species of Lepidoptera, and on the high end of genetic distances between geographic races (Fig. 3); evidence from genetic distance is therefore ambiguous. Field evidence of hybrid larvae (see Introduction) rests entirely on colour patterns of larvae and may also be unreliable. Studies in progress on mating isolation and gene flow in the field should resolve this issue. This work will be especially interesting in view of the possibility that host-associated differentiation contributes to the changes in morph frequency during build-up and crash of the populations on larch (Baltensweiler, 1993).

Regardless of the taxonomic status of genetic differentiation, another interesting question is whether these host races or sibling species have evolved in sympatry. Traditional dogma states that species-level differentiation can evolve only in allopatry (Mayr, 1970). However, there have been unprecedented challenges to this view recently (Bush, 1994). While Pleistocene refugia containing only larch and only pine could have been present, it is worth asking whether this would be likely, or even necessary for
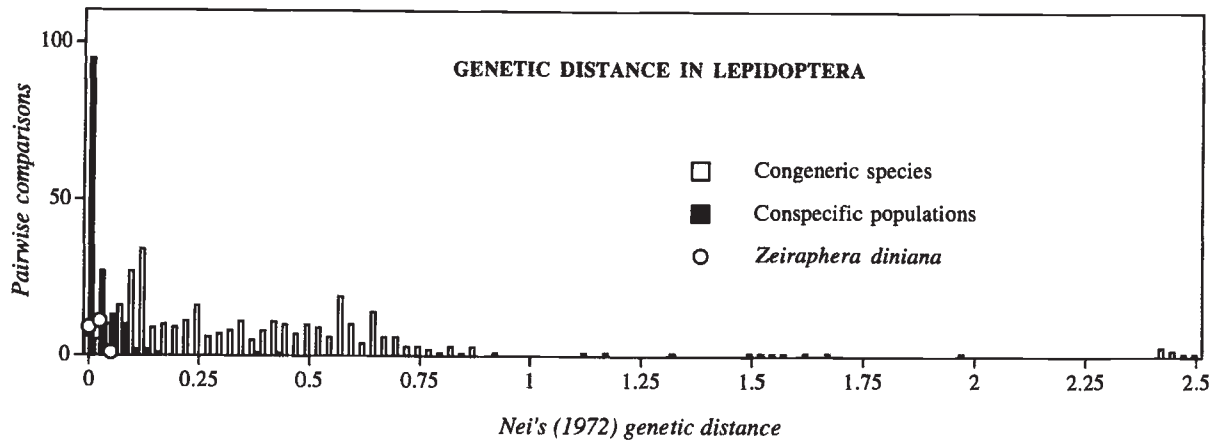
Fig. 3 Genetic distance in Lepidoptera. 65 species, 26 genera, 9 families. 344 pairwise comparisons of congeneric species, 152 of conspecific populations and 21 of *Z.diniana* populations. Compiled from original and published data; references are available from authors upon request.

the differentiation observed. In present-day Europe and during the last 13 000 years there have been plenty of pine forests without larch, but wherever larch is present, pine (particularly *Pinus cembra*) is intermixed (Huntley & Birks, 1983). Therefore it seems unlikely that there has ever been larch-only refugia in Europe. Even if, in the past, areas existed in which larch forms could develop in isolation, they would need to be many hundreds of km from pine populations as *Z. diniana* is well-known for mass-migrations over these kinds of distances (Baltens-weiler *et al.*, 1977).

As an alternative to slow, allopatric speciation in refugia, chance colonization of a refugium might lead to rapid host-associated speciation via a founder effect (Mayr, 1970). A founder event could thus provide host adaptation resistant to subsequent gene flow. However, this model is problematic; one of the main arguments against founder effect speciation in Hawaiian *Drosophila*, for example, has been the complete absence of any genetic evidence for population bottlenecks (Barton & Charlesworth, 1984). The same is true for the species under study here: the overall average heterozygosity of the larch form is 0.183, that of the pine form is 0.173 (Table 2). These values are similar to, although somewhat higher than, those found generally in species of Lepidoptera (0.120–0.145; see Menken, 1982; Graur, 1985; Emelianov & Goncharenko, 1992), and suggest that there has not been a bottleneck in the formation of either host biotype (see Mallet *et al.* (1993) for evidence of bottlenecks in other species). This lack of reduction in heterozygosity might be explained by hybridization after speciation occurred; but, if this is so, our data show that the host-associated forms can remain differentiated in sympatry in

the face of gene flow.

In contrast, divergence could easily have been sympatric. Pine shoots become acceptable to larvae (and pine form larvae hatch) about a month after larch acceptability (and hatch of larch form larvae) (Day, 1984); adult emergences are similarly staggered. This temporal difference ensures that larch-feeding and pine-feeding populations mate to some extent assortatively, leading to accumulations of larch-adapted alleles in early populations and pine-adapted alleles in late populations (see Butlin, 1990). In *Zeiraphera*, sympatric differentiation is at least as easy to contemplate as allopatric differentiation.

Acknowledgements

References

BALTENSWEILER, W. 1977. Colour-polymorphism and dynamics of larch bud moth populations (*Zeiraphera diniana* Gn., Lep. Tortricidae). *Mitteil. Schweiz. Entomol. Gesellsch.*, **50**, 15–23.

BALTENSWEILER, W. 1993. A contribution to the explanation of the larch bud moth cycle, the polymorphic fitness hypothesis. *Oecologia*, **93**, 251–255.

BALTENSWEILER, W., BENZ, G., BOVEY, P. AND DELUCCHI, V. 1977. Dynamics of larch bud moth populations. *Ann. Rev. Ent.*, **22**, 79–100.

BALTENSWEILER, W. AND PRIESNER, E. 1988. Studien zum

Pheromon Polymorphismus von *Zeiraphera diniana* Gn. (Lep., Tortricidae). 3. Anflugspezifität männlicher Falter zweier Wirtsrassen an synthetische Pheromonquellen. *J. Appl. Entomol.*, **106**, 217–231.

BARTON, N. H. AND CHARLESWORTH, B. 1984. Genetic revolutions, founder effects and speciation. *Ann. Rev. Ecol. Syst.*, **15**, 133–164.

BERRYMAN, A. A. 1986. *Forest Insects. Principles and Practice of Population Management.* Plenum Press, New York.

BOVEY, P. AND MAKSYMOV, J. K. 1959. Le problème des races biologiques chez la tordeuse grise du mélèze *Zeiraphera griseana* (Hb.). Note préliminaire. *Vierteljahrsschrift Naturforsch. Ges. Zürich*, **104**, 264–274.

BUSH, G. L. 1994. Sympatric speciation in animals: new wine in old bottles. *Trends Ecol. Evol.*, **9**, 285–288.

BUSH, G. L., FEDER, J. L., BERLOCHER, S. H., McPHERON, B. A., SMITH, D. C. AND CHILCOTE, C. A. 1989. Sympatric origins of *R. pomonella*. *Nature*, **339**, 346.

BUTLIN, R. 1990. Divergence of emergence time of host races due to differential gene flow. *Heredity*, **65**, 47–50.

DAY, K. 1984. Phenology, polymorphism and insect-plant relationships of the larch budmoth, *Zeiraphera diniana* (Guenée) (Lepidoptera: Tortricidae), on alternative conifer hosts in Britain. *Bull. ent. Res.*, **74**, 47–64.

DIEHL, S. R. AND BUSH, G. L. 1984. An evolutionary and applied perspective of insect biotypes. *Ann. Rev. Ent.*, **29**, 471–504.

EMELIANOV, I. M. AND GONCHARENKO, G. G. 1992. Genetic variation, population structure and gene flow in the pine lappet moth *Dendrolimus pini* L. (Lepidoptera: Lasiocampidae) in forests of Byelorussia and Ukraine. *Genetika*, **28**, 72–84.

FEDER, J. L., CHILCOTE, C. A. AND BUSH, G. L. 1988. Genetic differentiation between sympatric host races of the apple maggot fly *Rhagoletis pomonella*. *Nature*, **336**, 61–64.

FEDER, J. L., OPP, S. B., WLAZLO, B., REYNOLDS, K., GO, W. AND SPISAK, S. 1994. Host fidelity is an effective premating barrier between sympatric races of the apple maggot fly. *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 7990–7994.

FUTUYMA, D. J. 1983. Evolutionary interactions among herbivorous insects and plants. In: Futuyma, D.J. and Slatkin, M. (eds) *Coevolution*, pp. 207–231 Sinauer, Sunderland, MA.

GRANT, B. R. AND GRANT, P. R. 1979. Darwin's finches: population variation and sympatric speciation. *Proc. Natl. Acad. Sci., U.S.A.*, **76**, 2359–2363.

GRAUR, D. 1985. Gene diversity in Hymenoptera. *Evolution*, **39**, 190–199.

GUERIN, P. M., BALTENSWEILER, W., ARN, H. AND BUSER, H. R. 1984. Host race pheromone polymorphism in the larch budmoth. *Experientia*, **40**, 892–894.

GUTTMAN, S. I., WOOD, T. K. AND KARLIN, A. A. 1981. Genetic differentiation among host plant lines in the sympatric *Echenopa binotata* Say complex (Homoptera: Membracidae). *Evolution*, **35**, 205–217.

HUNTLEY, B. AND BIRKS, M. J. B. 1983. *An Atlas of Past and Present Pollen Maps for Europe 0–13,000 Years Ago.*

Cambridge University Press, Cambridge.

JAENIKE, J. AND SELANDER, R. K. 1980. On the question of host races in the fall webworm, *Hyphantria cunea*. *Entomologia exp. appl.*, **27**, 31–37.

MAKSYMOV, J. K. 1959. Beitrag zur Biologie und Ökologie des Grauen Lärchenwicklers *Zeiraphera griseana* (Hb.) (Lepidoptera, Tortricidae) im Engadin. *Mitt. Schweiz. Anst. forstl. Versuchw.*, **35**, 277–315.

MALLET, J., KORMAN, A., HECKEL, D. G. AND KING, P. 1993. Biochemical genetics of *Heliothis* and *Helicoverpa* (Lepidoptera: Noctuidae) and evidence for a founder event in *Helicoverpa zea*. *Ann. Entomol. Soc. Am.*, **86**, 189–197.

MAYR, E.. 1970. *Populations, Species, and Evolution.* Harvard University Press, Cambridge, MA.

McPHERON, B. A., SMITH, D. C. AND BERLOCHER, S. H. 1988. Genetic differences between host races of *Rhagoletis pomonella*. *Nature*, **336**, 64–66.

MENKEN, S. B. J. 1981. Host races and sympatric speciation in small ermine moths, Yponomeutidae. *Entomologia exp. appl.*, **30**, 280–292.

MENKEN, S. B. 1982. Biochemical genetics and systematics of small ermine moths (Lepidoptera, Yponomeutidae). *Z. Zool. Syst. Evolut. -forsch.*, **20**, 131–143.

MITTER, C. AND FUTUYMA, D. J. 1979. Population genetic consequences of feeding habits in some forest Lepidoptera. *Genetics*, **32**, 1005–1021.

NEI, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, **89**, 583–590.

PASHLEY, D. P. 1988. Current status of fall armyworm host strains. *Flor. Entomol.*, **71**, 228–234.

PRIESNER, E. 1979. Specificity studies on pheromone receptors of noctuid and tortricid Lepidoptera. In: Ritter, F.J. (ed.) *Chemical Ecology: Odour Communication in Animals*, pp. 57–71. Elsevier, Amsterdam.

RICHARDSON, B. J., BAVERSTOCK, P. R. AND ADAMS, M. 1986. *Allozyme Electrophoresis. A Handbook for Animal Systematics and Population Studies.* Academic Press, London.

SMITH, D. C. 1988. Heritable divergence of *Rhagoletis pomonella* host races by seasonal asynchrony. *Nature*, **336**, 66–67.

SWOFFORD, D. L. AND SELANDER, R. B. 1989. *BIOSYS-1. A Computer Program for the Analysis of Allelic Variation in Population Genetics and Biochemical Systematics*, Release 1.7. Illinois Natural History Survey, Champaign, IL.

VARLEY, G. C., GRADWELL, G. R. AND HASSELL, M. P. 1973. *Insect Population Ecology. An Analytical Approach.* Blackwell Scientific Publications, Oxford.

VERDON, B. 1978. Genetische Untersuchungen über Isozyme im Zusammenhang mit dem gradologischen Verhalten des Lärchenwicklers *Zeiraphera diniana* (Gn.) (Lep., Tortricidae). Ph.D. Dissertation, Eidgenössischen Technischen Hochschule, Zürich.

WRIGHT, S. 1978. *Evolution and the Genetics of Populations*, vol. 4, *Variability Within and Among Natural Populations.* University of Chicago Press, Chicago.