THE ROYAL SOCIETY

Received 8 July 2003
Accepted 1 September 2003
Published online 20 November 2003

Genomic evidence for divergence with gene flow in host races of the larch budmoth

Igor Emelianov[1*], František Marec[2] and James Mallet[3]

[1]*Plant and Invertebrate Ecology Division, Rothamsted Research, Harpenden AL5 2JQ, UK*
[2]*Institute of Entomology, Czech Academy of Sciences, Ceské Budejovice CZ-370 05, Czech Republic*
[3]*Galton Laboratory, Department of Biology, University College London, London NW1 2HE, UK*

Ecological divergence in the face of gene flow has recently become implicated as a potentially important cause of speciation and adaptive radiation. Here, we develop a genomic approach to test for divergent selection in sympatric host races of the larch budmoth *Zeiraphera diniana* (Lepidoptera: Tortricidae). We analysed hundreds of amplified fragment length polymorphism markers in 92 individuals in sympatric and allopatric populations, and in two backcross broods used to map the markers to individual chromosomes. The results directly confirm the existence of natural hybridization and demonstrate strong heterogeneity between chromosomes in terms of molecular divergence between host races (the average level of divergence was $F_{ST} = 0.216$). However, genomic heterogeneity was not found when we analysed divergence between geographically separated populations of the same host race. We conclude that the variance of the level of sympatric divergence among chromosomes is the footprint of divergent selection acting on a few linkage groups, combined with appreciable gene flow that homogenizes between-race variation at the remaining linkage groups. These results, coupled with other recent multilocus analyses of sister species pairs, demonstrate that selection-driven sympatric phase of genetic divergence in the presence of gene flow is a likely feature of speciation.

Keywords: speciation; gene flow; selection; genomic analysis; larch budmoth

1. INTRODUCTION

The role of ecological divergence in speciation and adaptive radiation has been much discussed recently (Schluter 2000). Ecological divergence could be extremely important if it enables the genetic integrity of diverging populations to be maintained and enhanced during the crucial phase before reproductive isolation is complete. Recent theory suggests that selection driving ecological divergence can achieve just that (Dieckmann & Doebeli 1999; Kondrashov & Kondrashov 1999; Kirkpatrick & Ravigné 2002), but empirical evidence is still scarce (Turelli *et al.* 2001; Via 2001).

Selection acting to maintain genetic divergence of habitat-specific forms can be inferred from patterns of genetic variation across the genome. Theory predicts that loci linked to regions under divergent selection will hitchhike towards fixation of different alleles even in the face of gene flow (Maynard Smith & Haigh 1974; Barton & Bengtsson 1986; Charlesworth *et al.* 1997; Fey & Wu 2000). When populations associated with different habitats hybridize, between-habitat variation at neutral markers will tend to be retained in genomic regions linked to loci under selection, even though the rest of the genome may become homogenized. Genome-wide analysis of between-habitat variance at neutral markers might, therefore, provide useful evidence about the role of ecological divergence in speciation. In cases where selection drives and maintains genetic divergence of hybridizing populations, we expect heterogeneity characterized by a clustered genomic distribution of between-habitat differences. By contrast, divergence in geographical isolation, or under conditions of complete reproductive isolation, should produce a more random genomic distribution of habitat-specific alleles. Heterogeneity of molecular divergence has recently been demonstrated in genomes of *Helianthus* (Rieseberg *et al.* 1999), *Drosophila* (Ting *et al.* 2000; Kliman *et al.* 2000; Wu 2001; Machado *et al.* 2002; Machado & Hey 2003) and between *Homo* and *Pan* (Navarro & Barton 2003). These studies found significant heterogeneity of divergence between genomic regions, which implies some between-species gene flow after speciation. However, in these cases, it is unknown whether speciation was driven by ecological divergence, or whether other factors of disruptive selection such as sterility and non-viability of hybrids was responsible for the observed genomic heterogeneity.

The power of natural selection to drive speciation of ecologically diversified populations in the absence of postmating incompatibility can be unequivocally demonstrated only by genomic analysis of ecologically differentiated incipient species that produce viable and fertile hybrids in nature, and where genomic heterogeneity can reasonably be attributed to adaptive divergence with gene flow. To our knowledge, no genomic studies of this type have yet been made. Insect herbivores provide good models of such a process (Jaenike 1981; Diehl & Bush 1984): several species have differentiated ecologically and genetically, sympatric and potentially hybridizing 'host races' associated with different plants (Via 2001; Berlocher & Feder 2002; Drès & Mallet 2002). Host races of Lepidoptera are especially attractive model systems, because analysis of genomic heterogeneity is particularly tractable, for the following reasons. Achiasmatic meiosis in Lepidoptera females (Marec & Traut 1993) allows unambiguous

* Author for correspondence (igor.emelianov@bbsrc.ac.uk).

assignment of markers to individual chromosomes even in relatively small mapping broods. In addition, selection-induced hitchhiking of neutral markers is expected to occur in regions spanning large chromosomal segments (Charlesworth *et al.* 1997), potentially leading to substantial heterogeneity among chromosomes and greatly simplifying tests of genomic heterogeneity, especially when each chromosome is small relative to the total genome size, as in Lepidoptera. The relatively large number of chromosomes in Lepidoptera (the mode is 31; see Robinson 1971) gives ample statistical power for such a test.

We studied larch- and pine-feeding host races of the larch budmoth, *Zeiraphera diniana* (Lepidoptera, Tortricidae). These host races cross freely in the laboratory and are commonly found in sympatry in nature (Emelianov *et al.* 1995, 2001, 2003). Despite probable hybridization, races remain strongly differentiated at genetic marker loci (Emelianov *et al.* 1995), as well as in traits involved in host utilization (Bovey & Maksymov 1959; Day 1984; Emelianov *et al.* 2003), morphology (particularly coloration of larvae; see Baltensweiler 1993) and mate choice via differentiated female pheromones (Baltensweiler *et al.* 1978; Priesner & Baltensweiler 1987; Emelianov *et al.* 2001). If these racial differences are maintained by selection in the face of continuous gene exchange, we expect an excess of differences between races on chromosomes under divergent selection, and a deficit of divergence elsewhere in the genome. However, intrinsic differences between chromosomes in the underlying rates of mutation or molecular evolution during divergence might also cause heterogeneous patterns in complete geographical or reproductive isolation. These would be difficult to distinguish from those due to heterogeneity in selection and gene flow (Hofker *et al.* 1986; Kliman *et al.* 2000; Filatov & Charlesworth 2002; Machado *et al.* 2002; Navarro & Barton 2003). We control for intrinsic chromosomal divergence heterogeneity by comparing patterns of divergence between hosts in sympatry and pattern of divergence between geographically isolated populations on the same host (cf. the similar use of *Drosophila bogotana* as a geographically isolated control for studies of divergence heterogeneity across the genomes of the sympatric *D. pseudoobscura* and *D. persimilis*; Machado & Hey 2003). We also test directly for the existence of natural hybridization in sympatric populations.

Various outcomes are possible, depending on the relative roles of divergent selection and of intrinsic chromosomal divergence heterogeneity (table 1). Only if divergence is driven and maintained by selection in the face of gene flow, and if it outweighs any intrinsic genomic heterogeneity in rates of molecular marker divergence, can we expect the second outcome at column 1 of table 1.

Specifically, we ask the following questions.

(i) Is there direct evidence of hybridization between larch and pine races in the wild?
(ii) Is there any heterogeneity across the genome in levels of molecular marker divergence between host races in sympatry?
(iii) Is there a similar pattern of heterogeneity between geographically isolated populations on the same host?

2. MATERIAL AND METHODS

(a) *Samples from natural populations*

We obtained data from multiple polymorphic genetic markers in a total of 92 larch budmoth larvae from five populations. Out of these, 84 larvae were collected from larch and pine trees in mixed larch–pine forests in the east and in the west of the Alps. In the east (Upper Engadine, Switzerland, 46°30′ N, 9°53′ E) we analysed 60 individuals, 36 from larch and 24 from pine. In the west (Briançon, France, 44°50′ N, 6°39′ E) 24 individuals were analysed, 12 from larch and 12 from pine. We also analysed eight pine-feeding larvae in a larch–pine forest in the centre of the Sredinny Khrebet mountain ridge, Kamchatka peninsula, 55°00′ N, 158°00′ E.

(b) *Molecular marker analysis*

We analysed amplified fragment length polymorphism (AFLP) and allozyme markers, as described by Vos *et al.* (1995) and Emelianov *et al.* (1995), respectively. In our study, AFLP markers were generated using 31 primer combinations, all based on a pair of core primer sequences, 5′-GACTGCGTA CCAATTC-3′ and 5′-GATGAGTCCTGAGTAA-3′, but differing in a 2–3 nucleotide 'selective' extension at the 3′-end. This technique allows screening of many hundreds of polymorphic loci. Each polymorphic AFLP locus in this study had two alleles, presence of a band on the gel (dominant allele) and absence of a band (recessive allele), and three possible genotypes (presence/presence and absence/absence homozygotes, and presence/absence heterozygotes). Allozyme markers were limited to *Mdh*, *Pgm* and *Idh*; all three are known to have different levels of differentiation between host races (Emelianov *et al.* 1995).

(c) *Linkage mapping*

Linkage analysis was used to determine the number of linkage groups and the number of differentiated and non-differentiated loci in each group. We used a standard backcross design (Manly & Olson 1999). Segregating markers were analysed in grandparents, parents and in two unrelated backcross broods, BC1 ([larch female × pine male] mother × [larch × larch] father, family size: 26) and BC2 ([pine female × larch male] mother × [pine × pine] father, family size: 21). Genotypes of F_1 parents of these broods were established by analysing the same marker loci in pure-race grandparents and verified by analysing the segregation pattern in backcrosses. Achiasmatic meiosis in *Z. diniana* females means that maternally heterozygous backcross markers can be unambiguously assigned to linkage groups even in relatively small broods. Therefore, only markers with a presence/absence genotype in the mother and absence/absence genotype in the father were used in this analysis. A lack of recombination in females resulted in a lack of information on locus order within linkage groups. However, this limitation did not affect our objective, because we were interested in studying the variance of divergence among chromosomes rather than among segments within chromosomes.

Markers of the same molecular mass generated by the same pair of primers, but in different individuals, were assumed to be homologous. This assumption has been tested previously in an orthopteran by sequencing putative homologues (Parsons & Shaw 2001), but we tested the assumption within our study because correct assignment of divergence values to mapped markers depends on the homology of loci analysed in natural populations and in mapping broods. The frequency of false homology in our study was estimated by performing 129 pairwise linkage

Table 1. Four possible outcomes of a test for ecologically driven divergence between sympatric ecological races.
(Only the case in bold type would be strongly indicative of divergence with gene flow.)

	chromosomes lacking an intrinsic tendency to accumulate divergence at different rates	chromosomes intrinsically heterogeneous in accumulation of divergence[a]
divergence without gene flow, and driven by neutral evolution	hybridization test: negative divergence in sympatry: not heterogeneous across the genome little similarity between replicates little similarity in genomic pattern to that between allopatric populations from the same habitat types (the latter is not heterogeneous)	hybridization test: negative divergence in sympatry: heterogeneous across the genome similar between replicates similar genomic pattern to that between allopatric populations from the same habitat types
divergence maintained by divergent selection, in the face of current gene flow	**hybridization test:** **positive** **divergence in sympatry:** **heterogeneous across the genome** **similar genomic pattern between replicates** **little similarity to that between allopatric populations from the same habitat (the latter not heterogeneous in genomic pattern)**	hybridization test: positive divergence in sympatry: heterogeneous across the genome similar genomic pattern between replicates similar to that between allopatric populations from the same habitat types

[a] Intrinsic genomic heterogeneity in rates of divergence across the genome could be due to a number of causes. For example, heterogeneity of gene duplications and/or gene conversion and concerted evolution, heterogeneity of neutral molecular evolution due to heterogeneity of mutation rates.

tests between the broods. In each case, we tested for linkage between a pair of BC1 markers apparently homologous to a pair of markers known to be linked in BC2. Lack of linkage between putatatively homologous markers in BC1 meant that the homology of at least one of the markers must be false. Only one case of false homology was detected, giving an estimated frequency of false homology of 0.008.

An additional product of AFLP locus homology tests between broods was the establishment of homology between BC1 and BC2 linkage groups.

(d) *Cytogenetic analysis*

It was important to compare the results of our linkage analysis with karyology to determine whether linkage groups in our mapping broods corresponded to actual chromosomes. To count the number of chromosomes, squash preparations of mitotic and meiotic metaphase chromosomes were made from testes using the methods of Marec & Traut (1994). Chromosome spreads were also made from pachytene oocytes using ovaries obtained from fifth instar larvae using the methods developed by Traut (1976), stained according to Marec & Traut (1994).

(e) *Statistical analyses*

(i) *Analysis of amplified fragment length polymorphism marker divergence within and between races*

Only polymorphic loci were used for this analysis (loci were judged polymorphic if the frequency of the rarest allele in the global sample of 92 individuals was greater than 5%). To compare the between- and within-race variation of allele frequencies at each mapped polymorphic locus, we estimated between-sample standardized phenotype frequency variance (F_{ST}), by analysing samples taken from different hosts in the same area and from the same host but in different areas. Five such pairwise analyses were made: (i) east Alps larch race versus east Alps pine race; (ii) west Alps larch race versus west Alps pine race; (iii) east Alps larch race versus west Alps larch race; (iv) east Alps pine race versus west Alps pine race; and (v) the east Alps pine race versus Kamchatka pine race. Each pairwise comparison was replicated by using two independent sets of loci: the set mapped in BC1 and the set mapped in BC2. Thus, we were able to obtain, overall, four replicates comparing different host races from the same area, and six replicates comparing populations of the same host race from different areas. F_{ST} was partitioned by using analysis of molecular variance (AMOVA; Excoffier *et al.* 1992), and the significance of F_{ST} values was tested by using permutation (Excoffier *et al.* 1992; Schneider *et al.* 2000). Loci with $p < 0.05$ for F_{ST} were considered significantly differentiated. AFLP markers are dominant, and so we treated AFLP phenotypes as haplotypes. This procedure causes some loss of allelic information, but because similar numbers of dominant alleles were at high frequency on larch and on pine, these phenotypic F_{ST} values provide a valid measure of differentiation for the comparative purposes needed here.

(ii) *Identification of hybrids*

For each individual in our sample we estimated the posterior probability that it belongs to larch race, to pine race or to hybrid classes (F_1, F_2 or backcross) using a Bayesian method on the AFLP marker data. This procedure uses a Markov-chain Monte Carlo method and is implemented in the program 'NewHybrids' (Anderson & Thompson 2002).

(iii) *Genomic heterogeneity*

To discriminate between the effect of divergent selection with gene flow and intrinsic genomic heterogeneity we compared: (i) divergence between races in sympatry; and (ii) divergence between same-race populations in allopatry. Each type of test was replicated (see § 2e(i)). A $2 \times N$ G-test (i.e. a likelihood ratio test) of homogeneity was used to detect heterogeneity among chromosomes, examining numbers of differentiated ($p < 0.05$) and non-differentiated ($p \geqslant 0.05$) loci for each of the N larch budmoth chromosomes. Chromosomes that contributed significantly ($p < 0.05$) to heterogeneity were found by using jackknife-after-bootstrap of variance of q, where q is the fraction of differentiated loci on individual chromosomes (MathSoft Inc. 1999). To test whether average levels of F_{ST} per chromosome were consistent in each homologous chromosome between replicates, we estimated the correlation coefficient and its 95% confidence limits as follows. In each replicate, we averaged marker F_{ST} values for individual chromosomes that could be homologized between broods BC1 and BC2, and then used a bootstrap resampling technique (MathSoft Inc. 1999) to repeatedly calculate correlation coefficients. In our study, we have more than two replicates (four between-race replicates and six within-race replicates). To obtain global correlation coefficients for all within-race and all between-race replicates we used the same bootstrap re-sampling technique, except that the data this time consisted of all possible average pairwise F_{ST} values for each chromosome.

3. RESULTS

(a) *Divergence between and within races*

Each of the five population samples was analysed at up to 1291 polymorphic AFLP markers. Within-race geographical divergence was found both on a transcontinental and on a regional scale. Pine-feeding larvae from the Alps and from Kamchatka differed significantly at 158 out of 1187 polymorphic loci, with average $F_{ST} = 0.104$ ($p < 0.001$). Significant divergence between east and west Alps was found in the pine race (118 out of 1169 loci, average $F_{ST} = 0.057$, $p < 0.001$) and even in the highly migratory (Baltensweiler 1993) larch race (51 out of 1266 loci, average $F_{ST} = 0.005$, $p = 0.038$). Finally, sympatric pine and larch host races in the Alps differed significantly at 228 out of the 1291 loci (average $F_{ST} = 0.216$, $p < 0.001$; significance was estimated using 1000 permutations in each case). For these comparisons between host races, the dominant AFLP allele was most common in the larch race for 112 loci, and in the pine race for 116 loci. No markers were completely fixed for alternative AFLP genotypes between host races.

(b) *Identification of hybrids*

The 92 individuals are displayed in figure 1*a*, based on their genotypes at 112 larch and 116 pine AFLP markers. Figure 1*a* shows that the larch host race is readily separable from the sympatric pine race, and that the pine race from the Alps is differentiated from the Kamchatka pine race. Samples within host races from different parts of the Alps are not clearly differentiable. Any hybrids between host races will be intermediate in terms of number of loci showing the larch phenotype on this display, and they will also be heterozygous at more loci than pure races, so we expect them to have a higher fraction of divergent AFLP loci with at least one dominant ('presence') allele. The three individuals from the Alps indicated by letters a, b and c seem: (i) clearly intermediate in terms of numbers of larch alleles, and outside either larch or pine race clusters; and (ii) to have a generally higher fraction of 'presence'-containing loci; thus on both counts, they appear to

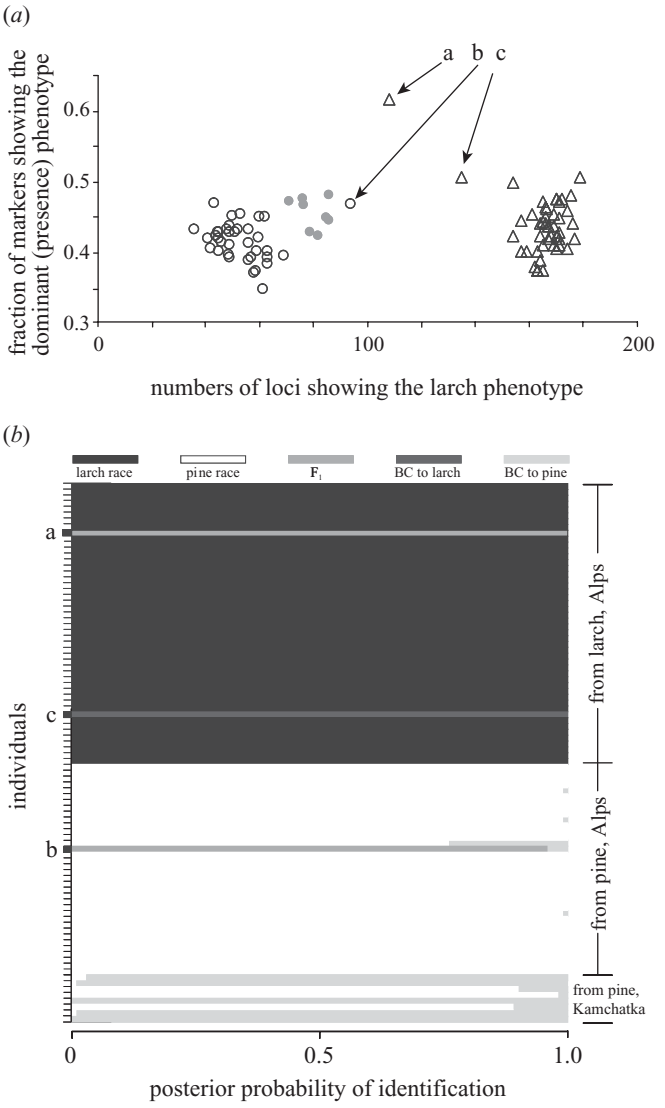

Figure 1. Evidence for hybridization in the larch budmoth. (*a*) The two-dimensional positions of 92 *Zeiraphera diniana* collected as larvae from larch (open triangles) or pine (open circles, Alps; filled circles, Kamchatka) are plotted on an axis representing individual heterozygosity (the fraction of markers showing the dominant, i.e. presence of band, phenotype) against an axis partitioning larch and pine races (a 'hybrid index' score representing the numbers of differentiated loci showing the larch phenotype). Hybrids should show up with intermediate hybrid index scores, but with elevated heterozygosity, and therefore high on the intercept. Arrows indicate putative hybrids a, b and c. (*b*) Bayesian posterior probabilities that the 92 individuals, represented by horizontal shaded bars, belong to parental races or various hybrid classes. The length of each bar reflects the fraction of permutations with corresponding diagnosis. For instance, individual b was diagnosed as an F_1 hybrid in 97% of permutations, but as a backcross to pine in 3% of permutations. Individuals a, b and c correspond to individuals a, b and c in (*a*).

be hybrids. Individual a appears in the greatly elevated position and is also almost exactly intermediate in terms of the number of larch alleles between larch and pine clusters, as expected for an F_1 hybrid. This individual female was collected as a larva on larch but had an unusual pale head capsule colour typical of the pine race. Dark head capsule colour is a recessive trait of the larch race, inherited at a single locus. Progeny of this individual, when crossed to a male with typically dark larch head coloration, segregated 1 : 1 for head capsule colour, as expected if she was heterozygous for this race-specific trait. Individual b appears to be a backcross to pine, whereas individual c appears to be a backcross to larch. Bayesian analysis of the 92 individuals using the NewHybrids algorithm based on the 228 between-race divergent markers produced similar results (figure 1*b*). Individuals a, b and c were identified, respectively, as: F_1 hybrid, F_1 hybrid (or possibly a backcross to pine), and backcross to larch. All other individuals from areas of sympatry were assigned to parental races. The divergent population on pine from Kamchatka, when examined using these diallelic markers, appeared to have a significant fraction of apparent 'larch race alleles' (figure 1*a*), and to have many individuals classed as 'backcross to pine' rather than pure pine race (figure 1*b*). However, these Kamchatkan results are certainly due to interpopulation differentiation rather than actual hybridization, because larch race is not known to occur in this area, and because the larch versus pine marker classifications were based only on populations from the Alps.

(c) *Linkage analysis*

We analysed individuals of each of the backcross broods at 1609 segregating loci, 887 of which were presence/absence heterozygous in the mother and absence/absence homozygous in the father. Out of the 887 maternally heterozygous markers, 347 were polymorphic in our population samples; 154 of these segregated only in BC1, 130 segregated only in BC2, and 63 segregated in both broods. Linkage analysis of maternally heterozygous markers produced 27 autosomal and two sex (Z and W) linkage groups in both BC1 and BC2. We were able to locate two or more of the 63 markers shared between BC1 and BC2 on Z and W sex chromosomes and on 19 autosomes homologous between the two broods. The remaining eight autosomes shared only one marker or none, so that pairwise homology tests were not possible.

(d) *Karyology*

Cytological analysis revealed 28 metaphase I bivalents (figure 2) and 56 mitotic metaphase elements (not shown) in spermatocytes. We also found 28 bivalents in pachytene nuclei of oocytes (not shown). The correspondence between linkage data (27 autosomal linkage groups + Z + W, see above and figures 3 and 4) and the cytological data show that linkage groups in our study correspond to chromosomes. *Zeiraphera diniana* thus appears to have $n = 28$ pairs of chromosomes.

(e) *Test for genomic heterogeneity*

Results of the test of genomic homogeneity are shown in figures 3 and 4. The four panels on figure 3 show F_{ST} values at loci segregating in either BC1 or BC2 and analysed in either eastern or western Alps areas of sympatry. Visual examination of this pattern, supplemented by the jackknife-after-bootstrap analysis shows that some chromosomes (marked by stars) have reached a far more advanced stage of divergence than others. The hypothesis of genomic homogeneity of divergence is strongly rejected: $G_{28} = 83.77$ (figure 3*a*); $G_{28} = 80.98$ (figure 3*b*); $G_{28} = 83.72$ (figure 3*c*); $G_{28} = 76.24$ (figure 3*d*); $p \ll 0.001$ in all four replicates. Strong divergence is consistently associated with the same set of chromosomes, regardless of the brood or the geographical site used for the analysis. The consistent non-randomness of the pattern is highlighted by a strong correlation of average chromosomal F_{ST} among replicates ($r = 0.740$ with 0.647–0.814 being the 95% confidence limits). The correlation remained significant ($r = 0.687$ (0.592–0.763)) even after removal of highly differentiated (starred) chromosomes from the analysis, confirming that repeatable heterogeneity of divergence between chromosomes is widespread across the genome. The non-randomness of the pattern in figure 3 is also highlighted by the fact that the number of mapped markers varies among linkage groups within replicates and covaries with average chromosomal F_{ST} ($r = 0.768$ (0.657–0.835)). The likelihood that a locus will segregate in a backcross brood increases with between-race differentiation at that locus, therefore such covariance is expected only when genomic distribution of divergence is clumped.

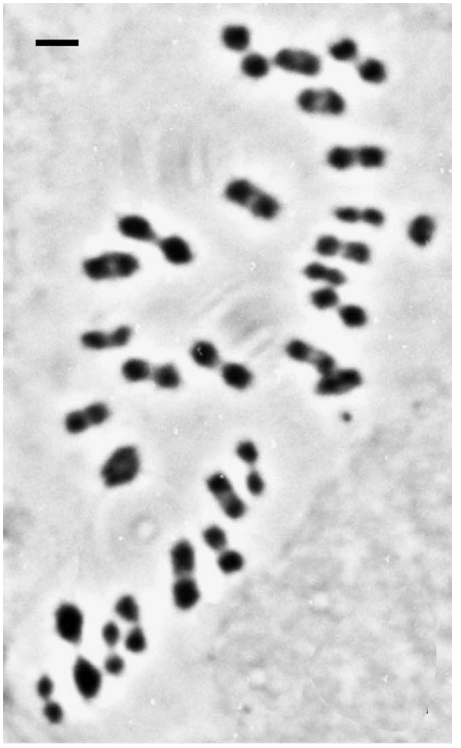

Figure 2. Light micrograph of a meiotic metaphase I spermatocyte nucleus from *Zeiraphera diniana* larvae (larch race), showing 28 bivalents. Scale bar, 2 μm.

The strongly heterogeneous genomic pattern revealed above is in stark contrast with the genomic distribution of geographical differences within each host race (figure 4). Loci differentiated between isolated same-race populations were not distributed heterogeneously among the 29 linkage groups either in BC1 or in BC2: $G_{28} = 23.44$,

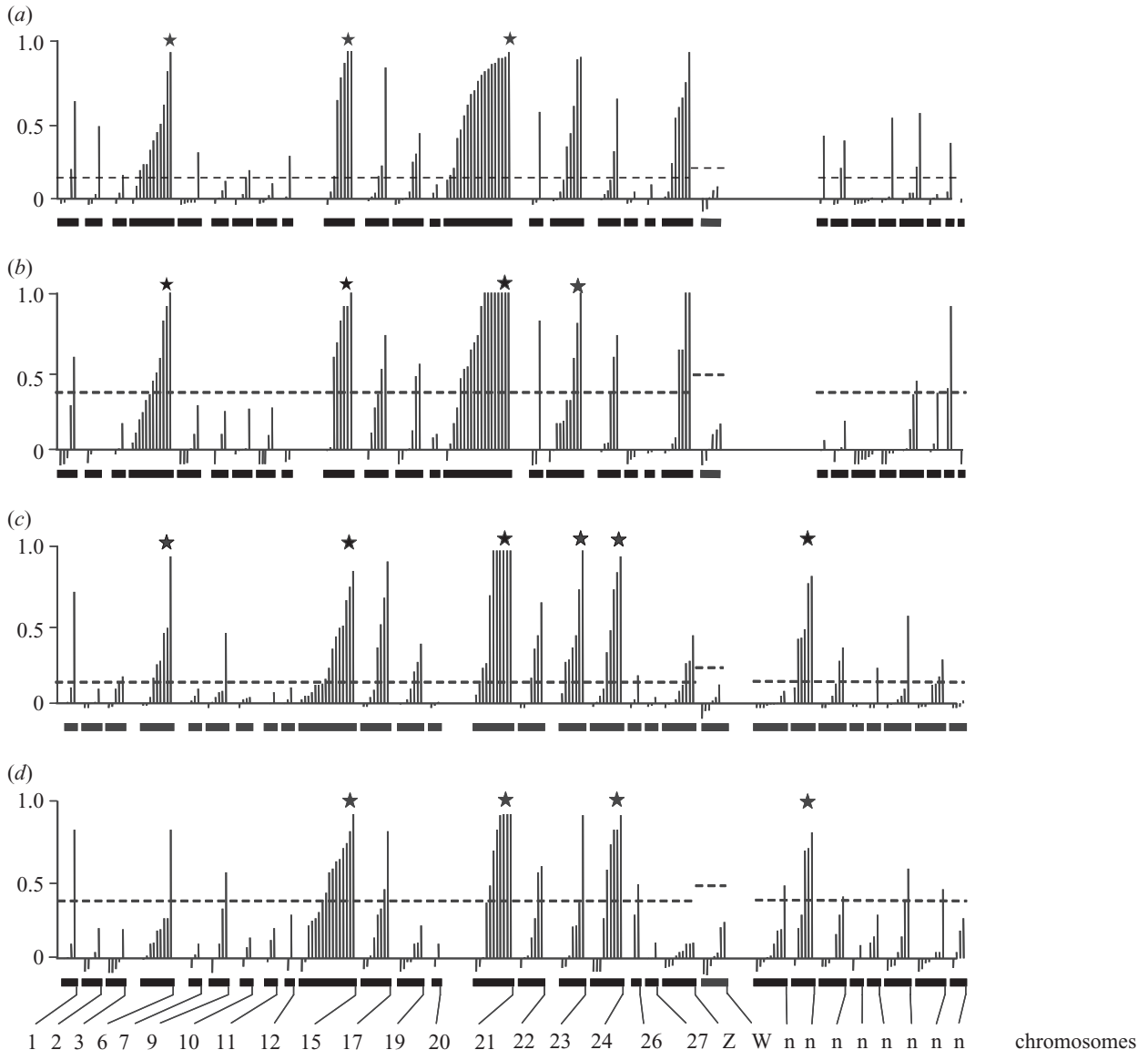

Figure 3. Between-race (sympatric) F_{ST} at loci across *Zeiraphera diniana* genome. (*a*)–(*d*) correspond to four replicates based on two different groups of loci segregating either in BC1 or BC2 and analysed in two different areas of sympatry: larch versus pine (*a*) eastern Alps, BC1, 152 loci; (*b*) western Alps, BC1, 154 loci; (*c*) eastern Alps, BC2, 186 loci; (*d*) western Alps, BC2, 179 loci. The height of vertical needles corresponds to between-race F_{ST} at individual loci. Chromosomes are shown in each replicate by a horizontal row of 29 solid bars. Between-brood homology has been established for the 19 numbered autosomes, as well as for W and Z sex chromosomes. Where between-brood homology could not be established, chromosomes are designated by the letter 'n'. Chromosomes contributing significantly to genomic heterogeneity are starred. Loci within groups are sorted by F_{ST} value and do not represent linkage order. Dashed lines show the significance cut-off level of F_{ST} ($p < 0.05$) for each AFLP band for the given sample size. W-linked markers have reduced sample size and an inflated cut-off level because they can be analysed only in females. After subtraction of error variance, slightly negative F_{ST} values can occur in the absence of population structure, and should be treated as zero. The sample contains two F_1 hybrids therefore no loci could have completely fixed divergence. However, strongly differentiated loci with only absence/absence genotypes in one race and presence/presence or presence/absence genotypes in another race are very likely to have presence/absence genotype in F_1. In some such cases, F_1 was collected from the host with a high frequency of 'presence' allele. F_{ST} in these cases assumes a value of 1, which is clearly a slight overestimate. The numbers of loci displayed differ across the four panels, because not all loci segregate in all broods and because not all loci were polymorphic at all sites. The 63 markers segregating in both broods are shown only on BC2 panels to ensure independence of the four replicates.

17.64, 33.61, 21.99, 23.18, 34.67, $p > 0.1$ in all six cases (figure 4*a–f*, respectively). In addition to the lack of heterogeneity of divergence within replicates, the chromosomal pattern of divergence is uncorrelated between replicates ($r = -0.008$ (-0.053–0.036)).

4. DISCUSSION

We use multiple markers in a genomic approach to study genetic divergence between two host races in sympatry, and between geographical populations of the same

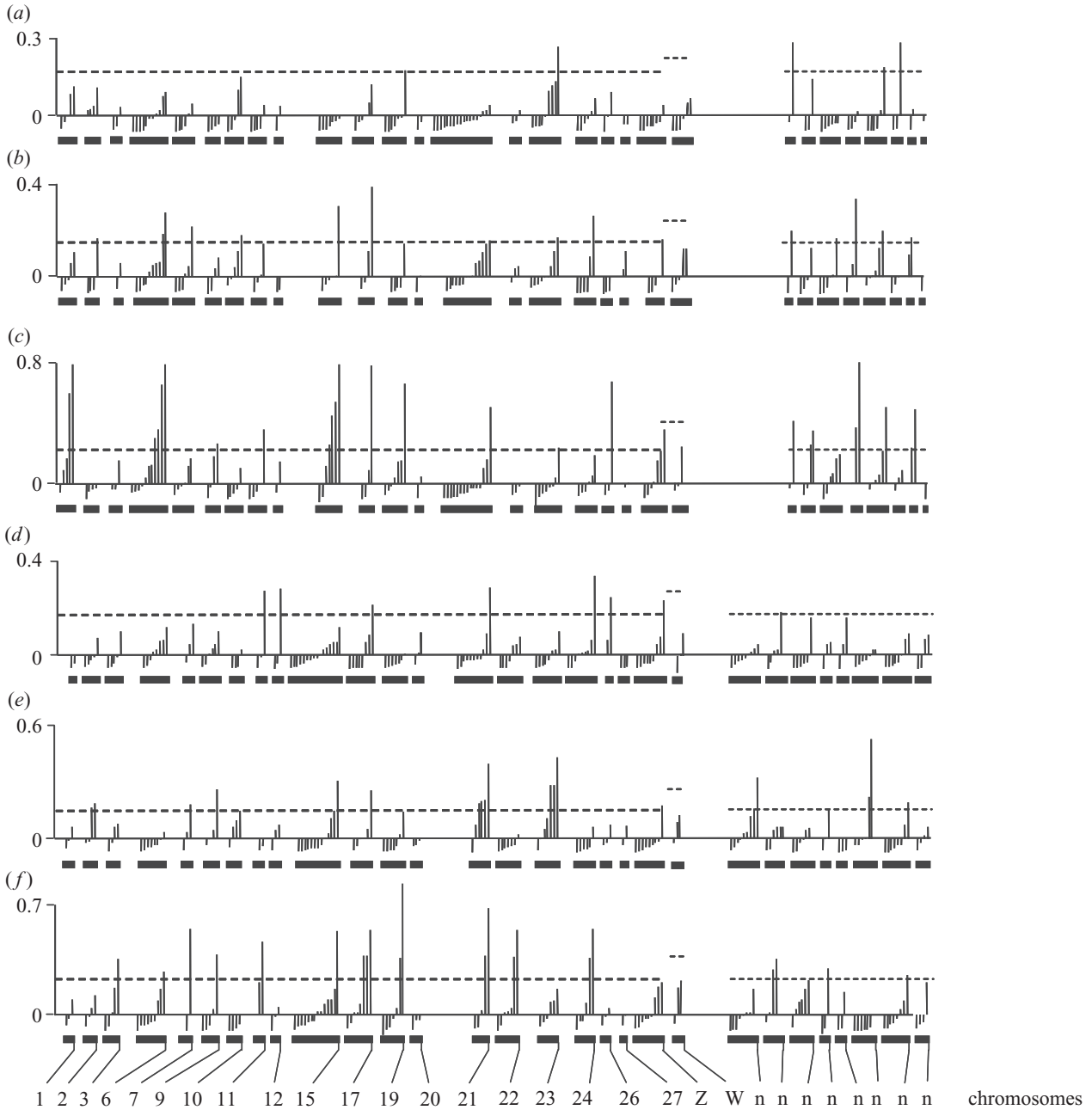

Figure 4. Within-race (geographical) F_{ST} at loci across *Zeiraphera diniana* genome. Panels correspond to six replicates based on two different groups of loci (segregating either in BC1 or BC2) analysed in three different geographical pairs of populations of the same host race. East versus west Alps: (*a*) larch, BC1, 147 loci; (*b*) pine, BC1, 132 loci; (*d*) larch, BC2, 177 loci; (*e*) pine, BC2, 158 loci. East Alps versus Kamchatka: (*c*) pine, BC1, 140 loci; (*f*) pine, BC2, 159 loci. The order of chromosomes and legend is as for figure 3.

host race. The pine and larch host races of *Zeiraphera diniana* were already known to be divergent at several morphological and molecular traits, and this study based on 1291 markers confirms this pattern for AFLP loci. The two forms are considered host races rather than full species because of the potential for hybridization. Among the 84 individuals collected in sympatric populations, we identified three individuals whose Bayesian posterior probability indicates that they are hybrids (figure 1). Our earlier behavioural studies indicate that the probability of hybridization between sympatric host races is *ca*. 2–3% (Emelianov *et al.* 2003). The current study provides direct evidence of hybrids, and a confirmation that earlier estimates of hybridization rate were sound.

We have found strong genomic heterogeneity of molecular divergence between host races in areas where hybridization occurs in sympatry (figure 3). However, there is no genomic heterogeneity in divergent geographical populations of the same host race (figure 4). This situation, represented by the bold type in table 1, strongly suggests that sympatric differentiation is maintained by selection, but that hybridization rates are sufficient to homogenize much of the genetic variation in genomic regions neutral in terms of host adaptation. The lack of a pronounced heterogeneous pattern of divergence in allopatric populations of the same host race (figure 4), in spite of sometimes substantial, and always significant, levels of geographical divergence ($F_{ST} = 0.005$–0.104), shows that

selection-driven divergence with gene flow explains the genomic heterogeneity we observe in sympatry better than other factors (see table 1).

Making inferences about the role of selection in divergence and speciation solely based on the patterns of divergence is not a simple task (Vitalis *et al.* 2001; Wilding *et al.* 2001) (see also recent genomic work on sibling species of *Drosophila* (Kliman *et al.* 2000; Machado *et al.* 2002; Machado & Hey 2003) and *Pan* versus *Homo* (Navarro & Barton 2003)). One of the main difficulties is that genomic heterogeneity similar to that observed here might be produced if chromosomes have intrinsically different rates of molecular evolution. The resulting non-random distribution of divergent AFLP loci might be due to systematic differences among chromosomes in the rate of mutation or rate of gene conversion between duplicated genes. However, our allopatric versus sympatric comparisons controlled for any intrinsic tendency towards chromosomal divergence heterogeneity, and we found none when we looked at geographically separated populations that did not differ ecologically. In addition, other types of marker display similar heterogeneity of differentiation. The strongly differentiated allozyme marker *Mdh* (Emelianov *et al.* 1995) maps to the strongly differentiated chromosome 6, whereas the mildly differentiated *Idh* and even less divergent *Pgm* (Emelianov *et al.* 1995) map, respectively, to the more weakly differentiated chromosomes Z and 11. Clearly, we cannot rule out the possibility that some genomic heterogeneity is due to intrinsic differences between chromosomes, but it seems clear that the very strong heterogeneity found only in sympatry must be explained chiefly by something other than intrinsic chromosomal differences.

Genomic heterogeneity in the larch budmoth appears to be largely due to strong divergence of a relatively small number of chromosomes, but the degree of divergence varies even within this small group of genomic segments (figure 3). It is likely that this is due to variation in the level of divergent selection on individual regions, but at least two other factors—linkage clustering of host-use genes, and association of these genes with rearranged sections of the genome—may also be involved. The contribution of a cluster of linked adaptive genes can be greater than the sum of contributions of individual genes due to nonlinear interactions between the amount of gene clustering and the width of flanking regions affected by selection-induced hitchhiking (Kelly 2000). Linkage clustering of adaptive epistatically interacting genes has been reported in various insects where divergence is accompanied by probable gene flow (Hawthorne & Via 2001; Naisbit *et al.* 2003), and may have some importance in *Zeiraphera*. Divergent selection on a gene embedded in a large rearranged region, for instance an inversion, will also make a disproportional contribution towards genomic heterogeneity, for the following reasons. Selection should produce faster divergence if epistatically interacting genes are associated with structural chromosomal rearrangements rather than with collinear chromosomal regions, as found in *Homo* versus *Pan* apes (Navarro & Barton 2003) and in species of *Drosophila* (Noor *et al.* 2001). In addition, all markers situated within rearranged regions are likely to recombine at a reduced rate and will hitchhike to fixation if at least one gene in the region is under selection.

At present, we cannot identify precisely the factors that cause the strong concentration of between-race divergence on a few *Zeiraphera* chromosomes (figure 3). Single genes may be under exceptionally strong divergent selection for host adaptation or other host-race specific effects (such as genes affecting divergent pheromonal communication), or there may be clusters of genes under divergent selection, and/or rearrangement of large chromosomal segments. However, it seems unlikely that these factors alone would cause such strong genomic heterogeneity without appreciable gene flow acting to homogenize, or partly homgenize, the remaining chromosomes. The heterogeneous genomic pattern is thus a footprint of selection in sympatric, hybridizing populations, which is driving and maintaining racial divergence in *Zeiraphera*, and which leads, through pleiotropy, to assortative mating (Emelianov *et al.* 2001, 2003) and ultimately to speciation. We cannot tell how long it will be before, or even if, gene flow between sympatric *Zeiraphera* races will cease completely. However, this study suggests that an exchange of genes within many genomic regions not under strong divergent selection is continuing long after separation into ecologically, genetically and morphologically differentiated races. Our results, coupled with those of other studies (Machado *et al.* 2002; Machado & Hey 2003; Navarro & Barton 2003) demonstrate that a long and relatively stable sympatric phase of genetic divergence in the presence of gene flow is a probable feature of speciation.

The authors thank John Kelly, Brian Charlesworth, Allen Orr and Andrew Pomiankowski for discussion, Andrei Gourov for help with field collections, and Fraser Simpson for laboratory assistance. This research was supported by a grant from the Biotechnology and Biological Sciences Research Council.

REFERENCES

Anderson, E. C. & Thompson, E. A. 2002 A model-based method for identifying species hybrids using multilocus genetic data. *Genetics* **160**, 1217–1229.

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

Baltensweiler, W., Priesner, E., Arn, H. & Delucci, V. 1978 Untershiedliche Sexuallockstoffe bei Larchen- und Arvenform des Grauen Larchenwicklers (*Zeiraphera diniana* Gn., Lep. Tortricidae). *Mitt. Schweiz. Entomol. Ges.* **51**, 133–142.

Barton, N. H. & Bengtsson, B. O. 1986 The barrier to genetic exchange between hybridising populations. *Heredity* **56**, 357–376.

Berlocher, S. H. & Feder, J. L. 2002 Sympatric speciation in phytophagous insects: moving beyond controversy? *A. Rev. Entomol.* **47**, 773–815.

Bovey, P. & Maksymov, J. K. 1959 Le problème des races biologiques chez la tordeuse grise du mélèze *Zeiraphera griseana* (Hb.). Note préliminaire. *Vierteljahresschr. Naturforsch. Ges. Zur.* **104**, 264–274.

Charlesworth, B., Nordborg, M. & Charlesworth, D. 1997 The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genet. Res.* **70**, 155–174.

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. Entomol. Res.* **74**, 47–64.

Dieckmann, U. & Doebeli, M. 1999 On the origin of species by sympatric speciation. *Nature* **400**, 354–357.

Diehl, S. R. & Bush, G. L. 1984 An evolutionary and applied perspective of insect biotypes. *A. Rev. Entomol.* **29**, 471–504.

Drès, M. & Mallet, J. 2002 Host races in plant-feeding insects and their importance in sympatric speciation. *Phil. Trans. R. Soc. Lond.* B **357**, 471–492. (DOI 10.1098/rstb.2002.1059.)

Emelianov, I., Mallet, J. & Baltensweiler, W. 1995 Genetic differentiation in *Zeiraphera diniana* (Lepidoptera, Tortricidae, the larch budmoth)—polymorphism, host races or sibling species? *Heredity* **75**, 416–424.

Emelianov, I. M., Drès, M., Baltensweiler, W. & Mallet, J. 2001 Host-induced assortative mating in host races of the larch budmoth. *Evolution* **55**, 2002–2010.

Emelianov, I., Simpson, F., Narang, P. & Mallet, J. 2003 Host choice promotes reproductive isolation between host races of the larch budmoth *Zeiraphera diniana*. *J. Evol. Biol.* **16**, 208–218.

Excoffier, L., Smouse, P. & Quattro, J. 1992 Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**, 479–491.

Fey, J. & Wu, C.-I. 2000 Hitchhiking under positive Darwinian selection. *Genetics* **155**, 1405–1413.

Filatov, D. A. & Charlesworth, D. 2002 Substitution rates in the X- and Y-linked genes of the plants, *Silene latifolia* and *Silene dioica*. *Mol. Biol. Evol.* **19**, 898–907.

Hawthorne, D. J. & Via, S. 2001 Genetic linkage of ecological specialization and reproductive isolation in pea aphids. *Nature* **412**, 904–907.

Hofker, M. H., Skraastad, M. I., Bergen, A. A. B., Wapenaar, M. C., Bakker, E., Millingtonward, A., Vanommen, G. J. B. & Pearson, P. L. 1986 The X-chromosome shows less genetic variation at restriction sites than the autosomes. *Am. J. Hum. Genet.* **39**, 438–451.

Jaenike, J. 1981 Criteria for ascertaining the existence of host races. *Am. Nat.* **117**, 830–834.

Kelly, J. K. 2000 Epistasis, linkage and balancing selection. In *Epistasis and evolutionary process* (ed. J. B. Wolf, E. D. Brodie & M. J. Wade), pp. 146–157. New York: Oxford University Press.

Kirkpatrick, M. & Ravigné, V. 2002 Speciation by natural and sexual selection: models and experiments. *Am. Nat.* **159**, S22–S35.

Kliman, R. M., Andolfatto, P., Coyne, J. A., Depaulis, F., Kreitman, M., Berry, A. J., McCarter, J., Wakeley, J. & Hey, J. 2000 The population genetics of the origin and divergence of the *Drosophila simulans* complex species. *Genetics* **156**, 1913–1931.

Kondrashov, A. S. & Kondrashov, F. A. 1999 Interactions among quantitative traits in the course of sympatric speciation. *Nature* **400**, 351–354.

Machado, C. A. & Hey, J. 2003 The causes of phylogenetic conflict in a classic *Drosophila* species group. *Proc. R. Soc. Lond.* B **270**, 1193–1202. (DOI 10.1098/rspb.2003.2333.)

Machado, C. A., Kliman, R. M., Markert, J. A. & Hey, J. 2002 Inferring the history of speciation from multilocus DNA sequence data: the case of *Drosophila pseudoobscura* and close relatives. *Mol. Biol. Evol.* **19**, 472–488.

Manly, K. F. & Olson, J. M. 1999 Overview of QTL mapping software and introduction to Map Manager QT. *Mamm. Genome* **10**, 327–334.

Marec, F. & Traut, W. 1993 Synaptonemal complexes in female and male meiotic prophase of *Ephestia kuehniella* (Lepidoptera). *Heredity* **71**, 394–404.

Marec, F. & Traut, W. 1994 Sex chromosome pairing and sex chromatin bodies in W–Z translocation strains of *Ephestia kuehniella* (Lepidoptera). *Genome* **37**, 426–435.

MathSoft Inc. 1999 *S-PLUS 2000 professional edition for Windows*. Release 2.

Maynard Smith, J. & Haigh, J. 1974 The hitch-hiking effect of a favorable gene. *Genet. Res.* **23**, 23–35.

Naisbit, R. E., Jiggins, C. D. & Mallet, J. 2003 Mimicry: developmental genes that contribute to speciation. *Evol. Dev.* **5**, 269–280.

Navarro, A. & Barton, N. H. 2003 Chromosomal speciation and molecular divergence—accelerated evolution in rearranged chromosomes. *Science* **300**, 321–324.

Noor, M. A. F., Grams, K. L., Bertucci, L. A. & Reiland, J. 2001 Chromosomal inversions and the reproductive isolation of species. *Proc. Natl Acad. Sci. USA* **98**, 12 084–12 088.

Parsons, Y. M. & Shaw, K. L. 2001 Species boundaries and genetic diversity among Hawaiian crickets of the genus *Laupala* identified using amplified fragment length polymorphism. *Mol. Ecol.* **10**, 1765–1772.

Priesner, E. & Baltensweiler, W. 1987 A study of pheromone polymorphism in *Zeiraphera diniana* Gn. (Lep., Tortricidae). 1. Male pheromonal response types in European wild populations, 1978–85. *J. Appl. Entomol.* **104**, 234–256.

Rieseberg, L. H., Whitton, J. & Gardner, K. 1999 Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics* **152**, 713–727.

Robinson, R. 1971 *Lepidoptera genetics*. Oxford: Pergamon Press.

Schluter, D. 2000 *The ecology of adaptive radiation*. New York: Oxford University Press.

Schneider, S., Roessli, D. & Excoffier, L. 2000 ARLEQUIN: a software for population genetics data analysis, v. 2.000. Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva.

Ting, T.-C., Tsaur, S.-C. & Wu, C.-I. 2000 The phylogenetic history at the *Odysseus* locus of reproductive isolation. *Proc. Natl Acad. Sci. USA* **97**, 5313–5316.

Traut, W. 1976 Pachytene mapping in the female silkworm, *Bombyx mori* L. (Lepidoptera). *Chromosoma* **58**, 275–284.

Turelli, M., Barton, N. H. & Coyne, J. A. 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.

Vitalis, R., Dawson, K. & Boursot, P. 2001 Interpretation of variation across marker loci as evidence of selection. *Genetics* **158**, 1811–1823.

Vos, P. (and 10 others) 1995 AFLP—a new technique for DNA-fingerprinting. *Nucleic Acids Res.* **23**, 4407–4414.

Wilding, C. S., Butlin, R. K. & Grahame, J. 2001 Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers. *J. Evol. Biol.* **14**, 611–619.

Wu, C.-I. 2001 The genic view of the process of speciation. *J. Evol. Biol.* **14**, 851–865.

As this paper exceeds the maximum length normally permitted, the authors have agreed to contribute to production costs.