Genet. Res., Camb. (2005), **85**, *pp.* 23–46. *With 7 figures.* © 2005 Cambridge University Press
doi:10.1017/S0016672305007317 Printed in the United Kingdom

Evolutionary history of the *Drosophila bipectinata* species complex

ARTYOM KOPP* AND OLGA BARMINA
Section of Evolution and Ecology and Center for Genetics and Development, University of California – Davis, Davis, CA 95616, USA

(*Received 9 August 2004 and in revised form 24 October 2004*)

Summary

Groups of recently diverged species offer invaluable glimpses into the history and genetic basis of speciation and phenotypic evolution. In this report, we combine phylogenetic and population-genetic approaches to reconstruct the evolutionary history of the *Drosophila bipectinata* species complex. This complex is a group of four closely related, largely sympatric species – *D. bipectinata*, *D. parabipectinata*, *D. malerkotliana* and *D. pseudoananassae*. Using the sequences of one mitochondrial and six nuclear loci, we show that *D. bipectinata* and *D. parabipectinata* are the two most closely related species, and that together with *D. malerkotliana* they form a monophyletic clade to which *D. pseudoananassae* is a relatively distant outgroup. Genetic divergence among *D. bipectinata*, *D. parabipectinata* and *D. malerkotliana* is extremely low, and we estimate that these species diverged only 283 000–385 000 years ago. We also find that mitochondrial DNA shows evidence of recent gene flow across species boundaries. Despite the low genetic divergence, species of the *bipectinata* complex show an unusually high degree of morphological differentiation. This contrast underscores the importance of understanding the genetic basis of functional differentiation among closely related species.

1. Introduction

Evolutionary studies in the genus *Drosophila* have made major contributions to our understanding of speciation and species divergence (Patterson & Stone, 1952; Dobzhansky, 1970; Powell, 1997). This is due both to the power and convenience of *Drosophila* as an experimental system, and to the existence of 'species complexes' – groups of closely related species that have diverged enough to acquire reproductive isolation and phenotypic differences, but not enough to preclude hybridization and genetic analysis. In particular, much of what we know about the genetic basis of speciation and phenotypic evolution is based on the work in the *simulans* and *pseudoobscura* species complexes (Coyne, 1992; Laurie *et al.*, 1997; Kliman *et al.*, 2000; Noor *et al.*, 2000; Machado *et al.*, 2002; Wu & Ting, 2004). In this report, we describe another promising group of closely related species, the *Drosophila bipectinata* species complex.

* Corresponding author. Tel: +1 (530) 752 8328 (laboratory), +1 (530) 752 8657 (office). Fax: +1 (530) 752 9014. e-mail: akopp@ucdavis.edu

The *bipectinata* species complex is a group of four closely related species – *D. bipectinata*, *D. parabipectinata*, *D. malerkotliana and D. pseudoananassae* (Bock, 1971*b*, 1978; Singh & Singh, 2001). These species occur throughout Southeast Asia, extending into northeastern Australia, the Indian subcontinent and South Pacific (Fig. 1). All species are sympatric over most of their geographic ranges. In each of the latter two species, two allopatric subspecies have been described: *D. m. malerkotliana* and *D. m. pallens*, and *D. p. pseudoananassae* and *D. p. nigrens* (Fig. 1). The *bipectinata* complex is part of the *ananassae* species subgroup, which in turn belongs to the large and diverse *melanogaster* group (Bock & Wheeler, 1972; Bock, 1980; Lemeunier *et al.*, 1986; Lachaise *et al.*, 1988; Kopp & True, 2002*b*; Schawaroch, 2002).

Based on isozyme variation (Yang *et al.*, 1972), polytene chromosome morphology (Bock, 1971*a*; Jha & Rahman, 1972) and the patterns of reproductive isolation (Bock, 1978), *D. malerkotliana, D. bipectinata* and *D. parabipectinata* appear to be very close relatives, whereas *D. pseudoananassae* is more distantly related to the other three species. All species of

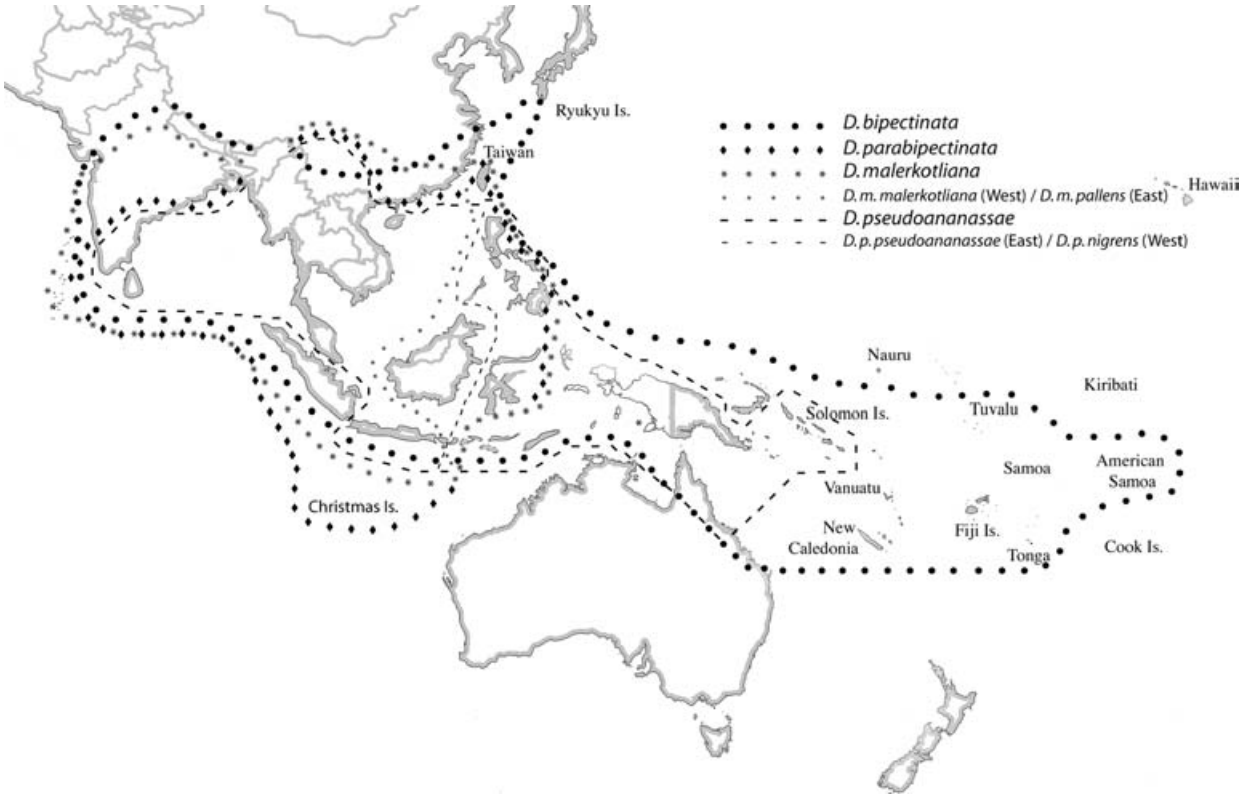

Fig. 1. Geographic distribution of the *bipectinata* species complex. The map is based on published reports (Bock, 1971*b*, 1980; Okada, 1979; Lemeunier *et al*., 1986), collection catalogs (*Drosophila* Species Stock Center, University of Arizona; Ehime University, Tokyo Metropolitan University and Tsukuba University *Drosophila* stock collections; and the Bishop Museum and Lund University museum collections) and personal communications from Drs B. N. Singh, S. McEvey, M. Polak, M. Toda and S.-C. Tsaur.

the *bipectinata* complex can be hybridized in the laboratory. Consistent with Haldane's rule (Haldane, 1922), the hybrid males are invariably sterile, whereas all hybrid females are fertile when backcrossed to either parent. There are no apparent reproductive barriers separating the subspecies of *D. malerkotliana* and *D. pseudoananassae*.

The *bipectinata* complex is distinguished from many other *Drosophila* species complexes by the high degree of morphological differentiation among and within species. The most obvious morphological features distinguishing the species and subspecies of the *bipectinata* complex are abdominal pigmentation and sex comb morphology. *D. parabipectinata*, *D. m. malerkotliana* and *D. p. nigrens* are pigmented in a sexually dimorphic pattern, where the last three abdominal segments are completely melanized in males but not in females. However, males of *D. bipectinata*, *D. p. pseudoananassae* and *D. m. pallens* lack such pigmentation, and are similar to females (Fig. 2). *D. bipectinata* and *D. parabipectinata* are also unique among the *ananassae* subgroup (but similar to several more distant evolutionary lineages) in having rotated, obliquely oriented sex combs, while *D. malerkotliana* and *D. pseudoananassae* have transverse sex combs that are typical of the *ananassae* subgroup (Fig. 3). Sex comb morphology also varies within species, especially in *D. bipectinata* (Chatterjee & Singh, 1993; Polak *et al*., 2004).

The extensive morphological variation within and among species together with the ease of hybridization makes the *bipectinata* species complex an excellent model for investigating the genetic basis of morphological evolution. To develop a historical framework for these studies, we have combined phylogenetic and population-genetic approaches to reconstruct the order and timing of speciation events, the extent of genetic differentiation among species and variation within species, and the levels of gene flow across species boundaries. We find that the high degree of morphological divergence among species and subspecies of the *bipectinata* complex stands in sharp contrast with an exceptionally low genetic differentiation at randomly chosen, neutrally evolving loci.

2. Materials and methods

(i) *Taxon sample*

The main taxon sample used in the analysis of nuclear loci consisted of 29 strains representing all six species

https://doi.org/10.1017/S0016672305007317 Published online by Cambridge University Press

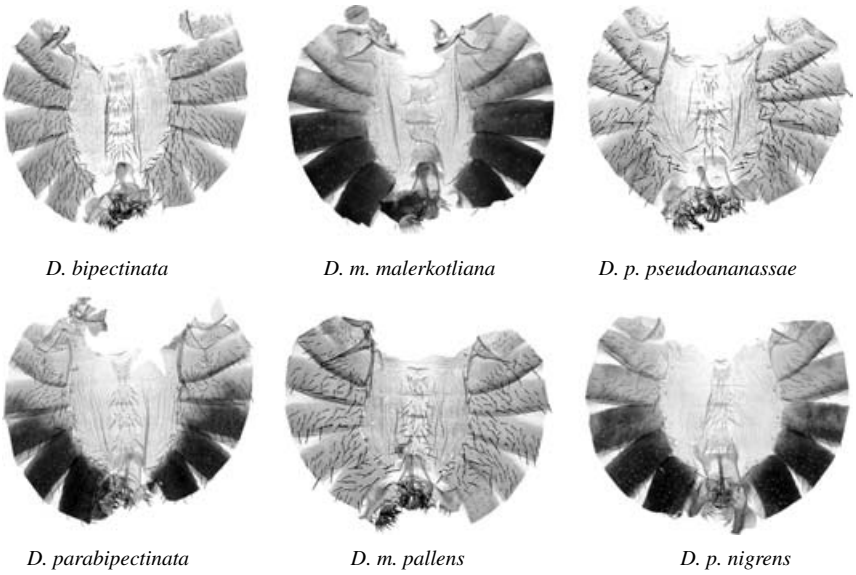

Fig. 2. Male abdominal pigmentation in the *bipectinata* species complex. Adult abdominal cuticles were cut open along the dorsal midline and mounted flat as described (Duncan, 1982), so that dorsal cuticle is on the outside and ventral cuticle is in the middle. Females of all species and subspecies lack dark abdominal pigmentation.

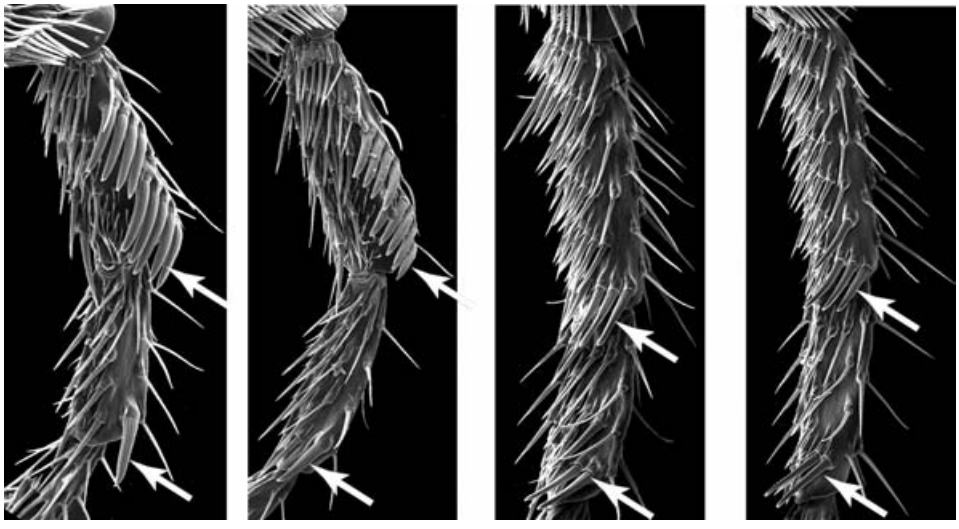

Fig. 3. Sex comb morphology in the *bipectinata* species complex. The two most proximal tarsal segments of the prothoracic male legs are shown for each species. Sex combs are indicated by arrows. Female chaetotaxy is identical in all species.

and subspecies of the *bipectinata* complex, including seven strains of *D. m. malerkotliana*, five strains of *D. m. pallens*, eight strains of *D. bipectinata*, five strains of *D. parabipectinata*, two strains of *D. p. pseudoananassae* and two strains of *D. p. nigrens*. Twenty-six additional strains of *D. bipectinata*, *D. parabipectinata* and *D. m. pallens* were included in the analysis of mitochondrial DNA. Geographic origin and collection dates for all strains are listed in Appendix Table 1. Strains used in this work represent the entire geographic range of the *bipectinata* species complex (Fig. 1). In most cases, only one strain from each location was available, so that population subdivision within species could not be investigated.

(ii) *Sequences*

Species of the *bipectinata* complex show very low levels of sequence divergence, and only non-coding sequences contain sufficient variation to be useful for phylogenetic analysis. Our emphasis was therefore on isolating intronic sequences of nuclear genes. Suitable fragments of six nuclear loci were amplified and sequenced, including *aristaless* (*al*), *bric a brac 2*

https://doi.org/10.1017/S0016672305007317 Published online by Cambridge University Press

(*bab2*), *Glycerol 3 phosphate dehydrogenase* (*Gpdh*), *pale* (*ple*), *Superoxide dismutase* (*Sod*) and *upheld* (*up*). In addition, we sequenced a fragment of the mitochondrial *Cytochrome oxidase subunit I* (*COI*) gene. In *D. melanogaster*, the *up* locus is located on the X chromosome, while the other five nuclear loci are autosomal. We have not examined the chromosomal locations of these loci in the *bipectinata* complex. Given the high degree of conservation of chromosomal arms in *Drosophila* (Ranz *et al*., 1999; Gonzalez *et al*., 2002), we treated the *up* locus as X-linked and the other loci as autosomal, for the purposes of population-genetic analysis. For each locus, the total length of aligned sequence fragments, the number of non-coding positions, the primers used to amplify the locus and the recommended annealing temperatures are listed in Appendix Table 2. Genomic DNA was isolated from a single male of each strain. Amplified PCR fragments were either sequenced directly using the same primers, or TA-cloned and sequenced using vector primers. All sequences have been deposited in EMBL/GenBank under accession numbers AJ844670–AJ844698 (*al*), AJ844699–AJ844727 (*bab2*), AJ844728–AJ844756 (*Gpdh*), AJ844757–AJ844809 (*COI*), AJ844810–AJ844838 (*ple*), AJ844839–AJ844867 (*Sod*) and AJ844868–AJ844892 (*up*).

ABI trace chromatograms were examined and the sequences corrected, if needed, using EditView and Contig Express software (from Applied Biosystems and Invitrogen, respectively). Some individuals were found to be heterozygous at one or more nucleotide positions, which were then represented using IUPAC ambiguity codes. Sequences were aligned using ClustalW multiple alignment algorithm (Thompson *et al*., 1994). Some small stretches of the *Sod* and *up* introns could not be aligned with confidence, and were excluded from analysis. All other sequences aligned unambiguously. For each locus, we prepared separate alignments for each species, as well as a combined file that included sequences from all species. Single-species alignments, which contained fewer gaps and were therefore shorter, were used to analyse sequence variation, allele frequency spectrum and recombination within species. Multi-species alignments were used for the analysis of species divergence and phylogenetic reconstruction.

(iii) *Phylogenetic analysis*

Phylogenetic reconstructions were performed using PAUP (Swofford, 2000) and MrBayes (Huelsenbeck & Ronquist, 2001) programs. For *Gpdh*, *Sod* and *COI*, *D. ananassae* and *D. melanogaster* sequences were used to provide outgroup information. Each locus and the combined data set consisting of concatenated sequences of all six nuclear loci were analysed using maximum parsimony, minimum evolution (based on total distance estimates) and Bayesian inference (Yang & Rannala, 1997; Larget & Simon, 1999; Huelsenbeck *et al*., 2002). One hundred bootstrap replicates with random order of sequence addition and TBR branch swapping were used to evaluate clade support in the maximum parsimony and minimum evolution analyses. For several loci, the number of equally parsimonious trees was very large, so that exhaustive bootstrapping could not be performed. In such cases, the number of trees swapped at each replicate was limited to 500. We found that minimum evolution analysis was highly sensitive to the choice of distance estimates. Typically, simple measures of distance (such as total uncorrected distances) produced the highest bootstrap values, while more highly parameterized models of sequence evolution (K2P or HKY85) yielded lower support. We suspect that recombination within loci causes models of sequence evolution to perform erratically. Bayesian analysis was carried out with uniform priors. For each data set, the analysis was repeated two to four times, and the chain was allowed to run for 2–3 million generations. Trees were sampled every 100 generations after a burn-in of 200 000 generations. Consensus tree topologies produced by replicate runs were always identical, and the posterior probabilities of taxon partitions were almost identical as well. In the combined analysis of the multi-locus data set, individual loci were treated as separate partitions with different models of sequence evolution. However, simplified analysis in which a single model of sequence evolution was applied to all partitions produced virtually identical results.

Incongruence length difference (ILD) and Shimodaira–Hasegawa (SH) statistical tests (Farris *et al*., 1995; Shimodaira & Hasegawa, 1999; Goldman *et al*., 2000) were performed in PAUP. The ILD test was carried out for the combined data set consisting of six nuclear loci, treated as separate partitions. One hundred replicates were performed to obtain a null distribution of tree length differences. For SH tests, the optimal tree reconstructed for each locus using maximum likelihood was compared with alternative trees reconstructed from all other individual loci and from the combined nuclear data set, under the model of sequence evolution estimated for the locus being tested. Since multiple comparisons were involved, statistical significance of test values was assessed using Bonferroni-corrected *P* values.

(iv) *Population genetic analysis*

Most population-genetic analyses were performed using DnaSP (Rozas & Rozas, 1999), ProSeq (Filatov, 2002) and SITES (Hey & Wakeley, 1997) software. In the assessments of sequence variation,

polymorphic sites segregating within indels were included in the analysis, whereas indels themselves were not considered as characters. Nucleotide diversity (π) (Tajima, 1983) and Watterson's estimate of the population mutation rate θ (Watterson, 1975) were calculated for each species and locus. To obtain mean estimates of sequence variation for each species, the values of π and θ at each locus were weighted by the length of that locus. These weighted averages were essentially identical to the values obtained from the concatenated data set.

The presence and extent of intra-locus recombination were assessed using the γ estimator of the population recombination rate $4Nc$ (Hey & Wakeley, 1997) and the four-gamete test, which measures the minimum number of recombination events in the sequence sample (Hudson & Kaplan, 1985). These analyses were performed both for single-species and for multi-species alignments, producing similar results. It should be noted, however, that the precision of $4Nc$ estimates is probably quite low due to the small number of sequences and the short length of some loci.

Allele frequency spectrum was examined using Tajima's D and Fu & Li's D and F test statistics (Tajima, 1989; Fu & Li, 1993). Tajima's test is based on the difference between two estimates of θ: one based on the number of segregating sites and one based on average pairwise allele differences (Watterson, 1975; Tajima, 1989). Significance of Tajima's D values is determined by comparing them with a distribution obtained by coalescent simulations under the neutral model of sequence evolution (Hudson, 1990). Significantly negative Tajima's D values indicate an excess of low-frequency variants, which may reflect either directional or purifying selection acting on the locus, or recent population expansion. Significantly positive values indicate an excess of intermediate-frequency variants, which may reflect balancing selection or population subdivision. We performed coalescent simulations separately for each locus and species, conditioning them on the length of the sequence, the number of strains and the number of segregating sites. Since the expected values of D depend on the population recombination rate, we repeated these simulations using $4Nc$ per base pair values ranging from zero to 0·015. In addition to single-locus analyses, we calculated weighted average values of Tajima's D across loci for each species following the same procedure as for π and θ. These average values were then used to compare the empirical results to the test distribution obtained by coalescent simulations.

Fu & Li's tests are similar in concept to Tajima's, and measure the excess or deficiency of nucleotide substitutions on the terminal branches of allele genealogy (i.e. singleton polymorphisms) (Fu & Li, 1993). To polarize nucleotide substitutions, allele genealogies in each species were rooted using outgroup sequences. The outgroup was chosen so as to minimize the number of polymorphisms shared with the ingroup: *D. pseudoananassae* was used as the outgroup for *D. bipectinata, D. parabipectinata* and *D. malerkotliana*, while *D. parabipectinata* was used as the outgroup for *D. pseudoananassae*. Significance of the test statistics was assessed using critical values listed by Fu & Li (1993). These values assume the absence of recombination and are conservative when recombination is present, as in our data set. The power of both Tajima's and Fu & Li's tests in our analysis is limited by the relatively small number of sequences, especially for *D. parabipectinata* and *D. pseudoananassae*.

Fst values were calculated for each locus and for each pair of species to assess the extent of genetic differentiation. Given high levels of recombination in our data set, each polymorphic site was treated as a separate locus (Hudson *et al*., 1992*b*). Significance of the Fst values was determined by permutation tests (Hudson *et al*., 1992*a*).

Isolation model fitting and tests of gene flow were performed using SITES and WH programs (Wang & Hey, 1996; Hey & Wakeley, 1997; Wang *et al*., 1997; Kliman *et al*., 2000; Machado *et al*., 2002). This model assumes that a single ancestral species gives rise instantaneously to two descendant species represented by the sequence samples; that the effective population size has been constant for each of the three species; and that the two descendant species evolved in complete isolation, i.e. without gene flow. Under these assumptions (some of which may be violated in our case: see below), the expected number of fixed differences and shared polymorphisms at each locus follows a distribution that can be estimated by coalescent simulations. If one or more loci have experienced recent gene flow across species, the observed numbers of fixed differences and shared polymorphisms at these loci will deviate from the simulated distribution: specifically, the number of shared polymorphisms will be elevated, while the number of fixed differences is reduced. The significance of this deviation is determined using a χ^2 test or the WWH statistic based on the count of fixed differences and shared polymorphisms (Wang *et al*., 1997; Kliman *et al*., 2000; Machado *et al*., 2002).

3. Results

(i) *Polymorphism levels*

As in other *Drosophila* species (Kliman *et al*., 2000; Machado *et al*., 2002), nuclear polymorphism levels in the *bipectinata* species complex vary considerably

https://doi.org/10.1017/S0016672305007317 Published online by Cambridge University Press

Table 1. *Polymorphism levels and allele frequency spectrum*

Locus	Taxon	*N*	*P*	*S*	*M*	π	θ(W)	Tajima *D*	Fu & Li *D*	Fu & Li *F*
al	*D. malerkotliana*	12	530	18	18	0·0083	0·0115	−1·22315	−1·37659	−1·59483
			432	17	17	0·0098	0·0134	−1·16252	−1·22614	−1·44028
	D. m. malerkotliana	7	518	13	13	0·0086	0·0102	−0·85656	−1·13639	−1·26423
			420	13	13	0·0107	0·0126	−0·85656	−1·13639	−1·26423
	D. m. pallens	5	519	10	10	0·0085	0·0092	−0·59633	−0·48589	−0·60616
			421	9	9	0·0095	0·0103	−0·52643	−0·30029	−0·41351
	D. bipectinata	8	518	17	17	0·0094	0·0127	−1·33693	**−2·09504***	**−2·29744***
			420	15	15	0·0104	0·0138	−1·26301	**−2·26013***	**−2·41947***
	D. parabipectinata	5	517	5	5	0·0043	0·0046	−0·56199	−1·11782	−1·18305
			419	5	5	0·0053	0·0057	−0·56199	−1·11782	−1·18305
	D. pseudoananassae	4	526	17	17	0·0168	0·0176	−0·48366	−0·97081	−1·03528
			428	16	16	0·0191	0·0204	−0·65448	−1·17360	−1·26691
bab2	*D. malerkotliana*	12	2237	89	90	0·0078	0·0132	**−1·97506****	**−3·12529****	**−3·37852****
	D. m. malerkotliana	7	2220	54	54	0·0074	0·0099	−1·27839	**−2·09345***	**−2·26165***
	D. m. pallens	5	2234	44	44	0·0080	0·0095	−0·99781	**−1·82099***	**−1·94844***
	D. bipectinata	8	2245	65	66	0·0079	0·0112	−1·57833	**−2·31769***	**−2·54914***
	D. parabipectinata	5	2247	29	29	0·0053	0·0062	−0·89450	−0·74182	−0·90964
	D. pseudoananassae	4	2252	41	41	0·0093	0·0099	−0·55663	−1·04248	−1·04526
Gpdh	*D. malerkotliana*	12	1125	69	80	0·0155	0·0203	**−1·78818***	**−2·85518****	**−3·11855****
			371	43	51	0·0315	0·0384	−1·63192	**−2·47139***	**−2·73864****
	D. m. malerkotliana	7	1123	44	55	0·0159	0·0160	−1·37585	−1·87608	−2·13456
			369	30	38	0·0336	0·0332	−1·34288	−1·70122	−1·96794
	D. m. pallens	5	1109	36	36	0·0144	0·0156	−0·62513	−1·32454	−1·39887
			355	21	21	0·0270	0·0284	−0·46201	−1·13663	−1·21527
	D. bipectinata	8	1116	63	66	0·0184	0·0218	−1·10413	−1·83196	−2·03849
			362	42	45	0·0408	0·0447	−0·85830	−1·32992	−1·52564
	D. parabipectinata	5	1123	14	14	0·0055	0·0060	−0·56398	0·07790	−0·12126
			369	9	9	0·0108	0·0117	−0·52644	0·73446	−0·60800
	D. pseudoananassae	4	1107	22	27	0·0129	0·0108	0·06671	−1·03735	−1·03085
			397	14	17	0·0231	0·0192	0·27911	−0·80320	−0·78423
ple	*D. malerkotliana*	12	961	34	34	0·0073	0·0117	−1·71156	**−2·88899****	**−3·10179****
			698	28	28	0·0082	0·0133	−1·69923	**−2·81674****	**−3·03282****
	D. m. malerkotliana	7	959	28	28	0·0092	0·0119	−1·28073	**−2·19380****	**−2·38104****
			696	22	22	0·0100	0·0129	−1·27303	**−2·17383***	**−2·35803***
	D. m. pallens	5	961	11	11	0·0050	0·0055	−0·65430	−1·39122	−1·46866
			698	10	10	0·0063	0·0069	−0·59632	−1·29567	−1·36388
	D. bipectinata	8	960	20	20	0·0062	0·0080	−1·20474	**−2·02484***	−2·19473
			697	17	17	0·0072	0·0094	−1·22454	**−2·09500***	**−2·26154***
	D. parabipectinata	5	959	19	20	0·0094	0·0095	−0·69305	−0·97882	−1·09843
			696	16	17	0·0109	0·0110	−0·59444	−0·73002	−0·83888
	D. pseudoananassae	4	961	12	12	0·0064	0·0068	−0·58365	−0·51161	−0·61642
			698	11	11	0·0081	0·0086	−0·55826	−0·75678	−0·84449
Sod	*D. malerkotliana*	12	1037	85	86	0·0201	0·0271	−0·95413	−1·03229	−1·20941
			626	63	64	0·0267	0·0339	−0·94042	−1·02977	−1·20029
	D. m. malerkotliana	7	1020	56	56	0·0193	0·0224	−0·44925	−0·65313	−0·73243
			626	43	43	0·0260	0·0288	−0·48416	−0·72290	−0·80684
	D. m. pallens	5	1037	47	47	0·0191	0·0218	−0·84029	−1·00738	−1·15082
			626	36	36	0·0246	0·0276	−0·81659	−1·06172	−1·19255
	D. bipectinata	8	1001	71	77	0·0247	0·0274	−1·13411	−1·45171	−1·67088
			590	57	63	0·0342	0·0379	−1·11964	−1·43333	−1·63129
	D. parabipectinata	5	1020	47	49	0·0208	0·0221	−0·77829	−1·26887	−1·44052
			609	46	48	0·0342	0·0363	−0·77829	−1·26691	−1·44052
	D. pseudoananassae	4	667	35	36	0·0290	0·0286	−0·11848	−0·02926	−0·02552
			602	35	36	0·0321	0·0317	−0·11848	−0·02926	−0·02552
up	*D. malerkotliana*	11	549	14	14	0·0082	0·0082	−0·14547	−0·59853	−0·57007
	D. m. malerkotliana	6	545	12	12	0·0095	0·0090	0·14539	0·31263	0·32808
	D. m. pallens	5	541	9	9	0·0062	0·0074	−1·18441	−0·07771	−0·36417
	D. bipectinata	8	562	15	15	0·0089	0·0112	−1·00801	−1·81464	−1·93532
	D. parabipectinata	4	553	0	0	0·0000	0·0000	NE	NE	NE
	D. pseudoananassae	2	564	8	8	0·0154	0·0154	NE	NE	NE

https://doi.org/10.1017/S0016672305007317 Published online by Cambridge University Press

Table 1. (*Cont.*)

Locus	Taxon	*N*	*P*	*S*	*M*	π	θ(W)	Tajima *D*	Fu & Li *D*	Fu & Li *F*
CO1	*D. malerkotliana*	20	553	15	15	0·00511	0·00894	−1·58175	**−3·11964****	**−3·18085****
	D. m. malerkotliana	7	553	4	4	0·00302	0·00345	−0·59756	−0·31269	−0·44473
	D. m. pallens	13	553	11	11	0·00512	0·00745	−1·26444	**−2·49140***	**−2·57779***
	D. bipectinata	23	553	24	24	0·00841	0·01395	−1·48497	**−2·44047***	**−2·57698***
	D. parabipectinata	8	553	6	6	0·00435	0·00486	−0·48874	−1·09521	−1·12848
	D. pseudoananassae	4	476	13	13	0·01821	0·01467	**2·24818***	**2·33023***	**2·72118***
All nuclear	*D. malerkotliana*	12	6439	309	322	0·0111	0·0159	−1·52078		
			4913	254	264	0·0123	0·0172	−1·50224		
	D. m. malerkotliana	7	6385	207	218	0·0113	0·0158	−1·00768		
			4876	174	182	0·0127	0·0146	−0·98507		
	D. m. pallens	5	6401	157	157	0·0103	0·0117	−0·83937		
			4875	129	129	0·0112	0·0127	−0·85804		
	D. bipectinata	8	6402	251	261	0·0123	0·0152	−1·30059		
			4876	211	221	0·0138	0·0169	−1·32591		
	D. parabipectinata	5	6419	114	117	0·0079	0·0085	−0·68427		
			4893	105	108	0·0095	0·0103	−0·68003		
	D. pseudoananassae	4	6077	135	141	0·0128	0·0128	−0·34174		
			4941	125	129	0·0128	0·0129	−0·38127		

N, number of sequences for each species; *P*, number of aligned nucleotide positions; *S*, number of polymorphic sites; *M*, total number of mutations; π, average nucleotide diversity per base pair (Tajima, 1983); θ(W), Watterson (1975) estimate of the population mutation rate per base pair. For loci that contain both coding and non-coding sequences, *P*, *S*, *M*, π, θ, Tajima's *D* (Tajima, 1989) and Fu & Li's *D* and *F* statistics (Fu & Li, 1993) were calculated for all sites (top line in each cell) and for non-coding sequences only (bottom line). Fu & Li's *D* and *F* statistics were calculated using outgroup sequences. Tajima's *D* and Fu & Li statistic values that are significant under the conservative assumption of no recombination within the locus are indicated by single asterisks ($P<0·05$) or double asterisks ($P<0·025$). For the *up* locus, some values could not be estimated (NE) due to insufficient number of sequences or lack of variation. For each species, average values of π, θ and Tajima's *D* across nuclear loci were calculated by weighting the value for each locus by the length of that locus (for all sites and for non-coding positions).

from gene to gene (Table 1). To obtain estimates of nucleotide diversity that are comparable across loci, as well as to similar estimates in other *Drosophila* species, we measured π values (Tajima, 1983) for non-coding sequences only (Table 1). In a combined sample of all six nuclear loci, non-coding π values were the highest for *D. bipectinata* (0·0138) and the lowest for *D. parabipectinata* (0·0095). These differences between species are not as pronounced as those found in the *pseudoobscura* and *simulans* species complexes (Kliman *et al.*, 2000; Machado *et al.*, 2002). Polymorphism levels were also similar in the two subspecies of *D. malerkotliana* (0·0127 and 0·0112).

In *D. malerkotliana, D. bipectinata* and *D. parabipectinata*, polymorphism levels at the mitochondrial *COI* locus are considerably lower than in the nuclear sequences (0·0044–0·0084; Table 1). The higher diversity in *D. pseudoananassae* (0·0182) is due entirely to the divergence between its two subspecies; no sequence differences were observed within either subspecies. To test whether low variation is unique to the *COI* locus, we sequenced a second mitochondrial DNA fragment that spanned the non-coding origin of replication region and several tRNA genes from five strains of *D. m. malerkotliana* and *D. m. pallens*. No nucleotide differences were found (data not shown), suggesting that low intraspecific polymorphism is a general feature of mitochondrial DNA in these species.

(ii) *Recombination*

γ estimates of the population recombination parameter $4Nc$ (Hey & Wakeley, 1997) and the minimum number of recombination events for each species and locus (Hudson & Kaplan, 1985) are shown in Appendix Table 3. Recombination was detected at all loci. However, the amount of variation at some loci was not always sufficient to estimate recombination for each individual species. Since much of the variation is shared among *D. malerkotliana, D. bipectinata* and *D. parabipectinata*, and since many alleles pre-date the separation of these species, we estimated $4Nc$ for combined samples consisting of sequences from *D. malerkotliana* and *D. bipectinata, D. bipectinata* and *D. parabipectinata*, and all three species together. The resulting estimates were roughly similar to those obtained for individual species (Appendix Table 3). Thus, although our data are not sufficient to obtain precise estimates of $4Nc$, it appears that all loci in our sample have undergone recombination. This finding has important implications for

https://doi.org/10.1017/S0016672305007317 Published online by Cambridge University Press

the analysis of allele frequency spectrum, speciation model fitting and phylogenetic analysis.

(iii) *Neutrality tests*

Since the sequences in our sample are predominantly non-coding, we did not use neutrality tests based on contrasts between synonymous and non-synonymous nucleotide substitutions. Instead, we applied the multilocus version of the HKA test (Hudson *et al*., 1987; Kliman *et al*., 2000), which is based on the prediction of the neutral model that the numbers of intraspecific polymorphism and fixed interspecific differences should be correlated across loci. No significant deviations from the neutral expectation were found for any species or loci (results not shown), suggesting that the patterns of sequence variation in our sample have not been influenced by a strong directional or purifying selection.

We also used Tajima's (1989) and Fu & Li's (1993) statistics to test each locus for deviations from the allele frequency spectrum expected under the neutral model of sequence evolution (Table 1). Test statistics for nuclear loci are negative in almost all cases. Several loci produce significant values of Tajima's and/or Fu & Li's statistics for some species even under the conservative assumption of no recombination. Especially low values are observed at the *bab2* locus, but some significantly negative values are also seen at the *al, Gpdh* and *ple* loci (Table 1). At the mitochondrial *COI* locus, significantly negative values are observed in *D. malerkotliana* and *D. bipectinata*, while significantly positive values are found in *D. pseudoananassae*. The latter observation is due to the high divergence between the two subspecies, *D. p. pseudoananassae* and *D. p. nigrens*.

The negative values of Tajima's and Fu & Li's statistics for all nuclear loci suggest that some or all species may have experienced recent population expansion. To test whether the observed pattern deviates significantly from the neutral expectation, we first calculated the average Tajima's *D* values for each species (Table 1). Two different measures were used: a mean value of *D* across loci weighted by the size of each locus, and the *D* value estimated from a single data set consisting of concatenated sequences of all six nuclear loci. The two measures were nearly identical; the value shown in Table 1 is the weighted average. The average values of *D* were negative for each species, ranging from −1·5208 in *D. malerkotliana* to −0·3417 in *D. pseudoananassae*. The significance of these values was assessed by coalescent simulations conditioned on 4*Nc* parameters ranging from zero to 0·01 per base pair. The empirical estimates of 4*Nc* in our sample are in excess of 0·02 for most loci (Appendix Table 3), indicating that our simulations remain conservative.

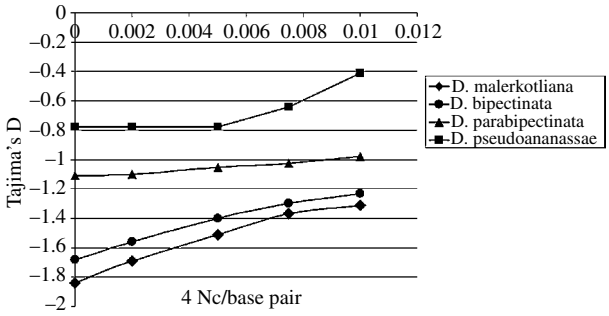

Fig. 4. Tajima's *D* values as a function of 4*Nc*. The γ estimate of 4*Nc* (Hey & Wakeley, 1997) was used in all simulations. Simulations were performed separately for each locus, conditioned on the number of sequences, length of the locus and the number of segregating sites. Average Tajima's *D* values for each species were then computed by weighting the value for each locus by the length of that locus. One thousand simulations were performed for each value of 4*Nc*. The plots show the lower 2·5% bound of the simulated *D* values, corresponding to $P=0{\cdot}05$ for one-tailed tests.

The results of coalescent simulations for each species are shown in Fig. 4. A comparison of actual and simulated values of Tajima's *D* indicates that the average values observed in *D. malerkotliana* and *D. bipectinata* fall below the neutral expectations for any reasonable level of recombination ($P<0{\cdot}05$). For *D. pseudoananassae* and *D. parabipectinata*, the observed average *D* values are not significant at $4Nc=0{\cdot}01$.

Among individual loci, the strongest negative values of Tajima's *D* are observed for *bab2*, suggesting that combined analysis may be influenced by the inclusion of that locus. We therefore estimated average *D* values and repeated the coalescent simulations after excluding *bab2* from the data set. In this analysis, the observed average *D* values for *D. malerkotliana* and *D. bipectinata* were significant ($P<0{\cdot}05$) at $4Nc=0{\cdot}015$, but not at $4Nc=0{\cdot}01$ (data not shown). Overall, both individual and combined analysis of nuclear loci suggests a significant excess of low-frequency polymorphisms in *D. malerkotliana* and *D. bipectinata*. The simplest interpretation of this pattern is that these two species have experienced recent population expansion.

(iv) *Divergence and shared variation among species*

To assess the extent of genetic differentiation among species, we first compared the average pairwise distances between alleles sampled from different species to the distances between alleles sampled from each individual species. The distances were calculated separately for each locus, as well as for a combined data set consisting of only non-coding sequences from

https://doi.org/10.1017/S0016672305007317 Published online by Cambridge University Press

Table 2. *Average pairwise distances between alleles* (*in per cent*)

Locus	*D. malerkotliana*	*D. bipectinata*	*D. parabipectinata*	*D. pseudoananassae*	*D. malerkotliana–D. bipectinata*	*D. malerkotliana–D. parabipectinata*	*D. bipectinata–D. parabipectinata*	*D. pseudoananassae–* others
al	0·828 (0·336)	0·945 (0·290)	0·427 (0·343)	2·104 (0·639)	1·890 (0·368)	1·556 (0·289)	0·815 (0·313)	2·611 (0·416)
bab2	0·959 (0·284)	1·295 (0·333)	0·779 (0·388)	1·618 (0·609)	1·091 (0·346)	1·005 (0·380)	1·142 (0·358)	2·530 (0·598)
Gpdh	1·340 (1·267)	1·462 (0·854)	0·989 (0·281)	1·198 (0·325)	1·569 (1·382)	2·666 (1·371)	2·981 (1·256)	3·463 (1·893)
ple	0·733 (0·292)	0·625 (0·362)	0·909 (0·503)	0·643 (0·259)	0·692 (0·347)	0·889 (0·290)	0·866 (0·292)	1·729 (0·272)
Sod	2·773 (0·292)	2·881 (0·664)	1·289 (0·577)	2·821 (0·525)	3·065 (0·709)	3·226 (0·564)	1·289 (0·577)	7·455 (1·060)
up	0·900 (0·494)	0·739 (0·341)	0·535 (0·325)	1·536 (N/A)	1·027 (0·366)	0·607 (0·197)	0·535 (0·325)	1·127 (0·3991)
All non-coding	1·486 (0·244)	1·702 (0·290)	1·350 (0·534)	1·790 (0·226)	1·942 (0·296)	2·093 (0·441)	2·154 (0·425)	4·243 (0·521)

Means and standard deviations are shown for each locus, as well as for a combined data set consisting of all non-coding nuclear sequences.

all six nuclear loci (Table 2). Average pairwise allele differences are very similar within each species. For *D. malerkotliana*, *D. bipectinata* and *D. parabipectinata*, the distances between alleles sampled from different species are only slightly (approximately 20–25%) higher on average than the intraspecific distances, suggesting that these species diverged very recently and/or continue to exchange alleles through gene flow. On the other hand, alleles sampled from *D. pseudoananassae* show much higher divergence from the other three species (Table 2), indicating a more distant relationship. *D. bipectinata*, *D. malerkotliana* and *D. parabipectinata* are almost equally diverged from *D. pseudoananassae* (4·384%, 4·080% and 4·408% non-coding sequence divergence, respectively).

We also calculated fixation indices (Fst) (Hudson *et al*., 1992*a*, *b*) and the numbers of fixed differences and shared polymorphisms for each pair of species (Table 3). All pairs of species show significant Fst values at most loci. However, genetic differentiation among *D. malerkotliana*, *D. bipectinata* and *D. parabipectinata* is quite low (Fst = 0·1824–0·3139 for nuclear loci), while the differentiation between these three species and *D. pseudoananassae* is greater (Fst = 0·4259–0·4822). There are many more shared polymorphisms than fixed differences among *D. malerkotliana*, *D. bipectinata* and *D. parabipectinata*, while in the comparison between these species and *D. pseudoananassae* this pattern is reversed (Table 3). There is no observable genetic differentiation between the two subspecies of *D. malerkotliana* (mean Fst = 0·0034, no fixed differences and no significant Fst values at any of the nuclear loci).

Among *D. malerkotliana, D. bipectinata* and *D. parabipectinata*, Fst values for the mitochondrial *COI* locus are not significant, and are almost an order of magnitude lower than for the nuclear loci (Table 3). This pattern is unlikely under neutral coalescence in the absence of gene flow, where fixed differences between species should accumulate faster in mitochondrial DNA due to its smaller effective population size. We believe that this result reflects a recent episode of gene flow that spread a single mitochondrial haplotype across all three species (see below). High differentiation between the two subspecies of *D. malerkotliana* is probably an artifact of taxon sampling, since the majority of *D. m. pallens* alleles were isolated from a single natural population, while the *D. m. malerkotliana* alleles represent strains collected at multiple geographic locations (Appendix Table 1).

(v) *Tests of gene flow*

The sharing of genetic variation may be due either to the segregation of ancestral polymorphisms, or to

https://doi.org/10.1017/S0016672305007317 Published online by Cambridge University Press

Table 3. *Genetic differentiation among species*

	malerkotliana–bipectinata	*malerkotliana–parabipectinata*	*bipectinata–parabipectinata*	*malerkotliana–pseudoananassae*	*bipectinata–pseudoananassae*	*parabipectinata–pseudoananassae*	*m. malerkotliana–m. pallens*
Genetic differentiation (Fst)							
al	0·5350**	0·6003**	0·1599*	0·4456**	0·5619**	0·6045**	−0·0713
bab2	0·0050	0·0433*	0·0344*	0·3082**	0·2976**	0·3233**	−0·0024
Gpdh	0·3113**	0·5510**	0·5363**	0·6030**	0·6366**	0·7676**	0·0240
ple	0·0243	0·0698	0·1061	0·5980**	0·6175**	0·5822**	−0·0482
Sod	0·0462*	0·2192**	0·1584**	0·5928**	0·5859**	0·6153**	0·0629
up	0·2520**	0·4000**	0·0994*	0·0078	0·0050	0·0000	0·0552
CO1	0·0283	0·0080	0·0352	0·2708**	0·2367**	0·3051**	0·2252**
Mean nuclear	0·1956	0·3139	0·1824	0·4259	0·4508	0·4822	0·0034
Number of fixed differences/shared polymorphisms							
al	3/1 (36)	3/0 (26)	0/0 (22)	2/0 (37)	7/1 (40)	7/0 (29)	0/5 (18)
bab2	0/14 (136)	0/10 (108)	0/6 (90)	9/1 (141)	7/1 (121)	8/4 (92)	0/8 (30)
Gpdh	2/5 (124)	6/2 (81)	6/2 (81)	22/0 (104)	23/0 (104)	30/0 (111)	0/11 (71)
ple	0/4 (50)	0/8 (45)	0/4 (35)	8/1 (52)	8/1 (39)	8/2 (37)	0/5 (34)
Sod	0/23 (131)	2/10 (124)	3/11 (110)	20/2 (129)	22/3 (120)	28/0 (103)	0/17 (87)
up	0/1 (28)	0/0 (14)	2/0 (16)	0/3 (19)	0/2 (20)	0/0 (8)	0/7 (14)
CO1	0/6 (34)	0/4 (18)	0/4 (27)	0/3 (25)	0/4 (32)	0/1 (18)	0/0 (15)
Total nuclear	5/48 (505)	11/30 (398)	11/23 (354)	61/7 (482)	67/8 (444)	81/6 (380)	0/53 (254)

The top part of the table shows Fst values for each locus, as well as the mean value for nuclear loci. Statistically significant values are indicated by single asterisks ($P<0·05$) or double asterisks ($P<0·01$). The bottom part shows the numbers of fixed differences and shared polymorphisms at each locus, in that order. The total number of segregating sites in both species is shown in parentheses.

Table 4. *Isolation model fitting and parameter estimates*

Species 1	Species 2	$\theta 1$	$\theta \varepsilon 2$	θA	τ	χ^2	$P(\chi^2)$	WWH	P(WWH)
malerkotliana	*bipectinata*	139·6 (101–215)	70·3 (91–211)	129·9 (42–101)	69·3 (28–41)	44·873 248·797	0·227 1·000	25 25	0·211 0·759
malerkotliana	*parabipectinata*	162·3 (84–336)	48·3 (28–96)	69·7 (28–114)	50·0 (17–35)	46·498 581·376	0·389 0·002	16 16	0·650 0·705
malerkotliana	*pseudoananassae*	110·1 (84–145)	71·1 (46–106)	58·7 (18–104)	134·3 (50–86)	22·829	0·743	22	0·492
bipectinata	*parabipectinata*	184·9 (109–535)	55·2 (37–85)	57·0 (28–85)	55·9 (21–28)	38·498 393·710	0·659 0·000	17 17	0·425 0·457
bipectinata	*pseudoananassae*	102·2 (61–193)	73·3 (43–153)	88·8 (18–172)	114·8 (29–85)	71·233	0·422	26	0·513
parabipectinata	*pseudoananassae*	46·2 (8–101)	64·9 (13–205)	101·5 (24–193)	80·4 (5–67)	72·929	0·576	34	0·498

For each pair of species, $\theta 1$ is the estimate of the population mutation rate for the first species in the pair, $\theta 2$ is the estimate for the second species and θA is the estimate for the common ancestor of the two species. τ is the estimate of the time since speciation, scaled for the effective population size. Point estimates and 95% confidence intervals (in parentheses) are shown for each parameter. χ^2 and WWH statistics (Wang *et al*., 1997) were used to test the fit of the isolation model. P values for each statistic indicate the percentage of coalescent simulations that produced more extreme values. For three species pairs, the χ^2 and WWH statistics were calculated for nuclear loci only (top value in each cell), and with the mitochondrial *COI* locus included (bottom values).

post-speciation gene flow (or both). *D. malerkotliana, D. bipectinata* and *D. parabipectinata* hybridize easily in the laboratory, suggesting that the exchange of alleles across species boundaries is not impossible. We examined the sequence data for evidence of recent gene flow by fitting it to the isolation speciation model (Wang & Hey, 1996; Wakeley & Hey, 1997; Wang *et al*., 1997). The results for each pair of species are shown in Table 4. If only nuclear loci are included in the analysis, the isolation model (i.e., a complete absence of gene flow) cannot be rejected for any of the six species pairs. This general result is at least somewhat robust to the assumed amounts of recombination. Higher recombination reduces the variance of parameter estimates, yielding lower P values (Wang *et al*., 1997; Kliman *et al*., 2000; Machado *et al*., 2002). P values shown in Table 4 correspond to the γ estimates of $4Nc$ obtained separately for each species (see Appendix Table 3). However, even if the highest $4Nc$ values estimated from combined data sets are used for all species, the χ^2 and WWH statistics still do not attain significance (results not shown). On the other hand, if the mitochondrial *COI* locus is included in the analysis, the χ^2 statistic reveals a significant deviation from the pattern expected under strict isolation for *D. malerkotliana* and *D. parabipectinata*, and for *D. bipectinata* and *D. parabipectinata* (Table 4). This suggests that some flow of mitochondrial DNA across species boundaries may have occurred.

(vi) *Ancestral population parameters*

The isolation speciation model (Wang & Hey, 1996; Wakeley & Hey, 1997; Wang *et al*., 1997) is based on four main parameters: the population mutation rate θ for the ancestral species and the two descendant species (a measure of effective population sizes) and the time since speciation (τ) scaled for the effective population size. The estimates of these parameters for each pair of species are shown, along with their 95% confidence intervals, in Table 4. By using the average of values estimated in different pairwise comparisons, we can obtain a rough estimate of the population mutation parameter for each of the four extant species. The mean $4Nu$ estimates are 137·324 for *D. malerkotliana*, 138·970 for *D. bipectinata*, 49·907 for *D. parabipectinata* and 69·757 for *D. pseudoananassae*. Thus, *D. malerkotliana* and *D. bipectinata* seem to have had larger effective population sizes than the other two species.

Phylogenetic analysis indicates that the common ancestor of *D. bipectinata* and *D. parabipectinata* diverged from *D. malerkotliana* prior to the separation of the former two species (Figs 5, 6; see below). We used this observation and the data from Table 4 to estimate $4Nu$ for the common ancestor of *D. bipectinata* and *D. parabipectinata* ($\theta = 56{\cdot}990$), the common ancestor of *D. malerkotliana, D. parabipectinata and D. bipectinata* ($\theta = 69{\cdot}985$) and the common ancestor of all four species ($\theta = 83{\cdot}0$). A comparison of $4Nu$ estimates between ancestral and descendant species suggests that *D. malerkotliana* and *D. bipectinata*, but not *D. parabipectinata* or *D. pseudoananassae*, have experienced population expansion. This is consistent with the strongly negative Tajima's D values observed in the former two species (Table 1; Fig. 4).

The mean τ (divergence time parameter) estimates are 55·939 for *D. bipectinata* and *D. parabipectinata*,

https://doi.org/10.1017/S0016672305007317 Published online by Cambridge University Press

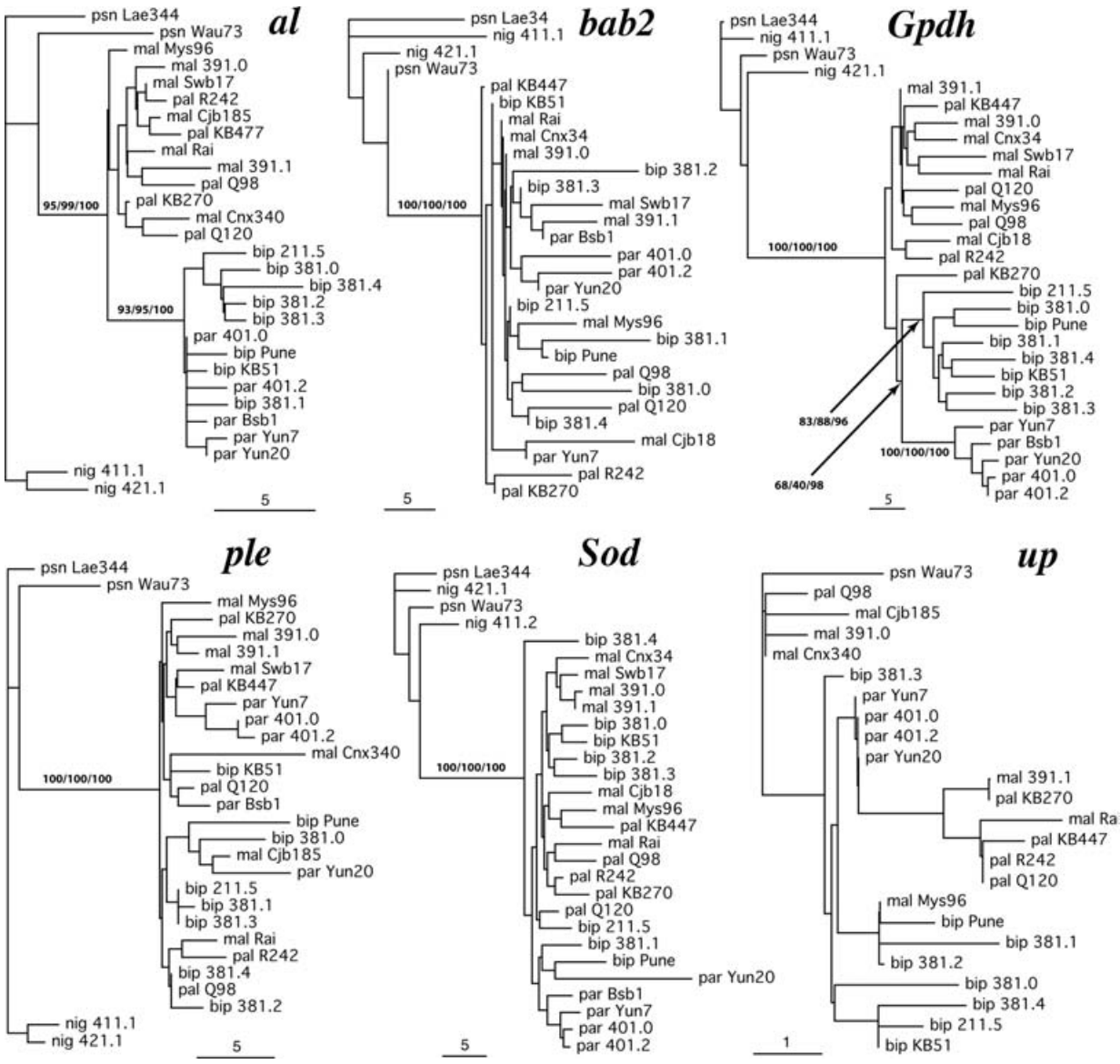

Fig. 5. Phylogenetic trees reconstructed from individual loci. Abbreviations: psn, *D. pseudoananassae pseudoananassae*; nig, *D. pseudoananassae nigrens*; mal, *D. malerkotliana malerkotliana*; pal, *D. malerkotliana pallens*; bip, *D. bipectinata*; par, *D. parabipectinata*. See Appendix Table 1 for strain information. Each locus was analysed using maximum parsimony (MP), minimum evolution based on total distances (ME) and Bayesian inference. Consensus trees, including branch lengths, are from Bayesian analysis. Measures of statistical support are shown for selected nodes. These values are, in order: MP bootstrap value, ME bootstrap value and Bayesian posterior probability. The *Gpdh* and *Sod* trees were rooted using *D. ananassae* and *D. melanogaster* sequences. In both cases, the root is located on the branch separating *D. pseudoananassae* from the other three species. All other trees are unrooted.

59·640 for *D. malerkotliana* and the common ancestor of *D. bipectinata* and *D. parabipectinata* and 109·8 for the common ancestor of all four species. Since the assumption of constant effective population sizes is clearly violated in two of the species, the τ estimates are probably not precise. Nevertheless, we can conclude that there was only a very brief interval between the two consecutive speciation events that separated *D. bipectinata*, *D. parabipectinata* and *D. malerkotliana*, whereas *D. pseudoananassae* diverged from the common ancestor of the other three species at a much earlier date.

(vii) *Phylogenetic analysis of nuclear loci*

Phylogenetic reconstruction assumes that all nucleotides within a locus share a unique, strictly bifurcating history – in other words, that there is no recombination within loci. This assumption is frequently violated in data sets from closely related species, and

https://doi.org/10.1017/S0016672305007317 Published online by Cambridge University Press

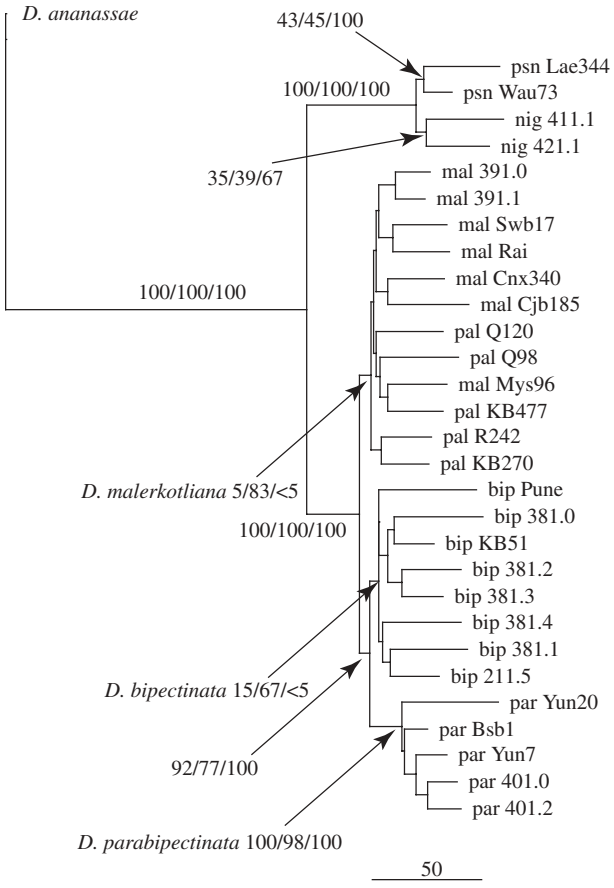

Fig. 6. Phylogenetic tree reconstructed from the combined nuclear data set. See Fig. 5 legend for details.

the *bipectinata* complex is no exception. Due to the high amounts of recombination at nuclear loci, the 'tree' reconstructed from each locus is not a real gene tree, but rather a superposition of multiple genealogies. Despite this important caveat, phylogenetic analysis may offer a valuable glimpse into the history of speciation.

Phylogenetic trees reconstructed from each of the six nuclear loci are shown in Fig. 5. Statistical support for monophyletic species and supra-specific clades is indicated in Table 5. Several conclusions can be drawn from this analysis. First, *D. malerkotliana*, *D. parabipectinata* and *D. bipectinata* clearly form a monophyletic group to the exclusion of *D. pseudoananassae*, with long branches separating *D. pseudoananassae* from the other three species. Second, *al* and *Gpdh* provide some support for monophyletic *D. bipectinata* and *D. parabipectinata*, which also form a monophyletic clade together to the exclusion of *D. malerkotliana*; on the other hand, none of the individual loci supports a monophyletic *D. malerkotliana*. Finally, in most analyses the sequences from *D. malerkotliana*, *D. parabipectinata* and *D. bipectinata* are intermingled, indicating a general lack of differentiation among these species. Internal branches tend to be very short, consistent with a nearly simultaneous separation of the three species (Table 4), as well as the presence of intra-locus recombination (Schierup & Hein, 2000). The main exception to this pattern is the *al* tree, in which a long branch separates *D. bipectinata* and *D. parabipectinata* from *D. malerkotliana* (Fig. 5). These species differ by several fixed inversions (Narda, 1969; Bock, 1971*a*; Jha & Rahman, 1972; Banerjee & Singh, 1996; A. Kopp, unpublished). One possibility is that the *al* locus is located within one of these inversions, in which case the divergence of *al* alleles may have pre-dated the divergence of species.

A combined analysis of all six nuclear loci supports the tree shown in Fig. 6. There is very strong support for a monophyletic clade composed of *D. malerkotliana*, *D. parabipectinata* and *D. bipectinata*, whereas *D. pseudoananassae* is the most basal species in the *bipectinata* complex. A monophyletic *D. parabipectinata*, and a clade composed of *D. parabipectinata* and *D. bipectinata*, are also well supported. On the other hand, there is little support for the monophyly of *D. malerkotliana* or *D. bipectinata*. No intraspecific nodes are supported by bootstrap values or posterior probabilities above 50%. The results of combined analysis should not be interpreted literally as a representation of historical relationships among strains, since it is virtually certain that different nuclear loci have different genealogies. The ILD (partition homogeneity) test (Farris *et al.*, 1995) reveals significant incongruence among nuclear loci ($P<0{\cdot}01$). The Shimodaira–Hasegawa test (Shimodaira & Hasegawa, 1999; Goldman *et al.*, 2000) also shows that some loci are incompatible with trees reconstructed from other loci (Appendix Table 4). However, all nuclear loci are compatible with the tree reconstructed from the combined data set (Fig. 6; Appendix Table 4). Despite the caveats associated with recombination and incongruence among loci, we can accept the topology (((*bipectinata*, *parabipectinata*) *malerkotliana*) *pseudoananassae*) as an approximation of the history of speciation in the *bipectinata* complex.

(viii) *Mitochondrial phylogeny*

Genetic differentiation among *D. malerkotliana, D. parabipectinata* and *D. bipectinata* is much lower at the mitochondrial *COI* locus than at the nuclear loci (Table 3). Nucleotide diversity at *COI* within species is also unexpectedly low (Table 1). To understand the reasons behind these findings, we reconstructed the phylogeny of *COI* sequences from 55 strains representing all four species. The resulting haplotype network is shown in Fig. 7A, and the geographic distribution of each haplotype is indicated in Fig. 7B. The most striking feature of this phylogeny is that

Table 5. *Bootstrap and posterior probability support for proposed monophyletic clades*

	#char[a]	#var[b]	#par[c]	#trees[d]	*bipectinata* complex	(*D. malerkotliana*, *D. bipectinata*, *D. parabipectinata*)	(*D. bipectinata*, *D. parabipectinata*)	*D. bipectinata*	*D. parabipectinata*	*D. malerkotliana*	*D. pseudo-ananassae*
al	530	59	19	101	N/A	95/99/100	93/95/100	<5/65/<5	<5/62/<5	<5/18/27	N/A
bab2	2275	211	54	>5000	N/A	100/100/100	–	–	–	–	N/A
Gpdh	1176	167	63	>5000	100/100/100	100/100/100	68/40/98	83/88/96	100/100/100	6/13/2000	100/100/100
ple	968	78	26	>5000	N/A	100/100/100	–	–;	–	–	N/A
Sod	1058	198	99	5	100/100/100	100/100/100	–	–	30/30/70	–	100/100/100
up	583	33	13	>5000	N/A	N/A	–	–	–	–	N/A
Combined	6590	746	274	10	100/100/100	100/100/100	92/77/100	15/67/<5	100/98/100	5/83/<5	100/100/100

The combined data set consisted of all six nuclear loci. The three values shown for each clade are the maximum parsimony bootstrap support, minimum evolution bootstrap support and Bayesian posterior probability, in that order. Some clades could not be evaluated due to insufficient taxon sampling (N/A). For *Gpdh* and *Sod*, trees were rooted using sequences from *D. ananassae* and *D. melanogaster*.

[a] Total number of nucleotide positions in the aligned sequences.

[b] Number of polymorphic characters.

[c] Number of parsimony-informative characters.

[d] Number of equally parsimonious trees.

nine strains of *D. bipectinata*, seven strains of *D. malerkotliana* and two strains of *D. parabipectinata* all share the same mitochondrial haplotype. This haplotype is found over the geographic area from Nepal to South Pacific islands. Although *D. ananassae* is too far diverged from the *bipectinata* complex to root the haplotype network reliably, it seems that the single most widespread haplotype is close to the ancestral *bipectinata* complex sequence (Fig. 7A). Most of the other haplotypes found in our sample are separated from the major haplotype by only one or two nucleotide substitutions. Some of these haplotypes are also shared by *D. malerkotliana*, *D. bipectinata* and *D. parabipectinata*, and are found at multiple geographic locations.

Strains of different species that share the same mitochondrial haplotype show much greater divergence at nuclear loci (Figs 5, 6; Table 2). This pattern is the opposite of that expected under random coalescence, given the smaller effective population size for mitochondrial loci. The most likely explanation for this discrepancy is gene flow. The spread of a single mitochondrial haplotype across three species and over a wide geographic area suggests that this process may have been favoured by natural selection. Such selection could result, for example, from the spread of a *Wolbachia* or similar female-inherited endosymbiotic infection (Turelli & Hoffmann, 1995; Werren *et al.*, 1995; Charlat *et al.*, 2003). We therefore attempted to amplify *Wolbachia* sequences from each of the 55 strains represented in Fig. 7A using conserved primers against 16S ribosomal RNA and *Wolbachia* Surface Protein (Zhou *et al.*, 1998). Only two strains (*D. m. pallens* KB447 and KB786, both from Brunei) tested positive for *Wolbachia*, so the evolutionary forces behind the spread of mitochondrial DNA across species boundaries remain unknown.

4. Discussion

(i) *Evolutionary history of the* bipectinata *species complex*

In this report, we combine phylogenetic and population-genetic approaches in an effort to reconstruct the evolutionary history of the *Drosophila bipectinata* species complex. Our sequence sample, while not large, is sufficient to draw several conclusions. First, we find that *D. bipectinata*, *D. parabipectinata* and *D. malerkotliana* form a monophyletic group, while *D. pseudoananassae* is much more distantly related to the other three species. This phylogeny is consistent with the relationships proposed on the basis of isozyme variation (Yang *et al.*, 1972), chromosome morphology (Bock, 1971*a*; Jha & Rahman, 1972) and interspecific hybridization (Bock, 1978). In particular, *D. bipectinata*, *D. parabipectinata* and

https://doi.org/10.1017/S0016672305007317 Published online by Cambridge University Press

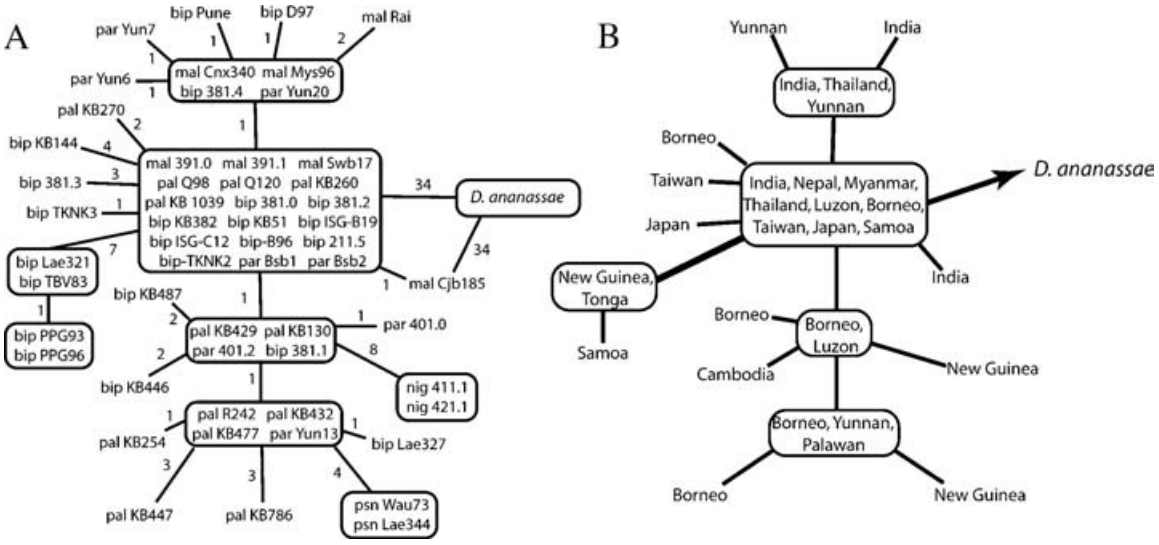

Fig. 7. (A) Mitochondrial haplotype network based on the *COI* locus. Numbers next to the edges denote the numbers of nucleotide substitutions that distinguish nearest haplotypes. (B) Same network, showing the geographic distribution of haplotypes.

D. malerkotliana hybridize easily in the laboratory and produce abundant and morphologically normal progeny. On the other hand, *D. pseudoananassae* is difficult to hybridize to any of the other three species, and the resulting F1 progeny are semi-lethal and frequently have morphological abnormalities (Bock, 1978; A. Kopp, unpublished). The relationship among *D. bipectinata*, *D. parabipectinata* and *D. malerkotliana* is extremely close (Tables 2, 3). Phylogenetic analysis suggests that the common ancestor of *D. bipectinata* and *D. parabipectinata* diverged from *D. malerkotliana* prior to the separation of the former two species (Fig. 6). However, estimates of divergence time indicate that the two speciation events were separated by a very brief interval (Table 4). Although an apparently closer relationship between *D. bipectinata* and *D. parabipectinata* could in principle be due to a greater post-speciation gene flow between these two species, this explanation appears unlikely, since we do not find any evidence of gene flow in our sample of nuclear sequences. A sister-group relationship between *D. bipectinata* and *D. parabipectinata* is also supported by their similar sex comb morphology (see below).

In contrast to the species phylogeny, analysis of present-day variation reveals greater differentiation between *D. malerkotliana* and *D. parabipectinata* than either between *D. malerkotliana* and *D. bipectinata*, or between *D. parabipectinata* and *D. bipectinata* (Table 3). Interestingly, this observation is paralleled by the pattern of reproductive isolation among these species. Hybridization in the laboratory between *D. malerkotliana* and *D. bipectinata*, and between *D. parabipectinata* and *D. bipectinata*, occurs more readily than hybridization between *D. malerkotliana* and *D. parabipectinata* (Bock, 1978; A. Kopp, unpublished). Moreover, fertile male hybrids can be produced quite easily by repeated backcrossing for *D. malerkotliana* and *D. bipectinata*, and for *D. parabipectinata* and *D. bipectinata*, but only with difficulty for *D. malerkotliana* and *D. parabipectinata* (A. Kopp, unpublished).

After their divergence, the four species of the *bipectinata* complex seem to have had different demographic histories. *D. malerkotliana* and *D. bipectinata* have experienced strong population growth, as evidenced by significant negative values of Tajima's and Fu & Li's statistics at nuclear loci (Table 1, Fig. 4) and by comparison of population mutation rate estimates between these two species and their hypothetical ancestors (Table 4). On the other hand, evidence of recent population expansion is lacking in *D. parabipectinata* and *D. pseudoananassae*. The smaller effective population size in *D. parabipectinata* may be responsible for the greater genetic divergence between this species and *D. malerkotliana* than between *D. malerkotliana* and *D. bipectinata*.

Although we do not find any evidence of nuclear gene flow in our sequence sample, it is unlikely that divergence among *D. malerkotliana*, *D. bipectinata* and *D. parabipectinata* took place entirely without gene flow. First, the results of isolation model fitting should be interpreted with caution, since one of the main assumptions of the model – a constant effective population size – is violated in *D. malerkotliana* and *D. bipectinata*. Most importantly, mitochondrial DNA provides strong evidence of recent gene flow (Fig. 7; Table 4). Hybridization in nature is probably quite rare. F1 and backcross hybrids between *D. bipectinata* and *D. malerkotliana* can be distinguished

https://doi.org/10.1017/S0016672305007317 Published online by Cambridge University Press

from either parental species by their sex comb morphology. In the course of our field work, we have collected several thousand males of both species, without observing a single possible hybrid (A. Kopp & O. Barmina, unpublished). However, even rare episodes of hybridization would be sufficient to introgress some genetic material across species boundaries. Due to Haldane's rule (Haldane, 1922), the signature of hybridization would be most noticeable at mitochondrial loci, but parts of the nuclear genome that are not linked to genes responsible for reproductive isolation may also be introgressed (Wu, 2001).

One of the most interesting features of the *bipectinata* species complex is that its four members are almost completely sympatric (Fig. 1). This does not necessarily imply that speciation events that separated these species occurred in sympatry. The geographic distribution of the *bipectinata* complex includes the entire insular Southeast Asia, which contains thousands of large and small islands and is famous for its biological diversity (Holloway & Hall, 1998). This region has had a very active geological history (Hall, 2002), and was affected by repeated fluctuations in sea level (Voris, 2000). It is entirely possible that different species of the *bipectinata* complex originated on different islands and subsequently spread to other parts of Southeast Asia and nearby regions, achieving their current sympatric distribution. Unfortunately, the sequence data available to us do not shed any light on the historical biogeography of the *bipectinata* complex. Most of the historical information contained in the non-recombining mitochondrial DNA appears to have been wiped out by a recent spread of a single haplotype across three species, while the amount of recombination at nuclear loci is so high that we are unable to reconstruct allele trees.

(ii) *A role for phylogenetic methods in the study of speciation*

In the *bipectinata* species complex, as in the *simulans* and *pseudoobscura* complexes (Kliman *et al*., 2000; Machado & Hey, 2003), different loci have different genealogies (Fig. 5). This may be due either to gene flow during the early stages of speciation, or simply to lineage sorting and persistence of ancient polymorphisms that pre-date speciation. Regardless of the cause, there is no single dichotomous tree that can adequately describe the history of the species' genomes. Given high amounts of intra-locus recombination, even a gene tree becomes an almost meaningless concept. In practical terms, the most easily observable effect of recombination is to stretch out the terminal branches of the 'tree', causing it to approach a star phylogeny (Schierup & Hein, 2000). This obviously complicates the reconstruction of interspecific relationships, which are represented by the deepest nodes in the phylogeny. However, the effect of recombination on tree topology depends mainly on the degree of relatedness of the recombining sequences. If recombination occurs between closely related sequences, as is likely in a sample drawn from several reproductively isolated species, the recovered tree will have the topology supported by the majority of individual characters (Posada & Crandall, 2002). Then, while each individual locus can be highly misleading, a combined analysis of a sufficiently large multi-locus data set may give us a reasonable approximation of the order of speciation events. In effect, we may still be able to recover parts of the species tree even though gene trees do not exist.

Such relaxed interpretation of phylogenetic trees is quite different from the traditional approach to combining multiple data partitions (Huelsenbeck *et al*., 1996; Wiens, 1998), and is open to many criticisms. However, sometimes we need a tree-like representation of species' histories – in particular, to trace the patterns of morphological evolution. While we cannot draw any conclusions about intraspecific relationships, phylogenetic analysis shows that *D. bipectinata* and *D. parabipectinata* are sister species (Figs 5, 6). This conclusion is further corroborated by morphological similarities (Fig. 3) and by patterns of reproductive isolation. We suggest that while methodological caveats should always be kept in mind, they should not necessarily prevent us from applying phylogenetic tools to analyse the history of recent speciation events.

(iii) *Estimating the dates of speciation*

We can use the levels of genetic variation within and among species to obtain a very rough estimate of the dates of speciation events. Since the island of Kauai in the Hawaii archipelago is known to be 5·1 million years (MY) old, we can calibrate the divergence of non-coding nuclear sequences in *Drosophila* by comparing sequences from *D. picticornis*, which occurs only on Kauai, and from the clade composed of *D. planitibia*, *D. silvestris, D. heteroneura* and *D. differens*, which occur only on the younger islands (Rowan & Hunt, 1991; Russo *et al*., 1995). The mean pairwise divergence of non-coding nuclear sequences between these two taxa, averaged across three loci (*Gpdh*, *Adh* and *yp1*), is 10·663 %, which yields an estimate of the divergence rate of 1·045 % per lineage per million years. The average divergence of non-coding nuclear sequences between *D. pseudoananassae* and the other species of the *bipectinata* complex is 4·243 %, suggesting that *D. pseudoananassae* split off from the other three species approximately 2·03 MY ago.

https://doi.org/10.1017/S0016672305007317 Published online by Cambridge University Press

This method of estimating speciation dates cannot be applied to *D. malerkotliana, D. bipectinata* and *D. parabipectinata*, since these species carry many ancestral alleles that pre-date speciation. Instead, we can use an approach similar to the one proposed by Kliman *et al.* (2000). If differences between alleles accumulate at a rate roughly proportional to the absolute divergence time, then the average pairwise distance between alleles sampled from two different species ($d_{1/2}$) should be equal to the sum of the average distance between alleles that were present in the common ancestor of these species (d_A) and the subsequent divergence $2ut$, where u is the rate of accumulation of nucleotide differences per lineage per unit of time and t is the time since speciation. Since all three species have diverged from *D. pseudoananassae* at a similar rate, we can use a single u estimate for all lineages. $d_{1/2}$ values for each pair of species are known (Table 2), and u can be assumed to equal the neutral divergence rate estimated above (1·045% per lineage per MY); the biggest uncertainty lies in the estimate of d_A. The effective population size of the common ancestor of *D. malerkotliana, D. bipectinata* and *D. parabipectinata* was probably slightly higher than in *D. parabipectinata*, but lower than in *D. malerkotliana* or *D. bipectinata* (Table 4). If we assume that average sequence divergence within species is roughly proportional to its effective population size, then d_A for the common ancestor of these three species should be equal to or greater than the average divergence within *D. parabipectinata*. The $2ut$ parameter can then be estimated at 0·592%, 0·743% and 0·804% from the pairwise comparisons between *D. malerkotliana*/*D. bipectinata*, *D. malerkotliana*/*D. parabipectinata* and *D. bipectinata*/*D. parabipectinata*, respectively. According to these estimates, the common ancestor of all three species existed 283 000–385 000 years ago – more recently if d_A was higher than the average sequence divergence within *D. parabipectinata*. Needless to say, this is a very rough estimate that depends on several unverified assumptions, including the constancy of the molecular clock in different *Drosophila* lineages.

(iv) *Comparison with other species complexes*

Much of what we know about the population genetics of speciation and species divergence is based on the work in two groups of closely related *Drosophila* species: the *simulans* species complex (which includes *D. simulans*, *D. mauritiana* and *D. sechellia*) and the *pseudoobscura* complex (including *D. p. pseudoobscura*, *D. p. bogotana* and *D. persimilis*) (Coyne, 1992; Palopoli & Wu, 1994; Wang *et al.*, 1997; Kliman *et al.*, 2000; Noor *et al.*, 2000; Machado *et al.*, 2002; Wu & Ting, 2004). To help us understand the evolutionary history of the *bipectinata* complex, it is worth comparing with to these better-known *Drosophila* groups (Table 6).

Several conclusions can be drawn from this comparison. First, population bottlenecks do not appear to have played a major role in any of the *bipectinata* complex species, as they did in *D. sechellia* and *D. p. bogotana* (Kliman *et al.*, 2000; Machado *et al.*, 2002). Second, genetic differentiation at nuclear loci among *D. malerkotliana*, *D. bipectinata* and *D. parabipectinata* is much lower than in the *simulans* species complex, and slightly lower than in the *pseudoobscura* complex (Table 6). The most likely explanation is that species of the *bipectinata* complex diverged more recently and/or had larger effective population sizes than species of the *pseudoobscura* and *simulans* complexes. Finally, the exchange of mitochondrial DNA among sympatric *Drosophila* species appears to be common. For example, although *D. p. pseudoobscura* is more closely related to *D. p. bogotana* at nuclear loci, the differentiation at mitochondrial loci between the two subspecies of *D. pseudoobscura* is far greater than differentiation between *D. pseudoobscura* and *D. persimilis* (Table 6). Since *D. p. pseudoobscura* is partially sympatric with *D. persimilis* but completely allopatric with *D. p. bogotana*, gene flow appears to be the most likely explanation (Powell, 1983; Machado *et al.*, 2002). Mitochondrial gene flow is also likely to have occurred among *D. simulans*, *D. mauritiana* and *D. sechellia* (Ballard, 2000*a*, *b*; Ballard *et al.*, 2002; Solignac, 2004). In fact, interspecific introgression of mitochondrial DNA has been documented in a wide variety or animals, especially insects (Willett *et al.*, 1997; Sota & Vogler, 2001; Sota *et al.*, 2001; Martin *et al.*, 2002; Shaw, 2002) and vertebrates (Ferris *et al.*, 1983; Mukai *et al.*, 1997; Crochet *et al.*, 2003; Rognon & Guyomard, 2003). In male-heterogametic organisms in particular, the exchange of mitochondrial DNA is made easier by the fact that recently diverged species can continue to produce fertile female hybrids for a long time after the initial speciation event (Coyne & Orr, 1989; Orr, 1993; Coyne, 1994; Turelli, 1998; Ting *et al.*, 2000; Wu, 2001). On a practical note, while the non-recombining mitochondrial DNA is useful for population-genetic and phylogeographic studies, it should not be used as an approximation of species phylogeny in groups of closely related taxa.

(v) *Morphological evolution*

The *bipectinata* species complex differs from the better-known *pseudoobscura* and *simulans* complexes in that its member species display a much greater degree of morphological divergence (Figs 2, 3). Phylogenetic analysis shows that taxa with light and dark male pigmentation do not cluster together

Table 6. *A comparison of three* Drosophila *species complexes*

	malerkotliana	*bipectinata*	*parabipectinata*	*pseudoananassae*	*malerkotliana–bipectinata*	*malerkotliana parabipectinata*	*bipectinata–parabipectinata*	*pseudoananassae–* others
Genetic distance[a] (nuclear)	1·486	1·702	1·350	1·790	1·942	2·093	2·154	4·243
Genetic distance (mitochondrial)	0·610	0·438	0·410	1·821	0·482	0·471	0·409	1·717
Fst (nuclear)					0·196	0·314	0·182	0·536
Fst (mitochondrial)					0·028	0·008	0·035	0·256
Nucleotide diversity π (nuclear)	0·0123	0·0138	0·0095	0·0128				
Nucleotide diversity π (mitochondrial)	0·0051	0·0084	0·0044	0·0183				

	simulans	*mauritiana*	*sechellia*	*melanogaster*	*simulans–mauritiana*	*simulans–sechellia*	*mauritiana–sechellia*	*melanogaster–* others
Genetic distance (nuclear)	2·116	0·987	0·685		2·406	3·085	3·077	6·857
Genetic distance (mitochondrial)	1·548	1·744	0·108		1·948	2·594	2·835	4·014
Fst (nuclear)					0·573	0·659	0·700	0·804
Fst (mitochondrial)					0·032	0·009	0·053	0·776
Nucleotide diversity π(nuclear)	0·0247	0·0144	0·0071					
Nucleotide diversity π (mitochondrial)	0·0147	0·0116	0·0170					

	pseudoobscura	*bogotana*	*persimilis*	*miranda*	*pseudoobscura–bogotana*	*pseudoobscura–persimilis*	*persimilis–bogotana*	*miranda–* others
Genetic distance (nuclear)	1·564	0·614	0·877	1·628	1·669	1·970	2·075	3·840
Genetic distance (mitochondrial)	0·325	0·097	0·338		1·592	0·376	1·580	3·572
Fst (nuclear)					0·291	0·381	0·624	0·528
Fst (mitochondrial)					0·865	0·113	0·865	0·936
Nucleotide diversity π (nuclear)	0·0172	0·0039	0·0079					
Nucleotide diversity π (mitochondrial)	0·0033	0·0015	0·0049					

Average pairwise distances between alleles (in per cent) and Fst and π values are shown for each species or species pair. To make the estimates comparable across taxa, genetic distances and π values were estimated for non-coding sequences only. For the *bipectinata* species complex, all six nuclear loci were combined for this analysis. For the *simulans* complex, 13 loci were used, including *Adh*, *ci*, *janA*, *OdsB*, *hb*, *In(2L)t*, *per*, *tra*, *yp2*, *z*, *Zw*, *Sxl* and *w* (data reported by Kliman *et al*. (2000). For the *pseudoobscura* complex, six loci were analysed, including *2001*, *2003*, *3002*, *4002*, *4003* and *X008* (data reported by Machado *et al*. (2002).

[a] Average pairwise distances between alleles, in per cent.

(Figs 2, 6). The distribution of pigmentation patterns on the phylogenetic tree is completely random, indicating that evolutionary transitions between light and dark male pigmentation occurred several times within the *bipectinata* complex. In view of the recent divergence among these species, the most likely explanation is that the alleles responsible for pigmentation differences pre-date speciation. In contrast, the evolution of a highly developed oblique sex comb in the *bipectinata* complex appears to be a unique event that occurred in the common ancestor of *D. bipectinata* and *D. parabipectinata* after its divergence from *D. malerkotliana* (Figs 3, 6). The oblique sex comb seen in *D. parabipectinata* and *D. bipectinata* is unique within the *ananassae* species subgroup, but strongly resembles the morphology seen in several other lineages in the *melanogaster* species group (Kopp & True, 2002*a*) and in the *obscura* species group (Lakovaara & Saura, 1982), presenting a clear example of convergent evolution.

The high degree of morphological divergence stands in sharp contrast with the exceptionally low genetic differentiation among *D. malerkotliana*, *D. bipectinata* and *D. parabipectinata* at randomly chosen, neutrally evolving loci. There are many examples in the animal world where the genealogy of neutrally evolving loci does not reflect the pattern of phenotypic differentiation (Brower, 1994; McMillan *et al*., 1999; Beltran *et al*., 2002; Shaw, 2002; Sota, 2002; Broughton & Harrison, 2003; Besansky *et al*., 2003; Glor *et al*., 2003; Ramon *et al*., 2003; Salzburger & Meyer, 2004). These observations are consistent with a view of speciation in which the genomes of closely related species are mosaics composed of blocks of high and low divergence (Ting *et al*., 2000; Wu, 2001; Ortiz-Barrientos *et al*., 2002; Machado & Hey, 2003; Wu & Ting, 2004). Even low levels of gene flow between emerging species will be sufficient to maintain a common gene pool at neutral loci (Crow & Kimura, 1970). However, genes involved in differential adaptation or reproductive isolation are likely to be eliminated from the recipient population by selection. Gene flow at these loci will cease earlier than in the rest of the genome, causing them to show a greater degree of interspecific differentiation than other parts of the species' genomes (Ting *et al*., 2000; Wu, 2001). An important observation in this regard is that the size of the sex comb appears to be under sexual selection in *D. bipectinata* (Polak *et al*., 2004). In the future, identification of genes responsible for phenotypic diversification in the *bipectinata* species complex may help us understand the evolutionary processes driving species divergence.

Appendix

Appendix Table 1. *Geographic origin of strains used in the analysis*

Species	Strain	Type[a]	Origin	Year[b]	Donor
D. p. pseudoananassae	psn Lae344*	Iso	Lae, Papua New Guinea	1981	Y. Fuyama
	psn Wau73*	Iso	Wau, Papua New Guinea	1981	Y. Fuyama
D. p. nigrens	nig 411·0*	Multi	Samut Songkhram, Thailand	1971	TDSSC[c]
	nig 421·0*	Multi	Chiang Mai, Thailand	1971	TDSSC
D. m. malerkotliana	mal 391·0*	Multi	Mysore, India	1971	TDSSC
	mal 391·1*	Multi	Samut Songkhram, Thailand	1971	TDSSC
	mal Swb17*	Iso	Schwebo, Myanmar	1981	Y. Fuyama
	mal Cnx 340*	Iso	Chiang Mai, Thailand	1981	Y. Fuyama
	mal Cjb185*	Iso	Coimbatore, India	1981	Y. Fuyama
	mal Mys96*	Iso	Mysore, India,	1981	Y. Fuyama
	mal Rai*	Multi	Raichuri, India	1991	B. N. Singh
D. m. pallens	pal R242*	Iso	Puerto Princesa, Palawan, Philippines	1979	Y. Fuyama
	pal Q98*, 120*	Iso	Los Banos, Luzon, Philippines	1979	Y. Fuyama
	pal KB270*, 477*	Iso	Temburong, Brunei	2003	our collection
	pal KB (other)	Iso	Temburong, Brunei	2003	our collection
D. bipectinata	bip 381·0*	Multi	Patan, Nepal	1954	TDSSC
	bip 381·1*	Multi	Cabuyao, Laguna, Luzon	?	TDSSC
	bip 381·2*	Multi	Pago-Pago, Amer. Samoa	1967	TDSSC
	bip 381·3*	Multi	Chia-i, Taiwan	1967	TDSSC
	bip 381·4*	Multi	Samut Songkhram, Thailand	1971	TDSSC
	bip Pune*	Multi	Pune, India	1999	B. N. Singh
	bip 211·5*	Iso	Taiwan	1989	S.-C. Tsaur
	bip KB51*	Iso	Temburong, Brunei	2003	our collection
	bip KB (other)	Iso	Temburong, Brunei	2003	our collection

Appendix Table 1. (*Cont.*)

Species	Strain	Type[a]	Origin	Year[b]	Donor
	bip Lae321, 327	Iso	Lae, Papua New Guinea	1981	M. Watada
	bip PPG93,96	Iso	Pago-Pago, Amer. Samoa	1981	M. Watada
	bip TBU83	Iso	Tongatapu, Tonga	1981	M. Watada
	bip B96	Iso	Chiang Mai, Thailand	1979	M. Watada
	bip D97	Iso	Coimbatore, India	1979	M. Watada
	bip ISG-B19, C12	Iso	Ishigakijima, Okinawa, Japan	2002	M. Watada
	bip TKNK2, 3	Iso	Tokunoshima, Kagoshima, Japan	2002	M. Watada
D. parabipectinata	par 401·0*	Multi	Ari Ksatr, Cambodia	?	TDSSC
	par 401·2*	Multi	Tagaytay, Philippines	1967	TDSSC
	par Yun7*, 20*	Iso	Yunnan, China	2002	S.-C. Tsaur
	par Yun (other)	Iso	Yunnan, China	2002	S.-C. Tsaur
	par Bsb1*, 2	Iso	Temburong, Brunei	2003	our collection

Strains included in the nuclear data set are marked by asterisks.
[a] Iso, isofemale strains; Multi, strains established from multi-female collections.
[b] Collection year.
[c] Tucson *Drosophila* Species Stock Center.

Appendix Table 2. *Sequenced loci and amplification primers*

Locus	Total length[a]	Non-coding[b]	Forward primer	Reverse primer	Temperature (°C)[c]
al	530	432	GCTGGCGATGAAAATTGGA TTAAC	TAGGGATTATACGGATGCG ACTGG	55
bab2	2275	2275	AATCGAACMTTTCCAGCAAA TTGCWTT	ATTTACAGGCMGACGGTGATT GTTTAAA	52
Gpdh	1176	422	GTGGTGCCCCACCAGTTCAT	GGCTTGAGCTGATTTGTGCA	55
ple	968	699	CATCTTCCAGAGCACCCAG TATGTG	GTAGATGGGCTGGTACTCC TGATCC	55
Sod	1058	646	CCTCTAGAAATGGTGG TTAAAGCTGTNTGCCTT	GCTGAGCTCGTGTCCACCC TTGCCCA	52
up	583	583	GACGAGAGTCCCAGCTGTTTTT	GCATCCAACGTACCATCACATT	55
COI	553	0	CCAGCTGGAGGAGGAGATCC	CCAGTAAATAATGGGTATCAGTG	55

[a] Length of the multi-species alignments (single-species alignments are shorter in most cases).
[b] Number of non-coding nucleotide positions in the multi-species alignment.
[c] Recommended annealing temperature for PCR.

Appendix Table 3. *Recombination estimates*

	al	*bab2*	*Gpdh*	*ple*	*Sod*	*up*
γ/base pair						
D. malerkotliana	0·035	0·052	0·107	0·031	0·063	0·020
D. bipectinata	NE	0	0·159	0·022	0·116	NE
D. parabipectinata	NE	NE	0	0	0·080	NE
D. pseudoananassae	NE	NE	0·049	NE	0·028	NE
D. malerkotliana + *D. bipectinata*	0·006	0·051	0·052	0·025	0·080	0·040
D. bipectinata + *D. parabipectinata*	0	0·020	0·071	0·010	0·075	NE
D. malerkotliana + *D. bipectinata* + *D. parabipectinata*	0·003	0·029	0·065	0·016	0·093	0·056
Minimum number of recombination events						
D. malerkotliana	1	8	7	3	10	1
	530	2237	1125	961	1037	549
D. bipectinata	0	1	8	1	12	0
	518	2245	1116	960	1001	553
D. parabipectinata	0	0	0	0	3	0
	517	2247	1123	959	1020	553
D. pseudoananassae	0	0	2	0	1	0
	526	2252	1107	961	667	476

https://doi.org/10.1017/S0016672305007317 Published online by Cambridge University Press

Appendix Table 3. (*Cont.*)

	al	*bab2*	*Gpdh*	*ple*	*Sod*	*up*
D. malerkotliana + *D. bipectinata*	1	9	12	3	17	1
	530	2275	1176	968	1058	583
D. bipectinata + *D. parabipectinata*	0	6	13	2	14	0
	530	2275	1176	968	1058	583
D. malerkotliana + *D. bipectinata* + *D. parabipectinata*	1	14	16	4	20	1
	530	2275	1176	968	1058	583

The top part of the table shows the γ estimate of the population recombination rate per base pair, which tends to underestimate the true rate (Hey & Wakeley, 1997). The bottom part shows the minimum number of recombination events in the locus inferred by the four-gamete test (Hudson & Kaplan, 1985) (top number) and the total length of aligned sequences (bottom number). Recombination was estimated separately for each species, as well as for combined sequence samples that included two or three species.

Appendix Table 4. *Results of Shimodaira–Hasegawa tests*

	al	*bab2*	*Gpdh*	*ple*	*Sod*	Combined
al	–	0·001	0·106	0·000	0·000	0·357
bab2	0·042	–	0·447	0·064	0·224	0·377
Gpdh	0·000	0·000	–	0·000	0·000	0·171
ple	0·002	0·004	0·007	–	0·009	0·029
Sod	0·000	0·000	0·001	0·000	–	0·025
up	0·420	0·139	0·539	0·097	0·046	0·086

For each locus (rows), a set of trees (columns) was evaluated under the model of sequence evolution estimated for that locus. *P* values are shown for each test. The Bonferroni-corrected critical *P* value for this analysis is 0·0016 (0·05/31).

This work would not be possible without the help of many Drosophilists around the world. We are grateful to I. Bock, Y. Fuyama, M. Kimura, S. McEvey, M. Polak, B. N. Singh, M. Toda, S.-C. Tsaur and M. Watada, and the Tucson *Drosophila* Species Stock Center, for fly strains and information on species distribution. Our own collections in Brunei Darussalam were made possible by the generous help of Kamariah Hj. Abu Salim and the staff of the Kuala Belalong Field Studies Centre. We thank M. Aguade, D. Bolnick, M. Polak, A. Templeton, M. Turelli, T. Turner and anonymous reviewers for advice and comments on the manuscript, J. Fu, C. Mah, B. Miller, A. Siegel and A. Wong for technical help, and Sean Carroll (University of Wisconsin – Madison) for allowing this project to be initiated in his laboratory. This work was supported by funds from the University of California – Davis to A. K.

References

Ballard, J. W. (2000*a*). Comparative genomics of mitochondrial DNA in *Drosophila simulans*. *Journal of Molecular Evolution* **51**, 64–75.

Ballard, J. W. (2000*b*). When one is not enough: introgression of mitochondrial DNA in *Drosophila*. *Molecular Biology and Evolution* **17**, 1126–1130.

Ballard, J. W., Chernoff, B. & James, A. C. (2002). Divergence of mitochondrial DNA is not corroborated by nuclear DNA, morphology, or behavior in *Drosophila simulans*. *Evolution* **56**, 527–545.

Banerjee, R. & Singh, B. N. (1996). Inversion polymorphism in natural populations of *Drosophila bipectinata*. *Cytobios* **87**, 31–43.

Beltran, M., Jiggins, C. D., Bull, V., Linares, M., Mallet, J., McMillan, W. O. & Bermingham, E. (2002). Phylogenetic discordance at the species boundary: comparative gene genealogies among rapidly radiating *Heliconius* butterflies. *Molecular Biology and Evolution* **19**, 2176–2190.

Besansky, N. J., Krzywinski, J., Lehmann, T., Simard, F., Kern, M., Mukabayire, O., Fontenille, D., Toure, Y. & Sagnon, N. (2003). Semipermeable species boundaries between *Anopheles gambiae* and *Anopheles arabiensis*: evidence from multilocus DNA sequence variation. *Proceedings of the National Academy of Sciences of the USA* **100**, 10818–10823.

Bock, I. R. (1971*a*). Intra- and interspecific chromosomal inversions in the *Drosophila bipectinata* species complex. *Chromosoma* **34**, 206–229.

Bock, I. R. (1971*b*). Taxonomy of the *Drosophila bipectinata* species complex. *University of Texas Publications* **6**, 273–280.

Bock, I. R. (1978). The bipectinata complex: a study of interspecific hybridization in the genus *Drosophila*. *Australian Journal of Biological Science* **31**, 197–208.

Bock, I. R. (1980). Current status of the *Drosophila melanogaster* species-group. (Diptera). *Systematic Entomology* **5**, 341–356.

Bock, I. R. & Wheeler, M. R. (1972). The *Drosophila melanogaster* species group. *University of Texas Publications* **7**, 1–102.

Broughton, R. E. & Harrison, R. G. (2003). Nuclear gene genealogies reveal historical, demographic and selective factors associated with speciation in field crickets. *Genetics* **163**, 1389–1401.

Brower, A. V. (1994). Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. *Proceedings of the National Academy of Sciences of the USA* **91**, 6491–6495.

Charlat, S., Hurst, G. D. & Mercot, H. (2003). Evolutionary consequences of *Wolbachia* infections. *Trends in Genetics* **19**, 217–223.

Chatterjee, S. & Singh, B. N. (1993). Variation in the male sex comb teeth number in four species of the *Drosophila bipectinata* complex. *Drosophila Information Service* **72**, 162.

Coyne, J. A. (1992). Genetics and speciation. *Nature* **355**, 511–515.

Coyne, J. (1994). Evolutionary biology. Rules for Haldane's rule. *Nature* **369**, 189–190.

Coyne, J. & Orr, H. A. (1989). Patterns of speciation in *Drosophila*. *Evolution* **43**, 362–381.

Crochet, P. A., Chen, J. Z., Pons, J. M., Lebreton, J. D., Hebert, P. D. & Bonhomme, F. (2003). Genetic differentiation at nuclear and mitochondrial loci among large white-headed gulls: sex-biased interspecific gene flow? *Evolution* **57**, 2865–2878.

Crow, J. F. & Kimura, M. (1970). *An Introduction to Population Genetics Theory*. New York: Harper and Row.

Dobzhansky, T. (1970). *Genetics of the Evolutionary Process*. New York: Columbia University Press.

Duncan, I. M. (1982). *Polycomblike*: a gene that appears to be required for the normal expression of the *bithorax* and *antennapedia* gene complexes of *Drosophila melanogaster*. *Genetics* **102**, 49–70.

Farris, J. S., Kallersjo, M., Kluge, A. G. & Bult, C. (1995). Constructing a significance test for incongruence. *Systematic Biology* **44**, 570–572.

Ferris, S. D., Sage, R. D., Huang, C. M., Nielsen, J. T., Ritte, U. & Wilson, A. C. (1983). Flow of mitochondrial DNA across a species boundary. *Proceedings of the National Academy of Sciences of the USA* **80**, 2290–2294.

Filatov, D. A. (2002). ProSeq: a software for preparation and evolutionary analysis of DNA sequence data sets. *Molecular Ecology Notes* **2**, 621–624.

Fu, Y. X. & Li, W. H. (1993). Statistical tests of neutrality of mutations. *Genetics* **133**, 693–709.

Glor, R. E., Kolbe, J. J., Powell, R., Larson, A. & Losos, J. (2003). Phylogenetic analysis of ecological and morphological diversification in Hispaniolan trunk-ground anoles (*Anolis cybotes* group). *Evolution* **57**, 2383–2397.

Goldman, N., Anderson, J. P. & Rodrigo, A. G. (2000). Likelihood-based tests of topologies in phylogenetics. *Systematic Biology* **49**, 652–670.

Gonzalez, J., Ranz, J. M. & Ruiz, A. (2002). Chromosomal elements evolve at different rates in the *Drosophila* genome. *Genetics* **161**, 1137–1154.

Haldane, J. B. S. (1922). Sex-ratio and unidirectional sterility in hybrid animals. *Journal of Genetics* **12**, 101–109.

Hall, R. (2002). Cenozoic geological and plate tectonic evolution of SE Asia and the SW Pacific: computer-based reconstructions, model and animations. *Journal of Asian Earth Sciences* **20**, 353–431.

Hey, J. & Wakeley, J. (1997). A coalescent estimator of the population recombination rate. *Genetics* **145**, 833–846.

Holloway, J. D. & Hall, R. E. (1998). *Biogeography and Geological Evolution of SE Asia*. Leiden, The Netherlands: Backhuys.

Hudson, R. R. (1990). Gene genealogies and the coalescent process. *Oxford Surveys in Evolutionary Biology* **7**, 1–44.

Hudson, R. R. & Kaplan, N. L. (1985). Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* **111**, 147–164.

Hudson, R. R., Kreitman, M. & Aguade, M. (1987). A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**, 153–159.

Hudson, R. R., Boos, D. D. & Kaplan, N. L. (1992*a*). A statistical test for detecting geographic subdivision. *Molecular Biology and Evolution* **9**, 138–151.

Hudson, R. R., Slatkin, M. & Maddison, W. P. (1992*b*). Estimation of levels of gene flow from DNA sequence data. *Genetics* **132**, 583–589.

Huelsenbeck, J. P. & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755.

Huelsenbeck, J. P., Bull, J. J. & Cunningham, C. W. (1996). Combining data in phylogenetic analysis. *Trends in Ecology and Evolution* **11**, 152–158.

Huelsenbeck, J. P., Larget, B., Miller, R. E. & Ronquist, F. (2002). Potential applications and pitfalls of Bayesian inference of phylogeny. *Systematic Biology* **51**, 673–688.

Jha, A. P. & Rahman, S. M. Z. (1972). Cytogenetics of natural populations of *Drosophila*. I. Role of chromosomal inversions in the evolution of the bipectinata species complex. *Chromosoma* **37**, 445–454.

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.

Kopp, A. & True, J. R. (2002*a*). Evolution of male sexual characters in the oriental *Drosophila melanogaster* species group. *Evolution and Development* **4**, 278–291.

Kopp, A. & True, J. R. (2002*b*). Phylogeny of the Oriental *Drosophila melanogaster* species group: a multilocus reconstruction. *Systematic Biology* **51**, 786–805.

Lachaise, D., Cariou, M. L., David, J. R., Lemeunier, F., Tsacas, L. & Ashburner, M. (1988). Historical biogeography of the *Drosophila melanogaster* species subgroup. *Journal of Evolutionary Biology* **22**, 159–225.

Lakovaara, S. & Saura, A. (1982). Evolution and speciation in the *Drosophila obscura* group. In *The Genetics and Biology of Drosophila, vol. 3b* (Eds. M. Ashburner, H. L. Carson & J. N. Thompson) pp. 1–59. Academic Press, London.

Larget, B. & Simon, D. (1999). Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. *Molecular Biology and Evolution* **16**, 750–759.

Laurie, C. C., True, J. R., Liu, J. & Mercer, J. M. (1997). An introgression analysis of quantitative trait loci that contribute to a morphological difference between *Drosophila simulans* and *D. mauritiana*. *Genetics* **145**, 339–348.

Lemeunier, F., David, J., Tsacas, L. & Ashburner, M. (1986). The *melanogaster* species group. In *The Genetics and Biology of Drosophila, vol. 3e* (Eds. M. Ashburner, H. L. Carson & J. N. Thompson) pp. 147–256. Academic Press, London.

Machado, C. A. & Hey, J. (2003). The causes of phylogenetic conflict in a classic *Drosophila* species group. *Proceedings of the Royal Society of London, Series B* **270**, 1193–1202.

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. *Molecular Biology and Evolution* **19**, 472–488.
Martin, J., Guryev, V. & Blinov, A. (2002). Population variability in *Chironomus* (*Camptochironomus*) species (Diptera, Nematocera) with a Holarctic distribution: evidence of mitochondrial gene flow. *Insect Molecular Biology* **11**, 387–397.
McMillan, W. O., Weigt, L. A. & Palumbi, S. R. (1999). Color pattern evolution, assortative mating, and genetic differentiation in brightly colored butterflyfishes (Chaetodontidae). *Evolution* **53**, 247–260.
Mukai, T., Naruse, K., Sato, T., Shima, A. & Morisawa, M. (1997). Multiregional introgressions inferred from the mitochondrial DNA phylogeny of a hybridizing species complex of gobiid fishes, genus *Tridentiger*. *Molecular Biology and Evolution* **14**, 1258–1265.
Narda, R. D. (1969). Genetical relationship and isolating mechanisms in *Ananassae* species subgroup (*Sophophora*, *Drosophila*). 3. Chromosome behaviour during meiosis in *D. bipectinata*, *D. malerkotliana* and their hybrid. *Cytologia* **34**, 126–132.
Noor, M. A., Johnson, N. A. & Hey, J. (2000). Gene flow between *Drosophila pseudoobscura* and *D. persimilis*. *Evolution* **54**, 2174–2175; discussion 2176–2177.
Okada, T. (1979). Subspeciation of the *Drosophila bipectinata*-complex in view of distribution. *Zoological Magazine, Tokyo* **88**, 673.
Orr, H. A. (1993). Haldane's rule has multiple genetic causes. *Nature* **361**, 532–533.
Ortiz-Barrientos, D., Reiland, J., Hey, J. & Noor, M. A. (2002). Recombination and the divergence of hybridizing species. *Genetica* **116**, 167–178.
Palopoli, M. F. & Wu, C. I. (1994). Genetics of hybrid male sterility between *Drosophila* sibling species: a complex web of epistasis is revealed in interspecific studies. *Genetics* **138**, 329–341.
Patterson, J. T. & Stone, W. S. (1952). *Evolution in the Genus* Drosophila. New York: Macmillan.
Polak, M., Starmer, W. T. & Wolf, L. (2004). Sexual selection for size and symmetry in a diversifying secondary sexual character in *Drosophila bipectinata* Duda (Diptera: Drosophilidae). *Evolution* **58**, 597–607.
Posada, D. & Crandall, K. A. (2002). The effect of recombination on the accuracy of phylogeny estimation. *Journal of Molecular Evolution* **54**, 396–402.
Powell, J. R. (1983). Interspecific cytoplasmic gene flow in the absence of nuclear gene flow: evidence from *Drosophila*. *Proceedings of the National Academy of Sciences of the USA* **80**, 492–495.
Powell, J. R. (1997). *Progress and Prospects in Evolutionary Biology: The Drosophila Model*. New York: Oxford University Press.
Ramon, M. L., Lobel, P. S. & Sorenson, M. D. (2003). Lack of mitochondrial genetic structure in hamlets (*Hypoplectrus* spp.): recent speciation or ongoing hybridization? *Molecular Ecology* **12**, 2975–2980.
Ranz, J. M., Caceres, M. & Ruiz, A. (1999). Comparative mapping of cosmids and gene clones from a 1·6 Mb chromosomal region of *Drosophila melanogaster* in three species of the distantly related subgenus *Drosophila*. *Chromosoma* **108**, 32–43.
Rognon, X. & Guyomard, R. (2003). Large extent of mitochondrial DNA transfer from *Oreochromis aureus* to *O. niloticus* in West Africa. *Molecular Ecology* **12**, 435–445.
Rowan, R. G. & Hunt, J. A. (1991). Rates of DNA change and phylogeny from the DNA sequences of the alcohol dehydrogenase gene for five closely related species of Hawaiian *Drosophila*. *Molecular Biology and Evolution* **8**, 49–70.
Rozas, J. & Rozas, R. (1999). DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**, 174–175.
Russo, C. A. M., Takezaki, N. & Nei, M. (1995). Molecular phylogeny and divergence times of drosophilid species. *Molecular Biology and Evolution* **12**, 391–404.
Salzburger, W. & Meyer, A. (2004). The species flocks of East African cichlid fishes: recent advances in molecular phylogenetics and population genetics. *Naturwissenschaften* **91**, 277–290.
Schawaroch, V. (2002). Phylogeny of a paradigm lineage: the *Drosophila melanogaster* species group. *Biological Journal of the Linnean Society* **76**, 21–37.
Schierup, M. H. & Hein, J. (2000). Consequences of recombination on traditional phylogenetic analysis. *Genetics* **156**, 879–891.
Shaw, K. L. (2002). Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: what mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. *Proceedings of the National Academy of Sciences of the USA* **99**, 16122–16127.
Shimodaira, H. & Hasegawa, M. (1999). Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution* **16**, 1114–1116.
Singh, S. & Singh, B. N. (2001). *Drosophila bipectinata* species complex. *Indian Journal of Experimental Biology* **39**, 835–844.
Solignac, M. (2004). Mitochondrial DNA in the *Drosophila melanogaster* complex. *Genetica* **120**, 41–50.
Sota, T. (2002). Radiation and reticulation: extensive introgressive hybridization in the carabid beetles *Ohomopterus* inferred from mitochondrial gene genealogy. *Population Ecology* **44**, 145–156.
Sota, T. & Vogler, A. P. (2001). Incongruence of mitochondrial and nuclear gene trees in the Carabid beetles *Ohomopterus*. *Systemic Biology* **50**, 39–59.
Sota, T., Ishikawa, R., Ujiie, M., Kusumoto, F. & Vogler, A. P. (2001). Extensive trans-species mitochondrial polymorphisms in the carabid beetles *Carabus* subgenus *Ohomopterus* caused by repeated introgressive hybridization. *Molecular Ecology* **10**, 2833–2847.
Swofford, D. L. (2000). *PAUP*: Phylogenetic Analysis Using Parsimony*, version 4.0. Sunderland, Mass: Sinauer.
Tajima, F. (1983). Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**, 437–60.
Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595.
Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). ClustalW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**, 4673–4680.
Ting, C. T., Tsaur, S. C. & Wu, C. I. (2000). The phylogeny of closely related species as revealed by the genealogy of a speciation gene, Odysseus. *Proceedings of the National Academy of Sciences of the USA* **97**, 5313–5316.
Turelli, M. (1998). The causes of Haldane's rule. *Science* **282**, 889–891.
Turelli, M. & Hoffmann, A. A. (1995). Cytoplasmic incompatibility in *Drosophila simulans*: dynamics and parameter estimates from natural populations. *Genetics* **140**, 1319–1338.

Voris, H. (2000). Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time durations. *Journal of Biogeography* **27**, 1153–1167.

Wakeley, J. & Hey, J. (1997). Estimating ancestral population parameters. *Genetics* **145**, 847–855.

Wang, R. L. & Hey, J. (1996). The speciation history of *Drosophila pseudoobscura* and close relatives: inferences from DNA sequence variation at the period locus. *Genetics* **144**, 1113–1126.

Wang, R. L., Wakeley, J. & Hey, J. (1997). Gene flow and natural selection in the origin of *Drosophila pseudoobscura* and close relatives. *Genetics* **147**, 1091–1106.

Watterson, G. A. (1975). On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology* **7**, 256–275.

Werren, J. H., Zhang, W. & Guo, L. R. (1995). Evolution and phylogeny of *Wolbachia*: reproductive parasites of arthropods. *Proceedings of the Royal Society of London, Series B*, **261**, 55–63.

Wiens, J. J. (1998). Combining data sets with different phylogenetic histories. *Systematic Biology* **47**, 568–581.

Willett, C. S., Ford, M. J. & Harrison, R. G. (1997). Inferences about the origin of a field cricket hybrid zone from a mitochondrial DNA phylogeny. *Heredity* **79**, 484–494.

Wu, C. I. (2001). The genic view of the process of speciation. *Journal of Evolutionary Biology* **14**, 851–865.

Wu, C. I. & Ting, C. T. (2004). Genes and speciation. *Nature Reviews Genetics* **5**, 114–122.

Yang, S. Y., Wheeler, M. R. & Bock, I. R. (1972). Isozyme variations and phylogenetic relationships in the *Drosophila bipectinata* species complex. *University of Texas Publications* **7**, 213–227.

Yang, Z. & Rannala, B. (1997). Bayesian phylogenetic inference using DNA sequences: a Markov chain Monte Carlo method. *Molecular Biology and Evolution* **14**, 717–724.

Zhou, W., Rousset, F. & O'Neil, S. (1998). Phylogeny and PCR-based classification of *Wolbachia* strains using *wsp* gene sequences. *Proceedings of the Royal Society of London, Series B*, **265**, 509–515.