Genet. Res., Camb. (2006), **88**, *pp.* 13–26. © 2006 Cambridge University Press
doi:10.1017/S0016672306008330 Printed in the United Kingdom

The speciation history of the *Drosophila nasuta* complex

DORIS BACHTROG*
Section of Ecology, Behavior and Evolution, University of California San Diego, La Jolla, CA 92093, USA

(*Received 10 January 2006 and in revised form 14 April 2006*)

Summary

The *Drosophila nasuta* subgroup of the *immigrans* species group is widely distributed throughout the South-East Asian region, consisting of morphologically similar species with varying degrees of reproductive isolation. Here, I report nucleotide variability data for five X-linked and two mtDNA loci in eight taxa from the *nasuta* subgroup, with deeper sampling from *D. albomicans* and its sister species *D. nasuta*. Phylogenetic relationships among these species vary among different genomic regions, and levels of genetic differentiation suggest that this species group diversified only about one million years ago. *D. albomicans* and *D. nasuta* share nucleotide polymorphisms and are distinguished by relatively few fixed differences. Patterns of genetic differentiation between this species pair are compatible with a simple isolation model with no gene flow. Nucleotide variability levels of species in the *nasuta* group are comparable to those in members of the *melanogaster* and *pseudoobscura* species groups, indicating effective population sizes on the order of several million. Population genetic analyses reveal that summaries of the frequency distribution of neutral polymorphisms in both *D. albomicans* and *D. nasuta* generally fit the assumptions of the standard neutral model. *D. albomicans* is of particular interest for evolutionary studies because of its recently formed neo-sex chromosomes, and our phylogenetic and population genetic analyses suggest that it might be an ideal model to study the very early stages of Y chromosome evolution.

1. Introduction

The *Drosophila nasuta* subgroup of the *immigrans* species group is widely distributed throughout the South-East Asian region (Wilson *et al.*, 1969), and members have also been recorded from Africa and Hawaii. The *nasuta* complex contains, depending on the taxonomic system, nine to 14 taxa, including *D. nasuta*, *D. albomicans*, *D. pulaua*, *D. kepulauana*, *D. kohkoa*, *D. sulfurigaster sulfurigaster*, *D. s. albostrigata*, *D. s. bilimbata*, *D. neonasuta*, *D. pallidifrons*, *D. niveifrons* and Taxon F, I and J (Kitagawa, 1991). This group exhibits several features that make it evolutionarily interesting, such as multiple speciation events in a short time period resulting in morphologically very similar species with varying degrees of reproductive isolation (Wilson *et al.*, 1969) and extensive chromosomal polymorphism (Hatsumi *et al.*, 1988; Suzuki *et al.*, 1990; Wilson *et al.*, 1969). The basic karyotype configuration of the *immigrans* group is a rod X and rod Y, a pair of dots (chromosome 4), a pair of Vs (chromosome 2) and a double length rod (chromosome 3). This double length rod is composed of two of the five long rods of the primitive *Drosophila* karyotype (Fig. 1). *D. albomicans* is of particular interest in the *nasuta* subgroup. The karyotype of this species is $2n=6$ instead of $2n=8$ (all other *nasuta* members), resulting from the fusion of the sex chromosomes with the third chromosomes (Patterson & Stone 1952; Wilson *et al.*, 1969). The fusion of an autosome to the ancestral Y chromosome parallels the karyotype of *D. miranda*, which has served as a useful model species to study the evolution of Y chromosomes (Bachtrog, 2005; Steinemann & Steinemann, 1998). The neo-Y chromosome of *D. albomicans* might be very young on an

* Division of Biological Sciences, University of California, San Diego, 9500 Gilman Drive, MC 0116, La Jolla, CA 92093, USA. Tel (office): +1(858) 8226762. Tel (lab): +1(858) 8221832. Fax: +1(858) 5347108. e-mail: dbachtrog@ucsd.edu

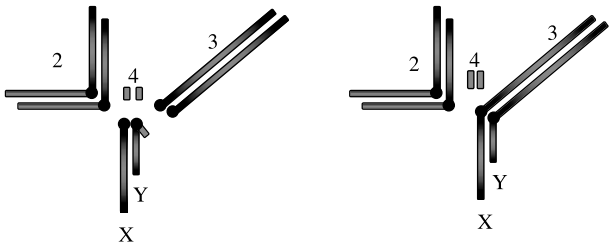

Fig. 1. Karyotype of the *D. nasuta* group (males are shown). The basic karyotype configuration of the *nasuta* group is a rod X and rod Y, a pair of dots (chromosome 4), a pair of Vs (chromosome 2), and the double length rod (chromosome 3), composed of two of the five long rods of the primitive *Drosophila* karyotype. In *D. albomicans*, the third chromosomes fused to the sex chromosomes.

evolutionary time scale, since *D. albomicans* and its close sister species *D. nasuta* (which lacks this Y–autosome fusion) are interfertile (Wilson *et al.*, 1969) and show little divergence at mtDNA (Yu *et al.*, 1999). Thus, *D. albomicans* could provide a potent model system to study the very first stages of sex chromosome evolution.

However, very little is known about general patterns of molecular variation and divergence in this species group. For evolutionary studies in this species group, it is necessary to develop an historical framework of the timing of speciation events and the degree of separation between taxa. Despite extensive analyses, including morphometric (Kitagawa *et al.*, 1982; Wilson *et al.*, 1969), reproductive isolation (Kitagawa, 1991; Wilson *et al.*, 1969), chromosomal evolution (Kitagawa *et al.*, 1982; Suzuki *et al.*, 1990; Wilson *et al.*, 1969), courtship song (Shao *et al.*, 1997), isozymes (Kitagawa, 1991) and mtDNA restriction and variability studies (Chang *et al.*, 1989; Yu *et al.*, 1999; Nagaraja *et al.*, 2004), the phylogenetic relationships among taxa of this group remain unclear.

Molecular variation within and between species contains information on a species' demographic (i.e. temporal and spatial) history. Here, I obtain multilocus sequence data from eight taxa and combine phylogenetic and population-genetic approaches to reconstruct the order of speciation events in this species group. Most previous studies considered only one individual per taxon for phylogenetic reconstruction. However, given the recent radiation of this group and a lack of reproductive isolation between several species, these taxa might not form monophyletic groups (Wilson *et al.*, 1969). In addition, no phylogeny based on nuclear DNA sequence markers has been performed, and mtDNA has been found to frequently introgress across species boundaries in other *Drosophila* species (Bachtrog *et al.*, 2006; Ballard, 2000; Powell, 1983). To overcome these limitations, I analyse two or more strains from most taxa, and investigate both mtDNA and nuclear sequence markers. In addition, with deeper sampling of individuals from *D. albomicans* and its sister species *D. nasuta*, I estimate the extent of genetic differentiation among this species pair and analyse patterns of molecular variation within species. This analysis provides a first important step towards characterizing the utility of the *D. albomicans* neo-sex chromosomes as a novel model system to study the very early stages of Y chromosome degeneration.

2. Materials and methods

(i) *Fly stocks*

A total of 23 lines of *D. albomicans*, six lines of *D. nasuta*, one line of *D. pulaua*, two lines of *D. kepulauana*, two lines of *D. kohkoa*, two lines of *D. s. albostrigata*, two lines of *D. s. bilimbata* and two lines of *D. s. sulfurigaster* were used for this study. A summary of all fly strains used, including their geographic origin, is given in Table 1. For a subset of the *D. albomicans* lines, we confirmed the karyotype using mitotic chromosome preparations (Sullivan *et al.*, 2000).

(ii) *DNA sequencing*

Genomic DNA was extracted from a single male of each line using the Puregene DNA extraction kit (Gentra Systems). PCR products were amplified as 360–1077 bp fragments from genomic DNA using standard methods (Sambrook *et al.*, 1989). Primer pairs were designed from EST sequences generated from a *D. albomicans* cDNA library (strain 15112-1751.4) for X-linked genes, or using conserved primers for mtDNA loci (Table 2). PCR products were used as sequencing templates after treatment with the SAP/EXO reagent. The original amplification primers, and if necessary internal sequencing primers, were used for sequencing with the BigDye 3.0 cycle sequencing kit (Applied Biosystems, Foster City, CA) following the manufacturer's protocol. Sequences were run on an ABI 3730 automated sequencer, and each fragment was sequenced at least once on both strands. I obtained polymorphism data for five regions from the X chromosome (a total of 2723 bp) and two regions from the mtDNA (1827 bp). Sequence trace files were proofread and aligned using Sequencher (Gene Codes Corporation).

(iii) *Phylogenetic reconstruction*

The presence of intragenic recombination, as observed in the data (see below) complicates interpretations of the phylogenetic history of a species group

https://doi.org/10.1017/S0016672306008330 Published online by Cambridge University Press

Table 1. *Origin of isofemale lines used in this study*

Species	Country	Origin	Isofemale lines
D. albomicans	Thailand	Bon Chakkrarat	15112-1751.04
		Chiang Mai	E-10814
	Japan	Fukui, Fukui	E-10826
		Haterumajima, Okinawa	E-10809
		Iriomotejima, Okinawa	E-10806, E-10807
		Ishigakijima, Okinawa	E-10808, E-10824
		Kumejima, Okinawa	E-10811, E-10812, E-10827, E-10828
		Okinawa Honto, Okinawa	E-10821, E-10822, E-10823
		Miyakojima, Okinawa	E-10802, E-10803, E-10804, E-10805
		Tokunoshima, Kagoshima	E-10819, E-10820
	Taiwan	Nakang	15112-1751.03
	Pescadores	Peng Hu Island	15112-1751.02
D. nasuta	Madagascar	Tananarive	15112-1781.09
	Kenya	Mombasa	15112-1781.06
	France	Seychelles Island	15112-1781.01
	India	Sri Lanka	15112-1781.11
		Mysore	15112-1781.00
	Malaysia	Penang	E-10813
D. pulaua	Malaysia	Semongok Forest Reserve, Sarawak	15112-1801.00
D. kepulauana	Malaysia	Sarawak	15112-1761.01
	Borneo	Brunei	15112-1761.02
D. kohkoa	Malaysia	Sarawak	15112-1771.01
	Philippines	Rizal	15112-1771.04
D. s. albostrigata	Malaysia	Kuala Lumpur	15112-1811.00
	Borneo	Brunei	15112-1811.05
D. s. bilimbata	Tonga Islands	Tongatapu	15112-1821.05
	Fiji	Vanua Levu	15112-1821.07
D. s. sulfurigaster	Australia	Queensland	15112-1831.00
	New Guinea	Wau	15112-1831.02

Table 2. *Sequence length and primer information for the loci studied*

Gene	No. of sites	Silent sites	Replacement sites	Forward	Reverse
CG10944	554	163·24	381·76	GTGCGTTTGCTGTTGAAGAA	TCTTAGACGATGTGGGCAGA
CG12056	549	130·36	418·64	AGACTACGCCACCACAAAGG	GGTCATCTAACTCCGCTTCG
CG2930	627	150·51	476·49	GGGTGGCATAAAACCTTGTG	TGGTAAGCTTTTGCAGTGGA
CG3699	633	159·92	473·08	CTGCCATTGCCGAGGTGTT	AAATTGTGCTTGCCACCATC
sesB	360	86·67	273·33	CTGCTGATACCGGCAAGG	ACAACACAAGCACGAAAGCA
COII	750	205·54	541·46	GTTTAAGAGACCAGTACTTG	ATGGCAGATTAGTGCAATGG
ND5	1077	239·32	835·68	TCTAACAATTTCTAAAATTATATCC	TTTGTGGTGTTAGTGATATGAA

(Felsenstein, 1988). In particular, recombination causes the different portions of a locus to have different evolutionary histories (Hudson, 1990). Thus, a tree reconstructed from each locus is not a real gene tree, but rather a superposition of multiple genealogies. Additionally, recombination shuffles variation among individuals within the species, generating the impression of homoplasy (Schierup & Hein, 2000). The presence of intragenic recombination particularly affects phylogenetic inference using maximum parsimony or maximum likelihood; thus I used a neighbour-joining (NJ) algorithm to reconstruct *composite gene trees*. I used the MEGA3.1 (Kumar *et al.*, 2004) software to create gene trees with

https://doi.org/10.1017/S0016672306008330 Published online by Cambridge University Press

Tamura–Nei distances (Tamura & Nei, 1993) for each gene individually and for the combined data set. *D. virilis* was used as an outgroup species to root the phylogenetic trees. The reliability of the inferred consensus trees was tested by bootstrapping using 1000 pseudoreplications. Note that the phylogenetic trees are used mainly for visual purposes, and no statistical inferences are drawn from them.

(iv) *Population genetic analysis*

The estimated number of synonymous sites, non-synonymous sites, average pairwise diversity (π), average pairwise divergence (D_{xy}), as well as counts of the number of polymorphic (P) and divergent (D) sites were performed with the DnaSP software package (version 4; http://www.ub.es/dnasp/). Multiply hit sites were included in the analysis but insertion–deletion polymorphisms were excluded. For locus- and species-specific estimates of variability see Table 3.

To detect non-equilibrium demography and/or selection in *D. albomicans* and *D. nasuta*, I used two multilocus statistical tests of the distribution of polymorphism frequencies. In particular, I used the means and variances of Tajima's *D* (Tajima, 1989) and Fay and Wu's *H* (Fay & Wu, 2000). The ancestral state of polymorphisms was inferred using a single *D. pulaua* sequence and standard parsimony criteria. *D. pulaua* was chosen as the outgroup since it clearly clusters outside the *D. albomicans*/*D. nasuta* clade in the inferred consensus tree.

All tests of neutrality were carried out using the neutral coalescent simulation program *ms* of Hudson (2002) and various auxiliary programs written in C and Perl by P. Andolfatto (see Haddrill *et al.*, 2005). These simulations are used to test a model of neutral sequence evolution (the standard neutral model, SNM) by first creating a null distribution for the test statistics of interest, and then comparing this distribution with the observed value of the test statistics in the real data. In the neutral simulations, I account for sample size, alignment length and θ for each locus. The parameter θ (i.e. the neutral population mutation rate) is estimated from the observed data using the HKA framework (Hudson *et al.*, 1987) based on the number of segregating sites and divergence to a single *D. pulaua* sequence.

Ignoring recombination when the data show evidence for genetic exchange is done at a great loss of power and is overly conservative when the fit of data to the SNM is tested (Przeworski *et al.*, 2001). I thus incorporate recombination into the simulations by using a point estimate of ρ/θ (the population recombination rate ρ versus the neutral population mutation rate), based on the mode of the posterior distribution of ρ/θ obtained for X-linked genes in *D. albomicans* using an approximate Bayesian method (Haddrill *et al.*, 2005, K. Thornton, unpublished). Posterior distributions of ρ and θ were jointly estimated based on summary statistics of the data (sample size, alignment length, number of segregating sites, number of haplotypes, and the minimum number of recombination events in the sample) and rejection sampling. This method assumes that ρ per site and θ per site are constant among loci. The prior distributions used were $\theta \sim$ Uniform (0·0025–0·01) and $\rho \sim$ Uniform (0·009–0·18), with ε (the tolerance) set to 0·1. The posterior densities for θ and ρ are contained within the limits of the uniform priors. I estimated ρ and θ based on all sites. Since estimates of ρ using a similar procedure were shown to be biased upward when the number of polymorphisms is small (Andolfatto & Wall, 2003), I only used loci that had a minimum of 10 segregating sites (i.e. three X-linked loci). The estimate of ρ and θ is based on 1000 draws from the posterior distribution.

P values for test statistics (i.e. Tajima's *D* and Fay & Wu's *H*) are based on 10 000 simulated replicates of the neutral coalescent. Note that since the simulations are run based on point estimates of both ρ and θ (instead of integrating over the uncertainty in these estimates) the true variance of the test statistics will be underestimated, which will be conservative for the conclusions drawn here (see Section3).

(v) *Estimating divergence times between* D. albomicans *and* D. nasuta

To estimate divergence times (T) between *D. albomicans* and *D. nasuta*, I used a summary likelihood approach based on the HKA framework (Hudson *et al.*, 1987), as implemented by Bachtrog *et al.* (2006). Briefly, this method assumes that an ancestral population with a neutral mutation rate θ splits into two populations at time T having neutral mutation rates θ and $f\theta$, where f represents a reduced effective population size (where $f \leqslant 1$). The parameters θ and f are estimated from the data for each locus and species. θ is estimated from the observed data using the HKA framework based on the number of segregating sites and divergence to a single *D. pulaua* sequence, and f is estimated as $f = \sum_{j=1}^{m} \theta_{1,j} / \sum_{j=1}^{m} \theta_{2,j}$ for the m segments surveyed on the X ($m=5$) and the mtDNA ($m=2$), respectively, where $\theta_{i,j}$ is the estimated population mutation rate in population i at locus j. Note that estimating f based on θ assumes that homologous loci have the same mutation rate, and I only use a point estimate of f. Following the HKA framework, I estimate the relative divergence time T_j for each locus j as

$$T_j = (D_{12,j}/\theta_{1,j}) - 1$$

https://doi.org/10.1017/S0016672306008330 Published online by Cambridge University Press

Table 3. *Polymorphism levels in the* nasuta *group for X-linked and mtDNA markers*

Locus	Taxon	*n*	S (silent)	S (replacement)	π (silent) %
		X-linked			
CG10944	*D. albomicans*	23	10	0	1·76
	D. nasuta	6	3	0	0·97
	D. kepulauana	2	2	0	1·60
	D. kohkoa	2	3	0	2·39
	D. s. albostrigata	2	0	0	0
	D. s. sulfurigaster	2	3	1	2·39
	D. s. bilimbata	2	0	0	0
CG12056	*D. albomicans*	23	9	3	2·49
	D. nasuta	6	1	1	0·46
	D. kepulauana	2	3	2	2·30
	D. kohkoa	2	4	1	3·07
	D. s. albostrigata	2	3	0	2·30
	D. s. sulfurigaster	2	1	0	0·77
	D. s. bilimbata	2	0	0	0
CG2930	*D. albomicans*	23	2	0	0·17
	D. nasuta	6	2	0	0·44
	D. kepulauana	2	3	1	1·99
	D. kohkoa	2	9	0	5·97
	D. s. albostrigata	2	3	0	1·99
	D. s. sulfurigaster	2	1	0	0·66
	D. s. bilimbata	2	0	1	0
CG3699	*D. albomicans*	23	18	0	2·85
	D. nasuta	6	10	1	2·46
	D. kepulauana	2	7	0	4·38
	D. kohkoa	2	6	1	3·75
	D. s. albostrigata	2	0	0	0
	D. s. sulfurigaster	2	0	0	0
	D. s. bilimbata	2	0	0	0
sesB	*D. albomicans*	23	0	0	0
	D. nasuta	6	1	0	0·39
	D. kepulauana	2	0	0	0
	D. kohkoa	2	0	0	0
	D. s. albostrigata	2	0	0	0
	D. s. sulfurigaster	2	1	0	1·15
	D. s. bilimbata	2	0	0	0
		mtDNA			
COII	*D. albomicans*	23	14	1	1·03
	D. nasuta	6	6	0	1·07
	D. kepulauana	2	5	0	2·42
	D. kohkoa	2	15	1	6·78
	D. s. albostrigata	2	3	2	1·45
	D. s. sulfurigaster	2	14	0	6·81
	D. s. bilimbata	2	3	0	1·45
ND5	*D. albomicans*	23	36	2	3·39
	D. nasuta	6	15	0	2·28
	D. kepulauana	2	15	0	6·26
	D. kohkoa	2	27	1	11·3
	D. s. albostrigata	2	9	0	3·76
	D. s. sulfurigaster	2	33	0	13·8
	D. s. bilimbata	2	8	1	3·33
		Average			
X-linked	*D. albomicans*	23	39	3	1·58
	D. nasuta	6	17	2	1·03
	D. kepulauana	2	15	3	2·26
	D. kohkoa	2	22	2	3·31
	D. sulfurigaster albostrigata	2	6	0	0·87
	D. sulfurigaster sulfurigaster	2	6	1	1·00
	D. sulfurigaster bilimbata	2	0	1	0

https://doi.org/10.1017/S0016672306008330 Published online by Cambridge University Press

Table 3. (*cont.*)

Locus	Taxon	*n*	S (silent)	S (replacement)	π (silent) %
mtDNA	*D. albomicans*	23	50	3	2·29
	D. nasuta	6	21	0	2·02
	D. kepulauana	2	20	0	4·48
	D. kohkoa	2	42	2	9·2
	D. sulfurigaster albostrigata	2	12	2	2·69
	D. sulfurigaster sulfurigaster	2	47	0	10·6
	D. sulfurigaster bilimbata	2	11	1	2·47

where $D_{12,j}$ is the average pairwise divergence between species 1 and 2 at locus j, and this relative time is transformed into units of absolute time as

$$T_{\text{abs},j} = (T_j \theta_{1,j})/(2\mu_j L_j)$$

where L_j is the number of silent sites surveyed and μ_j is the mutation rate per silent site per year for locus j. I assume an average mutation rate $\mu = 1{\cdot}5 \times 10^{-8}$ per silent site per year (Li, 1997) for loci on the X chromosome, and mutation rates for loci on the mtDNA are scaled based on Jukes–Cantor corrected average pairwise silent divergences to *D. pulaua* for mtDNA relative to the X chromosome.

Confidence intervals for T are estimated from simulated genealogical samples that closely resemble the actual data using the program *ms* (Hudson, 2002). For each locus, I generate a neutral sample from two subdivided populations with n_1 and n_2 individuals that split T_{abs} years ago, conditioning on $\theta_{1,j}$ and f. T_{abs} is varied over an interval from 0 to 3 million years in increments of 0·1. For each simulated dataset, I calculate $T_{\text{abs},j,\text{sim}}$ using the equation above and estimate the likelihood of T_{abs} as

$$Lik(T_{\text{abs},j} \mid \theta_{1,j}, f, T_j) \propto Prob\,[(T_{\text{abs},j,\text{sim}} - T_{\text{abs},j,\text{obs}}) < \varepsilon]$$

where $T_{\text{abs},j,\text{obs}}$ is the absolute divergence time of locus j estimated from the real data, and ε is an arbitrary rejection criterion set to 0·05. For each value of T_{abs}, I generated 100 000 samples. I incorporate intragenic recombination in the simulations for X-linked loci, as estimated by the approximate Bayesian method. This means that I set $\rho = 5{\cdot}2\theta$ for simulations of X-linked loci and $\rho = 0$ for simulations of the mtDNA. I generate a joint likelihood surface for T_{abs} as

$$Lik(T_{\text{abs},X}) = \prod_{j=1}^{m} Lik(T_{\text{abs},j}) | \theta_{1,j}, f, T_j).$$

The estimate of T is taken as the maximum of $Lik(T_{\text{abs},j})$ and the approximate 95% credibility interval (CI) as 2 units of log likelihood. Again, since only point estimates of f, ρ and θ are used, the true variance in T will be underestimated. Thus, the CI inferred using this method should be viewed as a lower bound for the CI of the divergence time between *D. nasuta* and *D. albomicans*.

3. Results

(i) *Phylogenetic relationships in the* D. nasuta *subgroup*

Phylogenetic analyses of X-linked markers and mtDNA sequences were performed to examine the inferred evolutionary relationships among the members of the *nasuta* subgroup. Neighbour-joining trees reconstructed from each individual X-linked marker are shown in Fig. 2. These trees should not be equated with gene trees, as most loci show evidence of recombination and thus do not have a simple bifurcating gene tree history. Instead, these trees are intended to serve as visual guides of locus-to-locus patterns of differentiation. For most genes, the pattern of relatedness among sequences is not consistent with reciprocally monophyletic species groups. The *D. albomicans* sequences are often intermingled with the *D. nasuta* sequences, but sometimes also cluster with more distantly related species (Fig. 2). Two of the X-linked loci, *sesB* and *CG2930*, support monophyly for the *D. albomicans* clade. The other species investigated also often do not cluster with conspecific individuals. These patterns are expected for very recently diverged species if multiple gene lineages persist in the descended species since the time of speciation (Felsenstein, 1988). Consistent with recent divergence of the *nasuta* group, many species pairs have several shared polymorphisms and few fixed differences (Table 4). Consensus trees from the combined analysis of all five nuclear and two mtDNA markers are shown in Fig. 3. In the consensus tree, most species form monophyletic groups. *D. albomicans* is most closely related to *D. nasuta*, and these two species form a monophyletic cluster that is supported by both the nuclear and the mtDNA consensus tree. In the consensus nuclear tree, *D. albomicans* is paraphyletic, and *D. nasuta* forms a monophyletic cluster within the *D. albomicans* sequences. Only the two *D. kohkoa* individuals do not cluster as a monophyletic group, in both the nuclear and the mtDNA phylogenies (Fig. 3).

https://doi.org/10.1017/S0016672306008330 Published online by Cambridge University Press

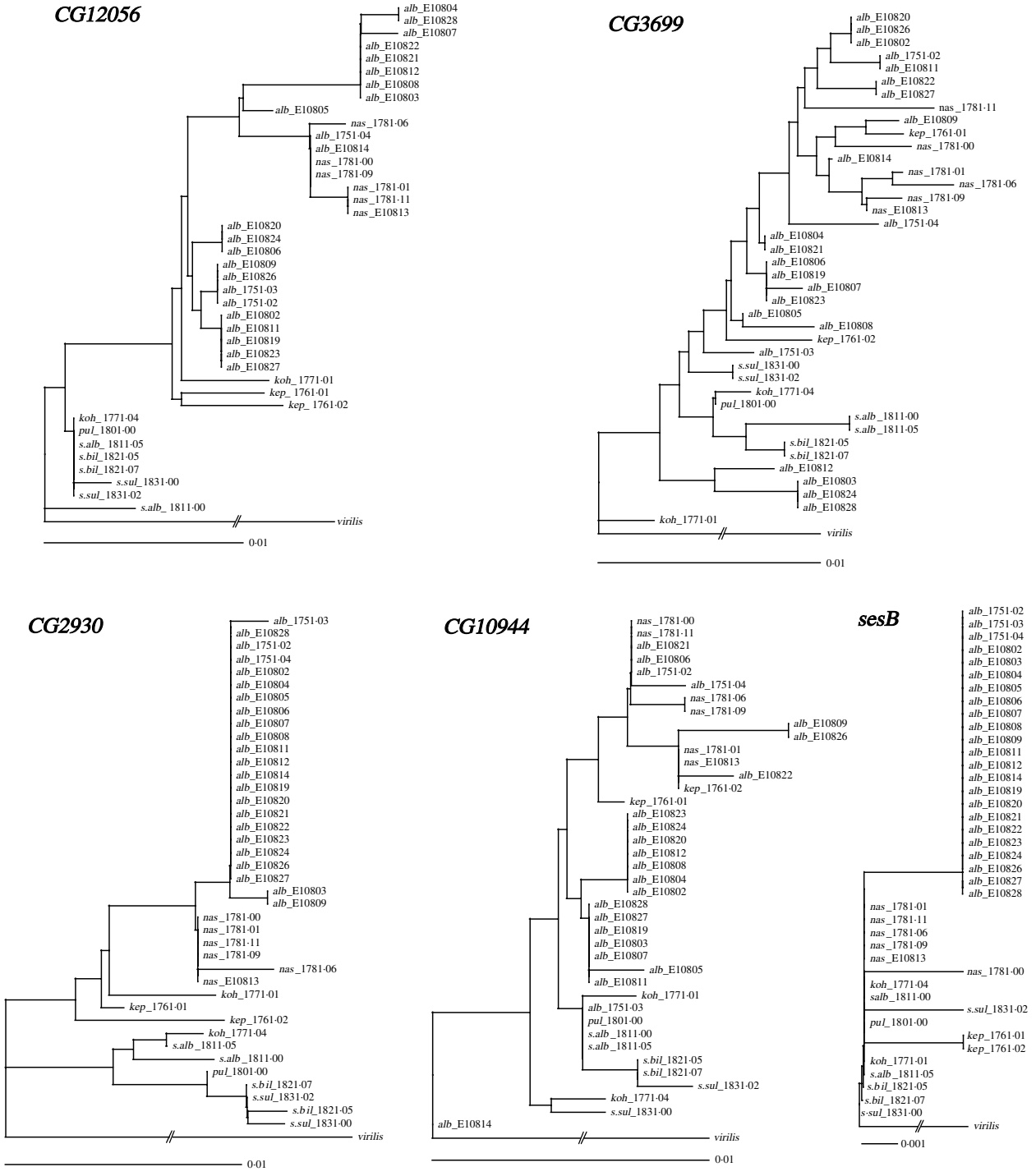

Fig. 2. Neighbour-joining trees constructed for individual X-linked loci. alb, *D. albomicans*; nas, *D. nasuta*; kep, *D. kepulauana*; koh, *D. kohkoa*; s. alb, *D. s. albostrigata*; s. bil, *D. s. bilimbata*; s. sul, *D. s. sulfurigaster*; pul, *D. pulaua*. All trees are rooted using *D. virilis* sequences.

(ii) *DNA polymorphism in the* nasuta *species group*

I obtained DNA sequence data for a total of five loci from the X chromosome and two loci from the mtDNA (Table 3), comprising a total of 4·6 kb sequenced in 40 lines of eight different taxa of the *nasuta* subgroup. Levels of diversity vary considerably among species and loci (Table 3). Estimates of levels of silent diversity of X-linked markers are highest in *D. kohkoa* and *D. kepulauna* (3·31% and 2·26%), followed by *D. albomicans, D. nasuta, D. s. sulfurigaster* and *D. s. albostrigata* (1·58%, 1·03%, 1·00% and 0·87%), while no silent variation was detected in *D. s. bilimbata*. These estimates of variation are in line

https://doi.org/10.1017/S0016672306008330 Published online by Cambridge University Press

Table 4. *Genetic differentiation between* D. albomicans *and other species of the* nasuta *group*

	nasuta	*kepulauna*	*kohkoa*	*s. albostrigata*	*s. bilimbata*	*s. sulfurigaster*
CG10944	1/0 (11)	2/0 (10)	1/0 (12)	0/0 (10)	0/1 (11)	1/0 (13)
CG12056	0/0 (14)	1/0 (16)	0/0 (17)	1/3 (17)	0/3 (15)	0/3 (16)
CG2930	0/1 (5)	0/4 (10)	0/4 (15)	0/9 (14)	0/12 (15)	0/12 (15)
CG3699	3/0 (26)	5/0 (20)	2/0 (23)	0/4 (22)	0/2 (20)	0/1 (19)
sesB	0/1 (2)	0/2 (2)	0/1 (1)	0/1 (1)	0/1 (1)	0/1 (2)
COII	2/0 (19)	1/3 (22)	1/0 (30)	1/0 (19)	1/5 (22)	2/2 (29)
ND5	4/0 (49)	5/2 (50)	6/4 (64)	5/2 (44)	1/15 (61)	6/8 (73)
X-linked	4/2 (58)	8/6 (58)	3/5 (68)	1/17 (64)	0/19 (62)	1/17 (65)
mtDNA	6/0 (68)	6/5 (72)	7/4 (94)	6/2 (63)	2/20 (83)	8/10 (102)

Shown are the number of shared/fixed (total) polymorphism between species pairs.

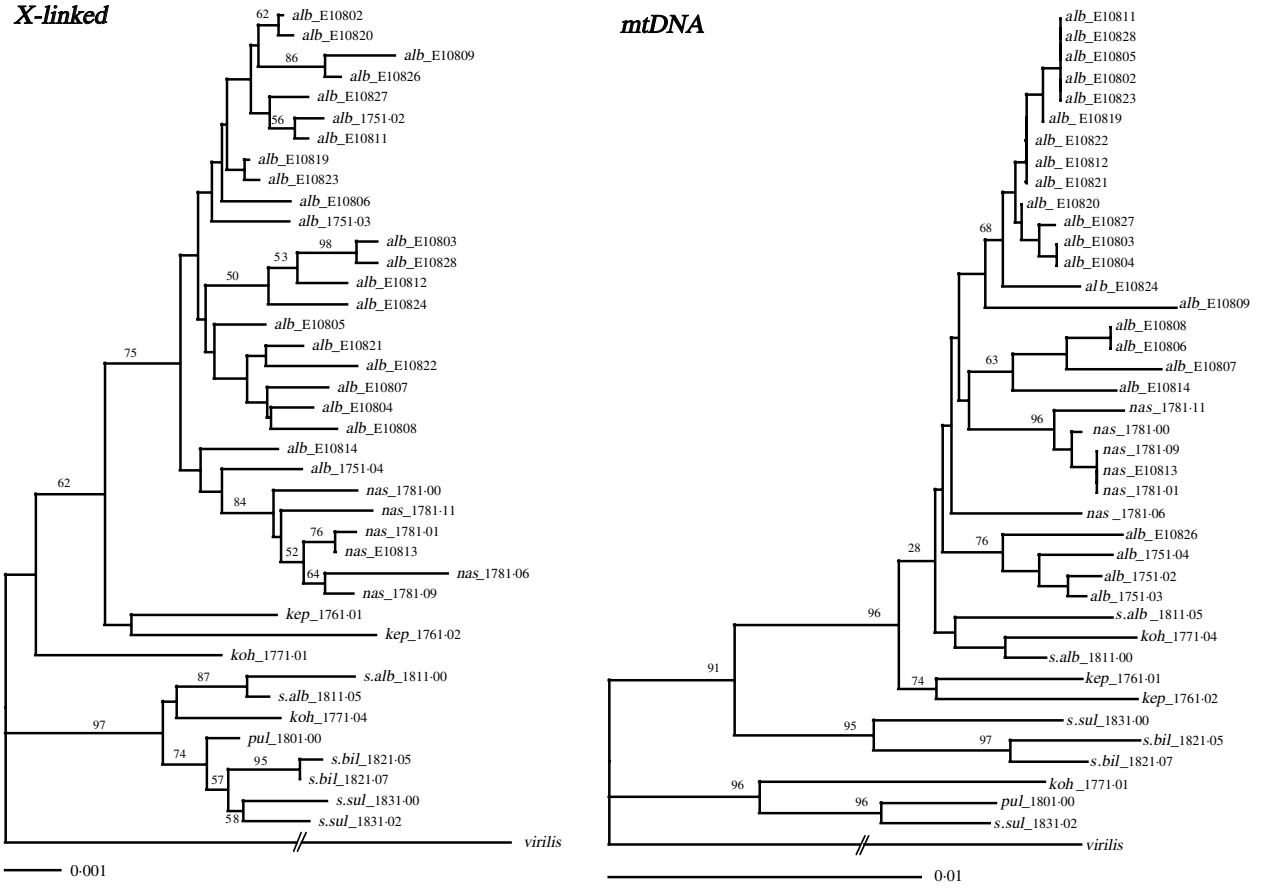

Fig. 3. Consensus trees constructed from the combined data set for X-linked loci and mtDNA. See Fig. 1 legend for details. Bootstrap values are shown for selected nodes.

with silent diversity estimated for several other members of the genus *Drosophila*, including the *melanogaster* species group (Andolfatto, 2005; Moriyama & Powell, 1996), the *yakuba* group (Bachtrog *et al*., 2006; Llopart *et al*., 2005), the *bipectinata* group (Kopp & Barmina, 2005) and the *pseudoobscura* group (Machado *et al*., 2002; Yi *et al*., 2003).

In most species investigated here, mtDNA shows higher levels of variability relative to nuclear genes (Table 3). For species with an equal number of males and females, the population size of mitochondria is one-third that of X chromosomes. However, mtDNA often has higher mutation rates than nuclear loci (Moriyama & Powell, 1997). Indeed, average pairwise divergence at silent sites between *D. albomicans* and *D. pulaua* is 3·4% for X-linked genes but 11·2% for the mtDNA (Table 5). Thus, the higher mutation rate of mtDNA, and possibly introgression of mtDNA across species boundaries, appears to increase diversity of mtDNA in the species investigated here.

https://doi.org/10.1017/S0016672306008330 Published online by Cambridge University Press

Table 5. *Polymorphism statistics for silent sites in* D. albomicans *and* D. nasuta

	D. albomicans			*D. nasuta*		
	D	*H*	K_s (%)	*D*	*H*	K_s (%)
CG10944	−0·64	0·67	1·72	1·03	0·53	1·86
CG12056	1·09	0·21	3·94	1·45	0·00	4·34
CG2930	−1·19	0·23	7·05	−1·13	0·53	7·19
CG3699	−0·24	1·12	2·43	−0·61	1·87	2·22
sesB	0	0	1·16	−0·93	0·27	0·19
COII	−1·56	−1·40	6·41	−0·93	0	7·01
ND5	−0·65	−9·52	15·39	−1·03	−9·87	16·12
X-linked	−0·24	0·56	3·40	−0·04	0·64	3·36
mtDNA	−0·96	−10·92	11·24	−1·04	−9·87	11·92

Silent divergence K_s (% per site) to *D. pulaua* is given.

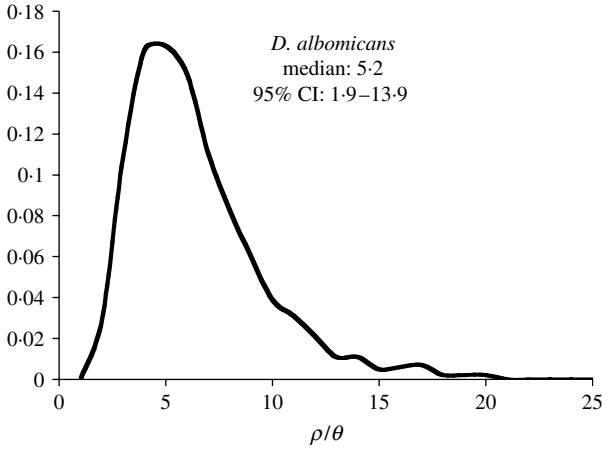

Fig. 4. Posterior distribution of ρ/θ (the population recombination rate scaled by the population mutation rate) for X-linked genes in *D. albomicans*, estimated using the approximate Bayesian method described by Haddrill *et al.* (2005). Priors were uniform and bounded by 0·0025 and 0·01 for θ and 0·009 and 0·18 for ρ. An arbitrary rejection criterion (ε) was set to 0·1. The distribution is based on 1000 acceptances. The median of the distribution is 5·2, and the 95% confidence interval is 1·9–13·9.

D. albomicans is of particular interest in this species group, because of its unusual karyotype. Thus, I included a large number of strains of *D. albomicans* (23 lines) and that of its close sister species *D. nasuta* (6 lines) in this analysis, to generate a first portrait of population-level nucleotide variability in the genome of these two species. Summary statistics, including locus-specific values of Tajima's *D* and Fay and Wu's *H*, and average silent divergence to *D. pulaua* are given in Table 5. To estimate rates of recombination in *D. albomicans*, I quantified levels of linkage disequilibrium scaled by the population mutation rate (ρ/θ) for X-linked markers using an approximate Bayesian method (Haddrill *et al.*, 2005). The posterior distribution for ρ/θ is shown in Fig. 4 and has a median value of 5·2 (with 95% confidence intervals ranging from 2·2 to 11·8). This value of ρ/θ is similar to that estimated from other *Drosophila* species (Bachtrog & Andolfatto, unpublished; Thornton & Andolfatto, 2006; Andolfatto & Przeworski, 2000) and was used in all simulations to account for recombination on the X chromosome.

Table 6. *Tests of the standard neutral model in* D. albomicans *and* D. nasuta *for X-linked genes, using* D. pulaua *as the outgroup*

	D	Var(*D*)	*H*	Var(*H*)
D. albomicans				
Observed	−0·24	0·94	0·56	0·19
SNM	−0·03	0·61	−0·03	1·99
	(0·28)	(0·20)	(0·81)	(0·95)
D. nasuta				
Observed	−0·04	1·42	0·64	0·52
SNM	−0·01	0·73	−0·03	1·17
	(0·48)	(0·08)	(0·91)	(0·60)

Multilocus means and variances are given for the observed and simulated data under the standard neutral model (see Section 2 for details). Probabilities of $X_{simulated} \leqslant X_{observed}$, where X is a given statistic, are given in parentheses.

To detect non-equilibrium demography or selection in *D. albomicans* and *D. nasuta*, I implemented multilocus statistical tests of the standard neutral model (SNM), following Haddrill *et al.* (2005). The mean and variance of Tajima's *D* and Fay and Wu's *H* do not statistically depart from the assumptions of the SNM in either *D. albomicans* or *D. nasuta* (Table 6). Tajima's *D* is slightly negative in both species, and Fay and Wu's *H* is slightly positive (Tables 5, 6). Note that only point estimates for ρ and θ were used in our simulations; this makes the conclusion that both *D. albomicans* and *D. nasuta* fit the SNM conservative.

(iii) *Divergence time of* D. nasuta *and* D. albomicans

As indicated by the phylogenetic analysis, *D. albomicans* and *D. nasuta* are a very closely related species pair. As incipient species begin to diverge from one another, they originally share genetic variation that was common to their ancestral species. Over time, shared genetic variability between species is lost, and fixed differences accumulate (Wakeley & Hey, 1997). Fig. 5 presents the polymorphic sites partitioned into variants that are shared between *D. albomicans* and *D. nasuta*, fixed between species, or exclusive to one species. Of the 58 variable sites from the X chromosome in this species comparison, only two are fixed between species, but four are shared (Fig. 5). Similar, of the 68 mtDNA variable sites, none are fixed

https://doi.org/10.1017/S0016672306008330 Published online by Cambridge University Press

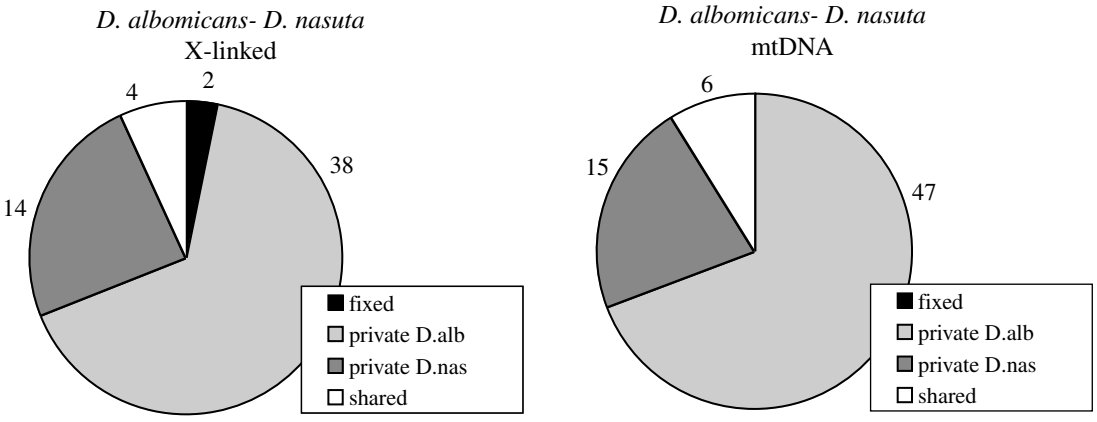

Fig. 5. Polymorphic sites shared between samples (white), fixed between samples (black), or exclusive to a sample (dark and light grey) for X-linked and mtDNA markers in *D. albomicans* and *D. nasuta*.

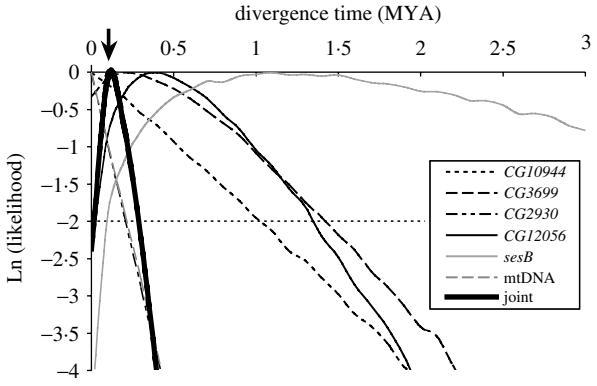

Fig. 6. Relative log-likelihoods of divergence time (in million years) for X-linked and mtDNA markers, between *D. albomicans* and *D. nasuta*. The dashed line indicates two log-likelihood units. The arrow indicates the joint maximum log-likelihood for all loci.

between species, and six are shared (Fig. 5). This pattern is indicative of a very recent time of splitting between this species pair. For species less closely related to *D. albomicans*, the number of shared polymorphisms decreases and the number of fixed differences increases (Table 4).

To infer divergence times between *D. albomicans* and *D. nasuta* at X-linked and mtDNA markers, I employed a summary-likelihood approach based on the HKA framework, following Bachtrog *et al.* (2006). The method estimates the most likely time of the species split (in years) based on two summaries of the data: the level of polymorphism within a species, and the average pairwise divergence between species. The most likely estimate of divergence time between *D. albomicans* and *D. nasuta* is about 120 000 years (with 95% confidence intervals ranging from 15 000 to 290 000 years; Fig. 6). Note that given uncertainties in the underlying mutation and recombination rates among loci, the true credibility intervals for the inferred divergence time are likely larger (see Section 2).

There is substantial variation in the most likely divergence time estimate for individual loci (Fig. 6). Heterogeneity in estimated speciation times across loci could result from either uncertainty in the estimates or some loci experiencing gene flow between species after the species split. A likelihood ratio test comparing the likelihood of six locus-specific speciation times (five X-linked loci and the pooled mtDNA loci) with the likelihood of the null model of one speciation time for all loci reveals no significant heterogeneity among loci ($-2\Delta(\ln Lik) = 9{\cdot}8$, d.f. $= 5$, $P = 0{\cdot}1$). Thus, the data are compatible with a simple isolation model of speciation, with little or no gene flow following the separation of *D. albomicans* and *D. nasuta*. Again, uncertainties in f, ρ and θ that are not accounted for in our simulations make this conclusion conservative.

4. Discussion

(i) *Radiation of the* nasuta *group*

The *Drosophila nasuta* species group is interesting from an evolutionary perspective, as multiple species formed within a short time period. In fact, many of the species pairs in this group are not yet reproductively isolated or show only partial reproductive isolation (Kitagawa, 1991; Wilson *et al.*, 1969). Here, I combine phylogenetic and population genetic approaches in order to reconstruct the evolutionary history of the *nasuta* species complex. Different loci in the *nasuta* complex do not appear to support the same genealogical history (Figs. 2, 3). This parallels patterns of differentiation seen in the *simulans*, the *pseudoobscura* and the *bipectinata* complex of *Drosophila* (Kliman *et al.*, 2000; Machado *et al.*, 2002; Kopp & Barmina, 2005), and could result from gene flow between species, or simply from lineage sorting and persistence of ancient polymorphism that pre-date speciation (Felsenstein,

https://doi.org/10.1017/S0016672306008330 Published online by Cambridge University Press

1988). Distinguishing between these models is challenging, and requires data from many loci (Kliman *et al*., 2000).

To obtain a very rough time estimate of the *nasuta* species group radiation, we can utilize levels of divergence among species. Average pairwise divergence between *D. pulaua* and *D. albomicans* is about 3·4% for X-linked loci. Assuming a substitution rate of synonymous sites of about 1·5% per million years in *Drosophila* (Li, 1997), this suggests that the *nasuta* species group radiated only about 1·1 million years ago. Interestingly, the patterns of differentiation among species are very similar for mtDNA and nuclear loci. The numbers of shared and fixed differences between *D. albomicans* and its relatives are similar for the mtDNA and nuclear DNA loci (Table 4). Also, the phylogenies for these two types of markers are similar, with most species showing a similar resolution at nuclear markers and mtDNA (Fig. 3). This is contrary to the observation of less mtDNA divergence relative to nuclear divergence in several other species complexes in the genus *Drosophila*, including the *simulans* clade (Ballard, 2000; Ballard *et al*., 2002), the *pseudoobscura* complex (Machado *et al*., 2002; Powell, 1983), the *yakuba* group (Bachtrog *et al*., 2006; Llopart *et al*., 2005) and the *bipectinata* complex (Kopp & Barmina, 2005).

Given the relatively short time span since this species group started to differentiate, it is perhaps not surprising that previous studies had difficulties in resolving the underlying species phylogeny, and often came to different conclusions (see Yu *et al*., 1999, for a discussion on different phylogenetic trees obtained by different studies). Morphological studies placed the taxa of the *D. nasuta* subgroup into three categories, based on the patterns of the male frons (Wilson *et al*., 1969). The first group includes *D. albomicans*, *D. nasuta*, *D. kepulauana* and *D. kohkoa*, which are characterized by a continuous silvery patch over the entire frons. The species in the second category, *D. pulaua* and the *D. sulfurigaster* subspecies, have pronounced whitish bands along the frontal orbits of the males. Species of the third category (none of which was investigated here) completely lack white frontal markings.

D. albomicans, *D. nasuta* and *D. kepulauana* were found to cluster together in the nuclear tree, as did *D. pulaua* and the *D. sulfurigaster* subspecies (Fig. 3); the position of *D. kohkoa* could not be resolved on the phylogenetic tree. Thus, the molecular phylogeny based on nuclear markers largely supports the classification based on morphological characters.

Studies on hybrid sterility (Kitagawa, 1991) and courtship song differences among species (Shao *et al*., 1997) found that *D. nasuta* and *D. albomicans* are sister species, and *D. kepulauana* forms the outgroup to this species pair. Likewise, *D. s. bilimbata* and *D. s. sulfurigaster* were identified as sister species, with *D. s. albostrigata* being more distantly related to this species pair. While studies based on hybrid sterility identified *D. s. albostrigata* as the sister taxa to *D. s. bilimbata* and *D. s. sulfurigaster* (Kitagawa, 1991; in accordance with our nuclear phylogeny), it was placed inside the *D.albomicans*, *D. nasuta* and *D. kepulauana* clade based on courtship song differences (Shao *et al*., 1997).

A previous study, based on mtDNA, found that a *D. albomicans* strain from Thailand formed (together with *D. nasuta* and *D. s. albostrigata*) a separate cluster from three *D. albomicans* strains from China (Yu *et al*., 1999). This grouping, however, is not supported here based on more data and more strains. In particular, while *D. albomicans* was found to be paraphyletic based on both the nuclear and mtDNA markers, *D. s. albostrigata* clearly groups outside the *D. albomicans*/*D. nasuta* clade. However, while the nuclear phylogeny suggests that *D. kepulauna* forms the outgroup to the *D. nasuta* and *D. albomicans* clade, the mtDNA tree places *D. s. albostrigata* as the closest outgroup to this species pair. This could indicate introgression of mtDNA across species boundaries, as observed in other *Drosophila* taxa (Bachtrog *et al*., 2006; Ballard, 2000; Powell, 1983).

A more detailed population genetics analysis of the *D. albomicans–D. nasuta* species pair suggests that there is no significant heterogeneity in estimated speciation times among loci. Thus, patterns of polymorphism and divergence for this species pair are compatible with a simple isolation model with no gene flow and an estimated species split time of only about 120 000 years ago. *D. albomicans* is endemic in South-East Asia from Japan, through Taiwan, Indochina and Malaysia to north-eastern India (Wilson *et al*., 1969). *D. nasuta* is known from the eastern coast of tropical Africa, through Madagascar and the Seychelles to Sri Lanka and most of the Indian subcontinent. Whether the two species come into contact in the north-east of India is not known. Mate-choice experiments between *D. albomicans* and *D. nasuta* found no sexual isolation between the two species (Kitagawa *et al*., 1982), but F_2 hybrid breakdown was found in laboratory studies (Chang *et al*., 1989). Thus, while there might be potential for current or ancestral gene flow between *D. albomicans* and *D. nasuta* in nature, we find no evidence for it based on DNA sequence data. This conclusion, however, is based on relatively few loci, and sampling from many loci may be necessary to find locus-specific differences in levels of gene flow (Kliman *et al*., 2000). If there was gene flow at some loci, the inferred divergence time would underestimate the true species split time (Osada & Wu, 2005). Therefore, the obtained estimate should probably be viewed as the minimum species split time between *D. albomicans* and

D. nasuta. Considering only loci with the deepest estimated divergence times (Fig. 6) suggests that this species pair is probably not older than 0·5 million years.

(ii) *Population genetics of* D. albomicans *and* D. nasuta.

Levels of silent variability in *D. albomicans* (1·6%) and *D. nasuta* (1·0%) are comparable to those of other *Drosophila* species, suggesting similar effective population sizes. The genetic model species *D. melanogaster* seems to harbour slightly more genetic variability (about 2·7% synonymous diversity for X-linked loci; Andolfatto, 2005), while *D. miranda* is slightly less variable (0·41% synonymous diversity at X and neo-X linked genes; Bachtrog & Andolfatto, unpublished). Thus, the effective population sizes of *D. albomicans* and *D. nasuta* are probably in the millions. Large population sizes for these species are consistent with their wide distribution through South-East Asia.

Multilocus patterns of nucleotide variability reveal that two summaries of the frequency spectrum of mutations (Tajima's *D* and Fay and Wu's *H*) are compatible with the SNM. This fit of the allele frequency spectrum to the SNM suggests that both species are close to a demographic equilibrium, and is probably not simply due to a lack of power to detect such departures in a small number of loci. Similar studies of other *Drosophila* species that used comparable numbers of loci and individuals often did detect departures of the allele frequency spectrum. For example, Bachtrog *et al*. (2006) found a significant skew in the frequency spectrum towards rare alleles in all three species of the *yakuba* group using only six X-linked loci. Likewise, Kopp & Barmina (2005) detected a significant excess of rare alleles, based on six nuclear loci in several species of the *bipectinata* species group. Similar departures from the neutral model towards an excess of rare alleles have also been detected in the *pseudoobscura* (Machado *et al*., 2002) or the *simulans* (Kliman *et al*., 2000) species complex, and have been interpreted as possible signatures of recent population expansion. Demographic departures from the SNM complicate inferences of selection using nucleotide variability data (i.e. Haddrill *et al*., 2005). The lack of an obvious departure from the SNM in *D. albomicans* greatly increases its utility for studying selection-driven models of sex chromosome evolution (Charlesworth & Charlesworth, 2000).

(iii) *Prospects for studying Y chromosome degeneration*

One of the species in the *nasuta* group, *D. albomicans*, is of particular interest due to a chromosomal fusion between an autosome and the sex chromosomes, creating a so-called neo-sex chromosome pair. As a result of such a fusion, one member of the pair of autosomes (the neo-Y) is inherited from father to son in association with the true Y chromosome. Its partner (the neo-X) is fused to the true X. In *Drosophila*, the absence of crossing over in males ensures that such a neo-Y chromosome is genetically isolated from its partner and sheltered from recombination (Lucchesi, 1978). Evolutionary theory predicts that a lack of recombination triggers the loss of gene function on such a neo-Y (Charlesworth & Charlesworth, 2000). Consistent with this prediction, neo-Y chromosome systems in the genus *Drosophila* have been found to be degenerating, and have provided important model systems to study the process of Y-chromosome degeneration (Bachtrog, 2005; Carvalho & Clark, 2005).

Of particular interest has been the neo-Y chromosome of *D. miranda*, which was formed about 1 million years ago (Bachtrog & Charlesworth, 2002). About one-third of all genes on this chromosome already harbour null alleles (Bachtrog, 2005), and large parts of its former homologue, the neo-X, are already dosage compensated in males (Bone & Kuroda, 1996; Marìn *et al*., 1996). Even older neo-Y chromosomes, such as the neo-Y of *D. pseudoobscura* or *D. willistoni*, are almost entirely degenerated (Carvalho & Clark, 2005), and the neo-X is completely dosage compensated (Abraham & Lucchesi, 1974; Bone & Kuroda, 1996; Marìn *et al*., 1996). It is of great interest to examine a very young neo-Y chromosome, where there has as yet been little opportunity for degeneration. A lack of degeneration implies that purifying selection against deleterious mutations is not relaxed on such a neo-Y chromosomes. In addition, studying a more recently formed neo-Y chromosome will also allow us to put a chronological order on the processes of Y chromosome degeneration, and to calibrate the speed from the transition of an ordinary autosome to a degenerate Y. Moreover, different processes of Y chromosome degeneration may be acting at different stages in the evolution of a Y chromosome, and it is necessary to study systems of varying ages to obtain a full picture of the process of sex chromosome evolution.

One such young neo-Y system to study the initial stages of Y chromosome degeneration might be that of *D. albomicans*. This neo-Y chromosome must be younger than the time of speciation between *D. albomicans* and *D. nasuta* (which lacks this fusion) and might be $<100\,000$ years old. The small numbers of fixed differences observed at X-linked loci between *D. nasuta* and *D. albomicans* may mean that large genomic regions from the neo-sex chromosomes need to be studied to detect differences in rates of sequence

https://doi.org/10.1017/S0016672306008330 Published online by Cambridge University Press

evolution between the neo-X and neo-Y chromosomes. However, only X-linked and mtDNA markers were surveyed here, and it is possible that the neo-sex chromosome system is older than the inferred species divergence if gene flow at X-linked markers and mtDNA continued for some time after the formation of the neo-sex chromosomes. Thus, molecular data from the neo-sex chromosomes are needed to further assess the utility of this system for studying sex chromosome evolution. The young age of *D. albomicans*, its large population size and lack of drastic departures from equilibrium assumptions suggest that its neo-sex chromosome system might provide an ideal model to study the very early stages of Y chromosome degeneration.

I am grateful to P. Andolfatto for comments on the manuscript and software, to K. Thornton for software, and to M. Watada and the Tucson *Drosophila* Species Stock Center for fly strains. D.B. is supported by a Postdoctoral Fellowship from the Austrian Academy of Sciences.

References

Abraham, I. & Lucchesi, J. (1974). Dosage compensation of genes on left and right arms of X-chromosome of *Drosophila pseudoobscura* and *Drosophila willistoni*. *Genetics* **78**, 1119–11226.

Andolfatto, P. (2005). Adaptive evolution of non-coding DNA in *Drosophila*. *Nature* **437**, 1149–1152.

Andolfatto, P. & Przeworski, M. (2000). A genome-wide departure from the standard neutral model in natural populations of *Drosophila*. *Genetics* **156**, 257–268.

Andolfatto, P. & Wall, J. (2003). Linkage disequilibrium patterns across a recombination gradient in African *Drosophila melanogaster*. *Genetics* **165**, 1289–1305.

Bachtrog, D. (2005). Sex chromosome evolution: molecular aspects of Y chromosome degeneration in *Drosophila*. *Genome Research* **15**, 1393–1401.

Bachtrog, D. & Charlesworth, B. (2002). Reduced adaptation of a non-recombining neo-Y chromosome. *Nature* **416**, 323–326.

Bachtrog, D., Thornton, K., Clark, A. & Andolfatto, P. (2006). Extensive introgression of mitochondrial DNA relative to nuclear gene flow in the *Drosophila yakuba* species group. *Evolution* **60**, 292–302.

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

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

Bone, J. R. & Kuroda, M. I. (1996). Dosage compensation regulatory proteins and the evolution of sex chromosomes in *Drosophila*. *Genetics* **144**, 705–713.

Carvalho, A. & Clark, A. (2005). Y chromosome of *D. pseudoobscura* is not homologous to the ancestral *Drosophila* Y. *Science* **307**, 108–110.

Chang, H. Y., Wang, D. G. & Ayala, F. J. (1989). Mitochondrial DNA evolution in the *Drosophila nasuta* subgroup of species. *Journal of Molecular Evolution* **28**, 337–348.

Charlesworth, B. & Charlesworth, D. (2000). The degeneration of Y chromosomes. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **355**, 1563–1572.

Fay, J. C. & Wu, C. I. (2000). Hitchhiking under positive Darwinian selection. *Genetics* **155**, 1405–1413.

Felsenstein, J. (1988). Phylogenies from molecular sequences: inference and reliability. *Annual Review of Genetics* **22**, 521–565.

Haddrill, P., Thornton, K., Charlesworth, B. & Andolfatto, P. (2005). Multilocus patterns of nucleotide variability and the demographic and selection history of *Drosophila melanogaster* populations. *Genome Research* **15**, 790–799.

Hatsumi, M., Morishige, Y. & Wakahama, K. (1988). Metaphase chromosomes of four species of the *Drosophila nasuta* subgroup. *Japanese Journal of Genetics* **63**, 435–444.

Hudson, R. R. (1990). Gene genealogies and the coalescent process. In *Oxford Surveys in Evolutionary Biology* (ed. J. Antonovics & D. Futuyma), pp. 1–44. Oxford: Oxford University Press.

Hudson, R. (2002). Generating samples under a Wright–Fisher neutral model of genetic variation. *Bioinformatics* **18**, 337–338.

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

Kitagawa, O. (1991). *Evolutions of Populations*. Tokyo: Tokyo University Press.

Kitagawa, O., Wakahama, K. I., Fuyama, Y., Shimada, Y., Takanashi, E., *et al*. (1982). Genetic study of *Drosophila nasuta* subgroup, with notes on distribution and morphology. *Japanese Journal of Genetics* **57**, 113–141.

Kliman, R., Andolfatto, P., Coyne, J., Depaulis, F., Kreitman, M., *et al*. (2000). The population genetics of the origin and divergence of the *Drosophila simulans* species complex. *Genetics* **156**, 1913–1931.

Kopp, A. & Barmina, O. (2005). Evolutionary history of the *Drosophila bipectinata* species complex. *Genetical Research* **85**, 23–46.

Kumar, S. Tamura, K. & Nei, M. (2004). MEGA3: integrated software for molecular evolutionary Genetics Analysis and sequence alignment. *Briefings in Bioinformatics* **5**, 150–163.

Li, W. (1997). *Molecular Evolution*. Sunderland, MA: Sinauer Associates.

Llopart, A., Lachaise, D. & Coyne, J. (2005). Multilocus analysis of introgression between two sympatric sister species of *Drosophila*: *Drosophila yakuba* and *D. santomea*. *Genetics* **171**, 197–210.

Lucchesi, J. C. (1978). Gene dosage compensation and the evolution of sex chromosomes. *Science* **202**, 711–716.

Machado, C., Kliman, R., Markert, J. & 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.

Marin, I., Franke, A., Bashaw, G. & Baker, B. (1996). The dosage compensation system of *Drosophila* is co-opted by newly evolved X chromosomes. *Nature* **383**, 160–163.

Moriyama, E. N. & Powell, J. R. (1996). Intraspecific nuclear DNA variation in *Drosophila*. *Molecular Biology and Evolution* **13**, 261–277.

Moriyama, E. N. & Powell, J. R. (1997). Synonymous substitution rates in *Drosophila*: mitochondrial versus nuclear genes. *Journal of Molecular Evolution* **45**, 378–391.

Nagaraja, J. & Ranganath, H. (2004). Molecular phylogeny of the *nasuta* subgroup of *Drosophila* based on 12S rRNA, 16S rRNA and CoI mitochondrial genes, RAPD and ISSR polymorphisms. *Genes and Genetic Systems* **79**, 293–299.

Osada, N. & Wu, C. (2005). Inferring the mode of speciation from genomic data: a study of the great apes. *Genetics* **169**, 259–264.

Patterson, J. T. & Stone, W. S. (1952). *Evolution in the Genus Drosophila*. New York: Macmillan.

Powell, J. (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.

Przeworski, M., Wall, J. D. & Andolfatto, P. (2001). Recombination and the frequency spectrum in *Drosophila melanogaster* and *Drosophila simulans*. *Molecular Biology and Evolution* **18**, 291–298.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Schierup, M. & Hein, J. (2000). Consequences of recombination on traditional phylogenetic analysis. *Genetics* **156**, 879–891.

Shao, H.-G., Li, D., Zhang, X.-N., Yu, H.-J., Li, X., *et al.* (1997). Study on the recognition and evolutionary genetics of the courship song of species in *Drosophila nasuta* species group. *Chinese Journal of Genetics* **24**, 311–321.

Steinemann, M. & Steinemann, S. (1998). Enigma of Y chromosome degeneration: neo-Y and neo-X chromosomes of *Drosophila miranda* – a model for sex chromosome evolution. *Genetica* **102–103**, 409–420.

Sullivan, W., Ashburner, M. & Hawley, R. (2000). *Drosophila Protocols*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Suzuki, Y., Kitagawa, O. & Wakahama, K. (1990). Chromosomal analysis and phylogenetic relationships in the *Drosophila nasuta* subgroup. I. Phylogenetic relationships within the *Drosophila sulfurigaster* species complex. *Genetica* **80**, 53–66.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595.

Tamura, K. & Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution* **10**, 512–526.

Thornton, K. & Andolfatto, P. (2006). Approximate bayesian inference reveals evidence for a recent, severe, bottleneck in a Netherlands population of *Drosophila melanogaster*. *Genetics* **172**, 1607–1619.

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

Wilson, F. D., Wheeler, M. R., Harger, M. & Kambysellis, M. (1969). Cytogenetic relations in the *Drosophila nasuta* subgroup of the *immigrans* group of species. *University of Texas Publications* **6918**, 207–254.

Yi, S., Bachtrog, D. & Charlesworth, B. (2003) A survey of chromosomal and nucleotide sequence variation in *Drosophila miranda*. *Genetics* **164**, 1369–1381.

Yu, H., Wang, W., Fang, S., Zhang, Y. P., Lin, F. J., *et al.* (1999). Phylogeny and evolution of the *Drosophila nasuta* subgroup based on mitochondrial ND4 and ND4L gene sequences. *Molecular Phylogenetics and Evolution* **13**, 556–565.