Available online at www.sciencedirect.com
SCIENCE DIRECT
Molecular Phylogenetics and Evolution 38 (2006) 731–741
MOLECULAR PHYLOGENETICS AND EVOLUTION
www.elsevier.com/locate/ympev

Interspecific divergence, intrachromosomal recombination, and phylogenetic utility of Y-chromosomal genes in *Drosophila*

Artyom Kopp *, Amanda K. Frank, Olga Barmina

Section of Evolution and Ecology and Center for Genetics and Development University of California—Davis, Davis CA 95616, USA

Received 25 May 2005; revised 14 October 2005; accepted 15 October 2005
Available online 1 December 2005

Abstract

Reconstruction of phylogenetic relationships among recently diverged species is complicated by three general problems: segregation of polymorphisms that pre-date species divergence, gene flow during and after speciation, and intra-locus recombination. In light of these difficulties, the Y chromosome offers several important advantages over other genomic regions as a source of phylogenetic information. These advantages include the absence of recombination, rapid coalescence, and reduced opportunity for interspecific introgression due to hybrid male sterility. In this report, we test the phylogenetic utility of Y-chromosomal sequences in two groups of closely related and partially inter-fertile *Drosophila* species. In the *D. bipectinata* species complex, Y-chromosomal loci unambiguously recover the phylogeny most consistent with previous multi-locus analysis and with reproductive relationships, and show no evidence of either post-speciation gene flow or persisting ancestral polymorphisms. In the *D. simulans* species complex, the situation is complicated by the duplication of at least one Y-linked gene region, followed by intrachromosomal recombination between the duplicate genes that scrambles their genealogy. We suggest that Y-chromosomal sequences are a useful tool for resolving phylogenetic relationships among recently diverged species, especially in male-heterogametic organisms that conform to Haldane's rule. However, duplication of Y-linked genes may not be uncommon, and special care should be taken to distinguish between orthologous and paralogous sequences.

Keywords: Speciation; Phylogeny; Y chromosome; Gene conversion; Concerted evolution; *Drosophila simulans*; *Drosophila bipectinata*

1. Introduction

Reconstruction of phylogenetic relationships among recently diverged species is complicated by three biological problems: segregation of polymorphisms that pre-date species divergence, gene flow during and after speciation, and intralocus recombination. Under neutral coalescence in complete isolation, sequence variation that was present in the common ancestor of two species is lost in each of the descendant species at a rate that *on average* is proportional to its effective population size (Hudson, 1990). However, coalescence is a stochastic process, and there is wide variation in coalescence times of individual loci around the expected value: some loci may coalesce much faster, and some much slower, than average (Hudson and Turelli, 2003). When the time since speciation is short relative to the effective population size, gene trees reconstructed from individual loci are unlikely to be an accurate reflection of species phylogeny (Nichols, 2001). In fact, different loci are expected to have different genealogies, and there is no single dichotomous tree that applies to the entire genome. Reconstruction of species phylogeny is further complicated by the possibility of post-speciation gene flow. Reproductive isolation between nascent species is not achieved instantaneously, and the species may continue to hybridize at low frequency for a considerable time. As a result of this hybridization, genomic regions that are not linked to genes responsible for reproductive isolation or species-specific ecological adaptations may be introgressed across species boundaries, producing genealogies that are in conflict with

* Corresponding author. Fax: +1 530 752 9014.
E-mail address: akopp@ucdavis.edu (A. Kopp).

1055-7903/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2005.10.006

the population history (Ting et al., 2000; Wu, 2001). Finally, at short evolutionary distances, the effects of intralocus recombination cannot be ignored. Phylogenetic reconstruction assumes that all nucleotides within a locus share a unique, strictly bifurcating history. This assumption is violated in many nuclear genes sampled from closely related species, so that the "tree" reconstructed from each locus is not a real gene tree, but rather a superposition of multiple genealogies.

In light of these difficulties, Y-chromosomal sequences offer an excellent source of information for resolving phylogenetic relationships among recently diverged species. First, there should be no recombination on the Y chromosome. Thus, Y-chromosomal phylogenies can be regarded as true gene trees, and the data from multiple Y-linked loci can be pooled for combined analyses. Second, neutral coalescence times for the Y chromosome are on average four times shorter than for autosomal genes, and three times shorter than for X-linked loci ($0.5Ne$ vs. $2Ne$ and $1.5Ne$, respectively). Finally, Y-chromosomal sequences should be less susceptible to post-speciation gene flow than most other markers. Due to Haldane's rule, hybrid male sterility is often the first isolating mechanism to evolve during species divergence in male-heterogametic organisms, and may precede the evolution of hybrid female sterility or inviability by a long interval (Coyne, 1994; Haldane, 1922; Orr and Turelli, 1996; Turelli, 1998; Turelli and Orr, 2000). Therefore, gene flow across nascent species boundaries should cease much earlier for the paternally inherited Y chromosome than for the bi-parentally inherited autosomal and X-linked loci, or for the maternally inherited mitochondrial DNA (mtDNA). In this respect, the Y chromosome should be similar to genes that are directly responsible for the evolution of reproductive isolation (Ting et al., 2000).

The contrast between mtDNA and the Y chromosome is particularly striking. Although the two are similar in the short coalescence time ($0.5Ne$) and the absence of recombination, Haldane's rule predicts that mtDNA should be far more susceptible to interspecific introgression than nuclear markers in male-heterogametic organisms. Indeed, there is mounting evidence that exchange of mtDNA across species is common in a variety of taxa (Ballard, 2000b; Boissinot and Boursot, 1997; Crochet et al., 2003; Ferris et al., 1983; Kopp and Barmina, 2005; Machado et al., 2002; Martin et al., 2002; Mukai et al., 1997; Ruedi et al., 1997; Shaw, 2002; Sota, 2002). This makes mtDNA unsuitable for reconstructing phylogenetic relationships among closely related species.

The Y chromosome has received some attention as a phylogeographic marker (Boissinot and Boursot, 1997; Karn et al., 2002; Jobling and Tyler-Smith, 2003; Underhill et al., 2001; Underhill et al., 2000) and as a source of information about animal mating systems (Stone et al., 2002; Tosi et al., 2000; Verkaar et al., 2004). However, it has not been used explicitly for the purposes of phylogenetic reconstruction. The combination of rapid coalescence, early cessation of gene flow, and the absence of recombination suggest that the genealogy of Y-chromosomal sequences should reflect species phylogeny more accurately than randomly chosen nuclear markers or mtDNA. In this report, we test this hypothesis in two groups of recently diverged species—the *Drosophila simulans* and *D. bipectinata* species complexes. The three members of the *simulans* complex—*D. simulans*, *D. mauritiana*, and *D. sechellia*—are thought to have separated 263,000–413,000 years ago (Kliman et al., 2000). All three species can be hybridized in the lab, and produce fertile female and sterile male hybrids. They have been used extensively to study the genetic basis of reproductive isolation (Coyne, 1996; Fang et al., 2002; Greenberg et al., 2003; Hollocher and Wu, 1996; Macdonald and Goldstein, 1999; Takahashi et al., 2001; Tao et al., 2003a,b; Ting et al., 1998; True et al., 1996) and the population genetics of the origin and divergence of species (Ballard, 2000a; Kliman et al., 2000; Noor and Kliman, 2003). However, phylogenetic relationships among these species are not entirely clear, and all three possible species-level phylogenies have been proposed at some point (Caccone et al., 1996; Cariou, 1987; Harr et al., 1998; Hey and Kliman, 1993; Kliman et al., 2000; Ting et al., 2000). All three confounding factors (lineage sorting, gene flow, and intralocus recombination) combine to obscure interspecific relationships when autosomal and X-chromosomal sequences are considered (Kliman et al., 2000), while mtDNA shows strong evidence of interspecific introgression (Ballard, 2000a; Ballard, 2000b; Solignac, 2004). As a result, sequences isolated from different species are often related more closely than conspecific alleles, providing little support either for the monophyly of individual species, or for any two-species clades (Eanes et al., 1996; Hey and Kliman, 1993; Kliman et al., 2000; Noor and Kliman, 2003).

The *D. bipectinata* species complex is a group of four predominantly sympatric species: *D. bipectinata*, *D. parabipectinata*, *D. malerkotliana*, and *D. pseudoananassae* (Bock, 1971; Bock, 1978; Kopp and Barmina, 2005; Singh and Singh, 2001). Two allopatric subspecies, defined on the basis of male-specific pigmentation, are recognized within each of the latter two species: *D. m. malerkotliana* and *D. m. pallens*, and *D. p. pseudoananassae* and *D. p. nigrens*, respectively (Bock, 1971). Based on the patterns of reproductive isolation, *D. bipectinata* also appears to consist of at least two subspecies: *D. b. bipectinata*, which occurs on the Indian subcontinent and in Southeast and Northeast Asia, and *D. b. pacificae*, which is found in the South Pacific region from New Caledonia to American Samoa (Kopp and Frank, 2005; Matsuda et al., 2005). A zone of secondary hybridization between these subspecies may exist in New Guinea and Australia (Kopp and Frank, 2005). *D. bipectinata*, *D. parabipectinata*, and *D. malerkotliana* appear to have diverged even more recently than the *simulans* species complex, and the genetic differentiation among these species at autosomal and X-linked loci is extremely low (Kopp and Barmina, 2005). Species of the *bipectinata* complex hybridize readily in the laboratory, producing fertile female and sterile male hybrids. Not surprisingly,

phylogenetic relationships in the *bipectinata* complex have proven difficult to resolve. A combined analysis of six nuclear genes supports a sister-group relationship between *D. bipectinata* and *D. parabipectinata*. However, in the analysis of individual loci, sequences sampled from the same species rarely cluster together, and there is little or no support for any supra-specific clades (Kopp and Barmina, 2005). At the same time, mtDNA shows clear evidence of introgression across species boundaries, so that many mitochondrial haplotypes are shared among *D. bipectinata*, *D. parabipectinata*, and *D. malerkotliana*. The subspecies of *D. malerkotliana* and *D. bipectinata* show no evidence of genetic differentiation at nuclear loci, although the Asian and Pacific subspecies of *D. bipectinata* are well differentiated at the level of mtDNA, presumably reflecting prolonged geographic isolation (Kopp and Barmina, 2005; Kopp and Frank, 2005).

We have analyzed interspecific divergence of several Y-chromosomal loci in the *D. bipectinata* and *D. simulans* species complexes. In the *bipectinata* complex, the resulting phylogeny strongly supports the monophyly of each species, as well as a sister-group relationship between *D. bipectinata* and *D. parabipectinata*. This result is consistent with the phylogeny based on a much larger multi-locus data set (Kopp and Barmina, 2005), suggesting that Y-chromosomal sequences do in fact reflect species phylogeny. In the *simulans* complex, however, the situation is complicated by the duplication of at least one gene region. Intrachromosomal conversion between duplicate genes introduces many "recombination" events into their sequences, so that different sites have different genealogies, and the phylogeny of the gene as a whole is impossible to determine. We suggest that Y-chromosomal sequences are a useful tool for resolving phylogenetic relationships among recently diverged, partially interfertile species, especially in male-heterogametic organisms that conform to Haldane's rule. However, duplication of Y-linked genes may not be uncommon, and special care should be taken to distinguish between orthologous and paralogous sequences.

2. Materials and methods

2.1. Gene fragments

Three fragments of the Y chromosome were sequenced for the *simulans* species complex:

kl2-Left. A 400 base pair (bp) fragment of the *kl-2* gene (Carvalho et al., 2000) corresponding to positions 4433–4833 of the partial coding sequence of *kl-2* in *D. melanogaster* (Accession No. AF313479).

kl2-Right. A 777 bp fragment of *kl-2* corresponding to bases 10,049–10,697 of AF313479. This fragment includes two small introns: a 62 bp intron at position 10416 of AF313479 (positions 379–440 of the multi-species alignment) and a 51 bp intron at position 10522 of AF313479 (positions 549–599 of the alignment). The multi-species alignment is longer than the *D. melanogaster* sequence due to several insertions in *D. sechellia* and *D. mauritiana*.

ory. A 672 bp fragment of the *ory* gene (Carvalho et al., 2001) corresponding to positions 1009-1389 of the partial coding sequence of *ory* in *D. melanogaster* (Accession No. AF427496). This fragment includes a 292 bp intron at position 1093 of AF427496 (positions 84–375 of the alignment).

The three gene regions, including the two parts of *kl-2*, are separated by large stretches of heterochromatin whose exact sizes are unknown (Carvalho et al., 2000; Carvalho et al., 2001; Hoskins et al., 2002).

Two fragments were sequenced for the *bipectinata* complex:

kl2-Left. A 553 bp fragment of *kl-2* corresponding to positions 4440–4992 of the partial coding sequence of *kl-2* of *D. melanogaster* (AF313479).

kl2-Right. A 502 bp fragment of *kl-2* corresponding to positions 9554–10056 of AF313479.

Amplification primers and GenBank accession numbers for representative sequences of each species are listed in Table 1. Complete sequence alignments in Nexus format are available as supplementary information on the authors' website.

2.2. Drosophila strains

Seven strains of *D. simulans*, four strains of *D. mauritiana*, four of *D. sechellia*, and two of *D. melanogaster* were sequenced for the *simulans* species complex. Six strains of *D. bipectinata*, four strains of *D. parabipectinata*, five of *D. malerkotliana*, and four of *D. pseudoananassae* were used

Table 1
Amplification primers and sequence accession numbers

	Left primer	Right primer	Accession Nos.
simulans complex			
kl2-Left	CTGTAAGGATACGGGCCATTT	AACATAAGCTCCAACGCCTTT	DQ239728–731
kl2-Right	AAGAATACAAACCAGCATCGGAACG	GAACTGGCGGATCAACATATCGAG	DQ239721–727
ory	TCCGTGCACTCTGAGAAAATGA	TCGTACCATTTGCAATCCGAC	DQ239717–720
bipectinata complex			
kl2-Left	AATACTTTTCGGGCCAAATTGC	GCATCGAGAGTCGGAACTTTGT	DQ239703–709
kl2-Right	CACAAGGAAGTCGTTGGCCTTT	TAAAATTGCTGCGCGTTCAGAG	DQ239710–716

in the analysis of the *bipectinata* complex, with *D. ananassae* sequences used as outgroups. An additional 53 strains of *D. malerkotliana*, *D. bipectinata*, and *D. parabipectinata* were sequenced only for the *kl2-Right* locus to assess the amounts of intraspecific variation. The strains were chosen to represent the entire geographic range of each species (except for the island endemics *D. sechellia* and *D. mauritiana*). The names, collection locations, and Y-chromosomal haplotypes of all strains are listed in Table 2 and in Supplement Table 1.

2.3. DNA sequencing and analysis

Genomic DNA was isolated from a single male of each strain. Amplified gene fragments were sequenced directly using the forward and reverse amplification primers. ABI trace chromatograms were aligned and examined using Contig Express software (Invitrogen). We found that the *kl2-Right* locus in the *simulans* species complex contained numerous polymorphic nucleotide substitutions and indels, suggesting that the locus was duplicated. We therefore cloned the *kl2-Right* fragments from *D. simulans*, *D. mauritiana*, and *D. sechellia*, and sequenced 7–10 independently isolated clones for each species and strain. The sequences of cloned and directly sequenced fragments from each species were aligned and compared to identify and separate duplicate genes. In each species, all clones fell into two clearly distinct clusters, presumably corresponding to two duplicate loci. Most clones within each cluster were identical, with the exception of two sites in *D. simulans* that had different nucleotides in different strains, but were always monomorphic in each strain. We also noticed a small number of singleton polymorphisms within strains, which we assumed to be caused by PCR errors. We also observed two recombinant clones that combined regions from different sequence clusters. Given the low frequency of such clones (2 out of more than 50), we assumed them to reflect in vitro (PCR-mediated) recombination events, and eliminated them from analysis.

Table 2
Drosophila strains and collection locations

Species/subspecies	Strain	Collection location	Date	Collector or donor
simulans complex				
simulans	C167	Nanyuki, Kenya	?	A. V. Olembo
simulans	C22	Congo	?	W. Ballard
simulans	TT101TS	Papeete, Tahiti	?	K. Kaneshiro
simulans	Wolfskill 2	Winters, CA, US	1995	D. Begun
simulans	Florida	Florida, US	?	TDSSC[a]
simulans	SZ7	Harare, Zimbabwe	1994	T. Mutangadura
simulans	NC48	Noumea, New Caledonia	1999	M. Solignac
mauritiana	6	Mauritius	?	C. Jones[b]
mauritiana	72	Mauritius	?	C. Jones[b]
mauritiana	197	Mauritius	?	C. Jones[b]
mauritiana	207	Mauritius	?	C. Jones[b]
sechellia	4	Seychelles	?	C. Jones[b]
sechellia	15	Seychelles	?	C. Jones[b]
sechellia	77	Praslin Island, Seychelles	1987	C. Jones[b]
sechellia	81	Cousin Island, Seychelles	1985	C. Jones[b]
melanogaster	Wi89	Winters, CA, US	1999	S. Nuzhdin
melanogaster	Mw5	Malawi	?	W. Ballard
bipectinata complex				
b. bipectinata	TH01	Hsinchu, Taiwan	2004	C.-S. Ng
b. bipectinata	TH02	Hsinchu, Taiwan	2004	C.-S. Ng
b. bipectinata	KB382	Kuala Belalong, Temburong, Brunei	2003	A. Kopp & O. Barmina
b. bipectinata	KB446	Kuala Belalong, Temburong, Brunei	2003	A. Kopp & O. Barmina
b. pacificae	TBU83	Tongatapu, Tonga	1981	M. Watada
b. pacificae	PPG96	Pago-Pago, Amer. Samoa	1981	M. Watada
parabipectinata	401.2	Tagaytay, Philippines	1968	L. Throckmorton
parabipectinata	401.0	Ari Ksatr, Cambodia	1967	M. Delfinado
parabipectinata	Yun6	Yunnan, China	2002	S.-C. Tsaur
parabipectinata	BSB1	Bandar Seri Begawan, Brunei	2003	A. Kopp & O. Barmina
m. malerkotliana	391.0	Mysore, Karnataka, India	1971	N. B. Krishnamurthy
m. malerkotliana	391.1	Samut Songkhram, Thailand	1971	V. Baimai
m. malerkotliana	YA31	Yaounde, Cameroon	2004	J. Pool
m. pallens	KB130	Kuala Belalong, Temburong, Brunei	2003	A. Kopp & O. Barmina
m. pallens	R242	Puerto Princesa, Palawan, Philippines	1979	Y. Fuyama
p. pseudoananassae	Lae344	Lae, Papua New Guinea	1981	Y. Fuyama
p. pseudoananassae	Wau73	Wau, Papua New Guinea	1981	Y. Fuyama
p. nigrens	411.0	Samut Songkhram, Thailand	1971	V. Baimai
p. nigrens	421.0	Chiang Mai, Thailand	1971	TDSSC[a]

[a] Obtained from Tucson Drosophila Species Stock Center, collector unknown.
[b] Obtained from C. Jones, collector unknown.

Multi-species sequence alignments were examined by eye in MacClade (Maddison and Maddison, 2001) to identify polymorphic sites and record their allelic states in each species and strain. For the *simulans* species complex, phylogenetic reconstruction was performed in PAUP (Swofford, 2000) using maximum parsimony with random order of sequence addition and TBR branch swapping. The haplotype network for the *bipectinata* complex was reconstructed manually. The four-gamete test (Hudson and Kaplan, 1985) was performed in SITES (Hey and Wakeley, 1997), and the results were confirmed by eye.

3. Results

3.1. Y-chromosomal sequences resolve species relationships in the bipectinata complex

We obtained a total of 1055 bp of sequence from two Y-chromosomal gene regions, *kl2-Left* and *kl2-Right*, from 19 strains belonging to all four species and seven subspecies of the *bipectinata* complex. This data set contained 57 variable sites, including 53 silent sites and 4 changes resulting in amino acid substitutions (Fig. 1). Forty-five variable sites were phylogenetically informative, that is, had rare character states that were shared by at least two taxa. No recombination was detected either within or between the two gene regions. All informative sites were consistent with a single haplotype network, which is shown in Fig. 2. We used sequences from *D. ananassae*, a relative of the *bipectinata* complex, to root this network and determine the direction of character changes (Figs. 1 and 2). A total of seven haplotypes were observed. Six of these each corresponded exactly to a single species or subspecies: *D. b. bipectinata*, *D. b. pacificae*, *D. parabipectinata*, *D. p. pseudoananassae*, and *D. p. nigrens* (Fig. 2). Two distinct haplotypes, separated by three nucleotide substitutions, were found in *D. malerkotliana*. These haplotypes did not correspond to the two allopatric subspecies defined on the basis of male pigmentation (Bock, 1971; Kopp and Barmina, 2005). Instead, haplotype

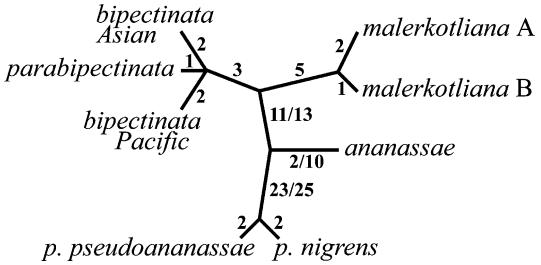

Fig. 2. Y-chromosomal haplotype network for the *bipectinata* species complex. See Fig. 1 for haplotype definitions. The numbers next to each branch indicate the number of nucleotide substitutions. The exact allocation of substitutions along the branches linking *D. pseudoananassae* with the other three species, or the complex as a whole with *D. ananassae*, could not be determined due to multiple substitutions at some sites. The minimum and maximum number of substitutions is given instead.

B was found in *D. m. pallens* and in the strain of *D. m. malerkotliana* from Thailand, while haplotype A was found in *D. m. malerkotliana* stains from India and West Africa (Figs. 1, 2, and Supplementary data).

The genealogy of Y-chromosomal sequences (Fig. 2) is consistent with the species phylogeny based on a much larger (>6.5 kb) data set consisting of six autosomal and X-linked loci (Kopp and Barmina, 2005). Namely, *D. bipectinata* and *D. parabipectinata* are the two most closely related species, followed by *D. malerkotliana*, while *D. pseudoananassae* forms a relatively distant outgroup (Fig. 2). Templeton test (Larson, 1994; Templeton, 1983) shows that the node separating *D. malerkotliana* from the (*D. bipectinata* + *D. parabipectinata*) lineage is significantly supported ($P = 0.025$). In contrast, in the analysis of individual autosomal and X-linked loci, alleles from the same species rarely form monophyletic clades. In particular, the monophyly of *D. malerkotliana* is not supported by any of the six loci, and alleles from *D. bipectinata*, *D. parabipectinata*, and *D. malerkotliana* are frequently intermingled. Phylogenetic analysis is further complicated by recombination, which is prevalent at most loci (Kopp and Barmina, 2005). In contrast, all species, including *D. malerkotliana*, are completely differentiated

Gene Region	kl2-L	kl2-L	kl2-L	kl2-L	kl2-L	kl2-L	kl2-L	kl2-L	kl2-L	kl2-L	kl2-L	kl2-L	kl2-L	kl2-L	kl2-L	kl2-L	kl2-L	kl2-L	kl2-L	kl2-L	kl2-L	kl2-L	kl2-L	kl2-L	kl2-L	kl2-L	kl2-L	kl2-L	kl2-L
Position	17	26	53	56	59	62	68	74	83	86	96	104	110	158	161	167	194	221	248	314	321	326	347	**387***	401	**422***	**427***	470	479
Species / Subspecies																													
bipectinata Asian	G	T	C	A	T	T	A	T	A	G	A	C	C	A	A	T	A	A	C	A	C	T	T	G	G	G	G	G	G
bipectinata Pacific	G	C	C	A	T	T	A	T	A	G	A	C	C	A	A	T	A	A	C	G	C	T	T	G	G	G	G	G	G
parabipectinata	G	T	C	G	T	T	A	T	A	G	A	C	C	A	A	T	A	A	C	G	C	T	T	G	G	G	G	G	G
malerkotliana A	A	T	C	A	T	T	A	C	A	G	A	T	C	A	A	A	A	A	C	G	C	T	T	G	G	G	G	G	C
malerkotliana B	G	T	C	A	T	C	A	C	A	G	A	T	C	A	A	T	A	A	C	G	C	T	T	G	G	G	G	G	C
p. pseudoananassae	G	T	T	A	C	T	C	T	T	A	C	C	T	T	C	T	G	T	T	G	A	C	T	T	A	A	A	A	G
p. nigrens	G	T	T	A	C	T	C	T	T	A	C	C	T	T	C	T	G	T	T	G	A	C	G	T	A	A	A	A	G
ananassae	A	T	T	A	T	T	A	G	A	T	C	C	T	A	A	A	G	A	C	A	C	T	T	G	G	G	A	A	G

(Continued)

Gene Region	kl2-L	kl2-L	kl2-L	kl2-L	kl2-L	kl2-L	kl2-R	kl2-R	kl2-R	kl2-R	kl2-R	kl2-R	kl2-R	kl2-R	kl2-R	kl2-R	kl2-R	kl2-R	kl2-R	kl2-R	kl2-R	kl2-R	kl2-R	kl2-R	kl2-R	kl2-R	kl2-R	kl2-R	kl2-R
Position	480	488	501	506	521	548	2	11	38	44	50	62	116	**148***	167	179	251	263	293	320	335	362	383	398	416	419	434	488	
Species / Subspecies																													
bipectinata Asian	C	G	C	A	C	A	A	G	A	T	A	T	A	C	C	T	C	G	T	A	T	A	C	T	C	T	G	C	
bipectinata Pacific	C	G	C	A	C	A	A	G	A	T	A	T	A	C	C	T	C	G	T	A	T	G	C	T	C	T	G	T	
parabipectinata	C	G	C	A	C	A	A	G	A	T	A	T	A	C	C	T	C	G	T	A	T	G	C	T	C	T	G	C	
malerkotliana A	C	A	C	A	C	A	A	A	A	G	A	T	G	C	A	T	C	G	T	A	T	G	C	T	C	T	G	C	
malerkotliana B	C	A	C	A	C	A	A	A	A	G	A	T	G	C	A	T	C	G	T	A	T	G	C	T	C	T	G	C	
p. pseudoananassae	A	A	T	G	A	C	G	G	T	G	G	C	A	C	A	C	A	A	G	T	C	G	T	C	T	C	A	C	
p. nigrens	A	A	T	G	A	C	G	G	T	G	G	C	A	G	A	C	A	A	G	T	C	G	C	T	T	C	A	C	
ananassae	C	A	T	A	T	A	A	A	A	G	A	T	T	C	A	T	C	G	G	T	C	G	C	T	C	A	A	G	

Fig. 1. Variable nucleotide positions in the *bipectinata* species complex. Derived character states (inferred from a comparison with *D. ananassae*) are shaded. Replacement substitutions are indicated by asterisks.

from each other at both Y-chromosomal loci (Figs. 1 and 2), and the sister-group relationship between *D. bipectinata* and *D. parabipectinata* is supported unambiguously (Fig. 2). The interspecific divergence of Y-chromosomal sequences is very low (~0.3% between *D. bipectinata* and *D. parabipectinata*, and 1.0–1.1% between these species and *D. malerkotliana*; Table 3), consistent with their recent radiation. In fact, interspecific divergence of Y-chromosomal sequences is considerably lower than the average distances between autosomal and X-chromosomal alleles within species, supporting the view that most genetic variation segregating in *D. bipectinata*, *D. parabipectinata*, and *D. malerkotliana* predates the origin of these species (Kopp and Barmina, 2005).

To confirm that exact correspondence between species and Y-chromosomal haplotypes is not an artifact of limited sampling, we sequenced *kl2-Right* from an additional 53 lines of *D. bipectinata*, *D. parabipectinata*, and *D. malerkotliana* collected throughout the geographic range of each species. No new haplotypes were discovered, and each haplotype was strictly confined to a single species or subspecies (Supplement Table 1). These results indicate that few, if any, ancestral polymorphisms are segregating within species, and there is no gene flow across species boundaries at Y-chromosomal loci.

3.2. Duplication and intrachromosomal recombination of Y-chromosomal sequences in the simulans species complex

It is a well-known observation that randomly chosen autosomal and X-chromosomal sequences do not resolve phylogenetic relationships among *D. simulans*, *D. mauritiana*, and *D. sechellia* (Eanes et al., 1996; Kliman et al., 2000; Kliman and Hey, 1993; Noor and Kliman, 2003). To summarize the information available to date, we performed a phylogenetic analysis of 16 previously published data sets (Kliman et al., 2000; Kulathinal et al., 2003; Ting et al., 2000). The results are shown in Supplement Table 2. As noted before, most loci support monophyly for *D. sechellia* and *D. mauritiana*, but not for *D. simulans*. Typically, *D. simulans* appears to be paraphyletic with respect to *D. mauritiana*, *D. sechellia*, or both. This is consistent with the scenario of a widespread ancestral "proto-*simulans*" population giving rise to the two island endemics (Kliman et al., 2000). What interests us, however, is the order of speciation events. Most autosomal and X-chromosomal loci provide little support for any of the possible two-species clades, although evidence appears to be the strongest in favor of the (*D. simulans* + *D. mauritiana*) clade, with *D. sechellia* as the outgroup (Supplement Table 2).

Y-chromosomal sequences also fail to resolve species relationships, but for a very peculiar reason. At least one gene region (*kl2-Right*) is duplicated in *D. simulans*, *D. mauritiana*, and *D. sechellia*. Phylogenetic analysis shows that gene duplication predates the radiation of the *simulans* species complex (Fig. 3). Neither duplicate can be amplified from females, indicating that both remain Y-linked. One of the gene copies is quite conserved relative to *D. melanogaster*, while the other copy is highly diverged both from the *melanogaster* sequence and from the conserved duplicate

Table 3
Divergence between Y-chromosomal haplotypes

Species/subspecies	*bipectinata* A	*bipectinata* P	*parabipectinata*	*malerkotliana* A	*malerkotliana* B	*p. pseudoananassae*	*p. nigrens*
(A)							
bipectinata (Asian)	—	0.379	0.284	1.137	1.043	4.265	4.265
bipectinata (Pacific)	4	—	0.284	1.137	1.043	4.265	4.265
parabipectinata	3	3	—	1.043	0.948	4.171	4.171
malerkotliana A	12	12	11	—	0.284	4.455	4.455
malerkotliana B	11	11	10	3	—	4.360	4.360
p. pseudoananassae	45	45	44	47	46	—	0.379
p. nigrens	45	45	44	47	46	4	—
Species/gene	*simulans*	*mauritiana*	*sechellia*	*simulans*-Dup	*mauritiana*-Dup	*sechellia*-Dup	*melanogaster*
(B)							
simulans	—	0.783	0.653	3.417	3.804	5.587	3.655
mauritiana	6	—	0.653	3.417	3.668	5.587	3.394
sechellia	5	5	—	3.285	3.668	4.888	3.264
simulans-Dup	26	26	25	—	4.348	5.866	5.782
mauritiana-Dup	28	27	27	32	—	6.657	6.793
sechellia-Dup	40	40	35	42	46	—	7.402
melanogaster	28	26	25	44	50	53	—
Species	*simulans*	*mauritiana*	*sechellia*				*melanogaster*
(C)							
simulans	—	0.547	0.656				4.654
mauritiana	10	—	0.656				4.432
sechellia	12	12	—				4.654
melanogaster	84	80	84				—

(A) *bipectinata* species complex, both gene regions combined. (B) *kl2-Right*, including both paralogs, in the *simulans* species complex. (C) *simulans* species complex, all three gene regions combined, excluding the diverged paralog of *kl2-Right* ("Dup"). For each pair of taxa, absolute genetic distance (in percent) is shown above the diagonal, and the total number of nucleotide differences below the diagonal.

(Table 3). The more conserved duplicates have uninterrupted open reading frames in all three species. The more diverged duplicates may be functional in *D. simulans* and *D. sechellia*, but contain multiple stop codons in *D. mauritiana* (see Supplementary Data).

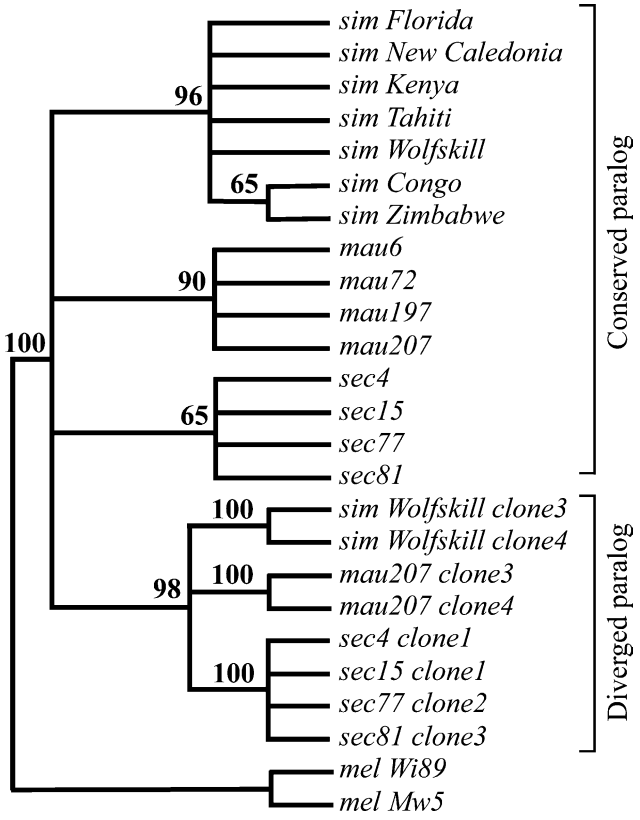

Fig. 3. Phylogenetic relationships among duplicated *kl2-Right* sequences in the *simulans* species complex. The "tree" was reconstructed by maximum parsimony. Bootstrap support values above 50% are given at the appropriate nodes. This is not a true gene tree, since the two paralogs have undergone multiple recombination events.

Each of the duplicate *kl2-Right* gene regions is monomorphic within each species, with the exception of the more conserved copy in *D. simulans*, which contains two variable sites. All *simulans* complex sequences cluster together with strong support (Fig. 3). Within this group, the diverged duplicates from *D. simulans*, *D. sechellia*, and *D. mauritiana* clearly form a monophyletic clade, whereas the conserved duplicates do not (Fig. 3). Reconstruction of a haplotype network has proven to be impossible because different sites favor different genealogies. Four-gamete test (Hudson and Kaplan, 1985) reveals a minimum of six recombination events in the 777 bp *kl2*-Right region (Fig. 4). Although this observation could in principle be explained by an exceedingly high frequency of recurrent mutations, we do not find any homoplastic sites in the other two sequenced regions (*kl2-Left* and *ory*), arguing against this scenario. Intrachromosomal gene conversion between the two duplicate copies of *kl2-Right* appears to be a more plausible explanation. Variable sites in the *kl2-Right* region can be subdivided into three classes (Fig. 4). The first type (such as positions 466 through 595) are consistent with a simple history of duplication followed by independent divergence of paralogous genes. Other sites (e.g., positions 72 and 110) involve a single gene conversion event in one of the species. Finally, sites 144, 620, and 740 have more complex histories that require multiple conversion events (Fig. 4). In the end, intrachromosomal recombination complicates the history of *kl2-Right* to such an extent that its phylogeny cannot be reconstructed (Fig. 3).

In contrast to *kl2-Right*, we do not observe any intraspecific variation in *kl2-Left* or *ory* in any of the species (except *D. melanogaster*), suggesting that these regions are either not duplicated, or if they are, the two duplicates remain completely identical. Once the more diverged duplicates of *kl2-Right* are excluded from analysis, the remaining data set consisting of 3 gene regions and 1849 bp of sequence contains 18 sites that are variable within the *simulans* complex (Fig. 5). Two of these (position 18 of *kl2-Left* and

Position	72	110	144	161	162	184	238	466-470	511	514	517	595	620	679	740	750
Site Type	CON	CON	CON	DUP	DUP	CON	CON	DUP	DUP	DUP	DUP	DUP	CON	CON	CON	CON
Species/Gene																
simulans	G	T	A	T	T	A	C	ATAAT	T	A	C	A	T	T	C	G
mauritiana	A	T	G	T	T	A	C	ATAAT	T	A	C	A	T	T	C	G
sechellia	A	T	A	T	T	A	C	ATAAT	T	A	C	A	T	T	T	T
simulans-Dup	G	T	G	A	A	T	T	Deletion	A	G	A	G	T	A	T	G
mauritiana-Dup	A	G	A	A	A	A	T	Deletion	A	G	A	G	T	T	C	G
sechellia-Dup	A	G	A	A	A	T	C	Deletion	A	G	A	G	C	A	T	T
melanogaster	A	T	G	T	T	T	C	ATAAT	T	A	C	A	C	T	T	G
Recombination events																

Fig. 4. Variable sites in the duplicated *kl2-Right* region in the *simulans* species complex. Derived character states (inferred from a comparison with *D. melanogaster*) are shaded. Recombination events inferred by the four-gamete test are indicated by brackets at the bottom of the figure. Some sites ("DUP") are consistent with simple gene duplication followed by independent divergence of the two paralogs. Other sites ("CON") appear to have been altered by gene conversion.

Gene Region	kl2-L	kl2-L	kl2-L	kl2-R	kl2-R	kl2-R	kl2-R	kl2-R	kl2-R	kl2-R	kl2-R	kl2-R	ory	ory	ory	ory	ory	ory	ory
Position	17	**18***	347	72	144	315	351	556	667	679	712	**713***	319	334	339	363	450	480	664
Species / strain																			
sim Congo	A	A	C	G	A	C	G	C	T	C	C	G	G	C	C	A	A	A	A
sim Florida	A	A	C	G	A	G	G	C	T	C	C	G	G	C	C	A	A	A	A
sim Kenya	A	A	C	G	A	G	G	C	T	C	C	G	G	C	C	A	A	A	A
sim New Caledonia	A	A	C	G	A	G	G	C	T	C	C	G	T	C	C	A	A	A	A
sim Tahiti	A	A	C	G	A	G	G	C	T	C	C	G	G	C	C	A	A	A	A
sim Wolfskill	A	A	C	G	A	G	G	C	T	C	C	G	G	C	C	A	A	A	A
sim Zimbabwe	A	A	C	G	A	C	G	C	T	C	C	G	G	C	C	A	A	A	A
mauritiana 6	A	A	C	A	G	G	C	T	A	T	C	G	G	T	A	A	A	G	A
mauritiana 72	A	A	C	A	G	G	C	T	A	T	C	G	G	T	A	A	A	G	A
mauritiana 197	A	A	C	A	G	G	C	T	A	T	C	G	G	T	A	A	A	G	A
mauritiana 207	A	A	C	A	G	G	C	T	A	T	C	G	G	T	A	A	A	G	A
sechellia 4	G	G	G	A	A	G	G	T	T	T	T	T	G	C	A	T	G	A	A
sechellia 15	G	G	G	A	A	G	G	T	T	T	T	T	G	C	A	T	G	A	A
sechellia 77	G	G	G	A	A	G	G	T	T	T	T	T	G	C	A	T	G	A	A
sechellia 81	G	G	G	A	A	G	G	T	T	T	T	T	G	C	A	T	G	A	A
melanogaster Wi89	A	A	C	A	G	G	G	T	T	T	T	G	G	T	A	A	A	A	T
melanogaster Mw5	A	A	C	A	G	G	G	T	T	T	T	G	G	T	A	A	A	A	A

Fig. 5. Variable nucleotide positions in the *simulans* species complex, excluding the diverged paralogs of *kl2-Right*. Derived character states (inferred from a comparison with *D. melanogaster*) are shaded in light gray. Synapomorphic characters are boxed and shaded in darker gray. However, two of the synapomorphies (sites 144 and 712 of *kl2-Right*) are potentially false, since these sites have been influenced by gene conversion (Fig. 4).

position 713 of *kl2-Right*) are replacement substitutions, while the rest are silent. The use of *D. melanogaster* sequences as outgroups allows us to polarize nucleotide substitutions and assign them to specific internal branches of the phylogeny. 13 of the variable sites are autapomorphies, i.e., fixed changes limited to a single species. In contrast to most autosomal and X-chromosomal genes (Kliman et al., 2000), no segregating sites are shared across species boundaries, so that the monophyly of each species, including *D. simulans*, is well supported (Fig. 3). 2 sites segregate in *D. simulans*, while *D. mauritiana* and *D. sechellia* are completely monomorphic (Fig. 5). Of the three apparent synapomorphies, two (positions 144 and 712 of *kl2-Right*) result from gene conversion, and are therefore not informative in inferring species phylogeny. The remaining synapomorphy (position 334 of *ory*) supports the grouping of *D. simulans* with *D. sechellia*. This grouping is at odds with the plurality of autosomal and X-linked loci (Supplement Table 2), and conflicts with the species phylogeny based on the proposed "speciation gene" *Odysseus* (Ting et al., 2000). However, the Templeton test (Templeton, 1983; Larson, 1994) indicates that the ((*simulans*, *sechellia*) *mauritiana*) tree is not significantly better supported than the more widely accepted ((*simulans, mauritiana*) *sechellia*) tree ($P = 0.564$). Thus, the small amount of evidence precludes us from making any inferences about species relationships in the *simulans* complex based on the Y chromosome.

4. Discussion

Our expectations for the phylogenetic utility of Y-chromosomal sequences are based on a combination of three factors: rapid neutral coalescence, early cessation of gene flow due to Haldane's rule, and the absence of recombination. An additional factor that needs to be considered is that the Y chromosome may experience frequent episodes of positive selection due to its importance for male reproductive success (McAllister and Charlesworth, 1999; Charlesworth and Charlesworth, 2000; Clark et al., 2000; Chippindale and Rice, 2001; Bachtrog, 2004; Gerrard and Filatov, 2005). If this is the case, coalescence times for all Y-linked loci may be considerably shorter than under the neutral model of sequence evolution. This resulting decrease in intraspecific variation would further reduce the probability of lineage sorting during speciation. When several species diverge in rapid succession, selective sweeps affecting the entire Y chromosome would increase the probability of coalescence during brief intervals between consecutive speciation events, allowing us to recover species phylogeny from a sample of Y-linked loci.

We have tested this idea in two groups of recently diverged *Drosophila* species. We find intraspecific variation on the Y chromosome to be very low (or non-existent) in all examined species. This result is in agreement with an earlier report of Zurovcova and Eanes, who sequenced a 1738 bp fragment of a different Y-chromosomal gene (*kl-5*) in *D. simulans*, and found only one polymorphic site (Zurovcova and Eanes, 1999). Polymorphism levels on the Y chromosome are also greatly reduced in a variety of mammalian species (Hellborg and Ellegren, 2004; McPherson et al., 2001). Such reduced variation is expected in a non-recombining region, especially if it undergoes frequent selective sweeps. Selective sweeps might also contribute to the complete absence of polymorphisms on the Y chromosome that are shared between species (Figs. 1 and 5). As a result, all species, including those with large effective population sizes (such as *D. simulans* and *D. malerkotliana*), appear as monophyletic clades in the analysis of Y-linked loci (Figs. 1, and 3). In this respect, Y-chromosomal genes are quite different from the majority of autosomal and X-chromosomal loci (Kliman et al., 2000; Kopp and Barmina, 2005).

The contrast between Y-chromosomal and mitochondrial sequences is particularly striking. mtDNA often

suggests false species relationships due to post-speciation gene flow in *Drosophila* (Ballard, 2000a,b; Machado et al., 2002; Solignac, 2004) and in a variety of other animal taxa (Ferris et al., 1983; Martin et al., 2002; Mukai et al., 1997; Ruedi et al., 1997; Shaw, 2002; Sota, 2002). In the house mouse *Mus musculus*, mtDNA shows evidence of secondary introgression between well differentiated subspecies at several hybrid zones (Boissinot and Boursot, 1997; Ferris et al., 1983). Y-chromosomal sequences, however, are confined to each subspecies even inside the hybrid zones, a pattern that has been attributed to selection against the introgressed Y chromosomes (Boissinot and Boursot, 1997; Dod et al., 1993; Karn et al., 2002; Vanlerberghe et al., 1986). On the other hand, in birds (which are female-heterogametic), mtDNA shows less evidence of gene flow across species than do nuclear markers, suggesting that selection against female hybrids is operating (Crochet et al., 2003; Helbig et al., 2001). In both *D. simulans* and *D. bipectinata* species complexes, interspecific introgression of mtDNA appears to be common (Kopp and Barmina, 2005; Solignac, 2004), whereas Y-chromosomal haplotypes are strictly species-specific. The difference between matrilineal and patrilineal markers is also evident in the Asian and Pacific subspecies of *D. bipectinata*. Hybridization between the two subspecies produces partially or completely sterile males, but the hybrid females are invariably fertile (Kopp and Frank, 2005). Several strains of *D. bipectinata* collected in New Guinea and Queensland behave as the Pacific subspecies in reproductive isolation tests but carry mtDNA haplotypes of the Asian subspecies, or vice versa, suggesting that the two subspecies may hybridize at some frequency. In contrast to mtDNA, Y-chromosomal haplotypes coincide exactly with the subspecies defined on the basis of reproductive isolation (Supplement Table 1). This observation highlights the importance of Haldane's rule in determining the phylogenetic value of molecular markers in recently diverged taxa.

We were most interested in the usefulness of Y-chromosomal sequences for reconstructing interspecific relationships. In the *D. bipectinata* species complex, consistent with our expectations, the Y-chromosomal gene tree is almost certainly an accurate reflection of species phylogeny. The sister-group relationship between *D. bipectinata* and *D. parabipectinata*, which is supported by a large multilocus data set as well as by morphological similarities (Kopp and Barmina, 2005), is recovered unambiguously by the Y-chromosomal haplotype network (Fig. 2). In the *simulans* complex, the situation is greatly complicated by gene duplication and intrachromosomal conversion between duplicate genes. The mechanism of this duplication is not clear, but the presence of introns in both paralogs argues against retrotransposition. Ectopic intrachromosomal recombination, possibly mediated by transposable elements or other repetitive sequences, is a more likely explanation. Such recombination is known to occur on the human Y chromosome, often resulting in gene deletions (Blanco et al., 2000; Repping et al., 2002; Sun et al., 2000). Gene conversion among paralogous genes has most often been studied in the context of concerted evolution of gene families (Bettencourt and Feder, 2002; Liu et al., 1987; Ohta, 1985; Ohta, 2000; Schlotterer and Tautz, 1994), but may also contribute to the maintenance of allelic diversity at each paralogous locus (Lazzaro and Clark, 2001). In the duplicated Y-chromosomal genes, the most obvious effect of conversion is to introduce recombination into an otherwise non-recombining part of the genome. Thus, in contrast to the *bipectinata* complex, Y-chromosomal sequences in the *simulans* complex are not related by a single unambiguous genealogy, and are of little use in reconstructing species relationships.

In conclusion, we believe that Y-linked loci will prove useful for resolving phylogenetic relationships in many groups of recently diverged, partially interfertile species. In male-heterogametic organisms, they should be far more useful than the commonly used mitochondrial DNA. However, gene duplication in the *simulans* complex serves as an important cautionary tale. In fact, we found that another Y-chromosomal gene, *kl-3*, is duplicated in the *bipectinata* complex (Frank and Kopp, data not shown), suggesting that duplication of Y-linked genes may not be uncommon. For a full exploration of species histories, multi-locus data sets including genes from different chromosomes, including the Y chromosome, are indispensable.

Acknowledgments

We thank D. Begun, Y. Fuyama, C. Jones, C.-S. Ng, J. Pool, S.-C. Tsaur, and M. Watada for fly strains, A. Clark for advice on Y-chromosomal gene structure, J. Fu and A. Wong for technical help, and D. Begun, M. Turelli, and two anonymous reviewers for discussions and comments on the manuscript. This work was supported by NSF Grant IOB-05-18654 and funds from UC-Davis to A.K.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2005.10.006.

References

Bachtrog, D., 2004. Evidence that positive selection drives Y-chromosome degeneration in Drosophila miranda. Nat. Genet. 36, 518–522.

Ballard, J.W., 2000a. Comparative genomics of mitochondrial DNA in *Drosophila simulans*. J. Mol. Evol. 51, 64–75.

Ballard, J.W., 2000b. When one is not enough: introgression of mitochondrial DNA in *Drosophila*. Mol. Biol. Evol. 17, 1126–1130.

Bettencourt, B.R., Feder, M.E., 2002. Rapid concerted evolution via gene conversion at the *Drosophila* hsp70 genes. J. Mol. Evol. 54, 569–586.

Blanco, P., Shlumukova, M., Sargent, C.A., Jobling, M.A., Affara, N., Hurles, M.E., 2000. Divergent outcomes of intrachromosomal recombination on the human Y chromosome: male infertility and recurrent polymorphism. J. Med. Genet. 37, 752–758.

Bock, I.R., 1971. Taxonomy of the *Drosophila bipectinata* species complex. Univ. Texas Publs. 6, 273–280.

Bock, I.R., 1978. The *bipectinata* complex: a study of interspecific hybridization in the genus *Drosophila*. Aust. J. biol. Sci. 31, 197–208.

Boissinot, S., Boursot, P., 1997. Discordant phylogeographic patterns between the Y chromosome and mitochondrial DNA in the house mouse: selection on the Y chromosome? Genetics 146, 1019–1034.

Caccone, A., Moriyama, E.N., Gleason, J.M., Nigro, L., Powell, J.R., 1996. A molecular phylogeny for the *Drosophila melanogaster* subgroup and the problem of polymorphism data. Mol. Biol. Evol. 13, 1224–1232.

Cariou, M.L., 1987. Biochemical phylogeny of the eight species in the *Drosophila melanogaster* subgroup, including *D. sechellia* and *D. orena*. Genet. Res. 50, 181–185.

Carvalho, A.B., Dobo, B.A., Vibranovski, M.D., Clark, A.G., 2001. Identification of five new genes on the Y chromosome of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA 98, 13225–13230.

Carvalho, A.B., Lazzaro, B.P., Clark, A.G., 2000. Y chromosomal fertility factors kl-2 and kl-3 of *Drosophila melanogaster* encode dynein heavy chain polypeptides. Proc. Natl. Acad. Sci. USA 97, 13239–13244.

Charlesworth, B., Charlesworth, D., 2000. The degeneration of Y chromosomes. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 355, 1563–1572.

Chippindale, A.K., Rice, W.R., 2001. Y chromosome polymorphism is a strong determinant of male fitness in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA 98, 5677–5682.

Clark, A.G., Dermitzakis, E.T., Civetta, A., 2000. Nontransitivity of sperm precedence in *Drosophila*. Evol. Int. J. Org. Evol. 54, 1030–1035.

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

Coyne, J.A., 1996. Genetics of sexual isolation in male hybrids of *Drosophila simulans* and *D. mauritiana*. Genet. Res. 68, 211–220.

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? Evol. Int. J. Org. Evol. 57, 2865–2878.

Dod, B., Jermin, L.S., Boursot, P., Chapman, V.M., Nielsen, J.T., Bonhomme, F., 1993. Counterselection on sex chromosomes in the *Mus musculus* European hybrid zone. J. Evol. Biol. 6, 529–546.

Eanes, W.F., Kirchner, M., Yoon, J., Biermann, C.H., Wang, I.N., McCartney, M.A., Verrelli, B.C., 1996. Historical selection, amino acid polymorphism and lineage-specific divergence at the G6pd locus in *Drosophila melanogaster* and *D. simulans*. Genetics 144, 1027–1041.

Fang, S, Takahashi, A., Wu, C.I., 2002. A mutation in the promoter of desaturase 2 is correlated with sexual isolation between *Drosophila* behavioral races. Genetics 162, 781–784.

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. Proc. Natl. Acad. Sci. USA 80, 2290–2294.

Gerrard, D.T., Filatov, D.A., 2005. Positive and negative selection on Mammalian y chromosomes. Mol. Biol. Evol. 22, 1423–1432.

Greenberg, A.J., Moran, J.R., Coyne, J.A., Wu, C.I., 2003. Ecological adaptation during incipient speciation revealed by precise gene replacement. Science 302, 1754–1757.

Haldane, J.B.S., 1922. Sex-ratio and unidirectional sterility in hybrid animals. J. Genet. 12, 101–109.

Harr, B., Weiss, S., David, J.R., Brem, G., Schlotterer, C., 1998. A microsatellite-based multilocus phylogeny of the *Drosophila melanogaster* species complex. Curr. Biol. 8, 1183–1186.

Helbig, A.J., Salomon, M., Bensch, S., Seibold, I., 2001. Male-biased gene flow across an avian hybrid zone: evidence for mitochondrial and microsatellite DNA. J. Evol. Biol. 14, 277–287.

Hellborg, L., Ellegren, H., 2004. Low levels of nucleotide diversity in mammalian Y chromosomes. Mol. Biol. Evol. 21, 158–163.

Hey, J., Kliman, R.M., 1993. Population genetics and phylogenetics of DNA sequence variation at multiple loci within the *Drosophila melanogaster* species complex. Mol. Biol. Evol. 10, 804–822.

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

Hollocher, H., Wu, C.I., 1996. The genetics of reproductive isolation in the *Drosophila simulans* clade: X vs. autosomal effects and male vs. female effects. Genetics 143, 1243–1255.

Hoskins, R.A., Smith, C.D., Carlson, J.W., Carvalho, A.B., Halpern, A., Kaminker, J.S., Kennedy, C., Mungall, C.J., Sullivan, B.A., Sutton, G.G., Yasuhara, J.C., Wakimoto, B.T., Myers, E.W., Celniker, S.E., Rubin, G.M., Karpen, G.H., 2002. Heterochromatic sequences in a *Drosophila* whole-genome shotgun assembly. Genome Biol., 3. RESEARCH0085.

Hudson, R.R., 1990. Gene genealogies and the coalescent process. Oxf. Surv. Evol. Biol. 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., Turelli, M., 2003. Stochsticity overrules the "three-times rule: genetic drift, genetic draft, and coalescence times for nuclear loci versus mitochondrial DNA. Evolution 57, 182–190.

Jobling, M.A., Tyler-Smith, C., 2003. The human Y chromosome: an evolutionary marker comes of age. Nat. Rev. Genet. 4, 598–612.

Karn, R.C., Orth, A., Bonhomme, F., Boursot, P., 2002. The complex history of a gene proposed to participate in a sexual isolation mechanism in house mice. Mol. Biol. Evol. 19, 462–471.

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.

Kliman, R.M., Hey, J., 1993. DNA sequence variation at the period locus within and among species of the *Drosophila melanogaster* complex. Genetics 133, 375–387.

Kopp, A., Barmina, O., 2005. Evolutionary history of the *Drosophila bipectinata* species complex. Genet. Res. 85, 23–46.

Kopp, A., Frank, A., 2005. Speciation in progress? A continuum of reproductive isolation in *D. bipectinata*. Genetica 125, 55–68.

Kulathinal, R.J., Skwarek, L., Morton, R.A., Singh, R.S., 2003. Rapid evolution of the sex-determining gene, transformer: structural diversity and rate heterogeneity among sibling species of *Drosophila*. Mol. Biol. Evol. 20, 441–452.

Larson, A., 1994. The comparison of morphological and molecular data in phylogenetic systematics. In: Schierwater, B., Streit, B., Wagner, G.P., DeSalle, R. (Eds.), Molecular ecology and evolution: Approaches and applications. Birkhauser Verlag, Basel, Switzerland.

Lazzaro, B.P., Clark, A.G., 2001. Evidence for recurrent paralogous gene conversion and exceptional allelic divergence in the Attacin genes of *Drosophila melanogaster*. Genetics 159, 659–671.

Liu, T.J., Liu, L., Marzluff, W.F., 1987. Mouse histone H2A and H2B genes: four functional genes and a pseudogene undergoing gene conversion with a closely linked functional gene. Nucleic Acids Res. 15, 3023–3039.

Macdonald, S.J., Goldstein, D.B., 1999. A quantitative genetic analysis of male sexual traits distinguishing the sibling species *Drosophila simulans* and *D. sechellia*. Genetics 153, 1683–1699.

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

Maddison, D.R., Maddison, W.P., 2001. MacClade 4: Analysis of phylogeny and character evolution. Version 4.03. In. Sinauer Associates, Sunderland, Massachusetts.

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. Mol. Biol. 11, 387–397.

Matsuda, M., Tomimura, Y., Tobari, Y.N., 2005. Reproductive isolation among geographical populations of *Drosophila bipectinata duda* (Diptera, drosophilidae) with recognition of three subspecies. Genetica 125, 69–78.

McAllister, B.F., Charlesworth, B., 1999. Reduced sequence variability on the Neo-Y chromosome of *Drosophila americana americana*. Genetics 153, 221–233.

McPherson, J.D. et al., 2001. A physical map of the human genome. Nature 409, 934–941.

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. Mol. Biol. Evol. 14, 1258–1265.

Nichols, R., 2001. Gene trees and species trees are not the same. Trends Ecol. Evol. 16, 358–364.

Noor, M.A., Kliman, R.M., 2003. Variability on the dot chromosome in the *Drosophila simulans* clade. Genetica 118, 51–58.

Ohta, T., 1985. A model of duplicative transposition and gene conversion for repetitive DNA families. Genetics 110, 513–524.

Ohta, T., 2000. Evolution of gene families. Gene 259, 45–52.

Orr, H.A., Turelli, M., 1996. Dominance and Haldane's rule. Genetics 143, 613–616.

Repping, S., Skaletsky, H., Lange, J., Silber, S., Van Der Veen, F., Oates, R.D., Page, D.C., Rozen, S., 2002. Recombination between palindromes P5 and P1 on the human Y chromosome causes massive deletions and spermatogenic failure. Am. J. Hum. Genet. 71, 906–922.

Ruedi, M., Smith, M.F., Patton, J.L., 1997. Phylogenetic evidence of mitochondrial DNA introgression among pocket gophers in New Mexico (family Geomyidae). Mol. Ecol. 6, 453–462.

Schlotterer, C., Tautz, D., 1994. Chromosomal homogeneity of *Drosophila* ribosomal DNA arrays suggests intrachromosomal exchanges drive concerted evolution. Curr. Biol. 4, 777–783.

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. Proc. Natl. Acad. Sci. USA 99, 16122–16127.

Singh, S., Singh, B.N., 2001. *Drosophila bipectinata* species complex. Indian J. Exp. Biol. 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 Ecol. 44, 145–156.

Stone, A.C., Griffiths, R.C., Zegura, S.L., Hammer, M.F., 2002. High levels of Y-chromosome nucleotide diversity in the genus Pan. Proc. Natl. Acad. Sci. USA 99, 43–48.

Sun, C., Skaletsky, H., Rozen, S., Gromoll, J., Nieschlag, E., Oates, R., Page, D.C., 2000. Deletion of azoospermia factor a (AZFa) region of human Y chromosome caused by recombination between HERV15 proviruses. Hum. Mol. Genet. 9, 2291–2296.

Swofford, D.L., 2000. PAUP*: Phylogenetic analysis using parsimony, version 4.0. Sinauer, Sunderland, MA.

Takahashi, A., Tsaur, S.C., Coyne, J.A., Wu, C.I., 2001. The nucleotide changes governing cuticular hydrocarbon variation and their evolution in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA 98, 3920–3925.

Tao, Y., Chen, S., Hartl, D.L., Laurie, C.C., 2003a. Genetic dissection of hybrid incompatibilities between *Drosophila simulans* and *D. mauritiana*. I. Differential accumulation of hybrid male sterility effects on the X and autosomes. Genetics 164, 1383–1397.

Tao, Y., Zeng, Z.B., Li, J., Hartl, D.L., Laurie, C.C., 2003b. Genetic dissection of hybrid incompatibilities between *Drosophila simulans* and *D. mauritiana*. II. Mapping hybrid male sterility loci on the third chromosome. Genetics 164, 1399–1418.

Templeton, A.R., 1983. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. Evolution 37, 221–244.

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. Proc. Natl. Acad. Sci. USA 97, 5313–5316.

Ting, C.T., Tsaur, S.C., Wu, M.L., Wu, C.I., 1998. A rapidly evolving homeobox at the site of a hybrid sterility gene. Science 282, 1501–1504.

Tosi, A.J., Morales, J.C., Melnick, D.J., 2000. Comparison of Y chromosome and mtDNA phylogenies leads to unique inferences of macaque evolutionary history. Mol. Phylogenet. Evol. 17, 133–144.

True, J.R., Weir, B.S., Laurie, C.C., 1996. A genome-wide survey of hybrid incompatibility factors by the introgression of marked segments of *Drosophila mauritiana* chromosomes into *Drosophila simulans*. Genetics 142, 819–837.

Turelli, M., 1998. The causes of Haldane's rule. Science 282, 889–891.

Turelli, M., Orr, H.A., 2000. Dominance, epistasis and the genetics of postzygotic isolation. Genetics 154, 1663–1679.

Underhill, P.A., Passarino, G., Lin, A.A., Shen, P., Mirazon Lahr, M., Foley, R.A., Oefner, P.J., Cavalli-Sforza, L.L., 2001. The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. Ann. Hum. Genet. 65, 43–62.

Underhill, P.A., Shen, P., Lin, A.A., Jin, L., Passarino, G., Yang, W.H., Kauffman, E., Bonne-Tamir, B., Bertranpetit, J., Francalacci, P., Ibrahim, M., Jenkins, T., Kidd, J.R., Mehdi, S.Q., Seielstad, M.T., Wells, R.S., Piazza, A., Davis, R.W., Feldman, M.W., Cavalli-Sforza, L.L., Oefner, P.J., 2000. Y chromosome sequence variation and the history of human populations. Nat. Genet. 26, 358–361.

Vanlerberghe, F., Dod, B., Boursot, P., Bellis, M., Bonhomme, F., 1986. Absence of Y-chromosome introgression across the hybrid zone between *Mus musculus domesticus* and *Mus musculus musculus*. Genet. Res. 48, 191–197.

Verkaar, E.L., Nijman, I.J., Beeke, M., Hanekamp, E., Lenstra, J.A., 2004. Maternal and paternal lineages in cross-breeding bovine species. Has wisent a hybrid origin? Mol. Biol. Evol. 21, 1165–1170.

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

Zurovcova, M., Eanes, W.F., 1999. Lack of nucleotide polymorphism in the Y-linked sperm flagellar dynein gene Dhc-Yh3 of *Drosophila melanogaster* and *D. simulans*. Genetics 153, 1709–1715.