Copyright © 1995 by the Genetics Society of America

Molecular Evolution of the Duplicated *Amy* Locus in the *Drosophila melanogaster* Species Subgroup: Concerted Evolution Only in the Coding Region and an Excess of Nonsynonymous Substitutions in Speciation

Hiroki Shibata and Tsuneyuki Yamazaki

Department of Biology, Faculty of Science, Kyushu University, Fukuoka 812, Japan

Manuscript received April 29, 1993
Accepted for publication June 9, 1995

ABSTRACT

From the analysis of restriction maps of the *Amy* region in eight sibling species belonging to the *Drosophila melanogaster* species subgroup, we herein show that the patterns of duplication of the *Amy* gene are almost the same in all species. This indicates that duplication occurred before speciation within this species subgroup. From the nucleotide sequence data, we show a strong within-species similarity between the duplicated loci in the *Amy* coding region. This is in contrast to a strong similarity in the 5′ and 3′ flanking regions within each locus (proximal or distal) throughout the species subgroup. This means that concerted evolution occurred only in the *Amy* coding region and that differentiated evolution between the duplication occurred in the flanking regions. Moreover, when comparing the species, we also found a significant excess of nonsynonymous substitutions. In particular, all the fixed substitutions specific to *D. erecta* were found to be nonsynonymous. We thus conclude that adaptive protein evolution occurred in the lineage of *D. erecta* that is a "specialist" species for host plants and probably also occurs in the process of speciation in general.

α-AMYLASE (EC 3.2.1.1) is one of the most extensively studied enzymes in the fruit fly, Drosophila. The amylase of *D. melanogaster* in natural populations is highly polymorphic in both isozyme frequencies and activities (Abe 1958; Kikkawa 1964; de Jong *et al.* 1972; Hickey 1979; Singh *et al.* 1982; Yamazaki *et al.* 1984; Langley *et al.* 1988). Molecular cloning of the *Amy* region of *D. melanogaster* verified that the gene consists of a duplication as inverted repeats and that the products of both *Amy* gene copies are enzymatically active (Gemmill *et al.* 1985, 1986; Levy *et al.* 1985). The nucleotide sequences of *Amy* gene copies from several strains have been determined (Boer and Hickey 1986; Okuyama and Yamazaki 1988; Inomata *et al.* 1995). The *Amy* structural gene is composed of one exon of 1482 nucleotides and is predictably translated into a polypeptide with a length of 494 amino acid residues. Duplications of the *Amy* gene of *D. melanogaster* sibling species were found using genomic Southern blotting (Payant *et al.* 1988). Therefore, this duplication is thought to have been an ancestral event before the radiation of this species subgroup.

There have been many studies showing that members of a multigene family evolved not independently but in a concerted manner (*e.g.*, Smith 1976; Arnheim *et al.* 1980; Coen *et al.* 1982; Matsuo and Yamazaki 1989). Hickey *et al.* (1991) have found evidence for concerted evolution of *Amy* locus in *D. melanogaster* and in *D. erecta* from nucleotide sequence data.

In this study, we present the molecular structure and the nucleotide sequence of the *Amy* region in eight sibling species belonging to the *D. melanogaster* species subgroup. This subgroup consists of three complexes: (1) the *melanogaster* complex that consists of four species, *D. melanogaster, D. simulans, D. mauritiana* and *D. sechellia*, (2) the *yakuba* complex that consists of *D. teissieri* and *D. yakuba* and (3) the complex that consists of *D. erecta* and *D. orena* (Lemeunier and Ashburner 1976; Cariou 1987; Lachaise *et al.* 1988). We show by molecular cloning that the *Amy* genes are indeed commonly duplicated throughout the species subgroup. We also show that in all eight species the coding sequences of the duplicated *Amy* genes have evolved in a concerted manner in contrast to independent evolution in their flanking sequences. In addition, we show a significant excess of nonsynonymous substitution fixed in both loci of each species. We present two plausible explanations for our results: adaptive protein evolution involving alteration of the ecology and a bottleneck effect that permits fixation of slightly deleterious mutations.

MATERIALS AND METHODS

Fly stocks: *D. mauritiana, D. sechellia, D. erecta, D. orena, D. teissieri* and *D. yakuba* were all obtained from the National Institute of Genetics, Japan. *D. simulans* was collected at Munakata near Fukuoka City in 1986. All strains are from isofemale lines. All fly stocks were reared on a corn meal-molasses medium at 25°. The mobilities of amylase isozymes of the sibling species studied here (*D. simulans, D. mauritiana, D. sechellia,*

Corresponding author: Tsuneyuki Yamazaki, Department of Biology, Faculty of Science, Kyushu University, 6-10-1 Hakozaki, Higashiku, Fukuoka City, Japan 812.
E-mail: tyamascb@mbox.nc.kyushu-u.ac.jp

D. erecta, D. orena, D. teissieri and *D. yakuba*) were 0.78, 0.78, 0.78, 0.53, 0.40, 0.85 and 0.78, respectively, relative to the mobility of *Amy*[1] of *D. melanogaster* using polyacrylamide gel electrophoresis at pH 8.9. Each isozyme of the sibling species appeared to be equivalent to the most common isozymes assigned by DAÏNOU *et al.* (1987).

Construction of genomic DNA library: Drosophila genomic DNAs were purified on a $CsCl_2$ gradient as described by BINGHAM *et al.* (1981). Genomic libraries were constructed from stocks of each of *D. simulans, D. mauritiana, D. sechellia, D. erecta, D. orena, D. teissieri* and *D. yakuba* according to FRICHAUF *et al.* (1983) using the commercially available packaging extract Giga-Pack (Strategene) according to the manufacturers recommendations. Screening for positive phage clones and isolation of those clones were conducted according to MANIATIS *et al.* (1982) using an *Eco*RI 3.8-kb fragment from λDm65 (GEMMILL *et al.* 1985) as a probe (shown in Figure 1).

Molecular cloning and sequencing: The Southern blot analyses were performed as described by MANIATIS *et al.* (1982). The nucleotide sequences of cloned *Amy* genes from the seven sibling species were determined. After subcloning into plasmid vector Bluescript SK+, sequencing reactions were performed according to the modified dideoxy method (SANGER *et al.* 1980) using a commercially available sequencing kit [sequenase (USB)] and several synthetic primers (17mer). For the sake of precision, the nucleotide sequences were determined in both strands of subcloned DNA.

Data deposition: The nucleotide sequence data reported in this article appear in the DDBJ, EMBL and GenBank Nucleotide Sequence Databases with the following accession numbers D17727-D17738, D21128, D21129.

Nucleotide data analysis: The calculations of genetic distance between coding sequences were performed using DNADIST in PHYLIP. The maximum matchings of the flanking sequence and the calculations of genetic distance were performed using GENETYX. All genetic distances were calculated using the method of JUKES and CANTOR (1969). Neighbor joining trees were constructed using NEIGHBOR in PHYLIP. A bootstrap analysis with 100 replications was performed to test the significance of our phylogenetic tree of coding sequence using SEQBOOT and CONSENSE in PHYLIP. We did not attempt a bootstrap analysis for the flanking sequence, because numerous small inversions and deletions tended to lead to misleading results in the multiple alignment.

RESULTS

A comparison of the restriction maps: Figure 1 shows the alignment of the restriction maps of the *Amy* regions. In all seven sibling species studied here (*D. simulans, D. mauritiana, D. sechellia, D. erecta, D. orena, D. teissieri* and *D. yakuba*), the amylase genes were duplicated as in *D. melanogaster*. The nucleotide lengths between the proximal and distal copies were almost the same in all the species studied (4.5 kb). *Bam-Sal-Bam* triads suggest that the duplications are in opposite directions in *D. mauritiana, D. sechellia, D. simulans, D. orena* and *D. yakuba*. In *D. erecta* both first *Bam*HI sites of the triad were absent, and the *Sal-Bam* dyads suggest the same structure of the duplication. In addition, in *D. teissieri*, which lacks both the first *Bam*HI site and the second *Bam*HI site of the proximal copy, the conserved two *Sal*I sites suggest the same structure. These results are essentially consistent with PAYANT *et al.* (1988). On

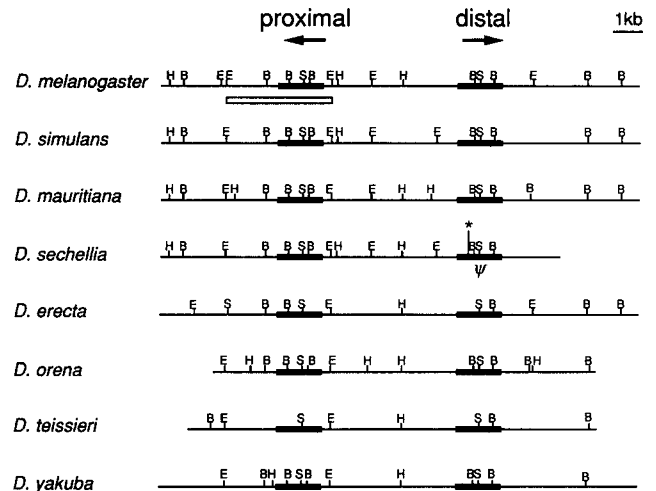

FIGURE 1.—Restriction endonuclease cleavage maps of the *Amy* regions of the eight species belonging to the *D. melanogaster* species subgroup. The duplicated structure genes are represented by shaded boxes and each copy is divergently transcribed, as shown by the arrows. A 3.8-kb *Eco*RI fragment, which was used as a probe in molecular cloning, is shown as an open box below the map of *D. melanogaster.* The restriction sites are designated as follows: B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; S, *Sal*I. * on the map of *D. sechellia* means a 35-bp deletion and a 4-bp insertion that causes shifting of its reading frame and produces a novel termination codon, therefore, the distal copy of *Amy* in *D. sechellia* does not encode a functional protein (ψ).

the other hand, there were a number of divergent sites in the flanking regions. LANGLEY *et al.* (1988) reported several large insertions in the flanking region of *Amy* in *D. melanogaster.* In this study, however, we could not find any large length differences in this region among these closely related species.

The nucleotide sequence of the coding region: The nucleotide sequences of the *Amy* loci in *D. simulans, D. mauritiana, D. sechellia, D. erecta, D. orena, D. teissieri* and *D. yakuba* were determined. There was a single exon (1482 bp) with no introns in all the sibling species studied here, except for the distal copy of *D. sechellia.* As shown in Figure 2, this copy contained a 35-bp deletion and a 4-bp insertion within the coding sequence. Since this deletion shifted the reading frame and produced a termination codon 30 bp downstream from the deletion, the distal copy of *D. sechellia* is considered to be a pseudogene. Since the insertion forms a part of a novel duplication of 10 bp, this insertion might have arisen through the duplication process. The nucleotide differences among these eight species are listed in Table 1. The data on the TN329 strain of OKUYAMA and YAMAZAKI (1988) were used as representative of *D. melanogaster.* Excluding the deletion and insertion found in *D. sechellia,* there are 150 sites with nucleotide substitutions, including eight sites with multiple substitutions. Within each species the duplicated *Amy* gene copies are strongly similar. In particular, the similarities were extremely strong within *D. erecta, D. orena* and *D. yakuba.*

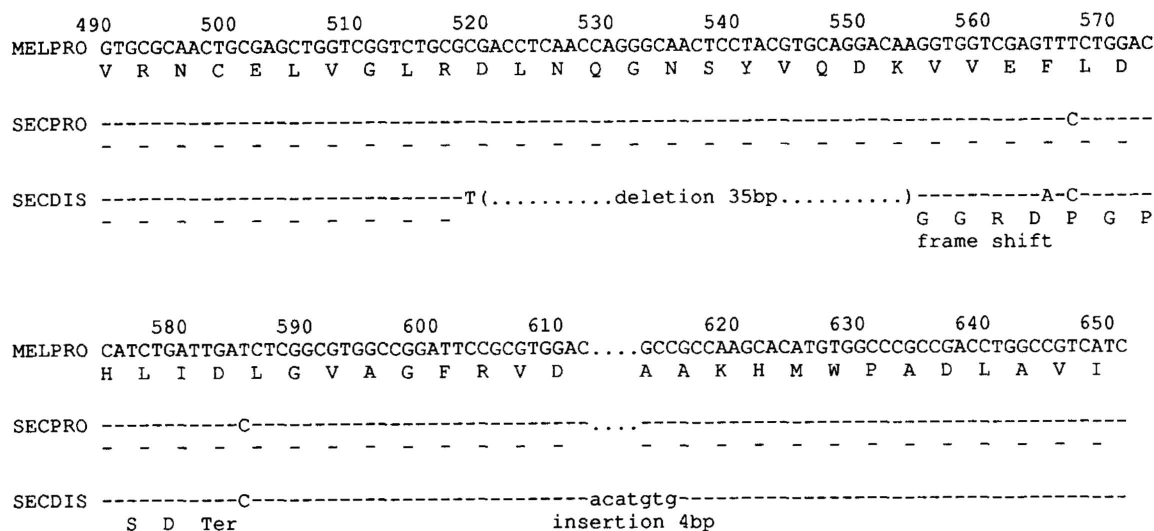

FIGURE 2.—A 32-bp deletion and 4-bp insertion were found at position 520–554 and at position 612, respectively, in the distal copy of *D. sechellia*. The predicted amino acid sequences are also shown in one letter symbols below the nucleotide sequences. The underlined areas show a potential duplication of 10 bp. MELPRO, *D. melanogaster* proximal; SECPRO, *D. sechellia* proximal; SECDIS, *D. sechellia* distal.

In *D. erecta*, there were three synonymous and only one nonsynonymous substitutions between the two *Amy* copies. The nonsynonymous substitution is located at position 40 in the 1482 nucleotide sequence. It was predicted that the first 18 amino acid residues from the N-termini would be cut off after translation, and the mature protein would then be composed of the remaining 476 amino acid residues (BOER and HICKEY 1986). The mature proteins encoded by the two copies of *D. erecta* are therefore considered to be identical. In *D. orena*, there were two synonymous and only one nonsynonymous substitution between the two *Amy* copies. This nonsynonymous substitution is located at position 1089, and the two copies in this species encoded polypeptides that differ by one amino acid site. In *D. yakuba*, there is only one synonymous substitution between the two *Amy* copies.

The phylogenetic trees from the coding region: We constructed a phylogenetic tree from the sequence data using the neighbor joining method (SAITOU and NEI 1987). Figure 3 shows the tree estimated from the sequence data of the 1482-bp coding region. The tree is striking evidence in support of the concerted evolution of duplicated *Amy* genes, because it is highly unlikely that duplication of *Amy* occurs repeatedly. The phylogeny consists of three large clusters that matched three species complexes, *i.e.*, the *melanogaster* complex, the *yakuba* complex and a third complex. This cladogram is generally consistent with the phylogenetic trees constructed from various interspecific characteristics of these species (*e.g.*, EISSES *et al.* 1979; OHNISHI *et al.* 1983; ASHBURNER *et al.* 1984; SOLIGNAC *et al.* 1986; CARIOU 1987). Our phylogeny also showed that the proximal copy of *D. simulans* is a sister sequence to the combined set of sequences including *D. mauritiana*, *D. sechellia* and the distal copy of *D. simulans*.

The nucleotide sequence of the 5′ and 3′ flanking region: Figure 4 shows the alignment of the 5′ flanking sequence of the duplicated *Amy* region. There are several conserved sequences that are potentially important for gene expression or regulation. The CAAT-like sequence: CAAAT was conserved at about −112 in the proximal copies, whereas it exists at about −100 in the distal copies. The TATA box: TATATAA was conserved at about −60 except in the proximal copies (TATAAAA) and the distal copies (TTTATAA) of *D. teissieri* and *D. yakuba*. Their locations were not accurately conserved because of small deletions and insertions. Putative transcription initiation sites were different between the proximal copies (ACCAG except for AACAT in the proximal copy of *D. orena*) and the distal copies (ATCAG), although their locations were accurately conserved. Figure 5 shows the alignment of the 3′ flanking sequence of the duplicated *Amy* region. The termination codon (TAA) was conserved in all sequences except in the proximal copy of *D. sechellia* (TGA). The putative polyadenylation signal (AATATA) was also conserved in all sequences except for the distal copies of *D. teissieri* (AATGTA) and *D. yakuba* (AATTTA). There were also a number of nucleotide substitutions, deletions and insertions that were either proximal-specific or distal-specific.

The phylogenetic trees from the flanking sequences: We estimated the nucleotide divergence of the 5′ and 3′ flanking sequences of *Amy* among the species. The alignments of these sequences were performed with the maximum matching method using a computer package, GENETYX, and all deletions and insertions were eliminated in this analysis. Since no appropriate sequence data were available for rooting the tree, we constructed unrooted trees by the neighbor joining method.

TABLE 1

Nucleotide substitutions in the coding region of *Amy* in the eight sibling species

		MEL		SIM		MAU		SEC		ERE		ORE		TEI		YAK		Repl. or	Equal or not
	Consensus	pro	dis	pro	dis	pro	dis	pro	dis	pro	dis	pro	dis	pro	dis	pro	dis	syn.	equal
12	C	—	—	T	—	T	—	T	A	—	—	—	—	—	—	—	—	Syn.	N.E.
19	A	—	—	—	C	—	C	—	C	—	—	—	—	—	—	—	—	Repl.	N.E.
21	A	—	—	T	—	—	—	—	—	T	T	T	T	—	—	—	—	Syn.	N.E.
36	C	—	—	—	—	G	—	—	—	—	—	—	—	—	T	T	T	Syn.	N.E.
37	C	—	—	—	—	—	—	—	—	—	—	—	—	—	—	T	T	Syn.	E.
40	G	—	—	—	—	—	—	—	—	—	C	—	—	—	—	—	—	Repl.	N.E.
57	A	—	—	—	—	—	—	—	—	G	G	—	—	G	G	—	—	Syn.	E.
61	G	—	—	—	—	—	—	—	—	—	—	A	A	—	—	—	—	Repl.	E.
72	C	—	—	—	—	—	—	T	—	—	—	—	—	—	—	—	—	Syn.	N.E.
75	C	A	—	—	—	—	—	T	—	—	—	—	—	—	—	—	—	Syn.	N.E.
84	T	—	—	—	—	—	—	—	—	C	C	C	C	C	C	C	C	Syn.	E.
96	C	—	—	—	—	—	T	—	—	G	G	—	—	G	G	—	—	Syn.	N.E.
126	C	—	—	A	—	—	—	—	—	—	—	—	—	—	—	—	—	Syn.	N.E.
129	T	—	—	—	—	—	—	—	—	—	—	—	—	C	C	C	C	Syn.	E.
141	G	A	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	Syn.	N.E.
150	G	T	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	Syn.	N.E.
164	T	A	A	—	—	—	—	—	—	—	—	—	—	—	—	—	—	Repl.	E.
171	C	T	T	T	—	T	T	—	—	—	—	T	T	—	—	—	—	Syn.	N.E.
174	C	T	T	—	—	—	—	—	—	—	—	—	—	—	—	—	—	Syn.	E.
183	C	—	—	G	G	—	—	G	G	—	—	—	—	T	—	—	—	Syn.	E./N.E.
186	T	—	—	—	—	—	—	—	—	—	—	—	—	C	C	C	C	Syn.	E.
192	C	—	—	—	—	T	—	—	—	—	—	—	—	—	—	—	—	Syn.	N.E.
201	C	—	—	G	—	—	—	—	—	—	—	—	—	—	—	—	—	Syn.	N.E.
216	C	—	—	—	—	—	—	T	T	—	—	—	—	—	—	—	—	Syn.	E.
219	C	—	—	—	—	—	—	—	—	—	—	T	T	—	—	—	—	Syn.	E.
228	A	—	—	—	—	—	—	—	—	G	G	G	G	G	G	G	G	Syn.	E.
231	T	—	—	—	—	—	—	—	—	C	C	C	C	C	C	C	C	Syn.	E.
237	G	A	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	Syn.	N.E.
243	C	—	—	—	—	—	—	—	—	T	T	T	T	—	—	T	T	Syn.	E.
257	A	—	—	—	—	—	—	—	—	—	—	T	T	—	—	—	—	Repl.	E.
258	G	—	—	—	—	—	—	—	—	C	C	—	—	—	—	—	—	Repl.	E.
276	G	A	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	Syn.	N.E.
277	C	G	G	—	—	—	—	—	—	—	—	—	—	—	—	—	—	Repl.	E.
297	C	—	—	—	T	—	—	—	—	—	—	—	—	—	—	G	G	Syn.	N.E.
299	A	—	—	—	—	—	—	—	—	G	G	—	—	—	—	—	—	Repl.	E.
303	A	C	C	C	T	T	T	T	T	C	C	—	—	—	—	—	—	Syn.	N.E.
312	C	—	—	—	T	T	T	T	T	—	—	—	—	—	—	—	—	Syn.	N.E.
315	C	—	—	—	G	G	G	G	G	—	—	—	—	—	—	—	—	Syn.	N.E.
326	T	C	C	C	—	—	—	—	—	—	—	—	—	C	C	C	C	Repl.	N.E.
336	C	—	—	—	T	T	T	T	T	—	—	—	—	—	—	—	—	Syn.	N.E.
340	G	—	—	—	A	A	A	A	A	—	—	A	A	—	—	—	—	Repl.	N.E.

TABLE 1

Continued

	Consensus	MEL pro	MEL dis	SIM pro	SIM dis	MAU pro	MAU dis	SEC pro	SEC dis	ERE pro	ERE dis	ORE pro	ORE dis	TEI pro	TEI dis	YAK pro	YAK dis	Repl. or syn.	Equal or not equal
360	C	—	—	T	—	—	—	—	—	—	—	—	—	—	—	—	—	Syn.	N.E.
378	C	—	—	—	—	—	—	—	—	A	A	A	A	A	A	A	A	Syn.	E.
381	G	T	T	—	T	T	T	—	—	—	C	—	—	—	—	—	—	Syn.	N.E.
393	C	—	—	—	T	—	—	—	—	—	—	—	—	—	—	—	—	Syn.	N.E.
407	G	—	—	—	—	—	—	—	—	—	—	—	—	C	C	C	C	Repl.	E.
416	A	—	—	—	—	—	—	—	—	—	—	—	—	T	T	T	T	Repl.	E.
417	C	T	T	—	T	T	T	T	T	—	—	—	—	—	—	—	—	Syn.	N.E.
420	C	—	—	—	—	—	—	T	T	—	—	—	—	—	—	—	—	Syn.	E.
432	C	—	—	—	—	—	—	T	T	—	—	—	—	—	—	—	—	Syn.	E.
438	G	—	—	—	—	C	—	—	—	—	—	—	—	—	—	—	—	Syn.	N.E.
456	C	—	—	—	—	—	—	T	T	—	—	—	—	—	—	—	—	Syn.	E.
460	G	—	—	—	—	—	—	—	—	—	—	—	—	T	T	T	T	Repl.	E.
461	C	—	—	—	—	—	—	—	—	G	G	—	—	—	—	—	—	Repl.	E.
466	A	—	—	—	—	—	—	—	—	—	—	—	—	G	G	—	—	Repl.	E.
467	A	G	G	—	—	—	—	—	—	G	G	—	—	—	—	—	—	Repl.	E.
468	C	—	—	—	—	—	—	—	—	—	—	—	—	G	G	—	—	Syn.	E.
476	A	—	—	—	—	—	—	—	—	—	—	—	—	—	G	—	—	Repl.	N.E.
487	C	G	G	—	—	—	—	—	—	—	—	—	—	—	—	—	—	Repl.	E.
510	C	—	—	—	G	G	—	—	—	—	—	—	—	—	—	—	—	Syn.	N.E.
519	C	—	—	—	T	—	—	—	T	—	—	—	—	—	—	—	—	Syn.	N.E.
520	G	—	—	—	—	—	—	—	*	—	—	—	—	—	—	—	—		
521	A	—	—	—	—	—	—	—	*	—	—	—	—	—	—	—	—		
522	C	—	—	—	—	—	—	—	*	—	—	—	—	—	—	—	—		
523	C	—	—	—	—	—	—	—	*	—	—	—	—	—	—	—	—		
524	T	—	—	—	—	—	—	—	*	—	—	—	—	—	—	—	—		
525	C	—	—	—	—	—	—	—	*	—	—	—	—	—	—	—	—		
526	A	—	—	—	—	—	—	—	*	—	—	—	—	—	—	—	—		
527	A	—	—	—	—	—	—	—	*	—	—	—	—	—	—	—	—		
528	C	—	—	—	—	—	—	—	*	—	—	—	—	—	—	—	—		
529	C	—	—	—	—	—	—	—	*	—	—	—	—	—	—	—	—		
530	A	—	—	—	—	—	—	—	*	—	—	—	—	—	—	—	—		
531	G	—	—	—	—	—	—	—	*	—	—	—	—	—	—	—	—		
532	G	—	—	—	—	—	—	—	*	—	—	—	—	—	—	—	—		
533	G	—	—	—	—	—	—	—	*	—	—	—	—	—	—	—	—		
534	C	—	—	—	—	—	—	—	*	—	—	—	—	—	—	—	—		
535	A	—	—	—	—	—	—	—	*	—	—	—	—	—	—	—	—		
536	A	—	—	—	—	—	—	—	*	—	—	—	—	—	—	—	—		
537	C	—	—	—	—	—	—	—	*	—	—	—	—	—	—	—	—		
538	T	—	—	—	—	—	—	—	*	—	—	—	—	—	—	—	—		
539	C	—	—	—	—	—	—	—	*	—	—	—	—	—	—	—	—		
540	C	—	—	T	—	—	—	—	*	—	—	—	—	—	—	—	—	Syn.	N.E.
541	T	—	—	—	—	—	—	—	*	—	—	—	—	—	—	—	—		

TABLE 1

Continued

	Consensus	MEL		SIM		MAU		SEC		ERE		ORE		TEI		YAK		Repl. or syn.	Equal or not equal
		pro	dis	pro	dis	pro	dis	pro	dis	pro	dis	pro	dis	pro	dis	pro	dis		
542	A	—	—	—	—	—	—	—	*	—	—	—	—	G	G	—	—	Repl.	E.
543	C	—	—	—	—	—	—	—	*	—	—	—	—	G	G	—	—	Syn.	E.
544	G	—	—	—	—	—	—	—	*	—	—	—	—	—	—	—	—		
545	T	—	—	—	—	—	—	—	*	—	—	—	—	—	—	—	—		
546	G	—	—	—	—	—	—	—	*	—	—	—	—	—	—	—	—		
547	C	—	—	—	—	—	—	—	*	—	—	—	—	—	—	—	—		
548	A	—	—	—	—	—	—	—	*	G	G	G	G	—	—	—	—	Repl.	E.
549	G	—	—	—	—	—	—	—	*	—	—	—	—	—	—	—	—		
550	G	—	—	—	—	—	—	—	*	—	—	—	—	—	—	—	—		
551	A	—	—	—	—	—	—	—	*	—	—	—	—	—	—	—	—		
552	C	—	—	—	—	—	—	—	*	—	—	—	—	—	—	—	—		
553	A	—	—	—	—	—	—	—	*	—	—	—	—	—	—	—	—		
554	A	—	—	—	—	—	—	—	*	—	—	—	—	—	—	—	—		
565	T	—	—	—	—	—	—	—	A	—	—	—	—	—	—	—	—	Repl.	N.E.
567	C	T	T	—	—	—	—	—	—	—	—	—	—	—	—	—	—	Syn.	E.
582	T	—	—	—	—	—	—	—	—	—	—	—	—	C	—	—	—	Syn.	N.E.
585	C	T	T	—	—	—	—	—	—	—	—	—	—	—	—	—	—	Syn.	E.
588	C	—	—	—	—	T	—	—	—	—	—	—	—	—	—	—	—	Syn.	N.E.
594	G	—	—	—	—	A	—	—	—	—	—	—	—	—	—	—	—	Syn.	N.E.
606	C	—	—	—	—	—	—	—	—	G	—	—	—	—	—	—	—	Syn.	N.E.
612	*	*	*	*	*	*	*	*	A	*	*	*	*	*	*	*	*		
	*	*	*	*	*	*	*	*	C	*	*	*	*	*	*	*	*		
	*	*	*	*	*	*	*	*	A	*	*	*	*	*	*	*	*		
	*	*	*	*	*	*	*	*	T	*	*	*	*	*	*	*	*		
614	C	—	—	—	—	—	—	—	T	—	—	—	—	—	—	—	—	Repl.	N.E.
615	C	—	—	—	—	—	—	—	G	—	—	—	—	—	—	—	—	Syn.	N.E.
645	C	—	—	A	—	—	—	—	—	—	—	—	—	—	—	—	—	Syn.	N.E.
647	T	—	—	—	—	—	—	—	—	—	—	C	C	—	—	—	—	Repl.	E.
660	C	—	—	—	—	—	T	—	—	—	—	—	—	—	—	—	—	Syn.	N.E.
668	A	—	—	—	C	—	—	—	—	—	—	—	—	C	C	C	C	Repl.	N.E.
672	G	—	—	—	A	—	—	—	—	—	—	—	—	—	—	—	—	Syn.	N.E.
681	C	—	—	—	—	—	—	—	—	—	—	—	—	G	G	—	—	Repl.	E.
691	G	—	—	—	—	—	—	—	—	—	—	—	—	A	A	A	A	Repl.	E.
692	C	—	—	—	—	—	—	—	—	—	—	—	—	A	A	A	A	Repl.	E.
696	G	—	—	—	—	—	—	—	—	—	—	—	—	—	—	C	C	Syn.	E.
699	A	—	—	—	—	—	—	—	—	C	C	—	—	C	C	—	—	Syn.	E.
704	A	—	—	—	—	—	—	—	—	—	—	G	G	—	—	—	—	Repl.	E.
708	A	—	—	—	—	—	—	—	—	C	C	C	C	C	C	C	C	Syn.	E.
717	C	—	—	—	—	—	—	—	—	—	—	—	—	G	G	—	—	Syn.	E.
718	C	—	—	—	—	—	—	—	—	—	—	—	—	—	—	G	G	Repl.	E.

TABLE 1

Continued

	Consensus	MEL		SIM		MAU		SEC		ERE		ORE		TEI		YAK		Repl. or syn.	Equal or not equal
		pro	dis	pro	dis	pro	dis	pro	dis	pro	dis	pro	dis	pro	dis	pro	dis		
720	G	—	—	—	—	A	—	—	—	—	—	—	—	—	—	—	—	Syn.	N.E.
738	T	C	C	C	C	A	A	C	C	—	—	—	—	—	—	—	—	Syn.	E.
759	C	—	—	—	—	—	—	—	—	—	—	T	T	—	—	—	—	Syn.	E.
770	G	—	—	C	—	—	—	—	—	—	—	—	—	—	—	—	—	Repl.	N.E.
771	A	—	—	—	—	—	—	—	—	C	C	C	C	—	—	C	C	Syn.	E.
772	C	—	—	—	—	—	—	G	G	—	—	—	—	—	—	—	—	Repl.	E.
777	C	—	—	—	—	—	—	T	T	—	—	—	—	—	—	—	—	Syn.	E.
781	A	—	—	—	—	—	—	—	—	—	—	—	—	G	G	G	G	Repl.	E.
786	C	—	—	—	—	—	—	—	—	—	—	T	—	—	—	—	—	Syn.	N.E.
818	T	—	—	—	—	—	—	—	—	C	C	C	C	—	—	—	—	Repl.	E.
819	G	C	C	—	—	—	—	—	—	C	C	C	C	—	—	—	—	Syn.	E.
835	C	—	—	—	—	—	—	—	—	—	—	A	A	—	—	—	—	Repl.	E.
837	G	—	—	—	A	—	—	—	—	—	—	—	—	—	—	—	—	Syn.	N.E.
846	C	—	T	—	—	—	—	—	—	—	—	—	—	—	—	—	—	Syn.	N.E.
850	A	—	—	—	—	—	—	—	—	C	C	—	—	—	—	—	—	Repl.	E.
851	C	—	—	—	—	—	—	—	—	G	G	G	G	—	—	—	—	Repl.	E.
877	G	—	—	—	—	—	—	—	C	—	—	—	—	—	—	—	—	Repl.	N.E.
879	C	T	T	—	—	—	—	—	—	T	—	—	—	—	T	—	—	Syn.	N.E.
900	C	A	A	—	—	—	—	—	—	—	—	—	—	—	—	—	—	Syn.	E.
921	C	T	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	Syn.	N.E.
933	C	T	T	—	—	—	—	—	T	—	—	—	—	—	—	—	—	Syn.	N.E.
942	A	—	—	T	—	—	—	—	—	—	—	—	—	C	C	—	—	Syn.	N.E.
954	G	T	T	—	—	—	T	—	—	—	—	—	—	—	—	—	—	Syn.	N.E.
961	T	—	—	—	—	—	—	—	—	—	—	C	C	—	—	—	—	Repl.	E.
974	A	—	—	—	—	—	—	—	—	—	—	G	G	—	—	—	—	Repl.	E.
991	T	—	—	—	—	—	—	—	—	—	—	A	A	—	—	—	—	Repl.	E.
1014	C	—	A	—	—	—	—	—	—	—	—	—	—	—	—	—	—	Syn.	N.E.
1032	G	—	—	—	—	—	—	—	—	—	—	—	—	—	—	C	—	Syn.	N.E.
1051	T	A	A	—	—	—	—	—	—	—	—	—	—	—	—	—	—	Repl.	E.
1054	G	—	—	—	—	—	—	—	—	—	—	A	A	—	—	—	—	Repl.	E.
1059	G	C	C	—	—	—	—	—	—	—	—	—	—	—	—	—	—	Syn.	E.
1062	C	T	T	—	—	—	—	—	—	—	—	—	—	—	—	—	—	Syn.	E.
1089	C	—	—	—	—	—	—	—	—	—	—	—	G	—	—	—	—	Repl.	N.E.
1105	G	A	A	—	—	—	—	—	—	—	—	—	—	—	—	—	—	Repl.	E.
1113	T	—	—	—	—	—	—	—	—	C	C	—	—	C	C	—	—	Syn.	E.
1120	A	—	—	—	—	—	—	—	—	C	C	—	—	—	—	—	—	Repl.	E.
1121	A	—	—	—	—	—	—	—	—	—	—	G	G	—	—	—	—	Repl.	E.
1146	C	T	T	—	—	—	—	—	—	—	—	—	—	—	—	—	—	Syn.	E.
1168	T	—	—	—	—	—	—	—	—	—	—	A	A	—	—	—	—	Repl.	E.
1188	G	A	A	—	—	—	—	—	—	—	—	—	T	—	—	—	—	Syn.	E./N.E.
1192	G	A	A	—	—	—	—	—	—	—	—	—	—	—	—	—	—	Repl.	E.
1202	C	—	—	—	—	—	—	T	—	—	—	—	—	—	—	—	—	Repl.	N.E.

TABLE 1

Continued

	Consensus	MEL		SIM		MAU		SEC		ERE		ORE		TEI		YAK		Repl. or syn.	Equal or not equal
		pro	dis	pro	dis	pro	dis	pro	dis	pro	dis	pro	dis	pro	dis	pro	dis		
1208	C	A	A	—	—	—	—	—	—	—	—	—	—	—	—	—	—	Repl.	E.
1230	C	—	—	—	—	—	—	—	T	—	—	—	—	—	—	—	—	Syn.	N.E.
1246	T	—	—	—	—	—	—	—	—	G	G	G	G	G	G	G	G	Repl.	E.
1257	A	—	—	—	—	—	—	—	—	C	C	C	C	C	—	C	C	Syn.	N.E.
1259	G	—	—	—	—	—	—	—	C	—	—	—	—	—	—	—	—	Repl.	N.E.
1278	C	—	—	—	T	—	—	—	—	—	—	—	—	—	—	—	—	Syn.	N.E.
1344	C	—	—	—	T	—	—	—	—	—	—	—	—	—	—	—	—	Syn.	N.E.
1353	C	—	T	—	—	—	—	—	—	—	—	—	—	—	—	—	—	Syn.	N.E.
1371	C	T	—	—	—	—	—	T	T	—	—	—	—	—	—	—	—	Syn.	N.E.
1380	C	—	—	T	—	—	—	—	—	—	—	—	—	—	—	—	—	Syn.	N.E.
1388	A	—	—	—	—	—	—	—	—	G	G	—	—	—	—	—	—	Repl.	E.
1397	G	C	C		—	—	—	—	—	—	—	—	—	—	—	—	—	Repl.	E.
1404	A	—	—	—	—	—	—	—	—	—	—	—	—	C	—	—	—	Syn.	N.E.
1416	G	—	—	—	T	—	—	—	—	—	—	—	—	—	—	—	—	Syn.	N.E.
1426	T	A	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	Repl.	N.E.
1427	A	—	—	—	—	—	—	—	—	C	C	C	C	—	—	—	—	Repl.	E.
1429	A	—	—	—	—	—	—	—	—	G	G	G	G	G	G	C	C	Repl.	E.
1431	T	—	—	—	—	—	—	—	—	—	—	—	—	G	G	G	G	Syn.	E.
1435	A	—	—	—	—	—	—	—	—	C	C	—	—	—	—	—	—	Repl.	E.
1436	G	—	—	—	—	—	—	—	C	—	—	—	—	—	—	—	—	Repl.	N.E.
1446	C	—	—	—	—	—	—	—	T	—	—	T	T	—	—	—	—	Syn.	N.E.
1452	A	—	—	—	—	—	—	—	—	—	—	C	C	—	—	—	—	Syn.	E.
1455	G	—	—	—	—	—	—	—	—	C	C	C	C	—	—	—	—	Syn.	E.
1461	C	—	—	—	—	—	T	—	—	—	—	—	—	—	—	—	—	Syn.	N.E.
1464	C	T	T	—	—	—	—	—	—	—	—	—	—	—	—	—	—	Syn.	E.
1467	C	—	—	—	T	—	—	—	—	—	—	—	—	—	—	—	—	Syn.	N.E.
1484	A	—	—	—	—	—	—	G	—	—	—	—	—	—	—	—	—	Syn.	N.E.

When we found one or more species that carry different nucleotides at the same position of two *Amy* copies, we classified such sites as not equal. Therefore, equal means that identical nucleotides were found at the same position between the two copies of each species. Any deletion and insertion found in the distal copy of *D. sechellia* was ignored in these classifications. The nucleotides identical to the consensus are shown by dashes; deletions are shown by asterisks.

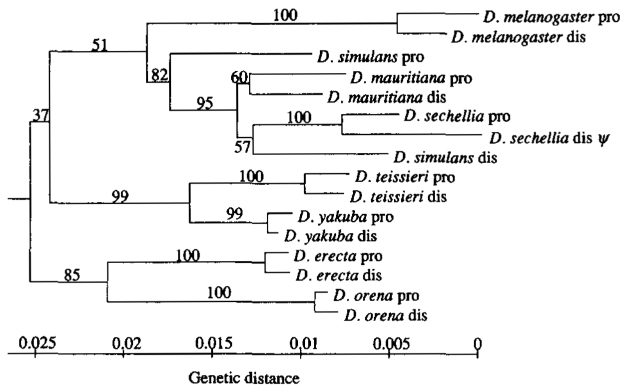

FIGURE 3.—A phylogenetic tree constructed by the neighbor joining method from the nucleotide sequence data of the coding regions (1482 bp) in *D. melanogaster, D. simulans, D. mauritiana, D. sechellia, D. erecta, D. orena, D. teissieri* and *D. yakuba*. The tree was rooted using the data of *D. pseudoobscura* presented by BROWN *et al.* (1990). The numbers above branches indicate the number of occurrences among the trees constructed from 100 bootstrapped data sets. pro, proximal copy; dis, distal copy.

Figure 6, A and B, shows the trees estimated from the 5′ and 3′ flanking sequences, respectively. These trees differ markedly from the trees estimated from the coding sequences. Instead of the proximal and distal loci of each species being similar, we can clearly recognize higher similarities within each locus (proximal or distal) between species than between the two loci within each species, except for *D. erecta* (lower left of Figure 6B). In other words, there are generally proximal-specific and distal-specific sequences in the flanking regions of the *Amy* genes, at least in the seven closely related species studied here, with the exception of the distal region of *D. erecta*. Hence, the flanking regions of the proximal copies and those of the distal copies appear to have evolved independently without being influenced by the concerted evolution of the coding region.

Chi-square test for synonymous and nonsynonymous substitutions: We classified all nucleotide substitutions between the coding sequences into either synonymous or replacement groups and also classified them into equal or not equal groups. The equal group was defined as substitutions between species at sites at which, within each of the eight species compared, both duplicated loci share an identical nucleotide. The rest of the substitutions were classified into the not equal group (see Table 1). By this definition substitutions are not equal, if the corresponding sites of duplicated loci differ in any of the eight species compared. All substitutions therefore fall into four classes. *D. sechellia* was omitted

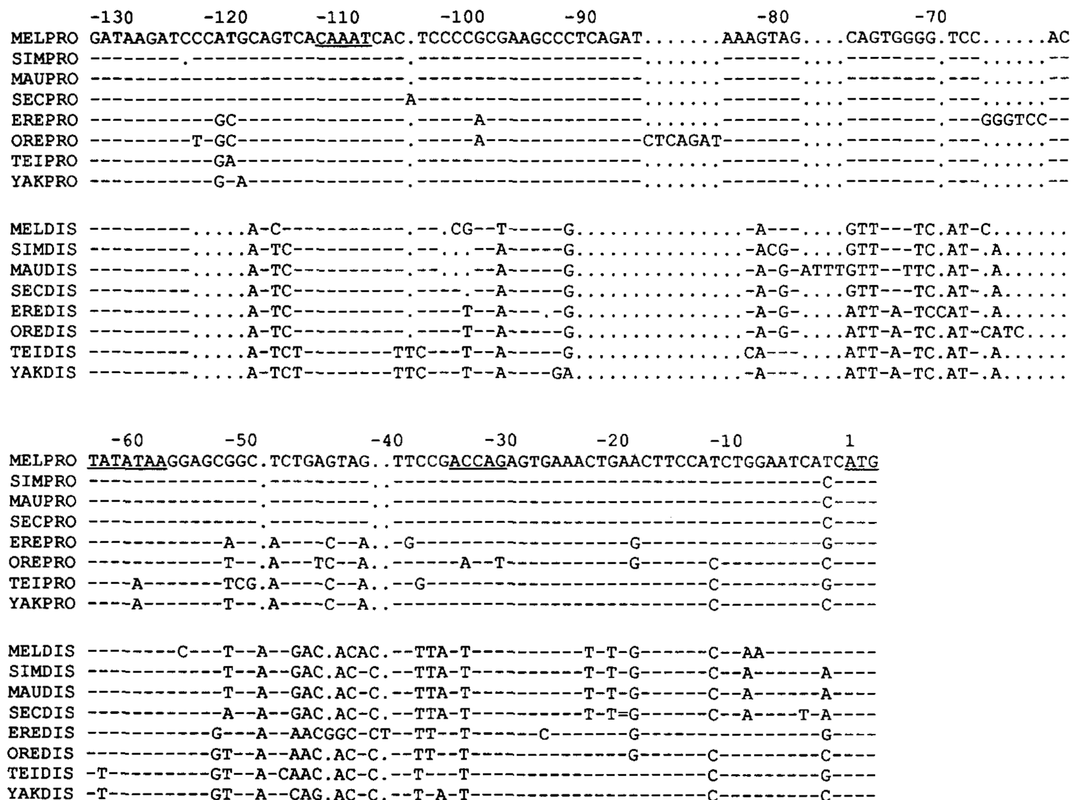

FIGURE 4.—The nucleotide sequences of 5′ flanking regions of the duplicated *Amy* loci in the eight species belonging to the *D. melanogaster* species subgroup. CAAT box, TATA box, transcription initiation site (ACCAG for proximal, while ATCAG for distal) and initiation codon (ATG) are underlined. MEL, *D. melanogaster;* SIM, *D. simulans;* MAU, *D. mauritiana;* SEC, *D. sechellia;* ERE, *D. erecta;* ORE, *D. orena;* TEI, *D. teissieri;* YAK, *D. yakuba*. Identical nucleotides and deletions are designated as dashes (-) and dots (·), respectively.

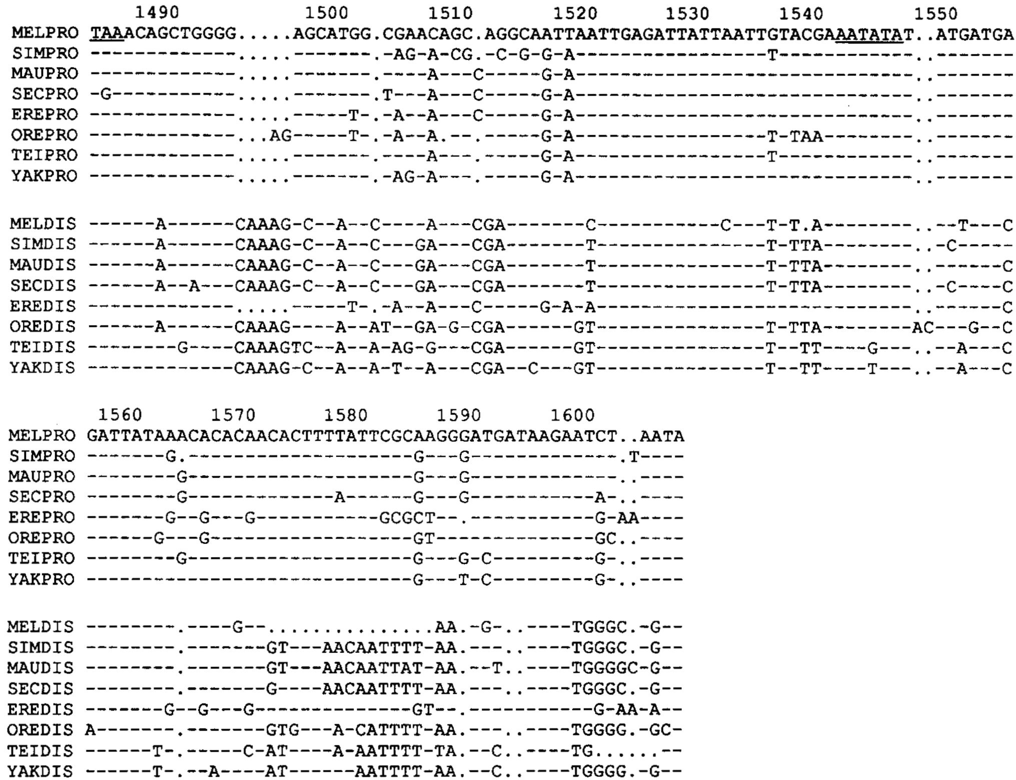

FIGURE 5.—The nucleotide sequences of 3′ flanking regions of the duplicated *Amy* loci in the eight species belonging to the *D. melanogaster* species subgroup. The termination codon (TAA) and putative polyadenylation signal (AATATA) are underlined. The symbols and abbreviations are the same as those used in Figure 4.

in the following analysis because one of the two genes is not functional. Three multiple substitution sites (position 96, 183 and 1188) were classified as not equal by our definition. If molecular evolution occurs only by fixation of neutral mutations (KIMURA and OHTA 1971), we expect that the relative fixation probabilities of synonymous and replacement changes will be the same in comparisons between any stages of phylogenetic trees, including those within and between the species or those between the equal and the not equal classes. To test

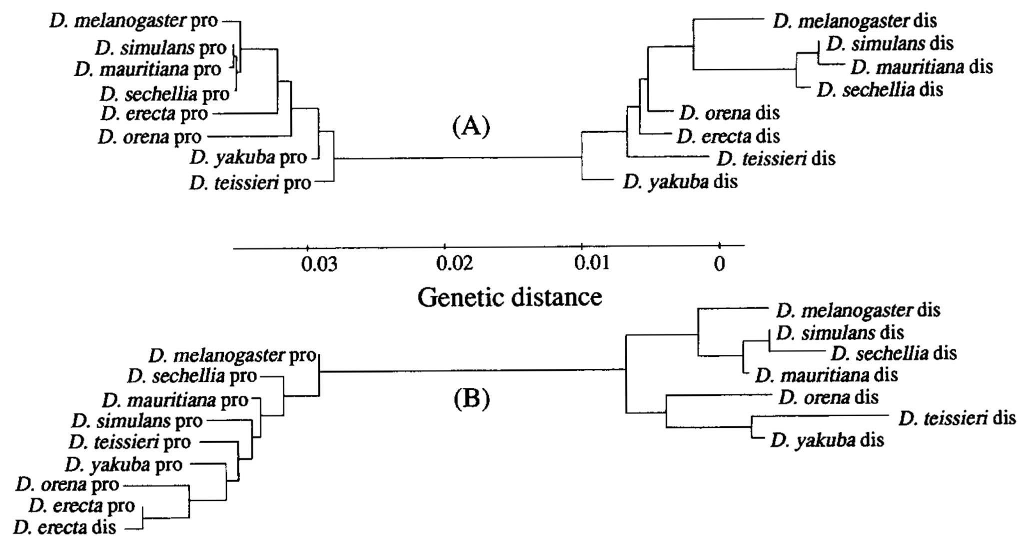

FIGURE 6.—(A) The phylogenetic relationships from the nucleotide sequence data of 5′ flanking regions (≤200 bp). There are two large clusters. One is for the proximal genes (left half) while the other is for the distal genes (right half). (B) The phylogenetic relationships from the nucleotide sequence data of 3′ flanking regions (≤200 bp). Two large clusters are observed in the same manner as in the phylogenetic relationship for the 5′ flanking region (A) except for the distal copy of *D. erecta*, which demonstrated a strong homology with the proximal copy of *D. erecta*.

TABLE 2

2 × 2 contingency table on nucleotide substitutions among the eight sibling species

	Repl.	Syn.	Total
Equal	43	32	75
Not equal	9	57	66
Total	52	89	141

$\chi^2 = 28.8$, with $P < 0.1\%$. The data from *D. sechellia* were excluded in this analysis, since the distal copy of *D. sechellia* was a pseudogene.

this null hypothesis, we performed a chi-square test of independence among the four classes of nucleotide substitution. This chi-square test is similar to the G test in McDONALD and KREITMAN (1991) but differs in that we included in the not equal substitution differences between the duplicated genes within one species. Table 2 shows the 2 × 2 contingency table. The chi-square value was 28.8 ($P < 0.1\%$) and thus the null hypothesis is rejected. The simple neutral theory does not suffice to explain the pattern of nucleotide substitutions in the *Amy* regions among the sibling species belonging to the *D. melanogaster* species subgroup. A large excess of equal replacement compared with not equal replacement (43 *vs.* 9 while the ratio of equal with not equal as a whole is 75 *vs.* 66) clearly indicates that amino acid changing substitutions are predominant in the class in which the equivalent sites of duplicated loci are replaced by the same base pairs. Namely, substitutions exclusively between species are likely to be not neutral with respect to fitness.

DISCUSSION

Concerted evolution of the coding region in contrast to the differential evolution of flanking region: We have demonstrated the existence of concerted evolution in the coding region of the duplicated *Amy* loci in the *D. melanogaster* species subgroup (Figure 3). We also showed differentiated evolution of the flanking region between the proximal and distal loci (see Figures 4–6). There are two different mechanisms to explain the above observations. The first hypothesis is unequal crossing over through a loop formation by asymmetrical pairing. When asymmetrical pairing occurs between duplicated genes on the same chromosome, a single crossover generates an inversion of the sequence intervening between the duplication (SCHWARTZ and DOANE 1989), whereas a double crossover conserves the order of the intervening sequence. When asymmetrical pairing occurs between the duplicated genes on different homologous chromosomes, a single crossover generates dicentric and acentric chromosomes that are then selected out. A double crossover within one locus conserves the order of the intervening sequence, while a double crossover including both loci generates an inversion of the intervening sequence. An alternative hypothesis is gene conversion through heteroduplex formation and mismatch repair. There are three major models: the HOLLIDAY model (HOLLIDAY 1964), the MESELSON-RADDING model (MESELSON and RADDING 1975) and the double-strand-break repair model (SZOSTAK *et al.* 1983). In any gene conversion model, an inversion of the sequence intervening between the duplication occurs with a 50% probability.

Thus, in either mechanism, be it unequal crossing over or gene conversion, inversion of the intervening sequence should be produced at a high frequency. However, such inversions are actually quite rare in natural populations [0.007(47/6697), unpublished results by S. H. SUNG *et al.*]. In fact, such inversions have been reported in two cases from *D. melanogaster.* One was a null mutant of amylase (OKUYAMA and YAMAZAKI 1988), while the other showed normal amylase activity (LANGLEY *et al.* 1988). We confirm here that the order of the intervening sequence between the duplication is conserved throughout the *D. melanogaster* species subgroup. We also show that the nucleotide sequences in the 5′ and 3′ flanking regions are conserved within each locus (proximal and distal) but diverged between the two loci (Figures 4 and 5). These data suggest that within each locus differential selection has played a role in the *Amy* flanking region. The flanking region is considered to be responsible for *Amy* gene expression. MATSUO and YAMAZAKI (1986) found evidence of differentiated regulation between duplicated *Amy* loci in *D. melanogaster.* From the above evidence we infer that such differentiated regulation may be common in this species subgroup.

Synonymous *vs.* amino acid changing substitutions: By looking at the sequence data of amylase coding regions, we noticed that nucleotide substitutions observed between species tend to be amino acid changing. If substitutions occurring during or after speciation are adaptive to the environments in which the species lives, they should be replacement changes. Some homogenizing mechanisms must also exist to ensure that both corresponding sites of the duplicated loci would be quickly substituted by the same base pairs. We therefore examined the possibility that amino acid changing substitution is more likely to be observed between species (when there is no nucleotide difference between the proximal and distal copies of the same species), namely the equal class, compared with one observed within species (when there is a nucleotide difference between two duplicated copies of at least one species), namely the not equal class. We performed a chi-square test on our sequence data to examine the independence between the types of nucleotide substitution (synonymous or replacement) and the types of evolutionary processes (equal or not equal). As shown in Table 2, we found a significant interaction between the above two factors. When we added 16 nucleotide sequence

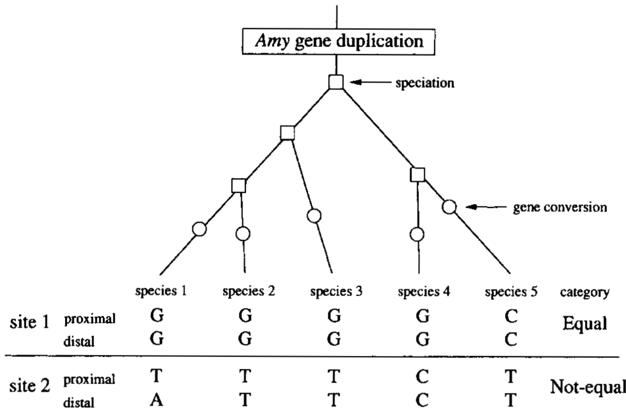

FIGURE 7.—A model phylogenetic tree showing the timing of base substitutions and homogenization events (gene conversions) leading to equal or not equal class. Equal class may arise by base substitutions occurring in any period before homogenizing events. In this category all species carry the same substitutions in both proximal and distal genes. Not equal class is produced by substitutions occurring after homogenization events. At least one of the species has different substitution in either proximal or distal gene in this group.

data of eight strains of *D. melanogaster* presented in INOMATA *et al.* (1994) to our analysis except *Amy*^{null} strain, the result remains unchanged ($\chi^2 = 47.4$, $P < 0.1\%$, contingency table not shown). In the calculation with *D. melanogaster* data, the not equal group is increased due to the presence of a large amount of genetic variability within the species between the loci, since a substitution is classified as not equal if any of the 16 sequences of *D. melanogaster* has substitutions within species (in other words, is polymorphic). When the data from *D. simulans* are excluded because of phylogenetic ambiguity (see Figure 3), the result also remains unchanged ($\chi^2 = 17.7$, $P < 0.1\%$, contingency table not shown). Note that both equal and not equal classes include the different substitutions that occurred in different clades of phylogenetical trees (Figure 7). Substitutions of the equal class include only the ones that occurred before the last homogenization event (gene conversion), but the not equal class contains substitutions occurring both before and after the homogenization event. However, the chi-square independence test is appropriate since the expected ratio of replacement to synonymous substitutions is the same between the equal and not equal classes in any clade of phylogenetic trees, if all substitutions are neutral.

Table 3 shows the number of synonymous and replacement substitutions in each of the two categories, equal and not equal, that are specific to each of the eight sibling species compared. These substitutions can be considered as the ones that occurred in the lineage of each species after divergence from other species. The test for independence was performed using the total of eight species. The chi-squre value was 26.0 (highly significant). For the equal groups, the replacement class was larger than the synonymous class in the lineages leading to *D. erecta* and *D. orena*, while a difference was not obvious in the lineages of other species. In *D. erecta* especially, all seven observed equal substitutions were replacements and no equal synonymous substitutions were found. In the species-specific substitutions in Table 3, the ratio (100% = 7/7) of replacement to synonymous substitutions in *D. erecta* is greater than the expected ratio (71%) under the assumption that all nucleotide substitutions are neutral, excluding termination codons (NEI 1975). Note that the ratio commonly observed is much lower than the neutral expectation because of selective constraints. For example, the overall ratio was 21% (10/47) in the case of polymorphism in *Amy* among nine strains of *D. melanogaster* (INOMATA *et al.* 1994). Both enzymes encoded by the duplicated *Amy* genes in *D. erecta* appear to be functional because of their complete gene structure in spite of many amino acid replacements compared with those of *D. melanogaster*. This indicates that amino acid substitution was highly accelerated in the lineage to *D. erecta*. The significant excess of equal replacements observed in both Tables 2 and 3 suggests the presence of accelerated amino acid substitution through speciation, while genetic variability between the loci within species is mainly synonymous variation. It seems that the adaptive replacement of amino acids may have occurred in the process of speciation not only in *D. erecta* but in other species as well.

There are two possible explanations for this phenomenon. One is adaptive protein evolution (MCDONALD and KREITMAN 1991; EANES *et al.* 1993), and the other is a bottleneck effect that allows fixation of slightly deleterious mutations (OHTA 1993). However, there are several pieces of experimental evidence that support the hypothesis of adaptive evolution of these substitutions but contradict the bottleneck hypothesis. First, although small population size increases the fixation probability of deleterious mutations (OHTA 1973), this is not a plausible explanation of our results, since it cannot lead to an excess of amino acid substitutions, compared with synonymous ones. The excess of the equal replacement class with two doses of amino acid change in an individual fly and presumably even more severe effects on the carriers, compared with flies with not equal replacements with a single dose of amino acid change, thus cannot easily be explained by the bottleneck hypothesis (see Tables 2 and 3). The other evidence that is contradictory to the hypothesis of a bottleneck effect is that, though all loci should be equally subject to this effect, there is no similar tendency for excess amino acid substitutions at the *Adh* locus of *D. erecta*, which is the only locus available for comparison (JEFFS, HOLMES and ASHBURNER, personal communication). Moreover, Table 3 shows that the proportion of equal replacements in *D. erecta* is greater than that in *D. orena* whose population size is considered to be extremely small (LACHAISE *et al.* 1988).

TABLE 3

Nucleotide substitutions specific to one species

	D. melanogaster[a]	*D. simulans*	*D. mauritiana*	*D. sechellia*[b]	*D. erecta*	*D. orena*	*D. teissieri*	*D. yakuba*	Total
Equal substitutions									
Synonymous	6	0	1	*	0	4	3	3	17
Replacement	6	0	0	*	7	11	3	2	29
Not equal substitutions									
Synonymous	31	16	8	9	1	2	3	1	71
Replacement	9	1	0	2	1	1	1	0	15

[a] Data from nine strains presented in INOMATA *et al.* (1993) except *Amy*^{null} strain.
[b] Only data from the proximal copy are shown.

All the above evidence suggests the hypothesis of adaptive protein evolution in the lineage of the *D. erecta* amylase and possibly also in the evolution of other species. MCDONALD and KREITMAN (1991) and EANES *et al.* (1993) came to similar conclusions from the nucleotide sequence data of the *Adh* gene and the *G6pd* gene in *Drosophila*, respectively. COUTURIER *et al.* (1985) reported that *D. melanogaster, D. erecta, D. teissieri* and *D. yakuba* exhibit a different dependence on host plants when they are sympatric. The number of plant species used as breeding sites was 7, 1, 4 and 14 for *D. melanogaster, D. erecta, D. teissieri* and *D. yakuba*, respectively. Thus *D. melanogaster* and *D. yakuba* are generalists for host plants, whereas the other two species are specialized for fewer species of host plants. In particular, *D. erecta* is a specialist, using only one genus of the Pandanaceae (screwpine Pandanus), a genus not used by other sympatric species (RIO *et al.* 1983). This specific host dependence is considered to have been established during or after the speciation of *D. erecta*. Alteration of the ecological environment should also change the mode of natural selection. In particular, alteration of host plants as feeding and breeding sites is likely to be very important for the evolution of digestive enzymes such as amylase. We conclude that adaptive protein evolution occurred in the process of adaptation of *D. erecta* to the new host plant Pandanus.

To obtain biochemical evidence of adaptation in this enzyme, we made a comparative study of hydropathy for the amino acid sequences of amylase in these species. No clear differences were observed in any comparisons between the species (data not shown). However, we found clear differences between species in the several biochemical properties of amylase such as specific activities under different temperature and pH ranges (SHIBATA and YAMAZAKI 1994). These differences provide supporting evidence for the hypothesis of adaptive protein evolution.

We are grateful to E. NITASAKA, C. H. LANGLEY, W. W. DOANE and H. TACHIDA for their valuable advice, and the National Institute of Genetics, Japan for providing the African sibling species fly stocks, namely, *D. mauritiana, D. sechellia, D. erecta, D. orena, D. teissieri* and *D. yakuba*.

LITERATURE CITED

ABE, K., 1958 Genetical and biochemical studies on amylase in *Drosophila melanogaster*. Jpn. J. Genet. **33:** 138–145.

ARNHEIM, N., M. KRYSTAL, R. SCHMICKEL, G. WILSON, O. RYDER *et al.*, 1980 Molecular evidence for genetic exchanges among ribosomal genes on nonhomologous chromosomes in man and apes. Proc. Natl. Acad. Sci. USA **77:** 7323–7327.

ASHBURNER, M., M. BODMER and F. LEMEUNIER, 1984 On the evolutionary relationships of *Drosophila melanogaster*. Dev. Genet. **4:** 295–312.

BINGHAM, P. M., R. LEVIS and G. M. RUBIN, 1981 Cloning of DNA sequences from the white locus of *D. melanogaster* by a novel and general method. Cell **25:** 693–704.

BOER, P. H., and D. A. HICKEY, 1986 The alpha-amylase gene in *Drosophila melanogaster:* nucleotide sequence, gene structure and expression motifs. Nucleic Acids Res. **14:** 8399–8411.

BROWN, C. J., C. F. AQUADRO and W. W. ANDERSON, 1990 DNA sequence evolution of the amylase multigene family in *Drosophila pseudoobscura*. Genetics **126:** 131–138.

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.

COEN, E., T. STRACHAN and G. DOVER, 1982 Dynamics of concerted evolution of ribosomal DNA and histone gene families in the *melanogaster* species subgroup of *Drosophila*. J. Mol. Biol. **158:** 17–35.

COUTURIER, G., D. LACHAISE and L. TSACAS, 1985 Le Drosophilidae et leurs gîtes larvaires dans la forêt dense humide de Taï en Côte-d'Ivoire. Rev. Fr. Entomol. (N. S.)(1986) **7:** 291–307.

DAÏNOU, O., M. L. CARIOU, J. R. DAVID and D. HICKEY, 1987 Amylase gene duplication: an ancestral trait in *Drosophila melanogaster* species subgroup. Heredity **59:** 245–251.

DE JONG, G. A., A. J. W. HOORN, G. E. W. THÖRIG and W. SCHARLOO, 1972 Frequencies of amylase variants in *Drosophila melanogaster*. Nature **238:** 452–453.

EANES, W. F., M. KIRCHNER and J. YOON, 1993 Evidence for adaptive evolutionof the *G6pd* gene in the *Drosophila melanogaster* and *Drosophila simulans* lineages. Proc. Natl. Acad. Sci. USA **90:** 7475–7479.

EISSES, K. T., H. V. DIJK and W. VAN DELDEN, 1979 Genetic differentiation within the *melanogaster* species group the genus *Drosophila* (*Sophophora*). Evolution **33:** 1063–1068.

FRISCHAUF, A. M., H. LEHRACH, A. POUSTKA and N. MURRAY, 1983 Lambda replacement vectors carrying polylinker sequences. J. Mol. Biol. **170:** 827–842.

GEMMILL, R. M., J. N. LEVY and W. W. DOANE, 1985 Molecular cloning of α-amylase genes from *Drosophila melanogaster*. I. Clone isolation by use of a mouse probe. Genetics **110:** 299–312.

GEMMILL, R. M., P. E. SCHWARTZ and W. W. DOANE, 1986 Structural organization of the *Amy* locus in seven strains of *Drosophila melanogaster*. Nucleic Acids Res. **14:** 5337–5352.

HAWLEY, S. A., R. A. NORMAN, C. J. BROWN, W. W. DOANE, W. W. ANDERSON *et al.*, 1990 Amylase gene expression in intraspecific and interspecific somatic transformants of *Drosophila*. Genome **33:** 501–508

HEY, J., and R. M. KLIMAN, 1993 Population genetics and phyloge-

netics of DNA sequence variation at multi loci within the *Drosophila melanogaster* species complex. Mol. Biol. Evol. **10:** 804–822.

HICKEY, D. A., 1979 The geographical pattern of an enzyme polymorphism in *D. melanogaster*. Genetica **51:** 1–4.

HICKEY, D. A., L. BALLY-CUIF, S. BUKASHAWA, V. PAYANT and B. F. BENKEL, 1991 Concerted evolution of duplicated protein-coding genes in *Drosophila*. Proc. Natl. Acad. Sci. USA **88:** 1611–1615.

HOLLIDAY, R., 1964 A mechanism for gene conversion in fungi. Genet. Res. Camb. **5:** 282–304.

INOMATA, N., H. SHIBATA, E. OKUYAMA and T. YAMAZAKI, 1995 Evolutionary relationships and sequence variation of α-amylase variants encoded by duplicate genes in the *Amy* locus of *Drosophila melanogaster*. Genetics **141:** 237–244

JUKES, T. H., and C. R. CANTOR, 1969 Evolution of protein molecules, pp. 21–132 in *Mammalian Protein Metabolism*, edited by H. N. MUNRO. Academic Press, New York.

KIKKAWA, H., 1964 An electrophoretic study on amylase in *Drosophila melanogaster*. Jpn. J. Genet. **39:** 401–411.

KIMURA, M., and T. OHTA, 1971 Protein polymorphism as a phase of molecular evolution. Nature **229:** 467–469.

LACHAISE, D., M. L. CARIOU, J. R. DAVID, F. LEMEUNIER, L. TSACAS *et al.*, 1988 Historical biogeography of the *Drosophila melanogaster* species subgroup. Evol. Biol. **22:** 159–225.

LANGLEY, C. H., A. E. SHRIMPTON, T. YAMAZAKI, N. MIYASHITA, Y. MATSUO *et al.*, 1988 Naturally occurring variation in the restriction map of the *Amy* region of *Drosophila melanogaster*. Genetics **119:** 619–629.

LEMEUNIER, F., and M. ASHBURNER, 1976 Relationships within the *melanogaster* species subgroup of the genus *Drosophila* (*Sophophora*). II. Phylogenetic relationships between six species based upon chromosome banding sequences. Proc. R. Soc. Lond. **193:** 275–294.

LEVY, J. N., R. M. GEMMILL and W. W. DOANE, 1985 Molecular cloning of α-amylase genes from *Drosophila melanogaster*. II. Clone organization and verification. Genetics **110:** 313–324.

MANIATIS, T., E. F. FRITSCH and J. SAMBROOK, 1982 *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

MATSUO, Y., and T. YAMAZAKI, 1986 Genetic analysis of natural populations of *Drosophila melanogaster* in Japan. VI. Differential regulation of duplicated amylase loci and degree of dominance of amylase activity in different environments. Jpn. J. Genet. **61:** 543–558.

MATSUO, Y., and T. YAMAZAKI, 1989 Nucleotide variation and divergence in the histone multigene family in *Drosophila melanogaster*. Genetics **122:** 87–97.

MATSUURA, Y., M. KUSUNOKI, W. HARADA and M. KAKUDO, 1984 Structure and possible catalytic residues of Taka-Amylase-A. J. Biochem. **95:** 697–702.

MCDONALD, J. H., and M. KREITMAN, 1991 Adaptive protein evolution at the *Adh* locus in *Drosophila*. Nature **351:** 652–654.

MESELSON, M. S., and C. M. RADDING, 1975 A general model for genetic recombination. Proc. Natl. Acad. Sci. USA **72:** 358–361.

NEI, M., 1975 *Molecular Population Genetics and Evolution*. North Holland Publishing, Amsterdam.

OHNISHI, S., M. KAWANISHI and T. K. WATANABE, 1983 Biochemical phylogenies of *Drosophila*: protein differences detected by two-dimensional electrophoresis. Genetica **61:** 55–63.

OHTA, T., 1973 Slightly deleterious mutant substitutions in evolution. Nature **246:** 96–98.

OHTA, T., 1993 Amino acid substitution at the *Adh* locus of *Drosophila* is facilitated by small population size. Proc. Natl. Acad. Sci. USA **90:** 4548–4551.

OKUYAMA, E., and T. YAMAZAKI, 1988 Nucleotide sequence of the duplicated amylase structural gene in *Drosophila melanogaster*: one active and one null strain. Proc. Jpn. Acad. **64:** 274–277.

PAYANT, V., S. ABUKASHAWA, M. SASSEVILLE, B. F. BENKEL, D. A. HICKEY *et al.*, 1988 Evolutionary conservation of the chromosomal configuration and regulation of amylase genes among eight species of the *Drosophila melanogaster* species subgroup. Mol. Biol. Evol. **5:** 560–567.

RIO, B., G. COUTURIER, F. LEMEUNIER and D. LACHAISE, 1983 Evolution d'une spécialsation saisonière chez *Drosophila erecta* (Dipt., Drosophilidae). Ann. Soc. Ent. Fr. (N. S.) **19:** 235–248.

SAITOU, N., and M. NEI, 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:** 406–425.

SANGER, F., A. R. COULSON, B. G. BARRELL, A. J. H. SMITH and B. A. ROE, 1980 Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. **59:** 33–49.

SCHWARTZ, P. E., and W. W. DOANE, 1989 Molecular cloning of alpha-amylase genes from *Drosophila melanogaster*. III. an inversion at the *Amy* locus in an amylase-null strain. Biochem. Genet. **27:** 31–46.

SINGH, R. S., D. A. HICKEY and J. DAVID, 1982 Genetic differentiation between geographically distant populations of *Drosophila melanogaster*. Genetics **101:** 235–256.

SHIBATA, H., and T. YAMAZAKI, 1994 A comparative study of the enzymological features of α-amylase in the *Drosophila melanogaster* species subgroup. Jpn. J. Genet. **69:** 251–258.

SMITH, G. P., 1976 Evolution of repeated DNA sequences by unequal crossover. Science **191:** 528–535.

SOLIGNAC, M., M. MONNEROT and J. C. MOUNLOU, 1986 Mitochondrial DNA evolution in the melanogaster species subgroup of *Drosophila*. J. Mol. Evol. **23:** 31–40.

SZOSTAK, J. W., T. L. ORR-WEAVER and R. J. ROTHSTEIN, 1983 The double-strand-break repair model for recombinnation. Cell **33:** 25–35.

YAMAZAKI, T., Y. MATSUO, Y. INOUE and Y. MATSUO, 1984 Genetic analysis of natural populations of *Drosophila melanogaster* in Japan. I. Protein polymorphism, lethal gene, sterility gene, inversion polymorphism and linkage disequilibrium. Jpn. J. Genet. **59:** 33–49.

Communicating editor: D. CHARLESWORTH