Evolutionary Relationships and Sequence Variation of α-Amylase Variants Encoded by Duplicated Genes in the *Amy* Locus of *Drosophila melanogaster*

Nobuyuki Inomata, Hiroki Shibata, Eisaku Okuyama and Tsuneyuki Yamazaki

Laboratory of Population Genetics, Department of Biology, Faculty of Science, Kyushu University, Fukuoka 812, Japan

Manuscript received April 29, 1993
Accepted for publication June 12, 1995

ABSTRACT

To infer the genealogical relationships of α-amylase electromorphs of *Drosophila melanogaster*, we determined the nucleotide sequences of a collection of electromorphs sampled throughout the world. On average there were 1.0 amino acid substitutions between identical electromorphs and 3.9 between different electromorphs, respectively. We found that the evolution of AMY^1 through AMY^6 electromorphs occurred by sequential accumulation of single amino acid substitutions each causing one charge difference. The nucleotide diversities at synonymous sites within Amy^1,Amy^2,Amy^3,Amy^4 and Amy^6 were 0.0321, 0.0000, 0.0355, 0.0059 and 0.0030, respectively. We also obtained evidence of genetic exchanges, such as intrachromosomal recombination, interchromosomal recombination or gene conversion, between the two duplicated *Amy* genes as well as among the alleles.

THE α-amylase (AMY, EC 3.2.1.1, α-1,4-glucan-4-glucanohydrolase) of *Drosophila melanogaster* is highly polymorphic in natural populations (KIKKAWA 1964; DOANE 1969; HICKEY 1979; SINGH *et al.* 1982; YAMAZAKI *et al.* 1984). Twelve different electromorphs were detected by using polyacrylamide gel electrophoresis (DAÏNOU *et al.* 1987). For convenience, the electromorphs that can be distinguished on polyacrylamide gels have been numbered −1 and 1–7, with −1 observed to migrate toward the anode most rapidly and 7 the least (KIKKAWA 1964; DOANE 1969; HICKEY 1979; SINGH *et al.* 1982; YAMAZAKI *et al.* 1984). Many of these electromorphs are rare and only found locally (HICKEY 1979). Usually the AMY^1 electromorph is predominant in natural populations in Japan and North America (YAMAZAKI *et al.* 1984; LANGLEY *et al.* 1988). Although there are considerable differences in α-amylase activity between the different electromorph classes and even within the same electromorph class (BENKEL and HICKEY 1986; DOANE *et al.* 1987; LANGLEY *et al.* 1988), it is not clear whether these differences are due to the effects on the catalytic function caused by amino acid substitutions or to regulating effects due to the differences in *trans*-acting factors or *cis*-acting elements linked with a particular electromorph. While there was a report that the regulating effect (inducibility) correlated with fitness (MATSUO and YAMAZAKI 1984), the adaptive significance of α-amylase electromorphs is not yet well understood (DEJONG *et al.* 1972; DEJONG and SCHARLOO 1976; HICKEY 1977; POWELL and ANDJELKOVIC 1983; MATSUO and YAMAZAKI 1984).

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

The *Amy* locus has been cloned and characterized (*Amy*, 2–80; 54A-B1) (GEMMILL *et al.* 1985, 1986; LEVY *et al.* 1985; BOER and HICKEY 1986). The locus is duplicated. The two genes are ~4 kb apart, transcribed divergently and each consists of 1482-bp ORF without introns. The genes proximal and distal to the centromere are called the proximal gene and the distal gene, respectively.

The duplication of the genes is one of the simplest cases of a multigene family. There are many examples of small multigene families in the genomes of higher organisms (NAKANISHI *et al.* 1979; BHAT *et al.* 1980; WAHLI *et al.* 1981). The members of a multigene family do not evolve independently but exhibit an evolutionary process called concerted evolution or coincidental evolution (ARNHEIM 1983). Several molecular mechanisms and theoretical models for concerted evolution have been proposed (LEIGH BROWN and ISH-HOROWICZ 1981; OHTA 1983; NAGYLAKI 1984). The *Amy* gene is duplicated throughout the *melanogaster* species subgroup (DAÏNOU *et al.* 1987; PAYANT *et al.* 1988; SHIBATA and YAMAZAKI 1995). Evidence of concerted evolution in the coding region of the *Amy* gene but not in the flanking region has been found (HICKEY *et al.* 1991; SHIBATA and YAMAZAKI 1995). However, within a species the amount and nature of variation in this gene is not known. The *Amy* gene is thus a model for genealogical studies in such small multigene families.

To infer the history of the electrophoretic phenotypes (electromorphs), we sequenced the *Amy* genes coding for different electromorphs of *D. melanogaster*. In this paper we present the amino acid and nucleotide sequences of various α-amylase electromorphs of *D. melanogaster* in worldwide populations. In the sample, genes encoding six different electromorphs and two

TABLE 1

A list of strains and assignment of expressed electromorphs

Strain	Origin	Haplotype	Expressed electromorphs	
			Proximal	Distal
TN329	Tanushimaru, Japan	1-1	1	1
KO123	Gwacheon, Korea	1-2	2	1
AO168	Aomori, Japan	1-3	1	3
KO140	Gwacheon, Korea	1-4	4	1
TN22	Tanushimaru, Japan	1-6	1	6
J87	Tananarive, Madagascar	2-3	2	3
TN256	Tanushimaru, Japan	3-6	3	6
1420#1	Drosophila Stock Center	4-6	4	6
L16	Nairobi, Kenya	5-6	5	6
NULL	Brownsville, Texas	Null	None	None

null variants are presented. We also present evidence for genetic exchanges such as intrachromosomal recombination, interchromosomal recombination or gene conversion.

MATERIALS AND METHODS

Strains: Ten strains of *D. melanogaster* with various α-amylase electromorphs (haplotypes) were collected from all over the world and were made homozygous for chromosome *II* using a balancer chromosome *Cy* and routine crosses. The 1420#1 ($Amy^{4,6}$) strain was supplied by the National Drosophila Species Resource Center at Bowling Green State University. The NULL strain, which was collected from Brownsville, Texas, was obtained from Dr. D. A. HICKEY, University of Ottawa, Canada, while the others were collected from natural populations. These strains are listed in Table 1.

Polyacrylamide gel electrophoresis: Adult flies of each strain were homogenized by sonication in 0.1 M Tris-borate/5 mM $MgCl_2$/10% sucrose buffer (pH 8.9). Homogenized samples were immediately applied to the polyacrylamide gels (5% acrylamide/0.2% Bis-acrylamide/20 mM $CaCl_2$/0.1 M Tris-borate), and the electrophoresis was done in 0.1 M Tris-borate (pH 8.9) for 3 hr at 0°, 300 V. After running, the gels were incubated in 2% starch solution for 1 hr at 25°. They were then washed with water and stained in an I_2-KI solution. The band mobilities were determined by comparing the bands with those of a standard marker strain, Oregon-R (Amy^1) or Canton-S ($Amy^{1,3}$). The relative mobilities of AMY^1 through AMY^6 were 1.00, 0.92, 0.87, 0.80, 0.75 and 0.68, respectively.

Molecular analyses: Genomic DNA libraries of each isogenic strain were constructed according to FRICHAUF *et al.* (1983) using the commercially available lambda packaging extract, Gigapack II (Stratagene). The molecular cloning described below was conducted according to MANIATIS *et al.* (1982). The phage clones containing the *Amy* region were isolated from the genomic libraries by plaque hybridization using the plasmid pDm3.8 containing the *Amy*-coding sequence of the Canton-S strain as a probe. DNA fragments of phage clones were digested by *Eco*RI or *Eco*RI/*Xho*I. Then each of the DNA fragments containing one of the duplicated *Amy* genes was subcloned into the polylinker site of the pBluescript phagemid vector (Stratagene). The sequencing of the *Amy* structural genes was performed using the modified dideoxy method (SANGER *et al.* 1980) with [α-^{35}S]dCTP and the synthetic 17 base oligonucleotide primers. The nucleotide sequences were then determined for both strands of DNA.

Data deposition: The nucleotide sequence data reported in this article appear in the DDBJ, EMBL and GenBank Data Libraries under the accession numbers L22716–L22735.

Data analysis: Since the strains used in our study were not chosen randomly, we estimated the average number of pairwise amino acid substitutions in natural populations as $(\Sigma p_i p_j A_{ij})/(\Sigma p_i p_j)$, where p_i and p_j are the frequencies of electromorph i and j ($i \leq j$) given in Table 2, and A_{ij} is the average number of pairwise amino acid substitutions between electromorph i and j.

RESULTS

Nucleotide variation of *Amy* structural genes: We sequenced sixteen *Amy* structural genes, each containing eight proximal and eight distal genes of *D. melanogaster*. Four sequences of TN329 and NULL strains reported by OKUYAMA and YAMAZAKI (1988) were added to the sixteen sequences in the analyses. Data of a *D. simulans* strain used in this study were obtained from SHIBATA and YAMAZAKI (1995). Figure 1 lists the segregating sites among them. Nucleotide sequences were numbered from the initiation codon. There were 23, 33 and 37 synonymous segregating sites and 9, 10 and 10 nonsynonymous segregating sites among the nine proximal genes, the nine distal genes and the total of the 18 *Amy* structural genes of *D. melanogaster*, respectively (Figure

TABLE 2

The frequency of AMY electromorphs of natural populations throughout the world

Electromorph	Frequency[a]
AMY^1	0.8225
AMY^2	0.0030
AMY^3	0.1265
AMY^4	0.0145
AMY^5	0.0095
AMY^6	0.0240
SUM	1.0000

[a] Calculated from genotype frequencies reported by HICKEY (1979) assuming that the genotypes showing only one band have two identical electromorphs.

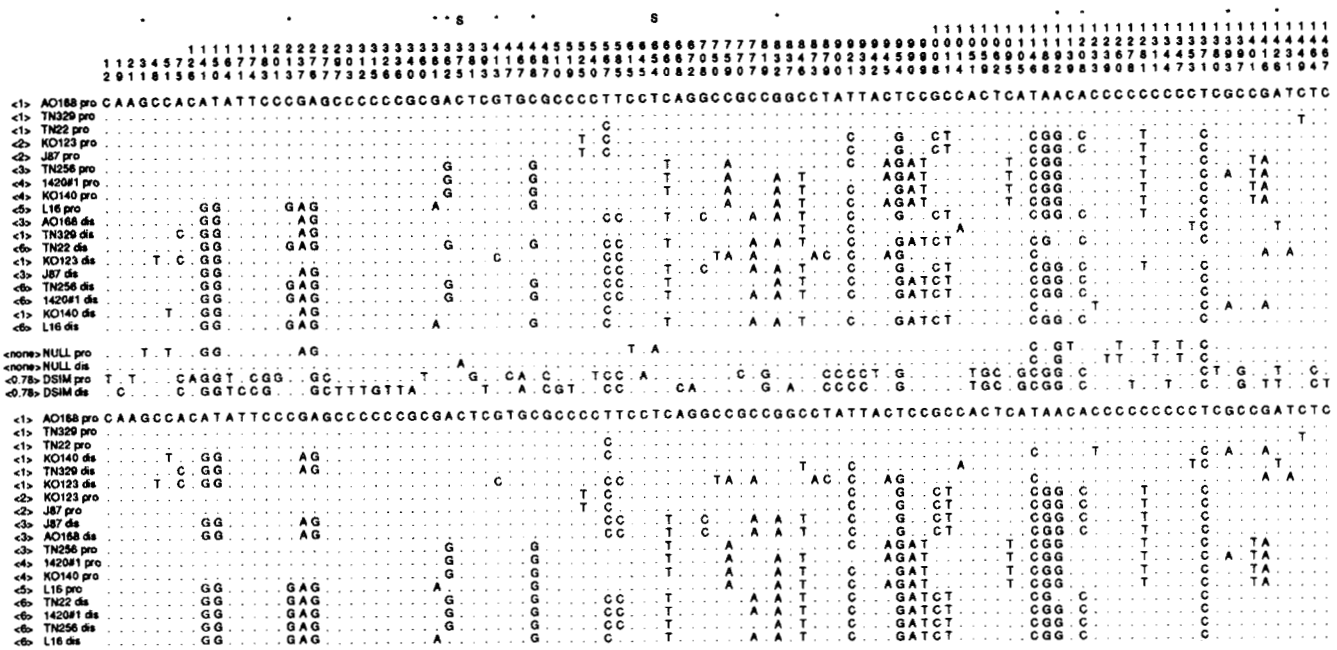

FIGURE 1.—Segregating sites among nucleotide sequences of *Amy* structural genes in *D. melanogaster* and *D. simulans*. The proximal genes are separated from the distal genes in above block. Below block is arranged in order of electromorph class. pro and dis after each strain denote the proximal gene and distal gene, respectively. The DSIM indicates *D. simulans*. The AO168 proximal gene found on the top line was used as a consensus sequence. The nucleotide positions are numbered from the initiation codon and are given by the numbers above the consensus sequence AO168 pro. The expressed electromorphs are shown in the leftmost column. Note that the expressed electromorphs of *D. simulans* are presented by the mobility relative to that of AMY[1] of *D. melanogaster*. The dots (.) indicate the bases identical to the AO168 pro sequence. The nonsynonymous sites in *D. melanogaster* and the sites causing stop codons are labeled by (*) and (S) over the nucleotide positions, respectively. In addition to the stop codon the NULL distal gene has a single base (guanine) insertion at the third base position, producing a frame shift mutation from the second amino acid (data not shown in Figure 1, see OKUYAMA and YAMAZAKI 1988). Note that the 361st and 362nd nucleotide substitutions cause different amino acid substitutions at the same amino acid position.

1), excluding the genes of NULL strain, since the genes of this strain must be pseudogenes (see Figure 2 and also OKUYAMA and YAMAZAKI 1988). There were six nonsynonymous fixed sites out of 15 fixed site differences between *D. melanogaster* and *D. simulans* (Figure 1).

We quantified the nucleotide variability using the nucleotide diversity measure (NEI and LI 1979; NEI and TAJIMA 1981; NEI 1987). The nucleotide diversity, π, is the average heterozygosity per site of the sequences. If m sites are found, then πm gives an estimate of $4N_e u = M$ (WATTERSON 1975), where N_e is the effective population size and u is the mutation rate per generation per DNA sequence examined. The NULL strain was not included in the computation of π. First, we examined the nucleotide diversity considering the proximal and distal genes separately. Since our sampling was not random and the flies were collected from all over the world, the nucleotide diversity was computed by weighting with the AMY electromorph frequencies in the worldwide populations (HICKEY 1979) presented in Table 2. Table 3 shows the results. The nucleotide diversity within identical electromorphs was also estimated (Table 4). There are many synonymous differences within the AMY[1] and the AMY[3] electromorphs, and many nonsynonymous differences within the AMY[3] electromorph, compared with those within the other electromorphs.

The pattern of nucleotide substitutions: The pattern of nucleotide substitution at the 921st, 954th, 1146th, 1188th, 1192nd and 1371st sites shown in Figure 1 is sufficient to distinguish the AMY[1] electromorph class from the non-AMY[1] electromorph classes, although some *Amy*[1] distal genes share the same nucleotides at several sites with the non-*Amy*[1] genes. The pattern of nucleotide substitution at the 141st, 150th, 237th and 276th sites shown in Figure 1 distinguishes the proximal from the distal gene locus with high probability, though there are some exceptions that the proximal gene of the L16 strain shares the same nucleotides with the distal genes at all those sites and the distal gene of the KO123 strain shares the same nucleotides with the proximal genes at the 237th and 276th sites. The pattern of nucleotide substitution of the proximal gene (*Amy*[3]) of the TN256 strain is very similar to that of the *Amy*[4] genes rather than the other two *Amy*[3] genes but is different from that of the *Amy*[4] genes around the 832nd and 846th sites, which are the same as either the *Amy*[1] or *Amy*[2] genes, as shown in Figure 1. The patterns of nucleotide substitution at the other sites in our samples are more complex.

Nucleotide divergence between *D. melanogaster* and *D. simulans*: *Amy* genes of seven species of the *melanogaster* species subgroup were cloned and sequenced by

Electromorph	Strain	Positions of Amino Acid Substitution																						
		7	11	55	71	93	109	114	121	125	138	156	163	218	223	257	278	351	369	398	403	465	466	476
1	AO168 pro	I	A	Y	S	E	T	V	D	Y	S	S	E	Y	N	G	D	T	I	T	E	V	T	N
	TN22 pro	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	TN329 pro	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	TN329 dis	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y
	KO140 dis	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	I	-	-
	KO123 dis	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-
2	KO123 pro	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	A	-	-	-
	J87 pro	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	A	-	-	-
3	AO168 dis	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	N	-	-	A	A	-	-	-
	J87 dis	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	N	-	-	A	A	-	-	-
	TN256 pro	-	-	-	-	-	-	-	G	-	-	R	-	-	-	-	-	-	-	A	-	-	-	-
4	KO140 pro	-	-	-	-	-	-	-	G	-	-	R	-	-	-	-	N	-	-	A	-	-	-	-
	1420#1 pro	-	-	-	-	-	-	-	G	-	-	R	-	-	-	-	N	-	-	A	-	I	-	-
5	L16 pro	-	-	-	R	-	-	-	N	-	-	R	-	-	-	-	N	-	-	A	-	-	-	-
6	TN22 dis	-	-	-	R	-	-	-	G	-	-	R	-	-	-	-	N	-	-	-	A	-	-	-
	TN256 dis	-	-	-	R	-	-	-	G	-	-	R	-	-	-	-	N	-	-	A	A	-	-	-
	L16 dis	-	-	-	R	-	-	-	N	-	-	R	-	-	-	-	N	-	-	A	A	-	-	-
	1420#1 dis	-	-	-	R	-	-	-	G	-	-	R	-	-	-	-	N	-	-	A	A	-	-	-
none	NULL pro	-	S	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	A	-	-	-	-
	NULL dis	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	A	-	-	-	-
0.78	DSIM pro	-	-	F	-	Q	-	-	-	-	-	N	Q	-	-	A	-	S	V	A	A	-	S	Y
	DSIM dis	L	-	F	-	Q	I	I	-	-	-	N	Q	-	T	-	-	S	V	A	A	-	S	Y

FIGURE 2.—Amino acid substitutions of various electromorphs. pro and dis after each strain denote the proximal gene and distal gene, respectively. The numbers above the AO168 pro sequence are the positions at amino acid sequences. The leftmost column shows the expressed electromorphs numbered from the mobility on the polyacrylamide gel. -, amino acids identical to the AO168 pro sequence. The shadowed columns indicate the positions with charge change. The DSIM represents *D. simulans*. +, the emergence of a stop codon.

SHIBATA and YAMAZAKI (1995). Here we use only the sequences of *D. simulans*, a sibling species of *D. melanogaster*. *D. simulans* has two *Amy* gene copies in inverted sequence, and each copy consists of 1482-bp ORF without introns as in *D. melanogaster*. We estimated the number of nucleotide substitutions per site (JUKES and CANTOR 1969) and their standard errors (KIMURA and OHTA 1972) between *D. melanogaster* and *D. simulans*. The estimates of the number of nucleotide substitutions per synonymous site in the proximal and the distal genes are 0.0977 ± 0.0179 and 0.1079 ± 0.0189, respectively. The difference between the proximal and distal genes is not significant. The number of nucleotide substitutions per synonymous site between *D. melanogaster* and *D. simulans* in the *Amy* genes is about twice as large as that in the *Adh* gene, 0.0539 ± 0.0172 (KREITMAN and HUDSON 1991), and is also similar to that in the Histone H3 genes, 0.1145 ± 0.0353 (MATSUO and YAMAZAKI 1989).

Amino acid variation: From the DNA sequences putative amino acid sequences were determined. The positions of amino acid substitution are summarized in Figure 2. Each clone representing various electromorphs can be designated as either proximal or distal (Table 2) according to the size of the clone because the restriction maps and polymorphic sites of the *Amy* regions are well characterized (LEVY *et al.* 1985; GEMMILL *et al.* 1986; BOER and HICKEY 1986; LANGLEY *et al.* 1988). Excluding the NULL strain there are amino acid substitutions at nine sites. To determine the relative charge change of each electromorph, we scored +1, −1 or 0, when the amino acid residue at the substituted position was charged positively, negatively or neutrally, respectively, at the pH value usually used in electrophoresis. As a result the relative charge changes of the AMY^1, AMY^2, AMY^3, AMY^4, AMY^5 and AMY^6 electromorphs were scored as −3, −2, −1, 0, +1 and +2, respectively. This observation is consistent with the band mobilities and the numbering of each AMY electromorph.

One or more amino acid substitutions distinguishes each AMY electromorph. For example, the AMY^1 and AMY^2 electromorphs differed at two amino acid positions, 398 (Thr/Ala) and 403 (Glu/Ala), and the latter position distinguishes the two electromorphs electrophoretically because Glu has a negative charge and Ala is neutral in normal electrophoresis. Electromorphs with the same mobility in polyacrylamide gel electrophoresis do not necessarily have the same amino acid sequences. For example, in this sample the six AMY^1 electromorphs included four different sequence variants. The substitutions are at three different positions (138 Ser/Thr, 465 Val/Ile, 476 Asn/Tyr), all of which conserve the charge. All three AMY^3 electromorphs differed in amino acid sequence from one another, although the relative charge of all AMY^3 electromorphs is scored as −1, and the positions that caused charge changes with respect to other alleles were quite differ-

TABLE 3

Estimates of π in the proximal and distal genes

	k*	π*
Within the proximal gene		
Synonymous sites	5.0	0.0148 (0.0315)
Nonsynonymous sites	0.9	0.0008 (0.0027)
Total sites	19.2	0.0129 (0.0093)
Within the distal gene		
Synonymous sites	15.1	0.0448 (0.0302)
Nonsynonymous sites	2.7	0.0023 (0.0035)
Total sites	17.8	0.0120 (0.0096)
Between the proximal and distal genes		
Synonymous sites	14.3	0.0424 (0.0452)
Nonsynonymous sites	1.9	0.0017 (0.0032)
Total sites	16.3	0.0110 (0.0128)
All comparisons		
Synonymous sites	12.6	0.0372 (0.0384)
Nonsynonymous sites	1.9	0.0017 (0.0032)
Total sites	14.5	0.0098 (0.0112)

The length of synonymous sites, nonsynonymous sites and total sites are 338.3, 1143.7 and 1482 bp, respectively.

* k is the average number of pairwise nucleotide differences and π is the nucleotide diversity (NEI and LI 1979; NEI and TAJIMA 1981; NEI 1987) weighted by *Amy* allele frequency in natural populations (HICKEY 1979). The unweighted nucleotide diversity is represented within parentheses.

TABLE 4

Estimates of π within identical electromorph

	n*	k*	π*
AMY[1]	6		
Synonymous sites		10.87	0.0321
Nonsynonymous sites		1.00	0.0009
Total sites		11.87	0.0080
AMY[2]	2		
Synonymous sites		0.00	0.0000
Nonsynonymous sites		0.00	0.0000
Total sites		0.00	0.0000
AMY[3]	3		
Synonymous sites		12.00	0.0355
Nonsynonymous sites		2.67	0.0023
Total sites		14.67	0.0099
AMY[4]	2		
Synonymous sites		2.00	0.0059
Nonsynonymous sites		1.00	0.0009
Total sites		3.00	0.0020
AMY[6]	4		
Synonymous sites		1.00	0.0030
Nonsynonymous sites		1.50	0.0013
Total sites		2.50	0.0017
Average[a]	17		
Synonymous sites		7.92	0.0235
Nonsynonymous sites		1.27	0.0011
Total sites		9.23	0.0062

The length of synonymous sites, nonsynonymous sites and total sites are 338.3, 1143.7 and 1482 bp, respectively.

* n is the number of samples used in this study, k is the average number of pairwise nucleotide differences and π is the nucleotide diversity (NEI and LI 1979; NEI and TAJIMA 1981; NEI 1987).

[a] This is the weighted average of each k and π value.

ent from the AMY[1] electromorph. This suggests that the AMY[3] electromorphs have different ancestors and have converged to become electrophoretically identical (see below). In both sequences of the NULL strain, stop codons appeared, in addition to a frame shift mutation at the third position of the distal gene. Moreover, the restriction map survey suggested the presence of an inversion around the *Amy* region. The sequence data indicated that the break points of the inversion were located downstream of the coding region for both the proximal and distal gene. The former events are considered to be the reason why these genes have no α-amylase activity. Previously, OKUYAMA and YAMAZAKI (1988) discussed this strain in detail.

Figure 3 shows the hydropathy profile (KYTE and DOOLITTLE 1982) of two representative electromorphs of *D. melanogaster* and that of *D. simulans*. The positions of the amino acid substitutions are almost completely in the hydrophilic regions and are outside the four amino acid sequence motifs conserved in both the Drosophila and mouse α-amylase proteins, as pointed out by BOER and HICKEY (1986). No hydropathic changes were seen among them. The same was also found among the electromorphs of *D. melanogaster*. The disulfide bridges between cysteine residues are important in determining the conformation of proteins. There were no cysteine substitutions. These observations support the above conclusion that the mobilities of α-amylase electromorphs are determined mostly by the net charge differences and not by the three-dimensional structure of the protein.

DISCUSSION

There were several amino acid substitutions not only among the different electromorphs but also within electromorphs of the α-amylase proteins of *D. melanogaster*. This is in contrast with ADH electromorphs, which have the same amino acid sequences except for a single amino acid substitution distinguishing Adh^F and Adh^S alleles (KREITMAN 1983). For α-amylase, the average number of pairwise amino acid substitutions in natural populations is 1.9. In other words, when we randomly sample two α-amylase electromorphs from natural populations, there will be on average 1.9 amino acid substitutions between them. The average number of pairwise amino acid substitutions between identical electromorphs, and between different electromorphs in natural populations, are 1.0 and 3.9, respectively. Thus, even when we randomly choose two of the same electromorph, there will be on average 1.0 amino acid substitution between them.

It is apparent that the duplication of the *Amy* loci occurred before the divergence of the *melanogaster* species subgroup (DAÏNOU *et al.* 1987; PAYANT *et al.* 1988; SHIBATA and YAMAZAKI 1995). If the duplicated (proxi-

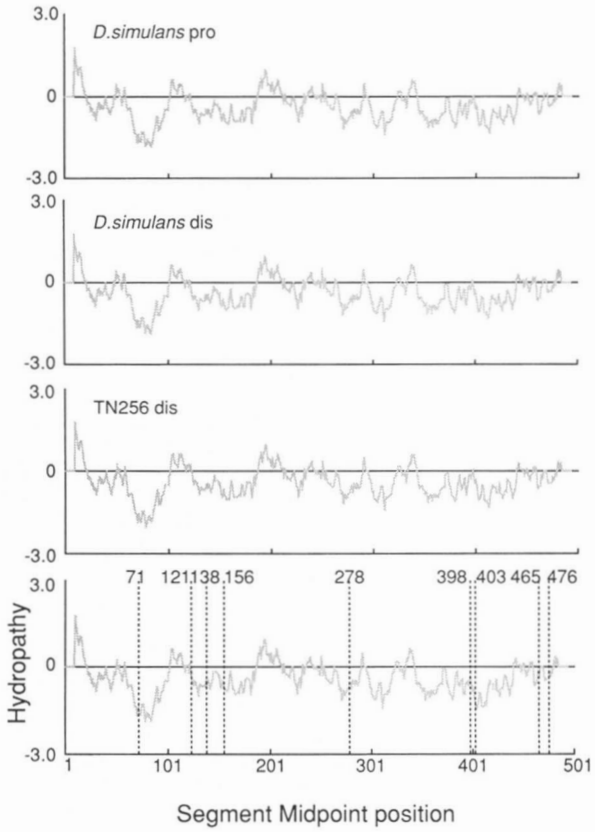

FIGURE 3.—Hydropathy plots for representative AMY electromorphs of *D. melanogaster* and *D. simulans*. Positive values indicate hydrophobicity, negative values hydrophilicity. The segment size is 21 amino acids. The lowest figure shows the hydropathy of the TN22 proximal protein. The positions of the amino acid substitutions in *D. melanogaster* are indicated in the lowest figure.

mal and distal) genes evolved independently since their origination, the nucleotide diversity between them should be much larger than that within either the proximal or the distal genes. However, their values were similar to each other (Table 3). This suggests that there has been genetic exchange between the duplicated genes. In fact, there were many sites with the same nucleotide substitutions in both of the duplicated genes. For example, C/T substitutions at sites 567, 660, 846, 921, 999, 1146, 1281 and 1371 were found in both. The nucleotide substitutions common to both of the duplicated genes must result from genetic exchange between them, because the occurrence of the same mutation at the same site in both loci is very unlikely.

The results of SAWYER's (1989) test for detecting gene conversion also found the occurrence of genetic exchanges. The test is based on imbalances in the distribution of maximal segments in which pairs of sequences agree. The test defines four kinds of statistics, SSCF, MCF, SSUF and MUF and the significance of each is judged by making 10,000 random permutations of the orders of the polymorphic sites in the sample. We applied these tests to the 16 *Amy* sequences of our samples. Sequences of the J87 and the NULL strains were excluded since the proximal and distal genes of the J87 strain were identical to the KO123 proximal and AO168 distal genes, respectively, and the genes of the NULL strain are thought to be pseudogenes.

When all polymorphic sites among the 16 *Amy* sequences were used for the test, the SSCF statistic was significant ($0.01 < P < 0.05$). We did not obtain any significant test values when the eight distal genes out of the 16 genes were used. On the other hand, when the eight proximal genes were used, the SSCF and SSUF statistics were highly significant ($P < 0.01$). In addition, the test indicated that the region between sites 1 and 361 of the L16 strain came from the foreign sequences other than the proximal genes examined, by gene conversion or reciprocal recombination ($P \ll 0.01$). It is apparent from the sequence data (Figure 1) that the foreign sequences are from distal genes by some kind of homogenizing mechanisms, since the nucleotide sequences common to both of proximal gene of the L16 strain and distal genes were found in this region. If we assume that one genetic exchange around this region took place, the outside materials are from distal gene of the L16 strain because the first 566 bp is exactly the same between both the proximal and distal genes including the sequences at the 361st and 362nd sites specific to the L16 strain.

Close examination of the substitution patterns indicates that many genetic exchanges, such as gene conversion or reciprocal recombination, are common both between the duplicated genes and within each locus. To visualize this more clearly, we expanded the region between sites 921 and 1146 of Figure 1 and showed each sequence schematically (Figure 4). It is clear from the figure that each strain is composed of mosaics of DNA sequences from other strains. If each site has experienced at most one mutation and there has been no recombination, we expect pairs of segregating sites to show a pattern consistent with the nested set of relationships defined by the genealogy of the sample. The observation of all four possible dinucleotide pairs is evidence for recombination between two sites. Three different locations of exchange, indicated by arrows in Figure 4, were detected by this method in the region between sites 921 and 1146: between 921 and 945, between 945 and 990, and between 999 and 1008. In the region covering the whole coding region of the *Amy*, 14 different locations were found in total.

The AMY^3 electromorph encoded by the TN256 proximal gene fortuitously has the same mobility as the other AMY^3 electromorphs, despite having four amino acid substitutions compared with those sequences (see Figure 2). The nucleotide sequence of the TN256 proximal gene is very similar to that of the 1420#1 and KO140 proximal genes that encode for the AMY^4 electromorph. In the region between sites 759 and 921, the nucleotide sequence of the TN256 proximal gene is

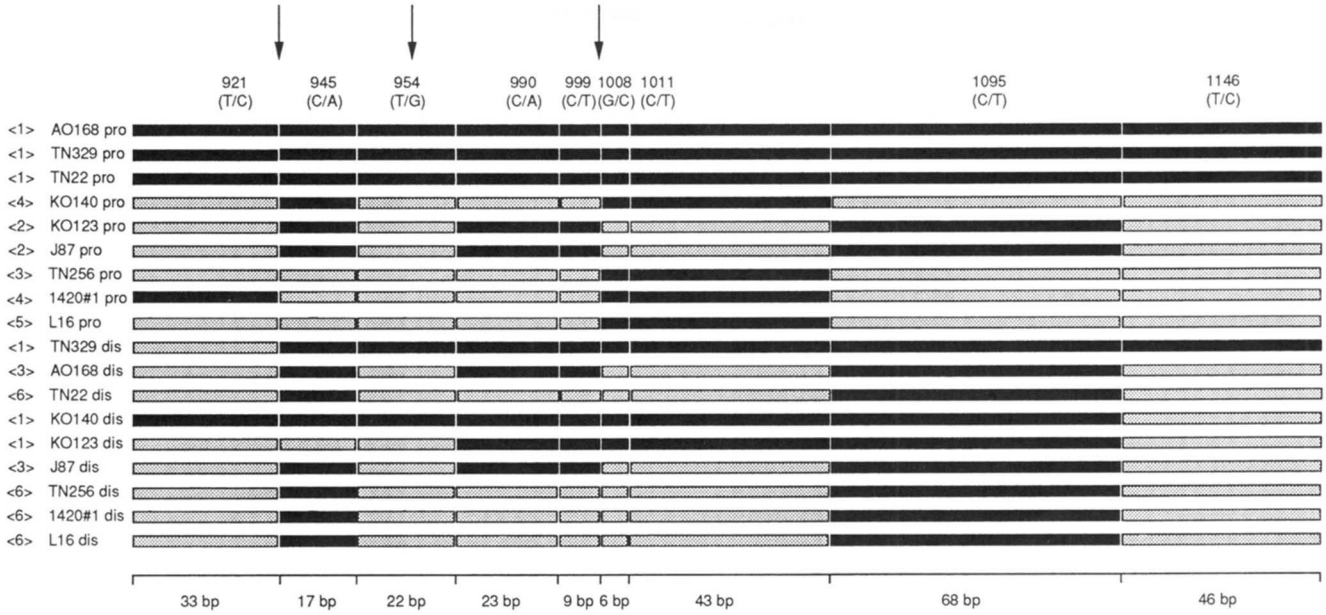

FIGURE 4.—A schematic figure of the nucleotide substitution patterns in the region between 921st and 1146th sites. Each fragment represents the nucleotide sequences containing a segregating site. Black fragments show the sequences identical to the AO168 proximal gene and gray ones a substitution. A boundary between the neighboring fragments is considered to be in the middle of two segregating sites concerned. A substitution at the 1014th site is not shown since it was not found in any other strains except for the TN329 distal genes. Vertical arrows indicate the locations of recombination. The leftmost column shows the expressed electromorphs. pro and dis after each strain denote the proximal gene and distal gene, respectively. The positions of the segregating sites are above the fragments and substitutions in the parentheses. The length of fragments are shown in the bottom.

identical to that of the proximal genes that encode for the AMY^1 and AMY^2 electromorphs (see Figure 1). A nucleotide substitution at the 832nd site (G/A) corresponds to an amino acid substitution with a charge difference at site 278 (Asp/Asn). This suggests that gene conversion between the Amy^4 and Amy^1 or Amy^2 genes has occurred around a small fragment including the amino acid substitution with a charge difference and resulted in the two types of AMY^3 electromorphs.

HICKEY *et al.* (1991) and SHIBATA and YAMAZAKI (1995) found clear evidence of concerted evolution at the coding region of *Amy* locus. They showed that sequence identity between the duplicated genes was observed only in the coding region but not in the flanking region, indicating that some kind of homogenizing mechanism such as gene conversion and/or unequal crossing over operated in the coding region. Moreover, data by SHIBATA and YAMAZAKI (1995) show that noncoding regions (both 5′ and 3′) around the two duplicated *Amy* loci are diverged not only independently but also specifically within each locus (proximal or distal) throughout the eight members of the *melanogaster* subgroup, suggesting absense of gene conversion or the presence of different selective forces. Our data in this study clearly show that sequence homogenity between the duplicated genes is maintained by frequent genetic exchange in various portions of the coding region.

We are grateful to H. TACHIDA, F. TAJIMA and C. H. LANGLEY for their valuable suggestions with analyses of data. We express our gratitude to T. S. TAKANO and E. NITASAKA for their many helpful comments and technical suggestions. Thanks are due to S. SAWYER for providing the computer program of his test. We thank O. OGASAWARA in our laboratory for his technical assistance. We also thank M. M. GREEN and J. WAKELEY for revising our manuscript. This work was supported in part by research grants to T.Y. from the Ministry of Education, Science and Culture of Japan.

LITERATURE CITED

ARNHEIM, N., 1983 Concerted evolution of multigene families, pp. 38–61 in *Evolution of Genes and Proteins*, edited by M. NEI and K. KOEHN. Sinauer Associates, Sunderland, MA.

BENKEL, B. F., and D. A. HICKEY, 1986 The interaction of genetic and environmental factors in the control of amylase gene expression in *Drosophila melanogaster*. Genetics **114:** 943–954.

BHAT, S. P., R. E. JONES, M. A. SULLIVAN and J. PIATIGORSKY, 1980 Chicken lens crystalline DNA sequences show at least two δ-crystalline genes. Nature **284:** 234–238.

BORE, 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.

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

DEJONG, G., and W. SCHARLOO, 1976 Environmental determination of selective significance or neutrality of amylase variants in *Drosophila melanogaster*. Genetics **84:** 77–94.

DEJONG, G., A. J. W. HOORN, G. E. W. THORIG and W. SCHARLOO, 1972 Frequencies of amylase variants in *Drosophila melanogaster*. Nature **238:** 542–453.

DOANE, W. W., 1969 Amylase variants in *Drosophila melanogaster:* linkage studies and characterization of enzyme extracts. J. Exp. Zool. **171:** 321–342.

DOANE, W. W., R. M. GEMMILL, P. E. SCHWARTZ, S. A. HAWLEY and R. A. NORMAN, 1987 Structural organization of alpha-amylase gene locus in *Drosophila melanogaster* and *Drosophila miranda*. Isozymes Curr. Top. Biol. Med. Res. **14:** 229–266.

FELSENSTEIN, J., 1993 PHYLIP (phylogeny inference package) version 3.5c. Distributed by the author, Department of Genetics, University of Washington, Seattle.

FRICHAUF, 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 alpha-amylase genes from *Drosophila melanogaster*. I. Clone isolation by use of 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.

HICKEY, D. A., 1977 Selection for amylase allozymes in *Drosophila melanogaster*. Evolution **31:** 800–804.

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. ABUKASHAWA, V. PAYANT and B. F. BENKEL, 1991 Concerted evolution of duplicated protein-coding genes in *Drosophila*. Proc. Natl. Acad. Sci. USA **88:** 1611–1615.

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, 1972 On the stochastic model for estimation of mutational distance between homologous proteins. J. Mol. Evol. **2:** 87–90.

KREITMAN, M., 1983 Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. Nature **304:** 412–417.

KREITMAN, M., and R. R. HUDSON, 1991 Inferring the evolutionary histories of the *Adh* and *Adh-dup* loci in *Drosophila melanogaster* from patterns of polymorphism and divergence. Genetics **127:** 565–582.

KYTE, J., and R. F. DOOLITTLE, 1982 A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157:** 105–132.

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

LEIGH BROWN, A. J., and D. ISH-HOROWICZ, 1981 Evolution of the 87A and 87C heat-shock loci in Drosophila. Nature **290:** 677–682.

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

MATSUO, Y., and T. YAMAZAKI, 1984 Genetic analysis of natural populations of *Drosophila melanogaster* in Japan. IV. Natural selection on the inducibility, but not on the structural genes, of amylase loci. Genetics **108:** 879–896.

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

NAGYLAKI, T., 1984 Evolution of multigene families under interchromosomal gene conversion. Proc. Natl. Acad. Sci. USA **81:** 3796–3800.

NAKANISHI, S., A. INOUE, T. KITA, M. NAKAMURA, A. C. Y. CHANG *et al.*, 1979 Nucleotide sequences of cloned cDNA for bovine corticotropin-β-lipotropin precursor. Nature **278:** 423–427.

NEI, M., 1987 *Molecular Evolutionary Genetics*. Columbia University Press, New York.

NEI, M., and W.-H. LI, 1979 Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. USA **76:** 5269–5273.

NEI, M., and F. TAJIMA, 1981 DNA polymorphism detectable by restriction endonucleases. Genetics **97:** 145–163.

OHTA, T., 1983 On the evolution of multigene families. Theor. Popul. Biol. **23:** 216–240.

OKUYAMA, E., and T. YAMAZAKI, 1988 Nucleotide sequences of the duplicated amylase structural genes 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 and J. DAVID, 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.

POWELL, J. R., and M. ANDJELKOVIC, 1983 Population genetics of Drosophila amylase. IV. Selection in laboratory populations maintained on different carbohydrates. Genetics **103:** 675–689.

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. **143:** 161–178.

SAWYER, S., 1989 Statistical tests for detecting gene conversion. Mol. Biol. Evol. **6:** 526–538.

SHIBATA, H., and T. YAMAZAKI, 1995 Molecular evolution of the duplicated *Amy* locus in *Drosophila melanogaster* species subgroup: concerted evolution only in coding region and an excess of nonsynonymous substitutions in speciation. Genetics **141:** 223–236.

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

WAHLI, W., I. B. DAWID, G. U. RYFFEL and R. WEBER, 1981 Vitellogenesis and the vitellogenin gene family. Science **212:** 298–304.

WATTERSON, G. A., 1975 On the number of segregating sites in genetic models without recombination. Theor. Popul. Biol. **7:** 256–276.

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