J Mol Evol (2003) 56:121–130
DOI: 10.1007/s00239-002-2384-3

JOURNAL OF MOLECULAR EVOLUTION

Phylogeny and the Evolution of the *Amylase* Multigenes in the *Drosophila montium* Species Subgroup

Ze Zhang,[1,*] Nobuyuki Inomata,[1] Marie-Louise Cariou,[2] Jean-Luc Da Lage,[2] Tsuneyuki Yamazaki[1]

[1] Laboratory of Molecular Population Genetics, Department of Biology, Graduate School of Science, Kyushu University, Fukuoka 812-8581, Japan
[2] Populations, Genetique et Evolution, Centre National de la Recherche Scientifique, 91198 Gif sur Yvette Cedex, France

Received: 19 March 2002 / Accepted: 6 August 2002

Abstract. To investigate the phylogenetic relationships and molecular evolution of α-amylase (*Amy*) genes in the *Drosophila montium* species subgroup, we constructed the phylogenetic tree of the *Amy* genes from 40 species from the *montium* subgroup. On our tree the sequences of the *auraria*, *kikkawai*, and *jambulina* complexes formed distinct tight clusters. However, there were a few inconsistencies between the clustering pattern of the sequences and taxonomic classification in the *kikkawai* and *jambulina* complexes. Sequences of species from other complexes (*bocqueti*, *bakoue*, *nikananu*, and *serrata*) often did not cluster with their respective taxonomic groups. This suggests that relationships among the *Amy* genes may be different from those among species due to their particular evolution. Alternatively, the current taxonomy of the investigated species is unreliable. Two types of divergent paralogous *Amy* genes, the so-called *Amy1*- and *Amy3*-type genes, previously identified in the *D. kikkawai* complex, were common in the *montium* subgroup, suggesting that the duplication event from which these genes originate is as ancient as the subgroup or it could even predate its differentiation. Thc *Amy1*-type genes were closer to the *Amy* genes of *D. melanogaster* and *D. pseudoobscura* than to the *Amy3*-type genes. In the *Amy1*-type genes, the loss of the ancestral intron occurred independently in the *auraria* complex and in several Afrotropical species. The GC content at synonymous third codon positions (GC3s) of the *Amy1*-type genes was higher than that of the *Amy3*-type genes. Furthermore, the *Amy1*-type genes had more biased codon usage than the *Amy3*-type genes. The correlations between GC3s and GC content in the introns (GCi) differed between these two *Amy*-type genes. These findings suggest that the evolutionary forces that have affected silent sites of the two *Amy*-type genes in the *montium* species subgroup may differ.

Key words: Amylase — Gene duplication — Phylogeny — *Drosophila montium* — Codon bias

Introduction

The *Amylase* (*Amy*) genes encoding α-amylase proteins, which break down starch into glucose and maltose, constitute a relatively small multigene family in *Drosophila* species. Different species have from two to at least seven copies of these genes (Bahn 1967; Brown et al. 1990; Da Lage et al. 1992; Shibata and Yamazaki 1995; Popadic et al. 1996; Steinemann and Steinemann 1999; Inomata and Yamazaki 2000; Da Lage et al. 2000). The *Amy* genes in *Drosophila* provide a good model for the study of the adaptive evolution of multigene families. This is because the

*Current address: Laboratory of Biometry and Bioinformatics, Graduate School of Agriculture and Life Sciences, University of Tokyo, Yayoi 1-1-1, Bunkyo-ku, Tokyo 113-8657, Japan
Correspondence to: Nobuyuki Inomata; *email:* ninomscb@mbox.nc.kyushu-u.ac.jp

Table 1. Data on species used in this study

Complex	Species	Stock No.[a]	Distribution[b]	Clone No.[c]	*Amy* type	Intron (bp)	GCi (%)	GC3s (%)	ENC
auraria	*D. auraria*	Shanghai-1	Ori (Shanghai)	6	1	—	—	94.2	26.7
		NS	Ori	US23	1	—	—	93.4	27.1
	D. biauraria	NS	Ori	IS15	1	—	—	94.2	26.4
				IS16	1	—	—	94.7	26.1
				IR9	3	67	40.3	66.0	48.2
	D. quadraria	NIG	Ori	23	1	—	—	93.8	26.3
				50	1	—	—	94.2	26.0
		NS	Ori	IS16	1	—	—	93.4	26.1
	D. subauraria	NS	Ori (Sapporo)	6	1	—	—	95.1	26.0
	D. triauraria	Yakushima-3	Ori (Yakushima Is.)	8	1	—	—	94.7	25.4
		Nippara	Ori (Tokyo)	3	1	—	—	92.6	27.1
		NS	Ori	IS2	1	—	—	93.4	26.5
kikkawai	*D. bocki*	A65	Ori (Taiwan)	1*	1	63	47.6	88.9	29.1
				2*	1	59	49.2	88.5	29.4
				3*	3	67	37.3	67.9	42.8
				4*	3	67	38.8	65.4	45.8
	D. cauverii	PGE	Ori (India)	2	1	65	43.1	84.0	31.3
	D. kikkawai	Naha-1	Ori (Okinawa)	1*	1	63	46.0	87.7	30.6
				2*	1	63	46.0	88.1	30.5
				3*	3	63	38.1	67.9	43.6
				4*	3	63	38.1	67.9	43.6
	D. leontia	RGN210-8	Ori (Rangoon)	1*	1	63	47.6	87.3	29.4
				2*	1	63	46.0	86.9	30.1
				3*	3	67	37.3	65.8	43.5
				4*	3	67	37.3	67.5	42.9
	D. lini	14028-0581.0	Ori (Taiwan)	1*	1	63	44.4	89.3	29.2
				2*	1	63	46.0	86.9	30.1
				3*	3	67	38.8	69.5	40.5
		NS	Ori	IS1	1	63	46.0	85.6	31.7
				IR1	3	67	38.8	68.7	40.7
				IR5	3	67	40.3	68.3	42.1
				IR6	3	67	38.8	68.3	41.0
	D. lini-like	NS	Ori	8	1	63	42.9	88.5	29.1
	D. diplacantha	14028-0586.0	Afr (Cameroun)	2	1	63	47.6	89.3	28.3
				4	1	63	47.6	89.3	28.5
jambulina	*D. jambulina*	NS	Ori (India)	US17	1	64	48.4	91.0	28.1
				US5	1	64	50.0	92.2	27.4
				US1	3	68	45.6	75.4	37.1
	D. punjabiensis	MDL177	Ori (Myanmar)	1	1	57	40.4	80.7	34.2
		14028-0531.0	Ori (Thailand)	24	1	55	41.8	92.2	27.6
				k46	3	66	40.9	64.3	45.2
		NS	Ori	IR4	1	55	36.4	91.0	27.5
				IS6	1	55	36.4	89.8	28.0
				IS5	1	56	42.9	89.3	29.1
				IR3	3	56	41.1	70.1	41.9
				IF1	3	67	42.4	67.6	43.5
				IF2	3	67	42.4	67.6	43.5
	D. watanabei	SWB248	Ori	18	1	63	38.1	92.2	27.9
				k46	3	66	40.9	65.2	44.0
serrata	*D. serrata*	14028-0681.0	Aus (Queensland)	1	1	64	48.4	77.5	34.1
		NS	Aus	US5	3	68	42.7	72.5	40.8
	D. birchii	14028-0521.0	Aus (Cairns)	21	1	64	46.9	82.8	32.5
				25	1	59	45.8	83.6	32.2
bocqueti	*D. bocqueti*	L139	Afr	21	1	64	48.4	89.3	28.3
				23	1	64	48.4	89.8	28.3
	D. bocqueti-like	PGE	Afr (Congo)	3	1	64	48.4	90.2	28.3
	D. burlai	L111	Afr	7	1	64	46.9	82.8	32.5
				8	1	64	46.9	82.8	32.5
bakoue	*D. greeni*	PGE168-5	Afr (Ivory Coast)	15	1	—	—	96.3	25.0
				IS8	1	—	—	96.7	25.0
	D. vulcana	14028-0711.0	Afr (Rhodesia)	9	1	64	45.3	91.4	27.8
	D. seguyi	K2	Afr (Kenya)	13	1	68	45.6	93.0	27.4
	D. tsacasi	14028-0701.0	Afr (Ivory Coast)	4	1	64	48.4	90.2	28.3

nikananu	*D. nikananu*	14028-0601.0	Afr (Ivory Coast)	1	1	64	46.9	86.9	29.4
		NS	Afr	IS11	1	64	45.3	86.1	30.2
				IF5	3	67	43.3	72.1	41.2
				1S3	3	67	43.3	73.4	37.8
kikkawai or *jambulina*	*D. barbarae*[d]	CTB207	Ori (India)	1	1	67	44.8	84.8	31.1
Others									
	D. baimaii	14028-0481.0	Ori (Malaysia)	25	1	67	58.2	97.5	24.6
				11	1	65	44.6	85.2	30.4
	D. bicornuta	14028-0511.1	Ori (Taiwan)	IS1	1	64	43.8	87.7	29.2
				IS3	1	64	43.8	85.2	30.6
	D. davidi	PGE183.1	Afr (Congo)	27	1	65	47.7	88.1	29.4
	D. dossoui	PGE274	Afr (Benin)	7	1	65	47.7	88.1	29.4
				10	1	—	—	89.3	28.7
				US24	1	—	—	89.3	28.6
				IF1	3	67	40.3	67.2	43.2
				1	3	67	40.3	67.6	43.2
				IF5	3	67	40.3	67.3	42.8
				9	3	67	40.3	66.8	43.5
				IR1	3	67	40.3	66.8	43.7
	D. kanapiae	14028-0541.0. dup	Ori (Philippines)	3-3	1	57	38.6	88.1	29.3
				14	3	67	35.8	67.6	46.8
	D. khaoyana	NS	Ori (Taiwan)	19	1	—	—	89.8	27.1
	D. lacteicornis	14028-0571.0	Ori (Okinawa)	32	1	64	51.6	92.1	26.6
	D. mayri	14028-0591.0	Aus (New Guinea)	6	1	66	48.5	83.6	33.0
	D. nagarholensis	PGE	Ori (India)	8	1	56	41.1	88.5	29.2
				k46	3	66	42.4	67.6	43.5
	D. orosa	14028-0611.0	Ori (Thailand)	3	1	64	43.8	90.2	29.9
	D. parvula	CNX312	Ori	18	1	57	45.6	86.5	29.4
	D. rufa	NS	Ori (Miyakejima Is.)	3-13	1	63	55.6	91.7	26.4
				5	3	66	34.9	63.9	49.2
		NS	Ori	IS2	1	—	—	86.8	28.8
				US13	1	63	55.56	91.3	26.4
				IR2	3	66	34.9	64.3	49.7
				IR4	3	66	33.3	64.3	50.2
	D. tani	NS	Ori	9	1	—	—	84.4	29.4
	D. truncata	RGN210-17	Ori	14	1	62	46.8	85.7	31.4

Note. Species used are listed in alphabetical order within each complex. Unclassified species are listed at the bottom. The geographic distribution and origin of the populations studied (parentheses), intron size, GC content of the intron (GCi), GC content at third codon positions (GC3s), and codon usage bias (ENC) are indicated.

[a] NS indicates that there is no stock number.

[b] Ori, Oriental; Afr, Afrotropical; Aus, Australasian.

[c] Sequences already published (Inomata and Yamazaki 2000) are indicated by an asterisk.

[d] *D. barbarae* belongs to the *kikkawai* (Lemeunier et al. 1986) or the *jambulina* (Kim et al. 1989) complex.

proteins encoded by these genes interact directly with food environments. Amylase activity is repressed by dietary glucose (Benkel and Hickey 1986; Inomata et al. 1995) but induced by dietary starch (Inomata et al. 1995). Thus, differences in gene structure, regulation, or the nature of proteins can be more readily interpreted in terms of adaptation. Indeed, there have been an increasing number of studies devoted to the evolution of the *Amy* multigene family in *Drosophila*. In particular, the organization and molecular evolution of the *Amy* multigene family in the *melanogaster* species subgroup and several other species have been characterized in detail (Doane et al. 1987; Brown et al. 1990; Hickey et al. 1991; Da Lage et al. 1992; Shibata and Yamazaki 1995; Popadic et al. 1996; Inomata et al. 1997; Steinemann and Steinemann 1999; Inomata and Yamazaki 2000; Da Lage et al. 2000). Furthermore, most previous studies suggested that the members of the *Amy* multigene family evolved in a concerted fashion. Da Lage et al. (1998) found a new paralogous amylase gene, *Amyrel*, that has diverged from the *Amy* gene to a striking extent. However, its function remains unknown. Recently, Inomata and Yamazaki (2000) also found divergent paralogous *Amy* genes encoding active amylase isozymes in *Drosophila kikkawai* and its sibling species. *Drosophila kikkawai* has two duplicated paralogous gene groups (*Amy1–Amy2* and *Amy3–Amy4* genes) located at different chromosomal positions. The sequences were very similar within each group, but the two groups diverged significantly from each other. The *Amy1*-type genes were more similar to the *Amy* gene of *D. melanogaster*. Regulatory changes between the two groups of genes occurred after gene dupli-

cation (Inomata and Yamazaki 2000). We therefore decided to investigate the occurrence and evolution of these two types of *Amy* genes throughout the whole *montium* species subgroup.

The wide geographical distribution and the large number of species make the *montium* species subgroup an attractive system for evolutionary studies. The *montium* subgroup includes 81 of the 156 known species of the *melanogaster* species group. It is distributed throughout Northeast Asia (Japan, Korea, and China), the South Pacific Islands (Borneo, Sumatra, Java, and Australia), and Indian and Afrotropical areas (Lemeunier et al. 1986; Ashburner 1989). The *montium* subgroup is divided into seven species complexes, *auraria*, *bakoue*, *bocqueti*, *jambulina*, *kikkawai*, *nikananu*, and *serrata*. It also contains a number of unclassified species that are too poorly known for their affinities to be apparent (Lemeunier et al. 1986; Ashburner 1989; Kim et al. 1989). Many studies of the phylogenetic relationships among the *montium* species have already been done using morphological, genetic (crossing), biochemical, mtDNA, and nucleotide sequencing tests (Ayala 1965; Bock and Wheeler 1972; Triantaphyllidis et al. 1978; Bock 1980; Ohnishi et al. 1983; Ohnishi and Watanabe 1984; Tsakas and Tsacas 1984; Kalantzi-Makri et al. 1985; Kim et al. 1989, 1993; Lamnissou and Zouros 1989; Pissios and Scouras 1993; Nikolaidis and Scouras 1996). All these studies, however, have limitations: (1) the number of species used was relatively small; and (2) because the species used in each study were different, the congruence of data cannot be confirmed.

In this study we determined 83 sequences of the *Amy* genes in 36 representative species of the *montium* species subgroup. In total we analyzed 98 sequences of the *Amy* genes from the 40 *montium* species. The main purpose of this study is to understand the structure of the *Amylase* gene family in the *montium* species subgroup, to investigate patterns of molecular evolution of the members of this family, and to draw inferences about the relationships between species based on the phylogeny of the genes. Our findings suggest that different evolutionary forces have affected the two *Amy*-type gene groups.

Materials and Methods

Drosophila *Species and DNA Sequence Data*

Table 1 lists the *Drosophila* species used in this study, along with the species complexes and distribution, intron size, GC content of the intron (GCi), GC content at synonymous third codon positions (GC3s), and codon bias (ENC). *Drosophila auraria*, *D. quadraria*, *D. rufa*, *D. subauraria*, *D. triauraria*, *D. lini-like, D. punjabiensis*, *D. watanabei*, *D. bocqueti*, *D. burlai*, *D. seguyi*, *D. barbarae*, *D. khaoyana*, *D. parvula*, *D. tani*, and *D. truncata* were obtained from Tokyo Metropolitan University. *Drosophila lacteicornis*, *D. diplacantha*, *D. punjabiensis*, *D. serrata*, *D. birchii*, *D. vulcana*, *D. tsacasi*, *D. nikananu*, *D. baimaii*, *D. bicornuta*, *D. kanapiae*, *D. mayri*, and *D. orosa* were obtained from the National *Drosophila* Species Resource Center (NDSRC) at Bowling Green State University, Bowling Green, OH. On the basis of a morphological analysis, the species with the stock number 14028-0531.0 was regarded in this study as *D. punjabiensis*, even though it was recorded as *D. jambulina*. *Drosophila auraria*, *D. biauraria*, *D. quadraria*, *D. rufa*, *D, triauraria*, *D. cauverii*, *D. jambulina*, *D. punjabiensis*, *D. serrata*, *D. bocqueti-like, D. greeni*, *D. nikananu*, *D. davidi*, *D. dossoui,* and *D. nagarhplensis* were from the Populations, Genetique et Evolution (PGE) stock of the Centre National de la Recherche Scientifique (CNRS), France. Sequences of different strains of *D. auraria*, *D. quadraria*, *D. rufa*, *D. triauraria*, *D. lini*, *D. punjabiensis*, *D. serrata*, and *D. nikananu* from Tokyo Metropolitan University, the NDSRC, or the PGE collection were determined. The complete *Amy* sequences of *D. kikkawai*, *D. bocki*, *D. leontia,* and *D. lini* (accession numbers AB035055–AB035069) were obtained from Inomata and Yamazaki (2000). The *Amy* sequence of *D. virilis* (accession number U02029), the *Amy* sequence of *D. melanogaster* (accession number L22730), and the sequence of *D. pseudoobscura* (accession number X76240) were used as references. The accession numbers of the 83 *Amy* sequences obtained in this study are AB048693–AB048700, AB048763–AB048787, AB048800–AB048810, AF119471–AF119474, AF119477–AF119484, AF119486–AF119488, AF119492–AF119495, and AF393786–AF393805.

Extraction of Genomic DNA, PCR Amplification, and DNA Sequencing

Five adult flies were homogenized and suspended in a 100 mM NaCl, 100 mM Tris–HCl (pH 8.0), and 100 mM EDTA solution with proteinase-K. After incubation for 30 min at 37°C, the genomic DNA was extracted with phenol/chloroform, precipitated with ethanol, and suspended in TE solution. The *Amy* genes were amplified by the PCR method with synthetic oligonucleotide primers (Table 2). The ancestral intron (Da Lage et al. 1996; Inomata et al. 1997) was always included when present. The reaction conditions for 32 to 40 cycles were denaturation 95 or 94°C for 30 s, annealing at 55 to 60°C for 30 s, and polymerization at 72°C for 1 or 2 min. The amplified fragments were cloned into pGEM-T vector. To eliminate PCR artifacts and to detect all the *Amy* gene types, more than two clones (2 to 10) were sequenced for both DNA strands. For the *Amy3*-type genes of *D. nagarholensis* and *D. watanabei*, we purified PCR products and sequenced them directly. Sequencing was performed using an ABI Model 377 automated sequencer with a DNA sequencing kit (BigDye terminator cycle sequencing ready reaction; ABI) and synthetic oligonucleotide primers.

Data Analyses

We studied 98 *Amy* sequences of the *montium* species subgroup. Because the *Amy* gene sequences differed in length, we analyzed a common part of the exon, 759 bp long, from position +175 to position +933 (the first Met is +1), and the intron. The position of the ancestral intron is at the 177th site (Da Lage et al. 1996; Inomata et al. 1997). The Clustal W program (Thomson et al. 1994) was used to align sequences and to produce a neighbor-joining (NJ) tree (Saitou and Nei 1987) based on distances estimated by the two-parameter method (Kimura 1980) with 1000 bootstrap samples. The codon bias measured by ENC (Wright 1990) and the GC content of introns (GCi) and at synonymous third codon positions (GC3s) were calculated using the DnaSP program version 3.14 (Rozas and Rozas 1999).

Because a phylogeny of the sequences of a particular gene is hierarchically structured, we cannot regard them as independent

Table 2. PCR primers used in this study[a]

Primer		Position of the 3′ end
Forward		
intr1	5′-AGTGCGAGAACTTCCTGGG-3′	+152
intr1U	5′-GTTCACCTCTTCGAGTGG-3′	+111
k46for	5′-GCTTCCTGGCCTTGGCTACT-3′	+51
Reverse		
rev1230	5′-TTGCTGCCGTTGTCCCACC-3′	+1220
Steph	5′-GGAGGCCATCTTGTACTGC-3′	+975
Finclas	5′-CTTGGCGTTGACTTGAATGG-3′	+1460

[a] The first nucleotide of the start codon (ATG) is +1. Nucleotide positions refer to *D. melanogaster* sequences (+1 to +1482).

samples for statistical purposes. To eliminate the effect of nonindependence on the correlation, we used Felsenstein's (1985) pairwise independent contrast test. This test was performed with the CONTRAST program in the computer package PHYLIP 3.5 (Felsenstein 1993). Given a phylogenetic tree with n sequences, a total of $n-1$ independent components ("contrasts" in terms of Felsenstein's test) can be obtained for each of the characters X and Y. We investigated the association between the independent components of X and those of Y. Three NJ trees were constructed: (i) for 68 sequences of the *Amy1*-type genes with and without the intron, (ii) for 50 sequences of the *Amy1*-genes with the intron, and (iii) for 27 sequences of the *Amy3*-type genes (excluding three completely identical sequences). These trees were used to estimate the correlation coefficients by Felsenstein's pairwise independent contrast test. The tree including 68 *Amy1*-type sequences and that including 27 *Amy*3-type sequences were used separately to estimate the correlation coefficients between the contrast values of ENC and those of GC3s. The tree including 50 *Amy1*-type sequences with the intron and that including 27 *Amy3*-type sequences were also used separately to estimate the correlation coefficients between the contrast values of GCi and those of GC3s.

Results

Molecular Structure and Phylogeny of the Amy *Genes in the* montium *Species Subgroup*

We constructed an NJ tree of 98 partial sequences of the *Amy* genes, including 15 published sequences (Inomata and Yamazaki 2000), from 40 *montium* species. The sequences of the *Amy* genes of *D. melanogaster* (*melanogaster* species subgroup), *D. pseudoobscura* (*obscura* species group), and *D. virilis* (subgenus *Drosophila*) were included. Our tree revealed two types of widely divergent *Amy* genes in the *montium* species subgroup (Fig. 1). They correspond to the two duplicated paralogous groups of genes, *Amy1–Amy2* and *Amy3–Amy4*, previously found in *D. kikkawai* and its siblings (Inomata and Yamazaki 2000). By comparing partial sequences the *Amy1* gene cannot be distinguished from the *Amy2* gene. Similarly, the *Amy3* gene cannot be distinguished from the *Amy4* gene. Therefore, for convenience, the *Amy1* and *Amy2* genes are referred to as *Amy1*-type genes and the *Amy3* and *Amy4* genes are referred to as *Amy3*-type genes.

Several clusters supported by high bootstrap values indicated clear differentiation of the *Amy1*-type genes among species (Fig. 1). Seventeen sequences of the *Amy1*-type genes formed a monophyletic clade supported by a high bootstrap value (92%). It includes the *auraria* complex *sensu stricto* and four non-*auraria* complex species: *D. khaoyana*, *D. tani*, *D. rufa*, and *D. lacteicornis*. Within the *auraria* complex (*D. auraria*, *D. biauraria*, *D. quadraria*, *D. subauraria*, *D. triauraria*), all *Amy1*-type genes are intronless and closely related. *Drosophila khaoyana* and *D. tani* also have intronless *Amy1*-type genes. *Drosophila rufa* had genes both with and without an intron. The three genes with the intron of *D. lacteicornis* and *D. rufa* were the first to cluster, and then they clustered together with two intronless genes of *D. rufa* and *D. tani*. However, the cluster of these five genes had low bootstrap support (42%), suggesting that all the intronless genes within this cluster composed of the 17 sequences could be monophyletic. Another very tight cluster with strong bootstrap support (99%) contained 10 genes with the intron from the *kikkawai* complex. The *Amy1*-type genes of *D. kikkawai*, *D. leontia*, and *D. bocki* clustered first, followed by those of *D. lini* and *D. lini-like*. This branching order was consistent with results from crossing experiments, 2DE-electrophoretic analysis of proteins, and mitochondrial DNA analysis (Ohnishi and Watanabe 1984; Kim et al. 1989, 1993). Similarly, the *Amy1*-type genes from the *jambulina* complex, *D. nagarholensis*, *D. punjabiensis*, and *D. watanabei*, formed a very tight cluster with a 100% bootstrap value. This cluster matched the taxonomic grouping, except that the *Amy1*-type genes of *D. jambulina* were not included in the cluster of the *jambulina* complex (see Fig. 1).

For the remaining *Amy1*-type genes, the grouping was more or less unrelated to the taxonomic classification. The last well-supported cluster (bootstrap value, 88%) included genes from species belonging to different taxonomic complexes (*serrata*, *bocqueti*) and some unclassified species, such as *D. mayri*, *D. bicornuta*, and *D. truncata*. There were no clear branching orders for these *Amy1*-type genes, indicating a lack of strong differentiation.

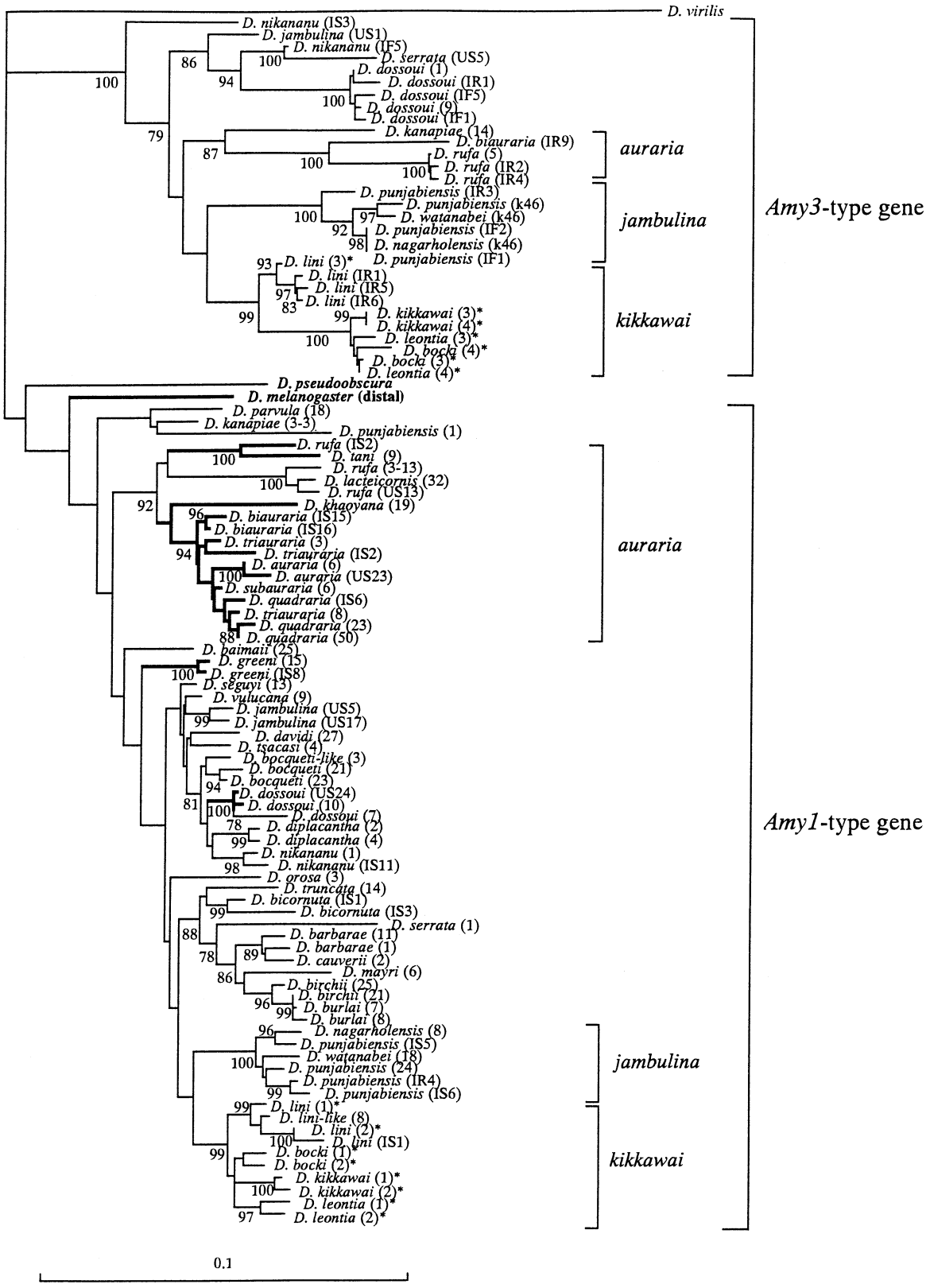

Fig. 1. Phylogenetic tree based on 1000 neighbor-joining trees from bootstrapped data sets. Bootstrap values higher than 75% are shown on the corresponding nodes. An *asterisk* indicates that the sequence was obtained from a genome library (Inomata and Yamazaki 2000). The *boldface* branches indicate sequences without the intron. Numbers and symbols in parentheses following species names indicate clone numbers or sequence symbols.

The distribution of the intronless *Amy1*-type genes revealed in our study suggests that the loss of the ancestral intron has occurred independently several times but exclusively in the *Amy1*-type genes from the *montium* species subgroup. Although, as described above, in the cluster composed of the 17 sequences, in which the *auraria* complex species are involved, the intronless *Amy1*-type genes might have a single origin, the intron of the *Amy1*-type gene has also been lost in *D. greeni* and *D. dossoui*, which belong to different complexes and are scattered within the tree (see Fig. 1). Independent loss of the ancestral intron has occurred in several *Drosophila* lineages (Da Lage et al. 1996; Inomata et al. 1997). The size of the intron ranged from 55 to 68 bp in the *Amy1*-type genes (Table 1). The *Amy* genes of *D. melanogaster* and *D. pseudoobscura* were closer to *Amy1*-type genes.

The *Amy3*-type genes (see Table 1 and Fig. 1) were detected in 14 *montium* species, including species reported by Inomata and Yamazaki (2000), which belong to different complexes and come from different geographical areas (oriental and Afrotropical). This suggests that the *Amy3*-type genes are very common throughout the *montium* species subgroup. All *Amy3*-type genes had the ancestral intron, ranging in size from 56 to 68 bp, which had an exon–intron structure consistent with that reported previously (Doane et al. 1987; Brown et al. 1990; Inomata and Yamazaki 2000). All *Amy3*-type genes diverged widely from the *Amy1*-type and the *Amy* genes of *D. melanogaster* and *D. pseudoobscura*. They formed a statistically well-supported monophyletic cluster (100% bootstrap value), indicating that they had a single origin. The *Amy3*-type genes from the *auraria*, *jambulina*, and *kikkawai* complexes formed very tight clusters with high bootstrap values (see Fig. 1). Their branching orders were similar to those of the *Amy1*-type genes. The *Amy3*-type genes of *D. jambulina* were, once again, not included in the cluster of the *jambulina* complex.

Patterns of Molecular Evolution in the Two Amy-*Type Genes*

The GC content at synonymous third codon positions (GC3s), GC content of the intron (GCi), and codon usage bias in terms of ENC (Wright 1990) are listed in Table 1. The GC3s of the *Amy1*-type genes was strikingly higher than that of the *Amy3*-type genes, ranging from 77.5% (*D. serrata*) to 97.5% (*D. baimaii*) in the *Amyl*-type genes (average, 89.2%) and from 63.9% (*D. rufa*) to 75.4% (*D. jambulina*) in the *Amy3*-type genes (average, 67.8%). The codon usage of the *Amyl*-type genes was more biased than that of the *Amy3*-type genes, with the ENC ranging from 24.6 (*D. baimaii*) to 34.2 (*D. punjabiensis*) in the *Amy1*-type genes (average, 28.8) and from 37.1 (*D. jambulina*) to 50.2 (*D. rufa*) in the *Amy3*-type genes (average, 43.5). The GCi of the *Amy1*-type genes was slightly higher than that of the *Amy3*-type. genes. It ranged from 36.4% (*D. punjabiensis*) to 58.2% (*D. baimaii*) in the *Amy1*-type genes (average, 45.8%) and from 33.3% (*D. rufa*) to 45.6% (*D. jambulina*) in the *Amy3*-type genes (average, 39.5%).

Because the sequences used in this study generally cannot be regarded as being independent, we used Felsenstein's (1985) independent contrast test to study the relationships among GC3s, GCi, and ENC. The *Amy1*- and *Amy3*-type genes were analyzed separately (see Materials and Methods). ENC was negatively correlated with GC3s ($r = -0.894$, df = 65, $p \ll 0.01$ in *Amy1* and $r = -0.745$, df = 24, $p \ll 0.01$ in *Amy3*). The *Amy3*-type genes showed a weak positive correlation between GC3s and GCi ($r = 0.349$, df = 24, $0.05 < p < 0.10$), but no correlation was found in the *Amy1*-type genes ($r = 0.133$, df = 47, $p > 0.10$). These findings indicate that the evolutionary patterns of these two *Amy*-type gene groups might be different.

Discussion

Taxonomic Inferences from the Amylase *Gene Tree*

The present study included 40 species from the *montium* species subgroup, which is about half the subgroup. A gene tree is not always consistent with a species tree, and a gene tree constructed from multigenes sometimes yields a complex topology. However, the phylogenetic tree of the *Amy* genes presented here gives at least a general picture of phylogenetic relationships among the *montium* species.

First, the sequence data for the *Amy1*- and *Amy3*-type genes are generally consistent with the species designation, and three main clusters, *auraria*, *jambulina*, and *kikkawai*, supported by high bootstrap values, are mainly in agreement with the taxonomically accepted relationships. We found that the *jambulina* complex is closer to the *kikkawai* complex than to the *auraria* complex. This is consistent with previous studies (Ohnishi et al. 1983; Ohnishi and Watanabe 1984; Kim et al. 1989, 1993). Several species in these three complex clusters require more detailed studies. An unexpected finding is that the *auraria* complex clearly comprises two sublineages, the *auraria* sublineage *sensu stricto* and the *rufa–tani–lacteicornis* sublineage. However, the positions of *D. khaoyana* and *D. tani* on the tree, combined with the presence of intronless *Amy1*-type genes, suggest that they are closely related to the *D. auraria* sibling species, even though previously they had not been assigned to the *auraria* complex. We suggest that *D. nagarholensis* should be assigned to the *jambulina*

complex, because it formed a very tight cluster with *D. punjabiensis* and *D. watanabei* (*jambulina* complex), with 100% bootstrap value for both the *Amy1*- and the *Amy3*-type genes. However, two *D. jambulina* and one *D. punjabiensis Amy1*-type genes were not included in this cluster and had very low bootstrap values. These *Amy* genes should be investigated further to determine whether they correspond to another paralogous gene. The *kikkawai* complex shows a consistent grouping of *D. kikkawai*, *D. leontia*, *D. bocki*, and *D. lini*. However, although *D. barbarea* and *D. cauveri* are included in this complex, they are included in a poorly resolved cluster composed of species belonging to different taxonomic complexes and a few unclassified species. It was originally suggested on the basis of the morphology of its male genitalia that *D. barbarae* belonged to the *kikkawai* complex (Lemeunier et al. 1986). Later, Kim et al. (1989) assigned this species to the *jambulina* complex on the basis of cross experiments. However, in our tree *D. barbarae* did not cluster with either the *kikkawai* complex or the *jambulina* complex. Further investigation is therefore called for to determine the position of this species. The position of *D. diplacantha* is consistent with its morphological/taxonomic characters, which show that the species is clearly distinct from other members of the *kikkawai* complex. This suggests that it should be removed from the *jambulina* complex (Lachaise, personal communication).

Second, species belonging to other complexes had more uncertain positions and did not form clear complex clusters. The taxonomy of many species within the *montium* subgroup is difficult and often relies on tenuous morphological characters. Our tree suggests that *D. birchii* (belonging to the *serrata* complex) and *D. burlai* (belonging to the *bocqueti* complex) are closely related (bootstrap value, 97%). The Afrotropical species tend to cluster together (see Fig. 1), although *D. greeni* (from the *bakoue* complex) and *D. burlai* were not included in the Afrotropical cluster and had low bootstrap support. Note that *D. jambulina* is also included in this cluster, despite the fact that it is not found in an Afrotropical area.

Molecular Evolution of the Amy *Genes*

Inomata and Yamazaki (2000) found that there were two types of widely divergent paralogous *Amy* genes that encode active amylase isozymes in *D. kikkawai* and its sibling species and that the regulatory changes between these two gene types have occurred in association with gene duplication. A similar case was also found in *D. ananassae* (Da Lage et al. 2000). In the *montium* species subgroup to which *D. kikkawai* belongs, we found the same two types of *Amy* genes in 10 species from different complexes and different geographic areas. The parsimonious hypothesis based on the data is that the two *Amy*-type genes originated through duplication at the origin of the subgroup. However, we cannot rule out the possibility that the duplication predates the diversification of the *montium* subgroup. Within the *melanogaster* group, an investigation *Drosophila* species belonging to the Oriental subgroups could give new information on this subject.

Our tree suggests that the loss of the ancestral intron had occurred only in the species in the *auraria* complex and in several Afrotropical species. *Drosophila rufa* and *D. dossoui* have three types of *Amy* genes: *Amy1*-type genes with and without the intron and *Amy3*-type genes. *Drosophila biauraria* had the intronless *Amy1*-type gene and the *Amy*3-type gene. The *Amy3*-type genes always included the ancestral intron. Together with previous observations (Da Lage et al. 1996; Inomata et al. 1997), this suggests that the loss of the intron has occurred fairly often in the evolution of the *Amy* genes in *Drosophila*. The mechanism by which introns are lost is poorly understood. Kliman and Eyre-Walker (1998) have suggested that in *D. melanogaster*, genes with and without introns show strikingly different trends in base composition at silent sites. We found a similar trend in the *Amy* genes analyzed in the present study. The average GC3s of the *Amy1*-type genes with the intron (50 genes) and without it (18 genes) was 88 and 93%, respectively. The evolution of the *Amy* genes with and without the intron is an interesting topic which remains to be studied.

ENC was negatively correlated with GC3s. This is consistent with previous reports, which indicated that in *Drosophila* the base composition at silent sites is positively correlated with the codon usage bias among genes as well as among species for the same genes (Shield et al. 1988; Moriyama and Hartl 1993; Inomata et al. 1997; Akashi et al. 1998). We found not only striking interspecific variations within the same gene type, but also apparent intraspecific variations between the two groups of paralogous genes with regard to GC3s and codon bias. There are at least two plausible explanations for such differences. One is the mutation bias model, in which the sources of variations are due to differences in regional and/or lineage-specific mutation rates or mutational pattern. This model predicts that the base composition of synonymous sites is the same as that of adjacent noncoding regions, such as the intron, which presumably evolves in a neutral fashion. Thus, the GC content of synonymous sites and that of adjacent noncoding regions can be expected to show a positive correlation. The *Amy3*-type gene showed a weakly positive correlation between GC3s and GCi. Therefore, the *Amy3*-type gene evolution may be partly

explained by the mutation bias model. However, contrary to our expectations, GCi is lower than GC3s (see Table 1). One possible explanation for the deviation from the predictions of the model is the remnant of past selection on synonymous sites. Following gene duplication, the *Amy3*-type genes have evolved in a neutral fashion, but they have not yet reached an equilibrium state. On the other hand, no correlation was found between GC3s and GCi in the *Amy1*-type gene, suggesting that the mutation bias model should probably be rejected in this case. Another possible explanation is the major codon-preference model. This model argues that evolutionary patterns at silent sites are determined by the balance of mutation, selection, and random genetic drift (Sharp and Li 1989, Akashi 1995). One piece of evidence that supports this model is the fact that highly expressed genes have a higher codon bias, whereas genes with lower expression levels show less codon bias (Shields et al. 1988). Since major codons are thought to confer fitness benefits, and more biased genes generally have more major codons, highly expressed genes have a higher fitness than less biased genes. The activities of amylase isozymes encoded by the *Amy1*-type genes change more in response to food environment and developmental stage than those encoded by the *Amy3*-type genes (Inomata and Yamazaki 2000), suggesting stronger regulation (expression) of the *Amy1*-type genes. In this sense, the greater ability to respond to different food environments appears to be advantageous for fruit fly adaptation to severe environments. If so, the major codon-preference model might explain the observed patterns of the *Amy1*-type genes.

Acknowledgments. We are grateful to Drs. H. Tachida and A.E. Szmidt for helpful discussion and Drs. D. Lachaise, M.T. Chassagnard, and Y. Fuyama for species diagnosis. We would like to thank F. Maczkowiak for technical assistance. This work was supported by research grants to N.I. and T.Y., by a research fellowship awarded to Z.Z. from the Ministry of Education, Science and Culture of Japan, and by a CNRS research cooperative program (PICS 607) awarded to M.-L.C. and J.-L.DL.

References

Akashi H (1995) Inferring weak selection from patterns of polymorphism and divergence at "silent" sites in Drosophila DNA. Genetics 139:1067–1076

Akashi HR, Kliman RM, Eyre-Walker A (1998) Mutation pressure, natural selection, and the evolution of base composition in *Drosophila*. Genetica 102/103:49–60

Ashburner M (1989) *Drosophila*: A laboratory handbook. Cold Spring Harbor Laboratory, Cold Spring Harbor,NY

Ayala JF (1965) Sibling species of the *Drosophila serrata* group. Evolution 19:538–545

Bahn E (1967) Crossing over in the chromosomal region determining amylase isozymes in *Drosophila melanogaster*. Hereditas 58:1–12

Benkel B, Hickey DA (1986) Glucose repression of amylase gene expression in *Drosophila melanogaster*. Genetics 114:137–144

Bock IR (1980) Current status of the *Drosophila melanogaster* species group (Diptera). Syst Entomol 5:341–346

Bock IR, Wheeler MR (1972) The *Drosophila melanogaster* species group. Univ Tex Publ 7213:1–102

Brown CJ, Aquadro CF, Anderson WW (1990) DNA sequence evolution of the amylase multigene family in *Drosophila pseudoobscura*. Genetics 126:131–138

Da Lage J-L, Lemeunier F, Cariou M-L, David JR (1992) Multiple amylase genes in *Drosophila ananassae* and related species. Genet Res 59:85–92

Da Lage J-L, Wegnez M, Cariou M-L (1996) Distribution and evolution of introns in Drosophila amylase genes. J Mol Evol 43:334–347

Da Lage J-L, Renard E, Chartois F, Lemeunier F, Cariou M-L (1998) *Amyrel*, a paralogous gene of the amylase gene family in *Drosophila melanogaster* and the *Sophophora* subgenus. Proc Natl Acad Sci USA 95:6848–6853

Da Lage J-L, Maczkowiak F, Cariou M-L (2000) Molecular characterization and evolution of the amylase multigene family of *Drosophila ananassae*. J Mol Evol 51:391–403

Doane WW, Gemmill RM, Schwartz PE, Hawley SA, Norman R (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 (1985) Phylogenies and the comparative method. Am Nat l25:1–15

Felsenstein J (1993) PHYLIP—Phylogeny inference package, v. 3.5c. University of Washington, Seattle

Hickey DA, Bally-Cuif L, Abukashawa S, Payant V, Benkel BF (1991) Concerted evolution of duplicated protein-coding genes in *Drosophila*. Proc Natl Acad Sci USA 88:1611–1615

Inomata N, Yamazaki T (2000) Evolution of nucleotide substitutions and gene regulation in the amylase multigenes in *Drosophila kikkawai* and its sibling species. Mol Biol Evol 17:601–615

Inomata N, Kanda K, Cariou M-L, Tachida H, Yamazaki T (1995) Evolution of the response pattern to dietary carbohydrates and the developmental differentiation of gene expression of α-amylase in *Drosophila*. J Mol Evol 41:1076–1084

Inomata N, Tachida H, Yamazaki T (1997) Molecular evolution of the *Amy* multigenes in the subgenus *Sophophora* of *Drosophila*. Mol Biol Evol 14:942–950

Kalantzi-Makri MC, Margaritis H, Sourdis J (1985) Phylogenetic relationships within the *montium* species subgroup of the genus *Drosophila* (*Sophophora*) based on electrophoretic data of the major chorion proteins. Ann Soc Entomol Fr (New Ser) 21:357–366

Kim BK, Watanabe TK, Kitagawa O (1989) Evolutionary genetics of the *Drosophila montium* subgroup 1. Reproductive isolations and the phylogeney. Jpn J Genet 64:177–190

Kim BK, Aotsuka T, Kitagawa O (1993) Evolutionary genetics of the *Drosophila montium* subgroup 2. Mitochondrial DNA variation. Zool Sci 10:991–996

Kimura M (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120

Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge

Kliman RM, Eyre-Walker A (1998) Patterns of base composition within the genes of *Drosophila melanogaster*. J Mol Evol 46:534–541

Lamnissou K, Zouros E (1989) Interspecific ovarian transplantations in *Drosophila*: Vitellogenin uptake as an index of evolutionary relatedness. Heredity 63:29–35

Lemeunier F, David JR, Tsacas L, Ashburner M (1986) The *melanogaster* species group. In: Ashburner M, Carson HL, Thompson JN (eds) The genetics and biology of *Drosophila*, Vol 3e. Academic Press, London, Orlando, pp 147–256

Moriyama EN, Hartl DL (1993) Codon usage bias and base composition of nuclear genes in *Drosophila*. Genetics 134:847–858

Nikolaidis N, Scouras ZG (1996) The *Drosophila montium* subgroup species: Phylogenetic relationships based on mitochondrial DNA analysis. Genome 39:874–883

Ohnishi S, Watanabe TK (1984) Systematics of the *Drosophila montium* species subgroup: A biochemical approach. Zool Sci 1:801–807

Ohnishi S, Kawanishi M, Watanabe TK (1983) Biochemical phylogenies of *Drosophila*: Protein differences detected by two-dimensional electrophoresis. Genetica 61:55–63

Pissios P, Scouras ZG (1993) Mitochondrial DNA evolution in the *montium*-species subgroup of *Drosophila*. Mol Biol Evol 10:375–382

Popadic A, Morman RA, Doane WW, Anderson WW (1996) The evolutionary history of the amylase multigene family in *Drosophila pseudoobscura*. Mol Biol Evol 13:883–888

Rozas J, Rozas R (1999) DnaSP version 3: An integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics 15:174–175

Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Sharp PM, Li W-H (1989) On the rate of DNA sequence evolution in *Drosophila*. J Mol Evol 28:398–402

Shibata H, Yamazaki T (1995) Molecular evolution of the duplicated *Amy* locus in *Drosophila melanogaster* species subgroup: Concerted evolution only in coding region and excess of nonsynonymous substitutions in speciation. Genetics 141:223–236

Shields DC, Sharp PM, Higgins DG, Wright F (1988) "Silent" sites in Drosophila genes are not neutral: Evidence of selection among synonymous codons. Mol Biol Evol 5:704–716

Steinemann S, Steinemann M (1999) The amylase gene cluster on the evolving sex chromosomes of *Drosophila miranda*. Genetics 151:151–161

Thomson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Triantaphyllidis CD, Panourgias JN, Scouras ZG (1978) Isozyme variation and phylogenetic relationships among six species of the *montium* subgroup of the *Drosophila melanogaster* species group. Genetica 48:223–227

Tsakas SC, Tsacas L (1984) A phenetic tree of eighteen species of the *melanogaster* group of *Drosophila* using allozyme data as compared with classifications based on other criteria. Genetica 64:139–144

Wright F (1990) The 'effective number of codons' used in a gene. Gene 87:23–29