J Mol Evol (1996) 43:334–347

JOURNAL OF MOLECULAR EVOLUTION

Distribution and Evolution of Introns in Drosophila Amylase Genes

Jean-Luc Da Lage,[1] Maurice Wegnez,[2] Marie-Louise Cariou[1]

[1] UPR 9034: Populations, Génétique et Évolution, CNRS, 91198 Gif sur Yvette cedex, France
[2] Laboratoire Développement et Morphogenèse, UPS centre d'Orsay, Bât. 445, 91405 Orsay cedex, France

Received: 23 October 1995 / Accepted: 5 March 1996

Abstract. While the two amylase genes of *Drosophila melanogaster* are intronless, the three genes of *D. pseudoobscura* harbor a short intron. This raises the question of the common structure of the *Amy* gene in Drosophila species. We have investigated the presence or absence of an intron in the amylase genes of 150 species of Drosophilids. Using polymerase chain reaction (PCR), we have amplified a region that surrounds the intron site reported in *D. pseudoobscura* and a few other species. The results revealed that most species contain an intron, with a variable size ranging from 50 to 750 bp, although the very majoritary size was around 60–80 bp. Several species belonging to different lineages were found to lack an intron. This loss of intervening sequence was likely due to evolutionarily independent and rather frequent events. Some other species had both types of genes: In the *obscura* group, and to a lesser extent in the *ananassae* subgroup, intronless copies had much diverged from intron-containing genes. Base composition of short introns was found to be variable and correlated with that of the surrounding exons, whereas long introns were all A-T rich. We have extended our study to non-Drosophilid insects. In species from other orders of Holometaboles, Lepidoptera and Hymenoptera, an intron was found at an identical position in the *Amy* gene, suggesting that the intron was ancestral.

Key words: Alpha-amylase — Intron — *Drosophila ananassae* — Obscura species group — *Zaprionus* — Drosophilids — Base composition — Multigene families

Correspondence to: J-L. Da Lage

Introduction

In the field of evolutionary biology, alpha-amylases (EC 3.2.1.1) are of great interest since these enzymes interact directly with the environment through food substrates. Therefore, differences in gene structure, regulation, or in proteins may be more easily interpretated in terms of adaptation. Indeed, the amylase system has been widely studied in Drosophila and other organisms for the last 30 years. In *D. melanogaster*, the duplicated structure of the *Amy* gene was first evidenced by Bahn (1967). A high level of isozyme polymorphism was found in this species, enhanced by the presence of two active copies (Hickey 1979; Daïnou et al. 1987). The coding region was cloned using a mouse probe (Gemmill et al. 1985) and the nucleotide sequence revealed that the two genes, 1.5 kb long each and divergently transcribed, were monoexonic (Boer and Hickey 1986). Further studies showed that the gene structure observed in *D. melanogaster* was not the rule for all Drosophila species. Evidence arose that in several species, the amylase genes were interrupted by a short intron: In *D. pseudoobscura* (Brown et al. 1990), *D. virilis* (D. Hickey, unpublished), *D. eugracilis* and *D. ficusphila* (Tadlaoui-Ouafi 1993), and some copies of *D. ananassae* (Da Lage et al. in prep.). These data lead us to investigate further the presence of introns in Drosophila amylase genes. In the present controversy over the origin of introns (see, e.g., Sharp 1985; Doolittle 1987; Rogers 1990; Cavalier-Smith 1991, for reviews), increasing data suggest a very ancient existence of introns, although more recent insertional events have also been evidenced. However, few studies have focused on comparisons of introns in a single gene between many related species, which may be

helpful in understanding the short-term dynamics of intron evolution. Drosophila amylases are a model of choice for this purpose, since the data available suggest that amylase introns could have been either inserted recently in some species (in hotspots) or removed in some other taxa. Using a PCR assay, we have screened over 150 Drosophilid species for the presence of an intron at the position where it had been found in the species cited above. We also attempted to detect the presence of an intron in non-Drosophilid insect species. Some of the PCR products were sequenced to characterize further the introns and the surrounding coding genes.

Materials and Methods

Species of *Drosophila*, *Zaprionus*, and other genera of Drosophilids are listed in Table 1. A number of species from our laboratory living stocks and many species from various origins, kept frozen (−80°C) for years, were assayed. Some DNAs (about 30 species) were already available in the laboratory and then used directly. These DNAs had been mass-extracted so that there could be some significant molecular polymorphism. For the other species, DNA was roughly extracted from single flies, according to Gloor and Engels (1991).

Primers for in vitro amplification were designed using an alignment performed by Tadlaoui-Ouafi (1993) between a few Drosophila *Amy* sequences already available: *D. eugracilis*, *D. ficusphila* (Tadlaoui-Ouafi 1993), *D. melanogaster* (Boer and Hickey 1986), *D. ananassae* (J.-L. Da Lage, unpublished), *D. pseudoobscura* (Brown et al. 1990), *D. virilis* (D.A. Hickey, unpublished). The primers were chosen in highly conserved regions surrounding the putative intron site (position 177 in *D. melanogaster*). Primer sequences are: INTR1 (upstream): $^{5'}$AGTGCGAGAACTTCCTGGG$^{3'}$; INTR2 (downstream, reverse-complementary strand): $^{5'}$CGGGCCACATGTGCTTGGC$^{3'}$. They were checked with the AMPLIFY 1.2 PCR simulator for Macintosh by W. Engels. Standard PCR procedure was the following: initial denaturation 95°C, 5 min; cycles: 95°C, 30 s; 60°C, 1 min; 72°C, 2 min; 25 cycles; terminal elongation: 72°C, 2 min. Negative controls were always performed and filtertips were always used to ensure the absence of contamination. PCR products were visualized on 1.3% agarose overnight gels in 0.5 × TBE buffer. In the absence of intron, the expected size was 500 bp. A few words of caution about this method: We have observed a few errors of Taq polymerase controlled by sequencing several cloned PCR products amplified from a known cloned DNA fragment. This might explain slight differences between intron sequences of species known to have a single *Amy* gene. A more important problem may be due to primer specificity. We have checked that in *D. ananassae* the cloned *Amy-4N* gene with an intron was successfully amplified separately, although the PCR product from fly amplification showed only a very faint band at 560 bp. On the other hand, *Amy-c1* was not amplified due to mismatch with INTR1 primer (simulated with AMPLIFY 1.2). In multicopy species, this problem might affect the results.

Non-Drosophilid insect species were also assayed, but other primers were used for this purpose. The species were: Diptera: *Ceratitis capitata* (Tephritidae); *Hymenoptera: Vespula vulgaris*, *Bombus* sp., *Vespa crabro*, *Halictes* sp. (DNAs supplied by L. Garnery); Lepidoptera: *Spodoptera dolichos*, *Spodoptera frugiperda*, *Spodoptera ornithogalli*, *Spodoptera latifascia* (Noctuidae, DNAs supplied by J-F. Silvain); Coleoptera: *Blaps mucronata* (supplied by S. Prigent). Primers were chosen using an alignment between amylase sequences from *Drosophila ananassae*, *Aedes aegypti* (cDNA, GenBank access number L03637), and *Tribolium castaneum* (cDNA, EMBL access number X06905). The alignment showed few suitable conserved regions surrounding the intron putative site. The following sequences were selected: 1U (upstream): $^{5'}$GTTCACCTCTTCGAGTGG$^{3'}$; 2U (downstream, reverse-complementary strand): $^{5'}$GGTTGGTAACGCTCCCACC$^{3'}$. The 3′ ends of these primers correspond to tryptophan codons TGG (one *Trp* for 1U, two successive *Trp* for 2U). Since these bases are critical for conservation of the tryptophan amino acid (encoded by a single codon), one expects this codon to be well conserved even in poorly related species, and then a better matching of the primers. PCR cycles were modified as follows: Annealing temperature was lowered to 55°C and 35 cycles were run. In the absence of intron, the expected size is about 150 bp. Some PCR products were cloned for sequencing: After gentle precipitation in the presence of ammonium acetate to eliminate primers and dNTPs, about 5–10 ng was ligated to 40 ng of the pGEM-T vector (Promega) and transformed in DH5α *E. coli* bacterial strain. Positive clones were sequenced using the dideoxy termination method (Sanger et al. 1977). Sequences were analyzed with the SEQAPP program for Macintosh by Don Gilbert. Sequence alignments were performed with the CLUSTALV program by Des Higgins. Coding sequences were used for constructing a tree with the PAUP 3.1.1 program by David Swofford. All the sequences were deposited to the Genbank/EMBL databases with accession numbers U31121 to U31158.

Results

The Structure of the Amylase Genes in Drosophilids

Over 150 Drosophilid species were assayed for intron presence in the *Amy* gene. Table 1 shows the results of PCR for Drosophilids. Most *Drosophila* species and all the other Drosophilids tested have an intron in the amylase gene within the putative insertion region. The most common intron size is about 60–80 bp, but there is considerable variation in size between species (Fig. 1A). The smallest ascertained intervening sequences are about 50 bp and the shortest *Amy* intron sequenced so far is 54 bp long in *D. eugracilis* (Tadlaoui-Ouafi 1993). We have found some PCR products ca. 510 bp long, suggesting the presence of a very short intron, although a few additional codons in the coding sequence could result in this slight difference. A few long introns were found; the longest ones (ca. 750 bp) occurred in the related species *D. kuntzei* and *D. limbata* (*quinaria* group).

Interestingly, some species or lineages, all within the Sophophora subgenus, are intronless (Table 1). This is the case for the whole *melanogaster* subgroup (Boer and Hickey 1986; Payant et al. 1988; Shibata and Yamazaki 1995) and a few species, interspersed throughout the subgenus: *D. lutescens* and three species from the *montium* subgroup, *D. greeni*, *D. auraria*, *D. triauraria*. Two cases deserve particular attention: The *obscura* group and the *ananassae* subgroup. In the *obscura* group (Fig. 2A) there are always two bands in the amplified DNA: One is 500 bp long (no intron); the other is 560–580 bp long (common-size intron). In *D. pseudoobscura*, three *Amy* genes were described (Brown et al. 1990), all of them with an intron. However, the authors reported an additional in situ signal, suggesting another gene cluster. We have probably amplified a gene from the other locus,

Table 1. List and systematic position of Drosophilid species used in this study and approximate sizes of *Amy* introns[a]

Genus/Subgenus	Group	Subgroup	Species	Size (bp)
Drosophila				
sensu stricto	annulimana	–	*aracataca**	60
			talamancana	60
	cardini	cardini	*cardini*	100
			cardinoides	60
		dunni	*arawakana*	80
			caribiana	60
	dreyfusi	–	*camargoi*	80
	funebris	–	*funebris**	188
	histrio	–	*sternopleuralis*	80
	immigrans	immigrans	*immigrans*	60
		hypocausta	*hypocausta*	80
			rubida	80
		nasuta	*albomicans*	60
			kepulauana	60
			*nasuta**	66
			pallidifrons	60
			sulfurigaster	60
		ungrouped	*trilimbata*	150/300
	melanica	–	*melanica*	0/80
			tsigana	60
	mesophragmatica	–	*gaucha*	60
	pallidipennis	–	*pallidipennis*	80
	peruviana	–	*peruviana*	80
	polychaeta	–	*hirtipes(iri)*	100
			polychaeta	100
	quinaria	–	*adamsi*	60
			*kuntzei**	757
			*limbata**	737
			nigromaculata	700?
			transversa	80/700
	repleta	repleta	*repleta** (4)	75
		hydei	*bifurca*	60
			hydei	60
		mercatorum	*mercatorum**	77
		mulleri	*aldrichi*	60
			buzzatii	80
			huayalasi	30/80
			martensis	80
			mojavensis	100/120
			mulleri	100
			nigrodumosa	110
	testacea	–	*testacea*	450
	tripunctata	III	*mediopictoides*	60
		IV	*metzii*	250
	virilis	–	*americana*	60
			littoralis	60/680
			*virilis**(5)	59
	Ungrouped		*aracaea*	550
			pruinosa	60
			repletoides	750
			wheeleri	60
	Hawaii?		*mimica*	60
Sophophora	fima	–	*fima**	60
	melanogaster	ananassae	*ananassae** (0)	0/56–61
			atripex	0/60?
			*bipectinata**	0/65
			*ercepeae**	0/60
			malerkotliana	0/60
			monieri	0/60
			pallidosa	0/60/180?
			pallidosa-like	0/60?
			parabipectinata	0/60

Table 1. Continued

Genus/Subgenus	Group	Subgroup	Species	Size (bp)
			phaeopleura	0/60
			pseudoananassae	0/60
			nigrens	
			*vallismaia**	0/78
			varians	60
		elegans	*elegans*	60
			subelegans	60
		eugracilis	*eugracilis** (1)	54
		ficusphila	*ficusphila** (1)	72
		melanogaster	*melanogaster** (2)	0
		montium	*asahinai*	10?/70
			auraria	0
			barbarae	70
			bocqueti	60
			burlai	50
			cauverii	60
			cf. bakoue	60
			chauvacae	60
			davidi	60
			diplacantha	60
			dossoui	0/60
			greeni	0
			jambulina	70
			kikkawai	60
			leontia	60
			malagassya	60
			nagarholensis	60
			nikananu	80
			rufa	0/70
			serrata	62
			triauraria	0
			vulcana-like	60
		suzukii	*lucipennis*	0/64
			biarmipes (rajasekari)	0/60
		takahashii	*lutescens*	0
			pseudotakahashii	60
			*takahashii**	0/57–59–64
	obscura	obscura	*ambigua*	0/60
			bifasciata	10/70
			*guanche**	0/67
			imaii	0/60
			lowei	0/60
			*obscura**	0/64
			subobscura	0/60
		pseudoobscura	*miranda**	0/67
			persimilis	10/80
			*pseudoobscura** (3)	0/70–71–81
		affinis	*affinis*	0/60
			algonquin	0/60
			azteca	0/60
			helvetica	20/100
		microlabis	*kitumensis**	0/64
			*microlabis**	0/64
	saltans	cordata	*neocordata*	80
		elliptica	*emarginata*	10
		saltans	*prosaltans*	60
	willistoni	bocainensis	*fumipennis*	60
			nebulosa	60
		willistoni	*willistoni*	60
Dorsilopha	–	–	*busckii*	60
Hirtodrosophila				
	quadrivitata	confusa	*confusa*	60
Lordiphosa				
	fenestrarum	–	*andalusiaca*	60

Table 1. Continued

Genus/Subgenus	Group	Subgroup	Species	Size (bp)
Scaptodrosophila				
	latifasciaeformis	–	*latifasciaeformis*	60
	saba	–	*pugionata*	60
	victoria		*deflexa*	60
			lebanonensis	80
			rufifrons	60
	ungrouped		*bryani*	10/50
Chymomyza				
	fuscimana		*amoena*	60
Liodrosophila	–	–	*aerea*	60
Scaptomyza				
Parascaptomyza	–	–	*pallida*	60
Zaprionus				
Zaprionus	inermis	–	*badyi*	300
			ghesquieri	60
			inermis	180
			kolodkinae	320
			verruca	300
	armatus	armatus	*vrzydaghi (vouidibioi)*	50
		vittiger	*indianus*	60
			megalorchis	60
			taronus	60
			vittiger	60
		tuberculatus	*mascariensis*	250/400
			*sepsoides**	521
			*tuberculatus**	272
anaprionus			*lineosa*	80

[a]Asterisks indicate sequenced clones with exact sizes of introns: (0) J.L. Da Lage (unpublished), (1) Tadlaoui-Ouafi (1993), (2) Boer and Hickey (1986), (3) Brown et al. 1990, (4) S. Prigent (unpublished), (5) D. Hickey (unpublished), all the others: this study.

which has no intron at the putative site. Another interesting result is that American species of the *obscura* group show the strongest band at 560–580 bp, and a weaker signal at 500 bp. In the Afro-European species and the Japanese *D. imai*, both bands are equally intense. This suggests that the ratio of the number of genes with and without intron is different between the American and the Afro-European species. The *ananassae* subgroup is known for multiple *Amy* genes (Da Lage et al. 1989, 1992; Da Lage and Cariou 1993) and also shows two amplified DNA fragment (Fig. 2B). Species of the *bipectinata* and *ercepeae* complexes have two equally intense bands that indicate the lack of intron in some copies. In the *ananassae* complex, the intron missing seems still more severe since the 560-bp band is always very faint. However, genomic cloning of *Amy* genes in *D. ananassae* itself has shown that several copies have an intron (Da Lage et al. in prep.). The status of *Drosophila varians* is not clear regarding the presence of intron. This species has diverged early within the subgroup (Cariou et al. unpublished), but its PCR pattern is similar to that of *D. ananassae* with a faint band at 560 bp suggesting the presence of the two types of genes. However, the PCR products cloned from this species showed no intron (see below).

Analysis of Intron Sequences in Drosophila *Amylase Genes*

Twenty-nine *Amy* introns were cloned and sequenced (Fig. 3 and Table 1). All of them were located at an identical position in the coding sequence, corresponding to position 177 of the *D. melanogaster* gene, between two codons in the conserved site: CAG[intron]GT. This site fits the consensus reported by Mount (1982) and Mount et al. (1992). At least two or three clones were sequenced for a single PCR product in order to estimate molecular variability due to diploidy or multicopy structure. A low variability (2–4%) was found between different clones from species with single amylase gene. For species like *D. takahashii*, where several *Amy* genes have been reported (Tadlaoui-Ouafi 1993), three introns which differed in sequence and size were found. In *D. ananassae*, seven copies have been sequenced after genomic cloning (Da Lage et al. in prep.), three of which have an intron. *Amy-c1* and *Amy-4N* introns have very different sequences. Differences in size between the introns of *Amy* genes of *D. pseudoobscura* have also been evidenced (Brown et al. 1990).

Two 5′ splice-site consensus sequences were found at the 5′ splicing site: GTAAG (14 over 29) and GTGAG

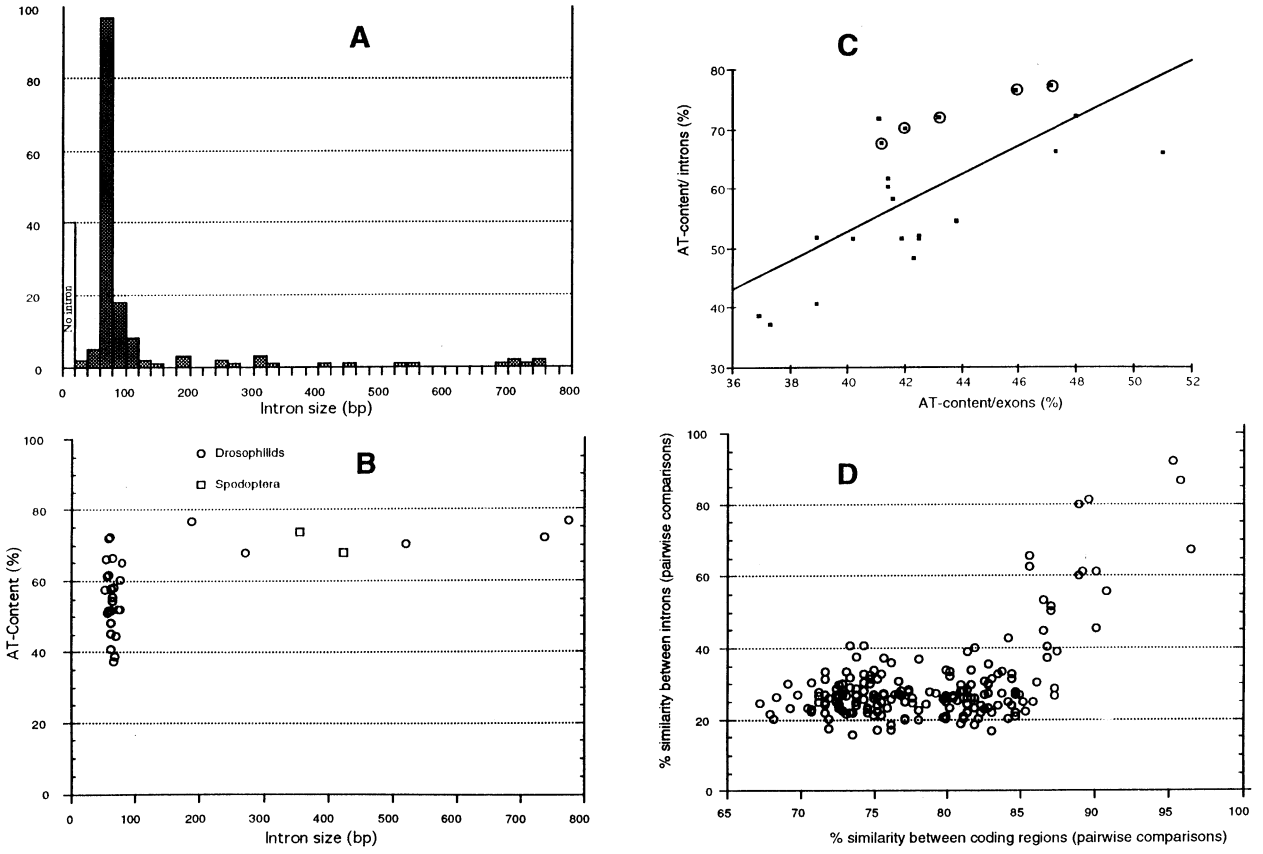

Fig. 1. **A** Frequency distribution of intron sizes for Drosophilid amylase genes, as from data of Table 1. Several sizes may be possible for a single species, due to multicopy structure. **B** Base composition of the introns sequenced from amylase genes: *circles:* Drosophilids; *squares: Spodoptera* species (Lepidoptera Noctuidae). Variability of composition is striking in the cluster of short introns, while long introns are all A-T rich. **C** Correlation between A-T contents in introns and surrounding coding sequences in amylase genes of Drosophilids; 22 sequences were analyzed. Genes with long introns are circled. **D** Plot of similarity scores of the coding sequences vs the introns in pairwise comparisons of genes containing introns of similar sizes (short introns only: 20 genes used; *D. takahashii* and *D. lucipennis* were excluded).

(12 over 29). The three remaining sequences were GTGCG (*D. bipectinata*), GTACA (*D. ananassae*, gene *Amy-4N*), and GTAGT (*D. vallismaia*). The last three bases of *Amy* introns were CAG (17 over 29) or TAG (11 over 29), except in one *D. takahashii* intron (AAG). The branchpoint sequence was difficult to determine and rarely fitted the CTAAT consensus perfectly (Keller and Noon 1985; Mount et al. 1992). In many species a polypyrimidine stretch was found between the putative branchpoint and the 3′ splicing site: In *D. mercatorum*, the intron contains a microsatellite-like $(CT)_{11}$ sequence, which is slightly modified in the related species *D. repleta*. In the *D. obscura* group, where there are 22 bases between the CTCAT putative branchpoint site and the AG intron termination, the (C,T) proportions vary from 16/22 to 20/22 according to the species. There is no clear pyrimidine-rich region in *D. fima*. Long introns are pyrimidine-rich in the 20–30 nucleotides before the 3′ splicing site but the branchpoint sites were not clearly identified.

Figure 1B shows that there is considerable variability in base composition for short introns. The lowest A-T contents are in the *obscura* group. On the contrary, long introns are all A-T rich. We have compared A-T richness in the introns and the surrounding coding sequences. Figure 1C shows that the variability in intron base composition is correlated to that of exons ($r = 0.695$; $P < 0.001$, for the 22 sequences used, including long introns; see legend of figure). Alignments of introns were possible only between closely related species, such as *D. ercepeae* and *D. vallismaia*, or within the *D. obscura* group. The *D. vallismaia* and *D. ercepeae* introns match almost perfectly, except for a 18-bp insertion in *D. vallismaia*, and the similarity is 95% in the coding region of the PCR product (440 bp sequenced). In the *obscura* group, the coding sequence similarities range between 95% and 90%. The intron similarity is also very high (92%) between the closely related species *D. kitumensis* and *D. microlabis* (Cariou et al. 1988) and the latter species and *D. obscura* (81%). It strongly decreases if we compare American and Afro-European species. The similarity is only 45% between *D. pseudoobscura* and *D. obscura* introns. However, these data should be considered cautiously, since there are several *Amy* gene copies in these species, all of which may have not been cloned. We are not sure whether we compare orthologous or

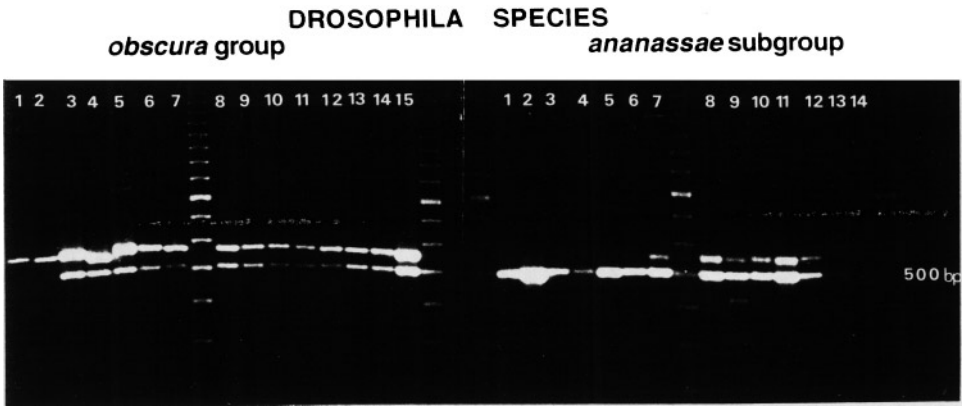

Fig. 2. Electrophoresis gels (1.3% agarose) of DNA fragments amplified with primers INTR1 and INTR2 (see Materials and Methods); 20% of the PCR products were loaded. **A** *obscura* group: 1, *D. affinis;* 2, *D. azteca;* 3, *D. algonquin;* 4, *D. lowei;* 5, *D. pseudoobscura;* 6, *D. persimilis;* 7, *D. miranda;* 8, *D. ambigua;* 9, *D. bifasciata;* 10, *D. helvetica;* 11, *D. subobscura;* 12, *D. obscura;* 13, *D. microlabis;* 14, *D. guanche;* 15, *D. imai* **B** *ananassae* subgroup: 1, *D. ananassae;* 2, *D. pallidosa;* 3, *D. pallidosa*-like; 4, *D. atripex;* 5, *D. phaeopleura;* 6, *D. monieri;* 7, *D. bipectinata;* 8, *D. malerkotliana;* 9, *D. pseudoananassae;* 10, *D. bipectinata;* 11, *D. parabipectinata;* 12, *D. ercepeae;* 13, *D. vallismaia;* 14, *D. varians. M:* 100-bp ladder.

paralogous genes. Such comparisons are even more difficult in the *D. ananassae* subgroup, where the *Amy* gene copies are more numerous. More generally, we have done pairwise comparisons of similarity scores for exons and introns from the sequenced clones for which intron sizes are similar (Fig. 1D). The graph shows that below 85% similarity in the coding sequence the intron comparison scores generally are within a range close to random. Thus, comparisons between more remote species are not suitable. In addition, problems can arise in interpreting alignments because of gaps due to size differences. Introns of very different sizes cannot obviously be aligned, except in the case of duplications of whole or internal sequences, such as in *Zaprionus sepsoides* where an internal duplication of about 190 bp was found. At first sight, comparison with *Z. tuberculatus* might suggest that it happened after the lineages split but similarity scores between the three sequences (two repeats in *Z. sepsoides* and one in *Z. tuberculatus*) indicate a much higher similarity between the distal repeat of *Z. sepsoides* and *Z. tuberculatus* than within *Z. sepsoides.* The data thus suggest that the duplication was ancient and was lost in the *Z. tuberculatus* lineage (Fig. 5). Clearly, internal duplications could be involved in increasing intron size but cannot explain all the long introns. In *D. limbata* and *D. kuntzei,* no internal direct nor inverted repeats were found.

In a number of species from the *Sophophora* subgenus, we have found *Amy* genes in which the intron has been lost. We have checked the accuracy of intron excision in these species. In *D. melanogaster,* both copies are intronless and comparisons with genes that retained an intron (e.g., *D. virilis*) show a high conservation of the region surrounding the intron site (not shown). That is, no deletion nor addition (relics from an ancient intron) was found. In species with the two types of genes (Fig. 4), for example, those of the *D. ananassae* subgroup (*D. ananassae, D. bipectinata, D. ercepeae, D. vallismaia*), the region is not altered. In the *D. obscura* group there is no change for *D. guanche,* but for *D. microlabis* and *D. kitumensis* the intronless gene sequence is modified just before the 5′ splicing site GT, with an A instead of G. This is also the case for intronless copies of *D. takahashii* and *D. lucipennis.* Thus, these intronless copies do not fit the upstream consensus MAG/GT . . . (M = A or C) (Mount et al. 1992). However, these authors have also reported a few cases of genes with a A at this point (−1 relative to the intron). Partial sequencing of the adjacent coding region (430–440 bp) mostly downstream to the intron shows no stop codon, so it is not possible from our data to conclude upon the functional or nonfunctional status of these intronless copies.

Introns of Amylase Genes in Non-Drosophilid Species

Amplification of the putative insertion region for an intron in amylase genes was assayed in non-Drosophilid Diptera and more remote insects or other Arthropods. Using 1U and 2U primers (see Materials and Methods), amplification was successful in *Ceratitis capitata,* in four moth species of the genus *Spodoptera* (Noctuidae), and in the common wasp *Vespula vulgaris* (Hymenoptera). The other Hymenoptera tested yielded many bands in PCR products and were not investigated farther. In Coleoptera, a weak amplification was obtained in *Blaps mucronata.* The intron region was cloned in *Ceratitis capitata, Vespula vulgaris, Spodoptera frugiperda* and *S. ornithogalli.* Due to the positions of primers, the coding region is short, but these four species have an intron at the expected position. Intron sizes are 80 bp for *C. capitata,* ca. 1,200 bp for *Vespula vulgaris* (not sequenced entirely), ca. 400 bp for noncloned *Spodoptera,* 422 bp for *S. frugiperda,* and 355 bp for *S. ornithogalli* (see

D. ananassae 4N, 61 bp
GTACATAAGATATATTTTTTAAAAGGCTTCCCTGAAGTTTAACAAATTTATTTGGTTCCAG

D. ananassae c1, 56 bp
GTAAGACTAGAGATGGCAGCTGAAATAGAATGATTTCAACTCACTTTTGTTAATAG

D. bipectinata, 65 bp
GTGCGTCAGATATATTCTTAAACGAGGTTCCTTAAAATTTGACAAATTTTGGTTTTGGTTGTCAG

D. ercepeae, 60 bp
GTAAGGTTTTTTTTAACGGATTTATGAACCGGCCTTAAAACCAATCTGTCTGGTTTGCAG

D. lucipennis, 64 bp
GTAAGGATTAGCCCCGCAATTACCCCCCATGAGCCACCGGAGTAACAGAATCCCCTGCCATTAG

D. vallismaia, 78 bp
GTAGTCATATACCGCCTTATTAAGGTTTTTTTTACGGGTTTATGAACCGGCCTTAAAACCAATCTGTCTGGTTTGCAG

D. ficusphila, 72 bp
GTAAGACCCTATTCCCCGAGTGGGCTTTTTACATCGATGCAGTGACAGCAATCTCCCCCGCTTCGCCCACAG

D. eugracilis, 54 bp
GTAAGAATCTAATTCTCCCAACAAGAACACCTTATTGAGTGGCTTCGACCACAG

D. serrata, 62 bp
GTGAGGATCAGCTGTAACTCTCTCTCTCTCGAAGGCACTTCTATATCCTAATCCCCCATCAG

D. takahashii 6, 59 bp
GTGAGGATTTCCTACCCTCTTCCCTTTACGAAAAAGGCTCTACCCCCTACGATCCTTAG

D. takahashii 7, 57 bp
GTGAGGATTCTATACCCACTTATAATCAAGAAAAAGACTCTACCCTACTCCCAATAG

D. takahashii 8, 64 bp
GTGAGGATTTGATAGCCACTTATCATTAAGAAAAAGGCTCTACCCCTATCCCCTACGATCTAAG

D. fima, 60 bp
GTAAGAGATTCATAGCCTCAATCTAAAAAATTATAAAATACCCATAAGTATTGTTGACAG

D. kitumensis, 64 bp
GTGAGAACCGTCGGCAGACTTGAAATCCACATCCACTCATTCTCTGCCTCTACGTCTGTCTCAG

D. microlabis, 64 bp
GTGAGAACCATCCGCAGACTTTAAATCCACATCCACTCATTTTCTGCCTCTACGCCTGTCTCAG

D. obscura, 64 bp
GTAAGATCCGTCCGCAGCCTTGAAATCCTCATCCACTCACCTGCTGCCTCTGCGCCCGTCTCAG

D. guanche, 67 bp
GTGAGATCCGTTCTCCGATTTTCAGAAATCCTCATCCACTCATATTTTGCTTATACGTCCATCTCAG

D. miranda, 67 bp
GTAAGATCCTTCCGGAGACACTCCCTCCTCGCCCATTCCTCATCTCCCGCCTCCGCCTCCCTCTCAG

D. pseudoobscura (Amy1), 70 bp
GTGAGATCCGTCCGGAGAGCTTCCCTCCACCGTCACCCAATCCTCATCTCCCGCCTCTACCTCTTCCCAG

D. virilis, 59 bp
GTGAGCCAAATTATAGCCATTGCCAGCACATTAGTTTCAATGCCAATCCCCCCGCTCAG

D. repleta, 75 bp
GTGAGAACCGAAAGGCTAAATGATACTCAAAGCAACAGTTGGCTAGAGAGATGTTCTCCCTCTCCTCCTCCCTAG

D. nasuta, 66 bp
GTAAGTTCAAGAGTTCAACACTTCGCAGCCACACAGATGCCTTTTAATTGGGCTACTCGCTTGTAG

D. mercatorum, 77 bp
GTGAGAAGCGAAGGGCTGAATGTTACCCAATGCAACAGTTTGAGAGCTATTCTCTCCTCTCTCTCTCTCTCCTAG

D. funebris, 188 bp
GTGAGCGAAATAAATGCATTCAATTAATTATAAATAAATTATTTAAATTATTTGAAATCAAATTAGATTGGTGTAATGAA
TTCATTAATCAGATCTCTATATATTCTTTTCAATTATATATATGATGTATATTGAGTCCCTTAGGTTTATCTAAAAGCAA
ATACTACTGCATCTTTCGTTGCATTTAG

D. aracataca, 60 bp
GTAAGCCAATTGCCAAAGACTCGACTGCAGAGCATGGTCTAAAGCCCCATATCTGGTTAG

D. kuntzei, 757 bp
GTAAGCTAGACCAATTATTTTAGTTTATAATTTACTTAATTTATTATTAGTTTCTTACTATACTCATTTAACGTAGAATA
AAATTATTTGACAATAGAAAAAAAATAAAATTGGAGATTAAAATTTAAAATATTAATTAAATTAAATTAAACATTTTGTA
ATTAAAACTGGTCAGTGGTTCTCTAGTTAGAGTCATTAGAGTCATAGAAAATTAATTGGGTTCGATTCGCCGTCATATTA
TTCTCAATGTGTTTTTAAGTTCCTAATTCACAATACTTCAATAATAAGAAAAATAAGAAAAGATTTAAATTAGCAATTTA
ATTAAAATATTCAACATTTTTAATTGTAAATATAAAAATACATTAATAAATTTATTAAGTTGTCCACTTAAATTCTAGCT
GAGAGATATTTTTGAATTGACATTAATTTCTATTCATTCAACAAAAGTTGAGTCTTAAGATTGGTACTAGAAATATCTAG
ATCTGCATGTAGTATATTCTTTCAAATACATATATAATTTTAATTTATTTTGCTTTTTTTGATACATTATTTATTATTAT
ATTAGAATTCTAAAAAGATAGGTATAGTACGGACTTATACGTGGCAGTGTGGGATTCGAACCCAGTATTTTTCAAACGAA
AATCGATTATGTACAGGTCGGTACTCTTACTACTTATAAATATTTATAATTACATTTTATATACTCTATTTTTTATCAATG
CATTTTGTTTAATTTAATCTCTTCCTGCTGCTTTTAG

Fig. 3. Sequences of introns of amylase genes cloned for this study in the pGEM-T vector and those of *D. ficusphila, D. eugracilis, D. pseudoobscura, D. virilis* already available (see text or Table 1 for references). D: *Drosophila;* Z: *Zaprionus;* C: *Ceratitis;* S: *Spodoptera.*

D. limbata, 737 bp
GTAAGCTAAGCTAATCATAAGAATTATAAATGAATTAACTAAATATTACGAAACTAATATCCTATCGTTGTCATGTTGCA
ATTGTTTACTTGATTGAGCTTTGGCTTGACAATAGAATAATCTAGGGGTCGTTATGGCGTCGATCTTTTCGTCTTCTTAT
GAAGAAATTACATTATCTGATCAATAAGATTAGATTAGTATTAGTGGGCCCGAGGGTCTCGGTTAACTGGTTTAACTGGG
TTCGATTTTTATTCTGTCATGTTACAAAAGAAGAAAGCAAAGAAAAAATAATAGTCTTATATAAAATGTTCTATCACCGT
CATCAACACAAGTGATATAAAAAAAAAATACGAGTCGATAGAATAATCATTCAGATTTCAAACTGAAACTGAAGACTTGA
TTTTAAATTAGCAATCAAAATAAATTATACAACATTTTGAATTTAAATAGAAAAGATTAAAAATAATCCACCATAAATTA
AGTCTTCTTTAAAACTGAAAAGATTAAAGCTAATCTACTATTGAAAAGTGAGTCAGTTTTTCACATAAATTCTAAGTTAG
AGTTATTTTTAAATCTGACACTAATTGTAACTTTATGTAACTTCTAAATATCTACCAAATTATCTATGACATAAGATTGG
TAGCTAAAATATTTAGATCTGCATATATTAAAAAAAAAGTGTGTTTTTTTAAGCATCGGAAATTGAATACAAAGTCTTCT
CCTCCCGCTGCTGTTAG

Z. tuberculatus, 272 bp
GTAAGCGGGAAGAAGGAGAGAGAGAAGCTATTATCGATTGCCAAACAGTACTATCGATTATATCGATCTAAGTTGCTTCT
TAGACTAAAACAAACGATAAATCGAAACAAAAAAAATCGTTAAATGAAAAACAGTTGAAAAGTTCTTTAAAATCCCTAGA
AACAATCGATAGTCCAACAGAAATAATCGATATATTGAGCAACTTAATCAAGCTTCGATATATTTTGAATTGTGCAACCT
TTTCGATTAATTAATCCCTATTTCCTTGCCAG

Z.sepsoides, 521 bp
GTAAGCAGGAAGAAAGAGAGAAAAATAACTTACCTTCGATTTGTCAAACAATACTATCGATTATATCTAAGATATATATC
TAAGTAAAGCTAAGTCGATTCTTACGATTGTGGCTTGAAAAATGCTTCAAATCAATCAAAAAATCGATACAAAAAAAAAA
CGATAAATGAAAAACGTTGAAAAGTTCTTTGAAATGGCTAGAAATAATCGATAGTGCCACAAAAACCATCGATATATCGT
GGAACTTAAACAAGCTTCTATATAGTATAATCAATTGTGGAATCTTTTCGATTAATCTATACCTACTTTCTTGCCAGGTT
TGGCTTGAAAAATACTTCAAAAAAACGATAAATCGAAACACAAAAAATCTATAAATAAAAAACATATTGAAAGTGCTTTA
AAACCCCTAGCAACAATCGATAGTGCAACAGAAACAATCGATATATCGAGCAACTTAATCAAGCTTCGATATTTTCAATT
GTGAAACCTTTTCGATTAATTTATCCCTTTTTTCTTGCCAG

C. capitata, 80 bp
GTGAGCTCGAGGCTTTTATTCAAACTTAATATAATCATCTGTGCATTAAATGAAATACCGAACATGTTTTCTTCGGCTAG

S. frugiperda, 422 bp
GTGAGTGGATTGGGATTCACTATTTTTATTTTTTAACTTAGAGACATGGAGTTTCTTGCTCGTTCTTCTCCATTAGAAGC
TACATTTTGGAACGAGCGCCTAGCTTCACCGACAGACAGACTGACGGACGATTCAATTTGACGTTTCAAAAATAGTTTCC
TAATTTAGGATTAATTGAAATAAATGTTTTGACTTTGACTTTGAAACTTAGACTTAGGTGCAAAAAATAAGTAGGTGAAA
AAAAAAACTTGGAGCATTTGCACGGAAATTAAACAATGGAATGGAAATTATTTTTATTCTGCCTCTAAGTATTGATATAT
TTAGAATCAATCAAAAAATACTTTCTTTTGACTTCACACCAAATAGATTTCTTTCTCCATTTTGTTCGGAAATATGTCTC
CAAGAACACAATTTACTTGTAG

S. ornithogalli, 355 bp
GTGAGGGTGTTAACAATATTTTTATTTACGTTTTAATTTGGATTTATTAGTACACCAAAAATAAAGGAACACTTGAACAG
GGCTTTCTTGTTTGATAATTAGGTACGTATTATGAGGAATATAAGTTGTTTTAAAGATATTTATTGTATTTATGGATAAA
TCAAATTCATTCATAATTAAAACAAAGGGCCAAGTTGAATAGGAGCTTACATTACAACTTGTTAATATTTCTGTGCAATA
AATCAATTTTTTCCGCCATGTTTTCTTTCTATTATTTTATTTTCTTCACAACAATTAGATTTCTGTCTTCATTTTGTTTG
AAAATGTGACTGAAGACAATACAATTGTATTCTAG

Fig. 3. Continued.

sequences in Fig. 3). The two introns of these last two species share a global similarity of 52% which is much higher in the last 80 nucleotides. As illustrated in Fig. 1B, these long introns are AT-rich, as is the one of *V. vulgaris* in the portion sequenced (about 600 bp, not shown). At the nucleotide level, the 5′ splicing site is similar to Drosophila: CAG/GT. At the end of the introns the sequences are AG/AT for non Diptera instead of AG/GT in Drosophilids and *C. capitata.*

Discussion and Conclusion

The Ancestral Amylase Gene Contained an Intron

We have investigated by PCR the presence of an intron in the amylase genes of numerous Drosophilid species. Our data show that an intron of 60–80 bp is the most common case in Drosophila species as well as neighbor genera, such as *Chymomyza* or *Hirtodrosophila.* Thus, the question that had arisen from comparisons between the first sequenced amylase genes of Drosophila (see Introduction) may have now a clear answer: the presence of an intron in position 177 (referring to *D. melanogaster*) is ancestral in Drosophilids and is not due to independent insertions in a hotspot. It must also be pointed out that these introns are only putative since no verification has been made on cDNAs. However, the intron sequences presented above are not compatible with a functional gene in the absence of splicing because all of them have nonsense codons or create frame-shifts. Also, as pointed out in Materials and Methods, PCR is a limited tool of investigation in that it can fail to detect some genes for which one primer has a critical mismatch or in the case of competition between several copies, like in *D. ananassae.* Some species found to have genes of only one type (intron or not) may harbor ‘‘hidden’’ copies. Work is in progress in our laboratory to check this point.

The lack of intervening sequences in some species of the *Sophophora* subgenus is likely due to independent excision in different taxa. Intron deletions were not found elsewhere in Drosophila but they may exist in species that were not tested. Intron loss seems to be a rather common event in evolution, but its mechanism is poorly documented. In the case of Drosophila amylase genes, it is unlikely that retrotranscripts were inserted in the genome in a position allowing gene expression. Mature messenger RNA may have interfered with the DNA complementary sequence, and the mismatched bases of the DNA intron would have been then deleted. In plant mitochondria, intron loss has been reported to be mediated by recombination between a cDNA and the genomic DNA (Geiss et al. 1994). In species where several amylase gene copies exist, the loss of intron may have spread

to other copies after the initial event through gene conversion or unequal crossover (concerted evolution, see Ohta 1983).

From our results on non-Drosophilid species, we can reasonably assume that the presence of an intron in the amylase gene at the position studied here is ancestral to the higher insects, which had diverged at the beginning of the Mesozoic era (240 Myr) (Kristensen 1991; Kukalova-Peck 1991). Comparisons with human amylase data show a concordance for the position of intron 2 (the first intron within the ORF) of a human salivary amylase gene (EMBL access number M18671) and our intron site. This suggests that this intron may be a relic of the very early animal alpha-amylase genes. Vertebrates would have gained several additional introns in the course of their evolution (see Gumucio et al. 1988, for amylase gene structure in human) or, on the other hand, insects would have lost most of the primordial intervening sequences, leading in some species to monoexonic structure (e.g., *D. melanogaster*). In addition, the region in which the intron lies has been reported to be well conserved among animals and Streptomycetes (Janacek 1994). Our data do not solve the controversy about the old or recent origin of introns (see Introduction for references), but we conclude a very ancient origin of the amylase intron is the case.

Evolutionary History of Intron Loss in the Sophophora *Subgenus*

Intron loss is most likely ancestral to the diversification of the *melanogaster* subgroup because the same structures and gene arrangements are conserved. For the *D. ananassae* subgroup and the *D. obscura* group, due to their phylogenetic vicinity within the *Sophophora* subgenus, with an uncertainty for their precise respective taxonomic position regarding the *D. melanogaster* group (Pélandakis et al. 1991), one could expect that the two-band patterns observed after PCR in many species of both taxa were of common ancestral origin. Some lineages such as *D. ananassae* would have lost some of their introns. However, our data indicate independent evolutionary history in the two lineages. The parsimonious tree (Fig. 6) reveals that intronless and intronic genes in the *D. ananassae* subgroup are clustered together, and thus the intron loss may be an internal event of this subgroup which would have occurred more recently in this taxon. The bootstrap value (0.62) at the basal node of the subgroup is not high, but the neighbor-joining method of Saitou and Nei (1987) gave the same branching with a bootstrap value of 0.88 (1,000 replicates; data not shown). We have good indications from our unpublished results that intron-containing genes as well as intronless genes are functional in *D. ananassae* itself. However, the ratio of genes with or without intron in the

```
                                    ↓
D. ananassae/n.i.    GGT GTG CAG GTA TCT CCT
D. ananassae/i.      GGT GTC CAG GTA TCA CCC
                     *** **  *** *** **  **

D. bipectinata/n.i.  GGT GTA CAG GTA TCT CCC
D. bipectinata/i     GGT GTC CAG GTA TCA CCC
                     *** **  *** *** **  ***

D. ercepeae/n.i.     GGT GTT CAG GTA TCC CCC
D. ercepeae/i        GGA GTT CAG GTA TCG CCC
                     **  *** *** *** **  ***

D. vallismaia/n.i.   GGT GTT CAG GTA TCC CCC
D. vallismaia/i      GGA GTT CAG GTA TCG CCC
                     **  *** *** *** **  ***

D. guanche/n.i.      GGC GTG CAG GTC AGT CCC
D. guanche/i         GGT ATT CAG GTG TCG CCC
                     **   *  *** **      ***

D. kitumensis/n.i.   GGC GTG CAA GTC AGT CCG
D. kitumensis/i      GGC CTT CAG GTA TCG CCA
                     ***  *  **  **      **

D. microlabis/n.i.   GGC GTG CAA GTC AGT CCG
D. microlabis/i      GGC CTT CAG GTA TCG CCA
                     ***  *  **  **      **

D. takahashii/n.i.   GGA GTG CAA GTC AGT CCC
D. takahashii/i      GGA GTG CAG GTC TCC CCT
                     *** *** **  ***     **

D. lucipennis/n.i.   GGT GTG CAA GTG AGC CCC
D. lucipennis/i      GGA GTG CAG GTC TCC CCC
                     **  *** **  **    * ***
                                    ↑
```

Fig. 4. Conservation of the insertion region of introns in the *Amy* genes of several Drosophila species. The coding sequences of intronless copies (*n.i.*) and copies with intron (*i*) are aligned. *Arrows* indicate the insertion site of introns. The alignments suggest accurate excisions. In intronless genes from *D. takahashii*, *D. kitumensis*, and *D. microlabis*, the base just before the 5′ splicing site is A instead of the consensual G.

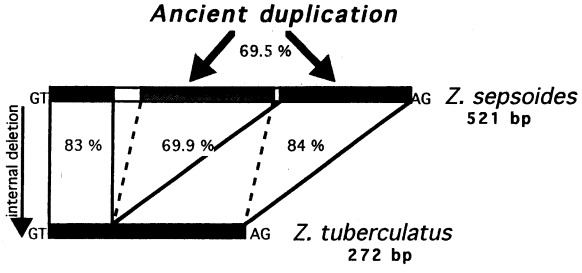

Fig. 5. Schematic representation of the internal duplication of the *Zaprionus sepsoides* intron, and relationships with *Z. tuberculatus*, showing the deletion of a duplicate in the latter species. Percentages of homologies between the different parts of the sequences are indicated, as computed by the CLUSTALV program.

subgroup seems variable, like in the *D. obscura* group. It may be due to a variable total number of amylase genes in the different species, as suggested by previous results (Da Lage et al. 1992). On the contrary, intronless copies

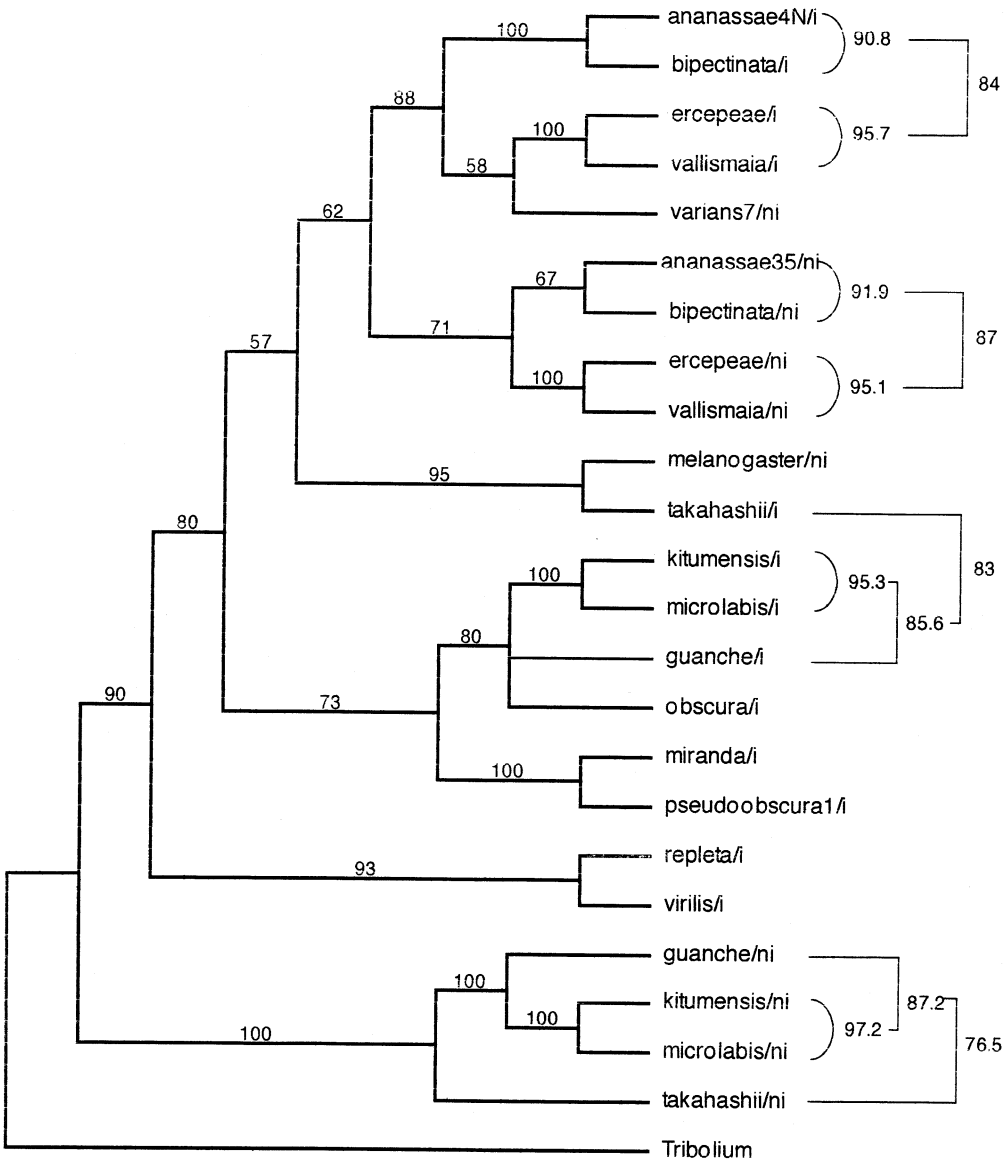

Fig. 6. Tree reconstruction with a parsimony method (PAUP program) using part of the coding sequences of the amylase genes and the sequences of *D. melanogaster*, *D. virilis*, *D. pseudoobscura*, and *Tribolium castaneum* already available (see text for references). Portions of 424 bp were first aligned with CLUSTAL prior to PAUP treatment. The tree is a consensus cladogram of 1,000 bootstrap replicates. Bootstrap values are indicated near the nodes. The tree was rooted with the flour beetle *Tribolium* (Coleoptera Tenebrionidae). To the right of the species-gene names some similarity scores are indicated computed by CLUSTALV (see text). *i:* genes with intron; *n.i.:* genes without intron.

of the *D. obscura* group are very divergent from all other *Amy* genes of *Drosophila* species, but remain clustered together, along with the *D. takahashii* intronless gene, outside the *Drosophila* cluster. In this case, intron loss occurred in a few genes before diversification of the group, probably a long time ago, although there may have been some accelerated changes in intronless copies. Hybridization of a blot from the gel in Fig. 2A with a *D. melanogaster* cDNA probe (pOR-M7, Boer and Hickey 1986) at a rather high stringency (0.5 × SSC; 63°C) has shown that lower bands (intronless genes) were barely labeled, confirming the divergence of these copies in all the species (not shown). The fact that these intronless copies remain clustered together (and with *D. takahashii*) far from the other Drosophila *Amy* genes does not support a pseudogenic status of these copies. Had they been orthologous pseudogenes, one could expect an important divergence between them. Instead, strong similarities have been kept, suggesting that these intronless copies have experienced selective pressures as strong as genes with introns. The differences of PCR patterns between American and Afro-European species could be due to differences in the number of copies, as suggested above, but also to differences in target sequences that would modify PCR results. Amylase evolution in the *D. obscura* group should be investigated further.

Size of Introns in Amylase Genes and Dynamics of Evolution

A result of our study is the high variability of intron size in amylase genes of Drosophilids. The ''standard'' size of the amylase intron of Drosophilids is consistent with previous surveys: Hawkins (1988) found that the most common size of Drosophila introns was about 70 bp. It is interesting to note that the overall pattern of intron size variation based on numerous genes from a single, or a few species, was similar to our findings on the amylase gene from numerous species (Fig. 1A). The increasing size from the ancestral 60 bp may have been progressive or brutal in different species through internal duplications of small or long fragments. Short repeats of a few bases may be undetectable because of the high evolutionary rate of introns (see Fig. 1D). Longer repeats such as the *Zaprionus sepsoides* one may increase intron size drastically but are likely to become less and less detectable as time goes on. In this species, the duplication, although previous to *Z. sepsoides/Z. tuberculatus* divergence, should be rather recent. Amplification through microsatellite expansion may have occurred in some cases. In *D. mercatorum,* although the intron is rather small, the (CT) repeat could be a hotspot for amplification. In *D. serrata,* a $(CT)_6$ repeat was also found. Intraspecific studies may be interesting to check the stability of intron size in these species. Long microsatellites within introns have been found in the second intron of the *engrailed* gene in *D. virilis,* but not in *D. melanogaster,* where the intron is shorter (Kassis et al. 1986). Exogenous insertions are also good candidates for increasing intron size. Brown et al. (1990) have described a 10-bp insertion within the introns of two of the three amylase genes they have sequenced in *D. pseudoobscura.* These three introns were otherwise almost identical due to concerted evolution. Another case of exogenous insertion is in *D. vallismaia* compared to *D. ercepeae.* Transposable elements have been reported to generate introns or to increase intron size (Purugganan and Wessler 1993; Giroux et al. 1992). However, we did not find any evidence for such a mechanism in amylase genes.

Although we have found introns of important size compared to the coding sequence (half of the coding sequence size in *D. kuntzei*), these sizes are not exceptional in insects. In many genes involved in development, long introns have been described: for example, 1,737 bp for intron 1 of *engrailed* in *D. virilis* (Kassis et al. 1986), 3.7 kb for intron 2 of *sevenless* in *D. melanogaster* (Bowtell et al. 1988), and over 20 kb in the *EcR* gene (Koelle et al. 1991), all in coding regions. In addition, very long introns may be found in 5′ untranslated regions but will not be considered here.

Regarding base compositions of introns, we have found a high variability of A-T content in short introns (standard size) which was positively correlated to the A-T content in surrounding exons. Base composition in these short introns may undergo the same selective pressure as the coding region. However, A-T enrichment of introns is twice that of exons (slope of the regression line, Fig. 1C). Long introns are all highly A-T rich, but also follow a positive correlation with exon A-T content, and the intercept is about 15% higher, as if there was in long introns a ''constitutive'' A-T richness. The origin of A-T enrichment of long introns is unclear. Several studies on various organisms have reported an overall AT-richness in introns (Mount et al. 1992 for Drosophila). Csank et al. (1990) reported that the overall ΔGC (%G-C exons–%G-C introns) is about 20% in *D. melanogaster,* and higher in *Caenorhabditis elegans* (Nematode), *Dictyostelium* (slime mold), and *Tetrahymena thermophila* (Protozoan). However in mammals, A-T content of introns may be low, and similar to flanking coding regions as well as noncoding regions, due to location of the considered gene in isochores (Ellsworth et al. 1994). Csank et al. (1990) have focused on *Tetrahymena,* in which introns have a very low G-C content. In contrast to our results, they have found that the A-T content was higher in small introns and that there was no correlation between intron and exon A-T contents. A-T richness has been reported to be important for splicing in plants (Goodall and Filipowicz 1989, quoted in Csank et al. 1990), but our results do not confirm that such base composition is important for splicing in animals since some of our sequences are A-T poor (*obscura* group). Guo et al. (1993) and Talerico and Berget (1994) have reported that a polypyrimidine tract was necessary at the 3′ splicing site in Drosophila for efficient splicing of long introns but not for short introns which often lack this polypyrimidine tract. Stephan et al. (1994) have compared the metallothionein genes of *D. ananassae* and *D. melanogaster* and found an important difference in intron sizes between these species, with the predicted difference in base composition at the 3′ end. In amylase genes, as expected, long introns have a C-T–rich end, but less expectedly, almost all short introns do have a polypyrimidine stretch, too.

Since the first papers on the evolutionary significance of introns (see Gilbert 1978), a large amount of data has been published on the possible functions of intervening sequences. It is now clear that genes-in-pieces allow a compaction and a modulation of genetic information by the multiplication of transcripts from a single genomic locus through alternative splicing. An example is the *Broad complex* (BR-C) of *Drosophila melanogaster,* a locus involved in the fly development (DiBello et al. 1991). In Drosophila, some introns have also been reported to be *cis*-regulators of gene expression, in the *rough* gene (Heberlein and Rubin 1990), *engrailed* (Kassis et al. 1986), *sevenless* (Michael et al. 1990), and *Xdh* (Riley 1989), a few examples among many others. These

authors have compared the genes in *D. melanogaster* and *D. virilis* or *D. pseudoobscura.* The presence of conserved blocks within the intron lead them to assume that these portions were functionally important. An on/off regulation of transcription at the splicing level was found in several Drosophila genes (Bingham et al. 1988). Some introns have been reported to contain other genes, like the *sina* gene, included in an intron of *Rh4* in *D. melanogaster* (Neufeld et al. 1991). However, many introns have not been found to have any function and were often considered as relics of assembly of different functional domains in a far past (Gilbert 1978). Other introns have been suspected to have been inserted secondarily through transposition (Purugganan and Wessler 1993). In the case of amylase genes, intron comparison between remote species failed to show any conserved regions, except the splicing sites, so we think that these introns have no function. In addition, we have seen no differences in amylase activity between species that retained an intron or not.

Other authors have compared introns of various genes in various and sometimes nonrelated organisms in order to identify consensus sequences and other signals involved in splicing mechanisms (Csank et al. 1990; Mount et al. 1992). From an evolutionary point of view, however, such comparisons must be handled cautiously since they imply more or less a sort of unity or homogeneity of introns as a whole, while we have seen above that the functions of introns, if any, are diverse. Thus, we have compared comparable sequences, i.e., introns of a single gene, inserted in the same position, which is more rigorous in an evolutionary sense. One can assume that these introns have the same evolutionary significance since the genes in which they are inserted likely undergo similar selective constraints. Because the amylase genes of the 150 species of Drosophilids tested here are under similar selective pressure, the fact that we found in a single gene a variability similar to that observed by comparing introns of many Drosophila genes (Mount et al. 1992) suggests that the variability of some intron features (size, consensus sequences) is not linked to gene-specific constraints.

Acknowledgments. We thank Jean-Christophe Luciani and Frédérique Chartois for technical help, and Leonidas Tsacas for guiding us in Drosophilid systematics. This work was supported in part by a grant from the Groupement de Recherches et d'Etudes sur les Génomes (GREG n° 73/94).

References

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

Bingham PM, Chou T-B, Mims I, Zachar Z (1988) On/off regulation of gene expression at the level of splicing. Trends Genet 4:134–138

Boer PH, Hickey DA (1986) The alpha-amylase gene in *Drosophila melanogaster:* nucleotide sequence, gene structure and expression motifs. Nucleic Acids Res 14:8399–8411

Bowtell DDL, Simon MA, Rubin GM (1988) Nucleotide sequence and structure of the *sevenless* gene of *Drosophila melanogaster.* Genes Dev 2:620–634

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

Cariou M-L, Lachaise D, Sourdis J, Tsacas L, Krimbas C, Ashburner M (1988) New African species in the *Drosophila obscura* species group: genetic variation, differentiation and evolution. Heredity 61: 73–84

Cavalier-Smith T (1991) Intron phylogeny: a new hypothesis. Trends Genet 7:145–148

Csank C, Taylor FM, Martindale DW (1990) Nuclear pre-mRNA introns: analysis and comparison of intron sequences from Tetrahymena thermophila and other eukaryotes. Nucleic Acids Res 18: 5133–5141

Da Lage J-L, Cariou M-L, David JR (1989) Geographical polymorphism of amylase in *Drosophila ananassae* and its relatives. Heredity 63:67–72

Da Lage J-L, Cariou M-L (1993) Organization and structure of the amylase gene family. In: Tobari YN (ed) *Drosophila ananassae,* genetical and biological aspects. Japan Scientific Societies Press, Karger, Tokyo, pp 171–181

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

Daïnou O, Cariou M-L, David JR, Hickey D (1987) Amylase gene duplication: an ancestral trait in the *Drosophila melanogaster* species subgroup. Heredity 59:245–251

Di Bello PR, Withers DA, Bayer CA, Fristrom JW, Guild GM (1991) The *Drosophila Broad-Complex* encodes a family of related proteins containing zinc-fingers. Genetics 129:385–397

Doolittle WF (1987) The origin and function of intervening sequences in DNA: a review. Am Nat 130:915–928

Ellsworth DL, Hewett-Emmett D, Li W-H (1994) Evolution of base composition in the insulin and insulin-like growth factor genes. Mol Biol Evol 11:875–885

Geiss KT, Abbas GM, Makaroff CA (1994) Intron loss from the NADH dehydrogenase subunit 4 gene of lettuce mitochondrial DNA: evidence for homologous recombination of a cDNA intermediate. Mol Gen Genet 243:97–105

Gemmill RM, Levy JN, Doane WW (1985) Molecular cloning of alpha-amylase genes from *Drosophila melanogaster.* I Clone isolation by use of a mouse probe. Genetics 110:299–312

Gilbert W (1978) Why genes-in-pieces? Nature 271:501

Giroux MJ, Clancy M, Baier J, Ingham L, McCarthy D, Hannah LC (1994) *De novo* synthesis of an intron by the maize transposable element *Dissociation.* Proc Natl Acad Sci USA 91:12150–12154

Gloor G, Engels W (1991) Single fly DNA preps for PCR. Dros Inf Serv 74:148–149

Goodall GJ, Filipowicz W (1989) The AU-rich sequences present in the introns of plant nuclear pre-mRNAs are required for splicing. Cell 58:473–483

Gumucio DL, Wiebauer K, Caldwell RM, Samuelson L, Meisler MH (1988) Concerted evolution of human amylase genes. Mol Cell Biol 8:1197–1205

Guo M, Lo PCH, Mount SM (1993) Species-specific signals for the splicing of a short intron in vitro. Mol Cell Biol 13:1104–1118

Hawkins JD (1988) A survey on intron and exon lengths. Nucleic Acids Res 16:9893–9905

Heberlein U, Rubin G (1990) Structural and functional comparisons of the *Drosophila virilis* and *Drosophila melanogaster rough* genes. Proc Natl Acad Sci USA 87:5916–5920

Hickey DA (1979) The geographical pattern of an enzyme polymorphism in *D. melanogaster.* Genetica 51:1–4

Janecek S (1994) Sequence similarities and evolutionary relationships of microbial, plant and animal alpha-amylases. Eur J Biochem 224: 519–524

Kassis JA, Poole SJ, Wright DK, O'Farrell PH (1986) Sequence conservation in the protein coding and intron regions of the *engrailed* transcription unit. EMBO J 5:3583–3589

Keller EB, Noon WA (1985) Intron splicing: a conserved internal signal in introns of Drosophila pre-mRNAs. Nucleic Acids Res 13:4971–4981

Koelle MR, Talbot WS, Segraves WA, Bender MT, Cherbas P, Hogness DS (1991) The Drosophila EcR gene encodes an ecdysone receptor, a new member of the steroid receptor superfamily. Cell 67:59–77

Kristensen NP (1991) Phylogeny of extant Hexapods. In: Nauman ID (ed) The insects of Australia. Carlton, Victoria, Melbourne, Univ. Press, pp 125–140

Kukalova-Peck J (1991) Fossil history and the evolution of Hexapod structures. In: Nauman ID (ed) The insects of Australia. Carlton, Victoria, Melbourne, Univ. Press, pp 141–179

Michael WM, Bowtell DDL, Rubin GM (1990) Comparison of the sevenless genes of *Drosophila virilis* and *Drosophila melanogaster*. Proc Natl Acad Sci USA 87:5351–5353

Mount SM (1982) A catalogue of splice junction sequences. Nucleic Acids Res 10:459–472

Mount SM, Burks C, Hertz G, Stormo GD, White O, Fields C (1992) Splicing signals in Drosophila: intron size, information content, and consensus sequences. Nucleic Acids Res 20:4253–4262

Neufeld TP, Carthew RW, Rubin GM (1991) Evolution of gene position: chromosomal arrangement and sequence comparison of the *Drosophila melanogaster* and *Drosophila virilis sina* and *Rh4* genes. Proc Natl Acad Sci USA 88:10203–10207

Ohta T (1983) On the evolution of multigene families. Theor Popul Biol 23:216–240

Payant V, Abukashawa S, Sasseville M, Benkel BF, Hickey DA, David J (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

Pélandakis M, Higgins DG, Solignac M (1991) Molecular phylogeny of the subgenus *Sophophora* of *Drosophila* derived from large subunit of ribosomal RNA sequences. Genetica 84:87–94

Purugganan M, Wessler S (1993) The splicing of transposable elements and its role in intron evolution. In: McDonald JF (ed) Transposable elements and evolution. Kluwer Academic, Dordrecht, pp 28–36

Riley MA (1989) Nucleotide sequence of the *Xdh* region in *Drosophila pseudoobscura* and an analysis of the evolution of synonymous codons. Mol Biol Evol 6:33–52

Rogers JH (1990) The role of introns in evolution. FEBS Lett 268: 339–343

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

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463

Sharp PA (1985) On the origin of RNA splicing and introns. Cell 42:397–400

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

Stephan W, Rodriguez VS, Zhou B, Parsch J (1994) Molecular evolution of the metallothionein gene *Mtn* in the *melanogaster* species group: results from *Drosophila ananassae.* Genetics 138:135–143

Tadlaoui-Ouafi A (1993) Evolution structurale et moléculaire de la famille multigénique Amylase chez quelques Drosophilidae. Thesis, Université Pierre et Marie Curie, Paris, 142 pp

Talerico M, Berget SM (1994) Intron definition in splicing of small *Drosophila* introns. Mol Cell Biol 14:3434–3445