Genetica (2006) 128:145–158
DOI 10.1007/s10709-005-5578-y

Origin and evolution of the *Amyrel* gene in the α-amylase multigene family of Diptera

Frédérique Maczkowiak & Jean-Luc Da Lage*
*Populations, génétique et évolution UPR 9034, CNRS, Gif sur Yvette, 91198 Cedex, France; *Author for correspondence: (Phone: +33-1-69-82-37-27, Fax: +33-1-69-82-37-36, E-mail: jldl@pge.cnrs-gif.fr)*

Received 9 May 2005; Accepted 30 November 2005

Key words: amylase inhibitor, gene duplication, intron loss, molecular adaptation, Muscomorpha

Abstract

Alpha-amylase genes often form multigene families in living organisms. In Diptera, a remote paralog, *Amyrel*, had been discovered in Drosophila, where this gene is currently used as a population and phylogenetic marker. The putative encoded protein has about 40% divergence with the classical amylases. We have searched the presence of the paralog in other families of Diptera to track its origin and understand its evolution. *Amyrel* was detected in a number of families of Muscomorpha (Brachycera-Cyclorrapha), suggesting an origin much older than previously thought. It has not been found elsewhere to date, and it is absent from the *Anopheles gambiae* genome. The intron–exon structures of the genes found so far suggest that the ancestral gene (before the duplication which gave rise to *Amyrel*) had two introns, and that subsequent, repeated and independent loss of one or both introns occurred in some Muscomorpha families. It seems that the Amyrel protein has experienced specific amino acid substitutions in regions generally well conserved in amylases, raising the possibility of peculiar, functional adaptations of this protein.

Introduction

The alpha-amylase gene family has long been used as a model in evolution of multigene families (Gumucio et al., 1988; Inomata, Tachida & Yamazaki, 1997; Da Lage, Maczkowiak & Cariou, 2000; Zhang et al., 2003a). Indeed, the *Amy* gene, which encodes an enzyme whose basic function is degrading starch and other polysaccharides, is duplicated in many organisms from animal, plant, fungal and bacterial kingdoms. Adaptive interpretations for this general trend to *Amy* gene multiplication have to date remained somewhat elusive. Indeed, the duplication events may have occurred independently, even in closely related taxa (Zhang et al., 2003a, 2003b). Levels of divergence between duplicates are variable among species. For instance, in human, there are five copies expanding on 200 kb along chromosome 1, three of which are salivary and two are pancreatic, with an aminoacid divergence limited to 3.3% (access number NG_004750). In the frog *Xenopus tropicalis*, two tandem copies have diverged up to 22% (http://genome.jgi-psf.org). In contrast, *Amy* copies may also keep high similarity with each other, especially when they are in physical vicinity. This is the case in a number of Drosophila species, where the *Amy* family has been investigated in details (Brown, Aquadro & Anderson, 1990; Popadic & Anderson, 1995 Da Lage, Maczkowiak & Cariou, 2000; Inomata & Yamazaki, 2000;). In *Drosophila ananassae*, the family is one of the most complex cases known so far in animals, with seven copies which diverged from one another from 0.5–41% in aminoacids (Da Lage, Maczkowiak & Cariou, 2000), and with or without an intron. The most divergent copy, named *Amyrel* (Amylase-related), was firstly described in some Drosophila species and was thought to be restricted to the *Sophophora* subgenus of Drosophila, thus making

it a recent acquisition (Da Lage et al., 1998). In Drosophila, the Amyrel protein is characterized by 40% divergence with the Amy protein, and has an putative additional disulfide bridge, locking the ''greek key'' of the domain C of the enzyme (Da Lage, van Wormhoudt & Cariou, 2002). The amylolytic ability of Amyrel was shown *in vitro* by transient expression in S2 cells (unpublished results). However, using living flies, no clear activity was evidenced in classical electrophoresis gels revealed by the starch–iodine reaction: the Amyrel protein could be recognized because analysis of the protein sequence shows that the electric charge of Amyrel is highly negative, compared to the classical alpha-amylase (Da Lage, van Wormhoudt & Cariou, 2002). The intron–exon structure is different from that of *Amy*, with a single intron in the middle of the coding sequence. In *D. melanogaster*, *Amyrel* is expressed only in larvae, unlike *Amy*, which is produced also in adults.

Amyrel was later found in many Drosophila species, but also in *Ceratitis capitata*, a member of the family Tephritidae. In this species, both classical and *Amyrel* genes share a two-exon structure, the first intron corresponding to that present in *Amy* genes of Drosophila and the second one corresponding to the intron present in *Amyrel* genes of Drosophila (Da Lage, van Wormhoudt & Cariou, 2002). The presence of *Amyrel* in a remote family demonstrated that *Amyrel* appeared much older than previously thought. In addition, the intron found in Drosophila was no more *Amyrel*-specific, and several amino acid substitutions considered characteristic of Amyrel in the first publication (Da Lage et al., 1998) were in fact not. This finding prompted us to look for *Amyrel* genes in other Diptera, to try discovering when *Amyrel* appeared and reconstituting the subsequent structural and functional evolution of this gene, including intron losses or gains. As the function of Amyrel remains unknown, we attempted to approach its putative functional specificities by looking for features specific to the encoded protein, such as changes in otherwise conserved amino acids or motifs. *Amyrel* has proved to be a good phylogenetic and phylogeographic marker in Drosophila (Rodriguez-Trelles, Tarrio & Ayala, 2000; Cariou et al., 2001). Indeed, a better knowledge of its origin could be valuable to use it for phylogenies in other Diptera.

Materials and methods

Biological samples

Most flies used in this study came from frozen or ethanol-preserved stocks, except *Megaselia scalaris* (Phoridae) and Drosophila species, which were bred in the laboratory. The species and their taxonomic positions are listed in Table 1.

Extraction and amplification of DNA

Genomic DNAs were extracted from single individuals as described in Da Lage, van Wormhoudt and Cariou (2002). For each sample, combinations of a set of primers designed in conserved regions of animal alpha-amylases (Table 2 and Figure 2) were used to amplify amylase sequences by PCR, without discrimination between putative *Amy* or *Amyrel* genes. After initial results, additional primers (not shown in Table 2) were designed for further extension of fragments in several species. Amplifications were performed with the AmpliTaq Gold polymerase (Applied Biosystems) with the following conditions: enzyme activation: 94°C, 6 mn; denaturation: 94°C, 30 s; annealing: 50–60°C, 50 s; elongation: 72°C, 1–2 mn; 45 cycles. The amplified fragments were cloned in TA-cloning vectors and sequenced on a ABI373 sequencer. Genbank accession numbers are indicated in Table 1. In addition to these partial sequences, an *Amy* gene from *Megaselia scalaris* (Phoridae) was completely sequenced using an internal PCR, followed by classical inverse PCR technique to get 5′ and 3′ ends of the gene (GenBank acccession AF467104). We obtained *Amy* genes from the Nematocera *Bibio marci* (Bibionidae) with the Universal Genome Walker kit (Clontech) after initial PCR with our sets of primers (GenBank AY082795-AY193771). Except sequences from mosquitoes and from Drosophila, all sequence data were obtained or updated for this study.

Analyses

Intron-containing sequences were first suspected by their length longer than expected in the absence of intron. Then, introns were detected by comparison of the translated sequences with known diptera amylases: the presence of an intron introduced stops or frameshifts, easily detected. Then,

Table 1. List of Brachycera species studied. Some species were not precisely identified (n.i.). The species named "Cypselidae" was a non identified Sphaeroceridae

Infra order	Superfamily	Family	Species	Acc. Numbers
Aschiza	Platypezoidea	Phoridae	*Megaselia scalaris*	AF467104
	Syrphoidea	Syrphidae	n.i.	DQ021951-54
		Pipunculidae	n.i.	DQ021944
Schizophora	Nerioidea	Neriidae	*Chaetonerius sp.*	DQ021917 DQ021926
	Diopsoidea	Diopsidae	n.i.	DQ021935
	Tephritoidea	Tephritidae	*Ceratitis capitata*	AF146757 AF146758
			Dacus oleae	DQ021932-34
	Sciomyzoidea	Sciomyzidae	*Tetanocera ferruginea*	DQ021949 DQ021950
		Coelopidae	*Coelopa frigida*	DQ021927 DQ021928
	Opomyzoidea	Asteiidae	*Echidnocephalus barbatus*	DQ021922 DQ021923
	Carnioidea	Chloropidae	*Thaumatomyia notata*	DQ021955 DQ021956
	Sphaeroceridea	Sphaeroceridae	*Anatalanta aptera*	DQ021918
			"*Cypselidae*"	DQ021931
	Ephydroidea	Curtonotidae	n.i.	DQ021929 DQ021930
		Drosophilidae	*Drosophila melanogaster*	X04569 AF022713
			Drosophila kikkawai	AB035055 U96156
			Drosophila ananassae	AF024691 U53698 U53477
			Drosophila pseudoobscura	U20331 U82556
			Drosophila funebris	AF335556 AF335557
			Drosophila repleta	AF048776 AY736496
			Hirtodrosophila confusa	AF335558 AF335559
			Leucophenga maculata	DQ021937-39
	Hippoboscoidea	Glossinidae	*Glossina morsitans*	DQ021936
	Muscoidea	Scathophagidae	*Scathophagida sp.*	DQ021946-48
		Anthomyiidae	*Delia radicum*	DQ021919-21
		Muscidae	*Musca domestica*	DQ021940-43
	Oestroidea	Calliphoridae	*Calliphora vomitaria*	DQ021924 DQ021925
		Sarcophagidae	*Sarcophaga carnaria*	DQ021945

Table 2. Primers used for PCR amplification in this study. Most may be used for other animals. Positions are numbered following *D. melanogaster Amy* sequence. DIPTEREV is Diptera-specific

Primer name	Séquence	Position	sense
1U+	GTNCAYCTNTTYGARTGG	94–111	+
12U+	TTYGARTGGAARTGGDVNGAYATHGC	103–128	+
WWERYQP	GGTGGGARCGTTAYCARCC	221–239	+
KHM	GAYGCNNNNAARCAYATGTGGC	610–631	+
DNHD	GAYAAYCAYGAYAAYCARCG	907–926	+
2U+	GGYTGRTANCNYTCCCACCA	220–239	–
INTR2+	CAGRTCTCCRGGCCACAT	625–642	–
DIPTEREV	ACGGAATCCAGCAACACCC	588–606	–
CEHREV	TGNCKCCANCGRTGYTCRCA	1144–1163	–
REV1230+	TTGCTGCCRTTRTCCCACCA	1219–1238	–
REV1400+	CCNGADATNAYRTCRCARTA	1339–1358	–

canonical intron boundaries (5′ splicing site, "donor" GT; 3′ splicing site "acceptor" AG) were identified. Translated protein sequences were easily aligned manually. To assign a sequence to either *Amy* or *Amyrel*, gene trees were constructed from aligned protein sequences, either by Neighbor joining (Saitou & Nei, 1987) using CLUSTALW (Thompson, Higgins and Gibson, 1994), or by maximum parsimony, using PAUP* (Swofford, 2002). Protein sequences were used for phylogenetic reconstruction, preferably to nucleotide sequences, because synonymous positions were largely saturated, except between Drosophila species (see the analysis of the rate of evolution in the Results section). Sequences which were clustered with *Amyrel* genes of Drosophila with high support were considered as *Amyrel* orthologs. Sequences which were clustered with *Amy* genes of Drosophila were assigned as "classical" *Amy* genes.

Results

Amyrel *is present in many Brachyceran lineages*

In most species of our sample, we found several gene copies, demonstrating that multiple Amylase gene copies are widespread in Diptera. In *Anopheles gambiae*, the genome analysis with an amylase sequence of Drosophila as a query shows the presence of four genes. The first cluster (*AmyA-AmyB*) is a tandem with an intergenic distance of 4 kb (contig AAAB01008859). Another contig on another chromosome (AAAB01008986) has a third copy (*Amy1*) with a pseudogene located 250 kb apart. In a third contig (AAAB01008960), there is a fourth complete amylase gene (*Amy2*), already described as a salivary amylase of adult females (AF469165, Francischetti et al., 2002). The amylase sequences obtained from the different species investigated here were used for drawing unrooted trees, with addition of some published sequences from Drosophila and *Aedes aegypti* (Figure 1). Due to the variable length of sequence available for the different genes, several trees were built. Only two are shown here. All methods gave a clear branching of *Amyrel* genes of *Drosophila* and *C. capitata* with genes amplified from flies belonging to the families Drosophilidae (Steganinae), Neriidae, Coelopidae, Sciomyzidae, Asteiidae, Chloropidae, Sphaeroceridae, Curtonotidae, Scathophagidae, Anthomyiidae, Muscidae, Calliphoridae, all these families being in the clade Schizophora. This enabled us to assign these sequences as orthologous to the *Amyrel* genes of Drosophila with no ambiguity, and thus, these genes were defined as *Amyrel*. Indeed, the *Amyrel* cluster in the trees is the most robust one. All the other sequences amplified with our primers are non-*Amyrel* amylase genes. The *Amyrel* cluster has no clear relationship with any of the other clusters of the trees. The bootstrap values are often rather low, but we still may assign the status of Drosophila *Amy* ortholog to a set of genes, clustered along with *Drosophila* and *C. capitata Amy*, used as a reference (bootstrap value 775/1000 in Figure 1b). The genes branched closest to the Drosophila *Amy* genes tend to reflect the phylogenetic relationships of the families, e.g. in the tree Figure 1b, where the closest relative of Drosophila *Amy* genes is the Drosophilidae (Steganinae) *L. maculata*, and then the Curtonotidae (another Ephydroidea) and farther, other Acalyptrates. In *Dacus oleae*, no *Amyrel* gene was detected, unlike its relative *C. capitata*, but three amylase copies were found, two of which are pseudogenes due to the presence of stop codons. One copy had also an insertion of 21 nucleotides in a usually conserved region. Another gene with a stop codon was found in *Scathophagida*. The three pseudogenes are not shown in the trees. The topologies of the trees show that some amylase genes in several species are not directly related to those of Drosophila (considered as our references) and experienced specific evolutionary histories. All genes from Nematocera branch off early in the trees. Interestingly, the absence of an *Amyrel* ortholog in the complete genome of *A. gambiae*, not just an absence of amplification, suggests that *Amyrel* could be restricted to Brachycera. In Nematocera, it is possible to clarify some orthology relationships: *Amy1* of *A. aegypti* and *Amy2* of *A. gambiae*; *AmyA-AmyB* of *A. gambiae* and *Amy2* of *A. aegypti*; *Amy1* of *A. gambiae* and *Amy2* of *B. marci*. Identifying orthology between the Nematocera *Amy* and those of Brachycera is not possible with our data.

It seems that a third kind of gene could be identified. Although it is not the main subject of this paper, it is worth noting it. The trees tend to cluster genes which diverge from both *Amyrel* and

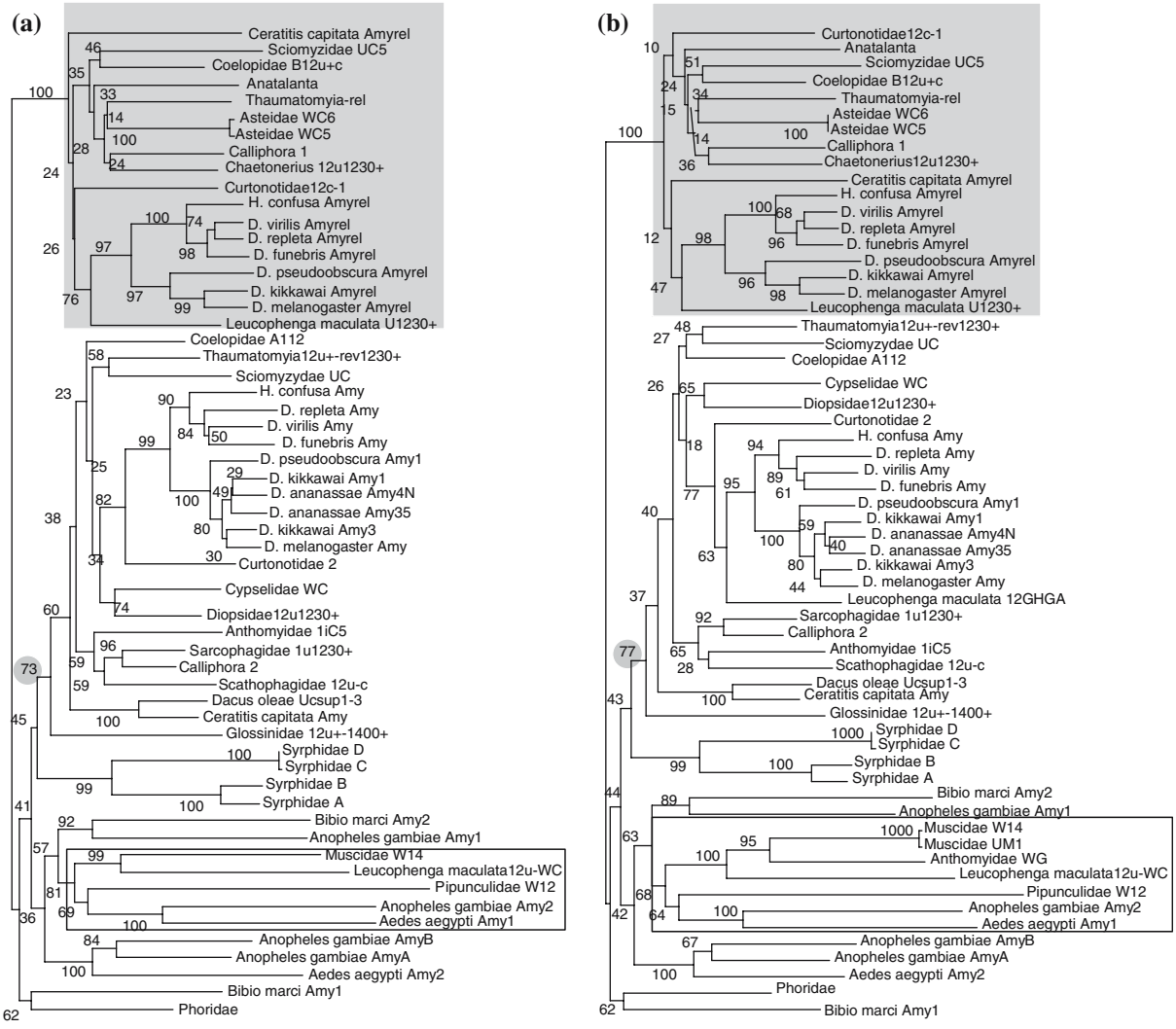


Figure 1. Two trees of partial amino acid sequences of Diptera amylases, reconstructed with NJ method. One thousand bootstrap replicates were done. Bootstrap values are shown as percentages. (a) from sequence data between the motifs ISYKL and RVMSS (282 amino acid positions); (b) between the motifs ISYKL and DNHD (239 amino acid positions), which lack the GHGA motif. Grey boxes: Amyrel cluster; open boxes: third type of amylase gene; the root of the "classical" Amy cluster is indicated by the grey circle surrounding the bootstrap value.

classical *Amy*. These genes have an novel structure with a third, phase 2 intron (intron 3). In Brachycera, these unusual genes were found, strongly clustered with a bootstrap value of 99%, (thus probably orthologous) in *Musca domestica* (Muscidae), *Delia radicum* (Anthomyidae) and quite surprisingly in the Drosophilidae *Leucophenga maculata*. This is surprising because this gene is clearly absent from Drosophila. In *L. maculata*, there is a third intron in phase 1, at the same position as in *Bibio marci Amy1*. We have to date no other sample of *Leucophenga* to definitely confirm the presence of this copy, but it was checked by several PCR on several individuals. Indeed, genes with the intron 3 may correspond to another paralog, whose presence is still not well documented. The intron 3 position was also found in Nematocera *Aedes aegypti* (*Amy1*), *Anopheles gambiae* (*Amy2*) and *Bibio marci* (*Amy1*), but this is in itself not an evidence of orthology. However,

the trees show these genes clustered together, except *Bibio marci Amy1*, but with moderate support. The gene found in the Pipunculidae shows another struture, with a novel intron absent from the other genes, but at a position already reported in e.g. Lepidoptera. Otherwise this gene seems related in the tree to the third kind of gene.

As a summary, we have detected *Amyrel* to date in a number of Brachycera families, all of which are Schizophora. Until now we did not detect this paralog in Phoridae, Pipunculidae, Syrphidae (Aschiza) nor in some Schizophora, such as Diopsidae, Glossinidae, Sarcophagidae. Figure 2 summarizes the diversity of intron–exon structures of the sequences found in each species, with their *Amy* or *Amyrel* assignation.

Origin and structural evolution of Amyrel

The proposed phylogenetic tree of Brachycera Muscomorpha on Figure 3 (McAlpine, 1989) shows the occurrence of *Amy* and *Amyrel* sequences found in our study. From this phylogenetical repartition, we propose that *Amyrel* arose from a duplication ancestral to Schizophora, from a template having both introns 1 and 2. It is clear that multiple intron losses occurred in the course of *Amy* and *Amyrel* evolution in Muscomorpha. Intron loss events are reported on Figure 3. For instance, it is likely that intron 1 was lost at least four times in *Amyrel*: in Nerioidea, Carnoidea, Ephydroidea, and in Muscoidea-Oestroidea. Concerning *Amy* genes, intron 1 was lost at least two times, in Nerioidea and some Drosophilidae (among which *D. melanogaster*). Intron 2 was lost in Phoridae and Drosophilidae for *Amy* genes, and only in *Coelopa frigida* (Coelopidae) for *Amyrel*.

Substitutions and motifs specific to the Amyrel *protein*

Amyrel genes have been until now identified by their clear branching with *Amyrel* of Drosophila. We have looked for changes in the protein, which could be specific to Amyrel, and thus might be indicative of adaptive specificities. The trees of Figure 1 have shown that *Amyrel* genes are tightly clustered, and well distinct from *Amy* genes. That is, the Amyrel proteins share a global similarity. All Amyrel protein sequences lack a GHGA motif in the "flexible loop" of the enzyme (Da Lage et al., 1998; Strobl et al., 1998). This motif, whose absence was once considered as a characteristic of Amyrel, is not diagnostic, because the GHGA motif is also lacking in clearly non-*Amyrel* genes, such as *A. gambiae Amy1*. More over, the tree built with sequences truncated upstream of this motif shows that it is not sufficient to change the clustering (Figure 1b), that is, the Amyrel cluster is not driven by the GHGA motif. In order to find out substitutions which could be specific to Amyrel, we have considered amylase sequences of not only Diptera, but also other animals. In the first description of *Amyrel* in Drosophila (Da Lage et al., 1998), the presence of a fifth putative disulfide bond, which was absent from Drosophila *Amy* genes was considered typical of Amyrel. In fact, these two additional cysteins are present in the *Amy* genes of *C. capitata*, *Megaselia scalaris*, *Calliphora* and several *Amy* genes of Nematocera: *B. marci*, *A. aegypti* and *A. gambiae* (except *Amy1*). This fifth disulfide bond of the protein could be ancestral in insects, since it is present in other insect Orders, and it may have been lost in some *Amy* genes, such as those of Drosophila. Using our data set, we have been able to find interesting amino acid modifications in Amyrel, in regions otherwise well conserved in animals (Figure 4). The first motif is a replacement of the conserved DV**V**F**N**HM by DV(**I/L**)**L**NHM in the β_3 sheet; the second motif is a replacement of the V**R**N motif, highly conserved in animals before the α_3 helix, by V**Q**(N/Q/R/H); the third motif, in β_4, is a modification of the conserved motif KHM**WP** by KHM**AA** (mainly) or other replacements of the bulky tryptophane and proline by smaller residues; the fourth motif, in the β_5 sheet, is the replacement of EVID(L/M)G(V/G/S) by EVID(H/Y)G(H/G/Y), i.e. replacement of small residues by bigger, cyclic ones; the last motif we have considered is the modification of the very well conserved GPP, after the β_8 sheet (see Figure 4). Another substitution deserves interest, although it is limited to a few Amyrel sequences. In the conserved motif **R**(V/I)MSSY (β8 sheet), the arginine is changed in a glutamine (**Q**(V/I)MSSY. This arginine is thought to play a role in chloride fixation by the enzyme (D'Amico, Gerday & Feller 2000). This

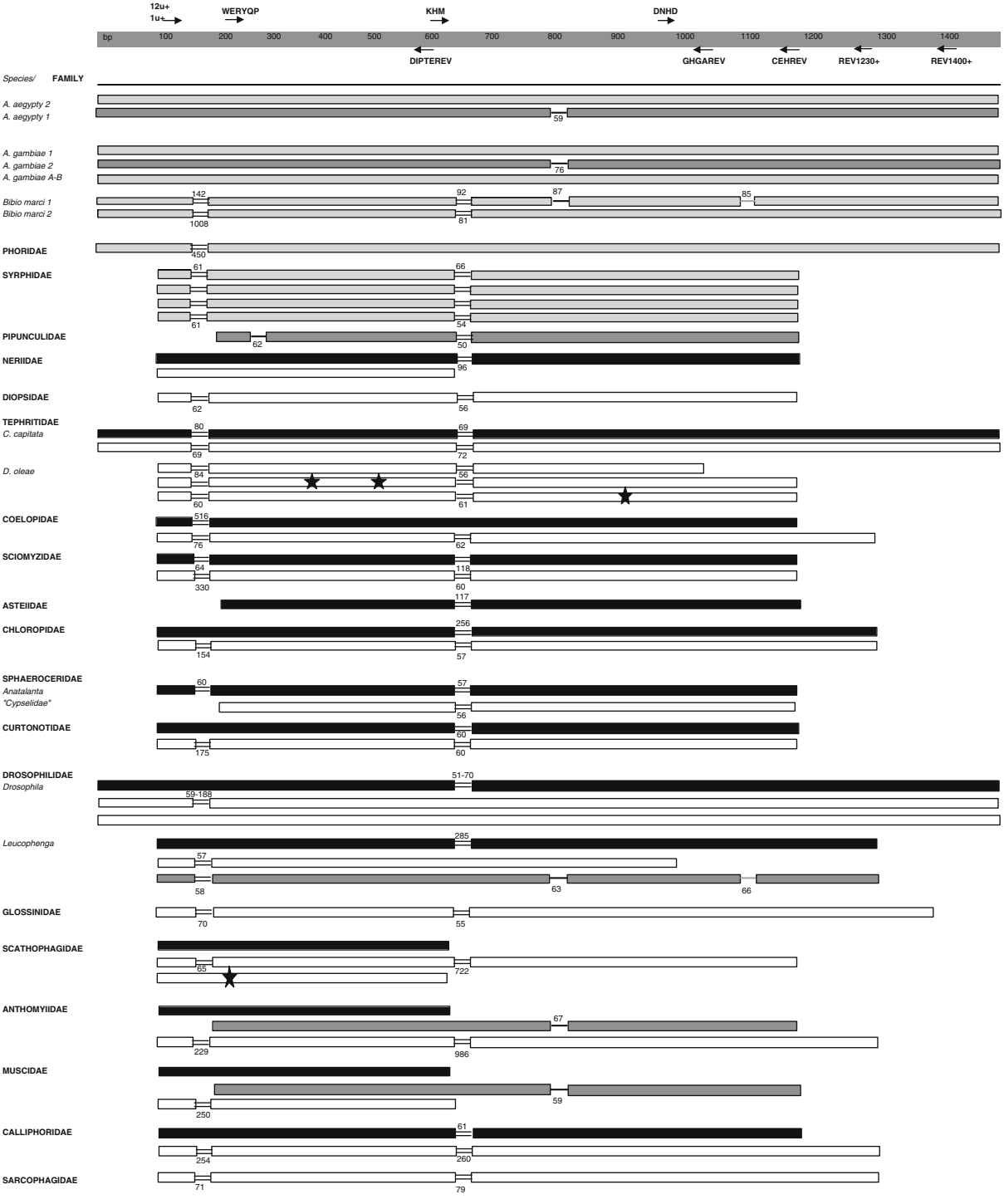
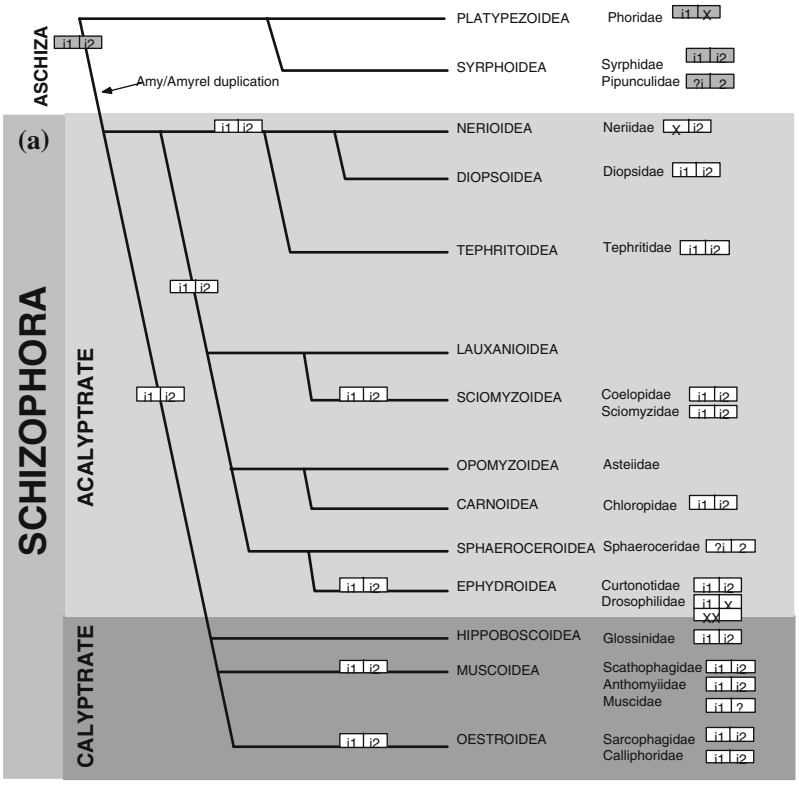


Figure 2. Diagram of the amylase genes found in Diptera. Black thick bars: *Amyrel*; white thick bars: *Amy*; dark grey thick bars: third kind of amylase; light grey thick bars: unclassified amylases. White thin bars: phase zero introns; grey thin bars: phase one introns; black thin bars: phase two introns. Stars represent stop codons. Numbers are intron lengths. The scale and positions of PCR primers are shown at the top of the figure.


(a)
SCHIZOPHORA
ASCHIZA
ACALYPTRATE
CALYPTRATE
Amy/Amyrel duplication
PLATYPEZOIDEA
SYRPHOIDEA
NERIOIDEA
DIOPSOIDEA
TEPHRITOIDEA
LAUXANIOIDEA
SCIOMYZOIDEA
OPOMYZOIDEA
CARNOIDEA
SPHAEROCEROIDEA
EPHYDROIDEA
HIPPOBOSCOIDEA
MUSCOIDEA
OESTROIDEA
Phoridae
Syrphidae
Pipunculidae
Neriidae
Diopsidae
Tephritidae
Coelopidae
Sciomyzidae
Asteiidae
Chloropidae
Sphaeroceridae
Curtonotidae
Drosophilidae
Glossinidae
Scathophagidae
Anthomyiidae
Muscidae
Sarcophagidae
Calliphoridae

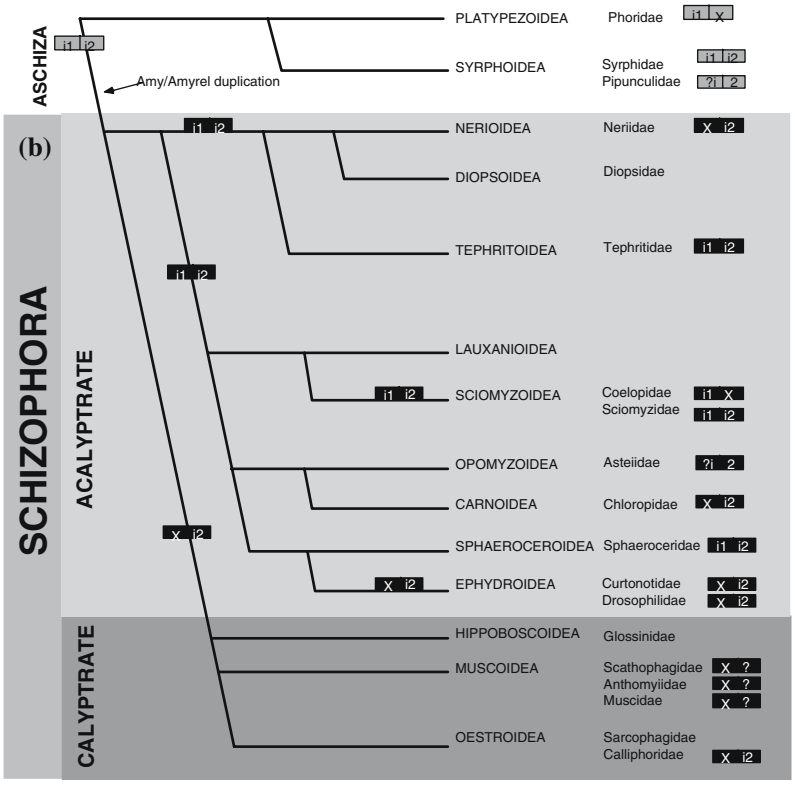

(b)
SCHIZOPHORA
ASCHIZA
ACALYPTRATE
CALYPTRATE
Amy/Amyrel duplication
PLATYPEZOIDEA
SYRPHOIDEA
NERIOIDEA
DIOPSOIDEA
TEPHRITOIDEA
LAUXANIOIDEA
SCIOMYZOIDEA
OPOMYZOIDEA
CARNOIDEA
SPHAEROCEROIDEA
EPHYDROIDEA
HIPPOBOSCOIDEA
MUSCOIDEA
OESTROIDEA
Phoridae
Syrphidae
Pipunculidae
Neriidae
Diopsidae
Tephritidae
Coelopidae
Sciomyzidae
Asteiidae
Chloropidae
Sphaeroceridae
Curtonotidae
Drosophilidae
Glossinidae
Scathophagidae
Anthomyiidae
Muscidae
Sarcophagidae
Calliphoridae

Figure 3. A tree of Muscomorpha (after McAlpine, 1989) showing the structural evolution of (a) *Amy* genes (white boxes) and (b) *Amyrel* genes (black boxes), after the *Amy*/*Amyrel* duplication. Introns 1 or 2 are indicated by i1 or i2, respectively. Intron losses are indicated by X. Amylase genes before duplication or unclassified types are shown as grey boxes. Boxes at branch tips show extant structures found in this study; boxes along branches are the deduced ancestral structures.

change is limited in our data set to Drosophila species which are not member of the *Sophophora* subgenus, and to *L. maculata*.

Comparison of base composition in Amy and Amyrel *genes*

In the first description of *Amyrel* (Da Lage et al., 1998), it was suggested that the base composition was somewhat less biased in *Amyrel* than in *Amy*. To check whether it is a general trend, we have compared the nucleotide composition of *Amy* and *Amyrel* in Muscomorpha species for which available sequence was sufficiently long for both paralogs. It concerns partial sequences of *Coelopa* (Coelopidae), *Tetanocera* (Sciomyzidae), *Thauma-*

Sus scrofa	VDAVINHM	YQVRDCQL	ASKHMWP	EVIDLGG	NQRGHGAGGS	GPPN
Crassostrea	ADVVINHM	NEVRNCRL	AAKHMWP	EVIDMGH	NQRGHGGGGG	GPPH
Tribolium	VDTVINHM	SNVRNCEL	AAKHMWP	EVIDLGG	NQR---TGGS	GPPQ
C. capitata	VDVVFNHM	YEVRNCEL	AAKHMWP	EVIDLGG	NQRGHGAGGA	GPPT
D. oleae	VDVVFNHM	YEVRNCEL	AAKHMWP	EVIDLGG	NQRGHGAGGA	GPPT
Scathophagidae	VDIILNHM	NQVRNCEL	AAKHMWP	EVIDLGG	NQRGHGAGGE	GPPT
Calliphoridae	VDIVFNHM	NQVRNCEL	AAKHMWP	EVIDMGG	NQRGHGAGGA	GPPT
Sarcophagidae	VDIVFNHM	NQVRNCEL	AAKHMWP	EVIDMGG	NQRGHGAGGA	GPPT
Coelopidae	VDVVFNHM	YQVRNCEL	AAKHMWP	EVIDLGG	NQRGHGAGGD	GPPT
Diopsidae	VDVVFNHM	ENVRNCEL	AAKHMWP	EVIDMGS	NQRGHGAGGA	GPPT
Cypselidae	VDVVFNHM	QQVRNCEL	AAKHMWP	EVIDMGG	NQRGHGAGGA	GPPT
Curtonotidae	VDVVFNHM	YNVRNCEL	AAKHMWP	EVIDLGS	NQRGHGAGGA	GPPT
L. maculata	VDVVFNHM	TEVRNCEL	AAKHMWP	EVIDLGG	???????????	????
D. melanogaster	VDVVFNHM	NEVRNCEL	AAKHMWP	EVIDMGG	NQRGHGAGGA	GPPT
D. ananassae 35	VDVVFNHM	NQVRNCEL	AAKHMWP	EVIDMGG	NQRGHGAGGA	GPPT
D. ananassae 4N	VDVVFNHM	NQVRNCEL	AAKHMWP	EVIDMGG	NQRGHGAGGA	GPPT
D. pseudoobscura	VDVIFNHM	NQVRNCEL	AAKHMWP	EVIDMGG	NQRGHGAGGA	GPPT
D. kikkawai 1	VDVIFNHM	NQVRNCEL	AAKHMWP	EVIDMGG	NQRGHGAGGA	GPPT
D. kikkawai 3	VDVVFNHM	NEVRNCEL	AAKHMWP	EVIDMGG	NQRGHGAGGA	GPPT
D. funebris	VDVVFNHM	TNVRNCEL	AAKHMWP	EVIDLGS	NQRGHGAGGA	GPPT
H. confusa	VDVVFNHM	ANVRNCEL	AAKHMWP	EVIDLGG	NQRGHGAGGA	GPPT
D. repleta	VDVVFNHM	TNVRNCEL	AAKHMWP	EVIDLGG	NQRGHGAGGA	GPPT
D. virilis	VDVVFNHM	TNVRNCEL	AAKHMWP	EVIDLGG	NQRGHGAGGA	GPPT
Sciomyzidae	VDVVFNHM	EQVRNCEL	AAKHMWP	EVIDMGG	NQRGHGAGGE	GPPT
Chloropidae	VDVVFNHM	NEVRNCEL	AAKHMWP	EVIDLGS	NQRGHGAGGA	GPPT
Glossinidae	VDIVFNHM	REVRNCEL	AAKHMWP	EVIDMGG	NQRGHGAGGS	GPPT
C. capitata	VDVLLNHM	YQVQNCEL	AAKHIPA	EVVDHGY	NQR----EGG	APPA
L. maculata	VDALLNHM	YQVQNCEL	AAKHMAA	EVIDHGH	NQR------D	APPM
D. melanogaster	VDVLLNHM	FQVQQCEL	AAKHMAS	EVIDHGH	NQR----DAG	PPPQ
D. pseudoobscura	VDVLLNHM	YQIQHCEL	AAKHMAS	EVIDHGG	NQR----DGG	APPQ
D. kikkawai	VDVLLNHM	FQVQQCEL	AAKHMAA	EVIDHGH	NQR----DMG	PPPQ
D. funebris	VDVLLNHM	FQVQQCEL	AAKHMAA	EVIDHGH	NQR----DGG	APPQ
H. confusa	VDVLLNHM	FQVQHCEL	AAKHMAA	EVIDHGH	NQR----DGG	APPQ
D. repleta	VDVLLNHM	FQVQQCEL	AAKHMAA	EVIDHGH	NQR----DGG	APPQ
D. virilis	VDVLLNHM	FQVQQCEL	AAKHMAA	EVIDHGH	NQR----DGG	APPQ
Curtonotidae	VDVILNHM	FQVQQCEL	AAKHMAA	EVIDHGH	NQR----DSG	PPPM
Sphaeroceridae	VDVLLNHM	KQVQRCEL	AAKHVAA	EVIDYGH	NQR----DGG	????
Neriidae	VDVLLNHM	YQVQNCEL	AAKHMAA	EVIDHGH	NQR----DGG	GPPA
Calliphoridae	VDVLLNHM	YQVQNCEL	AAKHMKA	EVIDHGH	NQR----DGG	PPPH
Scathophagidae	VDVILNHM	FQVQNCEL	???????	???????	???????????	????
Anthomyidae	VDVLLNHM	YEIQNCEL	???????	???????	???????????	????
Muscidae	VDVILNHM	YQVQNCEL	???????	???????	???????????	????
Chloropidae	VDVLLNHM	YQIQNCEL	AAKHMAA	EVIDHGG	NQR----DGG	SPPM
Asteidae	VDVILNHM	YQVQNCEL	AAKHMDA	EVIDHGH	NQR----DGG	PPPQ
Coelopidae	VDVILNHM	WQVQNCEL	AAKHMAA	EVIDHGH	NQR----DGG	PAPA
Sciomyzidae	VDVILNHM	YQVQNCEM	AAKHMEA	EVIDHGH	NQR----DGG	ASPM

Figure 4. Alignments of conserved regions of animal amylases, in which specific substitutions were found in Amyrel (light shading in Amyrel vs. dark shading in Amy). The box shows three non-diptera species: pig (*Sus scrofa*), oyster (*Crassostrea gigas*), beetle (*Tribolium castaneum*).

tomyia (Chloropidae), Curtonotidae, *Calliphora* (Calliphoridae), *Leucophenga* (Drosophilidae), with lengths of 950–1100 bp for *Amyrel* and 1040–1200 bp (except *L. maculata*:800 bp) for *Amy*; and full-length sequences (ca. 1500 bp) for all the *Drosophila* species and *C. capitata* (Tephritidae) (17 *Amy*/*Amyrel* pairs compared). That is, for each sequence, the maximum length available was used to compute the base composition. Note that shortening the sequences to the smallest one of the set does not change the values significantly. Content in G + C at third codon position (GC3) is positively correlated between *Amy* and *Amyrel* ($r = 0.90$; $P < 0.01$, $n = 17$). This probably reflects the general trend of each species for global base composition. GC3 values are generally higher in *Amy* genes than in *Amyrel* (Wilcoxon test is significant; $n = 15$, $P < 0.01$), indicating a more constrained codon usage.

In Drosophila, it has been shown that in *Amy* genes, most introns are short and have a variable base composition, correlated to the composition of surrounding coding sequence (Da Lage, Wegnez & Cariou, 1996). We have used our data set of Muscomorpha to compare *Amy* and *Amyrel* from this point of view. For *Amy* genes, we considered both introns 1 and 2 separately; for *Amyrel*, since there are few introns 1, we only considered introns 2. Introns are generally short, and show a high variability in their AT content; long introns, not frequent, are always AT-rich (Figure 5a and b). We focused on short introns only. In *Amy* genes, introns 1 (from 38 to 81% AT, mean = 63%, $\sigma = 12.3$, $n = 26$) are more variable than introns 2 (from 57 to 84% AT, mean = 71%, $\sigma = 8.2$, $n = 17$). Introns 2 are slightly more AT-rich than introns 1. In *Amyrel*, short introns 2 show a lower AT content (mean = 59%, $\sigma = 10.1$, $n = 23$). The base composition of introns is correlated to that of the coding sequence available (781 bp of coding region), for both introns in *Amy* genes, and for intron 2 in *Amyrel* (Figure 5c and d). Coefficients of correlation are $r = 0.782$ ($P < 0.01$, $n = 27$) for *Amy* intron 1, $r = 0.59$ ($P < 0.01$, $n = 18$) for *Amy* intron 2, $r = 0.83$ ($P < 0.01$, $n = 17$) for *Amyrel* intron 2.

Rate of evolution compared in Amy *and* Amyrel

The same set of species, for which available sequences were longest in both paralogs (see above), was used to measure the divergence between Amy and Amyrel proteins. The overall amino acid divergence between the paralogs within species is comprised between 31% (*Coelopa frigida*) and 40% (*Drosophila*) in the region available (273 amino acid positions, except *L. maculata*, 240 positions). We have also used the same set of species (keeping only three Drosophila species: *D. melanogaster*, *D. virilis*, *H. confusa*) to compare the rates of non-synonymous substitutions dN (Nei and Gojobori, 1986) between species for each paralog *Amy* or *Amyrel*. Using 800 nucleotide positions for *Amy* and 953 for *Amyrel*, we obtained a strong positive correlation in pairwise species comparisons (10 species, 45 pairwise comparisons) for dN *Amyrel* values plotted vs. dN *Amy* values: $r = 0.81$, $n = 45$, $P < 0.01$. The slope of the linear regression was 0.91, suggesting a similar rate of non-synonymous substitutions for *Amy* and *Amyrel*. Due to saturation of synonymous positions, the rates of synonymous substitutions dS, and therefore the ratio dN/dS, were not applicable for a number of pairwise comparisons, for both paralogs and with all methods implemented in MEGA3 (Kumar, Tamura & Nei, 2004). When computable, most dS values were well above 1, except between Drosophila species.

Discussion

In order to look for the origin of *Amyrel*, we have investigated a number of families of Diptera from the clade Muscomorpha (i.e. Brachycera Cyclorrapha), with also a sampling from Nematocera, using personal data and the genome of *Anopheles gambiae*. Muscomorpha are monophyletic, whereas Nematocera are probably not (Yeates & Wiegmann, 1999). We have been able to detect by PCR, and to sequence, partial amylase genes, which belonged to different types of amylase genes. Some of them could be classified without any ambiguity as *Amyrel* genes, due to their strong clustering to the *Amyrel* sequences of Drosophila and *Ceratitis capitata*, used as references. We have shown that *Amyrel* is present in Muscomorpha Schizophora, whereas its presence in Muscomorpha Aschiza has not been demonstrated. Non detection of *Amyrel* in some species of our sample may be due to limitations of PCR techniques. However, Southern blotting could not easily be used instead, first because of the low amount of

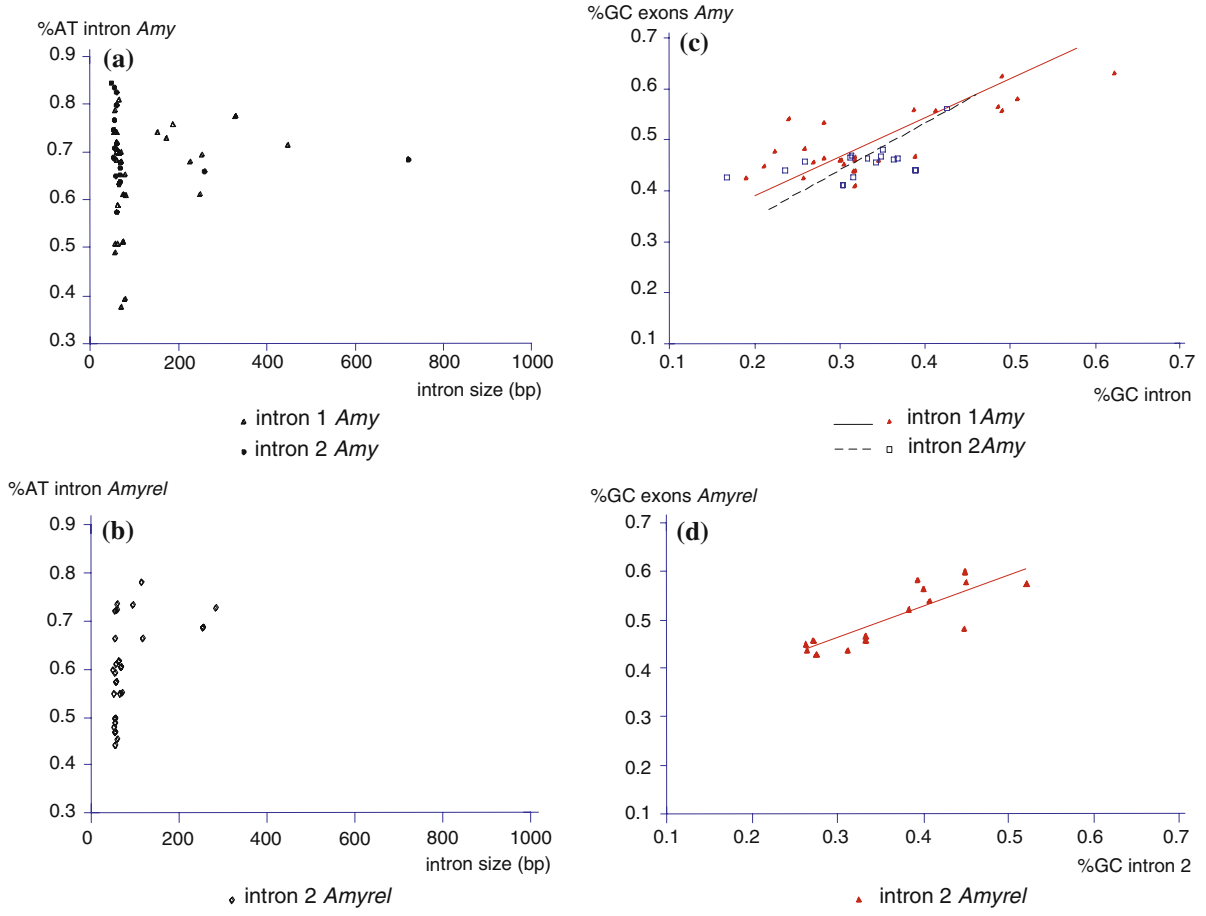


Figure 5. Analyses of the base compositions of introns 1 and introns 2 in *Amyrel* and *Amy* genes. (a) AT content in *Amy* introns 1 and 2; (b) AT content in *Amyrel* introns 2. Data from the *D. melanogaster* species subgroup were added (eight values); (c) correlation between base composition in introns and exons in *Amy* genes; (d) correlation between base composition in introns 2 and exons in *Amyrel* genes.

material for most species, and also because interspecific *Amyrel* probes could fail to hybridize to their putative target due to long-time sequence divergence in stringent conditions, or on the contrary, hybridize as well to non *Amyrel* copies at a rather necessary low stringency. An attempt with *Megaselia scalaris* (Phoridae), for which DNA was unlimited, with a Drosophila *Amyrel* probe, was unsuccessful. *Amyrel* might be absent from Nematocera, considering its absence in the Anopheles genome. That is, the duplication could have occurred during the Jurassic period (Wiegmann et al., 2003). Our data suggest that the gene ancestral to *Amy* and *Amyrel* in Schizophora had two introns (here referred to as intron 1 and intron 2), a structure that was preserved in some families, such as Tephritidae and Sciomyzidae. It is clear that for both paralogs, intron losses have occurred independently several times, as shown in Figure 3. It must be noted, however, that the phylogeny of Muscomorpha is still unclear, and that our counting of the loss events depends on the topology. In addition this count is underestimated. For instance, it is known that inside the Drosophila genus, intron 1 was lost several times in *Amy* (Da Lage, Wegnez & Cariou, 1996). In Muscomorpha, intron 2 was lost frequently in *Amy* genes, but not in *Amyrel*, except in *Coelopa frigida*. Strikingly, in the Drosophila genus, the *Amyrel* intron (intron 2) has never been lost in almost 200 species studied (unpublished). In Amylase genes, intron loss is probably not adaptive, as suggested earlier (Da Lage, Wegnez & Cariou, 1996). Our results show that intron loss is a frequent phenomenon, but it

affects each intron differently. The mechanism for intron loss is not ascertained, but it is often attributed to recombination in germline with a reverse transcript, which is rarely full-length, and thus affects preferably the 3′ part of the gene (Derr and Strathern, 1993; Geiss, Abbas & Makaroff, 1994). Therefore, the intron loss pattern in *Amyrel*, which makes intron 1 more prone to deletion than intron 2, is surprising. In contrast, for *Amy*, we have until now no example of gene with intron 2 only. It suggests that the mechanisms could be different for the two genes. For example, losing the first intron only could be due to recombination with an intronless amylase copy between intron 1 and intron 2. Also, are the protosplice sites (Long & Rosenberg, 2000; Sadusky, Newman & Dibb, 2004; Sverdlov et al., 2004) involved in this difference? ''Protosplice'' refers to the few coding nucleotides surrounding introns, which are probably involved in the efficacy of splicing, and are of the AG/GT or at least G/G type, where the slash indicates the intron position. The protosplice sites are often not canonical for intron 2 in many species (e.g. AA/AC in *Glossina Amy*, AA/AT in *Calliphora Amyrel*), while they are of the typical AG/GT sequence for intron 1 in most cases. Indeed, although the protosplice sequence seems to be involved in intron gain (Stoltzfus, 2004), its putative role in the mechanism and frequency of intron loss is unknown. The presence of these two introns is not limited to Diptera. Intron 1 is widespread and is considered to be very old in animals (Da Lage, Wegnez & Cariou, 1996), and intron 2 has been found also in the amylase gene of the shrimp *Litopenaeus vannamei* (Sellos & van Wormhoudt, 2002). In this latter case, it is not easy to infer whether the intron is ancestral to these Arthropods or if there have been two independent insertions, which, with the current data, is more likely. If we consider that introns 1 of *Amy* and *Amyrel* are homologous (respectively introns 2), given that they derive from a single ancestral gene, it would be appropriate to compare them between the two paralogs. We cannot perform this comparison between introns 1, because there are few data from *Amyrel* genes. But we have shown that introns 2 are more AT rich in *Amy* than in *Amyrel*. However, this is a global result. The base compositions should be compared for each species. Unfortunately, we have only five species for which intron 2 sequence is known for both paralogs (Calliphoridae, Curtonotidae, Chloropidae, Sciomyzidae, Tephritidae). We must also note that all intron losses are ''clean'', that is, they do not remove any coding nucleotide and do not leave any intron fragment.

We have shown that the base composition of *Amy* and *Amyrel* short introns is not independent from that of the surrounding coding sequences. This had been shown earlier for intron 1 in Drosophila *Amy* genes (Da Lage, Wegnez & Cariou, 1996), but no data were available for *Amyrel*, and for the second intron position. The previous result is then confirmed at a broader scale, although not clearly for intron 2 in *Amy* genes. The reason of such correlations has been investigated (Akashi, Kliman & Eyre-Walker, 1998; Kliman & Eyre-Walker, 1998). The authors have compared numerous genes within a genome, and suggest regional composition heterogeneity or mutational pattern bias. Our observations are somewhat different in that a single gene has been compared in several species, but the results are not inconsistent with the hypothesis of a consequence of specific mutation patterns, reflected in the global base composition of each species as well as in both exons and introns composition of amylase genes. It must be said also that *Amy* and *Amyrel* may be located in distant chromosomal regions. For instance, in *Drosophila ananassae*, they are on different chromosomes (Da Lage, Maczkowiak & Cariou, 2000).

Our data suggest a similar rate of protein (nonsynonymous) evolution for Amy and Amyrel. However, this average and long term estimate masks some irregularities: in the *Sophophora* subgenus of Drosophila, we have observed that Amyrel evolved almost twice as fast as Amy (Da Lage et al., 1998). And it is likely that after duplication, Amyrel underwent a series of quick changes corresponding to adaptive amino acid substitutions. Indeed, in search of clues for adaptation of the putative Amyrel protein to particular conditions, we have tried to find some specific amino acid changes, shared by all Amyrel sequences in otherwise highly conserved regions. The Amyrel protein is characterized first by the absence of the GHGA motif in the glycine-rich ''flexible loop'' (Strobl et al., 1998; Ramasubbu, Ragunath & Mishra, 2003). Although it is not specific to Amyrel (see above, Results), we think that this absence (probably loss) is indeed adaptative, perhaps in relation to

interactions with inhibitors (Da Lage, van Wormhoudt & Cariou, 2002; Kluh et al., 2005). Other amino acid changes have been pointed out above. Interestingly, the changes often occur in the vicinity of highly constrained regions. For instance, fixation of the vital calcium ion requires several interactions with some residues, three of which are very close to the Amyrel substitutions: the asparagine of the motif VFNHM, the arginine of VRN, the histidine of KHMWP. In the case of the arginine, this residue itself is replaced in Amyrel by a glutamine. However, the amino acid is bounded to calcium by the oxygen of the peptide bond (D'Amico, Gerday & Feller 2000), not by the lateral chain. Thus, the change of amino acid should not inhibit this fixation. Other changes occurred in regions of unknown function, but still highly conserved. Further knowledge of the amylase enzymology is needed to understand better the putative adaptive changes in Amyrel. Indeed, the function of Amyrel is still unknown, but since this paralog is widespread in flies, it must have a function. We can assume that this protein is adapted to other conditions or substrates, or as suggested above, against some inhibitors. It is also worth noting that we have very few examples of *Amyrel* pseudogenes, which suggests that it is fully functional. Some frameshift deletions have been found until now only in a Drosophila, *Zaprionus lineosus* (unpublished), and in a heterozygous state, in *Drosophila sechellia* (D. Legrand, personal communication). In contrast, *Amy* pseudogenes have been found in several Drosophila species (Popadic et al., 1996; Da Lage et al., 2003) and in this study too. Although probably active, *Amyrel* may be, in a number of species, expressed at a lower level than *Amy*, considering the GC3 as an indicator of expression level (Shields et al., 1988), since we have noted a lower GC3 in *Amyrel* than in *Amy*.

Duplications are frequent in amylase genes, and Diptera themselves show a variety of cases, concerning the number and the divergence of copies. Duplications may lead to tissue specificity, like in *Aedes aegypti* (Francischetti et al., 2002), or *Drosophila ananassae* (Da Lage, Klarenberg & Cariou, 1996). Tissue specificity, mainly due to changes in regulatory regions, may be itself associated to enzymological specificity, also not evidenced to date in such insects. In the case of the *Amy–Amyrel* pair in Muscomorpha, evolutionary forces have favoured the maintaining of the *Amyrel* copy, and its high divergence (up to 40%) in amino acids, with specific substitutions, which could confer original properties to the protein, such as withstanding inhibitors or toxins, or improved activity on new substrates. Data on biochemical properties of the Amyrel protein will help in the future, to understand what are the driving forces of this molecular adaptation.

Acknowledgements

We are grateful to V. Sarrant and A. Fradet for technical help and to two anonymous reviewers for their fruitful comments.

References

Akashi, H., R.M. Kliman & A. Eyre-Walker, 1998. Mutation pressure, natural selection, and the evolution of base composition in *Drosophila*. Genetica 102/103: 49–60.

Brown, C.J., C.F. Aquadro & W.W. Anderson, 1990. DNA sequence evolution of the amylase multigene family in *Drosophila pseudoobscura*. Genetics 126: 131–138.

Cariou, M.-L., J.-F. Silvain, V. Daubin, J.-L. Da Lage & D. Lachaise, 2001. Divergence between *Drosophila santomea* and allopatric or sympatric populations of *D. yakuba* using paralogous amylase genes and migration scenarios along the volcanic line. Mol. Ecol. 10: 649–660.

D'Amico, S., C. Gerday & G. Feller, 2000. Structural similarities and evolutionary relationships in chloride-dependent α-amylases. Gene 253: 95–105.

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

Da Lage, J.-L., C. Maisonhaute, F. Maczkowiak & M.-L. Cariou, 2003. A nested alpha-amylase gene in *Drosophila ananassae*. J. Mol. Evol. 57: 355–362.

Da Lage, J.-L., E. Renard, F. Chartois, F. Lemeunier & M.L. Cariou, 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., A. vanWormhoudt & M.-L. Cariou, 2002. Diversity and evolution of the alpha-amylase genes in Animals. Biologia Bratislava 57: 181–189.

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

Da Lage, J.-L., A. Klarenberg & M.-L. Cariou, 1996. Variation in sex-, stage- and tissue-specific expression of the amylase genes in *Drosophila ananassae*. Heredity 76: 9–18.

Derr, L.K. & J.N. Strathern, 1993. A role for reverse transcripts in gene conversion. Nature 361: 170–173.

Francischetti, I.M.B., J.G. Valenzuela, V.M. Pham, M.K. Garfield & J.M.C. Ribeiro, 2002. Toward a catalog for the transcripts and proteins (sialome) from the salivary gland of the malaria vector *Anopheles gambiae*. J. Exp. Biol. 205: 2429–2451.

Geiss, K.T., G.M. Abbas & C.A. Makaroff, 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.

Gumucio, D.L., K. Wiebauer, R.M. Caldwell, L.C. Samuelson & M.H. Meisler, 1988. Concerted evolution of human amylase genes. Mol. Cell. Biol. 8: 1197–1205.

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

Inomata, N. & T. Yamazaki, 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.

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

Kluh, I., M. Horn, J. Hyblova, J. Hubert, L. Doleckova-Maresova, Z. Voburka, I. Kudlikova, F. Kocourek & M. Mares, 2005. Inhibitory specificity and insecticidal selectivity of α-amylase inhibitor from Phaseolus vulgaris. Phytochemistry 66: 31–39.

Kumar, S., K. Tamura & M. Nei, 2004. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Brief Bioinform. 5(2): 150–163.

Long, M. & C. Rosenberg, 2000. Testing the ''proto-splice sites'' model of intron origin: evidence from analysis of intron phase correlation. Mol. Biol. Evol. 17: 1789–1796.

McAlpine, J., 1989. Phylogeny and classification of the Muscomorpha, pp. 1397–1518 In Manual of Neartic Dipterain, edited by M. Wood & J. McAlpine. Agriculture Canada Monograph, Ottawa.

Nei, M. & T. Gojobori, 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3: 418–426.

Popadic, A. & W.W. Anderson, 1995. Evidence for gene conversion in the amylase multigene family of *Drosophila pseudoobscura*. Mol. Biol. Evol. 12: 564–572.

Popadic, A., R.A. Norman, W.W. Doane & W.W. Anderson, 1996. The evolutionary history of the amylase multigene family in *Drosophila pseudoobscura*. Mol. Biol. Evol. 13: 883–888.

Ramasubbu, N., C. Ragunath & P.J. Mishra, 2003. Probing the role of a mobile loop in substrate binding and enzyme activity of human salivary amylase. J. Mol. Biol. 325: 1061–1076.

Rodriguez-Trelles, F., R. Tarrio & F.J. Ayala, 2000. Evidence for a high ancestral GC content in *Drosophila*. Mol. Biol. Evol. 17: 1710–1717.

Sadusky, T., A.J. Newman & N.N. Dibb, 2004. Exon junction sequences as cryptic splice sites: implication for intron origin. Curr. Biol. 14: 505–509.

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

Sellos, D.Y. & A. van Wormhoudt, 2002. Structure of the alpha-amylase genes in crustaceans and molluscs: evolution of the exon/intron organization. Biologia Bratislava. 57: 191–196.

Shields, D.C., P.M. Sharp, D.G. Higgins & F. Wright, 1988. ''Silent'' sites in *Drosophila* genes are not neutral: evidence of selection among synonymous codons. Mol. Biol. Evol. 5: 704–716.

Stoltzfus, A., 2004. Molecular evolution: introns fall into place. Curr. Biol. 14(9): R351–352.

Strobl, S, K. Maskos, M. Betz, G. Wiegand, R. Huber, F.X. Gomis-Ruth & R. Glockshuber, 1998. Crystal structure of the yellow meal worm alpha-amylase at 1.64A resolution. J. Mol. Biol. 278: 617–628.

Sverdlov, A.V., I.B. Rogozin, V.N. Babenko & E.V. Koonin, 2004. Reconstruction of ancestral protosplice sites. Curr. Biol. 14: 1505–1508.

Swofford, D.L., 2002. PAUP*. Phylogenetic Analysis Using Parsimony *and Other Methods. Version 4.0b10., Sinauer Associates, Sunderland, MA.

Thompson, J.D., D.G. Higgins & T.J. Gibson, 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22: 4673–4680.

Wiegmann, B.M., D.K. Yeates, J.L. Thorne & H. Kishino, 2003. Time flies, a new molecular time-scale for Brachyceran fly evolution without a clock. Syst. Biol. 52: 745–756.

Yeates, D.K. & B.M. Wiegmann, 1999. Congruence and controversy: toward a higher-level phylogeny of Diptera. Annu. Rev. Entomol. 44: 397–428.

Zhang, Z., N. Inomata, M.-L. Cariou, J.-L. DaLage & T. Yamazaki, 2003a. Phylogeny and evolution of the amylase multigenes in *Drosophila montium* species subgroup. J. Mol. Evol. 56: 121–130.

Zhang, Z., N. Inomata, T. Yamazaki & H. Kishino, 2003b. Evolutionary history and mode of the amylase multigene family in Drosophila. J. Mol. Evol. 57: 702–709.