Amyrel, a paralogous gene of the amylase gene family in Drosophila melanogaster and the Sophophora subgenus

Jean-Luc Da Lage\textsuperscript{a}, Emmanuel Renard, Frédérique Chartois, Françoise Lemeunier, and Marie-Louise Cariou

Populations, Génétique et Evolution, Centre National de la Recherche Scientifique, 91198 Gif sur Yvette cedex, France

Communicated by Wyatt W. Anderson, University of Georgia, Athens, GA, March 26, 1998 (received for review January 16, 1997)

ABSTRACT We describe a gene from Drosophila melanogaster related to the alpha-amylase gene Amy. This gene, which exists as a single copy, was named Amyrel. It is strikingly divergent from Amy because the amino acid divergence is 40%. The coding sequence is interrupted by a short intron at position 655, which is unusual in amylase genes. Amyrel has also been cloned in Drosophila ananassae, Drosophila pseudoobscura, and Drosophila subobscura and is likely to be present throughout the Sophophora subgenus, but, to our knowledge, it has not been detected outside. Unexpectedly, there is a strong conservation of 5' and 3' flanking regions between Amyrel genes from different species, which is not the case for Amy and which suggests that selection acts on these regions. In contrast to the Amy genes, Amyrel is transcribed in larvae of D. melanogaster but not in adults. However, the protein has not been detected yet. Amyrel evolves about twice as fast as Amy in the several species studied. We suggest that this gene could result from a duplication of Amy followed by accelerated and selected divergence toward a new adaptation.

Acquisition of new biological functions is a main process of evolution. Because a new function involves new genes or new regulations of existing genes, a great deal has been focused on gene duplication events (1–3). Several processes may follow a gene duplication: concerted evolution that retains the similarity between duplicates, accelerated divergence toward putative new functions, or pseudogene formation. As molecular data have accumulated, it has been shown that many genes are members of multigene families having active or pseudogenic “companions” (3).

The amylase gene is an interesting model for studying evolution of multigene families. For over 30 years, it has been investigated widely in many organisms. Its enzymatic activity is revealed easily on electrophoresis gels (4). During the past decade, Amy genes were cloned and sequenced in a number of bacteria, fungi, plants, and animals. Alignments of AMY proteins have shown that, despite a high variability, a few blocks of amino acids were conserved (5, 6). In animals, many more amino acid stretches are conserved between species, and alignments remain easy.

Multicopy structures with various gene arrangements were found in various taxa: man and other Primates (7), rodents (8), and Crustacea (9). In Drosophilids, Drosophila melanogaster has two copies (10), Drosophila pseudoobscura has been found to have one and three copies (11), Drosophila eugracilis and Drosophila ficsiphila have two copies (12), and Drosophila ananassae has at least seven copies (13). It seems that multiplications (and loss?) of Amy genes have occurred independently in many animal lineages, raising the question of an adaptive advantage.

Focusing on amylase evolution in Drosophilids, we observed that some species harbored two types of genes: those with an intron at position 177, which is supposed to be ancestral (14), and those without an intron at this position. This was the case in Drosophila takahashii, Drosophila lucipennis, and in the Drosophila obscura group. Phylogenetic trees clearly showed that “intronless” genes of these species were excluded from the Amy tree but remained clustered together, suggesting that they were paralogs (14). Given the tree topology, we suspected that such a divergent gene might be present also in D. melanogaster. By using PCR primers specific to these genes, a fragment then was amplified from D. melanogaster. The present work describes the structure, chromosomal localization, and expression pattern of this gene, named Amyrel (for Amylase-related), in D. melanogaster and several species of the Sophophora subgenus and its evolutionary relationship with the classical Amy genes.

MATERIALS AND METHODS

The Canton-S strain of D. melanogaster was used. D. subobscura was from Montenêvre, France, D. pseudoobscura was from Phoenix, Arizona, and D. ananassae was from Tai, Ivory Coast. DNA extraction and PCR conditions have been described (14). The primers used are listed in Table 1. Inverse PCR was performed by digesting genomic DNA with four-cutter restriction enzymes and religating the diluted cut DNA. The circularized DNA then was amplified with relevant primers. PCR products were cloned into the pGEM-T vector (Promega) and sequenced with an automated device, ABI 373 (Applied Biosystems).

Genomic clones were obtained for D. melanogaster, Drosophila subobscura, and D. pseudoobscura; minilibraries were obtained by cloning in pUC plasmid digestion fragments of required size (previously identified by Southern analysis). PCR fragments of Amyrel were used as probes for screening. Positive clones were treated for nested deletions and were sequenced. A genomic clone from D. ananassae had been obtained (J.-L.D.L., unpublished data).

For mRNA detection, flies were reared on axenic, sugar-free medium at 25°C, and several individuals were sampled at various time points from embryo to adult. The second- to third-instar molt was considered to occur 72 h after egg laying. For each time point, at least three individuals were assayed. Reverse transcription (RT)–PCR protocol was adapted from Huet et al. (15). RNAs were roughly extracted (16). Samples (1/100 of the extract) were treated with DNase/RNasin before RT-PCR. Negative controls were made on DNase-treated

Abbreviation: RT, reverse transcription.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database [accession nos. U69607 (Amyrel D. melanogaster), U53698 (Amy35 D. ananassae), U53477 (Amy4N D. ananassae), U53479 (Amy4 D. ananassae), U79724 (AmyD D. subobscura), U80035 (Amy D. subobscura), and U82556 (Amyrel D. pseudoobscura)].

*To whom reprint requests should be addressed. e-mail: jldl@pge.cnrs-gif.fr.
results

The Amyrel Gene. In D. melanogaster, an internal part of Amyrel first was sequenced from a PCR product between primers Amyrel1 and Rev1230 and the surrounding regions by inverse PCR from –90 to 1699, relative to the translation start. Further data in flanking regions were obtained from genomic clones (two 4-kb-long PstI fragments containing the left and right part of Amyrel, respectively). Using the full coding sequence as a probe, we performed Southern hybridizations that suggested that Amyrel is a single-copy gene (Fig. 1). Sequences from D. subobscura, D. pseudoobscura and D. ananassae were obtained readily from full-length genomic clones.

The coding sequence is 1482 bp long in both D. melanogaster and D. ananassae. In D. pseudoobscura and D. subobscura, an additional codon lies within the putative peptide signal. The gene is interrupted at position 655 (or 658) by a short intron (56 bp in D. melanogaster and D. ananassae, 60 bp in D. subobscura, and 68 bp in D. pseudoobscura), inserted between two codons. The intron position is quite unusual in amylase genes and seems not to be ancestral. The comparison between Amyrel and Amy (intron removed) shows high substitution rates per site and an overall divergence close to 40% in nucleotides and amino acids (see below). The length of Amyrel is equal to Drosophila Amy. Tree constructions based on either nucleic or protein data (Fig. 2 is from protein data) show that the divergence between Amy and Amyrel is higher than between any classical Amy genes within the Drosophila genus, but Amyrel remains inside the insect branch.

Nucleotide Divergence and Codon Usage in Amy and Amyrel. Table 2 shows that the divergence between Amy and Amyrel is similar in the four species with Ka values around 0.37, suggesting that Amyrel has undergone similar selective constraints at the protein level in the different taxa. An important divergence from Amy also may have occurred in their common ancestor. In contrast, the synonymous rates are different between species and are correlated negatively to the C content in the third position and to the codon bias, as often reported (20–22). It is known that, in Drosophila, Amy is highly biased for its codon usage, especially in D. melanogaster (23) and D. pseudoobscura (24). In these two species, 88% of the Amy codons end with G or C, the latter being most preferred. It seems that Amyrel genes are a bit less constrained because of a lower codon bias. In D. ananassae, in which several classical but divergent Amy genes exist [Amy35 and Amy4N in our study (J.-L.D.L., unpublished data)], the codon bias for Amy is lower than in D. melanogaster or D. pseudoobscura and similar to that of Amyrel (Table 2). However, comparisons between Amy and Amyrel indicate that for most synonymous groups, the same codons are preferred in the two genes for the species studied. This finding may reflect the general C-ending preference reported in the Sophophora subgenus (21). But, of interest, the high bias toward codon TTC (Phenylalanine), which is a typical trait in all Drosophila Amy genes known to date, is common to both types of genes. Tables available for codon usage in pooled Drosophila genes (25) or Xdh or Adh do not show such a bias. On the other hand, the low usage of GGG (glycine) or TTA (leucine) in Amy and Amyrel is in accordance with the Drosophila general usage.

The divergence between the Amyrel genes is higher than between the Amy genes (Table 3). Amyrel seems to be a fast
evolving gene if compared with Amy and also with Adh, Adh-dap, or Xdh, for instance. Ka values between D. pseudoobscura and D. melanogaster are 0.07 ± 0.008, 0.05 ± 0.008, 0.04 ± 0.008, and 0.07 ± 0.004 for these four genes, respectively. The Ka value for Amyrel is 0.13 ± 0.012. The higher synonymous rates (Ks) in Amyrel would be correlated to the lower codon bias of this gene compared with Amy, which is known for its low synonymous rate (like Adh; see ref. 21). Also, the high ratios Ks/Ka are indicative of protein-coding capacity (26).

The AMYREL Protein. The conceptual product AMYREL is a protein of 493 aa (55.3 kDa) in D. melanogaster that differs by 42% from the classical AMY protein. The divergence is higher within the first 30 residues, which makes the N-terminal part difficult to align, except for the putative cleavage site for the peptide signal (23) (Fig. 3). Several amino acid stretches highly conserved in animal or even bacterial alpha-amylases (5, 6) are present in AMYREL. Cysteine residues that are involved in disulfide bridges (6) are conserved particularly (asterisks in Fig. 3), suggesting that AMYREL may be an alpha-amylase. However, the AMYREL proteins from the four species studied have lost the conserved motif GHGA positions 260 to 594 to sequence-tagged site Dm0827 (28), that were used instead. Fig. 5 shows the chromosomal labeling of long stretches of noncoding sequences (Fig. 4) is a striking and unexpected result, given the situation in Amy, in which only a few short motifs have been found to be conserved upstream to the gene (27). Within the 3′ region, a polyadenylation site, AATAAA, is usually found in Amy genes. In Amyrel, this motif is included in an almost perfectly conserved stretch of 60–70 bp. Significant homologies between downstream sequences remain visible for several hundreds of base pairs (not shown on Fig. 4): AACTGGASTTAGCTCTAACA (241–260); TGCC-WCGACAACGASA (263–278); CAGCTGRCACWCTGT (391–406); CTGCAATAGAARAAGKTSGCAT (423–446); and TGAAATTTGAGTGGSTGYTCTGG (449–472). The 3′ region in D. melanogaster corresponds from positions 260 to 594 to sequence-tagged site Dm0827 (28), that maps at the same locus.

In the 5′ flanking regions, three long motifs have been preserved during evolution (Fig. 4): (i) −183 to −169; (ii) −129 to −84; and (iii) −80 to −69. The latter has been recognized by Magoulas et al. (29) to be involved in amylase regulation. But, except for this case, our search in databases for similar motifs was not conclusive. The putative TATA box was found in the four sequences, but no obvious CAAT box has been detected so far.

Chromosomal Localization. A PCR fragment of the Amyrel coding sequence was used as a probe for D. subobscura. For D. melanogaster and D. ananassae, genomic adjacent fragments were used instead. Fig. 5 shows the chromosomal labeling of the biotinylated probes. In D. melanogaster, Amyrel maps at 53D1–3, (Amy is at 54A1B1; ref. 30 and our data). In D. subobscura, Amyrel and Amy map at 74A and 55D, respectively, at the two opposite tips of the acrocentric chromosome E. In D. ananassae, Amyrel is at position 76C on the 3L arm (see ref. 31 for cytogenetical nomenclature), distant from a cluster of classical genes (81C; ref. 13).

Amyrel Expression in D. melanogaster. RT-PCR experiments revealed that Amyrel is transcribed at the second and third larval instars with a maximum in young and middle third-instar larvae (Fig. 6). In contrast to Amy, no transcript was detected in adults. Tissue activity was found in the midgut and perhaps in the fat body (not shown). More refined experiments will be necessary to study tissue expression precisely.

DISCUSSION

Gene amplification techniques are powerful tools to detect duplicated copies of genes. Amyrel was discovered in the course of a PCR study on classical genes (14). Its high divergence with Amy may explain why it had not been characterized earlier.

Table 2. Intraspecific comparison between Amy and Amyrel in the four species studied

<table>
<thead>
<tr>
<th></th>
<th>melanogaster</th>
<th>pseudoobscura</th>
<th>subobscura</th>
<th>ananassae Amy35</th>
<th>ananassae Amy4N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall nucleotide divergence (%)</td>
<td>43.3</td>
<td>41.1</td>
<td>37*</td>
<td>41.5</td>
<td>43.7</td>
</tr>
<tr>
<td>Overall aminoacid divergence (%)</td>
<td>41.8</td>
<td>43.3</td>
<td>40*</td>
<td>40.6</td>
<td>41.2</td>
</tr>
<tr>
<td>Kα Amyrel/Amy</td>
<td>0.37</td>
<td>0.36</td>
<td>0.31*</td>
<td>0.37</td>
<td>0.36</td>
</tr>
<tr>
<td>Ks Amyrel/Amy</td>
<td>1.32</td>
<td>0.96</td>
<td>0.85*</td>
<td>1.08</td>
<td>1.37</td>
</tr>
<tr>
<td>Overall GC richness, Amyrel/Amy (%)</td>
<td>55.5/62.8</td>
<td>59.6/62.8</td>
<td>61.5/61*</td>
<td>57.4/57.7</td>
<td>57.4/54.4</td>
</tr>
<tr>
<td>%C 3rd position, Amyrel/Amy</td>
<td>36.7/62.0</td>
<td>46.2/63.1</td>
<td>51.1/52.4*</td>
<td>43.6/50.0</td>
<td>43.6/42.4</td>
</tr>
<tr>
<td>%G+C 3rd position, Amyrel/Amy</td>
<td>67.1/87.6</td>
<td>79.2/87.6</td>
<td>81.3/80.3*</td>
<td>74.1/73.8</td>
<td>74.1/63.4</td>
</tr>
</tbody>
</table>

Asterisks indicate that the sequence available is incomplete for Amy in D. subobscura. Percentages of C and G+C at the third codon position were calculated excluding codons ATG (Met) and TGG (Tryp). These values are indicative of the level of codon bias (22).
Indeed, it is most likely that Amyrel had been cloned first by Gemmill et al. (30) along with Amy in D. melanogaster. The actual chromosomal localization of Amyrel is the same as that of their clone λ Dm32. However, Gemmill et al. found no expression by Northern blotting and considered this copy as a pseudogene. Brown et al. (24) observed an in situ hybridization signal in D. pseudoobscura additionally to the Amy locus and also suspected a pseudogene.

Here, we have shown that Amyrel has a full-length coding sequence and is transcribed in larvae. This divergent gene has remarkable features, making it a special case in the amylase family of Drosophila. The intron–exon structure is very original and does not correspond to an ancestral state because the intron site is not shared in other genomic sequences available (insects or vertebrates). Indeed, this intron is spliced correctly, as revealed by RT-PCR using primers that surround the intron site (data not shown). Other data, such as significant codon bias and high Ks/Ka ratios, also are in favor of an active gene. However, AMYREL has not been detected by the usual technique for amylase electrophoresis. Given the number of charged residues, AMYREL should migrate faster than the classical AMY1 allele.

Another unexpected and interesting result is the high interspecific conservation of noncoding flanking regions. D. melanogaster, D. ananassae, and D. subobscura are not closely related, and the divergence time between these species (20–35 million years) is too long to allow conservation without selection. The pattern of highly conserved blocks distributed along a much less conserved sequence suggests a functional role for these motifs. In addition, further sequencing of these regions indicates that Amyrel might be surrounded very closely by other putative coding genes (J.-L. D.L., unpublished results), which could have an influence on the conservation of the flanking sequences.

The chromosomal localizations of Amyrel genes compared with Amy are variable; whereas both genes are in rather close regions in D. melanogaster (53D vs. 54A), they can be very distant in other species. Chromosomal rearrangements are likely responsible for this variation, but it is difficult to know which situation is ancestral because of the uncertain interspecific correspondences between chromosomal arms. However, the large distances between Amyrel and Amy in each species (even in D. melanogaster) suggest that they have evolved without any contact (unequal crossover or gene conversion). Moreover, such events, common in classical Amy genes that often are arranged tandem (24, 32), are avoided by the single-copy structure of Amyrel.

Until now, we have not found Amyrel outside the Sophophora subgenus, but it has been detected by PCR in Drosophila willistoni, which is considered the most divergent member of this subgenus. The confirmation that Amyrel is restricted to the Sophophora subgenus would indicate that the gene has undergone a very fast evolution, with a high rate of nonsynonymous substitutions compared with Amy: 40% amino acid substitutions in <60 million years (according to the estimated divergence time between D. melanogaster and Dro-

Table 3. Interspecific substitution rates per sites (Amyrel/Amy)

<table>
<thead>
<tr>
<th>species</th>
<th>melanogaster</th>
<th>subobscura</th>
<th>pseudoobscura</th>
<th>ananassae Amy35</th>
<th>ananassae Amy4N</th>
</tr>
</thead>
<tbody>
<tr>
<td>melanogaster</td>
<td>—</td>
<td>0.92/0.52*</td>
<td>—</td>
<td>1.00/0.38</td>
<td>0.89/0.65</td>
</tr>
<tr>
<td>subobscura</td>
<td>0.13/0.06*</td>
<td>—</td>
<td>0.41/0.22*</td>
<td>0.83/0.63*</td>
<td>0.83/1.32*</td>
</tr>
<tr>
<td>pseudoobscura</td>
<td>0.13/0.07</td>
<td>0.05/0.04*</td>
<td>—</td>
<td>1.05/0.53</td>
<td>1.05/0.84</td>
</tr>
<tr>
<td>ananassae Amy35</td>
<td>0.08/0.04</td>
<td>0.13/0.06*</td>
<td>0.15/0.07</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>ananassae Amy4N</td>
<td>0.08/0.04</td>
<td>0.13/0.07*</td>
<td>0.15/0.07</td>
<td>—/—</td>
<td>—</td>
</tr>
</tbody>
</table>

Ks (synonymous) are above the diagonals; Ka (non-synonymous) are below. Both types of classical Amy genes of D. ananassae have been compared. The asterisks indicate that computing has been done by using the partial sequence of D. subobscura and the corresponding sequences of the other species.
sophila virilis (33). Initially relaxed constraints on the new copy followed by positive selection may have allowed these changes. A recent intron insertion would have accompanied the differentiation of {\it Amyrel}. We also may notice that the rate of evolution of {\it Amyrel} among the four species studied so far seems to be about twice as fast as {\it Amy} (also suggested by Fig. 2). Data from other species will help in estimating the evolutionary dynamics of this new gene. It would be of interest to know whether the substitution rates have varied during evolution or among lineages. A recent study on a duplication of the {\it Adh} gene in the {\it Drosophila repleta} group (26) shows parallel results. In this group, a duplicated gene, {\it Adh}-Ψ, physically close to {\it Adh} formerly had been reported to be a pseudogene, but, like {\it Amyrel}, it shows a full-length ORF (with recruitment of codons from the upstream region) and exhibits a significant codon bias and a high Ks/Ka ratio. Also, despite its physical vicinity with {\it Adh}, which could facilitate gene conversion, {\it Adh}-Ψ evolves faster, and it has been shown that the substitution rate was higher in the past than recently. The same phenomenon might have occurred in {\it Amyrel}.

Duplications of {\it Amy} frequently have been found in the {\it Sophophora} subgenus but are not documented in other subgenera (12). As far as we know, the duplications were not followed by strong differentiation. However, {\it Amy} encodes a major digestive enzyme and may be subject to adaptation; a case of accelerated divergence in the {\it Amy} gene was observed in {\it Drosophila erecta} (34) and was attributed to adaptation to new resources. Similarly, the fast divergence of {\it Amyrel} and its
maintenance as an active gene also suggest an adaptation toward a new function or substrate. Further evidence for a new function is suggested by the amino acid differences between Amyrel and Amy in a number of usually conserved positions. In Adh-V of the D. repleta group, substitutions in conserved regions of the protein also have been considered to be adaptive to new function (26). Other Adh-related genes of Drosophila, FBP2 (35) and jingwei (36), were reported to have turned to novel functions, thus increasing genome potentialities. The esterase multigene family is another well known example of diversifying multiplication (37). In the amylase family, although regulatory differences are known (38, 39), Amyrel is a new example of such a structure and sequence divergence. For instance, the mouse salivary and pancreatic proteins are only 15% divergent. Biochemical studies will help in understanding the evolutionary meaning of this duplication and divergence as a paradigm of physiological adaptation through gene duplication.

We thank Professors M. Matsuda and Y.N. Tobari and Dr. E. Hauschteck-Jungen for precise identification of chromosomal localizations in D. ananassae and D. subobscura, respectively; N. Chami-nade for in situ hybridizations; and Dr. W. Stephan and two anonymous referees for fruitful comments. This work was supported by a grant from the Groupement de Recherches et d’Etudes sur les Génomes n° 73/94.