Pergamon

PII: S0965-1748(97)00063-5

Insect Biochem. Molec. Biol. Vol. 27, No. 8/9, pp. 769–781, 1997
© 1997 Elsevier Science Ltd
All rights reserved. Printed in Great Britain
0965-1748/97 $17.00 + 0.00

Evidence for Two Distinct Members of the Amylase Gene Family in the Yellow Fever Mosquito, *Aedes aegypti*

GENELLE L. GROSSMAN[1],† YVONNE CAMPOS[2],† DAVID W. SEVERSON,‡ ANTHONY A. JAMES†*

Received 30 April 1997; revised and accepted 31 July 1997

Genomic DNA fragments encoding a salivary gland-specific α-amylase gene, *Amylase I* (*Amy I*), and an additional amylase, *Amylase II* (*AmyII*) of the yellow fever mosquito, *Aedes aegypti*, were isolated and characterized. Two independently isolated DNA fragments, G34-F and G34-14A, encode polymorphic alleles of *Amy I*. A 3.2 kilobase (kb) *Eco*R I fragment of G34-F, F2, has been sequenced in its entirety and contains 832 base pairs (bp) of the 5′-end, non-coding and putative promoter regions that are adjacent to 2.4 kb of the *Amy I* coding region. One intron, 59 bp in length, is found towards the 3′-end of the clone. A third genomic clone, 3A, corresponding to *Amy II*, was sequenced and shown not to contain the primary DNA sequence that encodes the 260 amino acid region that uniquely characterizes the amino terminal end of the *Amy I* product. *Amy I* was assigned by restriction fragment length polymorphism (RFLP) mapping to chromosome 2 (23.0 cM) and *Amy II* to chromosome 1 (44.0 cM). *Amy I* and *Amy II* are highly polymorphic and there may be multiple linked copies at each locus. Comparisons between *Amy I* and *Amy II* are presented for the putative promoter and conceptual translation products. The identification of two distinct amylase genes and their separate linkage assignments provides evidence for a multigene family of α-amylases in *Ae. aegypti*. © **1997 Elsevier Science Ltd. All rights reserved**

Mosquito *Aedes aegypti* Amylase gene Gene mapping

INTRODUCTION

Genes encoding α-amylases (E.C. 3.2.1.1) have a wide phylogenetic distribution and are found in plants, bacteria, fungi and animals. Many examples of multiple amylase genes have been reported in the Arthropoda. The isopods (Crustacea), *Sphaeroma* and *Asellus*, have two and six genes, respectively, and five of the latter are polymorphic (Oxford, 1986; Laulier, 1988). In the Insecta, two linked amylase genes were detected in the housefly, *Musca domestica* (Ogita, 1968), the silkworm, *Bombyx mori* (Kikkawa, 1953), and in several species of the flour beetle, *Tribolium* (Pope *et al.*, 1986).

In the genus *Drosophila*, the genomic arrangement and number of amylase genes varies among species and strains of the fruitfly. *Drosophila melanogaster* has two amylase genes, *AmyP* and *AmyD*, that are located approximately four kilobases apart on the same chromosome, and are transcribed divergently (Levy *et al.*, 1985). Both *AmyP* and *AmyD* are expressed in the midgut during the larval and adult stages of the insect. Amino acid polymorphisms in members of the duplicated amylase loci account for isozyme variants with different electrophoretic mobilities in different strains of this insect. In the tropical fruitfly, *Drosophila ananassae*, four potentially active amylase genes are situated at two loci found on different chromosome arms, with each locus consisting of two tightly linked, duplicated, amylase structural genes (Da Lage *et al.*, 1992). These diverse patterns of genomic organization appear to indicate that the amylase gene family is highly adaptive in species within which it is present (Hickey, 1977).

In studies of salivary gland-specific gene function in

*Author for correspondence. Tel: +1 714 824 5930; Fax: +1 714 824 2814; e-mail: aajames@uci.edu.
†Department of Molecular Biology and Biochemistry, University of California, BioSci II 3205, Irvine, CA 92697-3900, U.S.A.
‡Department of Biological Sciences, University of Notre Dame, 107 Galvin Life Sciences Building, Notre Dame, IN, 46556, U.S.A.
[1]Current address: Department of Entomology, University of Georgia, Athens, GA 30602, U.S.A.
[2]Current address: Temple University, School of Medicine, Wilmington, DE, 19803-0456, U.S.A.

the yellow fever mosquito, *Aedes aegypti*, a unique α-amylase, *Amylase I* (*Amy I*), with an extensive amino terminal region was found (Grossman and James, 1993). The primary sequence of *Amy I* is markedly different from other characterized insect amylases. In an effort to determine the relationship of *Amy I* with other members of the family, a genomic analysis of amylase encoding genes was performed. *Amy I* genomic DNA fragments were identified, as well as fragments corresponding to a second gene, *Amylase II* (*Amy II*). Restriction fragment length polymorphism (RFLP) mapping of these genes showed them to be unlinked. The coding regions and putative promoter sequences for both classes of mosquito amylase are presented and compared with other α-amylase genes.

MATERIALS AND METHODS

Mosquito maintenance

Ae. aegypti mosquitoes (Rockefeller strain) were reared using an 18:6 h light/dark circadian cycle, in a 27°C incubator at 80% relative humidity. All animals used in these experiments were between 3 and 10 days of age after adult emergence.

Genomic Southern analyses

Genomic DNA for Southern analyses (Southern, 1975) was prepared by the method of Bender *et al.* (1983). Whole mosquitoes were homogenized in lysis buffer, insoluble material was pelleted in a table-top micro-centrifuge and the DNA in the supernatant was collected by ethanol precipitation. Radioactive nucleic acid probes were generated using the Random priming kit from Boehringer–Mannheim (Indianapolis, IN) and incorporated [α-^{32}P] dCTP. Probes for *Amy I* hybridizing DNA fragments were derived from the *Amy I* cDNA clone, *λSGG34-1*, (Grossman and James, 1993). To avoid confusion in the following discussion, this cDNA will be referred to as the "*Amy I* cDNA".

Because of the salivary gland-specific expression of *Amy I*, it was presumed that *Ae. aegypti* would express an additional amylase gene similar to midgut amylase genes observed in other insects. An amylase cDNA from the mosquito, *Aedes atropalpus*, has been identified and oligonucleotide primers generated to its sequence were used to amplify a 551 base-pair (bp) fragment (size exclusive of the primers) from *Ae. aegypti* genomic DNA (gift of D. Hickey, University of Ottawa). The corresponding gene has been designated "*amylase II*". Probes for *Amy II* were generated by gene amplification of the 551 bp segment of DNA from *Ae. aegypti* genomic DNA and will be referred to as "*Amy II* 551 bp".

Aliquots containing 10 μg of total *Ae. aegypti* DNA derived from 70 or more adults were digested in separate reactions with different restriction endonucleases in the appropriate buffers and in the presence of 2 mM spermidine, final concentration. Digested DNA samples were electrophoresed in 0.8% agarose gels, visualized with ethidium bromide staining and photographed. Gels were depurinated for 15 min in 0.25 M HCl, denatured for 1 h in 1.5 M NaCl and 0.5 M NaOH and neutralized with two changes of 1.5 M NaCl and 0.5 M Tris (pH 7.5) for 20 min each. Nucleic acid was transferred overnight to Zetaprobe (Bio-Rad, Hercules, CA) nylon membrane in 10 × SSC (1.5 M NaCl, 0.15 M sodium citrate, pH 7.0), UV cross-linked at 120,000 mJ for 30 s and prehybridized for 4 h at 65°C. Hybridization of probes to filters was at 65°C in an aqueous hybridization solution, consisting of 5 × Denhardt's (0.1% w/v Ficoll, 0.1% w/v polyvinylpyrrolidine, 0.1% w/v bovine serum albumin), 5 × SSC, 100 mg/ml salmon sperm DNA and 0.5% sodium dodecyl sulfate (SDS). Southern analysis of phage DNAs was carried out as above, except for the following adjustments: after electrophoresis of digested phage DNA samples through 1% agarose gels, the DNA was depurinated by soaking the gels in 0.25 M HCl, denatured for 15 min and neutralized twice for 10 min with two changes of neutralization solution. Hybridization of phage DNA filters was carried out under stringent conditions at 65°C in a rotating hybridization oven. Filters were washed first for 30 min at room temperature in 2 × SSC with 0.1% SDS, then at 65°C in 0.1 × SSC, 0.1% SDS. Filters were monitored with a Geiger counter and exposed to Biomax or Fuji X-ray film at − 70°C for appropriate lengths of time.

Library screening

A λ Charon 4A (Blattner *et al.*, 1977) library constructed with *Ae. aegypti* (Rockefeller strain) genomic DNA (gift of H. Hagedorn, University of Arizona) was screened independently with random-primed *Amy I* cDNA and *Amy II* 551 bp probes. Screening was conducted under stringent conditions at 65°C in aqueous hybridization solutions including 6 × SSC, 5 × Denhardt's solution and 0.5% SDS. Washes at 65°C with 0.1 × SSC and 0.1% SDS ensured that only homologous sequences were obtained (Maniatis *et al.*, 1982).

RFLP analysis

Linkage and mapping data were derived from 85 F_2 intercross progeny from a pairwise mating between an *Ae. aegypti* Liverpool strain female and a RED strain male. The RED strain carries homozygous mutant morphological marker genes for each of the three *Ae. aegypti* chromosomes: *red-eye* (*re*) on 1; *spot abdomen* (*s*) on 2; and *black-tarsus* (*blt*) on 3. DNA was prepared from individual animals, digested with *Eco*R I, and evaluated by Southern analysis for linkage and segregation of the *Amy I* cDNA and the *Amy II* 551 bp markers. Map positions were assigned using procedures described in Severson *et al.* (1993, 1995). Chi-square goodness-of-fit values for the observed segregation ratios were calculated for each locus. Two-point recombination values and multipoint ordering of loci were determined using the MAPMAKER computer package (Lander *et al.*, 1987). A

LOD threshold score of 3.0 was used to detect linkage between loci, and a LOD threshold score of 2.0 to determine linear orders. *Amy I* and *Amy II* were integrated into a composite RFLP linkage map using the JoinMap computer program (Stam, 1993). Recombination frequencies were converted into map distances (cM) using the Kosambi function.

Sequencing of nucleic acids

All primary sequencing was done on DNA fragments subcloned into the plasmids pTZ19R or pBSSKII and using the chain-termination method of Sanger *et al.* (1977). Both the Sequenase kit (U.S. Biochemicals, Cleveland, OH), and dsDNA Cycle Sequencing kit (BRL, Gaithersburg, MO), utilizing a thermal cycler to facilitate sequencing reactions, were used. Commercially synthesized oligonucleotide primers and the generation of restriction enzyme-site deletion clones were both used to complete sequencing of both strands of DNA in all clones.

DNA and protein sequence analysis

Alignment of amylase DNA and protein sequences from other organisms was accomplished through the Altschul (Altschul and Erickson, 1986; Altschul *et al.*, 1990) and Doolittle (Feng and Doolittle, 1987) programs available through GCG (University of Wisconsin, U.S.A.). The conceptual translation products of the following DNA and cDNA sequences with accession numbers in GenBank (Bilofsky and Burks, 1988) were used for comparisons: *Anopheles gambiae* α-amylase, L04753 (Hickey and Des Groseillers, unpublished); α-amylase from *Anopheles merus*, U01210 (Foster *et al.*, unpublished); α-amylase from *Culex tarsalis*, U01211, (Hickey, unpublished); *D. melanogaster* proximal α-amylase, L22716 (Inomata *et al.*, 1995); and α-amylase from *Ae. atropalpus*, U012091 (Hickey, unpublished).

RESULTS

Genomic analysis by Southern blotting

As a first attempt to determine the copy number of α-amylase encoding genes in the mosquito genome, Southern analysis was performed with *Ae. aegypti* genomic DNA hybridized with the *Amy I* cDNA and the *Amy II* 551 bp probes (Fig. 1). Following digestion of genomic DNA with *Bam* HI, *Amy I* cDNA hybridization yielded four fragments, one at 5.4 kilobase-pairs (kb) in length, two larger than 10 kb, and a small fragment of less than 500 bases. Genomic DNA digested with *Eco*R I shows four major fragments at 3.8, 4.1, 4.5 and 5.4 kb in length and a faintly hybridizing fragment greater than 10 kb. The *Hin*d III digestion shows three major fragments at 3.5, 4.8 kb, and a large fragment greater than 10 kb in length. In addition, three lightly hybridizing fragments are evident at 4.3, ~8.0 and greater than 10 kb.

A Southern blot of DNA digested with *Bam* HI and probed with *Amy II* 551 bp shows a single major hybridizing fragment of ~6 kb in length and three fainter hybridizing bands greater than 10 kb (Fig. 1). Genomic DNA digested with *Eco*R I shows two major restriction enzyme fragments, one at 1 kb and one less than 500 bases in length, and lesser hybridizing fragments at 4.4, 4.7 and 5.4 kb. The *Hin*d III digestion showed two major fragments at ~8.0 and 10.0 kb in length and an additional lightly hybridizing fragment at 3.4 kb.

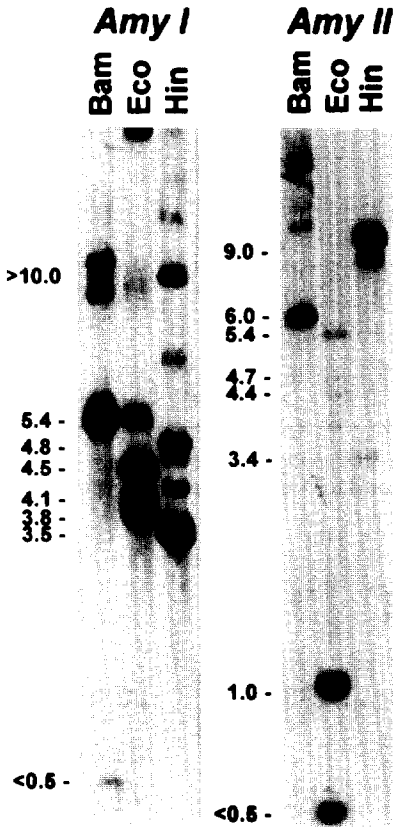

FIGURE 1. Genomic Southern analysis of *Amylase I* and *Amylase II*. Ten micrograms of *Ae. aegypti* DNA were digested with the enzymes shown, electrophoresed and Southern blotted (Southern, 1975). The filter labelled "*Amy I*" was probed with the *Amy I* cDNA, and "*Amy II*" was probed with the *Amy II* 551 bp gene amplification fragment. Select DNA fragments are labelled with a size marker on their left. Fragment sizes are based on extrapolations from the migration lengths of the 1.0 kb marker series (lanes not shown). Abbreviations: Bam, *Bam*H I, Eco, *Eco*R I, Hin, *Hin*d III.

Meiotic mapping with RFLP markers of Amy I *and* Amy II

The different hybridization patterns seen when using *Amy I* and *Amy II* probes suggested that there were at least two amylase-encoding genes in *Ae. aegypti*. We used genetic analysis to distinguish the two genes further. Segregation ratios for *Amy I* and *Amy II* were determined (Table 1) and map positions assigned relative to all RFLP marker data (Fig. 2). The composite RFLP map data can be accessed directly from the *Ae. aegypti* genome database (Mosquito Genomics, 1997). These data place *Amy I* on chromosome 2 at 23.0 cM, and *Amy II* at 44.0 cM on chromosome 1. All loci on chromosome 1 fit their

TABLE 1. Segregation of markers and recombination fractions between adjacent markers

	Marker	Segregation ratio	χ^2	Recombination fraction (0%)
1	LF198	26:40:19	1.45	
	LF178	19:49:17	2.08	19.9
	re	65:20	0.10	11.1
	Amy II	19:46:20	0.50	2.4
	AaCHT2	20:41:24	0.48	3.0
2	LF98	18:55:12	8.20*	
	AmyI	18:56:11	9.73**	1.8
	TyAD1	18:54:12	7.72*	1.8
	s	71:14	3.90*	16.6

Marker loci are arranged in chromosomal order.
All loci were scored for 1:2:1 segregation, except *re* and *s*, which were scored as 3:1.
*$P < 0.05$, **$P < 0.01$.

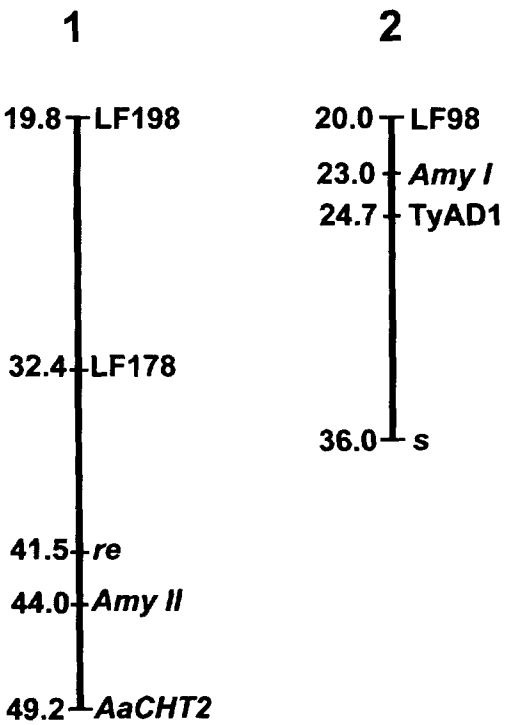

FIGURE 2. Linkage and map assignments of *Amy I* and *Amy II* genes on *Ae. aegypti* chromosomes 1 and 2. RFLP mapping (Severson *et al.*, 1993) produced the linkage map. All markers are described in the *Ae. aegypti* genome database (Mosquito Genomics, 1997).

expected segregation ratios. Loci on chromosome 2 exhibited significant deviations from the expected segregation ratios and this was a result of all loci showing reduced recovery in the RED strain genotype. Segregation distortions have been observed in other crosses involving the RED strain, however the observed bias did not influence determinations of linkage associations and linear orders (Severson *et al.*, 1993, 1995). The markers on chromosome 2 made it difficult to determine the linear order with certainty (e.g. the multipoint LOD threshold was <2.0). However, we are confident of the order presented in Table 1 and Fig. 2 because data from another cross-confirms the linear placements on chromosome 2 for LF98 and TyAD1 (Severson, unpublished data).

Isolation and characterization of Amy I *and* Amy II *genomic clones*

Having demonstrated separate loci for *Amy I* and *Amy II*, we wanted to know if we could further define the structure of each gene. The detection of multiple hybridizing fragments of *Amy I* in the genomic Southern analysis was interpreted to indicate the presence of multiple genes or alleles. The pattern for *Amy II* hybridization was much less complex and suggested a single gene. To determine whether these restriction enzyme fragments represent multiple, linked amylase genes, similar to the case for the *D. melanogaster* amylase genes, or polymorphic alleles, a genomic library was screened to isolate and identify amylase genes.

In an initial screen, nine genomic clones hybridizing to the *Amy I* cDNA were isolated. These clones were shown to have identical restriction enzyme patterns when digested with *Eco*R I. One of these clones was designated G34-F (F), and a 3.2 kb restriction fragment, F2, was determined to contain the DNA sequences hybridizing to the *Amy I* cDNA probe (Fig. 3(A) *Amy Clones* and *Amy I cDNA*). (Genomic fragments are labelled by the name of the clone and a number that is assigned progressively to individual fragments (large to small) that result from restriction enzyme digests. F2 is thus the second largest *Eco*R I fragment of clone F). Additional screening of the genomic library with *Amy I* cDNA identified a second class of genomic DNA fragments based on the restriction enzyme digestion patterns. A genomic clone representing this class was designated G34-14A (14A), and a 3.8 kb fragment, 14A2, contains the DNA sequences hybridizing to the *Amy I* cDNA probe (Fig. 3(A) *Amy Clones* and *Amy I cDNA*).

The two classes of clones represented by F and 14A exhibit distinct restriction enzyme patterns, but share significant similarity based on their equally strong hybridization signals with the *Amy I* cDNA probe. In addition, 14A genomic DNA was used to probe digested F DNA, and all but a 2.2 kb fragment, F4, hybridized with the 14A probe (Fig. 3(A) *Phage 14A*). Conversely, the isolated F phage *Eco*R I fragments, with the exception of F4, all hybridized with fragments of the 14A phage (Fig. 3(C)). F4 is located directly adjacent to the left arm of λ Charon 4A, and the 14A genomic clone probably does not include as large a portion of the 5′-end non-coding region of the *Amy* I gene as does the F clone, and therefore does not hybridize with the F4 fragment. The 3.8 kb fragment, 14A2, probably corresponds to the band of similar size observed in the genomic analysis (Fig. 1), however, we did not observe a 3.2 kb fragment that would correspond to F2.

Primary sequence data from within the coding region and at the 5′-end non-coding regions, show nearly identical sequences for F and 14A (Fig. 4). We interpret these data to indicate that the two clones represent genes

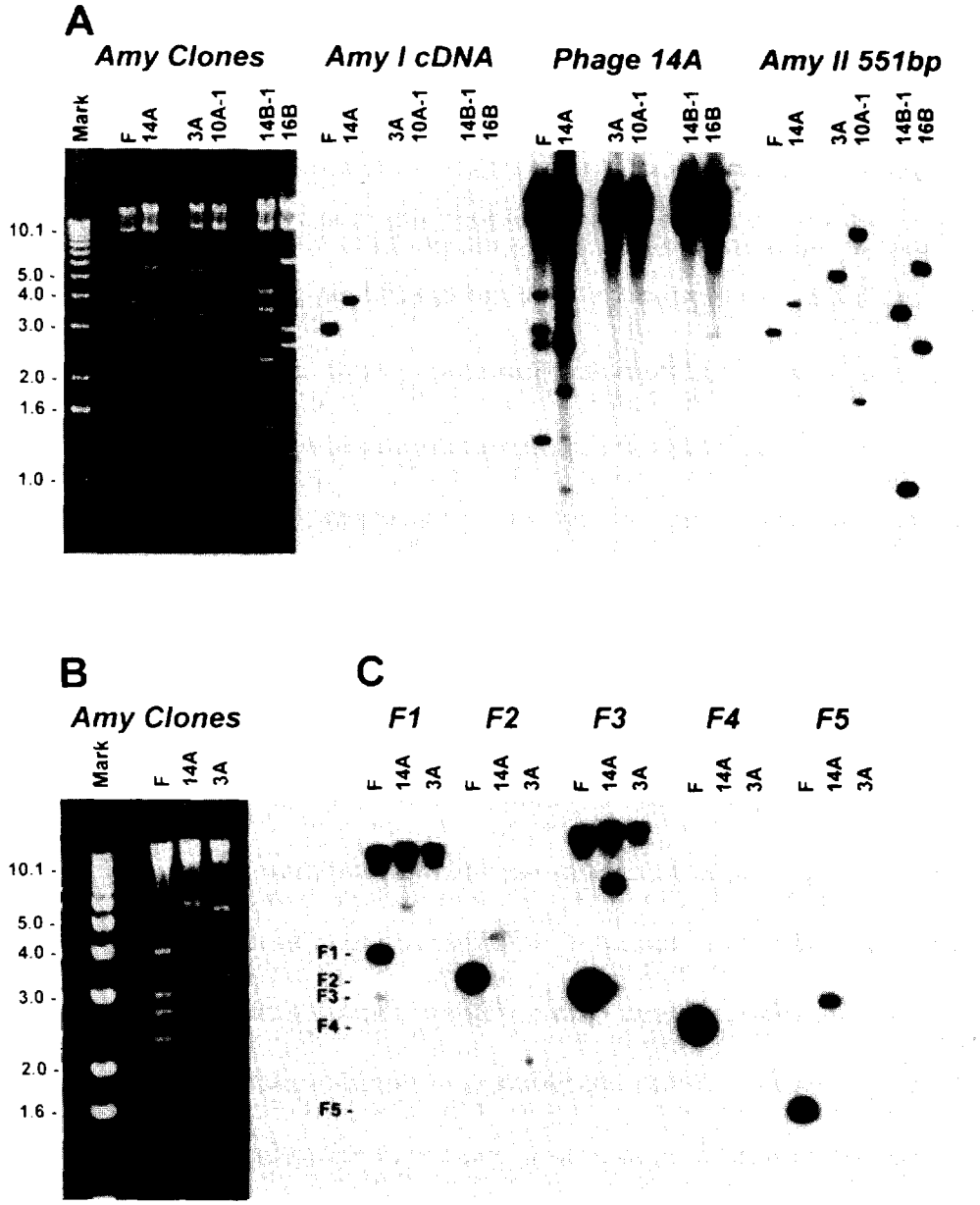

FIGURE 3. Analysis of the hybridization patterns of six genomic α-amylase clones, F, 14A, 3A, 10A-1, 14B-1 and 16B, isolated from an *Ae. aegypti* genomic DNA library. Isolated clones were digested with *Eco*R I, electrophoresed and Southern blotted (Southern, 1975). Horizontal labels describe panels (*Amy Clones*) or indicate which DNA fragments were used as probes (all others). Vertical labels indicate the phage clone that was digested or the marker lanes (Mark). (A) Ethidium bromide-stained patterns of DNA fragments from the amylase genes (*Amy Clones*). Note that the marker lane ran anomalously fast in this gel and in lane F, the fragment of apparent size 2.9 kb corresponds to F2. *Amy I cDNA*, genomic DNA clones probed with *Amy I* cDNA. *Phage 14A*, genomic DNA clones probed with total 14A genomic DNA to determine how many fragments cross-hybridized, indicating their homology at the nucleotide level. All fragments in the F digest, except for F-4, hybridized with 14A DNA. Contaminating phage DNA in the probe hybridizes with the phage arms in all lanes. *Amy II 551bp*, genomic DNA clones probed with the *Amy II* 551 bp fragment. (B) Gel electrophoresis of selected *Amy I* and *Amy II* clones. The clones F, 14A, and 3A were digested with *Eco*R I, resolved in an agarose gel and stained with ethidium bromide. The markers and clone fragments run at their respective molecular weights. (C) Hybridization analysis of clones F, 14A, and 3A using F phage fragments as probes. High molecular weight signals in F1 and F3 represent contamination of the probe with phage DNA. Low molecular weight signals in the "F" lane of the *F1* probe indicate some contamination of the probe.

located at the same locus in the genome, and are therefore alleles. However, the detection of other hybridizing fragments in the genomic Southern analysis may indicate the presence of multiple copies of *Amy I* in the genome.

The *Amy II* 551 bp probe was used in library screening to identify 29 mosquito genomic DNA fragments. Four of these clones, 3A, 10A-1, 14B-1 and 16B, have unique enzyme digestion patterns among themselves, and when compared to the F and 14A isolates (Fig. 3(A) *Amy Clones*). None of these genomic clones contain DNA fragments that hybridize with the *Amy I* cDNA probe under stringent conditions (Fig. 3(A) *Amy I cDNA*). How-

```
-832  TTAGCAAACATAATCGATAAAATTTCTGAGAAACCAAATCGGTACTACTGATATTTGACCATGCAACGGAACTCTAGTCCCATCCTTGTT  -743
-742  TCGCTATTAGTTTAGTCCCAGCAAATACTAGAAAAAAGGGGCTTTATGCTAGCAGCAAAACCGAATCAAGCTAGAAAATTTGTTTAGTAA  -653
-652  TCAGAATTCACGTGAGTCCTTGAATTACCAGTACTTACAGTCCTTGTTGAGTTAATACTCATGATTGTTATCCTGGCTTCAAGTGTAGTC  -563
-562  TCGGGGCTGACCAACTCCGAGTCTTTAACCATCTGCTAGGTAAAATAGATTAATTTTCAGAAACTGTTGCCAGTAACAAGGACTTTGTAC  -473
-472  GCGAATCTGAATTACCAGAACATATCGCTAGCCCGAAACAGATGAGACTAGGTCAATGTAGATCTACCGTACGCACACGAAGTGACTACC  -383
-382  CGCGTTACGGTGAACTGAGTAGATACTTAACGGGAATGGATATATACTTACTCATCTACCTTCAGTATGGTCTTTTCACAGTTGCCGCGA  -293
-292  TCGTTCACTAAGCTAAAAAGACCCGTAACACAATGTATCTGAGGTTAATTTCAAATTAATTTATCTTCACACATACGCTTATGATGTAAG  -203
      →
-202  AATATTCCTACGATATCAACGACCATAAAATAGTCACGTTTTGATTAATTTAAAATTCAGTCTGTTTAATGACACAGTTGATCAGTTAAT  -113
-112  AAGTGCAAAGAATACGATAGCAAACATCGATAGAATATGCAGATCGATCCACATGCTCCAGCTTAGAAAACAACATTCGATCCCTTATTA  -23
            T          +1        A              ←
 -22  TAAAAGTACGGTTTGGTAAAACAATTTCGTTAGTTTCTTATTCCATACTCGGCCAGCAAAATGAAGCAGTTATTGGTGATTTTGTACAGT   68
                                                                  M  K  Q  L  L  V  I  L  Y  S
        →                                                                    A
  69  GCTTTAGTGGTGACCAATGCCCAATTGCCGGGCTTCAGTTTTGGTCCTAGTGGATTTGGGGGCTTTTCCCTTTCGAATCCTTTGTTCAAC  158
      A  L  V  V  T  N  A  Q  L  P  G  F  S  F  G  P  S  G  F  G  G  F  S  L  S  N  P  L  F  N
                                                          ←
 159  GCTTCATTCGCGCCTGCAAATCCGGTAGATCGTATTTCGGACATTGCTTCAAGGTTTAGGGACGTAGCGAATAATGTGCAACGCAATCTT  248
      A  S  F  A  P  A  N  P  V  D  R  I  S  D  I  A  S  R  F  R  D  V  A  N  N  V  Q  R  N  L

 249  CCAAAAATACCGCAATTTACTAATCCTCTGTCAGCATTCTCTCGATTTGCAAGACCGACAAGTTCAAGTGGTTCGACAAATGCGAACTCC  338
      P  K  I  P  Q  F  T  N  P  L  S  A  F  S  R  F  A  R  P  T  S  S  S  G  S  T  N  A  N  S

 339  AACAACAATCAGAATTTCAATAGCAATCAAGCGAATGAGGGCACCAGTGGCGCCAGAAATGCCATCAACGAGGTGCTGAATCGCATACCA  428
      N  N  N  Q  N  F  N  S  N  Q  A  N  E  G  T  S  G  A  R  N  A  I  N  E  V  L  N  R  I  P

 429  GGACTATCAGGGGTACCACGTGTACCAGAAATACCTAAGGTGCCAGAAATACCGAGTTTCGGAAACCCGAGTGATGTTTTCGGTCCTTTT  518
      G  L  S  G  V  P  R  V  P  E  I  P  K  V  P  E  I  P  S  F  G  N  P  S  D  V  F  G  P  F

 519  GTGAACTCTTTGAGAAATATGACTAATCAAGGCATCAAAGACCCCGGGCAGATATTTAATCAGGTGAACACATTTTTACCCGATGTATCT  608
      V  N  S  L  R  N  M  T  N  Q  G  I  K  D  P  G  Q  I  F  N  Q  V  N  T  F  L  P  D  V  S

 609  CGCATTGCTAATAGCGTTGCTAACTATGCTAAGTCCGCTGCCAATTCTGGAATATGGAAGAATATATCCAACGACATCAAACATGAAATG  698
      R  I  A  N  S  V  A  N  Y  A  K  S  A  A  N  S  G  I  W  K  N  I  S  N  D  I  K  H  E  M

 699  CAGAATAGAATTGAGTCGATGACCCAAATACTATCGAACATAACTTCTGCCATGAAAAATATTGCCCAAAGTGTACCCAATTTAATGCAG  788
      Q  N  R  I  E  S  M  T  Q  I  L  S  N  I  T  S  A  M  K  N  I  A  Q  S  V  P  N  L  M  Q

 789  GATCAAAGCCTCAACACTACACTTCGCAATCAACCATTCTTCTTCCCTGGACATTCGGGAATTGTTCACCTGTTCGAGTGGAAATTCAGC  878
      D  Q  S  L  N  T  T  L  R  N  Q  P  F  F  F  P  G  H  S  G  I  V  H  L  F  E  W  K  F  S

 879  GATATTGCTGAAGAATGTGAGAAAGTGCTGGGACCTAACGGGTATGGCGGTGTGCAGGTTTCTCCGATCAACGAGTACTTGGTATCGCCA  968
      D  I  A  E  E  C  E  K  V  L  G  P  N  G  Y  G  G  V  Q  V  S  P  I  N  E  Y  L  V  S  P

 969  TCACGAGCTTGGTGGGAAAGATATCAACCTATATCTTTCGAAATAAAGTCACGATCGGGTAACGAGAAGCAATTCTCGGATATGGTGAAG 1058
      S  R  A  W  W  E  R  Y  Q  P  I  S  F  E  I  K  S  R  S  G  N  E  K  Q  F  S  D  M  V  K

1059  CGTTGCATGAAAGCCGGCGTTCGAATCTACGTGGATGTAGTAGTCAACCACATGGCTGCTCCTGGGGCTTCAGCGCCTTTGTATGGGACA 1148
      R  C  M  K  A  G  V  R  I  Y  V  D  V  V  V  N  H  M  A  A  P  G  A  S  A  P  L  Y  G  T

1149  GCTGGATCAACTTGTGATCCGCTGGCTAGGGATTATCCAGGTGTACCGTTCAACAGGTCCCACTTCCACGCTGATTGCCAGATCAACGAT 1238
      A  G  S  T  C  D  P  L  A  R  D  Y  P  G  V  P  F  N  R  S  H  F  H  A  D  C  Q  I  N  D

1239  TACAATAATGCGACTAATGTTCGGAATTGCGAGCTAGCTGCACTTCCGGATTTGGACCAATCCAATAGATTTGTCCAAAACAAGATCATC 1328
      Y  N  N  A  T  N  V  R  N  C  E  L  A  A  L  P  D  L  D  Q  S  N  R  F  V  Q  N  K  I  I

1329  CAATATTTGAATCATCTCCTCGATCTTGGAGTGGCAGGATTCAGGATGGACGCTTGCAAACATATGCGGCCAGAAGATCTTAAATCAATT 1418
      Q  Y  L  N  H  L  L  D  L  G  V  A  G  F  R  M  D  A  C  K  H  M  R  P  E  D  L  K  S  I

1419  TACGATCGACTAAAGCCAGTCAATGCAATGTTTCTTTTTCCACCTGGTGCAAGGCCTTTTATCTTTCAGGAAGTGATCGATTTGGGTACT 1508
      Y  D  R  L  K  P  V  N  A  M  F  L  F  P  P  G  A  R  P  F  I  F  Q  E  V  I  D  L  G  T

1509  GAAGGCGTTTCAGCGTAAGTAATTATGTAATAATGCAAAAACGTAGTTGTGTAATGTTATCTACACTTTTCAGTAAAGAATATACCAATC 1598
      E  G  V  S  A                                                                K  E  Y  T  N

1599  TTGGAGTTGTAACGGAGTTCAATTGGTGTATTGTAGTGGGAGGTGTCTTTCGAGGAACAACCAATGCGGATGCCTTAGAATTGCTTACGA 1688
      L  G  V  V  T  E  F  N  W  C  I  V  V  G  G  V  F  R  G  T  T  N  A  D  A  L  E  L  L  T

1689  AGAATGGTTCAGCTGGGGTTCTTCTTCCGTCAAGTCAGGCTCTGGTGTTCGTAGATAATCACGATAACCAACGAGGTCATGGAGCCGGAG 1778
      K  N  G  S  A  G  V  L  L  P  S  S  Q  A  L  V  F  V  D  N  H  D  N  Q  R  G  H  G  A  G
```

(a)

```
1779  GTGATTCTATCTTGACCTATAAGACAAAACCGCAGTACATCCAGGCAGTGGCGTTCACTCTTGCCACCGATTACGGAATTGCGCGGGTTA  1868
      G  D  S  I  L  T  Y  K  T  K  P  Q  Y  I  Q  A  V  A  F  T  L  A  T  D  Y  G  I  A  R  V

1869  TGAGTTCATACAATTTCTCGGATCCAGATCAAGGCCCCCCGCAAGATACCGTTCAAGTTATCAGATCACCCGGGTTCGCAGTGAATAACT  1958
      M  S  S  Y  N  F  S  D  P  D  Q  G  P  P  Q  D  T  V  Q  V  I  R  S  P  G  F  A  V  N  N

1959  CATGTAGCACCGGCTGGGTATGCGAACACCGTTGGCCAGAAATTAGGAAGATGATCCAATTCAAGAACTTCGTAGCTGGTTCATCGCTGG  2048
      S  C  S  T  G  W  V  C  E  H  R  W  P  E  I  R  K  M  I  Q  F  K  N  F  V  A  G  S  S  L

2049  ACCATATACAAGCGACACAAAACACGTTCGCTTTCTGTCGCGGAGAGAAGGGTTTCATCGTTTTCAACAATTCCGAAAACACGATCACGC  2138
      D  H  I  Q  A  T  Q  N  T  F  A  F  C  R  G  E  K  G  F  I  V  F  N  N  S  E  N  T  I  T

2139  AACAGTACCATACATGCCTTCCACAGGGTCAGTATTGTGACATTATTTCCGGAGAAGTGTCACAGAGCACATGTACAGGAACCGTTGTCA  2228
      Q  Q  Y  H  T  C  L  P  Q  G  Q  Y  C  D  I  I  S  G  E  V  S  Q  S  T  C  T  G  T  V  V

2229  ACGTTGATGCAAATGGTGATGCAGAAATAACGCTGCCCAAGAATTCAGTTGTAGCCATTTATATACCTTCAAGGCTTTCATAGAATGAAT  2318
      N  V  D  A  N  G  D  A  E  I  T  L  P  K  N  S  V  V  A  I  Y  I  P  S  R  L  S  *

2319  GAAACACTTGCTAAATTCAGCAGTAGGGTTTATTATTACAAATAAAATCATGTGGTTGGGTCGAATACTAAGCAAACTATTTTGGAATAG  2408

2409  CAGTTT  2414
```

(b)

FIGURE 4. Nucleotide sequence and conceptual translation product of the *Amylase I* genomic clone G34-F2. Nucleotide sequence of the 3.2 kb genomic fragment, F-2, corresponding to the salivary gland-specific amylase, *Amy I*, was determined by chain-termination sequencing of this subclone as well as sequence generated from double-strand sequencing of the total F clone in phage λ Charon 4A. This sequence has GenBank designation no. AF000569. Numbers adjacent to the sequence on both sides refer to the nucleotide position. The RNA transcription site is labelled + 1, and there is a consensus TATAAA sequence at − 24 nt upstream of the start of transcription. A consensus polyadenylation signal sequenced at 2360 nt and a stop codon at 2308 nt are also underlined. The 59 bp intron is underlined and is found at 1523 and 1581 nt. Underlined sequences in the 5′-end non-coding regions show at least 70% identity with *cis*-acting regulatory elements and are discussed in detail in the text. The arrows above the segments of DNA, − 201 to 30 and 76 to 207 nt, indicate homologous regions sequenced in phage 14A. Letters above the lines at − 15, 11, and 137 nt indicate sequence polymorphisms between F and 14A. The single letter code for amino acids represents the conceptual translation product and the asterisk denotes the end of translation.

ever, the *Amy* II 551 bp probe does cross-hybridize slightly with both F2 and 14A2 (Fig. 3(A) *Amy I cDNA* and *Amy II 551 bp*). Clone 3A was chosen for further analysis of the *Amy II* gene structure, and a 5.4 kb fragment, 3A1, was shown to contain the *Amy II* hybridizing sequences (Fig. 3(A) *Amy II 551 bp*).

The fact that numerous clones containing unique restriction enzyme-digestion patterns were isolated from the screen with *Amy II* can be interpreted to indicate that either there are multiple alleles of *Amy II* or that it exists as a linked multigene family. Interestingly, the 5.4 kb genomic fragment, 3A1, probably represents the lightly hybridizing signal in the genomic Southern analysis (Fig. 1). The 5.4 kb species may represent a rare polymorphic fragment in the *Amy II* gene that we recovered from the library. The lack of multiple hybridizing DNA fragments in the genomic Southern analysis, that would correspond to *Eco*R I fragments seen in the other clones identified in the screen (Fig. 3(A) *Amy II 551 bp*), can be taken to indicate that the *Eco*R I polymorphisms in the clones are not commonly represented in the pool of mosquitoes used to prepare the genomic DNA for the blots. From these data we are unable to distinguish whether there are multiple allelic variants of *Amy II* in this mosquito population, or if there are multiple, linked copies of *Amy II*. However, the former alternative is supported by the low number of *Eco*R I fragments seen in the genomic Southern analysis.

Features of the nucleotide sequence of the Amy I *genomic fragment, F2*

The genomic F2 fragment is 3246 bp in length (Fig. 4). The F2 sequence consists of 832 bp of 5′-end upstream sequence adjacent to the *Amy I* transcription initiation site. The transcription initiation site was identified previously using primer extension analysis (Grossman and James, 1993). There is a 59 bp intron at nucleotides (nt) 1523 to 1581, inclusive. A consensus TATA and polyadenylation signal sequence are evident at − 24 and 2360 nt, respectively. The cDNA clone for *Amy I* was previously shown to possess 780 bp of coding sequence at its 5′-end that is unique to the mosquito salivary gland amylase (Grossman and James, 1993). By sequencing the genomic clone, we have confirmed that this DNA is located directly adjacent to that region of *Amy I* that is more readily identified through sequence similarities to previously described amylase gene sequences. A number of nucleotide sequence motifs were found in the 5′-end non-coding region of the *Amy I* gene. Four regions, −645 to −637, −579 to −572, −253 to −246 and −211 to −204 nt, had six out of eight nucleotide identity to a cyclic AMP responsive element, 5′–TGACGTCA–3′, found adjacent to various mammalian genes (Silver *et al.*, 1987; Roestler *et al.*, 1988). Four regions, −644 to −637, −469 to −462, −257 to −251 and −210 to −203 nt, also showed six out of eight nucleotide identity to a cyclic AMP

receptor protein binding site, 5′–AATGTGAG–3′, found in the control regions of some bacterial catabolite-sensitive genes (de Crombrugghe *et al.*, 1984). Interestingly, three of the four cAMP-responsive elements overlap the receptor protein binding sites in the *Amy I* gene.

A sterol regulatory element, 5′–ACCCACTGC–3′, is implicated in the feed-back expression control of mammalian genes (Sudhof *et al.*, 1987), and a seven out of nine identity match was found at −535 to −527 nt, inclusive. AP-2 is a eukaryotic transcription factor and its binding site, 5′–TCCCCANGCG–3′, is found at the 5′ end of genes that respond to signal transduction (Imagawa *et al.*, 1987). Three sites in the *Amy I* gene −403 to −394, −288 to −279 and −65 to −56 nt, show at least seven out of ten nucleotide identity with the AP-2 binding site.

The *D. melanogaster* amylase genes were shown to be subject to glucose repression at the level of transcription, and a nucleotide sequence, 5′–CCAGTCAATAGGTCTG–3′, was determined to be responsible for this activity (Boer and Hickey, 1986; Benkel and Hickey, 1987). Two sequences in *Amy I*, −493 to −478 and −121 to −106 nt, were shown to have 69 and 63% identity, respectively.

Finally, the sequence, 5′–TCAGT–3′ was shown to be present near the transcription initiation site of many arthropod promoters (Cherbas and Cherbas, 1993). Absolute sequence identities with this motif are found at −321 to −317, −146 to −142 and −121 to −117 nt, but their distance from the transcription initiation site is farther than average.

Features of the partial nucleotide sequence of the Amy II genomic fragment, 3A1

The genomic fragment, 3A1, was determined to be 5.4 kb in length (Fig. 3(A) *Amy II 551 bp*). A 1770 bp region adjacent to one of the *Eco*R I sites was sequenced and shown to contain the *Amy II* gene (Fig. 5). We do not know which nucleotide represents the transcription initiation site, and thus the sequence has been numbered arbitrarily, starting at the beginning. The fragment has 305 bp of 5′-end non-translated sequence, and within this region there is a consensus TATA box at 237 nt. The coding region extends without interruption for 1458 bases.

The 5′-end non-coding region contains some of the sequence motifs observed in the *Amy I* gene. A single cAMP responsive element and cAMP receptor protein binding site overlap in the region 260 to 267 and 256 to 263 nt, respectively, and both have six out of eight nucleotide identity with their consensus sequences. A single AP-2 site, 41 to 51 nt, has a seven out of nine identity with the consensus motif. A single nucleotide insertion, T at 48 nt, interrupts the sequence.

Two potential glucose repression motifs are evident at 41 to 59 and 253 to 268 nt. Both have 56% similarity to the consensus, but the former one is interrupted by the T insertion. A short sequence, 5′–TCATT–3′, at 270 to 274 nt, is reminiscent of the arthropod initiation consensus sequence, 5′–TCAGT–3′ (Cherbas and Cherbas, 1993).

Developmental Northern blot analyses using RNA, isolated from embryos, all stages of larvae, pupae and adults, and both the *Amy II* 551 bp fragment and the subclone 3A1 as probes, were inconclusive. Although control hybridization probes recognized appropriate RNA species, we could not detect a signal with the *Amy II* probes (data not shown). Therefore we cannot confirm whether this amylase gene is expressed in the mosquito, and if so, at which stages of the life cycle. We expected that this gene would be expressed at all stages in both sexes of the mosquito, because we observe strong amylase activity at all points in the life cycle (Grossman, 1994).

Analysis of the conceptual translation products of Amy I *and* Amy II

The conceptual translation products of *Amy I* and *Amy II* are 737 amino acids (aa) and 486 aa in length, respectively (Figs 4 and 5). *Amy I* is expressed specifically in the adult salivary glands of adult *Ae. aegypti* and was shown to have a 260 aa amino terminal end that is unique to this gene product (Grossman and James, 1993). The predicted size of *Amy II* is more like the majority of described α-amylases. Both *Amy I* and *Amy II* have hydrophobic leader sequences that are characteristic of secretory signal peptides (Von Heijne, 1985), and are predicted to be approximately 16 and 20 aa in length, respectively (Grossman and James, 1993, and Fig. 5).

The primary amino acid sequences of the *Amy I* and *Amy II* products were aligned with the products of five other genes: two from the mosquito subfamily Anophelinae, *An. gambiae* α-amylase and *An. merus* α-amylase, two from the mosquito subfamily Culicinae, *Ae. atropalpus* α-amylase and *C. tarsalis* α-amylase (*Ae. aegypti* and the *Amy I* and *Amy II* gene products belong in this group), and the sequence of the fruit fly, *D. melanogaster*, proximal α-amylase (Fig. 6). The alignment program orders the amino acid sequences so that the most similar appear on the top line, and the others appear in order of decreasing identity. Conservation of five sequence motifs (2–6) common to all α-amylases (Svennson, 1988) indicated clearly that *Amy I* and *Amy II* are members of this large family of genes. Furthermore, conserved cysteine residues, predicted to exist as cystines in the mature proteins, are also evident.

Because the α-amylases used for comparison vary in size, absolute percentile differences are subject to adjustments. However, we have counted amino acid differences and expressed them as a percentage of the average size (494 aa) of the amylases used for comparison. The *Amy I* product was left out of the average, but differences in the truncated putative protein, 261–737 aa, were calculated. These numbers, albeit rough approximations, provide a useful method for comparing similarity among the amino acid sequences.

The two Anopheline proteins are most similar with

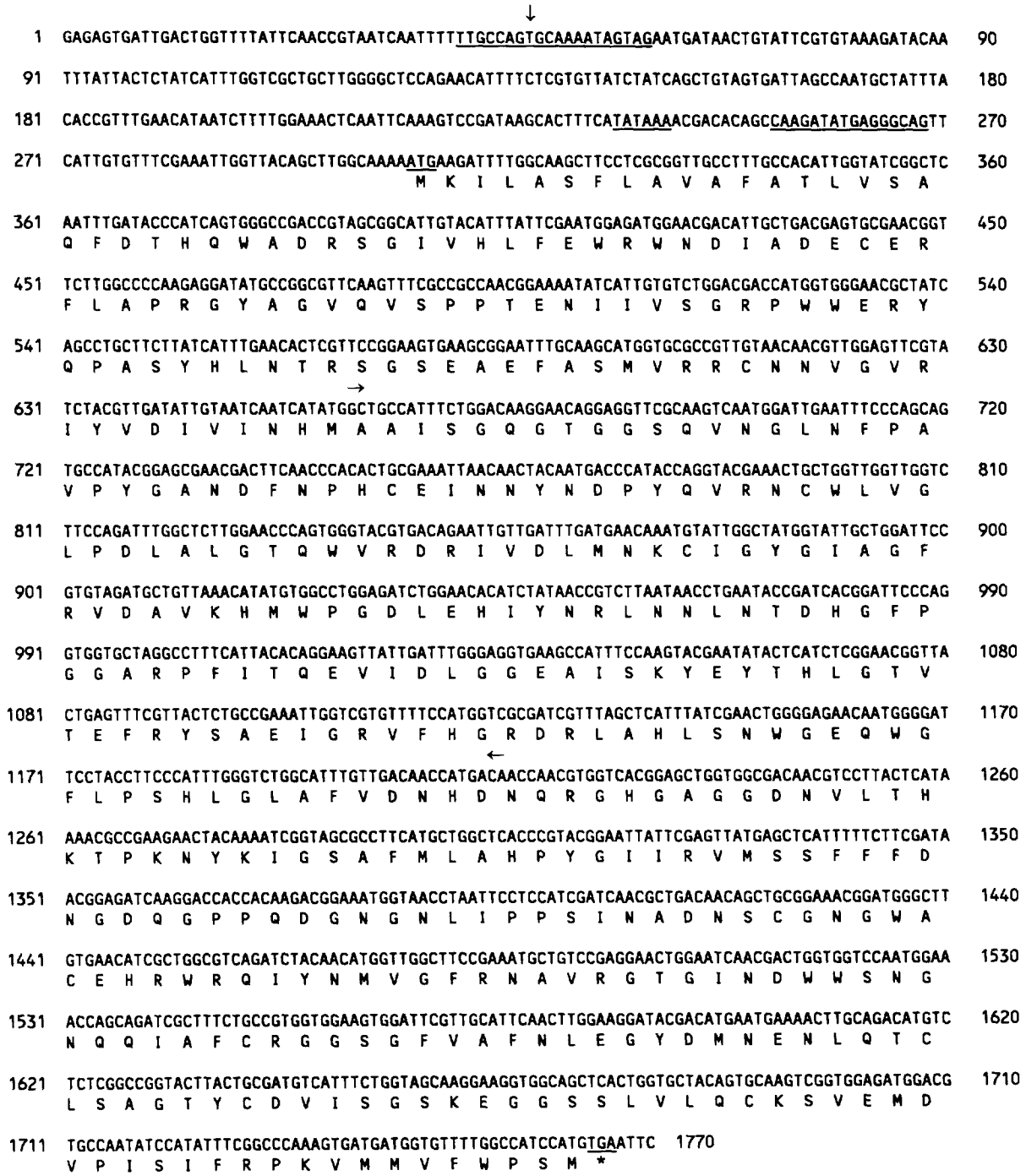

FIGURE 5. Nucleotide sequence and conceptual translation product of the *Amylase II* gene. Partial nucleotide sequence of the 3A1 5.4 kb *Eco*R I fragment was determined as described in Fig. 4. This sequence has GenBank designation no. AF000568. Arbitrary numbering of the nucleotide sequence is given on both sides of the figure. A consensus TATAAA sequence appears at 237 nt, and initiation and termination codons are evident at 306 and 1764 nt, respectively. Underlined sequences in the 5′-end non-coding region and the vertical arrow above the T residue at 48 nt are discussed in the text. The arrows above 662 to 1213 nt indicate the region that constitutes the *Amy II* 551 bp probe. The single letter code for the amino acids represents the conceptual translation product and the asterisk denotes the end of translation.

97% identity. The three Culicine α-amylases, *Amy II*, *Ae. atropalpus*, and *C. tarsalis*, show approximately 60, 64 and 60% identity to the Anopheline amylases. The *D. melanogaster* sequence is 63% identical to the Anopheline sequences. Of note is the fact that the *Amy I* product is only 52% identical.

Within the Culicinae, *Ae. atropalpus* and *C. tarsalis* α-amylases are 86 and 73% identical, respectively, to the *Amy II* product. However, *Amy I* is only 46% similar to *Amy II*. These data indicate that *Amy II* belongs to a grouping of α-amylases that despite being from different species share more similarity with one another than they do with *Amy I*.

```
                                                    ↓
Angamb    25 DPHFVRGHSTIVHLFEWKWSDIADECERFLGPKGYGGVQLSPVNENIVIRLADGSRPWWE
Anmeru    25
Aedae2    21  T QWADR G       R N          A R  A   V  PT   IVS-G---
Aedatp    21  T QWADR G       R N          A    A   V  PT    VS-G---
Cutars    20 QT QWADRTG       R N          A    A   V  PT  AIVW-SPVR
Dropro    21  TNYAS R GM        D   A   N    N  A   V      A ----KD
Aedae1   261 Q F FP   G        F    E   KV   N      V  I  YL ----SP  A

                                2          ↓          3
Angamb    85 RYQPISFKLDTRSGSEAEFADMSRRCNAAGVRLYVDIIINHMGATQPVEPAIGTGSSTAI
Anmeru    85
Aedae2    77     A YH N          S V    NV   I    V    A ---ISGQ-   G --Q
Aedatp    77     A YH N      S   S V    QV   I    V    A ---MSGQ-   G --S
Cutars    79     M YE T    N Q   S V    DV   I    V    A ---ITGDG   G  GS
Dropro    77       Y  E    N EQ  S VK    V   T   VVF   A ---DGGTY   G   S
Aedae1   317        EIKS   N KQ S  VK  MK    I   VVV   A PGASA LY  AG  CD

                                   ↓                ↓
Angamb   145 PSDRQFPAVPFGWPDFNPPCAINDWGNAEQIRNCELVGLHDLNQAVPWVRDRVVDFLNHL
Anmeru   145
Aedae2   131 VNGLN      Y AN    H E  NYNDPY V   W    P  ALGTQ      I  LM KC
Aedatp   131 VDGLN        PN    R D TNYNDKY V   W    P  ALGNQ      I  LM KC
Cutars   136 SQTLS  G   SAL     R PL NYN  IEV   W    P    G Q     KI ELF K
Dropro   134   SKSY G  YSSL     T   SNYND NEV        R    GNSY Q K  E  D
Aedae1   377  LA DY G   NRSH HAD Q   YN  TNV     AA P  D SNRF QNKIIQY

                  4                                          5
Angamb   205 IELGVAGFRVDAAKHMWPADLAVIFGRLNNLNTAYGFAPGSRAFLAQEVIDMGAHEAVRK
Anmeru   205                                                          V
Aedae2   191  GY I        V     G  EH YN        DH  PG A P IT     L G-  IS
Aedatp   191 VGY          V     G  EH YS        DH  PH AKP IT     L G-  IT
Cutars   196  GM          V     G  QA YS  DP P SH  P NAQ   T     L G-   TR
Dropro   194  D                      Y   K    DH   S  K YIV       G-  IS
Aedae1   437 LD       M  C   R E  KS YD  KPV AMFL P  A P IF     L -T G SA

                                                              6
Angamb   265 FEYTFLGTVTEFMFSHYLGRAFGGN---DALRWLSNFGEAWRLLASREAFVFVDNHDNQR
Anmeru   265                          ---              G     D
Aedae2   250 Y   H       RY AEI  V H R---  R AH   W  Q GF P HLGLA
Aedatp   250 Y   H       R  AEI  V R Y---  R AH   W  Q GF P HLGL
Cutars   255 D   H       R  AEI  V R H---  R AN R W    GF P HL L
Dropro   253 S   G  AI   RH DSI KV R K---  Q QY T W T  GFA  DRSL
Aedae1   496 K   N  V    NWCIVV GV R TTNA   EL TKN S GV  P SQ L

Angamb   322 GD----YGILTYKQPKPYKMATAFAAAYPYGQLRIMSSFAFTDFDQGPPSDAQGNLLSPI
Anmeru   322  N----
Aedae2   307  HGAGGDNV  H T  N  IGS  ML H   II V    F DNG     Q GN   IP S
Aedatp   307  HGAGGDNV  H V  N       ML H   IV      F  NG     Q GG   VP S
Cutars   312  HGAGGPN   HQLARN       ML F F IK V    Y D TE    Q  N   IP V
Dropro   310  HGAGGADV    V  Q    S  ML H F TP V    S   T     TTDGH IA
Aedae1   556  HGAGGDS     TKPQ IQ V  TL TD  IA V   YN S P     Q TVQVIR  G
```

(a)

```
                    ↓    ↓
Angamb  378 INPDKTCGGGWVCEHRWRQMYSMIHFRNLAWGTPLRHWWDNGNNQIAFARGDVGFVAFNN
Anmeru  378                                          Q              N
Aedae2  367   A NS  N  A       I N VG   AVR  GIND  S   Q    C  GS      L
Aedatp  367     NS  N  A       I N VG   AVR  N ND  T   M    C  AN      L
Cutars  372   ERGL DN Y         FN VE S VVR  GVND     A  M  C  NR  I   L
Dropro  370 F S NS S           I N VA   TVGSDEIQN     S   S S  SR
Aedae1  616 FAVNNS ST        PEIRK  Q K FVA SS D IQAT-Q TF  C  EK  IV

                              ↓          ↓
Angamb  438 EPFDMNVILQTGLPAGIYCDVISGAREGETCTGLQVIVEPNGFASISIRANAEDGVIAIH
Anmeru  438
Aedae2  427  GY   EN   C S  T       SK  GS---SL LQCKSVEMDVP SIFRPKVMMVFW
Aedatp  427  SY   ET   C    T       SK  GS   AT Q GGD R N Q P HGD   L
Cutars  432  S NLSQT   C    T       VKQ S    GT T GA  L Q T QTS W   L
Dropro  430 DNY L SS        T       SKS SS   KT T GSD R   N GSSED   L
Aedae1  675 SENTITQQYH C  Q Q   I   EVSQS    TV N DA  D E TLPK S--- V  Y

Angamb  494 SEVSVLVCVLPNEV*
Anmeru  494      C  CRMRSDGRV*
Aedae2  484 PSM*
Aedatp  487 INAKL*
Cutars  492 ADSRI*
Dropro  490 VNAKL*
Aedae1  732 IPSRLS*
```

FIGURE 6. Sequence alignment of the conceptual translation products of insect α-amylase genes. The amino acid sequences of seven α-amylase gene products have been aligned to show maximum sequence identities (Altschul and Erickson, 1986; Altschul *et al.*, 1990). The program lists the two most identical sequences at the top, followed by those with progressively less identity. The sequences of the Anopheline mosquitoes, *An. gambiae*, *Angamb*, (L04753) and *An. merus Anmeru*, (U012I0), are listed first. Following these are the conceptual translation products of the Culicine mosquitoes, *Ae. aegypti Amy II* (*Aedae2*) (AF000568), *Ae. atropalpus* (*Aedatp*) (U01209) and *C. tarsalis* (*Cutars*) (U01211). The *D. melanogaster* proximal α-amylase (*Dropro*) (L22716) comes next, followed by the *Ae. aegypti Amy I* sequence (*Aedae I*) (AF000569). The two *Ae. aegypti* sequences are in bold to facilitate comparisons. Dashes have been introduced to maximize alignment. Identities to the *An. gambiae* sequence are left as blanks. The arrows indicate conserved cysteine residues and the underlined and numbered amino acids are conserved regions in α-amylases (Svennson, 1988). Numbering on the left refers to the amino acid sequence of the conceptual translation product. Apparent secretory leader peptides have been left out of the figure.

A number of other notable features are evident in examining the mosquito α-amylases. The Anophelinae have apparently deleted a region that is highly conserved in the Culicinae and Drosophila, 307–313 aa in *Amy II* (Fig. 6). Furthermore, the Culicinae appear to have conserved among themselves regions 24 to 28 and 153 to 156 aa.

DISCUSSION

The partial characterization of the genomic structure of two α-amylase encoding genes, *Amylase I* and *Amylase II*, shows them to be unlinked. The presence of a complex pattern following Southern analysis of *Amy I* suggests the presence of multiple copies of this gene at the chromosome 2 locus. However, we were only able to recover two classes of genomic clones based on restriction site fragment polymorphisms. Furthermore, although representative members, F and 14A, of these two classes had distinct restriction enzyme fragment patterns, the fragments overlap one another based on hybridization, and the primary sequence of the coding and non-coding regions of 14A matched with few nucleotide differences. The small number of nucleotide differences, three out of 263, is a likely indication that F and 14A are alleles. Alternately, F and 14A may be duplicate copies of *Amy I* that have undergone concerted evolution. Human and Drosophila amylase genes have been shown to experience gene conversion (Gumucio *et al.*, 1988; Popadic and Anderson, 1995), although there is evidence in *D. melanogaster* that conversion occurs only in coding regions (Shibata and Yamazaki, 1995). It is possible that the *Amy I* probes cross-hybridize with *Amy II* coding fragments in the Southern analysis. However, we think that this is unlikely because we did not observe this cross-hybridization using isolated genomic clones. From our data it is thus not possible to determine how many copies of *Amy I* exist at the chromosome 2 locus.

The analysis of *Amy II* reveals a complexity that is in contrast to what we see with *Amy I*. *Amy II* has a relatively simple hybridization profile in Southern analysis, yet we were able to recover four different types of clones based on restriction enzyme digestion patterns. The recovery of these fragments indicates that the locus is

highly polymorphic with respect to restriction enzyme digestion sites, and may contain more than one copy of the *Amy II* gene.

One of the consistent observations we have made working with *Ae. aegypti*, and with the Rockefeller strain in particular, is that individual genes in populations appear to be highly polymorphic with respect to the primary DNA structure of alleles and numbers of alleles. In our work to characterize salivary gland-specific gene function in this vector mosquito, we constantly find nucleotide polymorphisms among cDNAs from one locus (James *et al.*, 1991; Grossman and James, 1993), among cDNAs and their corresponding genomic DNAs (James *et al.*, 1991; Grossman, 1994; Smartt *et al.*, 1995), and now in genomic DNAs. Much of the variation we see may represent the complexity of the population from which libraries are constructed, and from which the genes are isolated and characterized. An example of this is the 5.4 kb 3A1 fragment, which hybridizes faintly in the genomic Southern analysis, indicating that it is not highly represented in the population of mosquitoes used to make the DNA, yet it was straightforward to isolate the parent clone from the library.

The structure of the *Amy I* gene has previously been observed to be unique in the characterized α-amylase of insects (Grossman and James, 1993). Indeed, when the sequence was first reported, there were concerns expressed that it represented a cloning artifact. However, the size of the mRNA in the Northern analysis (Grossman and James, 1993) and the present sequence analysis of the genomic clone verify its unusual structure. In addition to its atypical length, we find that it has an intron about 70% of the distance through the coding region. The α-amylase genes of *D. melanogaster* do not exhibit any introns, however those of *Drosophila pseudoobscura* have one intron (Brown *et al.*, 1990), and Da Lage *et al.* (1996) found that only five out of 146 species of *Drosophila* lack an intron. In contrast to the mosquito, these introns are found in the 5'-end portion of the coding region, although in the Da Lage study no efforts were made to look for introns elsewhere in the gene.

The structure of *Amy II* is more consistent with that known from other α-amylases. We had initially speculated that *Amy II* would be expressed in the midgut, but all efforts to demonstrate an RNA product by Northern analysis were unsuccessful (Grossman, 1994). The open reading frame and structure of the promoter sequence appear to be intact, so we are unable to determine why we have not been able to demonstrate expression.

The promoter regions of both *Amy I* and *Amy II* have several sequence motifs that share at least 70% homology with *cis*-acting regulatory sequences found upstream of mammalian genes shown to be regulated by cAMP. In addition, there appear to be sequences similar to those found upstream of the *D. melanogaster* amylase gene that are involved in substrate feed-back regulation. However, the glucose repression elements are partly conserved in the mosquito amylase predicted promoter regions, and whether the mosquito amylases are regulated in a similar manner to the fruitfly amylase genes is not known. It may be that a 69% conservation at the nucleotide level (the highest percentage seen between the mosquito sequence of the fruitfly glucose repression element) is not sufficient to confer this same control over the mosquito transcript, as no glucose repression was seen for either the salivary gland-specific or the midgut amylase proteins (Grossman, 1994).

The comparisons of the conceptual translation products of *Amy II* with other deduced mosquito α-amylases reflect the systematics of the hematophagic subfamilies. Using all of the available mosquito sequence data, the α-amylases of the Anophelinae are more similar to one another than to those of the Culicinae. Conversely, the Culicinae are more similar to one another than the Anophelinae. The interesting observation comes from looking at the similarity of *Amy I* to the other sequences. The *Amy I* conceptual translation product is dissimilar to all mosquito α-amylase sequences, and is even more divergent than a *D. melanogaster* sequence. This indicates that *Amy I* is a highly diverged member of the α-amylase gene family. Most insects have multiple α-amylase genes, and one copy of the complement present in *Ae. aegypti* may have evolved to accommodate a role in the salivary glands by large-scale nucleotide substitutions and acquiring a novel amino-terminal domain. The use of α-amylase genes for systematic analyses (Hickey *et al.*, 1987; Da Lage *et al.*, 1996; Popadic *et al.*, 1996) must therefore be tempered by the presence of orthologous and paralogous members of the gene family, and for *Ae. aegypti*, *Amy I* and *Amy II* are clearly paralogous. This analysis has demonstrated two separate loci encoding α-amylases in *Ae. aegypti*, and this should contribute to an understanding of how genes evolve to take on new functions in those families of insects that have evolved hematophagy.

REFERENCES

Altschul S. R., Gish W., Miller W., Myers E. W. and Lipman D. J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410.

Altschul S. F. and Erickson B. W. (1986) Optimal sequence alignment using affine gap costs. *Bull. Math. Biol.* **48,** 603–616.

Bender W., Spierer P. and Hogness D. S. (1983) Chromosomal walking and jumping to isolate DNA from the Ace and rosy loci and the bithorax complex in *Drosophila melanogaster*. *J. Mol. Biol.* **168,** 17–33.

Benkel B. F. and Hickey D. A. (1987) A Drosophila gene is subject to glucose repression. *Proc. Natl. Acad. Sci. U.S.A.* **84,** 1337–1339.

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

Bilofsky H. S. and Burks C. (1988) The GenBank (R) Genetic Sequence Data Bank. *Nucl. Acids Res.* **16,** 1861–1864.

Blattner F., Williams B. G., Blechl A. E., Denniston-Thompson K., Faber H. E., Furlong L.-A., Grunwald D. J., Kiefer D. O., Moore D. D., Sheldon E. L. and Smithies O. (1977) Charon phages: safer derivatives of bacteriophage lambda for DNA cloning. *Science* **196,** 161–169.

Brown C. J., Aquadro C. F. and Anderson W. W. (1990) DNA

sequence evolution of the amylase multigene family in *Drosophila pseudoobscura*. *Genetics* **126,** 131–138.

Cherbas L. and Cherbas P. (1993) The arthropod initiator: the capsite consensus plays an important role in transcription. *Insect Biochem. Mol. Biol.* **23,** 81–90.

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

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

De Crombrugghe B., Busby S. and Buc H. (1984) Cyclic AMP receptor protein: role in transcription activation. *Science* **224,** 831–838.

Feng D.-F. and Doolittle R. F. (1987) Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *J. Mol. Evol.* **25,** 351–360.

Grossman, G. L. (1994) The amylase gene family of the vector mosquito, *Aedes aegypti*. Ph.D. Dissertation, Harvard University.

Grossman G. L. and James A. A. (1993) The salivary glands of the vector mosquito, *Aedes aegypti*, express a novel member of the amylase gene family. *Insect Mol. Biol.* **1,** 223–232.

Gumucio D. L., Wiebauer R. M., Caldwell R. M., Samuelson L. C. and Meisler M. H. (1988) Concerted evolution of human amylase genes. *Mol. Cell. Biol.* **8,** 1197–1205.

Hickey D. A. (1977) Selection for amylase allozymes in *Drosophila melanogaster*. *Evolution* **31,** 800–804.

Hickey D. A., Benkel B. F., Boer P. H., Genest Y., Abukashawa S. and Ben-David G. (1987) Enzyme coding genes as molecular clocks: the molecular evolution of animal alpha-amylases. *J. Mol. Evol.* **26,** 252–256.

Imagawa M., Chiu R. and Karin M. (1987) Transcription factor AP-2 mediates induction by two different signal-transduction pathways: protein kinase C and cAMP. *Cell* **51,** 251–260.

Inomata N., Shibata H., Okuyama E. and Yamazaki T. (1995) Evolutionary relationships and sequence variation of alpha-amylase variants encoded by duplicated genes in the Amy locus of *Drosophila melanogaster*. *Genetics* **141,** 237–244.

James A. A., Blackmer K., Marinotti O., Ghosn C. R. and Racioppi J. V. (1991) Isolation and characterization of the gene expressing the major salivary gland protein of the female mosquito, *Aedes aegypti*. *Mol. Biochem. Parasitol.* **44,** 245–254.

Kikkawa H. (1953) Biochemical genetics of *Bombyx mori* (silkworm). *Adv. Genet.* **5,** 107–140.

Lander E. S., Green P., Abrahamson J., Barlow A., Daly M., Lincoln S. E. and Newburg L. (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* **1,** 174–181.

Laulier M. (1988) Genetique et systematique evolutives du complexe d'especes *Sphaeroma hookeri* leach, *Sphaeroma levii* Argano et *Sphaeroma rugicauda* leach (Crustaces, Isopodes Flabelliferes). 1. Genetique formelle de onze locus enzymatiques. *Genet. Select. Evol.* **20,** 63–74.

Levy J. N., Gemmill R. M. and Doane W. W. (1985) Molecular cloning of alpha-amylase gene from *Drosophila melanogaster*. II. Clone organization and verification. *Genetics* **110,** 313–324.

Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, New York.

Mosquito Genomics (1997) The *Aedes aegypti* database. Available from the World Wide Web at URL, http://klab.agsci.colostate.edu.

Ogita Z. (1968) Genetic control of isozymes. *Ann. N.Y. Acad. Sci.* **151,** 243–262.

Oxford G. S. (1986) Multiple amylase loci in *Asellus* (Crustacea, Isopoda): genetics and linkage. *Heredity* **56,** 105–110.

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

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

Pope G. J., Anderson M. D. and Bremner T. A. (1986) Constancy and divergence of amylase loci in four species of *Tribolium* (Coleoptera, Tenebrionidae). *Comp. Biochem. Physiol.* **83B,** 331–333.

Roestler W. J., Vandenbark G. R. and Hanson R. W. (1988) Cyclic AMP and the induction of eukaryotic gene transcription. *J. Biol. Chem.* **263,** 9063–9066.

Sanger F., Nicklen S. and Coulson A. R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* **74,** 5463–5467.

Severson D. W., Mori A., Zhang Y. and Christensen B. M. (1993) Linkage map for *Aedes aegypti* using restriction fragment length polymorphisms. *J. Hered.* **84,** 241–247.

Severson D. W., Thathy V., Mori A., Zhang Y. and Christensen B. M. (1995) Restriction fragment length polymorphism mapping of quantitative trait loci for malaria parasite susceptibility in the mosquito *Aedes aegypti*. *Genetics* **139,** 1711–1717.

Silver B. J., Bokar J. A., Virgin J. B., Vallen E. A., Milsted A. and Nilson J. H. (1987) Cyclic AMP regulation of the human glycoprotein hormone A-subunit gene is mediated by an 18-base-pair element. *Proc. Natl. Acad. Sci. U.S.A.* **84,** 2198–2202.

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

Smartt C. T., Kim A. P., Grossman G. L. and James A. A. (1995) The *Apyrase* gene of the vector mosquito, *Aedes aegypti*, is expressed specifically in the adult female salivary glands. *Exp. Parasitol.* **81,** 239–248.

Southern E. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98,** 503–517.

Stam P. (1993) Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. *Plant J.* **3,** 739–744.

Sudhof T. C., Russell D. W., Brown M. S. and Goldstein J. L. (1987) 42 bp Element from LDL receptor gene confers end-product repression by sterols when inserted into viral TK promoter. *Cell* **48,** 1061–1069.

Svennson B. (1988) Regional distant sequence homology between amylases, α-glucosidases and transglucanosylases. *FEBS Lett.* **230,** 72–76.

Von Heijne G. (1985) Signal sequences — the limits of variation. *J. Mol. Biol.* **184,** 99–105.

Acknowledgements—The authors thank Dr Donal Hickey for sharing unpublished data and the *Amy II* 551 bp probe, Dr Kurt Yardley for help in the Southern analysis, Dr Craig Coates and Susan Fisher with help in the data analysis and figures, Theresa Schaub for help in sequencing, and Lynn Olson for help in preparing the manuscript. The work was supported by grants from the NIH NIAID AI 29746 (A.A.J.) and AI33127 (D.W.S.), and the John D. and Catherine T. MacArthur Foundation.