
Cristina P. Vieira, Jorge Vieira, and Daniel L. Hartl

Department of Organismic and Evolutionary Biology, The Biological Laboratories, Harvard University

We analyzed a 5,770-bp genomic region of Drosophila virilis that contains a cluster of two maltase genes showing sequence similarity with genes in a cluster of three maltase genes previously identified in Drosophila melanogaster. The D. virilis maltase genes are designated Mav1 and Mav2. In addition to being different in gene number, the cluster of genes in D. virilis differs dramatically in intron-exon structure from the maltase genes in D. melanogaster; the transcriptional orientation of the genes in the cluster also differs between the species. Our findings support a model in which the maltase gene cluster in D. virilis and D. melanogaster evolved independently. Furthermore, while in D. melanogaster the maltase gene cluster lies only 10 kb distant from the larval cuticle gene cluster, the maltase and larval cuticle gene clusters in D. virilis are located very far apart and on different chromosomes than that expected from the known chromosome arm homologies between D. virilis and D. melanogaster. A region of the genome containing the maltase and larval cuticle gene clusters appears to have been relocated between non-homologous chromosomes.

Introduction

Clusters of functionally related genes have, in some cases, a phylogenetically conserved or semi-conserved spatial organization of genes within the cluster. In other cases the spatial organization is not conserved (Martinez-Cruzado et al. 1988; Hooper et al. 1992; Droso-poulou and Scouras 1995; Stathakis et al. 1995; von Allmen et al. 1996; Wright 1996). With a few exceptions, when gene organization is conserved, it is not known whether the conservation results from functional constraints that preserve the organization or from the historical happenstance that no chromosomal rearrange-ments disrupting the organization have yet become fixed. To help distinguish between these possibilities, it is useful to compare homologous gene clusters in species that are sufficiently diverged to have had numerous chromosomal rearrangements become fixed, resulting in a reshuffling of the physical order and spacing of genes separated by an average distance of one megabase or less. Such appears to be the case with Drosophila virilis and Drosophila melanogaster. These species are suitable for studies of comparative genome organization because they diverged approximately 40–60 MYA (Powell and DeSalle 1995; Russo, Takezaki, and Nei 1995) and there has been extensive chromosome arm rearrangement. Comparative genome mapping of closely linked genes has so far failed to reveal any syntenic regions longer than could be explained by chance alone (Vieira et al. 1997a).

In this work, we examine genome organization on an even finer scale. We have used probes from the D. melanogaster “maltase gene cluster” of three maltase genes (H, D, and L) to isolate a segment of the D. virilis genome that contains the maltase gene cluster. In D. melanogaster, the “HDL gene cluster” containing the maltase H, D, and L genes comprises approximately 8 kb of DNA in polytene chromosomal region 44D in the right arm of chromosome 2 (Muller’s clement C; Snyder and Davidson 1983). The HDL gene cluster contains all of the known maltase genes in D. melanogaster, and each of the H, D, and L genes shows similarity in amino acid to the yeast maltase enzyme (Henikoff and Wallace 1988).

The 5,770-bp genomic region of D. virilis containing the maltase gene cluster was sequenced in its entirety, and its functional organization was inferred in part based on evidence from the sequence of reverse-transcription PCR products amplified from one of the maltase mRNAs. When the D. virilis maltase gene cluster is compared with that from D. melanogaster, we find major differences in intron-exon structure, gene number, and gene cluster organization. The D. virilis cluster contains only two maltase genes, designated Mav1 and Mav2, which are divergently transcribed.

The HDL gene cluster in D. melanogaster is separated by about 10 kb from another gene cluster that contains the Lep genes for larval cuticle proteins. The locations of the maltase gene cluster and the larval cuticle gene cluster in D. virilis violate the usually valid principle that, in Drosophila evolution, genes in the same chromosome element remain linked in the same chromosome element, although perhaps they are spatially rearranged (Hartl and Lozovskaya 1994). The D. virilis maltase and larval cuticle gene clusters were localized by in situ hybridization to polytene chromosome 4 (Muller’s element B) of D. virilis. This chromosome is not the homologous chromosome arm to that of polytene chromosome 2R (Muller’s element C) in D. melano-gaster (Ashburner 1989). Therefore, our data show that a region including the maltase gene cluster and the larval gene cluster changed position from one chromosome element to another. Moreover, the larval cuticle gene cluster and the maltase gene cluster are widely separated in D. virilis.

Key words: Drosophila virilis, maltase genes, gene cluster, genome organization.

Address for correspondence and reprints: Daniel L. Hartl, Department of Organismic and Evolutionary Biology, The Biological Laboratories, Harvard University, 16 Divinity Avenue, Cambridge, Massachusetts 02138. E-mail: dhartl@oeb.harvard.edu.
Materials and Methods

Strains and Nucleic Acid Preparation

*Drosophila virilis* wild-type strain 9 was originally collected in 1970 in Batumi, Georgia, in the former USSR (Lozovskaya, Scheinker, and Evgen’ev 1990). Wild-type stocks of *D. hydei* (15085–1641.12) and *D. pseudoobscura* (14011–0121.0) were obtained from the National *Drosophila* Species Stock Center in Bowling Green, Ohio. Genomic DNA was prepared as described in Nurminsky et al. (1996). DNA from *D. virilis* bacteriophage P1 clones was extracted using the Miniprep method (Hartl and Lozovskaya 1995). Plasmid DNA was extracted using the Qiaprep Plasmid Kit (Qiagen).

P1 Library Screening

We screened the *D. virilis* P1 library prepared by Lozovskaya, Petrov, and Hartl (1993) in order to isolate *D. virilis* P1 clones containing the maltase and larval cuticle genes. Filters were prepared for hybridization as described by these authors. Radioactive Southern hybridization at reduced stringency was performed at 42°C in 5 × SSPE (0.75 M NaCl, 0.05 M NaH₂PO₄, 0.005 M EDTA), 0.1% sodium lauryl sulfate (SDS), 1 × Denhardt’s solution (4 mg/ml hovine serum albumin, 4 mg/ml polyvinylpyrolidone, 4 mg/ml Ficoll), 0.1 mg/ml sonicated denatured salmon sperm DNA, and 29% formaldehyde (O’Neil and Belote 1992). The filters were washed three times at 50°C in 0.5 × SSPE containing 0.1% SDS. Hybridization probes were obtained from *D. melanogaster* genomic DNA by the polymerase chain reaction (PCR). Standard PCR amplification conditions were 25 cycles of denaturation at 94°C for 30 s, primer annealing for 30 s (temperature adjusted for individual primer pairs), and primer extension at 72°C for 3 min. The oligonucleotides used for PCR amplification were as follows: maltase gene *H*, 5’-AAACTCATCCAC-CAACTCA-3’ (position 2308 in entry K00045) and 5’-ACC1CAACTATGGCATTACCCACCCAG-3’ (position 2568 in entry K00045), annealing temperature 53°C, yielding a PCR product of 278 bp; maltase gene *D*, 5’-GGGAC-GATGGAAAACTGAA-3’ (position 5367 in entry K00045), annealing temperature 53°C, yielding a PCR product of 210 bp; maltase gene *L*, 5’-TCCACGCTCCTTCTCAAC-3’ (position 7673 in entry K00045) and 5’-AGGCCGATCATACCTACCTACCCG-3’ (position 8499 in entry K00045), annealing temperature 58°C, yielding a PCR product of 826 bp; larval cuticle protein 2 gene (*Lcp2*), 5’-TGTTAATCCGTTTTACGTTTC-3’ (position 1365 in entry V00203 J01081) and 5’-TTCCCCGCTCCCGATGATG-3’ (position 1575 in entry V00203 J01081), annealing temperature 56°C, yielding a product of size 210 bp. Primers were provided by Gibco/BRL.

Cytological Procedures

Localization of DNA probes was carried out by in situ hybridization with polytene chromosomes as described in Lozovskaya, Petrov and Hartl (1993) and Vieira et al. (1997a).

Subcloning from P1 into LambdaScan

Subclones averaging 10 kb were prepared from *D. virilis* bacteriophage P1 clone Dv66-88. The vector was the specialized LambdaScan vector described by Nurminsky and Hartl (1996), which is convenient because of the size selection for relatively large clones.

Southern Hybridization

DNA was digested with restriction enzymes, the restriction fragments separated by electrophoresis in agarose gels, and transferred onto Hybond N membrane (Amersham) as described in Sambrook, Fritsch, and Maniatis (1989). Labeling of the probe and Southern hybridization were performed as described in the Random Primer Fluorescein Labeling Kit (DuPont).

DNA Sequencing

Inserts of LambdaScan subclones were cut with restriction enzymes. The fragments of interest were transferred into the vector pSP72 (Promega) and sequenced using the transposon-facilitated DNA sequencing method of Strathmann et al. (1991). DNA sequencing was performed with an Applied Biosystems model 373A DNA sequencing system (Applied Biosystems) with the ABI PRISM Dye Termination Cycle-Sequencing Kit. Products amplified by PCR were sequenced after subcloning with the TA Cloning Kit (Invitrogen).

Computer Analysis

The coding region of the two maltase genes from *D. virilis* are deposited in GenBank (accession number AF006573). Alignment of the *Mavl* and *Mav2* amino acid sequences with other maltase sequences was based on the alignment of Zheng et al. (1995) using the computer program ESEE (the EyeBall Sequence Editor, version 1.04). DNA database screening was done using the updated EMBL and GenBank Nucleotide Sequence Data Library. The maltase genes we used in the comparison were as follows: the sequence data of *D. melanogaster* deposited in GenBank (accession number K00045 from Snyder and Davidson 1983); *Anopheles gambiae* (accession numbers X87410 and X87411 from Zheng et al. 1995); and *Aedes aegypti* (accession numbers M30443, M22322, and M22364 from James, Blackmer, and Racioppi 1989).

Results

Isolation and Cytological Localization of the Maltase Genes in *D. virilis*

We probed, by Southern hybridization at low stringency conditions, genomic DNA from *D. virilis* with four different probes. The probes were PCR amplification products of three maltase genes (*H*, *D*, and *L*) and one larval cuticle gene (*Lcp2*), all from *D. melanogas-
ter. The rationale for using the Lcp2 probe is that, in D. melanogaster, the HDL gene cluster lies 10 kb away from the larval cuticle gene cluster, which includes the Lcp2 larval cuticle gene. The probes for D. melanogaster genes H and D did not yield any hybridization signal. However, the probe for D. melanogaster gene L revealed two hybridization signals, and the D. melanogaster Lcp2 probe gave one hybridization signal. Therefore, we screened a D. virilis library that includes 10,080 P1 clones (Lozovskaya, Petrov, and Hartl 1993) with these D. melanogaster L and Lcp2 probes. When the L probe was used, only one P1 clone yielded a positive signal, clone Dv66-88. When the Lcp2 probe was used to screen the same P1 library, three positive P1 clones were obtained: Dv1-23, Dv2-45, and Dv2-47. All three clones yielded the same pattern of restriction fragments when digested and probed in Southern hybridizations with Lcp2.

The Dv66-88 and the Dv01-23 clones were hybridized with D. virilis genomic DNA. The size and number of hybridization bands were compatible with the size and number of bands observed when the D. melanogaster L and Lcp2 probes were hybridized with D. virilis genomic DNA.

Using Dv66-88 and Dv1-23 as probes for in situ hybridization with polytene chromosomes, we localized the maltase genes to D. virilis chromosome 4 section 43B/C (Muller’s element B) and the larval cuticle protein genes to chromosome 4 section 40A (Muller’s element B; fig. 1). The probe Dv1-23 also yielded a weak signal in the chromocenter. The chromosomal region designations follow those in the photograhic map of Gubenko and Evgen’ev (1984).

Judging from the localization of many other markers, polytene chromosome 4 (Muller’s element B) of D. virilis is not the homolog of D. melanogaster polytene chromosome 2R (Muller’s element C), in which the HDL gene cluster is located; it is the homolog of polytene chromosome 2L (Ashburner 1989). It should be noted that, in the D. melanogaster genome, no maltase genes could be detected by Southern hybridization other than the three in the maltase gene cluster (Snyder and Davidson 1983).

The Dv66-88 and Dv1-23 clones were also mapped by in situ hybridization in D. novamexicana, D. montana, and D. hydei. The localizations in D. novamexicana were at 43B (Muller’s element B; Dv66-88) and 40A (Muller’s element B; Dv1-23), using the chromosomal region designations of Vieira et al. (1997b). In D. montana, the localizations were at 45E (Muller’s element B; Dv66-88) and 41B (Muller’s element B; Dv1-23) using the designations of Vieira et al. (1997b). In D. hydei, the probes localized to sections 83 (Muller’s element B; Dv66-88) and 94 (Muller’s element B; Dv1-23) using the chromosomal region designations of Ananiev and Barsky (1982).

We also used a P1 clone from D. melanogaster (Dm1-66) that contains the maltase L gene (D. Nurminsky, personal communication) to determine the localization of the maltase genes in the D. pseudoobscura polytene chromosomes. Dm1-66 hybridized with D. pseudoobscura chromosome 3 section 65 (Muller’s element C) in the designations of Kastritis and Crumpacker (1966). Judging from figure 4b in Steinemann and Steinemann (1990), this hybridization signals conflict with the cuticle genes. This finding raises the possibility that the genomic region containing the maltase gene cluster and the larval cuticle protein gene cluster may be similar in D. pseudoobscura and D. melanogaster.
Genomic Structure of the Maltase Genes in *D. virilis*

We sequenced a region of 5,770 bp from Pl clone Dv66-88 that contains the coding regions of two maltase genes, here termed *Mavl* and *Mav2*. *Mavl* is 1,878 bp long and encodes a putative polypeptide of 586 amino acids, whereas *Mav2* is 1,709 bp long and encodes a putative polypeptide of 524 amino acids. The spacing between the genes is 1,920 bp. Both genes are oriented in the same direction (fig. 2). Hybridization with genomic DNA using *Mavl* and *Mav2* as probes shows that, in *D. virilis*, both genes are present in a single copy (fig. 3). The *Mavl* and *Mav2* coding sequences are 69% identical at the nucleotide level and 65% identical at the amino acid level.

From the analysis of the DNA sequence of the *Mavl* gene and the alignment of the putative amino acid sequence with other maltase proteins, we postulated the presence of two introns of sizes 53 bp (extending from nucleotide position 1296 to position 1348 in entry AFO06573) and 61 bp (extending from nucleotide position 1696 to position 1757 in entry AFO06573). A similar analysis of *Mav2* also suggested the presence of two introns of sizes 60 bp (nucleotide positions 5118–5178 in entry AFO06573) and 70 bp (nucleotide positions 5330–5403 in entry AFO06573). The splice and acceptor sequences that flank these putative introns match the consensus for small introns in *Drosophila* (Mount et al. 1992). These putative introns not only disrupt the open reading frames of *Mavl* and *Mav2*, but they also contain in-frame translation termination codons that would lead to truncated proteins if the intervening sequences were present in the mRNA. The inference that *Mavl* contains two introns was confirmed by analysis of the PCR-amplified cDNA sequence flanking the two putative introns of *Mavl*. The sequences corresponding to the introns are absent from the *Mavl* mRNA.

Several potential regulatory motifs were recognized in the DNA sequence that resemble but do not always completely match those identified by Breathnach and Chambon (1981), Arkhipova (1995), and Robin et al. (1996). They include putative TATA boxes (positions 32 and 3792 in entry AFO06573), cap-site sequences (positions 45 and 3862 in entry AFO06573), polyadenylation signals (positions 2107 and 5635 or 5655 in entry AFO06573), and a 5'-CCACCTTAG-3' motif (position 3203 in entry AFO06573) found upstream of each of the *H*, *D*, and *L* genes in *D. melanogaster* (Snyder and Davidson 1983).

Similarity to Other Maltase Genes

Database searches revealed that the encoded *Mavl* and *Mav2* putative polypeptides share significant similarities with maltases and other α-glucosidases. The highest similarities were found with maltases from *D. melanogaster* (Snyder and Davidson 1983), *Anopheles gambiae* (Zheng et al. 1995), and *Aedes aegypti* (James, Blackmer, and Racioppi 1989).

The maltase sequences from *Drosophila* and mosquitoes were aligned based on the alignment of Zheng et al. (1995). However, it was not possible to align *Mavl* and *Mav2* with the other sequences in the amino and...
carboxyl ends because of the low level of similarity in these regions and the difference in protein sizes (fig. 4).

Four regions that are conserved among all α-glucosidases (Svensson 1988) were identified. These are denoted regions 2, 3, 4, and 6 in figure 4. Both Mav1 and Mav2 in D. virilis possess a highly conserved substrate-binding histidine residue in region 3 and a catalytic aspartic acid residue in the NHD motif in region 6. These residues are marked with an asterisk.

In fact, the principal similarity among the sequenc- es results from the shared sequence motifs reflecting functional constraints on the maltase structure, which is illustrated by the observation that the similarity between the Drosophila and the mosquito sequences is not much greater than that between the sequences of the maltases from D. melanogaster and D. virilis (table 1).

Discussion

The lack of conservation of the maltase gene cluster in D. virilis and D. melanogaster has precedents in other systems. In species of the Drosophila montium subgroup, the organization of the beta-tubulin genes differs among species. In the North Oriental sibling species, three beta-tubulin genes are organized in a single cluster. On the other hand, in the South Oriental species, one of the three genes is separated and at a very distant location. Among Indian species of the group, as well in the Afrotropical species, still another arrangement is found, in which the three genes are completely dispersed (Drosopoulou and Scouras 1995).

There are also examples in which gene clusters are conserved. The autosomal chorion gene cluster of four chorion genes is present in the same order, spacing, and orientation in D. melanogaster, D. subobscura, D. virilis, and D. grimshawi (Martinez-Cruzado et al. 1988). The conserved organization of the cluster contrasts with the sequence divergence among the genes themselves. An example of a partially conserved cluster is the dopa decarboxylase (Ddc) gene cluster. This group of functionally related genes is clustered in one location in D. melanogaster (Statthakis et al. 1995; Wright 1996). However, in D. virilis and D. pseudoobscura, the gene cluster is separated into two subclusters, which are on different locations in the same chromosome, but each of the subclusters is intact (Statthakis et al. 1995).

The archetypes of conserved gene clusters are the homeotic gene clusters, which have been studied extensively in both invertebrates and vertebrates. In D. melanogaster, the homeotic genes are organized into two major clusters, the Antennapedia complex (ANT-C) and the Bithorax complex (BX-C). The organization of the ANT-C gene cluster appears to be conserved in D. melanogaster and D. virilis (Hopper et al. 1992). However, the BX-C cluster is split into two parts in D. virilis (von Allmen et al. 1996), although the functional sequelae of the split organization, if any, are unclear. Genetic studies with D. melanogaster suggest that the spatial organization of the genes in the BX-C cluster is important for wild-type function (Duncan 1987).

In D. melanogaster, the maltase gene family is composed of three clustered genes in a region of 8 kb (Snyder and Davidson 1983), whereas in D. virilis there are only two genes clustered in a region of 5.7 kb. The D. virilis genes Mav1 and Mav2 are transcribed in the same orientation, and each encodes a protein of approximately 520 amino acids. In D. melanogaster, the contiguous genes are transcribed in opposite directions. The sequence similarity of Mav1 and Mav2 indicates that the two genes probably arose by a single gene duplication event, but there is enough sequence divergence to indicate that the putative duplication was not recent. The position of the first intron is the same in Mav1 and Mav2, but the intron sequences are very divergent. Moreover, the position of the second intron is not the same in Mav1 and Mav2. None of the D. virilis Mav1 or Mav2 introns shares a position with any of the introns in the D. melanogaster maltase genes. Therefore, the
maltase gene clusters in Anopheles gambiae (Zheng et al. 1995); and D. melanogaster (James, Blackmer, and Davidson 1983). We find a similar sequence in D. virilis; H, D, and L genes only about 32% of synonymous nucleotides would have escaped substitution by chance alone (Hartl and Lozovsky 1995). Any block of six or more conserved nucleotides upstream of the ATG codon of the maltase gene. The function of this conserved sequence is unknown. Taking into account the divergence time between 40 Myr, it has been estimated that 86 conserved amino acids among all sequences analyzed are underlined. The bold letters identify four regions (2, 3, 4, and 6) conserved among the...
otides flanked by divergent sequences may be taken as also noteworthy. In *D. virilis*, *D. novamexicana*, *D. montana*, and *D. hydei*, the maltase and larval cuticle gene clusters are located in Muller's element B. In *D. melanogaster* and in *D. pseudoobscura*, the HDL maltase cluster and the larval cuticle gene cluster are located in Muller's element C. It is unclear whether this rearrangement took place in the lineage of the subgenus *Sophophora* or in the lineage of the subgenus *Drosophila*. The time to the last common ancestor between *D. virilis* and *D. melanogaster* has been estimated as approximately 40 Myr; between *D. pseudoobscura* and *D. melanogaster* it is approximately 25 Myr, and between *D. virilis* and *D. hydei* it is approximately 20 Myr (Russo, Takezaki, and Nei 1995; Kwiatowski et al. 1994). However, it should be noted that these estimates are based on the sequence of at most two genes under the assumption of a molecular clock. Therefore, the divergence times should be taken with due caution (Powell and DeSalle 1995). In any case, these dates support the age of the change in chromosome arm location as being at least 20

### Table 1: Matrix of Percent Pairwise Amino Acid Indentities of Insect Maltase Genes

<table>
<thead>
<tr>
<th></th>
<th>Mov2</th>
<th>D</th>
<th>H</th>
<th>L</th>
<th>Agm1</th>
<th>Agm2</th>
<th>Mal1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mov2</td>
<td>64</td>
<td>49</td>
<td>51</td>
<td>49</td>
<td>47</td>
<td>52</td>
<td>50</td>
</tr>
<tr>
<td>Mov1</td>
<td>49</td>
<td>51</td>
<td>51</td>
<td>47</td>
<td>47</td>
<td>52</td>
<td>50</td>
</tr>
<tr>
<td>D</td>
<td>47</td>
<td>52</td>
<td>54</td>
<td>54</td>
<td>52</td>
<td>50</td>
<td>54</td>
</tr>
<tr>
<td>H</td>
<td>47</td>
<td>52</td>
<td>54</td>
<td>54</td>
<td>52</td>
<td>50</td>
<td>54</td>
</tr>
<tr>
<td>L</td>
<td>47</td>
<td>52</td>
<td>54</td>
<td>54</td>
<td>52</td>
<td>50</td>
<td>54</td>
</tr>
<tr>
<td>Agm1</td>
<td>49</td>
<td>52</td>
<td>54</td>
<td>54</td>
<td>52</td>
<td>50</td>
<td>54</td>
</tr>
<tr>
<td>Agm2</td>
<td>49</td>
<td>52</td>
<td>54</td>
<td>54</td>
<td>52</td>
<td>50</td>
<td>54</td>
</tr>
</tbody>
</table>

**NOTES:** Due to the low level of similarity among the insect maltases in the carboxyl ends and to the difference in protein size, only the aligned regions shown in figure 4 were considered for the pairwise identities.

![Fig. 4 (Continued)](image-url)
Myr. Furthermore, in D. melanogaster, the maltase and larval cuticle genes are only 10 kb apart (Snyder and Davidson 1983), whereas, in all the species of the Drosophila subgenus analyzed, these two clusters are always in very different regions of the same chromosome. Once again, there is no evidence to support the hypothesis that any of the observed relative positions of these two gene clusters is the ancestral state. Considering the minimum age of the change in the location of the maltase and larval cuticle gene clusters (at least 20 Myr), it is possible that none of the observed relative positions is the ancestral state.

Acknowledgments

We are grateful to Dmitry Nurminsky for providing D. melanogaster clones, protocols, and helpful technical advice on how to use the sequence scanning method, and to E. Fidelma Boyd for helping with the computer programs. J.V. is supported by a grant from Programa de Doutoramento Gulbenkian em Biologia e Medicina, FLAD and JNICT. This work was supported in part by NIH grant HG01250.

LITERATURE CITED


JEFFREY R. POWELL, reviewing editor

Accepted June 13, 1997