Interchromosomal Exchange of Genetic Information between Gene Arrangements on the Third Chromosome of *Drosophila pseudoobscura*

Aleksandar Popadić, Danijela Popadić, and Wyatt W. Anderson
Department of Genetics, University of Georgia

During the last 60 years, the inversion polymorphism on the third chromosome of *Drosophila pseudoobscura* has become a case study of the evolution of linked blocks of genes, isolated from each other by the suppression of recombination in heterozygotes for different inversions. Due to its location within inverted regions in most gene arrangements, the amylase (*Amy*) gene region can be used to elucidate the molecular pattern of evolution in these inversions. We studied this region in the Tree Line phylad of gene arrangements, with regard to both restriction site polymorphisms (RSP) and nucleotide sequences. The analysis of restriction maps, encompassing 26 kb, corroborates the cytogenetic phylogeny established on the basis of inversion breakpoints. However, we found that the 2.7 kb of nucleotide sequences of the *Amy1* gene are identical in both Estes Park and Hidalgo arrangements, despite the fact that these inversions arose independently from Tree Line. These contrasting results suggest that a homogenizing force, most likely gene conversion, is able to bring about localized exchanges between otherwise isolated gene arrangements.

**Introduction**

The α-amylase (*Amy*) locus in *Drosophila pseudoobscura* is a multigene family of one, two, or three copies. Due to its location within an inverted region in most gene arrangements on the third chromosome, the evolution of the *Amy* multigene family is intertwined with the evolution of gene arrangements. These arrangements resulted from overlapping, paracentric inversions, with three of them—Standard (ST), Santa Cruz (SC), and Tree Line (TL)—occupying central positions in a phylogeny based on inversion breakpoints (Dobzhansky and Epling 1944). Each of these central arrangements, together with the inversions derived from it, constitutes a family, or phylad, of gene arrangements. Depending on the arrangement in question, the number of *Amy* genes varies from one (TL) to two (SC) to three (ST). This pattern of distribution reflects the evolutionary history of the regions contained within these gene arrangements, due to the suppression of recombination in heterozygotes for different inversions (Dobzhansky and Epling 1948). Recent analysis of restriction site polymorphism (RSP) within the inverted regions produced additional evidence that inversions are indeed an effective barrier against the homogenizing effects of recombination (Aquadro et al. 1991). Furthermore, the phylogeny derived from the RSP data provided a striking independent corroboration of the inversion phylogeny (Aquadro et al. 1991; Popadic and Anderson 1994).

In this report, we focus on the molecular evolution of the *Amy* locus in the TL phylad. The amylase region was mapped in three arrangements which are derived from TL: Olympic (OL), Estes Park (EP), and Hidalgo (HI), and in an additional TL strain as well. These four chromosomes were then phylogenetically analyzed jointly with the rest of the TL strains from the original RSP data set. In addition, the *Amy1* genes from these four chromosomes were cloned, sequenced, and analyzed. We then utilized the established evolutionary relationships between arrangements in the TL phylad, based on the inversion phylogeny, to evaluate results from both the RSP and DNA sequence data sets.

**Material and Methods**

*Drosophila pseudoobscura* lines homozygous for the third chromosome were constructed by using balancer stocks (Pavlovsky and Dobzhansky 1966). Salivary glands were dissected from third-instar larvae, 30 per line, and gene arrangements were diagnosed from squash preparations of the polytene chromosomes. Hybridization in situ to polytene chromosomes was carried out with probe DNA labeled with 16-dUTP, as described by
Lim (1993). Four strains were used for this study: TL, strain AH 73#2, from northern California; OL, strain s14AR-D, from British Columbia, Canada; EP, strain BC p430#4, from British Columbia, Canada; and HI, strain MEX a-176x, from Amecameca, Mexico. The frequencies of gene arrangements in a collection taken at Amecameca in January 1977 were calculated by studying a single larva from the offspring of each wild female with her mate in nature.

We carried out a restriction mapping of these four strains, as described earlier (Aquadro et al. 1991). Note that in the case of the EP and HI strains, the information generated from the restriction analysis was also used to infer diagnostic restriction patterns for the two strains. In particular, we used the fact that two HindIII sites (at locations 1.7 and 8.8) are present in the EP, but are absent in the HI, for that purpose (fig. 1). Digests were then carried out on the genomic DNA that was used for cloning, followed by Southern blotting and hybridization to a labeled probe (Aquadro et al. 1991). This additional step allowed us to eliminate the possibility of stock contamination, which was one conceivable explanation for the unexpected observation of identity in the EP and HI sequences.

To clone and sequence the Amy1 gene, which is located within a 5.6-kb HindIII/EcoRI fragment (fig. 1), we used an approach described previously (Popadic and Anderson 1994). Four recombinant phagemids were identified: pAP4 (containing TL Amy1); pAP5 (containing EP Amy1); pAP8 (containing OL Amy1); and pAP9 (containing HI Amy1). Sequence divergence estimates were calculated as direct counts of nucleotide sequence differences, since no correction is needed for differences as small as those in our study (Nei 1987, p. 64). Phylogenetic analysis was done by the neighbor-joining (NJ) method in the PHYLP package (Felsenstein 1993) and by the maximum parsimony method in the PAUP program (Swofford 1991).

Results and Discussion

The RSP Data

The genomic restriction maps of the Amy region, encompassing 26 kb, were determined for the four strains analyzed in this report (fig. 1). All four chromosomes carry the two 1.6-kb deletions, D2 and D5, relative to the ST chromosome of the probe. These two deletions remove the BamHI and Sal I restriction sites marking the coding regions of Amy2 and Amy3. This is the same

![Fig. 1.—Map of restriction site polymorphism for the amylase gene region of Drosophila pseudoobscura. Polymorphic and monomorphic restriction sites are indicated above and below the line, respectively, as follows: B, BamHI; E, EcoRI; H, HindIII; S, Sal I; X, Xba I; Z, Xho I. The scale is oriented with 0.0 at a Xho I site. The locations of Amy genes are boxed, and deletions relative to the ST map are shown as triangles above the map.](https://academic.oup.com/mbe/article-abstract/12/5/938/974555/12532856370555?guest=0)
pattern that was observed in all of the TL strains from the previous RSP analysis (Aquadro et al. 1991). The presence of a single amylase gene (Amy1) in arrangements derived from TL (OL, EP, and HI), as well as in the TL strains themselves, is then a common feature of the TL phylad.

Of the 45 restriction sites mapped within the 26-kb region, 30 sites are variable (this study; Aquadro et al. 1991). Among the TL derivatives, there are five restriction-site differences between OL and EP, four differences between OL and HI, and three differences between EP and HI (fig. 1). A single TL strain from this report has a restriction pattern that is identical to strain no. 22 from Aquadro et al. (1991).

The four chromosomes from this study were also phylogenetically analyzed, together with the rest of the TL strains from the original RSP data set. The resulting neighbor-joining tree is presented in figure 2A. The same tree topology was obtained with the maximum parsimony method (data not shown). The empirical tree in figure 2A agrees with the predictions we would make regarding relationships between TL and its derivatives on the basis of the inversion phylogeny (fig. 2B) and the estimated age of the TL phylad. Under such circumstances, the TL strains (with respect to RSP data) should exhibit a paraphyletic relationship with strains having inversion derivatives of TL (Neigel and Avise 1986). Thus, certain TL strains would be similar to one inversion derivative, say HI, while other TL strains would be more similar to EP or to OL. As figure 2A shows, this prediction is clearly borne out. Thus, the phylogenetic analysis of restriction map haplotypes provides an independent corroboration of evolutionary relationships among inversions in the TL phylad.

The DNA Sequence Data

We cloned the 5.6-kb HindIII/EcoRI fragment that contains the Amy1 region from all four arrangements (TL, OL, EP, and HI) and sequenced a total of 2,727 nucleotides (nt). The sequences generated include 716 nt in the 5' flanking region, the amylase coding region with two exons (177 and 1,308 nt long) and one intron (71 nt), and 455 nt in the 3' flanking region. The observed nucleotide sequence differences are presented in figure 3. Out of 2,727 nucleotides sequenced, 26 were polymorphic, excluding insertions and deletions. Of these 26 polymorphic sites, 10 are located within the combined flanking regions, 4 in the intron, and 12 in the coding region (fig. 3). This distribution pattern indicates that levels of polymorphism in the flanking and coding regions are about the same, despite the expectation of higher polymorphism in the flanking regions due to the absence of functional constraint. From the previous work by Brown et al. (1990) we know that the amylase protein sequence is not highly constrained, as indicated by the high frequency of nonsynonymous substitutions. This finding suggests that an unusual conservation of flanking sequences is responsible for the similar levels of polymorphism between them and the coding regions.

The percent sequence divergences among the four Amy1 genes are presented in table 1. Divergence between

![Fig. 2.-A. Neighbor-joining tree based on the RSP data. Taxa are labeled as in a previous report (Aquadro et al. 1991), with asterisks indicating strains used for cloning Amy1 genes in this study. Numbers refer to the bootstrap values as percentages for 100 replicates. The same tree topology was also obtained with the maximum parsimony method. B, Cytogenetic phylogeny of the Drosophila pseudoobscura gene arrangements examined in this study.](https://academic.oup.com/mbe/article-abstract/12/5/938/974555)
Interchromosomal Exchange of Genetic Information

Table 1
Pairwise Percent Sequence Divergence for the Coding (above the Diagonal) and Combined Flanking Regions (below the Diagonal)

<table>
<thead>
<tr>
<th></th>
<th>TL Amyl</th>
<th>EP Amyl</th>
<th>OL Amyl</th>
<th>HI Amyl</th>
</tr>
</thead>
<tbody>
<tr>
<td>TL Amyl</td>
<td>...</td>
<td>0.67</td>
<td>0.74</td>
<td>0.67</td>
</tr>
<tr>
<td>EP Amyl</td>
<td>0.51</td>
<td>...</td>
<td>0.20</td>
<td>0.00</td>
</tr>
<tr>
<td>OL Amyl</td>
<td>0.60</td>
<td>0.60</td>
<td>...</td>
<td>0.20</td>
</tr>
<tr>
<td>HI Amyl</td>
<td>0.51</td>
<td>0.00</td>
<td>0.60</td>
<td>...</td>
</tr>
</tbody>
</table>

NOTE.—The number of nucleotides compared is 1,482 in the coding region and 1,171 in the combined flanking regions (716 nt in the 5' flanking and 455 nt in the 3' flanking).

TL Amyl and its three derivatives in the coding region is about 0.70%, with approximately the same level (0.50%-0.60%) in the flanking regions. In contrast to the 2.9% sequence divergence between the TL and SC-ST phylads (Aquadro et al. 1991), the differences between arrangements within the TL phylad are of the order of 0.5%-0.7%. These data suggest that, while the inversion polymorphism is quite old (about 2 million years), some of the derived inversions have had a more recent origin. What is quite striking, however, is the sequence identity between Amyls from EP and HI. Note that these two genes are identical only in the 1,482 nt of the coding region (plus 71 nt in the intron) but also in the 1,171 nt of the combined flanking regions. Given the estimated age of 1.7 million years for the TL phylad (Aquadro et al. 1991), a fair, or at least detectable, level of divergence should have been observed. With respect to the RSP data this absence of divergence is certainly unexpected, as is obvious from the phylogenetic tree in figure 2.4.

Nonetheless, we can rule out the possibility that the observed sequence identity might be an artifact, due to stock contamination: First, the salivary slides of these two strains were diagnosed as EP and HI, respectively, on the basis of the differences in the banding patterns in the inverted region. Second, the genomic DNA used for cloning was diagnosed as bona fide EP and HI on the basis of their diagnostic restriction patterns. However, after finding that the EP and HI Amyls had identical sequences, we were so surprised that we rechecked the salivary chromosomes and genomic DNA, only to find once again that they were indeed EP and HI. As a final assurance, we started a new HI line from a single female. After diagnosing both the salivaries and the isolated DNA as HI, we then cloned and sequenced HI Amyl a second time, with exactly the same result. Thus, the sequence identity between EP and HI Amyls is real. How, then, can these seemingly contradictory results between the RSP and DNA sequence data be explained?

The suppression of recombination in inversion heterozygotes should lead to the formation of semi-isolated, coadapted gene complexes which are associated with the different gene arrangements (Dobzhansky 1970). Because the available evidence supports a monophyly of the third chromosome inversions in Drosophila pseudoobscura (Dobzhansky and Epling 1944; Aquadro et al. 1991), there are only two reasonable explanations for the observed sequence identity between our EP and HI strains: double crossover and gene conversion. Of the two mechanisms, gene conversion is much more likely to be responsible for the observed interchromosomal exchange for the following reasons. First, the size of the identical sequences (2.7 kb) is just a fraction of a much larger 26-kb region, analyzed by RSP, which did differ between the two arrangements (fig. 1). A double crossover event involving less than 3 kb is highly unlikely. In contrast, domains of gene conversion vary from a few base pairs to several kilobases, affecting loci whether or not they are on the same chromosome (Dover 1993). Second, in a study where it was possible to quantify the levels of interchromosomal exchange (Chovnick 1973), almost all of the exchanges were due to gene conversion, while only 1.5% were caused by double crossovers. Third, in agreement with this finding, recent studies of the rp49 region in D. subobscura have revealed that gene conversion is responsible for the transfer of genetic information between inversions in this species (Rozas and Aguade 1993, 1994). These observations suggest that the observed sequence identity between EP and HI Amyls is caused by gene conversion.

Of the two arrangements, EP occurs at moderate frequencies from Mexico to Canada, while HI is rare but widespread in Mexico. An interchromosomal exchange in an EP/HI heterozygote (fig. 4), therefore, probably occurred in Mexico and involved a Mexican EP. The inversion frequencies at the Amecameca location, where the HI strain was collected, are shown in table 2. As mentioned previously, HI is rare (0.9%) and is found almost exclusively as a heterozygote with other arrangements, where of course, interchromosomal exchange could occur. From table 2, we see that the two most common arrangements in Amecameca are Cuernavaca (CU) and TL (58.5% and 31%, respectively), while the frequency of EP is about 8%. Under random mating, the expected frequency of EP/HI heterozygotes would be 0.14%. The proposed gene conversion between EP and HI would clearly be an infrequent event, but one that could occur in a large population like that at Amecameca.

It is important to note that the EP strain used in this study was collected in British Columbia, Canada.
Nevertheless, we postulate that the putative gene conversion event occurred in Mexico and involved HI and Mexican EP. The suggested sequence identity in EP strains from these widely separated locations (Mexico and Canada) is intriguing, although not entirely surprising. Detailed surveys of restriction site and nucleotide variation in the Adh region of D. pseudoobscura (68 and 99 strains, respectively) by Schaeffer and Miller (1992) and Abdel-Megeed (1992) found evidence for extensive gene flow, sufficient to keep North American populations homogeneous. In addition, a separate study showed an absence of geographic patterning in the distribution of mtDNA haplotypes, where identical haplotypes were found in populations from Canada, Utah, Colorado, and Mexico (Abdel-Megeed 1992). Migration alone, however, cannot account for the lack of at least one nucleotide substitution in a 2.7-kb region. The observed sequence identity between EP and HI, in conjunction with high gene flow, would also suggest an appreciable level of interchromosomal exchange as a way of homogenizing regions sequestered within inversions.

Traditionally it has been thought that inversions prevent homogenizing recombination. Genes within an inversion could thus plausibly be conceived to evolve as an integral complex that would eventually become fine-tuned to particular environments under the influence of selection. The results presented here demonstrate, however, that the isolation that exists between such inversions can be violated—at least in certain cases. Our findings are consistent with the idea that the mechanism involved in the exchange between HI and EP was gene conversion rather than double crossover, but this matter cannot be fully resolved at present due to the limited nature of the available data. It would be especially in-

Table 2
Percent Frequencies of Gene Arrangements in Adult Females Collected at Amecameca, Mexico

<table>
<thead>
<tr>
<th>TL</th>
<th>CU</th>
<th>EP</th>
<th>HI</th>
<th>Other</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>30.9</td>
<td>58.5</td>
<td>7.8</td>
<td>0.9</td>
<td>1.9</td>
<td>660</td>
</tr>
</tbody>
</table>

Note. — n is the number of chromosomes examined. The gene arrangements are Tree Line (TL), Cuernavaca (CU), Estes Park (EP), and Hidalgo (HI); “other” includes rare endemics.

formative to obtain additional sequences for EP, both from Mexico and from areas to the north along the West Coast, and for TL, in order to obtain a better idea of the variation found in the inversion type ancestral to HI and EP.

Sequence Availability

The original sequences reported here have been deposited in GenBank under accession numbers U20334, U20335, and U20337.

Acknowledgments

We thank Eugene McCarthy for editorial assistance, and J. C. Avise, J. Hamrick, J. McDonald, R. Meagher, and two anonymous reviewers for comments on the manuscript.

LITERATURE CITED


Fig. 4.—In situ hybridization of the Amy region in the Estes Park (EP) and Hidalgo (HI) inversion heterozygote.

THOMAS EICKBUSCH, reviewing editor

Received March 15, 1995
Accepted May 12, 1995