ESTUDI COMPARATIU DE L'ESTRUCTURA DEL GEN
Adh A VÀRIES ESPÈCIES DE *Drosophila*

Gemma Marfany i Nadal

Departament de Genètica. Facultat de Biologia.
Universitat de Barcelona.

CAPÍTOL III: "The *Drosophila subobscura Adh* genomic region contains valuable evolutionary markers"

Molecular Biology and Evolution (en premsa)

The _Drosophila subobscura Adh_ genomic region contains valuable evolutionary markers.

Gemma Marfany and Roser Gonzàlez-Duarte[*]

Departament de Genètica. Facultat de Biologia. Universitat de Barcelona.

ABSTRACT

We have sequenced 4-kb of the genomic region comprising the _Adh_ (Alcohol dehydrogenase) gene of _Drosophila subobscura_. In agreement with other species which belong to the same subgenus, two structural genes, _Adh_ and _Adh-dup_, are contained in this region. The main features of these two genes of _D.subobscura_ have been inferred from the sequence data and compared with the homologous region of _D.ambigua_ and _D.pseudoobscura_. _D.subobscura Adh_ and _Adh-dup_ differ from _D.ambigua_ at a corrected estimation of 10.1% and 12.5%, respectively, while from _D.pseudoobscura_ , they differ 9.5% and 8.1%, respectively. Our data suggest that _Adh_ and _Adh-dup_ are evolving independently, showing a species-specific pattern. Moreover, particular features of some regions of these genes make them valuable evolutionary hallmarks. For instance, replacement substitutions in the third exon of _Adh_ may indicate the branching of the _melanogaster-obscura_ groups whereas replacement substitutions in the third exon of the _Adh-dup_ could assess speciation within the _obscura_ group.

KEYWORDS: _Adh, Adh-dup, Drosophila subobscura,_ gene evolution, nucleotide substitution rate, phylogenetic relationships.

INTRODUCTION

Phylogenetic relationships in _Drosophila_ have been the subject of much research to illustrate evolutionary trends among species and to determine the dynamics of the different groups and subgroups. Genetic distances among species have been evaluated through the analysis of morphological, chromosomal and biochemical traits. A wealth of information has been produced and many phylogenetic relationships have thus been established.

The evolution of the _obscura_ group of _Drosophila_ has been thoroughly reviewed at morphological, cytological and biochemical levels (Buzzati-Traverso and Scossiroli, 1955; Lakovaara and Saura, 1982; Krimbas and Loukas, 1984; Steneimann et al., 1984; Loukas et al., 1986; Hernández et al., 1988) and electrophoretic comparisons have also been drawn for the different subgroups (Cabrera et al., 1983; Loukas et al., 1984; Loukas et al., 1986). Lately, relationships between species in the _obscura_ subgroup have been examined through the

analysis of mitochondrial DNA (Latorre et al., 1988; González et al., 1990) and scDNA divergence (Goddard et al., 1990).

Comparative analysis at the genomic DNA level of one species of the obscura subgroup, D.pseudoobscura, has provided valuable information concerning evolutionary rates of individual genes (Schaeffer and Aquadro, 1987) as well as on the relationships among the members of a multigene family (Brown et al., 1990).

The aim of the present study is to establish the evolutionary pattern of two genes contained in the Adh genomic region of D.subobscura and to make inter-specific comparisons in the obscura subgroup: D.pseudoobscura (Schaeffer and Aquadro, 1987), and D.ambigua (Marfany and Gonzàlez-Duarte, 1991).

MATERIALS AND METHODS

Fly Stock: The source of DNA was a stock of flies of D.subobscura strain H27, with standard chromosomal arrangements, provided by R. de Frutos, Dept. of Genetics, University of Valencia, Spain.

Preparation of Genomic DNA and Construction of Libraries: Total genomic DNA was isolated using the guanidine isothyocyanate method initially described for RNA extraction (Chirgwin et al., 1989), with minor modifications. High molecular weight total DNA was partially digested with MboI, then fractionated by sucrose gradient to obtain DNA fragments of 15-20 kb. The library was constructed by cloning this DNA into the BamHI site of an EMBL4 phage vector. Between 150,000 and 200,000 recombinant plaques were screened for sequences homologous to the D.melanogaster Adh gene. A 2.7 kb HindIII-EcoRI restriction fragment of sAC1 (Goldberg, 1980) containing the complete Adh gene and the immediately adjacent 3' region was used as a probe. It was labeled either by nick-translation or by random-hexamer priming with $[\alpha-^{32}P]dCTP$ (NEN-DuPont) and hybridized to phage DNA on nitrocellulose filters (Hybond-C, Amersham) in 46% formamide at 42°C overnight in the presence of 10% Na dextran sulfate. Nonspecifically hybridized probe was removed with one wash in 2xSSC/0.1%SDS at room temperature for ten minutes, then two washes in 2xSSC/0.1%SDS at 65°C for ten minutes each, and finally two further washes in 1xSSC/0.1%SDS at 65°C for ten minutes each (1xSSC is 0.15M NaCl/0.015M sodium citrate, pH 7.5). After washing, filters were autoradiographed and the DNA from positive recombinant phages was isolated according to Maniatis et al. (1982, pp.76-85, 371-373).

Restriction Analysis of Positive Clones: Positive clones were characterized by restriction mapping with the following restriction enzymes: BamHI, EcoRI, HindIII, PvuII, PstI, SalI and XbaI (Boehringer Mannheim), using single and double digests (Maniatis et al.,1982, pp. 354-376). Adh was located on the restriction map by transferring digested and electrophoresed DNA to nitrocellulose membranes. Labeling of Adh probe, hybridization conditions, washes and autoradiography were performed as for plaque hybridization.

Nucleotide Sequence Analysis: Clones for sequencing were obtained by restriction digestion with suitable enzymes. These restriction fragments were subcloned into Bluescript +KS, +SK phasmids (Stratagene). Several series of nested deletions were generated using the method of Henikoff (1984) to sequence large regions. Sequencing was performed on either single or double stranded DNA using

suitable primers according to the dideoxy method of Sanger (Sanger et al., 1977), with [α-^{35}S)dATP (NEN-DuPont) and modified T7 phage polymerase (Sequenase USB, or Sequencing Kit from Pharmacia-LKB). After electrophoresis on TBE buffer, sequencing gels were fixed, dried and autoradiographed for 72 hours. Each nucleotide was sequenced at least 3 times. Sequences were read and aligned with the 4.185 kb. sequence of D.ambigua (Marfany and Gonzàlez-Duarte, 1991) and the 3.535-kb sequence of D.pseudoobscura (Schaeffer and Aquadro, 1987), as they are the only two members of the obscura group whose Adh gene has been sequenced. Alignments were determined by the Sequence Analysis Software Package of the Genetics Computer Group of the University of Wisconsin (Devereux et al., 1984). Unrooted trees were performed using the FITCH program of the PHYLIP Package (J. Felsenstein, University of Washington) based on the Least-Squares Distance Method (Fitch and Margoliash, 1967). This program was chosen because it does not assume synchronous growth of branches, thus allowing differences in evolutionary rates between different species. In addition, the program DNABOOT of the PHYLIP Package (J. Felsenstein, University of Washington) was used when calculating bootstraps for the trees based on total coding sequence data. This program does not allow one to work with only silent or replacement positions.

RESULTS

Clone characteristics: Different positive clones were isolated after the screening of the library. Some of them carried the functional Adh gene whereas the remainder shared the conventional features of retrosequences (also called retroposons) bearing the Adh coding region (Marfany and Gonzàlez-Duarte, in prep.). Restriction analysis of the functional Adh clones revealed polymorphic restriction sites possibly due to insertion/deletion events (data not shown). The cytological location of sequences homologous to Adh on polytene chromosomes in D.subobscura has been determined previously (Visa et al., 1991).

Sequence comparison of Adh coding regions: A total of 3980 bp of the genomic region containing Adh and Adh-dup was sequenced in D.subobscura (Fig.1) and then aligned with the homologous region of D.ambigua (Marfany and Gonzàlez-Duarte, 1991) and D.pseudoobscura (Schaeffer and Aquadro, 1987).

The position of the D.subobscura Adh adult and larval promoters, the proposed leaders, the transcriptional start sites and the polyadenylation signal were determined by homology with the sequences compared. Length differences were observed in the adult and larval leaders of the three species. D.subobscura leaders resembled D.ambigua not only in length but also in their position, although coding sequence similarity was greater between D.subobscura and D.pseudoobscura (Table 1). However, the overall structural comparison of the leaders of the three species showed that some of the regions were well conserved while others shared only reduced similarity. Variation in length was observed in the intron regions: for instance, some insertion/ deletion events were clearly detected in the adult intron, but functional sequences, such as consensus donor

68

<pre>
                                                                 *   *
                                                      CTGCAGTGTGTGTCTAT   -321

         *   *   *   *   *   *   *
AGGGGAGGGTGGCGAGCTAAAGGAGCGTGTTGGCGTGCGAGTGTTTGTTGATGTAGTATTGACAGATGAAGAAAAAATGT   -241

         *   *   *   *   *   *   *
AAAATTTCACAAATAACCGCTAAAGTGCAGATGTAGTACTGAGTGCCGGGTATAAAAGTTGTGACGCGTAAGAAGCGTCT   -161

         *   *   *   *   *   *   *
CACACGTCCCTTCTCGTTGGTTTTGAATATTTTGTATTTGCATTTTACATTATTACAAATTATTGCAGCTCCGGCGTCGC   -81

         *   *   *   *   *   *   *
GTTTGCGTTTGCGTATATAGAGGCGTTGATAGAGGCGCGGCAGAAGTGTATTTAAGGCACCGCACATTGCGAGGACAGCG   -1

         *   *   *   *   *   *   *
ATTATTGTCTCAGTACAGTTGCCAGGTGCAGTAGGCGGACCCCAGCATTTCTTCAAATCTACAGTTTCGCTCAAGTAAGT   80

         *   *   *   *   *   *   *
AAAGTAACTGTATGCCATGTACAGTCGACAGGCATATCACGCCCGACACCACTGAGAGAGGATTCGAGCACGGGAAGGTA   160

         *   *   *   *   *   *   *
AAGTTAATGTTCGATTTTCGATTGCAGAAGCTTCGAGCCTGACTTTGACAAAATACAAAATACTCCAAGTTTCAGTGAAT   240

         *   *   *   *   *   *   *
TTAACTGCAATAAACTACCCATCAACCCGACCTTGAACGGTAAAAATAGTACATATCAACAATCGTATGACGTATTCCCT   320

         *   *   *   *   *   *   *
CAGAGCAGTTTATAAAAATAATTCTCTCGATTTGGCGGACTAGGAAATCGTTCTGGCACTTGTCCAATTAATTTGTTTAC   400

         *   *   *   *   *   *   *
ACTTTTTCCTCCACCCCTGCTCAGAATATGTATGCCCTCACTTTCTGTGTGGTCGTATCAGGGAGCGCGCGTGTCGACTC   480

         *   *   *   *   *   *   *
TGATAGATCCCCAGACGCGCAGTATTTTTCCTCAAGAACCTGAACTCTAAACATAGACATAATTTACTACACTCGCAAAC   560

         *   *   *   *   *   *   *
ATATACAGATGTAAGAGTAGTGCCACTGATTAGGCACACGTATTAAAATACATTTATCGGCATAAAACCAAAACAAAGCG   640

         *   *   *   *   *   *   *
ATCCGAAACCGAGACGCTGTGAAGACGCAATCGAACGACACGTAATGCGAGAGATAAGAAAGCAAAAGCTTCCTTCACGC   720

         *   *   *   *   *   *   *
GAAATAAGCTTTTCGCGTGAAAGAGCTTTGCTTTGAAACGAAATAAATTCCCTATAAATAAGAGACTGAAACCAGCAGAA   800

                                                                      T C C
                                                                      G C
ATCTAACAAACCGTTGTTCCATCCTCCCCGATTTCCAGGTCAGGAACTACAAAAGCAAAGACTCAACATGTCACTTACA   880
                                                                   Met Ser LeuThr

   C C   C   C C   C   C   TA
   * GA   * C   * C C   *   * C   *C   C   * T   *
AACAAGAATGTTGTTTTCGTGGCTGGTCTGGGAGGTATTGGCCTGGACACCAGTCGGGAGTTGGTTAAGCGAGATCTGAA   960
AsnLysAsnVal Val PheValAlaGlyLeuGlyGly Ile GlyLeuAspThrSer Arg GluLeuValLysArg Asp LeuLy

                                                               C
         *   *   *   *   * C   *
GGTAAGAAAGAGGGAAATCTATTTCCATAGACTCTATGGAAAAAACTAATCCCAAATCCTCCCCTTATAGAATCTGGTCAT   1040
s   AsnLeuValIl
</pre>

```
                    C G T                                        A         T C
        *           C G T      *      G  *     C    *        *   G    *    T G*
CCTGGATCGCATTGACAATGCAGCGGCCATTGCCGAACTGAAGGCAATCAATCCCAAGGTGACCGTCACCTTCTACCCTT    1120
eLeuAspArgIleAspAsnAlaAlaAlaIleAlaGluLeuLysAlaIleAsnProLysValThrValThrPheTyrProT

        C           G C      T              G                 G G       A
        C      *G  G  CA *AGC  G*   G  G *          *         C G      G*      G  *
ATGATGTGACTGTACCTGTCGCAGAGACCACCAAACTCCTGAAGACCATCTTTGCCCAGATCAAGACCATCGATGTCCTG    1200
yrAspValThrValProValAlaGluThrThrLysLeuLeuLysThrIlePheAlaGlnIleLysThrIleAspValLeu

        C              G           C    G C              A G
        C     *  C    *G    T *     * G  C     *        *A G T   *            *
ATAAACGGTGCTGGCATCCTCGACGATCATCAGATTGAGCGTACTATTGCCGTTAACTACTCTGGCCTGGTCAACACCAC    1280
IleAsnGlyAlaGlyIleLeuAspAspHisGlnIleGluArgThrIleAlaValAsnTyrSerGlyLeuValAsnThrTh

                 C                                C                          C
     C    *       C*         *  .     *      C  *       *          *   C     *
CACAGCCATTCTGGACTTTTGGGACAAGCGCAAGGGCGGCCCAGGTGGCATCATTTGCAACATTGGCTCCGTTACCGGTT    1360
rThrAlaIleLeuAspPheTrpAspLysArgLysGlyGlyProGlyGlyIleIleCysAsnIleGlySerValThrGlyP

     C                    C  TC           C C  * G      T   TC
     C    *        *   C  *      *   C C  * G        *          *           *
TTAATGCCATCTACCAGGTGCCCGTTTACTCTGGCAGCAAGGCGGCGGTGGTTAACTTCACCAGCTCCCTGGCGGTAAGC    1440
heAsnAlaIleTyrGlnValProValTyrSerGlySerLysAlaAlaValValAsnPheThrSerSerLeuAla

                                                            G   T      T
     *        *         *          *        *         *      *         *       *
GCATCTCTCATCACATCTCTCTATTCGCAGAAACTAATTCTTAACTTATCCAAATCGTTTAGAAACTTGCACCCATCACT    1520
                                                            LysLeuAlaProIleThr

   T     T T          C       T      T        C
   T G  * C      *C    T  * T    *    G    *C  G *         *           *
GGAGTCACCGCATACACTGTGAATCCGGGCATCACCAAGACCACTCTGGTGCACAAATTCAACTCGTGGCTGGATGTGGA    1600
GlyValThrAlaTyrThrValAsnProGlyIleThrLysThrThrLeuValHisLysPheAsnSerTrpLeuAspValGl

   C T              C                                  C              G
   G AT *  C         *         *          *       *          *         *        *
GCCCAGAGTGGCGGAGAAGCTGCTGGAGCATCCCACCCAGACCTCTCAGCAGTGTGCCGAGAACTTTGTCAAGGCCATTG    1680
uProArgValAlaGluLysLeuLeuGluHisProThrGlnThrSerGlnGlnCysAlaGluAsnPheValLysAlaIleG

        C               G   TC    C CT     C               C
   *   C C  *           TC    *C         *C     *          C  *        *
AGCTGAACAAGAATGGTGCCATCTGGAAATTGGACTTGGGAACTCTGGAGTCCATCACATGGACCAAGCACTGGGATTCG    1760
luLeuAsnLysAsnGlyAlaIleTrpLysLeuAspLeuGlyThrLeuGluSerIleThrTrpThrLysHisTrpAspSer

     *        *         *          *        *         *      *         *       *
GGCATCTAAACGGGATACCCGCACCACAATGCATTCAATGGCTTTAAGCTTTTAGCTTCGTTTTTCCACACAATTGTTAC    1840
GlyIleEND

     *        *         *          *        *         *      *         *       *
GTATATATCTACATATGGCAATAAGGCTGATTTGATTCTCTTTAAATGGAACCCCGTTTTGAATATGATAATAAAAATTA    1920

     *        *         *          *        *         *      *         *       *
TATTTGAGAAATTTAAACACAAAGCAGATACGCAGTAGCTCTCTTTTAATTAAAAATAGATAAATAATGCCAGTGGCAGG    2000

     *        *         *          *        *         *      *         *       *
GGCACTGTATTCAGGCCAATAGCTCTATCGATTTCACACAACAAAACGTAACTTTAGTAATAGAAAAGAAGTCAAGAAAA    2080
```

```
                                                          G                                      C
           *         *    G    C         *         G  *         *         *         C  *
GCAGGAAAAATAATGTACGATCTGACGGGTAAGCATGTCTGCTATGTAGCTGACTGCGGTGGCATTGCACTGGAGACTAG       2160
              MetTyrAspLeuThrGlyLysHisValCysTyrValAlaAspCysGlyGlyIleAlaLeuGluThrSe


           *         *         *         *         *         *         *         *
CAAGGTTCTCATGACCAAGAATATAGCGGTGAGTGCGGTGTGTGGAGAGTGGAACAGAGATCTATGTTCTAGTTCCTAAA       2240
rLysValLeuMetThrLysAsnIleAla


           *         *         *         *         *         *         *         *
CAAGTCTCTCAGCTGCGGGACTGTCGAAACTAATTAAGTTAATGAATTTTGCACTCAATTTTTTCGCTGACTTTTTCTGT       2320


           *         *         *         *         *         *         *         *
TCGTGGCTACAATTAAGTTTAGTTTTAGAGTGATCTATTTTTCTTAAGGGAAATACTTTCGATTATGGATTATTGCTGCA       2400


           *         *         *         *         *    T         T         T  *    T    *
GAATAAAAAAATAGATACAAAGGAATGTACTCGTCTGTTGGTACCTTTTCTAGAAACTGGCAATCCTCCAGAGCGTGGAA       2480
                                                        LysLeuAlaIleLeuGlnSerValGlu


    CC                 A T A         G
     *                 A    A G G    G   *         *         *         *         *
AACCAACCGGCCATCGCTCGGCTTCAATCCATTAAGCACAGCACACAGATCTTCTTCTGGACCTTCGATGTGACCATGGC       2560
AsnGlnProAlaIleAlaArgLeuGlnSerIleLysHisSerThrGlnIlePhePheTrpThrPheAspValTheMetAl


    CG
    CG    A         A*         C  *         *    A         *         *G   T    *    G    C    G*
CCGACAGGAGATGAAGAAGTACTTCGATGAGGTCATGGTCCAGATGGACTACATCGATGTACTAATCAATGGAGCAACCC       2640
aArgGlnGluMetLysLysTyrPheAspGluValMetValGlnMetAspTyrIleAspValLeuIleAsnGlyAlaThrL


           *         C         *    G         C         *    G         *C         *
                         C G  *         C                         G
TGTGCGATGAGCGGAACATTGATGCCACCATCAATACAAATCTGACCGGAATGATGAACACCGTGGCCACTGTGCTGCCC       2720
euCysAspGluArgAsnIleAspAlaThrIleAsnThrAsnLeuThrGlyMetMetAsnThrValAlaThrValLeuPro


    T         T         T    G    C    T         C    C
    T    *  C         *    C    C  *    T    T*         *    G    G    T    C    C*   G    *
TACATGGACCGAAAGATGGGCGGATCGGGTGGATTGATCGTGAATGTCACCTCTGTCATAGGATTGGATCCATCGCCAGT       2800
TyrMetAspArgLysMetGlyGlySerGlyGlyLeuIleValAsnValThrSerValIleGlyLeuAspProSerProVa


    G                             C
    G    *         *         C    AC  *         *         *         *         *
CTTTTGTGCATACAGTGCCTCAAAGTTTGGTGTGATTGGGTTCACCAGAAGTCTAGCGGTGAGTCGAAGATCGTTACATT       2880
lPheCysAlaTyrSerAlaSerLysPheGlyValIleGlyPheThrArgSerLeuAla


           *         *         *         *    T    *    AT  *         *    G    *
GGCTTTTTGTACGCTAATAAGTATCTATTGTTTTATATAGGATCCCCTGTACTACACCCAAAATGGTGTGGCTGTCATGG       2960
                                   AspProLeuTyrTyrThrGlnAsnGlyValAlaValMetA


         T
T    T *    T    *         *         C         C         *    A         C    GT
CCGTCTGCTGTGCCCCACCAAAGTGTTTGTCGATCGGGAACTGAATGCCTTCCTGGAGTACGGTCAAACCTTTGCCGAT       3040
laValCysCysGlyProThrLysValPheValAspArgGluLeuAsnAlaPheLeuGluTyrGlyGlnThrPheAlaAsp


C         C              G                        A    C    G         G    C
C    G*TG G    *         G         T*         *T    A    *         *    C    G    C    *
CGCTTGCGTCGTGCACCCTGCCAATCGACTGCCGTCTGCGGCCAGAATATAGTAACTGCCATTGAAAGATCGGAAAATGG       3120
ArgLeuArgArgAlaProCysGlnSerThrAlaValCysGlyGlnAsnIleValThrAlaIleGluArgSerGluAsnGl
```

```
         C                      CC      C     G      C                                            G  G
         A   *    G    *        CC*     C     C      G    *          *            *      CA  C*
ACAGATTTGGATTGCCGACAAGGGCGGATTGGAGTTGGTGACCCTACACTGGTATTGGCATATGGCCGATCAGTTTTTAA            3200
yGlnIleTrpIleAlaAspLysGlyGlyLeuGluLeuValThrLeuHisTrpTyrTrpHisMetAlaAspGlnPheLeuS

A                      C    T    G         GTTT    T G---  C          G
          *       G    *         G    *    TGC  *  C      G CGGC  TATCGGAA  TA        *          *
GCTATATGCAGAGCACTGATGACGAGAATCAGGAACAGTTTGTATCAGGACAGCGATAAGGAGTATCGGAAATTATTTGT           3280
erTyrMetGlnSerThrAspAspGluAsnGlnGluGlnPheValSerGlvGlnArgEND

          *          *          *          *          *          *          *          *
AAATAAAGTATCAAATGTGACAGAAAAGGTGGTATGGCTAAAGATTTGATATACTTGTGGGTCATGTGAAAATTACACAT           3360

          *          *          *          *          *          *          *          *
ATGAAGGCGCGCCTGAAAGTATGCCTTGGATTGGATGGGTACTACGAGTGCTATTAATAGATTATGAACAGAGGAGGGTG           3440

          *          *          *          *          *          *          *          *
GCAGCGTTTGTTCTCTTTGGAATTTGAGGGAACTATTTATATTTCATATCGGTGGGACTCTTGGCAAATATGGCGCATTC           3520

          *          *          *          *          *          *          *          *
TACTACGTTTTTGGGATCTCCAGGAAAACACTCTCTCCAAACAATTTTCACTCGTTTCCTCTCTAAAAATTGTCTCTCTT           3600

          *          *          *          *          *          *          *          *
CCAACTTTTCATAAGTCAGTCGAAACCTTTTGACTTAGATTCC                                                3642
```

Fig.1.- Sequence of the Adh region of D.subobscura. Nucleotide position refers to the inferred adult transcription start site. The aminoacid sequence of Adh and Adh-dup genes is presented under the nucleotide sequence whereas gaps in this aminoacid sequence show intron positions. Nucleotide differences in Adh and Adh-dup coding sequences betwen D.subobscura and D.ambigua are shown immediately above the nucleotide sequence while the differences between D.subobscura and D.pseudoobscura coding sequences are shown in the second line above. Replaced aminoacids are also shown; underlined: replacements between D.subobscura and D.ambigua; wavy underlined: replacements between D.subobscura and D.pseudoobscura; double underlined: replacements occurred between D.subobscura with respect to the other two species; boxed, shared replaced aminoacids of the obscura species with respect D.mauritiana (melanogaster group species representative).

and acceptor splice sites appeared to be highly conserved.

The Adh coding sequence aligned perfectly among the species compared, as would have been observed with the melanogaster group species were it not for a deletion comprising six nucleotides corresponding to the third and fourth aminoacids of the protein, which constitutes a constant feature of the obscura group. The Adh coding region has accumulated 72 nucleotide differences since the divergence of D.subobscura and D.ambigua and 68 between D.subobscura and D.pseudoobscura. These substitutions were not randomly distributed in any codon position as it was evident that most of them affected the third nucleotide (D.subobscura vs. D.ambigua: $\chi^2=74.34$ and D.subobscura vs. D.pseudoobscura $\chi^2=67.19$; df=2 and P<0.00001 in each case). The ratio of transitions to transversions was 1:1, thus deviating significantly from the random ratio 1:2 (D.subobscura vs. D.ambigua $\chi^2=4.0$; df=1; P<0.05 and D.subobscura vs. D.pseudoobscura $\chi^2=15.56$; df=1; P<0.001), as expected for a coding sequence. Nucleotide differences in the D.subobscura sequence, affecting the third codon positions with four-fold degeneracy, led to an enrichment of the A+T content (Table 2). A test for the expected random ratio 1:1 of A+T/G+C of these differences showed a significant deviation (D.subobscura vs. D.ambigua $\chi^2=6.91$; df=1; P<0.01 and D.subobscura vs. D.pseudoobscura $\chi^2=6.43$; df=1; P<0.02. As coding bias towards G+C is clearly observed in the Adh gene of the Sophophora species analyzed to date (Starmer and Sullivan, 1990), D.subobscura Adh stands out among them for having a lower G+C content (Table 2).

Nucleotide substitutions that do not cause aminoacid replacements are synonymous or silent and thus they are expected to be more frequent than the rest. The number of potentially silent sites is different for each coding sequence, and for D.subobscura Adh it was 27.9% (that is 213.1 out of 765 nucleotide sites)(Table 1). But, according to our results, only 26.7% (57 nt) of these effectively silent sites differed between D.subobscura and D.ambigua and 28.2% (60 nt) between D.subobscura and D.pseudoobscura . In addition, these differences were randomly distributed throughout the three exons (D.subobscura vs. D.ambigua $\chi^2=1.2$; df=2; P>0.5; D.subobscura vs. D.pseudoobscura $\chi^2=1.76$; df=2; P>0.2).

When considering the potential replacement sites (551.9), 15 differences (2.7%) were detected between D.subobscura and D.ambigua, and 8 differences (1.4%) between D.subobscura and D.pseudoobscura. These replacement differences were also randomly distributed across the three exons whatever sequences were compared (D.subobscura vs. D.ambigua $\chi^2=0.46$; D.subobscura vs. D.pseudoobscura $\chi^2=0.40$; both cases df=2; P>0.7). Between D.subobscura and D.ambigua 12 aminoacid replacements, 10 of them conservative, were produced

TABLE 1.- Comparative evolutionary analysis of the Adh genomic region in D.subobscura, D.ambigua and D.pseudoobscura.

Region	D.subobscura	D.subobscura vs. D.ambigua			D.subobscura vs. D.pseudoobscura		
	Length (bp)	αNo. of changes	Percent differences	βCorrected percent divergence	γNo. of changes	Percent differences	βCorrected percent divergence
ADH							
5'noncoding	32.0	3	9.4	10.0	5	15.6	17.5
Adult leader	76.0	17	23.0	27.4	11	14.5	16.1
Adult intron	745.0	192	25.8	31.6	201	27.0	33.4
Adult and larval leader	64.0	18	28.1	35.2	12	18.7	21.6
Exon 1 silent	29.7	9	30.3	38.8	11	37.0	51.1
Replacement	63.3	1	1.6	1.6	1	1.6	1.6
Intron 1	68.0	13	19.1	22.1	21	30.9	39.8
Exon 2 silent	112.7	33	29.3	37.1	27	24.0	28.9
Replacement	292.3	9	3.1	3.2	5	1.7	1.7
Intron 2	68.0	25	36.8	50.5	22	32.4	42.3
Exon 3 silent	70.7	15	21.2	24.9	22	31.1	40.2
Replacement	196.3	5	2.5	2.6	2	1.0	1.0
INTERGENIC REGION	196.0	30	15.3	17.1	33	16.8	19.1
ADH-DUP							
5' noncoding	20.0	2	10.0	10.7	3	15.0	16.7
Leader	107.0	22	20.6	24.0	25	23.3	28.0
Exon 1 silent	24.0	3	12.5	13.7	2	8.3	8.8
Replacement	72.0	1	1.4	1.4	0	0.0	0.0
Intron 1	265.0	65	24.5	29.7	51	19.2	22.2
Exon 2 silent	108.3	41	37.9	52.7	22	20.3	23.7
Replacement	296.7	2	0.7	0.7	4	1.3	1.4
Intron 2	62.0	14	22.6	26.9	24	38.7	54.4
γExon 3 silent	86.7	33	38.1	53.1	24	28.2	35.4
Replacement	252.3	17	6.7	7.1	12	4.8	4.9
SUMMARY							
Noncoding	1703.0	401	23.5	28.3	408	24.0	28.9
Adh silent	213.1	57	26.7	33.1	60	28.2	35.3
Replacement	551.9	15	2.7	2.8	8	1.4	1.4
Total	765.0	72	9.4	10.1	68	8.9	9.5
Adh-dup silent	219.0	77	35.2	47.4	48	22.1	26.2
Replacement	621.0	20	3.2	3.3	16	2.6	2.6
Total	840.0	97	11.5	12.5	64	7.6	8.1

αThe effective number of silent sites for a gene depends on its aminoacid composition. It is calculated as the percentage of the potentially silent sites (possible substitutions that would not lead to an aminoacid replacement) out of the total coding nucleotides (Holmquist et al, 1972). Insertions and deletions are oversimplified to one single event. β Corrected percent divergence is estimated as d= 3/4 ln(1-4/3p), where p is the proportion of nucleotide sites that differ between two sequences (Jukes and Cantor,1969). γ Due to the variable length of the Adh-dup third exon we have used the figures for potentially silent and replacement sites of this table when comparing D.subobscura vs. D.ambigua, and 85.0 for silent and 251.0 for replacement sites when comparing D.subobscura vs. D.pseudoobscura.

by the 15 nucleotide differences whereas between D.subobscura and D.pseudoobscura, 7 aminoacid replacements, all of them conservative, were involved (Feng et al's scale, 1985). None of the essential aminoacids predicted for alcohol dehydrogenase function (four glycines and one aspartic acid, Benyajati et al,1981; Duester et al., 1986) has been replaced and the ADH protein of the three species shared extensive structural similarity, as could have been predicted from biochemical data (Hernández et al., 1988).

Replacement substitutions deviated from random distribution across the three exons when comparing the obscura species (Schaeffer and Aquadro, 1987; Marfany and Gonzàlez-Duarte, 1991) with D.mauritiana, a representative member of the melanogaster group. This deviation was not present in comparisons among other Drosophila species (Sullivan et al, 1990). Aminoacid replacements between D.mauritiana and any obscura subgroup species are boxed in Fig.1. Most replacements were clustered in the third Adh exon.

Analysis of non-coding regions defined some conserved sequence motifs. Whether or not they play a role in the regulation of the expression of this gene remains to be tested.

Adh-dup (3'ORF) sequence analysis: Downstream from Adh and adjacent to it, an open reading frame sequence was detected, the Adh-dup, already described in some Sophophora species (Schaeffer and Aquadro, 1987; Cohn and Moore, 1988: Marfany and González-Duarte, 1991). It shares considerable structural similarity with Adh. The main features of this gene were strongly conserved in all the species analyzed including D.subobscura. It contained three exons and had all the transcriptional and translational regulation signals of an expressed gene: a CCAAT and a TATA box, splicing consensus sequences and a polyadenylation signal, all of them located in the expected positions (Fig.1) The sequence evolution of this potential gene was analyzed in a similar way to Adh.

Comparison of the Adh-dup of D.subobscura with D.ambigua, and D.pseudoobscura showed considerable similarity in the first two exons, but in the third there were considerable length differences due to a different location of the STOP codon, which would account for the different number of aminoacids in the protein: 279 for D.subobscura, 281 for D.ambigua, and 278 for D.pseudoobscura (273 aminoacids for D.mauritiana and its sibling species). The C-terminal region of the 3'ORF presented increased variability which was reflected in the number of coding nucleotides, in the aminoacid sequence and in the variation of the STOP codon position.

The pattern of nucleotide substitutions resembled that of the Adh gene. The putative coding region had accumulated a total of 97 differences since the

TABLE 2.- Comparative analysis of the third codon positions with four-fold degeneracy for *Adh* and *Adh-dup* sequences.

A)

Adh	A	T	C	G	Total	%G+C
D.subobscura vs. *D.ambigua*(α)	12	20	8	7	47	31.9
D.subobscura vs. *D.pseudoobscura*(β)	11	14	5	5	35	28.6
Adh-dup						
D.subobscura vs. *D.ambigua*(α)	20	10	13	5	48	37.5
D.subobscura vs. *D.pseudoobscura*(β)	16	7	3	1	27	14.8

B)

Adh	A	T	C	G	Total	%G+C
D.subobscura	15	29	55	35	134	67.2
D.ambigua(α)	04	16	68	47	135	85.2
D.pseudoobscura(β)	05	25	70	37	137	78.1
D.mauritiana(γ)	08	19	76	35	138	80.4
Adh-dup						
D.subobscura	29	19	48	30	126	61.9
D.ambigua(α)	12	13	50	53	128	80.5
D.pseudoobscura(β)	13	17	57	43	130	76.9
D.mauritiana(γ)	22	19	34	39	114	64.0

A) Differences in third codon positions with four-fold degeneracy between *D.subobscura* and *D.ambigua* or *D.pseudoobscura* sequences. Figures are the number of times a change to this nucleotide has occurred in *D.subobscura*. **B)** Base composition of third codon positions with four-fold degeneracy in *D.subobscura*, *D.ambigua*, *D.pseudoobscura* and *D.mauritiana*. Sequence data from (α)Marfany and González-Duarte, 1991; (β)Schaeffer and Aquadro, 1987; (γ)Cohn and Moore, 1988.

divergence of D.subobscura and D.ambigua, and 64 between D.subobscura and D.pseudoobscura. These differences were not randomly distributed among the three positions of a codon, but they tended to accumulate significantly in the third position (D.subobscura vs. D.ambigua χ^2=73.84 and D.subobscura vs. D.pseudoobscura χ^2=36.55; each case df=2; P<0.0001). The ratio of transitions to transversions was close to 1:1 as was found when analyzing the Adh gene, significantly deviating from the random ratio of 1:2 (D.subobscura vs. D.ambigua χ^2=18.02; and D.subobscura vs. D.pseudoobscura χ^2=17.37; both cases df=1; P<0.0001). Nucleotide differences of D.subobscura Adh-dup which affect the third codon position with four-fold degeneracy increased the A+T content of the coding sequence so that this gene showed the lowest codon bias with respect to that of other species (Table 2).

Concerning this gene in D.subobscura (Table 1), 26.1% (219.0 nt) of the 840 coding nucleotides were effectively silent. Due to the variable length of the third exon, the number of coding nucleotides varied with species. So, when comparing D.subobscura to D.pseudoobscura. the number of effectively silent sites was 217.3 out of 837 coding nucleotides (26.0%). The number of effectively silent site differences was higher for D.subobscura vs. D.ambigua (77 nt, 35.2% of the total effectively silent sites) than for D.subobscura vs. D.pseudoobscura. (48 nt, 22.1%). Silent substitutions did not significantly deviate from random distribution among the three exons, either when comparing D.subobscura vs. D.ambigua (χ^2=3.93) or when comparing D.subobscura vs. D.pseudoobscura (χ^2=3.65; both cases df=2; P>0.1). Conversely, replacement substitutions clearly showed a significant deviation whatever species was compared (D.subobscura vs. D.ambigua χ^2=16.41; df=2; P<0.001; D.subobscura vs. D.pseudoobscura χ^2=8.31; df=2; P<0.02). Particularly, the first and second exons seemed to present a considerable constraint for replacement substitutions while the third exon showed a great accumulation.

Alignment of the protein sequences of the Adh-dup showed 15 replaced aminoacids between D.subobscura and D.ambigua and 13 between D.subobscura and D.pseudoobscura. Most of these replacements were conservative.

In contrast to the results for the Adh sequence, the Adh-dup aminoacid replacements between the melanogaster group representative (D.mauritiana) and any of the obscura species (Fig.1), were distributed randomly across the three exons (χ^2=0.02; df=2; P>0.99).

DISCUSSION

Molecular data are needed to determine evolutionary pathways and phylogenetic relationships among Drosophila species. Our aim has been to obtain more information on the genome evolution of three species of the obscura group: D.subobscura, D.ambigua and D.pseudoobscura. To this end, a 4-kb region containing the Adh gene and the Adhdup sequence has been analyzed in detail. A very recent report cited studies which demonstrate that Adh-dup is transcribed (Kreitman and Hudson, 1991). Furthermore, its great sequence conservation in the Sophophora species, where it was originally described (Schaeffer and Aquadro, 1987; Cohn and Moore, 1988; Marfany and Gonzàlez-Duarte, 1991), together with its many silent nucleotide substitutions, suggest that the Adh-dup could be a functional gene.

Nucleotide substitutions in exon and intron sites among either the Adh gene or the Adh-dup of D.subobscura, D.ambigua and D.pseudoobscura show the expected pattern. Coding regions considered as a whole are more conserved than non-coding regions, although when silent substitutions are considered, they are found to have occurred with a higher frequency than substitutions in non-coding regions (Table 1). These results should to be viewed cautiously as deletions/inversions in non-coding regions are difficult to evaluate and could introduce undetectable errors. When considering the Adh coding sequence of D.subobscura and D.ambigua, these two species appear to have diverged more than D.subobscura and D.pseudoobscura. This trend is even more noticeable when analyzing the Adh-dup sequences. Nevertheless, in non-coding regions the reverse seems to be true and, as already mentioned, insertion/deletion events could introduce errors.

The accumulation of nucleotide substitutions in the third codon position of Adh and Adh-dup and the deviated ratio of transitions to transversions is consistent with the high frequency of silent substitutions. Substitutions in the third position of a codon, particularly transitions, are frequently silent. If we assume that transversions would occur twice as often as the transitions and that selection removes most aminoacid replacements, then the final frequency of transitions will be greater than that of transversions

Nucleotide differences affecting the third codon position in D.subobscura (data not shown) led to an enrichment of A+T base composition. This trend is clearly detected when analysing the third codon position with four-fold degeneracy (Table 2). A recent analysis on third codon positions has been reported for all the Drosophila Adh sequences determined to date (Sullivan et al, 1989): the G+C

value for all <u>Sophophora</u> species is around 80%. <u>D.subobscura</u>, in contrast, showed a decreased G+C value of 68.7%, which is much closer to the distantly related <u>repleta</u> group than to its close relatives in its own group. That this also happens in other genes is supported by our data with the <u>Adh</u>-<u>dup</u> sequence, and so it may be a reasonably common feature of the <u>D.subobscura</u> genome. Nevertheless, it has been proposed that this tendency would not be favoured by selection, as synonymous substitutions giving rise to A and T seem to be selected against (Ticher and Graur, 1989). Whether this peculiarity in <u>D.subobscura</u> is a consequence of convergent evolution, genetic drift, or shifted codon usage due to different tRNA availability, remains to be elucidated.

Analysis of the frequency, type and distribution of substitutions shows that <u>Adh</u> and <u>Adh</u>-<u>dup</u> are evolving with different patterns, in agreement with other reports on several genes (Martínez-Cruzado et al., 1988; Sharp and Li, 1989). All types of substitutions in the <u>Adh</u> coding sequence, as well as silent nucleotide substitutions in the <u>Adh</u>-<u>dup</u>, were randomly distributed among the three exons, whatever sequence was compared. Conversely, replacement substitutions in the <u>Adh</u>-<u>dup</u> sequence showed a strong deviation from random distribution and most of them were clustered in the third exon. When comparing <u>D.mauritiana</u> (or any <u>melanogaster</u> group species) with the <u>obscura</u> species, non-random replacement substitutions are mostly located in the third exon of the <u>Adh</u> gene whereas in <u>Adh</u>-<u>dup</u> all types of substitutions are randomly distributed (Marfany and Gonzàlez-Duarte, 1991). This result is also supported at the protein level as revealed by cross-reactivity assays with one monoclonal antibody specific for the Cterminal domain of <u>D.melanogaster</u> ADH: positive reaction is observed with all the <u>melanogaster</u> group species while ADH from the <u>obscura</u> species is not detectable (Fibla and Gonzàlez-Duarte, in prep.).

All these results taken together suggest that selective constraints differ in these two genes, which are located in the same genomic region. Furthermore, these selective constraints have varied over time to produce a specific evolutionary pattern, as deduced by the differences in the distribution and type of nucleotide substitutions among the species. The third exon of <u>Adh</u> is the region of the two genes that has non-randomly accumulated the most differences during the divergence of the <u>melanogaster</u> and <u>obscura</u> groups, whereas within the <u>obscura</u> group radiation, the third exon of the <u>Adh</u>-<u>dup</u> stands out as being the most affected. Some attempts have been made to explain differences in the frequency and distribution of substitutions (Perler et al., 1980; Gillespie, 1984; Palumbi, 1989) and to illustrate the fact that a particular sequence can undergo variable substitution rates during evolution. Indeed, some genomic regions could

accumulate more replacement substitutions than others at specific periods in the course of speciation, but the rate of aminoacid replacement must have a limit to preserve functionality. If this does not happen, new functions would appear, as is the case when gene duplication occurs. These sequences can be particularly useful to evaluate evolutionary relationships in definite periods of time and constitute outstanding candidates for evolutionary hallmarks (Perler at al., 1980). In a similar way, replacement substitutions in the third exon of Adh could constitute a satisfactory marker for the melanogaster-obscura radiation, while these replacements in the third exon of Adh-dup could be used as a marker within the obscura group speciation.

Adh-dup has great structural similarity to Adh (Table 2), as has already been observed. This similarity is reflected at the nucleotide level, 48.2% identity, significantly deviating from the expected random value of 25% (χ^2= 287.1; df= 1; P<0.0001) and at the aminoacid level, with 36.6% of identities vs. a random value of 6.1% (χ^2= 412.7; df= 1; P<0.0001). As more data are available, it becomes increasingly feasible that both genes had diverged from an ancient precursor. Estimates for this duplication event give values around 130 MY, assuming a proposed divergence rate of 0.4% base substitution per million years in the obscura group (Caccone and Powell, 1990). This estimate is only an approximation, but it makes it plausible that Adh-dup was present in many, if not all, Drosophila radiations (Schaeffer and Aquadro, 1987; Albalat and Gonzàlez-Duarte, in prep.). Gene duplication events are not rare in the evolution of Drosophila. They involve not only the Adh region (Oakeshott et al., 1982; Batterham et al., 1984; Fisher and Maniatis, 1985; Atkinson et al., 1988) but also other genes (Gemmill et al., 1985; Levy et al., 1985; Bewley et al., 1989; Takano et al., 1989; Brown et al., 1990) and yield a wealth of new raw material for selection or neutral drift.

In summary, the essential features of the D.subobscura Adh region are: a) a lower nucleotide substitution rate for Adh-dup than for Adh in the comparison of D.subobscura with D.pseudoobscura in clear contrast to what is observed in other species comparisons, b) a higher percentage of A+T content in the third-codon position, which points to Adh and Adh-dup as being less codon biased than those of other species and, c) a higher similarity of the Adh and Adh-dup coding sequences with D.pseudoobscura while non coding regions are more similar to D.ambigua. These differences should be viewed cautiously as they may not be statistically significant and evolutionary trees based on them are thus open to question (Fig.2).

A)

TOTAL ADH

D.mauritiana ———————————————— |88| ┌——— D.ambigua
                                        ├——— D.pseudoobscura
                                        └——— D.subobscura

TOTAL ADH-DUP

D.mauritiana ————————————————— ┌——— D.ambigua
                                |65| ┌——— D.pseudoobscura
                                     └——— D.subobscura

1% Divergence
├──┤

B)

REPLAC. ADH-DUP

D.mauritiana ——— ┌— D.pseudoobscura
                 └┌— D.subobscura
                  └— D.ambigua
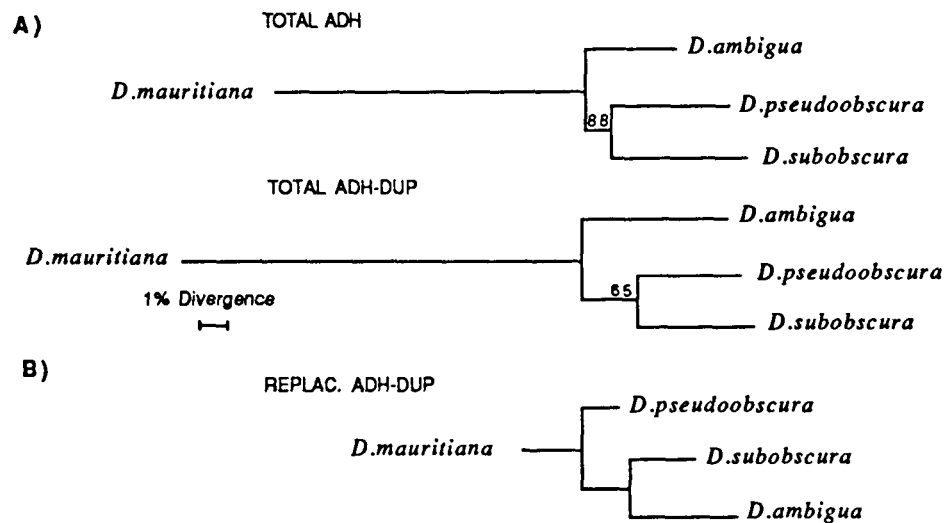
Fig.2- Phylogenetic unrooted trees were established with the Fitch-Margoliash method, assuming D.mauritiana as an outgroup species. It was chosen as a representative member of the melanogaster subgroup because more sequence data 3' of the Adh gene were available when our results were analyzed. (A) Distance matrices were constructed considering total, silent or replacement substitutions among all the species for Adh or the Adh-dup coding sequences, but only those obtained with total substitutions are depicted. Numbers indicate the bootstraps obtained for each branching (100 trees) considering the whole coding sequence of the genes. (B) Tree based on total replacement substitutions of the Adh-dup.

Our data supply more information for the phylogenetic location of D.ambigua, a paleoarctic species with a controversial position. The trees based on total substitutions locate D.subobscura nearer to D.pseudoobscura than to D.ambigua, which is in accordance with previous reports (Goddard et al., 1990). Indeed, bootstrap values seem to support a closer relationship for D.pseudoobscura and D.subobscura when using the Adh coding sequence data, although this is not that obvious when the Adh-dup coding sequence is considered. On the other hand, the use of statistically significant inter-specific differences overcomes the ambiguities generated in overall sequence comparisons. Replacement substitutions in the Adh-dup gene are not randomly distributed along the coding sequence and that is why they could be particularly valuable to establish phylogenetic relationships in our species. When these substitutions are considered, D.subobscura appears nearer to D.ambigua than to D.pseudoobscura (Fig.2) in agreement with the results obtained with mtDNA (González et al., 1990). At the same time, D.ambigua appears to have diverged faster within the obscura subgroup and this would explain the position of D.ambigua obtained by

scDNA divergence analysis (Goddard et al., 1990). Our contribution with the <u>Adh-dup</u> replacement substitutions as an evolutionary marker illustrates that a rapidly diverging species, such as <u>D.ambigua,</u> could distort its phylogeny, making it appear more distantly related to other species than it really is.

The molecular analysis of the <u>Adh</u> region of <u>D.subobscura</u> has supplied relevant information on the evolutionary features of this species. From our data, it can be inferred that: i) the <u>Adh</u> and the <u>Adh</u>-<u>dup</u> genes show different selective constraints; ii) their evolution is species-specific and iii) a particular gene, or even a specific coding region, may be under variable evolutionary constraints, yielding different nucleotide substitution rates and different distribution of these substitutions. These sequences are excellent genetic markers for evolutionary studies.

ACKNOWLEDGEMENTS

LITERATURE CITED

ATKINSON, P., W. MILLS, L. E. STARMER, and D. T. SULLIVAN. 1988. Structure and evolution of the <u>Adh</u> genes of <u>Drosophila mojavensis</u>. Genetics 120:713-723.

BATTERHAM, P., G. K. CHAMBERS, W. T. STARMER, and D.T. SULLIVAN. 1984. Origin and expression of an alcohol dehydrogenase duplication in the genus <u>Drosophila</u>. Evolution 38:644-657.

BENYAJATI, C., A. PLACE, D. POWERS, and W. SOFER. 1981. Alcohol dehydrogenase gene of <u>D.melanogaster</u>: relationship to intervening sequences to functional domains in the protein. Proc. Natl. Acad. Sci. USA 78:2717-2721.

BEWLEY, G. C., J. L. COOK, S. KUSAKABE, T. MUKAI, D. L. RIGBY, and G. K. CHAMBERS. 1989. Sequence, structure and evolution of the gene coding for sn-glycerol-3-phosphate dehydrogenase in <u>Drosophila melanogaster</u>. Nucleic Acids Res. 17:8553-8567.

BROWN, C. J., C. F. AQUADRO, and W. W. ANDERSON. 1990. DNA sequence evolution of the Amylase multigene family in <u>Drosophila pseudoobscura</u>. Genetics 126:131-138.

BUZZATI-TRAVERSO, A. A., and R. E. SCOSSIROLI. 1955. The <u>obscura</u> group in the genus <u>Drosophila</u>. Adv. Genet. 7:47-92.

CABRERA V. M., A. M. GONZALEZ, J. M. LARRUGA, and A. GULLON. 1983. Genetic distance and evolutionary relationships in the <u>Drosophila obscura</u> group. Evolution 37: 675-689.

CACCONE, A., and J. R. POWELL. 1990. Extreme rates and heterogeneity in insect DNA evolution. J. Mol. Evol. 30:273-280.

CHIRGWIN, J. M., A. E. PRZYBYLA, R. J. MACDONALD, and W. J. RUTTER. 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry 18:5294-5299.

COHN, V. , M. THOMPSON, and G. MOORE. 1984. Nucleotide sequence comparison of the Adh gene in three Drosophilids. J. Mol. Evol. 20:31-37.

DEVEREUX, J. P. HAEBERLI, and O. SMITHIES. 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 12:387-395.

DUESTER, G., H. JORNVALL, and G. W. HATFIELD. 1986. Intron-dependent evolution of the nucleotide-binding domains within alcohol dehydrogenase and related enzymes. Nucleic Acids Res. 14:1931-1941.

FENG, D. F., M. S. JOHNSON, and R. F. DOOLITTLE. 1985. Aligning amino acid sequences: comparison of commonly used methods. J. Mol. Evol. 21:112-125.

FISHER, J. A., and T. MANIATIS. 1985. Structure and transcription of the Drosophila mulleri alcohol dehydrogenase genes. Nucleic Acids Res. 13:6899-6917.

FISHER, J. A., and T. MANIATIS. 1988. Drosophila Adh: a promoter element expands the tissue specificity of an enhancer. Cell 53:451-461.

FITCH, W. M. and E. MARGOLIASH. 1967 Construction of phylogenetic trees. Science 155:279-284

GEMMILL, R, M., J. N. LEVY and W. W. DOANE. 1985. Molecular cloning of a-amylase genes from Drosophila melanogaster. I. Clone isolation by use of a mouse probe. Genetics 110: 299-312.

GILLESPIE, J. H. 1984. The molecular clock may be an episodic clock. Proc. Natl. Acad. Sci. USA 81:8009-8013.

GODDARD, K., A. CACCONE and J. R. POWELL. 1990. Evolutionary implications of DNA divergence in the Drosophila obscura group. Evolution 44:1656-1670.

GOLDBERG, D. A. 1980. Isolation and partial characterization of the Drosophila alcohol dehydrogenase gene. Proc. Natl. Acad. Sci. USA 77:5794-5798.

GONZALEZ, A. M. , M. HERNANDEZ, A. VOLZ, J. PESTANO, J. M. LARRUGA, D. SPERLICH, and V. M. CABRERA. 1990. Mitochondrial DNA evolution in the obscura species subgroup of Drosophila. 1990. J. Mol. Evol. 31:122-131.

HENIKOFF, S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene 28:351-359.

HERNANDEZ, J. J., LL. VILAGELIU, and R. GONZALEZ-DUARTE. 1988. Functional and biochemical features of alcohol dehydrogenase in four species of the obscura group of Drosophila. Genetica 77:15-24.

HOLMQUIST, R., C. CANTOR, and T. JUKES. 1972. Improved procedures for comparing homologous sequences in molecules of protein and nucleic acids. J. Mol. Biol. 64:145-161.

HOPP, T. P., and K. R. WOODS. 1981. Prediction of protein antigenic determinants from amino acid sequences. Proc. Natl. Acad. Sci. USA 78:3824-3828.

JUKES, T. H., and C. R. CANTOR. 1969. Evolution of protein molecules. Pp 21-132 in H. N. MUNRO ed. Mammalian protein metabolism, vol 3. Academic Press, New York.

KRIMBAS, C. B., and M. LOUKAS. 1984. Evolution of the obscura group Drosophila species. I. Salivary chromosomes and quantitative characters in D.subobscura and two closely related species. Heredity 53:469-482.

LAKOVAARA, S., and A. SAURA. 1982. Evolution and speciation in the Drosophila obscura group. Pp 2-9 in M. ASHBURNER, H. L. CARSON, and J. N. SU (eds). The genetics and biology of Drosophila, vol 3. Academic Press, London.

LATORRE, A., A. MOYA and F. J. AYALA. 1988. Mitochondrial DNA evolution in the Drosophila obscura group. Mol. Biol. Evol. 5:717-728.

LOUKAS, M., C. B. KRIMBAS, and Y. VERGINI. 1984 Evolution of the obscura group Drosophila species. II. Phylogeny of ten species based on electrophoretic data. Heredity 53:483-493.

LOUKAS, M., C. DELIDAKIS, F. C. KAFATOS. 1986. Genomic blot hybridization as a tool of phylogenetic analysis: evolutionary divergence in the genus Drosophila. J. Mol. Evol., 24:174-188.

LEVY, J. N., R. M. GEMMILL, and W. W. DOANE. 1985. Molecular cloning of a-amylase genes from Drosophila melanogaster. II. Clone organization and verification. Genetics 110:313-324.

MANIATIS, T., E. F. FRITSCH, and J. SAMBROOK. 1982. Molecular cloning: A laboratory manual, 1st edition. Cold Spring Harbor Laboratory. Cold Spring Harbor, New York.

MARFANY, G. and R. GONZALEZ-DUARTE. 1991. The Adh genomic region of Drosophila ambigua: evolutionary trends in different species. J. Mol. Evol.32:454-462.

MARTINEZ-CRUZADO, J. C., C. SWIMMER, M. G. FENERJIAN, and F. C. KAFATOS. 1988. Evolution of the autosomal chorion locus in Drosophila. I. General organization of the locus and sequence comparisons of genes s15 and s19 in evolutionarily distant species. Genetics 119:663-677.

OAKESHOTT, J., G. CHAMBERS, P. EAST, J. GIBSON, and J. BAKER. 1982. Evidence for a genetic duplication involving Adh in D.buzatii and related species. Aust. J. Biol. Sci. 35:73-84.

PALUMBI, S. R. 1989. Rates of Molecular evolution and the fraction of nucleotide positions free to vary. J. Mol. Evol. 29:180-187.

PERLER, F., A. EFSTRATIADIS, P. LOMEDICO, W. GILBERT, R. KOLODNER, and J. DODGSON. 1980. The evolution of genes: the chicken preproinsulin gene. Cell 20: 555-566.

ROWAN, R., and W. J. DICKINSON. 1988. Nucleotide sequence of the genomic region coding for alcohol dehydrogenase in Drosophila affinidisjuncta. J. Mol. Evol. 28:43-54.

SANGER, F., S. NICKLEN, and A. R. COULSON. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463-5467.

SCHAEFFER, S. W., and C. F. AQUADRO. 1987. Nucleotide sequence of the Adh gene region of Drosophila pseudoobscura: evolutionary differences and evidence for an ancient gene duplication. Genetics 117:61-73.

SHARP, P. M., and W-H. LI. 1989. On the rate of DNA sequence evolution in Drosophila. J. Mol. Evol. 28:398-402.

STARMER, W. T., and D. T. SULLIVAN. 1989. A shift in the third-codon-position nucleotide frequency in alcohol dehydrogenase genes in the genus Drosophila. Mol. Biol. Evol.6:546-552.

STEINEMANN, M., W. PINSKER, and D. SPERLICH. 1984. Chromosome homologies within the Drosophila obscura group probed by in situ hybridization. Chromosoma 91: 46-53.

SULLIVAN, D. T., P. W. ATKINSON, and W. T. STARMER. 1990. Molecular evolution of the alcohol dehydrogenase genes in the genus Drosophila. Pp 107-147 in M. K. HECHT, B. WALLACE, and R. J. MACINTYRE (eds). Evolutionary Biology. Plenum, New York.

TAKANO, T., S. KUSAKABE, A. KOGA, and T. MUKAI. 1989. Polymorphism for the number of tandemly multiplicated glycerol-3-phosphate dehydrogenase genes in Drosophila melanogaster. Proc. Natl. Acad. Sci. USA 86:5000-5004.

THROCKMORTON, L. 1975. The phylogeny, ecology, and geography of Drosophila. Pp 421-469 in R. C. KING ed. Handbook of Genetics, vol 3. Plenum, New York.

TICHER, A., and D. GRAUR. 1989. Nucleic acid composition, codon usage, and the rate of synonymous substitution in protein-coding-genes. J. Mol. Evol. 28:286-298.

VISA, N., G. MARFANY, LL. VILAGELIU, R. ALBALAT, S. ATRIAN and R. GONZALEZ-DUARTE. 1991. The Adh locus in Drosophila: chromosomal location and restriction analysis in species with different phylogenetic relationships. Chromosoma 100:315-322.

CAPÍTOL IV: "Evidence for retrotranscription of protein-coding genes in *Drosophila* "

Sotmés a publicació

Evidence for retrotranscription of protein-coding genes in the *Drosophila* genome.

Gemma Marfany and Roser Gonzàlez-Duarte
Departament de Genètica. Facultat de Biologia. Universitat de Barcelona.

ABSTRACT

Evidence is provided for the presence of retrosequences (also named retroposons) arising from *Adh* in the *Drosophila subobscura* genome. Restriction analysis of four different clones containing a retrosequence with flanking regions is reported. Besides, the primary structure of retrosequences S812 and S135 is presented. The fact that they lack introns and a recognizable promoter strongly supports their retrotranscriptional origin. Adjacent to all four analyzed retrosequences, a middle repetitive DNA element has been found which bears no clear similarity with any sequence reported to date in the GenBank/EMBL Data Library. A comparative analysis of these retrosequences with the functional *Adh* gene of *D.subobscura* is presented. In addition, a model concerning the origin, functionality and propagation of these genomic elements is discussed.

INTRODUCTION

The *Adh* gene/enzyme system in *Drosophila* has been the subject of many prominent evolutionary and genetic studies (for recent reviews Sofer and Martin,1987; Chambers, 1988; Sullivan et al., 1990). The genomic structure of the *Adh* region is rather complex and two different organizations have been reported so far. For *D.melanogaster* and its relatives, a single structural gene developmentally regulated by two different promoters has been described, whereas in several species of the *repleta* group two different genes with their own promoter, as well as a genomic pseudogene have been reported. In the *obscura* subgroup, only a single functional gene has been described for *D.pseudoobscura* (Schaeffer and Aquadro, 1987), *D.ambigua* (Marfany and Gonzàlez-Duarte, 1991) and *D.subobscura* (Marfany and Gonzàlez-Duarte, in press). In this paper, we provide evidence for *Adh* processed pseudogenes in this latter species.

Until recently processed pseudogenes, especially those derived from protein-coding genes, were thought to be restricted to mammalian genomes and strong arguments were advanced to support their confinement (for reviews, Rogers, 1985; Vanin, 1985; Weiner et al., 1986; Wagner, 1986). Only some types of genetic

elements in *Drosophila* were suggested as possible processed pseudogenes, e.g. the F family, consisting of mobile elements (DiNocera et al., 1983) or Type I and II rDNA insertions (whose origin is discussed in Eickbush and Robins, 1985). Nevertheless, retropseudogenes from structural genes in *Drosophila* have not yet been described except for a report on processed pseudogenes in two species of the melanogaster subgroup (Jeffs and Ashburner, 1991). The sequences that we have found in *D.subobscura* and other very closely related species (Marfany and Gonzàlez-Duarte, in preparation) clearly show some of the features associated with processed pseudogenes. Their structure contains neither introns nor promoter elements, but traces of the leader and a complete trailer surround the *Adh* sequence. The retrotranscriptional origin of these sequences seems evident. In addition, the presence of an adjacent inverted repeat which belongs to a middle repetitive DNA family could be related to the ontogeny of these genetic elements.

There is a growing wealth of information about families of sequences whose origin and propagation involve reverse transcription (retroelements). The first terms coined to name processed pseudogenes were retroposons (Rogers, 1983; Rogers, 1985) and retropseudogenes (Weiner et al., 1986). Later on, protein-coding processed pseudogenes as well as SINEs and LINEs were also included. In an attempt to avoid confusion, a general nomenclature has recently been put forward (Temin, 1989; Hull and Will, 1989). Accordingly, structural processed pseudogenes should be named retrosequences, retroposon being restricted to SINEs and LINEs. In this study, we will use the term retrosequences to refer to processed protein-coding pseudogenes.

MATERIALS AND METHODS

Genomic clones

Total genomic DNA from *D.subobscura* H27 adults was used to construct a library using the phage vector EMBL4. From 150,000 to 200,000 recombinant plaques were screened (14) with a 2.7 kb *Hind*III-*Eco*RI restriction fragment of sAC1 (15) containing the complete *Adh* gene and the immediately adjacent 3' region, in order to isolate *Adh* containing clones (for conditions, see below).

DNA mapping and sequencing

Positive clones were characterized by restriction mapping using standard techniques. *Adh* was located on the restriction map by transferring digested and electrophoresed DNA to nitrocellulose membranes. Labelling of *Adh* probe, hybridization conditions, washes and autoradiography were performed as for plaque hybridization (see below). Positive restriction fragments were subcloned into Bluescript phagemids (Stratagene) and sequenced on either single or double stranded DNA following the dideoxy method of Sanger (16), with [α-^{35}S)dATP (NEN-DuPont) and modified T7 phage polymerase (Sequenase USB, or Sequencing Kit from Pharmacia-LKB). A set of unidirectional deletions was generated using the method of Henikoff (17) to sequence large regions.

After electrophoresis on TBE buffer, sequencing gels were fixed, dried and autoradiographed for 72 hours. Each nucleotide was sequenced on both strands.

Southern blotting and hybridization

Restricted DNA fragments from phages, plasmids or total genomic DNA (4 to 8 mg) were electrophoresed through 0.8-1.0% agarose gels and transferred by Southern blotting into Hybond-C or Hybond-N membranes (Amersham). For hybridization, the probe was labelled either by nick-translation or by random-hexamer priming with $[\alpha\text{-}^{32}P]dCTP$ (NEN-DuPont) under the following conditions (for plaque as well as Southern blots): 46% formamide at 42°C overnight in the presence of 10% Na dextran sulfate. Nonspecifically hybridized probe was removed with one wash in 2xSSC/0.1%SDS at room temperature for ten minutes, then two washes in 2xSSC/0.1%SDS at 65°C for ten minutes each, and finally two further washes in 1xSSC/0.1%SDS at 65°C for ten minutes each (1xSSC is 0.15M NaCl/0.015M sodium citrate, pH 7.5). After washing, filters were autoradiographed.

DNA Sequence Analysis

Sequences were read and aligned with the 4.0-kb sequence of *D.subobscura* (Marfany and Gonzàlez-Duarte, in press. Accession Number for in GenBank/EMBL Data Library: M55545) using the Sequence Analysis Software Package of the Genetics Computer Group of the University of Wisconsin (Devereux et al., 1984). They are available in the GenBank/EMBL Data Library under the accession numbers: X55390 and X55391.

Genomic DNA PCR

1 µg of genomic DNA from the following species was used in a PCR experiment: *D.melanogaster*, *D.ambigua*, *D.subobscura*, *D.madeirensis*, and *D.guanche*, following the manufacturer's specifications (Taq polymerase from BRL). Two 17-mer oligonucleotides from highly conserved sequences of the *Adh* first and second exons were constructed: 5' ATTGGCCTGGACACCAG 3' and 5' ACGGGCACCTGGTAGAT 3'. The experimental conditions for one cycle were: 1 minute at 94°C for denaturing, 2 minutes at 55°C for annealing and 5 minutes at 72°C for DNA polymerization. This cycle was repeated 35 times. One tenth of the reaction was loaded in a 1% TBE-agarose gel.

In Situ Hybridization.- Salivary glands dissected from third instar larvae were fixed and squashed in 45% acetic acid. The DNA probe (see Fig.1) was labeled using Bio-11-dUTP or Bio-14-dATP (Enzo and BRL) and detection of biotin was carried out by incubating the slides with a streptavidin-peroxidase complex (Detek I-hrp Signal Generating System, Enzo). Peroxidase activity was detected using diaminobenzidine as a chromogen and afterwards, the preparations were stained with Giemsa. The detailed protocol is described in Visa et al. (1991).

RESULTS

Multiple copies of *Adh* processed pseudogenes

A *Drosophila subobscura* library was screened for *Adh* homologous sequences with a 2.7-kb *Hind*III-*Eco*RI restriction fragment of sAC1 (Goldberg, 1980). This screening yielded several positive clones, 15 of which were isolated and analyzed. Six of them contained the functional *Adh* gene (Marfany and Gonzàlez-Duarte, in press), and the rest contained sequences of processed *Adh* pseudogenes. The restriction maps of these latter sequences could be associated to two patterns of restriction sites, those of λS135 and λS812 (Fig.1). The isolated clones were partially sequenced while λS135 and λS812 were characterized for comparison

with the functional *Adh* gene. All of them bore an *Adh* homologous region but lacked either introns or promoter elements. Besides, their flanking regions contained a reduced leader and a highly conserved trailer without any recognizable polyA track. All these characteristics strongly suggest their retrotranscriptional origin.
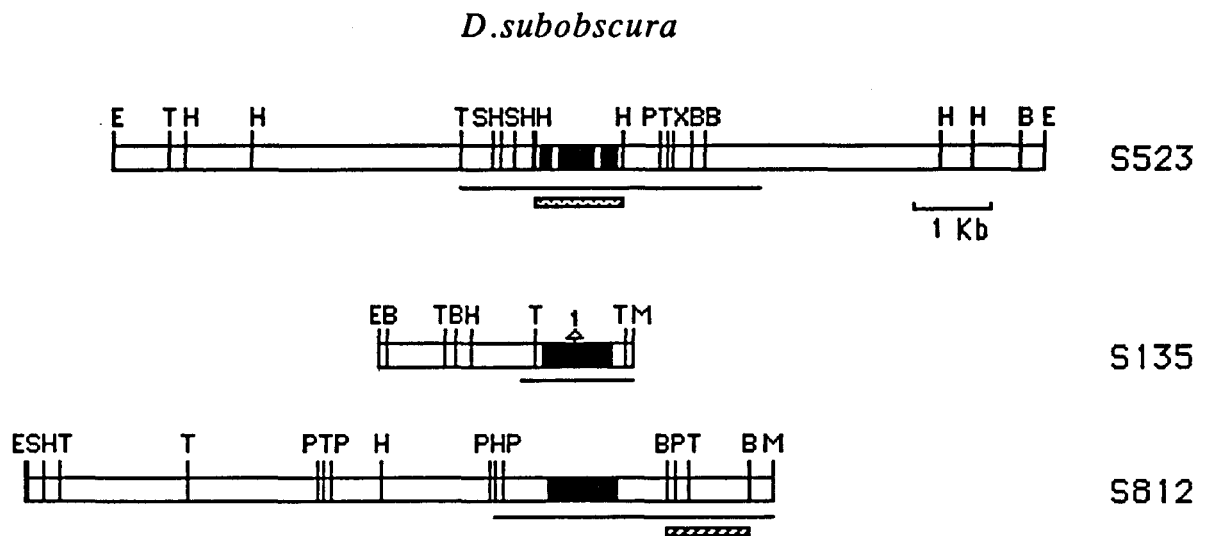
D.subobscura

Fig.1.- Restriction map of *D.subobscura* *Adh*-homologous sequences. Functional *Adh* and flanking genomic regions (S523). Different clones bearing *Adh* retrosequences (S120, S114, S135 and S812). Black boxes represent *Adh* coding regions. Sequenced regions are underlined. The hatched box shows the probe containing one terminal of the inverted repeat located downstream from the λS812 retrosequence (see Fig.3a). This probe was also hybridized to restricted clones that bore *Adh* and these positive fragments are shown by dotted boxes. Abbreviations: B-*Bam*HI, E-*Eco*RI, H-*Hin*dIII, P-*Pvu*II, T-*Pst*I, S-*Sal*I, X-*Xba*I, M-*Mbo*I (ending of the clone).

In situ hybridization on polytene chromosomes have confirmed the presence of different *Adh* homologous sequences on *D.subobscura* genome as two hybridization sites, on chromosomes U and E, have been observed (Visa et al., 1991) when a probe containing only the *Adh* gene was used (Fig.1). Previous analysis of genetic linkage revealed a single *Adh* functional locus, near the centromeric region, on chromosome U (Pinsker and Sperlich, 1984). According to these earlier results the hybridization site on chromosome U would correspond

to the functional *Adh* gene and the site on chromosome E might be due to non-functional *Adh* sequences (Visa et al., 1991).

Common retrotranscriptional origin

The nucleotide sequence of λS812 was aligned with homologous sequences of λS135 and also with the functional *Adh* gene (Fig.2), and all possible comparisons were drawn (Table I). The two retrosequences bore greater resemblance to each other than to the functional *Adh* although λS135 had accumulated more nucleotide substitutions than λS812. Their higher similarity, together with the fact that they shared homologous flanking regions at the boundaries of the *Adh*-derived sequences, independently of their genomic location, strongly suggested a common retrotranscriptional origin.

Sequence similarity extended to the boundaries of the *Adh*-derived sequences

Alignment of the two retrosequences revealed that sequence similarity extended to upstream and downstream regions outside the *Adh*-derived elements (Fig.2). The upstream region of λS135 analyzed (301 bp) aligned perfectly with the upstream sequence of λS812 except for three deletions of 1 bp each. Immediately adjacent to the trailer of both retrosequences, there was one terminal sequence of an inverted repeat of 192 bp which was identified in λS812. This sequence was perfectly aligned in both retrosequences but for a single deletion of 44 bp in λS135 (Fig.2). Further downstream, there was no detectable similarity between the two sequences.

TABLE 1.- Structural features and comparative analysis of *Drosophila subobscura* retrosequences S812, S135 and the functional *Adh* gene.

Region	S812 [2]L	S135 [2]L	[1]S812 vs. funct. Adh [3]NC	PD	[4]CPD	[1]S135 vs. funct. Adh [3]NC	PD	[4]CPD	S135 vs. S812 [3]NC	PD	[4]CPD
5' seq	304.0	301.0	–	–	–	–	–	–	11	3.6	3.7
Leader seq	26.0	26.0	7	26.9	33.4	7	26.9	33.4	0	0.0	0.0
Exon 1 seq Silent	29.7	29.7	5	16.8	19.1	4	13.5	14.8	0	0.0	0.0
Replacement	63.3	63.3	2	3.2	3.2	4	6.3	6.6	2	3.2	3.2
Exon 2 seq Silent	112.7	112.7	9	8.0	8.4	10	8.9	9.4	3	2.7	2.7
Replacement	292.3	292.3	16	5.5	5.7	17	5.8	6.1	1	0.3	0.3
deletions	–	1.0	–	–	–	1	1.1	1.1	1	1.1	1.1
Exon 3 seq Silent	70.7	112.7	6	8.5	9.0	5	7.1	7.4	1	1.4	1.4
Replacement	196.3	196.3	7	3.6	3.7	9	4.6	4.7	2	1.0	1.0
Trailer seq	171.0	171.0	12	7.0	7.4	14	8.2	8.7	3	1.8	1.8
Inverted repeat seq	192.0	148.0	–	–	–	–	–	–	6	3.1	3.2
SUMMARY											
Flanking regions	496.0	449.0	–	–	–	–	–	–	17	3.4	3.5
Leader and Trailer	197.0	197.0	19	9.7	10.3	21	10.7	11.5	3	1.5	1.5
Adh retro TOTAL	765.0	764.0	45	5.9	6.1	50	6.5	6.8	10	1.3	1.3

L- Length in base pairs; NC- Number of changes; PD- Percent differences; CPD- Corrected percent differences. [1] Comparison is restricted to coding *Adh* and related sequences: leader and trailer. [2] The number of silent sites for the functional *Adh* gene of *D.subobscura* has been calculated as the percentage of the potentially silent sites (possible substitutions that would not lead to an aminoacid replacement) out of the total coding nucleotides (22). [3] Insertions and deletions are oversimplified to one single change (event) and appear specified when they affect coding regions. [4] Corrected percent divergence is estimated as d= 3/4 ln(1-4/3p), where p is the proportion of nucleotide sites that differ between two sequences (23).

92

```
            *         *         *         *         *         *         *         *
AAGCTTCGGCAGCTGGACGAATTTCGATTGATTTACGAGATATACTCCGGGTGCGTAAAATGCATACCCAGCAAGCTGTA      80
            *         *         *         *         *         *         *         *
GCCAGTGTGGCAGTCCTTACTGTGTGAGAAATTTGAGTTGATAGTTTGCCCCAACCGCGTGAGTAGCGGGTATCCCACGC     160
            *         *         *         *         *         *         *         *
ACATTCATCGATGCATGTGTGCCGCTCAGGTGTGTGTGAGGGGAAATAACCAATCAAAGGTAAGACATCCCATTTGCGAA     240
            *         *         *         *         *         *         *         *
CCATAAATTCCAGGTACGATCTGCTACAGCGTCAGTGAACCGAGACTCAAAAGTCATTCGTTTGGACCTCTTGTTTTGTG     320
            *         *         *         *         *         *         *         *
CCTTTGACATTGTTTGGCAGTTGTTCCTTTCAGTCGAAATAAATACCCTCATTCTATTTCTGGTGTTCGTTCGGCAAACA     400
            *         *  .      *         *         *         *         *         *
TCTCTTTCTTTAATTCTAAATTTTCAGTAACATTTTTTTCTAATATTCAACATAATCCAAAAAATAATATTCAACATAAA     480
            *         *         *         *         *         *   *     • * G   *
GTACCTGGTCACTAGTCTCAAGTACTAAAACAAGAACTCTCTGAAAAACAATCGAAATTAATATTTAAAATTTTCACTGT     560
            *         *         *         *         *         *         *       _*
TTATATTGCGAGCAACCGAAGAGCATGTCGGGAGTCGGAATCAATGGCTTTGGCCGCATCGGTTCGCGTGCTTCTGCGCC     640
T           *  _ _     *         *         *         *         *         *         *
AATGTCTGTCGCGCAAGGATGCGCAGGTGCTGACCATCAATGACCCCTTAATGGTTCCCAAGTACTTGGCCTATACGCTG     720
            *         *       A *         *         *     * G       * G   C *
CGTTACGACTCCACTCACGGCATTTGCAACAGGAACAACACACTGGCGGTTGATGGTAAAAAGCTGCTGCAAGAATACGA     800
                                            G CT          C             CT   C
                                         ▲T │ │          │
            *  T     *A    *         *         *         *         *         *
TCCCAGGAAGCTCAAGTGGGGGGAACGGGGCGTTAAGACTCTGAAGGATTGCCAGAAACAAAAGAAAAGATCCAA|ATG|T     880
   T                                 G T          TCG          A
            *         *         *     *   *   * G        *         * G   *
CACTCACAAACAAGAATGTTGTTTTCGTGGCTGGTCTGGGATGCATTGGCCTGGACACCAGCAAGGAGTTGGTTAAGCGT     960
  GT.,AG
  │  T
            *      T.*  *         *      G  CG    *         *         *         *
GATCTGAAGAACCTGGTCATCCTGGATCGCATTGACAATCCAGATGCCATTGCCGAACTGAAGGCAATCAATCCCAAGGT    1040
          A         *    T   G     A           AC      G CC  A
            *         *         *         *         *         *         *
GACCGTCACCTTCTTCCCTTATGATGTGACTGCACCTCTCGCAGAGACCGCCAAACTCCTGAAGTGCATCTTTTCAAAGT    1120
          A    G    A    T        C        T
            *         *         *         *         *         *         *
TCAAGACCGTCGATATCCTGATCAACGGTGCGGGCATCCTGGACGATCATCAGATCGAGCGTACTATTGCCGTTAACTAC    1200
T                 C              T         G
            *     *C  *         *         *         *         *         *
ACTGGCCTGGTCAACACCACTACAGCCATTCTGGACTTCTGGGACAAGCGCAAGTGCGGCCCAGGTGGCATCATTTGCAA    1280
              .    C
            _       *         *         *         *         *         *
CATTGGCTCCGTTACCGGTTTTAATGCTATCTACCAGGTGCCCGTTTACTCTGGCAGCAAGGCGGCGGTGGTTAACTTCA    1360
  GT.,AG
  │                            A       *A       G
            *         *         *         *         *         *         *
CCAGCTCCCTGGCGAAACTTGCACCCATCACTGGAGTCACCGCCTACACTGTGAATCCTGGCATCACCAAGACCACTCTG    1440
  G C                          GA         C              T
  G         *         *         *         *         *         *         *
GTACAGAAATTCAACTCGTGGCTGGATGTGGAGCCCAAGGTGGCGGAGAAGCTGTTGGAGCATCCCACCCAGACCACTCA    1520
      G                                 A      C
            *         *    *A   *         *         *         *         *
GCAGTGTGCCAAGAACTTTGTCAAGGCCATTGAGCTGAACAAGAATGGTGCCCTCTGGAAATTGGATTTGGGAACTCTGG    1600
  T   C                                A  C  TG       ------
            *         *         *    *   A A   *  TG       *    C *
AGCCCATCAAATGGACCAAGCACTGGGATTCGGGCATC|TAA|ACGGGATACCCGCCCCATAACCCATTCAATGGGATAGGG    1680
C A                      A   TTACGTATATATCTACATATGGC          CTCT
            *         *         ---------------------  *         ----
TTTTTAGCTTTTAGCTTCGCTTTTCCACTCAATTG---------------------AAATAAGGCTGATTTGATT----    1734
TTAAATGGAACCCCGTTTTGAATA                              T -    ▲
------------------------  *         *         *         *         *
-----------------------TGATAATAAAAATTATATTTGAGAAATTAACAACAAAAAACGAGAAGGGACGTGTG    1790
       *         *    TTA -----------------------------------------------
AGACGCTTCTTACGCGTCACAACTTTTATACCCGGCACTCAGTACTACATCTGCAACTTAGCGGTTATTTGTCAAATTTT    1870
```

```
       T   *           *          T*           *           *           *           *           *           *
ACATTTCTTCTTCATCTGTCATCTACATCAACAACACTACTCACGCCAACACGCTCCTTTAGCTCGCCACCCTCCCCTAG    1950

       *●           *           *           *           *           *           *           *           *
AAACACACACTGTGAAACAGTGACACAGTGAAAAAACATGGAAACCAGTTACCTTCTACCGACCTCTGTTCGCCTGGACT    2030
       *           *           *           *           *           *           *           *           *
TGGTCTGAAAATTTTCCAGCTTTCGGCCATTTCAGTGAAACAATTTCAAAAGTTCACAGTGGCGTCCCAATTCTCTGATG    2110
       *           *           *           *           *           *           *           *           *
GGCCATTCTCGAACTACGCTATCTTAGTTCCAGAATGACCGTCTTCATCGCGTGGGCGCCACTGTGCAAAGAGGTGACTG    2190
       *           *           *           *           *           *           *           *           *
TCAAAATTTGACAGAGATAACAGGAGAGATTTGTTTTGTTGTTTTCTCCTTTATTTGTAGTCCGTCCCGGATGTGCCACG    2270
       *           *           *           *           *           *           *       ·   *           *
CCCCCTCGTGATCGGCACAAGTCCTGGATCCTTTGCGACCAGTTCGGGTGCAAGGACATTAAGGATAATGCCATTTTCAC    2350
       *           *           *           *           *           *           *           *           *
GTTGTTTTTTCTTTTTAATCTCTACAAGGCTATAGCTCAGCTGTTTCCTGTCGAAAAAGGACATGTCACGCCTCCTCGGG    2430
       *           *           *           *           *           *           *           *           *
ATCGGGAAAATGTCCCGGATCACTTCTGGTGCTAGGACATTAAGGATAATGCCATTTGAATATTTTTTTCTATATAATCC    2510
       *           *           *           *           *           *           *           *           *
CCTTCAACTCTATAACTCGGCTGTTTCCTGTCGGATCAGGACATGGCATGCCTCCTGGCGAAGGTACGAGTCTCCTGTAT    2590
       *           *           *           *           *           *           *           *           *
CGCTTGAGTCCATCAAGGACCACAAGGACTGCAGGATATGGCTGTCCTACGCTATTTTGTAATTGGGAACTTTTCAGCCT    2670
       *           *           *           *           *           *           *           *           *
GAAAGTATGCTATACTTTTGAGAACAAAAGCAGCCCCCAAAAGATACTTCAGTTTATTGGCCTCTGCCCCTGACACGTCT    2750
       *           *           *           *           *           *           *           *           *
CTGCTGCTGCCTCTGCCGCGACTCTGCCCTGACTCTGAAGTGTGTGTTTCTAGGGGAGGGTGGCGAGCTAAAGGAGCGTG    2830
       *           *           *           *           *           *           *           *           *
TTGGCGTGAGTAGTGTTGTTAATGTAGATGACAGATGAAGAAAAAATGTAAAATTTGACAAATTATCGCTAAGTTGCAGA    2910
       *           *           *           *           *           *           *           *           *
TGTAGTACTGAGTGCCGGGTATAAAAGTTGTGACGCGTAAGAAGCGTCTCACACGTCCCTTCTCGTTTTTATACCCGGTA    2990
       *           *           *           *           *           *           *           *           *
CTCGAAGCGTAAATAGGGTATATTGTATTTGTGCGGATAACGGTTGGTATGTAACGGCAGAGAAGAAACGTTTCCGACCC    3070
       *           *           *           *           *           *           *           *           *
CATAAGATATACATATATTCTTGATCAACATCAATAGCCGAGTCGATTGAGCCATGTCTGTCTGTCCGTCTGTCCTTCTG    3150
       *           *           *           *           *           *           *           *           *
TCCGTCTTGTTTGTCGGCTAGTTCTCAGAGCTATAAGAGCTAGAGCCACCAAATTTTGGCACCAGACTGCTGTATGCTCA    3230
       *           *           *           *           *           *           *           *           *
CACTGAAACCAAGGTATTTCAAAAATGAGCCCCGCCCCCTTCTGCCCCGGCAAAAAGACGAAAACCTCCCAAATCTACAA    3310
       *           *           *           *           *           *           *           *           *
AATACAATGAAGATAACAGAAAACTAAAAACGCCATTCCGTAGGGAATGACCAAATCTATCAGATCACCAAATTGGATCC    3390
       *           *           *           *           *           *           *           *           *
GATCGGATCATTATTATGGCCACAATGATTACGGGTATCTATCGCGATATTTTCCAAATTTTTCAATGTATCACTTACAA    3470
       *           *           *           *           *           *           *           *           *
ACATGTCTCGCTTAATAATTTACATACTTTACATACGTAGACAAAGTCTCGCACTCGCTGGCCCATTGCGTGGGCGCTGA    3550
       *           *           *           *           *           *           *           *           *
CTTGGCCATTGGGCGCTAGGCATAGCTGTGAAGTTCAAGGAGGTCTACGGCAAGGTCGATGATGCTGCGTGCCCAGAAGG    3630
       *           *           *           *           *           *           *           *           *
CGGCGGCCATCGGTATGTGGCTGTTGCTCAAGGATAGATCCGGGGAATTCGATATCAAGCTTATCGATCC    3700
```

Fig.2.- Nucleotide sequence of S812 (3.700-kb), aligned with S135 retrosequence (differences displayed immediately above) and with *D.subobscura* functional *Adh* (differences displayed in the second line above). The beginning and the end of sequence alignments are indicated, respectively, by a filled circle and a filled triangle. The first and STOP codons are boxed. Deletions are marked by hyphens, and insertions and intron positions are depicted above the corresponding line. The inverted repeat is underlined by empty arrows.
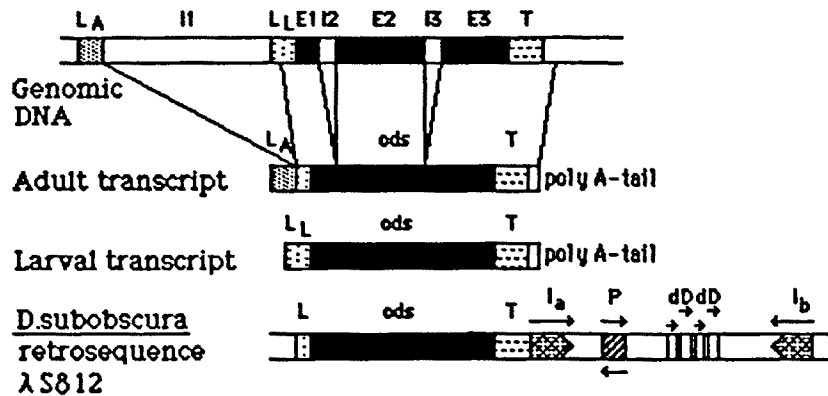
Fig.3).- Model illustrating the origin of *Adh* retrosequences through an RNA intermediate. The final structure of adult and larval transcripts as well as the structure of the S812 retrosequence with some of the features of the inverted repeat region are represented. Abbreviations: L-leader; L_A-Adult leader; L_L-Larval leader; E1, E2 and E3-1st, 2nd and 3rd exons; I1, I2 and I3-1st, 2nd and 3rd introns; cds -coding sequence; T-trailer; Ia and Ib-inverted repeat; P-palindromic sequence; d and D -two different direct repeats.

The inverted repeat belongs to a middle repetitive DNA family of *D.subobscura* genome

In λS812, the inverted repeat of 192 bp delimited an 827 bp sequence which also showed remarkable features, such as a 96 bp palindrome and two alternate direct repeats of 40 and 52 bp (Fig.3). The sequences linking the two different repeats have the same length and show traces of similarity, suggesting that they may once have belonged to a single repeat. The 1.1kb *Bam*HI probe from λS812 containing the inverted repeat (Fig.1) was used as a probe for hybridization to restriction blots of different retrosequences and functional *Adh* clones. Similarity was detected not only downstream from each retrosequence, but also upstream and downstream from the functional *Adh* gene (Fig1). In addition, hybridization with the same probe to Southern blots of total digested DNA yielded the multiple banding pattern characteristic of middle repetitive DNA families (Fig.4a). This result was totally confirmed by *in situ* hybridization analysis on polytene chromosomes (Fig.4b) as hybridization was detected near the centromeric regions of all chromosomes as well as in some dispersed bands along the chromosomal arms.

When the GenBank/EMBL database library was screened (Devereux et al., 1984) for sequence similarity with the inverted repeat and with the region in between, no clear ressemblance was found although interspersed similarity was observed with some *Drosophila* transposable elements.
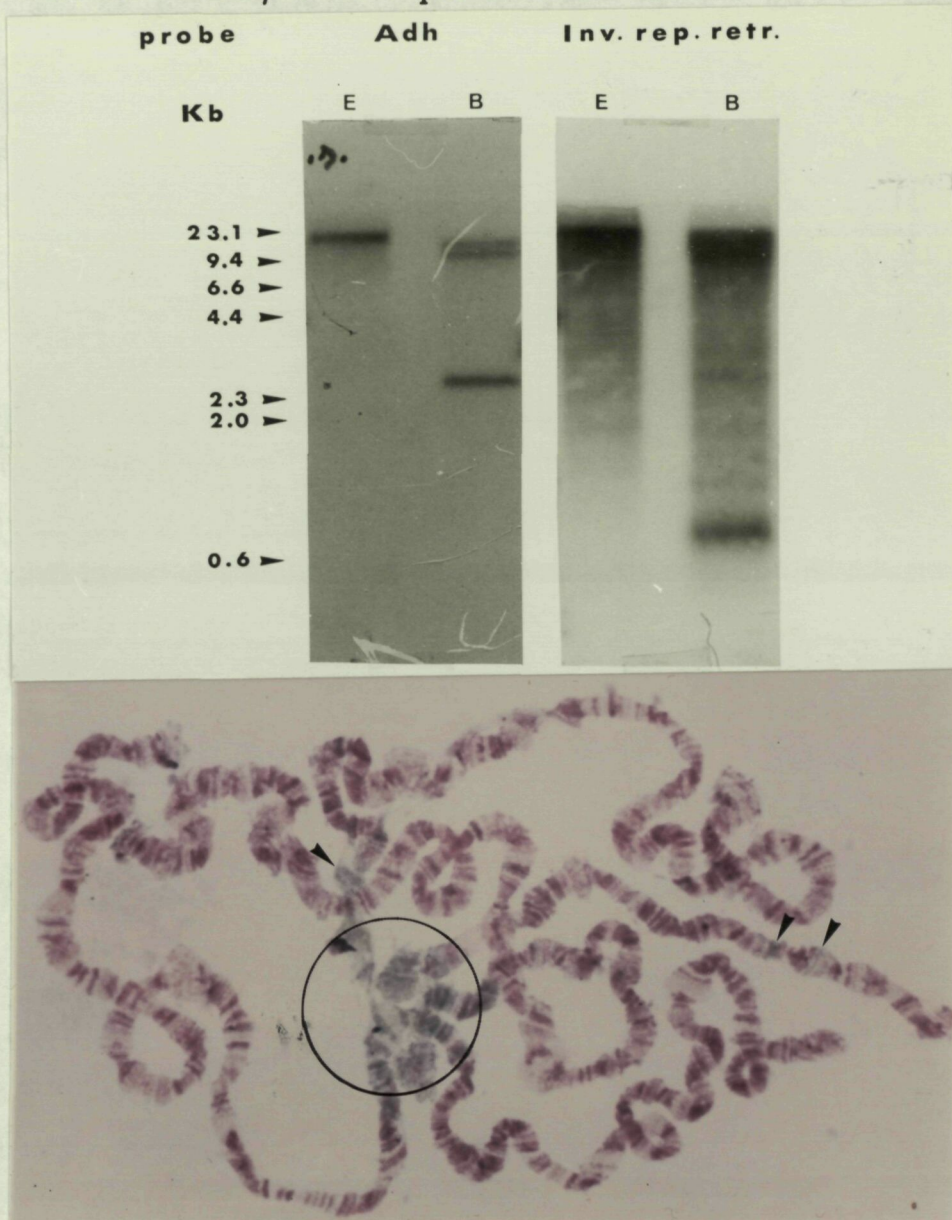
Fig.4a).- Southern blot of total genomic DNA from *D.subobscura* digested with *EcoRI* and *BamHI* and hybridized with two probes. *Adh*-Pattern obtained with the probe which contained only the *Adh* gene (*HindIII-HindIII*, 1.1 Kb, see Fig.1). Inv.rep.retr.-Pattern obtained with the probe which contained part of the inverted repeat of S812 (*BamHI-BamHI*, 1.1 Kb, see Fig.1).

Fig.4b).- *In situ* hybridization on the polytene chromosomes of *D.subobscura* with the probe containing the inverted repeat sequence (Fig.1). Hybridization signals clustered on the centromeric region are circled and those on the chromosomal arms are shown by arrows.

Sequence analysis of *Adh* retrosequences with respect to the functional gene

A detailed analysis of *Adh* retrosequences could shed light on their evolutionary pattern and possible functionality. It is worth pointing out that both retrosequences showed a high level of similarity with the functional *Adh* gene (94.11% for λS812 and 93.59% for λS135) although λS135 presented a small deletion of 1 bp in the coding sequence resulting in a frame-shift.

Comparison of the λS135 and λS812 sequences with functional *Adh* revealed 45 substitutions for λS812 and 49 plus a 1bp deletion for λS135 (Table I). These substitutions were not randomly distributed, as they were significantly fewer in the central position of a codon (λS812 vs funct. *Adh*, χ^2=6.54 and λS135 vs. funct. *Adh*, χ^2=6.43; both cases, d.f.=2 and P<0.05). Silent and replacement substitutions in λS812 seemed to deviate from random, although this is not so evident when considering λS135 (λS812 vs. funct. *Adh*, χ^2=6.17; d.f.=1; P<0.02 and λS135 vs. funct. *Adh*, χ^2=2.9; d.f.=1; P>0.2). However, both silent and replacement substitutions were randomly distributed throughout the whole sequence, without significant deviation across the three exons (for silent substitutions: λS812 vs. funct. *Adh*, χ^2=2.05; d.f.=2; P>0.3 and λS135 vs. funct. *Adh*, χ^2=0.96; d.f.=2; P>0.5; for replacement substitutions: λS812 vs. funct. *Adh*, χ^2=1.24; d.f.=2; P>0.5 and λS135 vs. funct. *Adh*, χ^2=0.43; d.f.=2; P>0.8).

Sequence analysis of *Adh* retrosequences λS812 and λS135

λS812 and λS135 showed a higher level of similarity with each other (98.8%) than when compared individually to functional *Adh*. The two retrosequences showed some differences in the nucleotide sequence (9 changes in addition to the single deletion in λS135) which supports their duplication and divergence some time after the retrotranscription and genomic insertion events.

The distribution of substitutions among the three positions of a codon appeared to be at random (χ^2=2.67; d.f.=2; P>0.2), as was the occurrence (χ^2=1.57; d.f.=1; P>0.3) and distribution of both silent (χ^2=1.0; d.f.=2; P>0.5) and replacement (χ^2=4.6; d.f.=2; P>0.1) substitutions along the whole sequence.

It is worth noticing that *Adh* retrosequences (including leader and trailer sequences) seem to be evolving more slowly than the flanking regions of λS812 and λS135 (summary of Table I).

Evolutionary history of *Adh* in *Drosophila subobscura*: a model

Different mathematical procedures have been described to calculate the timing of duplication or pseudogene formation (for review Vanin, 1986 and references therein). We have considered the mathematical model proposed by Miyata and Yasunaga (1981) to explain the present organization of *Adh* in *D.subobscura* (Fig.5).
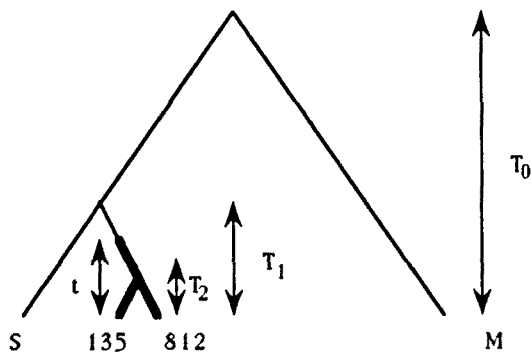
Fig.5.- Proposed model for the present organization of the sequences homologous to *Adh* in *D.subobscura* genome. Mel-*D.melanogaster Adh* gene, Sub-*D.subobscura Adh* functional gene, 812- S812 retrosequence, 135- S135 retrosequence. The inactive state of pseudogenes is indicated by the thick line. T_0-time since divergence of *D.subobscura* and *D.melanogaster* ; T_1-time since the retroposition event; T_2-time after the pseudogene duplication or multiplication and t-time since the original retropseudogene lost its functionality.

The equations of this model are based on three different evolutionary rates which are considered constant throughout the evolution of a gene. Two of them apply to functional genes, the rates for silent substitutions (v_s) and for replacement substitutions (v_a), and the third applies to substitutions in pseudogenes (v_0). Giving a value to T_0, time since divergence of *D.subobscura* and *D.melanogaster* and calculating the corrected percentage divergences for silent, replacement and overall substitutions among λS812 and λS135 retrosequences and among the functional *Adh* of *D.melanogaster* and *D.subobscura*, the following parameters were calculated: T_1, time since the retroposition event; T_2, time after the pseudogene duplication or multiplication; and t, time since the original retropseudogene lost its functionality. An independent loss of function of each retrosequence was also postulated, but this assumption was proved not to be correct under the conditions that we imposed. The final equations gave,

$$T_1 = T_0\{1 - (f(D.mel\text{-}S812) - f(D.sub\text{-}S812))/f(D.sub\text{-}D.mel)\} \quad (1)$$
$$t = 2T_0\{1 - (f(D.mel\text{-}S812)/f(D.sub\text{-}D.mel)\} \quad (2)$$

where $f(\ ) = Kc_s(\ ) - Kc_a(\ )$ and $Kc(\ )$ = corrected number of substitutions per nucleotide site. With T_1 and t, the corresponding substitution rates were deduced and, then,

$$T_2 = Kc_0(S812\text{-}S135)/2v_0 \quad (3)$$

Assuming that T_0 (divergence time for *D.melanogaster* and *D.subobscura*) is 25Myr (Throckmorton, 1975) and solving the whole set of equations, we obtained T_1 = 3.05 Myr (time for retropositional event), t = 1.46 Myr (time for loss of function of retrosequences) and T_2 = 0.23 Myr (time for duplication of retrosequences). The corresponding values for substitution rates were: v_S = 1.28 x 10^{-8}; v_a = 1.24 x 10^{-9} and v_O = 2.8 x 10^{-8} per nucleotide site per year.
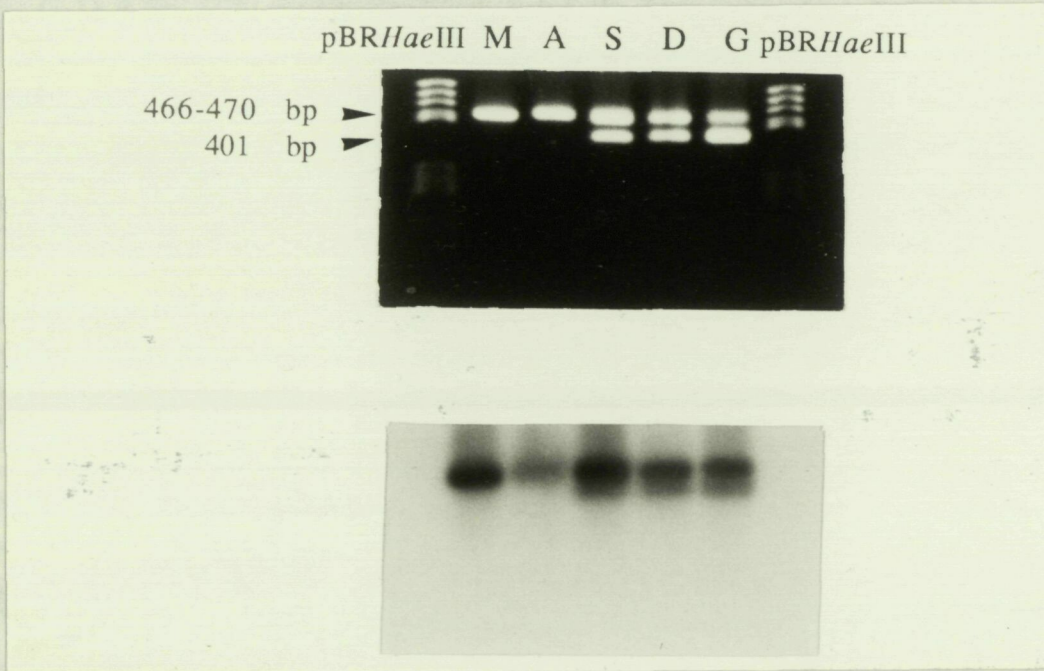
Fig.6.- PCR amplification of total genomic DNA from *D.melanogaster* (M), *D.ambigua* (A), *D.subobscura* (S), *D.madeirensis* (D) and *D.guanche* (G) and hybridization with the *Adh* probe (Fig.1). The longer fragments correspond to those yielded by the functional *Adh* genes (with an intervening sequence), whose size ranges from 466 to 470 bp, depending on the intron length of the different species. The shorter fragments correspond to *Adh* retrosequences, with a size of 401 bp.

Adh retrosequences are not restricted to *D.subobscura* genome

The main feature that defines a processed pseudogene is the lack of introns. Using the PCR technique in total genomic DNA with two suitable primers it is possible to detect the presence of processed pseudogenes in addition to the functional gene. Two 17-mer oligonucleotides corresponding to conserved coding sequences in the first and second exon were designed. Following PCR amplification and DNA electrophoresis, fragments of different length were observed. The longer ones corresponded to the functional gene, as they contained one intervening sequence. The data obtained (Fig.6) with total genomic DNA

from *D.melanogaster*, *D.ambigua*, *D.subobscura*, *D.madeirensis* and *D.guanche* (the latter four species belonging to the *obscura* subgroup) revealed that both *D.melanogaster* and *D.ambigua* had only one single functional gene whereas *D.subobscura* and the sibling species, *D.madeirensis* and *D.guanche*, all showed, in addition, *Adh* retrosequences (Marfany and Gonzàlez-Duarte, in prep.). Direct evidence for *D.madeirensis* and *D.guanche* retrosequences is also provided in Visa et al. (1991).

DISCUSSION

Not all the features associated with processed pseudogenes or retrosequences: variable chromosome location, lack of promoter elements, absence of intervening sequences (introns), traces of polyA trail at the 3'end and flanking direct repeats of usually chromosomal (A)-rich DNA (Rogers, 1985, Weiner et al., 1986) are always present in every processed pseudogene (Li and Johnson, 1989). In our case, λS812 and λS135 share most of them, and the lack of introns and of promoter elements are particularly notable.

Adh processed pseudogenes of *D.subobscura* appear to be located in different chromosomal sites than the functional *Adh* gene, as detected by *in situ* hybridization on polytene chromosomes (Visa et al., 1991). Furthermore, the similarity among these *Adh* retrosequences extends to flanking regions which are unrelated to the functional gene. This high similarity seems to suggest that the retrosequences analyzed have diverged from a common precursor following a duplication event.

Four different aspects are usually considered when dealing with processed pseudogenes: i) how they integrated into the genome, ii) how they became inheritable, iii) when they originated, and iv) whether or not some kind of selective pressure has ever acted upon them.

Of all the different mechanisms put forward to explain the origin of retrosequences (Rogers, 1985; Vanin, 1985; Wagner, 1986; Weiner et al., 1986) the most accepted view is based on integration of cDNA copies of the corresponding mRNA into the genome, probably at staggered chromosome breakpoints. The presence of reverse transcriptase activity in *Drosophila* has already been reported (Finnegan, 1989). However, small flanking direct repeats which would be the hallmarks of the integration processes have not been found for lS812 or lS135 retrosequences, although they may have been masked by sequence divergence or multiplication events.

The *Adh* transcripts must have been introduced into the germ line to become inheritable and this may have been accomplished by different means. One

possibility is that as *Adh* is expressed in the nurse cells of the ovary, its transcriptional product could be available for retrotranscription and insertion into the genome of the oocyte. In fact, immunohistochemical and histochemical analysis on *Adh* tissue distribution in *D.melanogaster* (Visa et al., in preparation) and in *D.grimshawii* and its relatives (Sullivan et al., 1990) has revealed the presence of ADH in the ovarian nurse cells, which form a syncithium with the immature oocyte.

Concerning the origin and functionality of these retrosequences, the high degree of sequence similarity with the functional gene seems to suggest that they are the result of a recent evolutionary event. The existence of *Adh* retrosequences in a cluster of very closely related species, *D.subobscura, D.madeirensis* and *D.guanche*, together with the lack of retrosequence counterparts in *D.ambigua* (also in the *obscura* subgroup) strongly supports the hypothesis that this retrotranscriptional event took place before the separation of the former three species but after the earlier radiations within the *obscura* subgroup. Further analysis on the number and sequence similarities among these retrosequences may shed light on the origin and evolutionary dynamics.

As in other reports (Miyata and Yasunaga, 1981; Li and Johnson, 1989), the evolutionary pattern of processed pseudogenes shows higher conservation than non-coding sequences and a restraint for replacement substitutions, especially those affecting the second-codon position has been detected. This trend is also found when considering λS812 and λS135 retrosequences and it stands out especially when comparing either of them with the functional *Adh* gene. Several questions remain open: first, whether the sequence is performing or has performed some function and is thus evolving or has evolved under selective pressures, although this is unlikely. Another possibility is that this phenomenon might reflect some constraints operating on protein-coding-like sequencea. Finally, although gene conversion could have played a role in the evolution of these sequences as it has been attributed to some retrosequences (Miyata and Yasunaga, 1981), it would be difficult to explain how the conversion of introns and flanking regions had been obviated.

It has been proved that most retrosequences bear no function, due to the lack of promoter elements (consequently they are unlikely to be transcribed), as well as to the acquisition of mutations impairing the original function. Very few exceptions have been reported of retrosequences being transcribed and even fewer of them gaining new functions (Perler et al., 1980; Stein et al., 1983; Soares et al.; 1985, Boer et al., 1987; Ashworth et al., 1990). In our case, the fact that these analyzed retrosequences lack a recognizable promoter strongly suggests that they

are not transcribed, at least in their present structure and organization. Analysis on *Adh* expression in *D.subobscura* have revealed one single transcript either on larvae or adults. Furthermore, there is one single protein with ADH activity in *D.subobscura* (Hernández et al., 1988). On the other hand, their high similarity with the functional *Adh* gene sequence would not support a different function and moreover, the single deletion of one base pair in λS135 causing a frame shift which produces a STOP codon precludes the functionality of this retrosequence.

Another possibility to be tested is the ancient functionality of these retrosequences. The model proposed by Miyata and Yasunaga (1981) for the evolution of the mouse α-globin gene and its homologous pseudogene, when applied to *D.subobscura Adh*, suggests that: i) *Adh* retroposition in this species was a recent evolutionary event, ii) the two retrosequences have diverged very recently (roughly 230,000 years ago) and iii) the ancestral retrosequence seemed to have evolved under selective pressure half of his life-time (thus explaining the distribution and types of substitutions in retrosequences S812 and S135). Although this model was formerly designed to analyze the evolutionary pattern of mammalian genes and constant substitution rates have been assumed, we believe that these approximate values shed some light on the timing of events. The figures obtained for the substitution rates are also worth noting: v_a is equivalent to that found for a mammalian gene, but v_S is twice that found for the α-globin genes. The higher evolutionary rate in insect DNA with respect mammalian DNA (Sharp and Li,1989; Caccone and Powell, 1990) would account for this difference. On the other hand, the substitution rate in these retrosequences, v_0, is roughly two-fold that obtained for silent substitutions in the Adh gene, a relative value similar to that reported for the α-globin genes.

The presence of an inverted repeat sequence immediately adjacent to the *Adh* retrosequence is noticeable. Furthermore, it is a member of a middle repetitive DNA family in *D.subobscura* genome whose distribution could be related to element mobility (Martin et al., 1983; Felger and Sperlich, 1989) as in mammalian genomes, where most of the major families of interspersed repeats have originated through a retroposition mechanism (Deininger and Daniels, 1986). In our case, this would provide a plausible explanation for the multiple presence of *Adh* retrosequences as well as for the similarity shared by flanking regions.

In summary, we suggest that three independent events took place to produce the present genomic organization of *Adh*-retrosequences in *D.subobscura*: firstly, an RNA intermediate was copied into cDNA, by an unknown reverse-transcriptase; secondly, this cDNA was integrated into the germ-line genome and has been accumulating substitutions with respect to the functional precursor

gene and thirdly, this retrosequence was propagated in the genome by a multiplication mechanism that also affected its flanking regions. Each retrosequence has been diverging independently since then. The two former events are the common basis for retrotranscriptional processes. We also suggest that this retroposition is a recent evolutionary event. The last step has been the spreading of these *Adh* retrosequences in the *D.subobscura* genome.

Thus, the same complex mechanisms underlying vertebrate genome evolution, mainly described in mammals, seem also to be of relevance in the dynamics and evolutionary patterns of invertebrate genomes.

ACKNOWLEDGEMENTS

REFERENCES

Ashworth A, Skene B, Swift S and Lovell-Badge R (1990) *Zfa* is an expressed retroposon derived from an alternate transcript of the *Zfx* gene. EMBO J 9:1529-1534.

Boer PH, Adra CN, Lau Y-F and McBurney MW (1987) The testis-specific phosphoglycerate kinase gene *pgk-2* is a recruited retroposon. Mol Cell Biol 7(9):3107-3112.

Caccone A and Powell JR (1990) Extreme rates and heterogeneity in insect DNA evolution. J Mol Evol 30:273-280.

Chambers GK (1988) The *Drosophila* alcohol dehydrogenase gene-enzyme system. In: Caspari EW and Scandalios JG (eds) Advances in Genetics, vol 25. Academic Press Inc, pp 39-107.

Deininger PL and Daniels GR (1986) The recent evolution of mammalian repetitive DNA elements. T.I.G. 2(3):76-80.

Devereux J, Haeberli P and Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12:387-395.

DiNocera PP, Digan ME and Dawid IB (1983) A family of oligoadenylated transposable sequences in *Drosophila melanogaster*. J Mol Biol 168:715-727.

Eickbush TH and Robins B (1985) *Bombyx mori* 28S ribosomal genes contain insertion elements similar to the Type I and II elements of *Drosophila melanogaster*. EMBO J 4:2281-2285.

Felger I and Sperlich D (1989) Cytological localization and organization of dispersed middle repetitive DNA sequences of *Drosophila subobscura*. Chromosoma 98:342-350.

Finnegan DF (1989) F and related elements in *Drosophila melanogaster*. In: Berg DE and Howe MM (ed) Mobile DNA. Washington, DC: American Society for Microbiology, pp 517-519.

Goldberg DA(1980) Isolation and partial characterization of the *Drosophila* alcohol dehydrogenase gene. Proc Natl.Acad Sci USA 77:5794-5798.

Henikoff S (1984) Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene 28:351-359.

Hernández JJ, Vilageliu LL and Gonzàlez-Duarte,R (1988) Functional and biochemical features of alcohol dehydrogenase in four species of the *obscura* group of *Drosophila*. Genetica 77:15-24.

Holmquist R, Cantor C and Jukes T (1972) Improved procedures for comparing homologous sequences in molecules of protein and nucleic acids. J Mol Biol 64:145-161.

Hull R and Will H (1989) Molecular biology of viral and nonviral retroelements. T.I.G. 5(11):357-359.

Jeffs PS and Ashburner M (1991) Processed pseudogenes in *Drosophila*. Proc R Soc London B 244:151-159.

Jukes TH and Cantor CR (1969) Evolution of protein molecules. In: Munro HN (ed) Mammalian protein metabolism, vol. 3. Academic Press, New York, pp 21-132.

Li D and Johnson LF (1989) A mouse thymidilate synthase pseudogene derived from an aberrantly processed RNA molecule. Gene 82:363-370.

Maniatis T, Fritsch EF and Sambrook J (1982) Molecular cloning: A laboratory manual, ed 1. Cold Spring Harbor Laboratory, Cold Spring Harbor New York.

Marfany G and Gonzàlez-Duarte R (1991) The *Adh* genomic region of *Drosophila ambigua*: evolutionary trends in different species. J Mol Evol 32:454-462.

Marfany G and Gonzàlez-Duarte R (1991) The *Drosophila subobscura Adh* genomic region contains valuable evolutionary markers. Mol Biol Evol (in press).

Martin G, Wiernasz D and Schedl P (1983) Evolution of *Drosophila* repetitive-dispersed DNA. J Mol Evol 19:203-213.

Miyata T and Yasunaga T (1981) Rapidly evolving mouse α-globin-related pseudo gene and its evolutionary history. Proc Natl Acad Sci USA 78:5450-453.

Perler F, Efstratiadis A, Lomedico P, Gilbert W, Kolodner R and Dodgson J (1980) The evolution of genes: the chicken preproinsulin gene. Cell 20:555-566.

Pinsker W, Sperlich D (1984) Cytogenetic mapping of enzyme loci on chromosomes J and U of *Drosophila subobscura*. Genetics 108: 913-926

Rogers JH (1983) Retroposons defined. Nature 301:460.

Rogers JH (1985) The origin and evolution of retroposons. Int Rev Cytol 93:187-279.

Sanger F, Nicklen S and Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463-5467.

Schaeffer SW and Aquadro CF (1987) Nucleotide sequence of the *Adh* region of *Drosophila pseudoobscura*: evolutionary changes and evidence for an ancient duplication. Genetics 117:61-73.

Sharp PM and Li W-H (1989) On the rate of DNA sequence evolution in *Drosophila*. J Mol Evol 28:398-402.

Soares M, Schon E, Henderson A, Karathanasis SK, Cate R, Zeitlin S, Chirgwin J and Efstratiadis A (1985) RNA-mediated gene duplication: the rat preproinsulin I gene is a functional retroposon. Mol Cell Biol 5(8):2090-2103.

Sofer W and Martin PF (1987) Analysis of alcohol dehydrogenase gene expression in *Drosophila*. Ann Rev Genet 21:203-225.

Stein JP, Munjaal RP, Lagace L, Lai EC, O'Malley BW and Means AR (1983) Tissue-specific expression of a chicken calmodulin pseudogene lacking intervening sequences. Proc Natl Acad Sci USA 80:6485-6489.

Sullivan DT, Atkinson PW and Starmer WT (1990) Molecular evolution of the alcohol dehydrogenase gene in the genus *Drosophila*. In: Hecht MK, Wallace B and MacIntyre RJ (ed) Evolutionary biology. Plenum, New York, pp 107-147.

Temin HM (1989) Retrons in bacteria. Nature 339:254-255.

Throckmorton L (1975) The phylogeny, ecology, and geography of *Drosophila*. In: King RC (ed) Handbook of genetics, vol 3. Plenum, New York, pp 421-469.

Vanin EF (1985) Processed pseudogenes: characterisitics and evolution. Ann Rev Genet 19:253-272.

Visa N, Marfany G, Vilageliu Ll, Albalat R, Atrian S and Gonzàlez-Duarte R (1991) The *Adh* in *Drosophila*: Chromosomal location and restriction analysis in species with different phylogenetic relationships. Chromosoma 100:315-322.

Wagner M (1986) A consideration of the origin of processed pseudogenes. T.I.G. 2(5):134-137.

Weiner AM, Deininger PL and Efstratiadis A (1986) Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. Ann Rev Biochem 55:631-661.

CAPÍTOL V: "Characterization and evolution of the *Adh* genomic region in *Drosophila guanche* and *Drosophila madeirensis*"

Sotmés a publicació

Characterization and evolution of the Adh genomic region in Drosophila guanche and Drosophila madeirensis.

Gemma Marfany and Roser Gonzàlez-Duarte
Departament de Genètica. Facultat de Biologia. Universitat de Barcelona.

ABSTRACT

We have sequenced the genomic region of the Adh (Alcohol dehydrogenase) gene of Drosophila guanche and Drosophila madeirensis. Two genes, Adh and Adh-dup, whose main features are shared by other Drosophila species, are contained in this region. Inter-species comparisons for this genomic region have been made for these two species with D.subobscura. Our data are in agreement with the phylogenetic relationship of the three species and particularly with the very close proximity of D.madeirensis and D.subobscura. The correct alignment of the noncoding as well as the coding sequences allows us to infer how this region has evolved. The degree of divergence of D.madeirensis and D.subobscura Adh sequences is low and replacement substitutions are as frequent as silent substitutions. In addition, heterogeneous divergence rates for coding and non-coding sequences have been observed. Our data also reveal that a recent evolutionary event is responsible for the branching of the three species.

INTRODUCTION

The Adh gene of Drosophila has been the subject of intense research and the great amount of information that has been gathered on this genomic region makes it particularly suitable to approach evolutionary trends and phylogenetic relationships (for molecular data review, see Sullivan, Atkinson and Starmer, 1990, and references therein). The melanogaster subgroup was the first to be analyzed (Kreitman, 1983; Bodmer and Ashburner, 1984; Cohn, Thompson and Moore, 1984; Eastel and Oakeshott, 1985; Stephens and Nei, 1985; Coyne and Kreitman, 1986) but more recently, data concerning other groups of the genus have been reported and they contribute to a much clearer picture of the evolutionary forces acting upon this genomic region (Schaeffer and Aquadro, 1987; Rowan and Dickinson, 1988; Sullivan, Atkinson and Starmer, 1990; Rowan and Hunt, 1991; Menotti-Raymond, Starmer and Sullivan, 1991; Kreitman and Hudson, 1991; Marfany and Gonzàlez-Duarte, 1991a; Marfany and Gonzàlez-Duarte, 1991b).

The <u>Adh</u> region in the genus <u>Drosophila</u> has been involved in complex genomic rearrangements (Oakeshott et al., 1982; Batterham et al., 1984; Fisher and Maniatis, 1985; Atkinson et al., 1988). Its present structural organization in the <u>Sophophora</u> subgenus appears to be the result of an ancient duplication yielding two different genes: <u>Adh</u> and adjacent to it, <u>Adh</u>-<u>dup</u>, whose function is still unknown (Schaeffer and Aquadro, 1987; Cohn and Moore, 1988; Kreitman and Hudson, 1991 Marfany and Gonzàlez-Duarte, 1991a; Marfany and Gonzàlez-Duarte, 1991b). The existence of two closely-linked genes sharing a common ancestor in a genomic region which encompasses no more than 4-kb provides a unique model for evolutionary studies. The purpose of this study is to analyze this region in a triad of closely related species of the <u>obscura</u> group in order to establish the evolutionary pattern between two recently described species, <u>D.madeirensis</u> and <u>D.guanche</u>, and a well-known member of the same group which we have previously analyzed, <u>D.subobscura</u> (Marfany and Gonzàlez-Duarte, 1991b). This latter is a widespread palearctic species whereas the other two are sibling species, endemic to the Madeira and the Canary Islands, respectively. The origin and phylogeny of <u>D.madeirensis</u> and <u>D.guanche</u> is still controversial as they are nearly indistinguishable by their morphology, although they do not interbreed. In addition, of the <u>obscura</u> species, they show the highest morphological resemblance to the <u>melanogaster</u> species. Phylogenetic relationships have been established for several species of the <u>obscura</u> subgroup including those mentioned, following morphological, cytological and biochemical traits (Buzzati-Traverso and Scossiroli, 1955; Lakovaara and Saura, 1982; Krimbas and Loukas, 1984; Steinemann, Pinsker and Sperlich, 1984; Loukas, Krimbas and Vergini, 1984; Hernández, Vilageliu and Gonzàlez-Duarte, 1988) and electrophoretic data (Cabrera et al., 1983; Loukas, Krimbas and Vergini, 1984; Loukas, Delidakis and Kafatos, 1986). Lately, relationships among species belonging to the <u>obscura</u> subgroup have been reviewed through the analysis of mitochondrial DNA (Latorre, Moya and Ayala, 1988; González et al., 1990) and scDNA divergence (Goddard, Caccone and Powell, 1990). The proposed phylogenetic proximity for the three species under study does not allow much time for divergence, thus providing a new insight into the early dynamics of genomic evolution. Our results clearly show heterogeneous rates of sequence divergence, both in the flanking regions and in coding and noncoding regions of the two genes.

109

MATERIALS AND METHODS

Fly Stock: The source of DNA for the two species was from stocks derived from wild-captured females of D.guanche (Monclús, 1976) and D.madeirensis (Monclús, 1984).

Preparation of Genomic DNA and Construction of Libraries: Total genomic DNA was isolated using the guanidine isothiocyanate method initially described for RNA extraction (Chirgwin et al., 1989), with minor modifications. High molecular weight total DNA was partially digested with MboI, then fractionated by sucrose gradient to obtain DNA fragments of 15-20 kb. The library was constructed by cloning this DNA into the BamHI site of an EMBL4 phage vector. Between 150,000 and 200,000 recombinant plaques were screened for sequences homologous to the D.melanogaster Adh gene. A 2.7 kb HindIII-EcoRI restriction fragment of sAC1 (Goldberg, 1980) containing the complete Adh gene and the immediately adjacent 3' region was used as a probe. It was labeled either by nick-translation or by random-hexamer priming with $[\alpha^{-32}P]dCTP$ (NEN-DuPont) and hybridized to phage DNA on nitrocellulose filters (Hybond-C, Amersham) in 46% formamide at 42°C overnight in the presence of 10% Na dextran sulfate. Nonspecifically hybridized probe was removed with one wash in 2xSSC/0.1%SDS at room temperature for ten minutes, then two washes in 2xSSC/0.1%SDS at 65°C for ten minutes each, and finally two further washes in 1xSSC/0.1%SDS at 65°C for ten minutes each (1xSSC is 0.15M NaCl/0.015M sodium citrate, pH 7.5). After washing, filters were autoradiographed and the DNA from positive recombinant phages was isolated according to Maniatis, Fritsch and Sambrook (1982).

Restriction Analysis of Positive Clones: Positive clones were characterized by restriction mapping as described in Visa et al., (1991).

Nucleotide Sequence Analysis: Clones for sequencing were obtained by restriction digestion with suitable enzymes. These restriction fragments were subcloned into Bluescript +KS, +SK phasmids (Stratagene). Several series of nested deletions were generated using the method of Henikoff (1984) to sequence long regions and across the restriction sites. Sequencing was performed on either single or double stranded DNA using suitable primers according to the dideoxy method of Sanger (Sanger, Nicklen and Coulson, 1977), with $[\alpha^{-35}S]dATP$ (NEN-DuPont) and modified T7 phage polymerase (Sequenase USB, or Sequencing Kit from Pharmacia-LKB). After electrophoresis with TBE buffer (Maniatis, Fritsch and Sambrook, 1982), sequencing gels were fixed, dried and autoradiographed for 72 hours. Each nucleotide was sequenced at least 3 times. Sequences were read and aligned with the 3.98-kb. sequence of D.subobscura (Marfany and Gonzàlez-Duarte, 1991b) and regions with substitutions were carefully confirmed. Alignments were determined by the Sequence Analysis Software Package of the Genetics Computer Group of the University of Wisconsin (Devereux, Haeberli and Smithies, 1984).

RESULTS

Clone characteristics: After the screening of the library, positive clones for each species were isolated and characterized. Restriction maps of the Adh genomic region of D.guanche and D.madeirensis and the cytological location of the gene on polytene chromosomes have already been reported (Visa et al., 1991). The organization of both Adh genes conformed to the general structure reported for

```
              *         *         *         *         *         *         *         *
      TCTAGATTGCATCACTCGTGCCGCCCTACGTTGTGAAGCACCACGCCCTGGACCCCGTTTACTTCGCTTAACCAC  -1033

              *         *         *         *         *         *         *         *
      TGTGGTCGAAGTAGAATCGAACAACGATAAATGGAACATTTGGGAAATGGTCAAGTAATAAATTAAAATAGAAGACAAGA  -953

              *         *         *         *         *         *         *         *
      TTTCTTTAGATATTAACACCTTTCAGTAAATATATAGATAGACAAATATATTTATGATTCACCTCTACGTTTGGTAACCA  -873

              *         *         *         *         *         *         *         *
      CCAATGGGCTATCATTTTACTGTAGCTGTTTTCTGTTTTCCTCTTTTACTTACATGCGTGCATTTTTGCTCTCCCTCTCT  -793

              *         *         *         *         *         *         *         *
      CTTGGGGCACACCCTTGAATCCGCTATTCATGCTCGGATCTGAAGTGGGCTTGGTTTCTCCTGTTCGACGGACAGACAAG  -713

              *         *         *         *         *         *         *         *
      CCGACAAGTAAGCGATTGCTAATGAGTTGCGTCTTAGGGGCGTGTAGTGTGTATTTTTTTAGGCTGAGTCTATACATTAG  -633

              *         *         *         *         *         *         *         *
      ATCTACTCCGCTCCTCCGTCTAGCGCAGAACCCCTTCTCCCCTAGCACACTATCTTCTCTACATGTGTAAATGTGCAAAT  -553

              *         *         *         *         *         *         *         *
      TAAGCCGAAGTTCAAGCCAAGCAAACACTACGAAAGGTCCACACTCTGCTCCTCACACGTTGCTTGACATTCACTGAAGG  -473

              *         *         *         *         *         *         *         *
      TTAAAGCTCTTGCTATCTCTCGCGCTGTGGTCTTGCTATTCTCTCCCACTTTATCAATCCACATTCCCGCTCCCTTGTTC  -393

              *         *         *         *         *         *         *         *
      CACCCAATTACAATATTTGTTAGCGCTCTGCACATTGCGGCAGATGATTTGTGTTTTTCTCCATAAGAGCATATCGACAT  -313

              *         *         *         *         *         *         *         *
      TGAATATTGAAAATATTTTTGAACGTATTCTTCAGTAAAATCTACAGATTAGATATGACATTTCCAGTCTCTCTGATATT  -233

              *         *         *         *         *         *         *         *
      ATTAAATATACCTCAATCAAATTTGATTAATATCGATGCTGGCCACCGTTCAGAAAGTGTATCAAGAGTCTCGACTTTCT  -153

              °    CG  G          TT   GTATTTGC  TT              A              T
              *°   CG  G*         TT  -GTATTTGC  TT          *    A    *         *        *
      AAGCAAACATTTCTTTTTAGTTTTGAATACATT--------ACATTACATTATTACGAATTATTGCAGCGCCGGCGTCGC  -81

                       T G             G                   T       G
              *        T G *      *    G   *       *        T   *   G  *
      GTTTGCGTTTGCGTATACATAGGCGTTGATAGAGGCTCGGCAGAAGTGTATTTAAGGCACCGCACATCGCGAGGACAACG  -1

                 T -              A                           AGTT C
              *  C   *      *     A  GCGGA       *        *   GTT  C*          *
      ATTATTGTCTCAGAA-CAGTTGCCAGGTGCAGTTG-----CCCCAGCATTTCTTCAAATCTACTAAATTGCTCAAGTAAG  74

                 T G C                C C  CA
              *  T   C *     *        *   T  C*          *          *          *
      TAAAGTAACTGAATTCGATGTACAGTCGACAGGCATATCATGCTCGATTCCACTGAGAGAGGATTCGAGCACGGGAAGGT  154

                       G  G  G      C        ACAAAAAT
              *        G  G *G   A   C*       ACAAAAAT  *          *          *
      AAAGTTAATGTTCGATTTTCGATTTCAAAAAACTTCGAGACTGACTTTG--------ACAAAATACTCCAAGTTTCAGTGA  226

                 C        A                A                A      A
              A *        A  *     *        A  *       *    *A      A*          *
      ATTTAAGTGCAATAATCTACCCATCAACCCGACCTTGGACGGTAAAAATAGTACATATCAGCAATCGTTTGACGTATTCC  306

              *          *          *          *          *          C     *
                                                                     C   *
      CTCAGAGCAGTTTATAAAAATAATTCTCTCGATTTGGCGGACTAGGAAATCGTTCTGGCACTTGT-CAATTAATTTGTTT  385

              C     ---------------  C          G     ----
              C   * ---------------  C*         *G    ----  *          *          *
      ATACTTTTTCCTCAAAAAGAATACCGTCTACCCCTGCTCAAAATATGGATGTATGCCCTCACTTTCTGTGTGGTCGTATC  465

                 G        C          *      -  CG
                 G   *    C    *      *     A  CG        *          *          *
      AGGCAGCGCGCGTGTAGACTCTGATAGATCCCCAG-ACGGCCAGTATTTTTCCTCAAGAACCTGAACTCTAAACATAGAC  544

                       A                  --       T                        A
              *        A         *        --   *  T      *     A  *       *A         *
      ATAATTTACTACACTCGCACACATATACAGATGTAGAAGAGAAGTGCCACTGATTAGGCACACGTATTAACATACATTTA  624

              T                              TG
              *          *          *        G *       *          *          *
      CCGGCATAAAACCAAAACAAAGCGATCCGAAACCGAGACGCTGCTAAGACGCAATCGAACGACACGTAATGCGAGAGATA  704
```

```
        GC            G                        G
        *             *            *           *G       *     G      *            *            *
AGAAACGAAAAGCTTCCTTCACGCGAAATAAGCTTTTCGCTTGAAAGAGCTTTTCTTTGAAACGAAATAAATTCCCTATA   784

        A                          A      TT
        A      *            *           * A      *            *            *            *
AATACGAGACTGAAACCAGCAGAAATCTAACAAGCCGTTGAACCATCCTCCCCGATTTCCAGGTCAGGAACTACAAAAGC   864

           C        T                                        T      C G
        * C        *A     *            *            *         T      C G*    *
AAAAGACTCAAAATGTCACTCACAAACAAGAATGTTGTTTTCGTGGCTGGTCTGGGAGGCATTGGCTTAGCACCAGTCG   944
                     MetSerLeuThrAsnLysAsnValValPheValAlaGlyLeuGlyGlyIleGlyLeuAspThrSerAr

              A                                        C*    A                   A    A
        *     A      *            *            *       C*    A      *            *A    *
GGAGTTGGTTAAGCGTGATCTGAAGGTAAGAAAGAGGGAAATCTATTTTCATTGACTCTATGGAAATACTTATCCCAAAT   1024
gGluLeuValLysArgAspLeuLys

              T                           G      G                         A
        *     T      *            *           * G*    *            *         *A     *
CCTCCCCTTATAGAACCTGGTCATCCTGGATCGCATTGACAATCCAGCTGCCATTGCCGAACTGAAGGCAGTCAATCCCA   1104
          AsnLeuValIleLeuAspArgIleAspAsnProAlaAlaIleAlaGluLeuLysAlaValAsnProL

        *            *            *            *            *            *            *
AGGTGACCGTCACCTTCTACCCTTATGATGTGACTGTACCTGTCGCAGAGACCACCAAACTCCTGAAGACCATCTTTGCC   1184
ysValThrValThrPheTyrProTyrAspValThrValProValAlaGluThrThrLysLeuLeuLysThrIlePheAla

        *            *            *            *            *            *            *
CAGATCAAGACCATCGATGTCCTGATAAACGGTGCTGGCATCCTCGACGATCATCAGATTGAGCGTACTATTGCCGTTAA   1264
GlnIleLysThrIleAspValLeuIleAsnGlyAlaGlyIleLeuAspAspHisGlnIleGluArgThrIleAlaValAs

  T                                     C      T
        *            *            *     C      T      *            *            *
CTACACTGGCCTGGTCAACACCACCACAGCCATTCTGGATTTCTGGGACAAGCGCAAGGGCGGCCCAGGTGGCATCATTT   1344
nTyrThrGlyLeuValAsnThrThrThrAlaIleLeuAspPheTrpAspLysArgLysGlyGlyProGlyGlyIleIleC

                                                                         T
        *            *            *            *      T *            *     T *
GCAACATTGGCTCCGTTACCGGTTTTAATGCCATCTACCAGGTGCCCGTTTACTCTGGCAGCAAGGCGGCGGTGGTAAAC   1424
ysAsnIleGlySerValThrGlyPheAsnAlaIleTyrGlnValProValTyrSerGlySerLysAlaAlaValValAsn

            G       TCACA  TC TC               G  A
        *   G       TCTCA  C  TC C      *G     A      *            *         G     *
TTCACCAGCTCCCTGGCGGTAAGCACATC-----TCATAAGTTTCTATTCTCTGAAACTAATTCTTAACTTATCCAAATC   1499
PheThrSerSerLeuAla

G
G
TTTTAGAAACTTGCACCCATCACTGGAGTCACCGCATACACTGTGAATCCGGGCATCACCAAGACCACTCTGGTGCACAA   1579
       LysLeuAlaProIleThrGlyValThrAlaTyrThrValAsnProGlyIleThrLysThrThrLeuValHisLy

                                           C
        *            *            *         C *            *            *            *
ATTCAACTCGTGGCTGGATGTGGAGCCCAGAGTGGCGGAGAAGCTGTTGGAGCATCCCACCCAGACCTCTCAGCAGTGTG   1659
sPheAsnSerTrpLeuAspValGluProArgValAlaGluLysLeuLeuGluHisProThrGlnThrSerGlnGlnCysA

                                           C                                   T
        *            *            *         C *            *            *     T *
CCGAGAACTTTGTCAAGGCCATTGAGCTGAACAAGAATGGTGCTATCTGGAAATTGGACTTGGGAACTCTGGAGCCCATC   1739
laGluAsnPheValLysAlaIleGluLeuasnLysAsnGlyAlaIleTrpLysLeuAspLeuGlyThrLeuGluProIle

                              C      A      TG    -------        C T A
        *            *           * C      *         T      * ------- *  C T A *
ACATGGACCAAGCACTGGGATTCGGGCATCTAAACGGGATATCCGCCCCACAACCCATTCAATGGGACATGGTTCTTAGC   1819
ThrTrpThrLysHisTrpAspSerGlyIleEND

              A                          C
        *     A      *            *         C *            *            *            *
TTTTAGCTTCGTTTTTCCACTCAATTGTTACGTATATATCTACATATGGAAATAAGGCTGATTTGATTCTCTTTAAATGG   1899

                                           C
        *            *            *         A *C      *      ------ *            *
AACCCCGTTTTGAATATGATAATAAAAATTATATTTGAGAAATTTAAACATAAAGCAGATACGCAGTAGCAGTAGCTCTC   1979

                              ------      T           T
        *            *           * ------      *T     * T      *         T *
TTTTAATTAAAAATAGATAAATAATGCCAGTGGCAGTGGCAGGGGCACTGGATTCAGGCCAAGAGCTCTATCGATTTCAC   2059
```

112

```
        G                              A           GA
        G                              A     *     G     *  C     *               *               *
ACAAAAAAACTTAACTTTAGTAATAGAAAAGAAGTCGAGAAAAGCAGCCAAAATAATGTACGATCTGACGGGTAAGCATGT  2139
                                                    MetTyrAspLeuThrGlyLysHisVa


        *           *           *           *           *           *           *           *           * ,
CTGCTATGTAGCTGACTGCGGTGGCATTGCACTGGAGACTAGCAAGGTTCTCATGACCAAGAATATAGCGGTGAGTGCGG  2219
lCysTyrValAlaaspCysGlyGlyIleAlaLeuGluThrSerLysValLeuMetThrLysAsnIleAla


              G           TCTATGTTCTAGTTCCTAAACAAG      TCA      G     T
        A     *           *    TCTGTGTTCTAGTTCCTAAACAAG  *T A    G*    T       *           *
TGTGTGGAGAGTGCAACAGAGA-----------------------TCTCCAGGCTGCTGGACGGTCGAAACTAATTAAG  2275


T           A     G  C CA   TTTTCGCTGACTTTTTCTGTTC              T
T           A  *  G  C A*   TTTTCGCTGACTTTTTCTGTTC        *     T     *           *           *
ATAATGACTTTTTCATTTTATT--------------------GTGGCTACAACTAAGTTTAGTTTTAGAGTGATCTAT  2333


        -                 C                 T        A                       --- GT
        -     *           C*    *           T  *     A*    *           *     --- GT    *
TTTTGCTTAAGGGAAATATTTTCGATTATGGATTATGGCTGCAGAATACAAAAATAGATACAAAGGAACATTCCACTCGT  2413


G                         A                                         G     T
G     *           *       A*    *           *           *           *  G     T      *
CTATTGGTACCTTTTCTAGAAACTGGCAGTCCTCCAGAGCGTGGAAAACCAACCGGCCATCGCTCAGCTACAATCCATTA  2493
           LysLeuAlaValLeuGlnSerValGluAsnGlnProAlaIleAlaGlnLeuGlnSerIleL


        *           *           *           *           *           *           *           *           *
AGCACAGCACACAGATCTTCTTCTGGACCTTCGATGTGACCATGGCCCGACAGGAGATGAAGAAGTACTTCGATGAGGTC  2573
ysHisSerThrGlnIlePhePheTrpThrPheaspValThrMetAlaArgGlnGluMetLysLysTyrPheAspGluVal


                    C                    A
        *          *C    *           *   A*    *           *           *           *           *
ATGGTCCAGATGGACTACATAGATGTACTAATCAATGGGCAACCCTGTGCGATGAGCGGAACATTGATGCCACCATCAA  2653
MetValGlnMetAspTyrIleaspValLeuIleAsnGlyAlaThrLeuCysAspGluArgAsnIleAspAlaThrIleAs


        C                    G
        C  *           *    *G    *           *           *           *          *G        *
TACAAATTTGACCGGAATGATGAACACCGTAGCCACTGTGCTGCCCTACATGGACCGAAAGATGGGCGGATCGGGTGGAT  2733
nThrAsnLeuThrGlyMetMetAsnThrValAlaThrValLeuProTyrMetAspArgLysMetGlyGlySerGlyGlyL


G
G     *           *           *           *           *           *           *           *           *
TGATCGTGAATGTCACCTCTGTCATAGGATTGGATCCATCGCCAGTCTTTTGTGCATACAGTGCCTCAAAGTTTGGTGTG  2813
euIleValAsnValThrSerValIleGlyLeuAspProSerProValPheCysAlaTyrSerAlaSerLysPheGlyVal


                                         T              G              AT G
        *           *           *         T A*    *         T     *    AT G    *
ATTGGGTTCACCAGAAGTCTAGCGGTGAGTCGAAGATCGTTACATCGGCTTTTTGTACTCTAATAAGTATCTTCTCTTTT  2893
IleGlyPheThrArgserLeuAla


                    C                           C
        *          C *    *           *     * C    *           *           *           *
ATATAGGATCCCCTGTATTACACCCAAAATGGTGTGGCTGTAATGGCCGTCTGCTGTGGCCCCACCAAAGTGTTTGTCGA  2973
           AspProLeuTyrTyrThrGlnAsnGlyValAlaValMetAlaValCysCysGlyProThrLysValPheValAs


                    C                               C              G
        *          C *    *           *           * C  *          *  T     G
TCGGGAACTGAATGCCTTTCTGGAGTACGGTCAAACCTTTGCCGATCGCTTGCGTTGTGCACCCTGCCAATCGACTGCCT  3053
pArgGluLeuAsnAlaPheLeuGluTyrGlyGlnThrPheAlaAspArgLeuArgCysAlaProCysGlnSerThrAlaS


T           G                    T     G                                G
T           TG    *           *G  *  T  G     *           *             G
CCTGCGGCCAAAATATAGTAACTGCCATTGAAAGATCGGAAAACGGACAAATTTGGATTGCCGACAAGGGCGGATTGGAA  3133
erCysGlyGlnAsnIleValThrAlaIleGluArgSerGluAsnGlyGlnIleTrpIleAlaAspLysGlyGlyLeuGlu


T                                          T
T           *           *           *      *T    *           *    * G     *
ATGGTGACCCTACACTGGTATTGGCATATGGCCGATCAGTTTTTAAGCTACATGCAGAGCACTGATGACGATAATCAGGA  3213
MetValThrLeuHisTrpTyrTrpHisMetAlaAspGlnPheLeuSerTyrMetGlnSerThrAspAspAspAsnGlnGl


                    A                       ------------------------------------
        *          A *    *           *     ------------------------------------
ACAGTTTGTATCAGGACGGCGATAAGGACTATCGGAAATTATTTGTAGGGCAGCTATGGGAAGAGAAACGGAAATAATAT  3293
uGlnPheValSerGlyArgArgEND


-----
-----     *           *           *           *
CCCATTAAATAAAGTATTAAACGCGACAGAAAA                                                  3326
```

Fig.1.- Sequence of the Adh region of D.guanche. Nucleotide position refers to the inferred adult transcription start site. The aminoacid sequence of Adh and Adh-dup genes is presented under the nucleotide sequence whereas gaps in this aminoacid sequence show intron positions. D.madeirensis sequence is shown immediatly above of that of D.guanche but only substitutions between these species are depicted. Changes between D.guanche and D.subobscura aligned sequences are shown in the second line above. The beginning of sequence alignment is marked by a filled circle (nucleotide position -142). Replaced aminoacids are also shown; underlined: replacements between D.guanche and D.madeirensis; wavy underlined: replacements between D.guanche and D.subobscura; double underlined: replacements between D.guanche and the other two species.

D.melanogaster, with a single functional gene, made up of three exons, whose expression is developmentally regulated by two different promoters.

Sequence comparison of Adh coding regions: A total of 4,467 bp and 3,888 bp of the genomic region containing Adh and Adh-dup was sequenced in D.guanche and D.madeirensis, respectively. Both sequences were aligned with the homologous region of D.subobscura (Marfany and Gonzàlez-Duarte, 1991b) (Fig.1).

The position of the main structural features of the Adh gene of both species with respect to D.subobscura was established by sequence homology. Length differences, due to small deletions/insertions, were only observed in non-coding regions such as the adult leader and introns. Nevertheless, the overall structural comparison supported the close phylogenetic relationships of the three species and the remarkable similarity of D.madeirensis and D.subobscura. In this case, and in contrast with most of the Drosophila species which have been compared to date, a correct sequence alignment was obtained even in upstream and downstream Adh sequences, which bear no relationship to function. The results of the comparisons among the three species are presented in Table 1. Sequence variation in non-coding regions is low and comparable to that in coding regions. The silent positions of the first Adh exon showed the highest degree of divergence (Table 1 and Fig.2).

The Adh coding region has accumulated 16 nucleotide differences since the divergence of D.guanche and D.subobscura, 15 between D.guanche and D.madeirensis and only 4 between D.madeirensis and D.subobscura. These substitutions were not randomly distributed in any codon position as most of them affected the third nucleotide when comparing D.guanche with the other two species (D.guanche vs. D.subobscura: χ^2=9.6 and D.guanche vs. D.madeirensis: χ^2=12.4; df=2 and P<0.001 in both cases). Nevertheless, the substitutions appear to be randomly distributed in any codon position when comparing the sequences of D.madeirensis versus D.subobscura (χ^2=2.06; df=2 and P>0.3). The ratio of transitions to transversions differed with the species compared. As expected for a coding sequence, it was significantly deviated from the random ratio 1:2 when comparing D.guanche with D.subobscura (χ^2=9.1; df=1; P<0.01) and D.guanche vs. D.madeirensis (χ^2=10.8; df=1; P<0.01) but there was no deviation when comparing D.madeirensis and D.subobscura (χ^2=0.1; df=1 and P>0.7).

Silent substitutions are expected to be more frequent than replacement substitutions as they have no effect upon the encoded protein. The number of potentially silent sites is different for each coding sequence: the calculated value

TABLE 1.- Comparative sequence analysis of the Adh genomic region of Drosophila guanche, D.madeirensis and D.subobscura.

Region	b Length (bp)	D.guanche vs. D.subobscura			D.guanche vs. D.madeirensis			D.madeirensis vs. D.subobscura [a]		
		c N. of changes	Percent differences	d Corrected percent divergence	c N. of changes	Percent differences	d Corrected percent divergence	c N. of changes	Percent differences	d Corrected percent divergence
Adh										
Upstream region	111.0	13	11.7	12.7	13	11.7	12.7	2	1.7	1.7
5' noncoding	32.0	2	6.2	6.4	2	6.2	6.5	0	0.0	0.0
Adult leader	69.0	8	11.6	12.1	7	10.1	10.9	3	4.0	4.1
Adult intron	777.0	42	5.4	5.5	37	4.8	4.9	16	2.1	2.1
Larval/adult leader	30.0	1	3.3	3.4	1	3.3	3.4	0	0.0	0.0
Exon 1 Silent	29.7	5	16.8	19.1	5	16.8	19.1	1	3.4	3.4
Replacement	63.3	0	0.0	0.0	0	0.0	0.0	0	0.0	0.0
Intron 1	68.0	4	5.9	6.0	3	4.4	4.5	1	1.5	1.5
Exon 2 Silent	112.3	5	4.5	4.5	6	5.3	5.5	1	0.9	0.9
Replacement	292.7	3	1.0	1.0	1	0.3	0.3	2	0.7	0.7
Intron 2	63.0	9	14.3	15.1	10	15.9	17.8	1	1.5	1.5
Exon 3 Silent	70.0	2	2.9	2.9	2	2.9	2.9	0	0.0	0.0
Replacement	197.0	1	0.5	0.5	1	0.5	0.5	0	0.0	0.0
INTERGENIC REGION	209.0	12	5.7	5.7	10	4.8	4.9	4	2.0	2.1
Adh-dup										
5' noncoding	20.0	0	0.0	0.0	0	0.0	0.0	0	0.0	0.0
Leader	113.0	8	7.1	7.3	8	7.1	7.4	3	2.7	2.7
Exon 1 Silent	24.0	0	0.0	0.0	0	0.0	0.0	0	0.0	0.0
Replacement	72.0	0	0.0	0.0	0	0.0	0.0	0	0.0	0.0
Intron 1	223.0	21	9.4	10.1	20	9.0	9.6	6	2.3	2.3
Exon 2 Silent	107.3	5	4.7	4.7	5	4.7	4.8	0	0.0	0.0
Replacement	297.7	2	0.7	0.7	4	1.3	1.4	2	0.7	0.7
Intron 2	62.0	5	8.1	8.3	6	9.7	10.4	3	4.8	5.0
Exon 3 Silent	85.7	8	9.3	9.7	9	10.5	11.3	3	3.4	3.5
Replacement	253.3	6	2.4	2.4	7	2.8	2.8	1	0.4	0.4
3'noncoding	88.0	4	4.5	4.6	4	4.5	4.7	0	0.0	0.0
SUMMARY										
Noncoding regions	1865.0	129	6.9	7.3	121	6.5	6.8	39	2.1	2.1
Adh silent	212.0	12	5.7	5.9	13	6.1	6.4	2	0.9	0.9
Replacement	553.0	4	0.7	0.7	2	0.4	0.4	2	0.4	0.4
Total	765.0	16	2.1	2.1	15	2.0	2.0	4	0.5	0.5
Adh-dup silent	217.0	13	6.0	6.2	14	6.5	6.7	3	1.4	1.4
Replacement	623.0	8	1.3	1.3	11	1.8	1.8	3	0.5	0.5
Total	840.0	21	2.5	2.5	25	3.0	3.0	6	0.7	0.7

a When comparing D.madeirensis versus D.subobscura, different lengths have been taken into account:for Adh, upstream region-118.0 bp; adult leader-75.0 bp; adult intron-762 bp; exon 2 silent-112.0 bp; exon 2 replacement-293.0 bp; intron 2-68 bp; exon 3 silent-70.7 bp; exon 3 replacement-196.3 bp; intergenic region- 196.0 bp; for Adh-dup, leader-107.0 bp; intron 1-265.0 bp; exon 2 silent-108.0 bp; exon 2 replacement-297.0 bp; exon 3 silent-88.0 bp; exon 3 replacement-251.0 bp and 3'noncoding-48.0 bp. b Comparison has been restricted to the sequences which could be aligned among the three species. c The number of silent sites has been calculated according to Holmquist, Cantor and Jukes, (1972). Insertions and deletions have been simplified to one single event. d Corrected percent divergence has been estimated as $d = 3/4 \ln(1-4/3p)$, where p is the proportion of nucleotide sites that differ between two sequences (Jukes and Cantor, 1969).

for D.guanche Adh was 27.7% and for D.madeirensis 27.8% (that is 212 and 212.4, respectively, out of 765 nucleotide sites)(Table 1). According to our results, 5.9% (12 nt) of these effectively silent sites have changed between D.guanche and D.subobscura, 6.4% (13 nt) between D.guanche and D.madeirensis, and only 0.9% (2nt) between D.madeirensis and D.subobscura. When comparing D.guanche with the other two species, the occurrence of silent vs. replacement substitutions was not at random (D.guanche vs. D.subobscura: χ^2=17.9 and D.guanche vs. D.madeirensis: χ^2=26.0; df=1 and P<0.0001 in both cases), neither was the distribution of silent substitutions among the three exons (D.guanche vs. D.subobscura: χ^2=8.81; df=2; P<0.02 and D.guanche vs. D.madeirensis: χ^2=6.9; df=2 and P<0.05). Nevertheless, silent and replacement substitutions between D.madeirensis and D.subobscura could have been produced randomly (D.madeirensis vs. D.subobscura: χ^2=0.98; df=1; P>0.3) and, in addition, silent substitutions were randomly scattered throughout the three exons (χ^2=2.5; df=1; P>0.2).

Concerning the potential replacement sites (553 for D.guanche and 552.6 for D.madeirensis), 4 differences (0.7%) were detected between D.guanche and D.subobscura and 2 (0.4%) between D.guanche and D.subobscura as well as between D.madeirensis and D.subobscura. These replacement changes were randomly distributed among the three exons whatever sequence was compared (D.guanche vs. D.subobscura: χ^2=0.96; df=2; P>0.5; D.guanche vs. D.madeirensis: χ^2=0.3; df=2; P>0.8; D.madeirensis vs. D.subobscura: χ^2=1.8; df=2; P>0.3). The lower number of replacement substitutions detected would account for the conservation of the biochemical features of the ADH enzyme in these three species (Hernández et al., 1988).

Adh-dup sequence analysis: Downstream from Adh and adjacent to it, there is another gene, the Adh-dup, whose main features appear to be strongly conserved not only in these species but also in the rest of species analyzed. It is made up of three exons and has all the transcriptional and translational regulation signals of an expressing gene: a CCAAT and a TATA box, splicing consensus sequences and a polyadenylation signal, all of them located in the expected positions. The sequence evolution of this potential gene was analyzed in a similar way to Adh.

Some of the regions of this gene showed strong conservation. For instance, the promoter and the first exon showed no differences among the three species. On the other hand, several other regions had diverged at a faster rate, as was the case of the two introns. In addition, an increased number of replacement and silent substitutions produced a higher degree of divergence in the third exon (Table 1 and Fig.2).
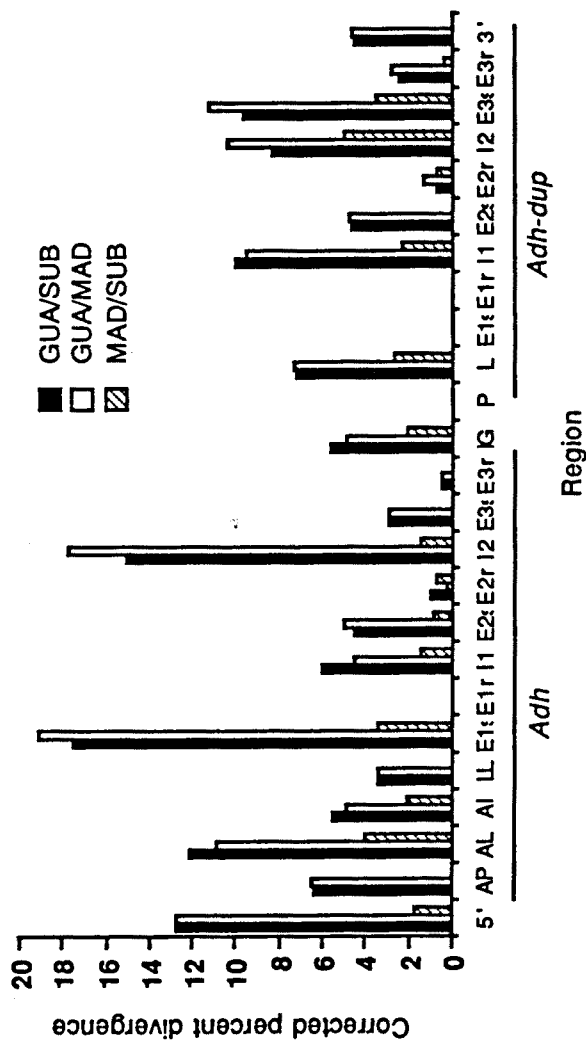
Fig. 2- Corrected percentage of divergence in coding and noncoding regions. See Table 1 for calculations. GUA-*D.guanche*; MAD-*D.madeirensis*; SUB-*D.subobscura*. Exons are divided into silent (s) and replacement (r) positions. 5'-upstream *Adh* noncoding region; AP-adult promoter; AL- Adult leader; AI- adult intron; LL- adult/larval leader; E1, E2 and E3- exons 1, 2 and 3; I1 and I2- introns 1 and 2; IG- intergenic region; 3'-downstream noncoding *Adh-dup* region.

The pattern of nucleotide substitutions ressembled that of the Adh gene. The coding region had accumulated a total of 21 differences since the divergence of D.guanche and D.subobscura, 25 for D.guanche and D.madeirensis and 6 for D.madeirensis and D.subobscura. When comparing D.guanche with the other two species, these substitutions were not randomly distributed among the three positions of a codon, as they tended to accumulate significantly in the third position (D.guanche vs. D.subobscura: χ^2=8.0; df=2; P<0.02 and D.guanche and D.madeirensis: χ^2=6.34; df=2; P<0.05). However, the substitutions between D.madeirensis and D.subobscura appeared to be distributed randomly among all codon positions (χ^2=1.0; df=2; P>0.5). The ratio of transitions to transversions deviated from the random 1:2 only when analyzing the sequence of D.guanche with respect to D.subobscura (χ^2=13.7; df=1; P<0.001) or to D.madeirensis (χ^2=10.7; df=1; P<0.01) but when comparing D.madeirensis with D.subobscura, this ratio did not show a significant deviation (χ^2=0.75; df=1; P>0.5).

Concerning the Adh-dup of D.guanche and D.madeirensis, 25.8% (217.0 nt) and 26.2% (220.0 nt), respectively, of the 840 coding nucleotides were effectively silent (Table 1). Thirteen of these silent sites (6.2%) changed between D.guanche and D.subobscura, 14 (6.7%) between D.guanche and D.madeirensis and only 3 (0.5%) between D.madeirensis and D.subobscura. As was observed when comparing the Adh, silent substitutions were significantly more frequent than replacement substitutions when comparing D.guanche and D.subobscura (χ^2=12.3; df=1; P<0.001) or D.guanche and D.madeirensis (χ^2=10.0; df=1; P<0.01) but this was not the case between D.madeirensis and D.subobscura (χ^2=1.5; df=1; P>0.2). Furthermore, the distribution of silent substitutions deviated significantly from random among the three exons whatever species was compared (D.guanche and D.subobscura: χ^2=5.54; df=1; P<0.1; D.guanche and D.madeirensis: χ^2=6.9; df=1; P<0.05; D.madeirensis and D.subobscura: χ^2=6.0; df=1; P<0.05).

On the other hand, the distribution of the replacement substitutions across the exons among all the species analyzed did not show a significant deviation (D.guanche and D.subobscura: χ^2=5.58; df=2; P>0.05; D.guanche and D.madeirensis: χ^2=4.25; df=2; P<0.1; D.madeirensis and D.subobscura: (χ^2=0.44; df=2; P>0.8). Although it was not statistically significant, there was a clear trend in the divergence of the Adh-dup of these species, so the first exon was clearly preserved while the third exon accumulated all types of substitutions.

Genomic sequence analysis: Comparison of inter-species divergence in a particular genomic region can be approached through the analysis of functional blocks, such as coding or noncoding sequences, silent or nonsilent substitutions, etc. (Fig.2 and Table 1). In another kind of approach, priority can be given to

sequence position within the whole structure, irrespective of the function of this sequence. Following this latter criterion, the compared sequences can be divided into sets with a defined number of nucleotides, and the inter-species differences for each set can then be calculated. The larger the sets are, the coarser the analysis will be. However, if the sets contain a low number of nucleotides, the dilution effect can render the analysis meaningless. To analyze 3.5-kb that were correctly aligned among the three species, we considered sets of 100 nt. The results of the corrected percent divergence against the position is displayed in Fig.3, in which the position of Adh and Adh-dup genes is also depicted.
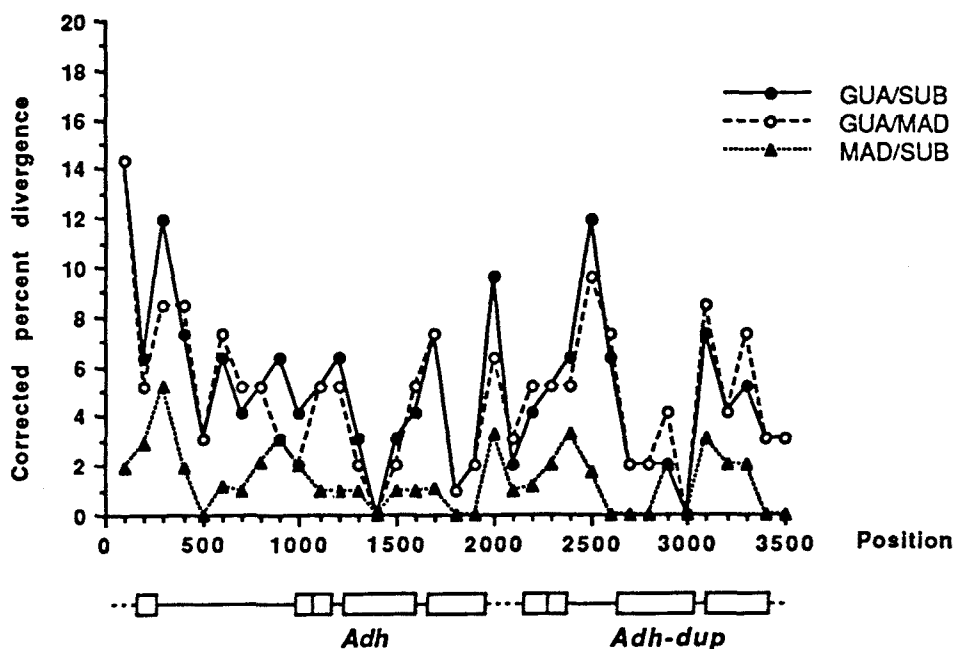
Fig.3- Distribution of nucleotide divergence along the sequence of Adh genomic region. Corrected percentage of divergence is shown for each of the three inter-specific comparisons in sets of 100-bp long. See footnotes of Table I for calculations. The positions of the structural features of Adh and Adh-dup genes are also depicted with respect to the sequence compared. GUA-D.guanche; MAD-D.madeirensis; SUB-D.subobscura.

The inter-species comparisons of the three sequences showed very similar patterns. Peaks of divergence and sequences of decreased or null divergence are generally shared. This was to be expected if divergence peaks corresponded with noncoding regions, and if regions with constrained divergence were within coding regions (Fig.3). However, two constrained sequences deserved attention: one corresponded to the central region of the second Adh exon, and the other was the 3' end of the second Adh-dup exon. Four other sequences showed moderate constraint: an internal region of the adult intron, the carboxy-terminal coding region of Adh, the promoter region of Adh-dup and the overlapping region

between the first intron and second exon of Adh-dup. Sequences of relaxed constraint showing a relatively high peak of divergence were: the adult leader, the region overlapping the second intron and third exon and the trailer of Adh, and the first intron and the third exon of Adh-dup.

DISCUSSION

In order to analyze the Adh genomic region of D.madeirensis and D.guanche we sequenced around 4-kb comprising two genes: the Adh and Adh-dup. We drew inter-species comparisons, and the two species have been compared with D.subobscura (Marfany and Gonzàlez-Duarte, 1991b) to deduce the most remarkable evolutionary trends.

The overall sequencing data confirm the phylogenetic proximity of this triad, D.madeirensis and D.subobscura being closest to each other. Correct alignment has been obtained for coding and for noncoding regions, in introns and boundary sequences of these genes. When more distantly related species are compared, alignment of noncoding regions becomes inaccurate, due not only to their relaxed constraint, but also to the unpredictable insertion/deletion events. In our case, insertions or deletions could be readily positioned along the sequence and most of them have probably arisen from a single event. Moreover, the degree of divergence observed for total noncoding sequences was very similar to that obtained for total silent substitutions in coding regions.

The heterogeneity in the divergence rates within coding and noncoding regions should be emphasized. In particular, within noncoding regions the leaders of both genes and the two introns of Adh-dup showed twice the values of divergence of the adult intron or of the first intron of Adh or even of the intergenic region (Table 1). It could be argued that the intergenic region contains sequences relevant for the regulation of the expression of both genes, and that the adult intron contains sequences for the larval promoter and leader, so they must be under some functional constraint which would impair a rapid evolution. But when the same homologous sequences were compared in other species (Schaeffer and Aquadro, 1987; Marfany and Gonzàlez-Duarte, 1991a), these regions present a degree of divergence comparable to the rest of noncoding sequences. With respect to coding regions, silent substitutions of the first exon of Adh and of the third exon of Adh-dup show the highest number of changes, in clear contrast with the first exon of the Adh-dup, which, although it has a similar size to that of Adh, remains unaltered whatever species is compared (Table 1).

Coding regions considered as a whole are more conserved than non-coding regions. Inter-specific comparisons yield two different patterns: one for D.guanche versus D.subobscura or D.madeirensis, with the types and distribution of substitutions that would be expected for the evolution of functional genes; and the other for D.madeirensis versus D.subobscura. When these comparisons are drawn for functional genes, the number of silent substitutions is usually higher than replacement substitutions. This is reflected in the type of substitutions: the ratio of transitions to transversions is usually significantly deviated towards 1 to 1, and the distribution of nucleotide substitutions is also significantly biased towards third-codon-positions. But these general trends do not apply when D.madeirensis and D.subobscura are compared. The small number of differences observed for these species (4 for Adh and 6 for Adh-dup) could bias the analysis, although the fact that substitutions are randomly distributed between silent and replacement sites, in both genes, does not support it. Further structural information of other genes of these species could illustrate whether this is a general trend in the evolution of their genomes or rather has been produced as an indirect effect of their recent divergence. Furthermore, the effect of genetic drift associated with a founder effect (D.madeirensis is confined in Madeira Islands) could not be discarded. Indeed, D.guanche does not show this trend although it is endemic to the Canary Islands. Another explanation would be that this random distribution constitutes a particular feature of this genomic region in this particular species, as has been suggested for the Amy1 gene of D.pseudoobscura when compared to that of D.melanogaster (Brown, Aquadro and Anderson, 1990).

Analysis of the frequency, type and distribution of substitutions of Adh and Adh-dup among other species of the obscura group (Schaeffer and Aquadro, 1987; Marfany and Gonzàlez-Duarte, 1991a; Marfany and Gonzàlez-Duarte, 1991b) has shown that these genes are evolving with different patterns and this is in agreement with other reports (Martínez-Cruzado et al., 1988; Sharp and Li, 1989). The Adh-dup shows more nucleotide changes than Adh. Both genes have undergone replacement substitutions which are randomly distributed among the three exons whereas silent substitutions are not so evenly distributed: in Adh, they are mainly accumulated in the first exon while in Adh-dup, they accumulate preferently in the third exon.

The comparison of species divergence through the data obtained from functional blocks allows us to approach the evolution of the different parts of a gene. Some regions seem to be more selectively constrained than others, as is the case of the first exon versus the third exon of Adh-dup. Another evolutionary

picture appears when the data are analyzed considering only sequence position, irrespective of function. In this case, coding and noncoding regions can be considered together and then sequence constraint prevails over functional constraint. The first exon of Adh-dup constitutes a good example: maximal functional constraint (no substitutions) in this sequence is coupled to a rapid diverging region (first intron) and both yield a moderate divergence value. On the other hand, sequence conservation in noncoding regions, i.e. the adult intron of Adh, could reveal the location of upstream regulatory sequences as in D.mulleri (Fisher and Maniatis, 1988). Coding sequences may evolve under different constraints as is shown for the second exon of Adh (Fig.3) thus providing further insight into the different contribution of genomic sequences to function (protein domains) and into the evolution of these sequences themselves. Obviously, both approaches complement each other and thus provide a comprehensive evolutionary analysis. Gene conversion has been proposed to account for the heterogeneous divergence rates in the Adh gene: i) in the two genes Adh-1 and Adh-2 of one single species, D.mojavensis (Atkinson et al., 1988) and ii) between two closely related hawaiian species, D.planitibia and D.differens (Rowan and Hunt, 1991). We do not favour this interpretation as a feasible explanation in our species, because it would require improbably complex events.

Total corrected divergence for silent positions in the Adh and Adh-dup sequences gives values of 5.7%/6.2% for D.guanche and D.subobscura, 6.4%/6.7% for D.guanche and D.madeirensis and 0.9%/1.4% for D.madeirensis and D.subobscura, respectively. Assuming a proposed silent substitution rate of 1.28 x 10^{-8} base substitution per silent nucleotide site and year for D.subobscura (Marfany and Gonzàlez-Duarte, 1991b) the approximate time for the branching of these species can be calculated: around 4.5 MY (milllion years) to 5.2 MY for the branching of D.guanche and the pair D.subobscura/D.madeirensis and about 1 MY for the separation of D.madeirensis and D.subobscura. This is in good agreement with the estimated divergence times reported for these species from electrophoretic data (Loukas, Krimbas and Vergini, 1984) and mtDNA analysis (Gonzàlez et al., 1990).

To summarize, we describe the essential features of the evolution of the Adh genomic region in a cluster of closely related species. D.madeirensis and D.subobscura appear to be much closer to each other than to D.guanche although all of them diverged recently in evolutionary time and this is in agreement with previous reports. The inter-specific comparison between D.guanche and the other two species show the expected evolutionary pattern, that is: increased number of

silent versus replacement substitutions, nonrandom ratio of transitions to transversion in coding regions and accumulation of substitutions in the third-codon-position. When D.madeirensis and D.subobscura are compared, random values are obtained for this type of analysis. On the other hand, when prevalence is given to sequence position, heterogeneous divergence rates are found within coding and noncoding sequences and reflect different evolutionary constraints. Both approaches contribute to a clearer picture of the specific nucleotide divergence along the gene.

Sequence availability - Accession Number for Sequence Data in GenBank/EMBL Data Library: X60112 for D.madeirensis and X60113 for D.guanche.

LITERATURE CITED

ATKINSON, P. W., MILLS, L. E., STARMER, W., and D. T. SULLIVAN, 1988 Structure and evolution of the Adh genes of Drosophila mojavensis. Genetics 120: 713-723.

BATTERHAM, P., CHAMBERS, G. K., STARMER, W. T., and D.T. SULLIVAN, 1984 Origin and expression of an alcohol dehydrogenase duplication in the genus Drosophila. Evolution 38: 644-657.

BODMER, M., and M. ASHBURNER, 1984 Conservation and change in the DNA sequences coding for alcohol dehydrogenase in sibling species of Drosophila. Nature 309: 425-430.

BROWN, C. J., AQUADRO, C. F., and W. W. ANDERSON, 1990 DNA sequence evolution of the Amylase multigene family in Drosophila pseudoobscura. Genetics 126: 131-138.

BUZZATI-TRAVERSO, A. A., and R. E. SCOSSIROLI, 1955 The obscura group in the genus Drosophila. Adv. Genet. 7: 47-92.

CABRERA, V. M., GONZALEZ, A. M., LARRUGA, J. M., and A. GULLON, 1983 Genetic distance and evolutionary relationships in the Drosophila obscura group. Evolution 37: 675-689.

CACCONE, A., and J. R. POWELL, 1990 Extreme rates and heterogeneity in insect DNA evolution. J. Mol. Evol. 30: 273-280.

CHIRGWIN, J. M., PRZYBYLA, A. E., MACDONALD, R. J. and W. J. RUTTER, 1979 Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry 18: 5294-5299.

COHN, V. , THOMPSON, M., and G. MOORE, 1984 Nucleotide sequence comparison of the Adh gene in three Drosophilids. J. Mol. Evol. 20: 31-37.

124

COHN, V., and G. MOORE, 1988 Organization and evolution of the alcohol dehydrogenasegene in Drosophila. Mol. Biol. Evol. 5: 154-166.

COYNE, J. A., and M. KREITMAN, 1986 Evolutionary genetics of two siling species, Drosophila simulans and Drosophila sechellia. Evolution 40: 673-691.

DEVEREUX, J. P., HAEBERLI, P., and O. SMITHIES, 1984 A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 12: 387-395.

EASTEL, S., and J. G. OAKESHOTT, 1985 Estimating divergence times of Drosophila species from DNA sequence comparisons (letter to the editor). Mol. Biol. Evol. 2: 87-91.

FISHER, J. A., and T. MANIATIS, 1985 Structure and transcription of the Drosophila mulleri alcohol dehydrogenase genes. Nucleic Acids Res. 13: 6899-6917.

FISHER, J. A., and T. MANIATIS, 1988 Drosophila Adh: a promoter element expands the tissue specificity of an enhancer. Cell 53: 451-461.

GODDARD, K., CACCONE, A., and J. R. POWELL, 1990 Evolutionary implications of DNA divergence in the Drosophila obscura group. Evolution 44: 1656-1670.

GOLDBERG, D. A, 1980 Isolation and partial characterization of the Drosophila alcohol dehydrogenase gene. Proc. Natl. Acad. Sci. USA 77: 5794-5798.

GONZALEZ, A. M. , HERNANDEZ, M., VOLZ, A., PESTANO, J., LARRUGA, J. M., SPERLICH, D. and V. M. CABRERA, 1990 Mitochondrial DNA evolution in the obscura species subgroup of Drosophila. J. Mol. Evol. 31: 122-131.

HENIKOFF, S, 1984 Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene 28: 351-359.

HERNANDEZ, J. J., VILAGELIU, LL., and R. GONZALEZ-DUARTE, 1988 Functional and biochemical features of alcohol dehydrogenase in four species of the obscura group of Drosophila. Genetica 77: 15-24.

HOLMQUIST, R., CANTOR, C., and T. JUKES, 1972 Improved procedures for comparing homologous sequences in molecules of protein and nucleic acids. J. Mol. Biol. 64: 145-161.

JUKES, T. H., and C. R. CANTOR, 1969 Evolution of protein molecules, pp 21-132 in Mammalian protein metabolism, vol 3, edited by H. N. MUNRO, Academic Press, New York.

KREITMAN, M., 1983 Nucleotide polymorphism at the alcohol dehydrogenase locus of Drosophila melanogaster. Nature 304: 412-417.

KREITMAN, M., and R. R. HUDSON, 1991 Inferring the evolutionary histories of the Adh and Adh-dup loci in Drosophila melanogaster from patterns of polymorphism and divergence. Genetics 127: 565-582.

KRIMBAS, C. B., and M. LOUKAS, 1984 Evolution of the obscura group Drosophila species. I. Salivary chromosomes and quantitative characters in D.subobscura and two closely related species. Heredity 53: 469-482.

LAKOVAARA, S., and A. SAURA, 1982 Evolution and speciation in the Drosophila obscura group, pp 2-9 in The genetics and biology of Drosophila, vol 3., edited by M. ASHBURNER, H. L. CARSON, and J. N. SU. Academic Press, London.

LATORRE, A., MOYA A., and F. J. AYALA, 1988 Mitochondrial DNA evolution in the Drosophila obscura group. Mol. Biol. Evol. 5: 717-728.

LOUKAS, M., KRIMBAS, C. B., and Y. VERGINI, 1984 Evolution of the obscura group Drosophila species. II. Phylogeny of ten species based on electrophoretic data. Heredity 53: 483-493.

LOUKAS, M., DELIDAKIS, C., and F. C. KAFATOS, 1986.Genomic blot hybridization as a tool of phylogenetic analysis: evolutionary divergence in the genus Drosophila. J. Mol. Evol. 24: 174-188.

MANIATIS, T., FRITSCH, E. F., and J. SAMBROOK, 1982 Molecular cloning: A laboratory manual, 1st edition. Cold Spring Harbor Laboratory. Cold Spring Harbor, New York.

MARFANY, G., and R. GONZALEZ-DUARTE, 1991a) The Adh genomic region of Drosophila ambigua: evolutionary trends in different species. J. Mol. Evol. 32: 454-462.

MARFANY, G., and R. GONZALEZ-DUARTE, 1991b) The Drosophila subobscura Adh genomic region contains valuable evolutionary markers. Mol. Biol. Evol. (in press).

MARTINEZ-CRUZADO, J. C., SWIMMER, C., FENERJIAN, M. G., and F. C. KAFATOS, 1988 Evolution of the autosomal chorion locus in Drosophila. I. General organization of the locus and sequence comparisons of genes s15 and s19 in evolutionarily distant species. Genetics 119: 663-677.

MENOTTI-RAYMOND, M., STARMER, W. T., and D. T. SULLIVAN, 1991 Characterization of the structure and evolution of the Adh region of Drosophila hydei. Genetics 127: 355-366.

OAKESHOTT, J., G., CHAMBERS, P., EAST, J., GIBSON, and J. BAKER, 1982 Evidence for a genetic duplication involving Adh in D.buzatii and related species. Aust. J. Biol. Sci. 35: 73-84.

ROWAN, R., and W. J. DICKINSON, 1988 Nucleotide sequence of the genomic region coding for alcohol dehydrogenase in Drosophila affinidisjuncta. J. Mol. Evol. 28: 43-54.

ROWAN, R., and J. A. HUNT, 1991 Rates of DNA change and phylogeny from the DNA sequences of the alcohol dehydrogenase gene for five closely related species of hawaiian Drosophila. Mol. Biol. Evol. 8: 49-70.

SANGER, F., NICKLEN, S., and A. R. COULSON, 1977 DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74: 5463-5467.

SCHAEFFER, S. W., and C. F. AQUADRO, 1987 Nucleotide sequence of the Adh gene region of Drosophila pseudoobscura: evolutionary changes and evidence for an ancient gene duplication. Genetics 117: 61-73.

SHARP, P. M., and W-H. LI, 1989 On the rate of DNA sequence evolution in Drosophila. J. Mol. Evol. 28: 398-402.

STEINEMANN, M., PINSKER, W. , and D. SPERLICH, 1984 Chromosome homologies within the Drosophila obscura group probed by in situ hybridization. Chromosoma 91: 46-53.

STEPHENS, J. C., and M. NEI, 1984 Estimation of evolutionary distance between nucleotide sequences. Mol. Biol. Evol. 1: 269-285.

SULLIVAN, D. T., ATKINSON, P. W., and W. T. STARMER, 1990 Molecular evolution of the alcohol dehydrogenase genes in the genus Drosophila, pp 107-147 in Evolutionary Biology, edited by M. K. HECHT, B. WALLACE, and R. J. MACINTYRE. Plenum, New York.

VISA, N., MARFANY, G., VILAGELIU, LL., ALBALAT, R., ATRIAN, S., and R. GONZALEZ-DUARTE, 1991 The Adh locus in Drosophila: chromosomal location and restriction analysis in species with different phylogenetic relationships. Chromosoma 100: 315-322.

CAPÍTOL VI: Apèndix

APÈNDIX.

Aquest apartat recull els resultats experimentals que no han estat inclosos en cap dels articles presentats o que han estat esmentats com a dades no presentades, però que han estat realitzats en el curs d'aquesta Tesi i han aportat informació complementària valuosa a alguns dels aspectes desenvolupats. Aquests experiments obeeixen a quatre objectius diferents: i) conèixer la localització cromosòmica de determinades seqüències genòmiques mitjançant hibridacions *in situ,* ii) ampliar el coneixement estructural de les seqüències de DNA i proteïna mitjançant *Southerns* de DNA genòmic, anàlisis de restricció, seqüenciació parcial i prediccions d'estructura secundària de proteïnes, iii) analitzar l'expressió dels gens *Adh* i *Adh-dup* a les espècies estudiades mitjançant *Northerns* i amplificacions de transcrits per PCR, iv) proposar un model matemàtic per a explicar l'evolució de les retroseqüències.

i) - Localització citològica del gen *Adh* a *Drosophila guanche.*

ii) - Detecció de seqüències homòlogues al gen *Adh* sobre *Southerns* de DNA genòmic total de les espècies: *D.melanogaster, D.ambigua, D.guanche, D.madeirensis* i *D.subobscura.*

 - Detecció de seqüències amb homologia a la repetició invertida situada a 3' de les retroseqüències de *D.subobscura* en *Southerns* de DNA genòmic total de les espècies: *D.melanogaster, D.ambigua, D.guanche, D.madeirensis* i *D.subobscura.*

 - Polimorfisme de restricció a la regió genòmica del gen *Adh* a *D.subobscura.*

 - Caracterització estructural de diferents retroseqüències del gen *Adh* al genoma de *D.subobscura.*

 - Evidència de l'existència de retroseqüències del gen *Adh* als genomes de *D.madeirensis* i *D.guanche.*

 - Prediccions d'estructura secundària per a les proteïnes codificades pels gens *Adh* i *Adh-dup* de *D.mauritiana, D.pseudoobscura, D.ambigua, D.guanche, D.madeirensis* i *D.subobscura.*

iii) - Estudi de l'expressió del gen *Adh* mitjançant transferències *Northern* a les espècies: *D.melanogaster, D.ambigua, D.guanche, D.madeirensis* i *D.subobscura* a diferents estadis del desenvolupament.

 - Estudi de l'expressió del gen *Adh-dup* mitjançant transferències *Northern* i PCR a les espècies: *D.melanogaster, D.ambigua, D.guanche, D.madeirensis* i *D.subobscura* a diferents estadis del desenvolupament.

iv) - Model matemàtic proposat per a l'evolució de les retroseqüències del gen *Adh* a *D.subobscura.*

LOCALITZACIÓ CITOLÒGICA DEL GEN *Adh* A *Drosophila guanche*

En el treball de Visa et al. (1991), es discuteix la localització citològica als cromosomes politènics del gen *Adh* de *D.guanche*. En aquesta espècie, a més del gen funcional, hem trobat d'altres seqüències homòlogues al gen *Adh* en diverses regions genòmiques. A *D.subobscura* i *D.madeirensis* hem descrit, entre d'altres, una banda d'hibridació en el centròmer del cromosoma U. D'altra banda, proves de lligament genètic (Pinsker i Sperlich, 1984) vàren demostrar que precisament en aquesta regió cromosòmica es troba ubicat el gen *Adh* funcional de *D.subobscura*, i això ha permès inferir que en aquesta regió es podria trobar el gen *Adh* funcional de *D.madeirensis*. La gran proximitat filogenètica d'aquestes tres espècies permetia hipotetitzar que a *D.guanche*, el gen funcional també es trobaria en una posició similar.

Per tal de verificar la localització corresponent al gen funcional de *D.guanche*, es van realitzar més hibridacions *in situ* sobre els cromosomes politènics seguint el protocol bàsic exposat a l'apartat 14 dels Materials i Mètodes emprant una sonda homòloga de *D.guanche*, representada a la Fig.17.

Els resultats obtinguts (Fig.18 i Fig.19) confirmen la hibridació prop del centròmer del cromosoma U. En la Fig.18 es mostra el cromosoma U sencer amb la hibridació distal, i a la Fig.19, es pot identificar la hibridació corresponent en la zona propera al centròmer.

La raó de la dificultat d'aquesta localització és que la zona centromèrica d'aquest cromosoma es troba poc compactada i a més, presenta un punt de trencament. Ambdós factors han incidit negativament a l'hora d'identificar la banda d'hibridació a *D.guanche*.

D.guanche G17E
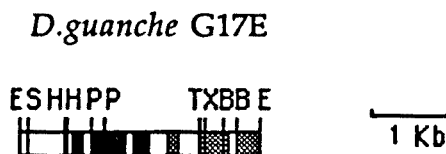
Fig.17.- Mapa de restricció de la sonda emprada per a la localització citològica del gen *Adh* sobre cromosomes politènics de *D.guanche*. Les caixes negres i les ratllades representen, respectivament, la regió codificant del gen *Adh* i del gen *Adh-dup*. E-*Eco*RI, B-*Bam*HI, H-*Hin*dIII, P-*Pvu*II, T-*Pst*I, S-*Sal*I, X-*Xba*I.

Fig.18.- Hibridació de la sonda homòloga del gen *Adh* al cromosoma U de *D.guanche*.
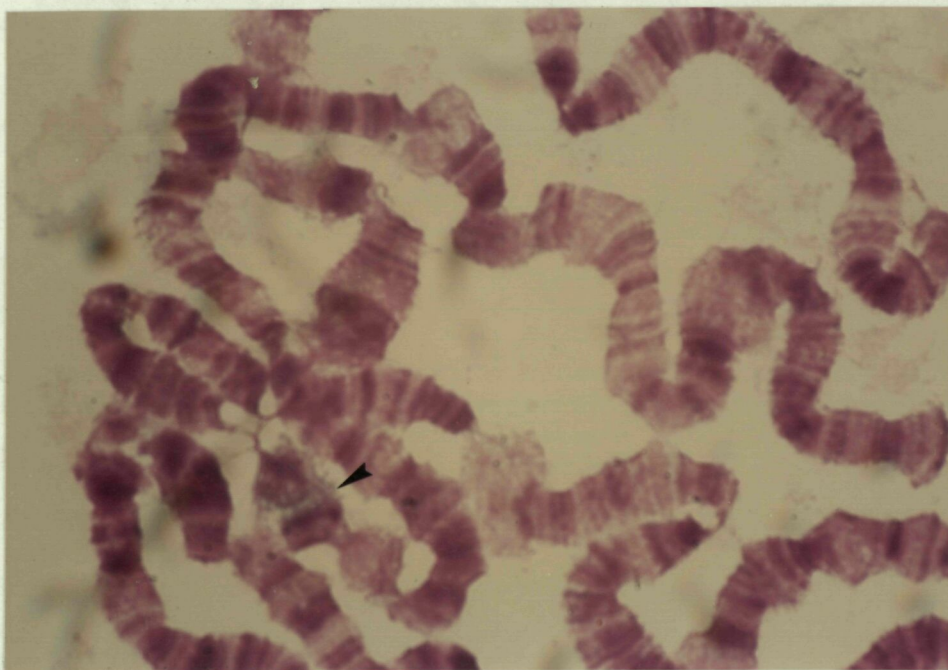
Fig.19.- Localització de l'hibridació de la sonda del gen *Adh* a la zona centromèrica del cromosoma U a *D.guanche*.

DETECCIÓ DE SEQÜÈNCIES HOMÒLOGUES AL GEN *Adh* EN *Southerns* DE DNA GENÒMIC TOTAL DE LES ESPÈCIES: *D.melanogaster, D.ambigua, D.guanche, D.madeirensis* I *D.subobscura.*

Per tal d'identificar la presència de seqüències homòlogues al gen *Adh* en les espècies esmentades, juntament amb *D.melanogaster* com a control, es va realitzar una hibridació sobre un *Southern* de DNA genòmic total digerit amb l'enzim *Eco*RI (segons les condicions especificades a l'apartat 5 de Material i Mètodes). La sonda emprada conté únicament el gen *Adh* de *D.ambigua* (fragment de restricció *Hind*III-*Cla*I de 1.3 kb) (Fig.20).

El resultat de la hibridació es troba a la Fig.21. S'observa la presència d'una única banda a *D.melanogaster* (coincident amb la mida descrita de 4.5 kb), a *D.ambigua* i a *D.subobscura*, encara que en aquesta última espècie la intensitat de la banda és superior i la seva mida pot superar les 20 kb. A *D.madeirensis* es distingeixen dues bandes i a *D.guanche*, tres.

D'aquests resultats es pot deduir que donat el mapa de restricció del gen funcional a *D.madeirensis* i a *D.guanche*, es detecta la presència de més d'una seqüència amb homologia a la sonda (que conté només el gen *Adh*). Per altra part, a *D.subobscura*, donat que tenim evidència directa de la presència de seqüències addicionals homòlogues a l'*Adh* i de localització genòmica diferent, deduim que no hi ha només una banda, sinó diverses, de mida molecular molt elevada, i que les condicions de resolució del gel d'agarosa no permeten la seva separació.
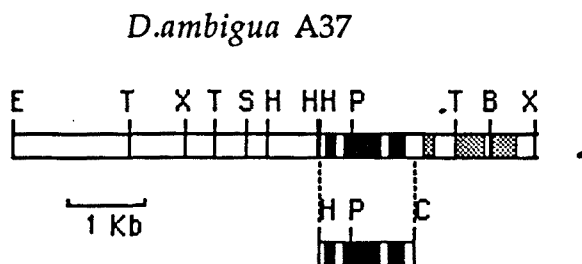
D.ambigua A37

Fig.20.- Mapa de restricció de la regió genòmica del gen *Adh* a *D.ambigua*. La sonda emprada per a la detecció del gen *Adh* sobre *Southerns* genòmics, es troba assenyalada. Les caixes negres i les ratllades representen, respectivament, la regió codificant del gen *Adh* i del gen *Adh-dup*. E-*Eco*RI, B-*Bam*HI, H-*Hind*III, P-*Pvu*II, T-*Pst*I, S-*Sal*I, X-*Xba*I, C-*Cla*I.
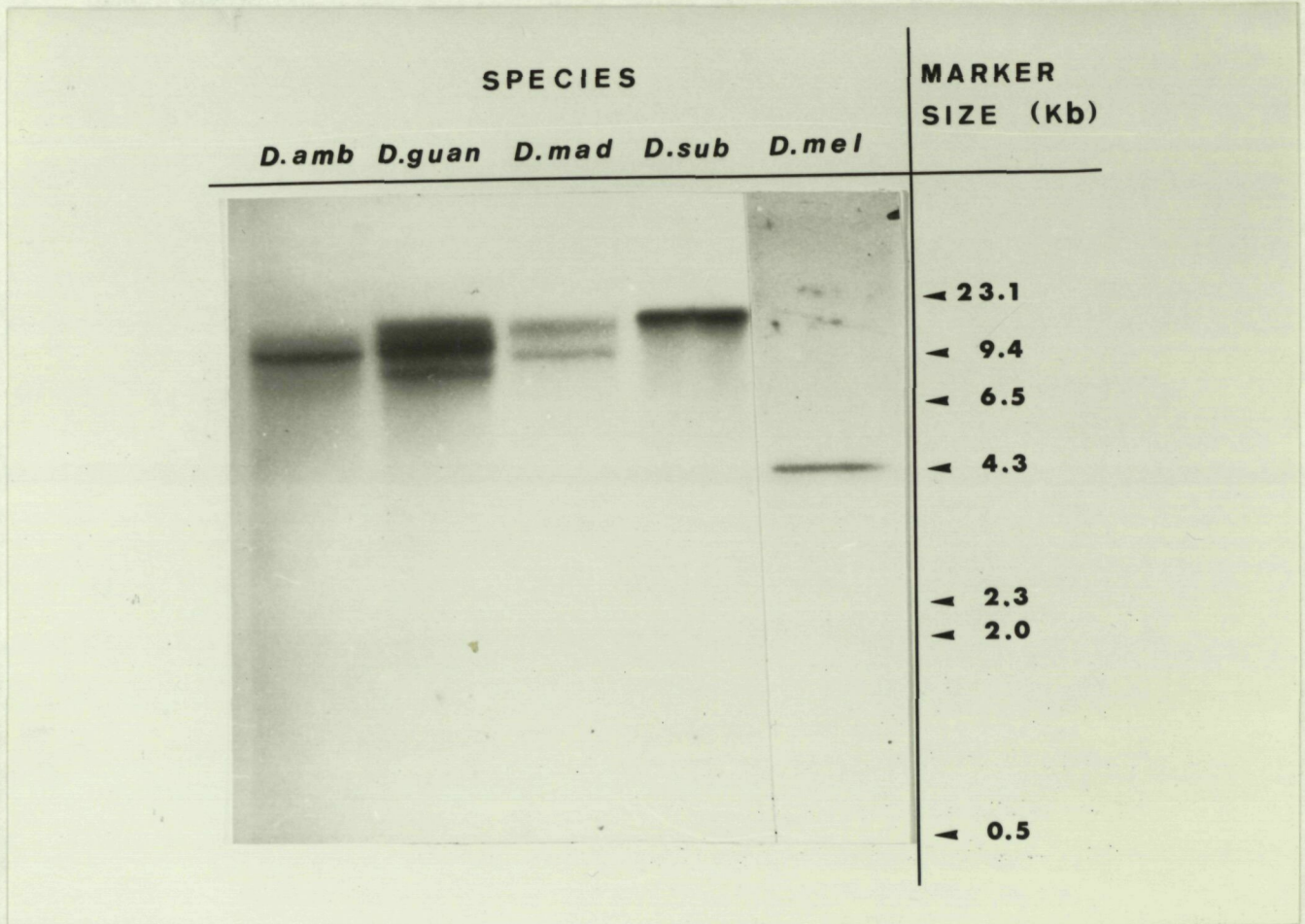
Fig.21.- Detecció de bandes amb homologia al gen *Adh* sobre *Southerns* genòmics de: *D.amb-D.ambigua, D.guan-D.guanche, D.mad-D.madeirensis, D.sub-D.subobscura* i *D.mel-D.melanogaster*.

DETECCIÓ DE SEQÜÈNCIES AMB HOMOLOGIA A LA REPETICIÓ INVERTIDA SITUADA A 3' DE LES RETROSEQÜÈNCIES A *D.subobscura* EN *SOUTHERNS* DE DNA GENÒMIC TOTAL DE LES ESPÈCIES: *D.melanogaster, D.ambigua, D.guanche, D.madeirensis* I *D.subobscura*.

A la Fig.22 es presenta la regió situada a 3' de les retroseqüències i que s'ha utilitzat com a sonda en aquest experiment: procedeix d'un fragment de restricció *Bam*HI-*Bam*HI, mesura 1.1 kb i conté un dels terminals de la repetició invertida.

Els resultats de la hibridació es presenten a la Fig.23, i confirmen el patró de bandes múltiple característic del DNA mitjanament repetit en el genoma de *D.subobscura*. Un altre resultat força interessant és la presència d'hibridació clara a les espècies *D.guanche* i *D.madeirensis*, que posa de manifest la presència de seqüències homòlogues a aquesta repetició invertida en les espècies esmentades. Amb una menor intensitat d'hibridació, i per tant, d'homologia, es distingeixen bandes a *D.ambigua*. Això contrasta amb el fet que a *D.melanogaster* no es pot observar hibridació.

Es pot, doncs, inferir que la seqüència de repetició invertida forma part d'una família de DNA moderadament repetit en el genoma, present no sols a *D.subobscura*, sinó també a *D.madeirensis* i a *D.guanche*, probablement amb menys similitud a *D.ambigua*, però totalment absent a *D.melanogaster*.
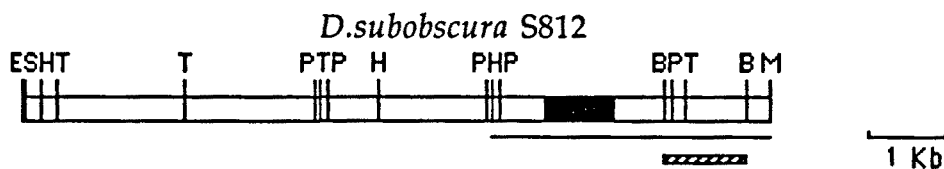
D.subobscura S812

Fig.22.- Mapa de restricció de la regió genòmica de la retroseqüència λS812 de *D.subobscura*. La regió seqüenciada està subratllada i la sonda que conté la repetició invertida està indicada. La caixa negra representa la regió codificant del gen *Adh* E-*Eco*RI, B-*Bam*HI, H-*Hind*III, P-*Pvu*II, T-*Pst*I, S-*Sal*I, X-*Xba*I.

Fig.23.- Detecció de bandes amb homologia a la sonda que conté la repetició invertida sobre *Southerns* genòmics de: M-*D.melanogaster*, A-*D.ambigua*, G-*D.guanche*, D-*D.madeirensis* i S-*D.subobscura*.
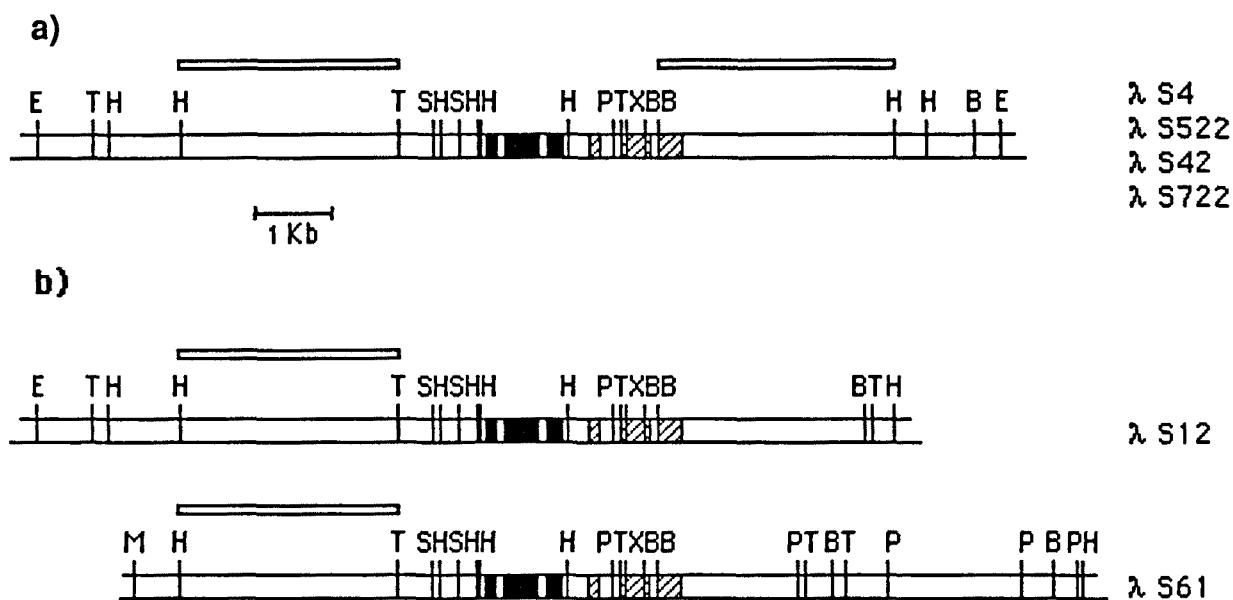
Fig.24.- Polimorfisme de restricció a la regió genòmica del gen *Adh* de*D.subobscura*. Les caixes negres i les ratllades representen, respectivament, la regió codificant del gen *Adh* i del gen *Adh-dup*. Les zones amb homologia d'hibridació amb la sonda que conté la repetició invertida (Fig.22) estan assenyalades. E-*EcoRI*, B-*BamHI*, H-*HindIII*, P-*PvuII*, T-*PstI*, S-*SalI*, X-*XbaI*.

POLIMORFISME DE RESTRICCIÓ A LA REGIÓ GENÒMICA DEL GEN *Adh* A *D.subobscura.*

En el crivellatge de la llibreria genòmica de *D.subobscura* per seqüències homòlogues al gen *Adh*, es van aïllar diversos clons positius que van ésser analitzats per restricció per evitar duplicats. Com ja s'ha esmentat, hi havia dos tipus de clons, uns que contenien el gen funcional i d'altres, que contenien retroseqüències del gen *Adh*. Dins dels clons amb el gen *Adh* funcional, es va detectar que hi havia polimorfisme en certes dianes de restricció. Quan es va realitzar un mapa de restricció més acurat d'aquests clons i es va situar el gen *Adh* i el gen *Adh-dup* dins del mapa, es va observar que aquest polimorfisme de restricció es trobava delimitat a les zones de la regió situada a 3' del conjunt *Adh/Adh-dup* (Fig.24).

Aquest polimorfisme de restricció no es deu a un canvi de dianes produit per mutacions puntuals sinó que sembla degut a insercions o delecions de fragments de DNA. Això es basa en els següents fets:

a) No es perd una única diana i es conserven les altres, sinó que la variació afecta a diverses dianes a la vegada.

b) Hi ha un fet més concloent: les diferències d'hibridació que es detecten entre els clons. Utilitzant la sonda de 1.1 kb *Bam*HI-*Bam*HI que conté la repetició invertida per hibridar en una transferència de DNA, digerit amb enzims de restricció, de fags que contenen seqüències homòlogues a l'*Adh*, es troba hibridació positiva no només a les retroseqüències (com ja està comentat), sinó també a dues zones properes al gen *Adh* funcional, situades una a 5' i l'altre a 3' del conjunt *Adh/Adh-dup*. En aquests clons de restricció polimòrfica, aquesta hibridació a 3' no hi és present (Fig. 24). Per tant, confirma la presència de diferents seqüències en aquesta regió, depenent del clon aïllat.

CARACTERITZACIÓ ESTRUCTURAL DE DIFERENTS RETROSEQÜÈNCIES DEL GEN *Adh* AL GENOMA DE *D.subobscura.*

En el crivellatge de la llibreria genòmica de *D.subobscura* per seqüències homòlogues al gen *Adh,* es van aïllar diversos clons positius que en ésser caracteritzats per restricció i parcial seqüènciació van mostrar trets propis de retroseqüències, és a dir, no presentaven seqüències promotores ni introns. Aquests clons es van anomenar: λS120, λS114, λS135, λS812. L'anàlisi inicial d'aquests clons va permetre agrupar-los en dos patrons diferents de restricció, els proporcionats per λS135 i per λS812. Aquests dos clons van ésser caracteritzats amb més detall (Capítol IV). A la Fig.25 es presenta el mapa de restricció de totes les retroseqüències aïllades, comparades amb la regió genòmica del gen funcional.

Per altra part, l'estructura d'aquestes retroseqüències és molt similar, ja que presenten també seqüències homòlogues a la repetició invertida, com s'ha detectat per hibridació emprant la sonda de 1.1 kb *Bam*HI-*Bam*HI del clon λS812 que conté la repetició invertida (Fig.25). Els clons λS120, λS114, λS135 poden contenir la mateixa regió genòmica, si bé amb dianes polimorfiques, ja que els mapes de restricció permeten la seva agrupació, inclús en regions no relacionades directament amb les retroseqüències.
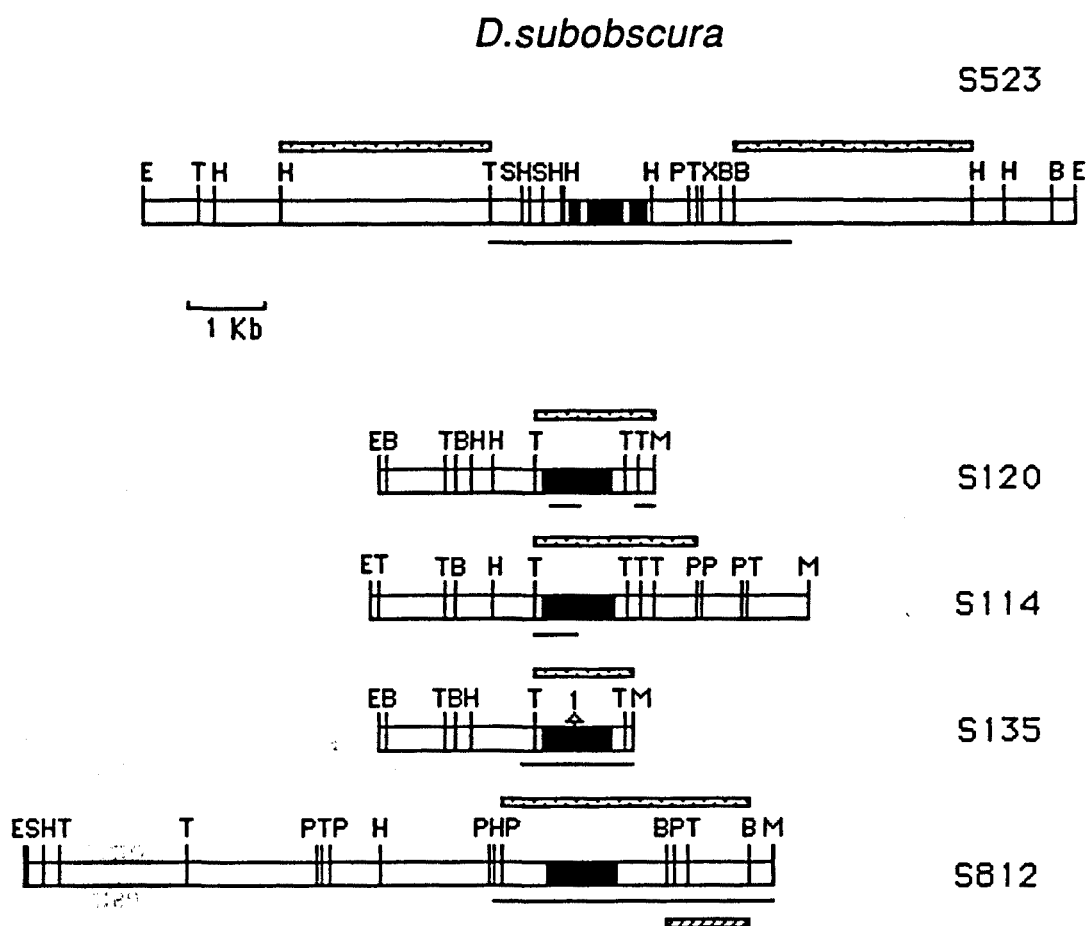
Fig.25.- Mapa de restricció de les seqüències amb homologia a l'*Adh* del genoma de*D.subobscura*. S523- Regió genòmica del gen*Adh* funcional. S120, S114, S135 i S812- Diferents clons que contenen retroseqüències. Les caixes negres representen les regions codificants d'*Adh* . Les zones seqüenciades estan subratllades. La caixa ratllada indica la sonda que conté la repetició invertida mentre que les regions que hibriden amb aquesta sonda estan assenyalades amb caixes puntejades (B-*Bam*HI, E-*Eco*RI, H-*Hind*III, P-*Pvu*II, T-*Pst*I, S-*Sal*I, X-*Xba*I, M-*Mbo*I).

EVIDÈNCIA DE L'EXISTÈNCIA DE RETROSEQÜÈNCIES DEL GEN *Adh* ALS GENOMES DE *D.madeirensis* I *D.guanche*.

Tres tipus d'evidències indirectes suggereixen l'existència de seqüències addicionals -retroseqüències- del gen *Adh* als genomes d'espècies properes a *D.subobscura*: *D.madeirensis* i *D.guanche*.

a) Hibridacions *in situ*, que han permès detectar diverses localitzacions cromosòmiques de seqüències homòlogues al gen *Adh*.

b) *Southerns* de DNA genòmic total i la comparació del patró de bandes obtingut amb l'esperat per al gen *Adh* a cada espècie. Es important assenyalar que no es troba cap diana *Eco*RI dins del gen *Adh*.

c) Experiments de PCR sobre DNA genòmic total amb dos encebadors corresponents a seqüències de dos exons diferents i que, per tant, permeten diferenciar la presència del gen funcional de les retroseqüències que procedeixen de transcrits processats (Fig.26 i Fig.27)

També hem obtingut evidències directes de l'existència de retroseqüències en aquestes espècies. Durant el crivellatge de les llibreries genòmiques de *D.guanche* i *D.madeirensis*, vam aïllar diversos clons amb seqüències homòlogues al gen *Adh*, a més dels que contenien el gen funcional. L'anàlisi de restricció d'aquests clons es presenta a la Fig.28. Les retroseqüències d'aquests clons no estan complertes degut al procediment d'obtenció de la llibreria, però la informació que es pot extreure d'aquests clons permet assegurar que es tracta de retroseqüències, ja que en cada cas, s'ha seqüenciat per sobre d'on es troba un intró en el gen funcional i en els dos clons apareix perfectament escindit, com és característic del cDNA procedent de tràlscrits processats. La seqüència de què es disposa de cada clon s'ha comparat amb la corresponent seqüència del gen funcional (Fig.29 i Fig.30), i també s'han alineat parcialment les retroseqüències de les tres espècies (Fig. 31; Taules III, IV i V).

Existeixen també proves d'hibridació que indiquen la presència en aquestes retroseqüències d'una seqüència homòloga a la repetició invertida que es troba a 3' de les retroseqüències de *D.subobscura*.

Tots aquests resultats, apunten cap al fet que el fenomen de retrotranscripció

del gen *Adh* i inserció en el genoma en aquestes tres espècies tan relacionades filogenèticament es va produir abans que es produís la seva divergència, però després que el seu ancestre comú es separés dins de la radiació del subgrup *obscura*.
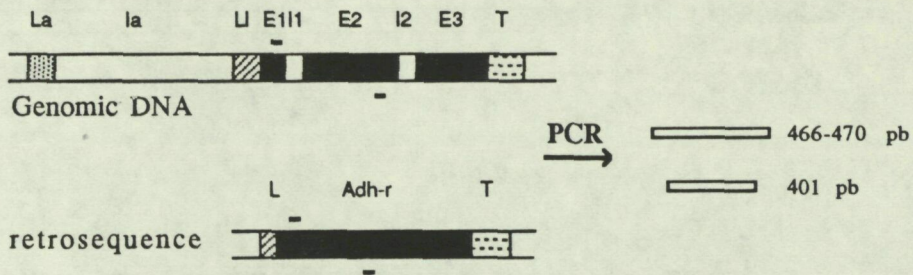
Fig.26.- Representació gràfica dels dos encebadors emprats per a la reacció de PCR sobre DNA genòmic ,i dels possibles productes d'amplificació.
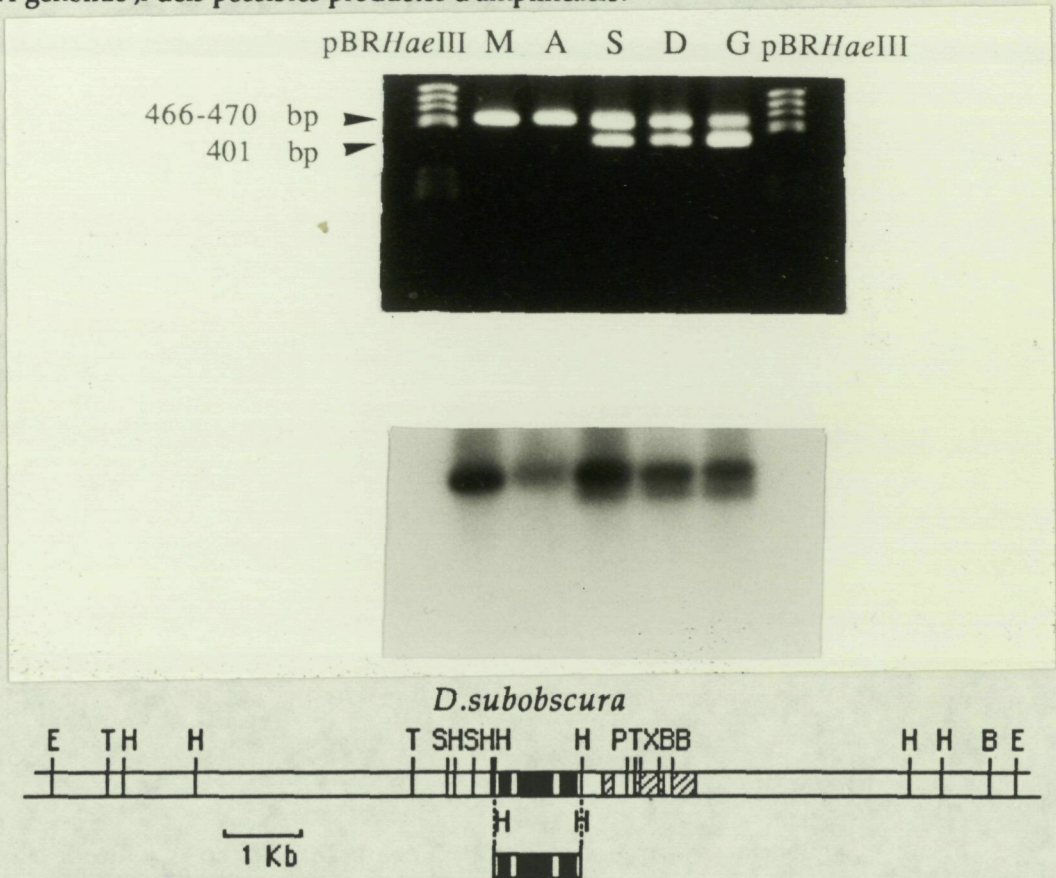
Fig.27.- a) Resultat de la reacció de PCR sobre DNA genòmic de les espècies: M-*D.melanogaster*, A-*D.ambigua*, G-*D.guanche*, D-*D.madeirensis* i S-*D.subobscura* i de la hibridació amb una sonda de *D.subobscura* que conté només el gen *Adh*.
b) Mapa de restricció de la regió genòmica del gen *Adh* de *D.subobscura* i de la sonda emprada en aquesta hibridació. Les caixes negres i les ratllades representen, respectivament, la regió codificant del gen *Adh* i del gen *Adh-dup*. E-*Eco*RI, B-*Bam*HI, H-*Hind*III, P-*Pvu*II, T-*Pst*I, S-*Sal*I, X-*Xba*I.

142

D.madeirensis

E T H   H   T SHSHH   H TXBB   H   B

D3EB

⊢ 1 Kb ⊣

retrosequence

M   TB   T   PH   H PPT   T   H THSE

D4b

D.guanche

X   ESHHPP   TXBB   HBH T   T P

G17

⊢ 1 Kb ⊣

retrosequence

M H   H S H T   TTHTH   E

G4
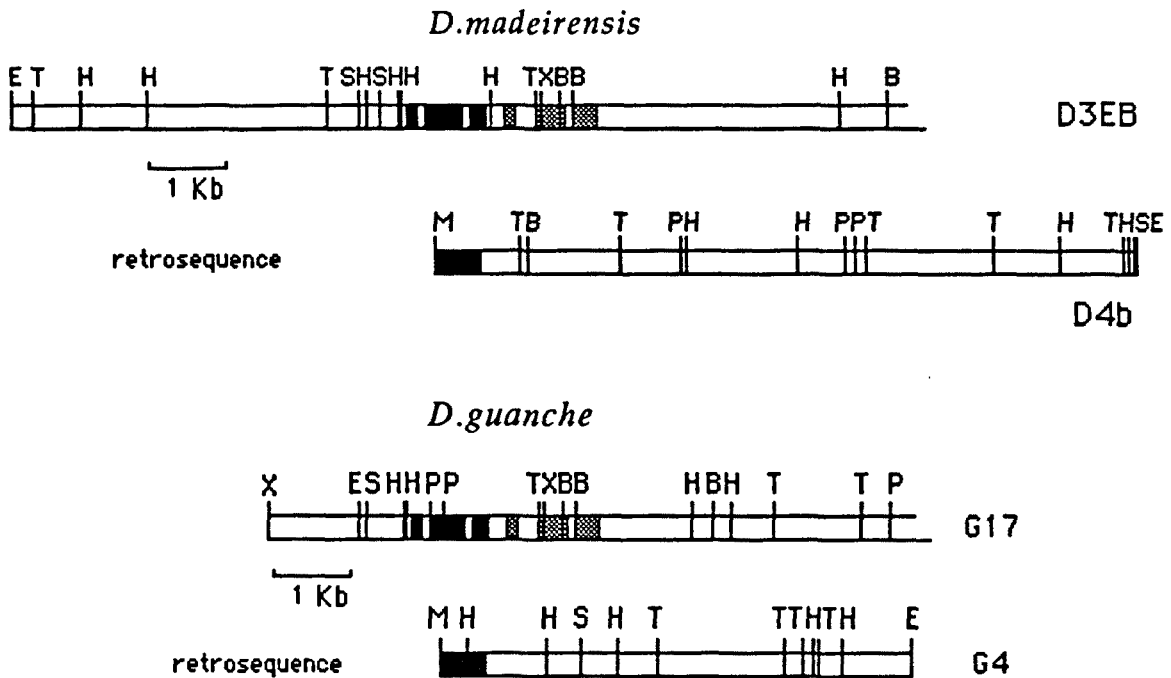
Fig.28.- Mapa de restricció de les retroseqüències del gen *Adh* comparades amb el respectiu gen funcional a les espècies *D.madeirensis* i *D.guanche*. Les caixes negres i les ratllades representen, respectivament, la regió codificant del gen *Adh* i del gen *Adh-dup*. Les dos zones subratllades indiquen la posició de la zona seqüenciada (Fig.30 i Fig.31). E-*EcoRI*, B-*BamHI*, H-*HindIII*, P-*PvuII*, T-*PstI*, S-*SalI*, X-*XbaI*.

```
GT....AG
      T                               CG
    *            *            *            *            *            *            *            *
GATCTGAAGAACCTGGTCATCCTGGATCGCATTGACAATCCAGATGCCATTGCCGAACTGAAGGCAATCAATCCCAAGGT      85-165

   A    C    C ACCC         T    G         A              AC       G    A
   A    C  *  C CCC  *         *            *              *            *    *   AA  *
GGCCGTTACCTTATTTAATTATGATGTGACTGCACCTCTCGCAGAGACCGCCAAACTCCTGAAGTGCATCTTTTCCCAGT         245

      A      G         T         C              T
      *            *         *       *G          *       *            *            *
TCAAGACCGTCGATATCCTGATCAACGGTGCGGGCATCCTAGACGATCATCAGATCGAGCGTACTATTGCCGTTAACTAC         325

      *
ACTGGC                                                                                331
```

Fig.29.- Seqüència de la retroseqüència aïllada a *D.madeirensis* alineada amb la retroseqüència de *D.subobscura* S812 (línia immediatament superior) i el gen funcional de *D.madeirensis* (línia superior). S'ha assenyalat la posició de l'intró en el gen funcional i els canvis nucleotídics detectats entre les tres seqüències. La numeració indica la posició dels nucleòtids des de l'inici de traducció.

```
AG           CC      T   A                    G              A          A
A        *   CC*     T   *        *       C   *       *   G  A *        *
GATCAATCCCAAGGTGAAGGTCACCGTCTTCCCTTATGATGTGACTGTACCTCTCGCAGAGACCACCGAACTCCTGAAGT  150-230

C   G             A           A                C          CAG  T       T
    *AA  T     *       A       *       G   *       *      CAG  *       T
GCATCTTTTCCCAGATCAAGACCGTCGATGTCCTGATCAACGGTGCTGGCATCCTGGACGATCATGCCATCGAGCGTACC       310

              C                     TC   T              C   G
    *         * C       *       T   *    TC   *C        *   C    *           *
ATTGCCGTTAACTACACTGGCTTGGTCAACACCACCACAGCCATGATGGAGTTCTGGGACAAGCGGAAGTGCGGCCCAGG       390

                        T
    *         *         *   T    *   T       *           *           *           *
TGGCATCATTTGCAACATTGGCTCCGTTACCGGTTTCAATGCCATCTACCAGGTGCCCGTTTACTCTGGCAGCAAGGCGG       470
                        GT....AG
    A                           A
    *         *         *   A            *
CGGTGGTTAACTTCACCAGCTCCCTGGCGAAGCTT                                                 506
```

Fig.30.- Seqüència de la retroseqüència aïllada a *D.guanche* alineada amb la retroseqüència de *D.subobscura* S812 (línia immediatament superior) i el gen funcional de *D.guanche*(línia superior). S'ha assenyalat la posició de l'intró en el gen funcional i els canvis nucleotídics detectats entre les tres seqüències. La numeració indica la posició dels nucleòtids des de l'inici de traducció.

```
                A   C     C   CCC                                               *
G                   AAG   C   G C CCC              T              A   G
AATCAATCCCAAGGTGGCCGTTACCTTATTTAATTATGATGTGACTGCACCTCTCGCAGAGACCGCCAAACTCCTGAAGT  150-230

    *AA       *                 *       *       G   *       *
    A                   G                   T            G           GCC         C
GCATCTTTTCCCAGTTCAAGACCGTCGATATCCTGATCAACGGTGCGGGCATCCTAGACGATCATCAGATCGAGCGTACT       310

    *         *         *
ATTGCCGTTAACTACACTGGC       331
```

Fig.31.- Alineament parcial de les retroseqüències D4b de *D.madeirensis*, G4b de *D.guanche* (línia immediatament superior) i S812 de *D.subobscura* (línia superior). S'han assenyalat els canvis nucleotídics detectats entre les tres retroseqüències. La numeració indica la posició dels nucleòtids des de l'inici de traducció.

TAULA III.- Anàlisi comparatiu preliminar.de les seqüències de les retroseqüències de *D.madeirensis* (Mr) i *D.subobscura* S812(Sr) amb el gen *Adh* funcional de *D.madeirensis* (Df).

[2]Longitud(pb)	Mr vs. Mf			Mr vs. Sr			Sr vs. Mf		
	[3]NC	PD	[4]CPD	[3]NC	PD	[4]CPD	[3]NC	PD	[4]CPD
246	22	8.9	9.5	8	3.3	3.3	18	7.3	7.7

TAULA IV.- Anàlisi comparatiu preliminar.de les seqüències de les retroseqüències de *D.guanche*(Gr) i *D.subobscura* S812(Sr) amb el gen *Adh* funcional de *D.guanche*(Gf).

[2]Longitud(pb)	Gr vs. Gf			Gr vs. Sr			Sr vs. Gf		
	[3]NC	PD	[4]CPD	[3]NC	PD	[4]CPD	[3]NC	PD	[4]CPD
355	28	7.9	8.3	25	7.0	7.4	22	6.2	6.5

TAULA V.- Anàlisi comparatiu preliminar.de les seqüències de les retroseqüències de *D.madeirensis* (Mr), *D.guanche*(Gr) i *D.subobscura* S812(Sr).

[2]Longitud(pb)	Mr vs. Gr			Mr vs. Sr			Gr vs. Sr		
	[3]NC	PD	[4]CPD	[3]NC	PD	[4]CPD	[3]NC	PD	[4]CPD
181	20	11.0	12.0	8	4.4	4.6	16	8.8	9.4

PREDICCIONS D'ESTRUCTURA SECUNDÀRIA PER A LES PROTEÏNES
CODIFICADES PELS GENS *Adh* I *Adh-dup* DE *D.mauritiana D.pseudoobscura,*
D.ambigua, D.guanche, D.madeirensis I *D.subobscura* .

La comparació de les prediccions d'estructures secundàries de proteïnes és una
eina valuosa per a l'avaluació de les diferències en les seqüències
aminoacídiques, ja que hi ha canvis de residus que són conservatius, mentre que
d'altres tenen un efecte més evident en la molècula que queda reflexat en
aquestes prediccions.

En primer lloc es mostren les prediccions de seqüència per al gen *Adh* de les
espècies esmentades (Fig.32 a Fig.37). No s'observen diferències relevants entre
espècies. Es posa de manifest l'elevat grau d'hidrofobicitat de la proteïna,
destacant només dues petites zones hidrofíliques.

Per últim, es mostren les prediccions d'estructura secundària per als gens *Adh-
dup* de les espècies esmentades (Fig.38 a Fig.43). La diferència observada amb el
perfil presentat per a l'*Adh-dup* de l'espècie *D.madeirensis* no correspón a una
diferència real sinó que es deu a una deformació per expansió dels gràfics en la
resta d'espècies, per raons tècniques del programa PLOTSTRUCTURE. Es tracta
també de proteïnes altament hidrofòbiques, amb molt poques zones
hidrofíliques. Les diferències més grans entre espècies es concentren en l'extrem
carboxi-terminal de la proteïna

Fig.32.- Predicció d'estructura per a la proteïna del gen *Adh* de *D.mauritiana*.
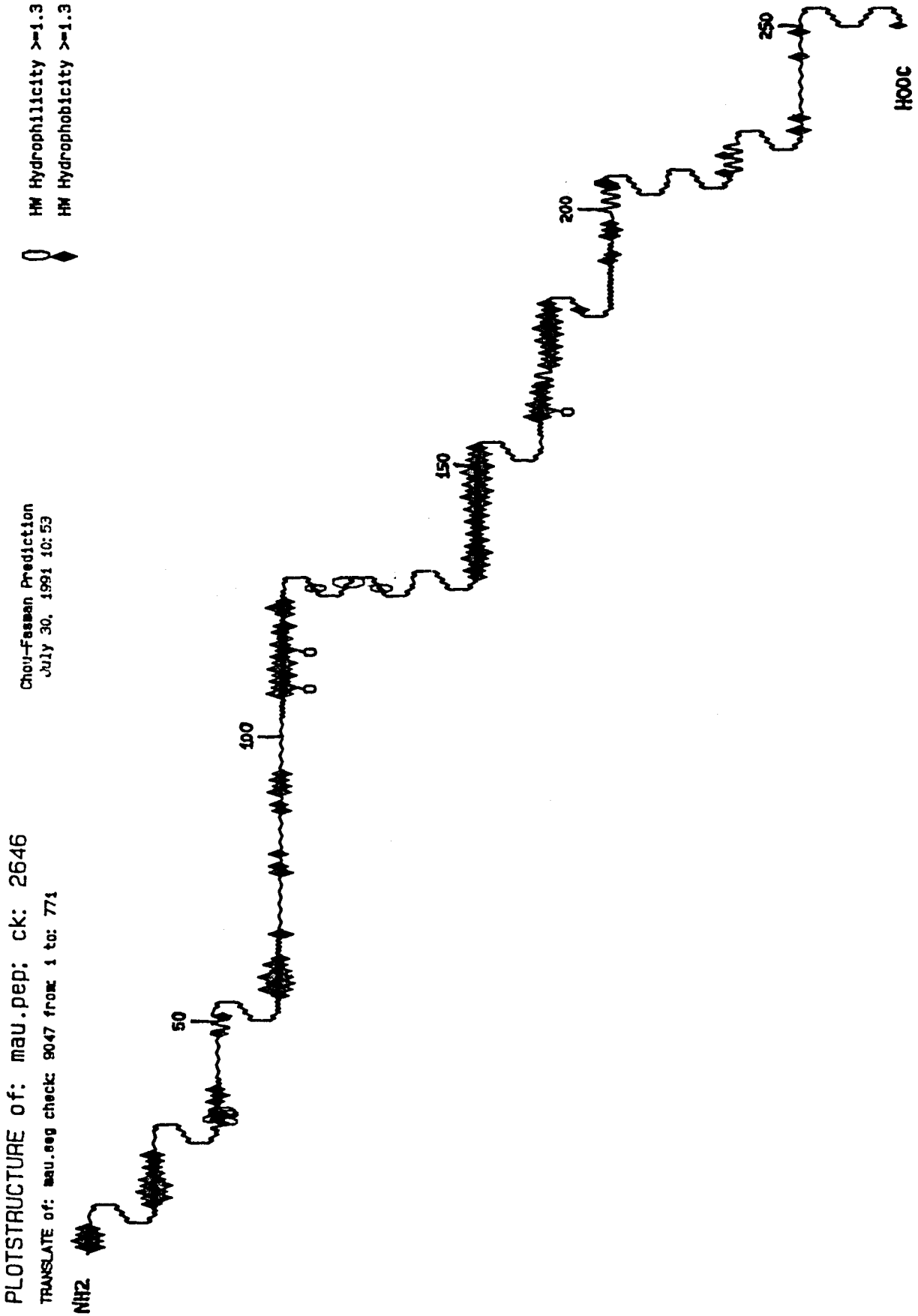
Fig.33.- Predicció d'estructura per a la proteïna del gen *Adh* de *D.pseudoobscura*.
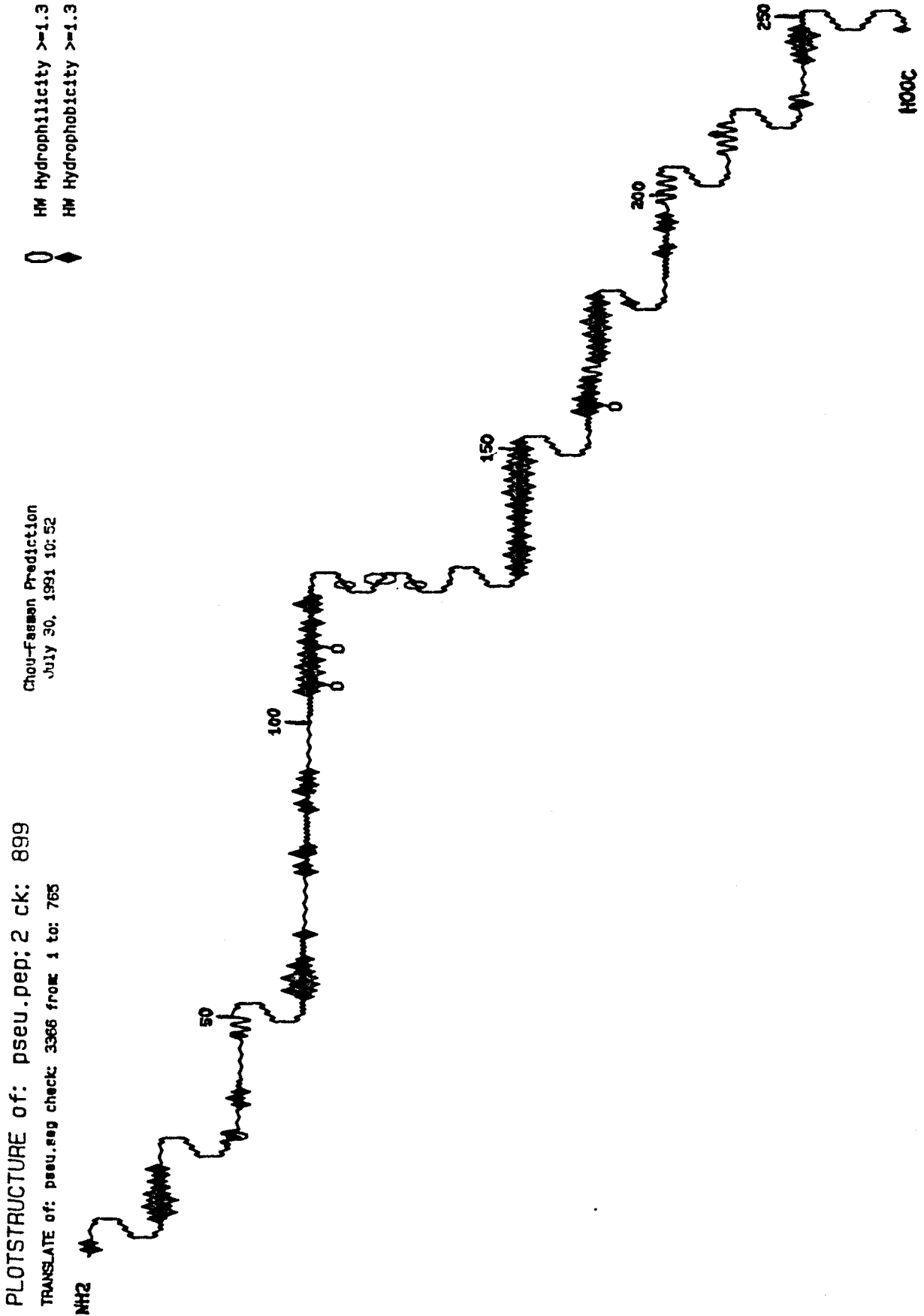
Fig.34.- Predicció d'estructura per a la proteïna del gen *Adh* de *D.ambigua*.
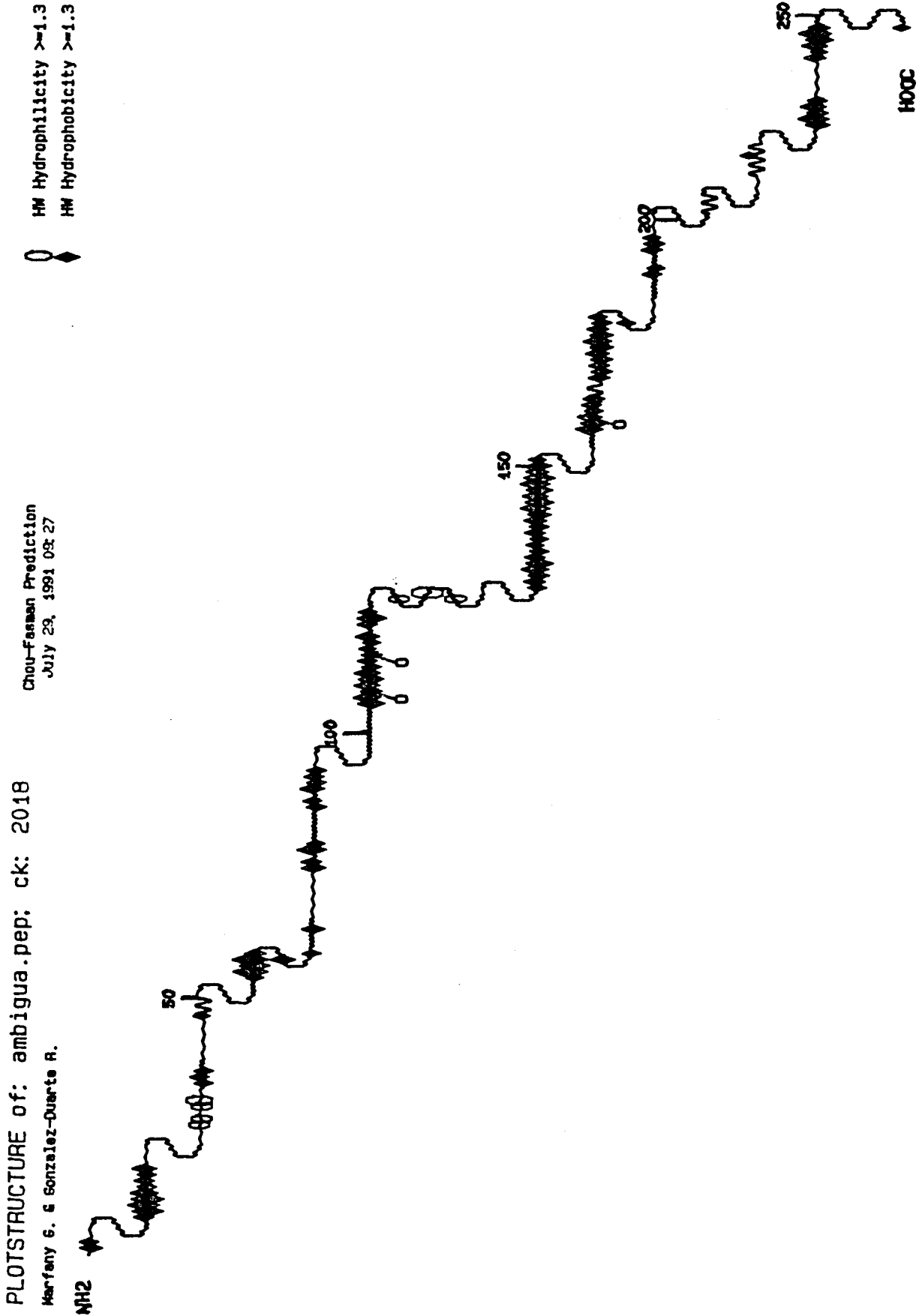
Fig.35.- Predicció d'estructura per a la proteïna del gen *Adh* de *D.subobscura*.
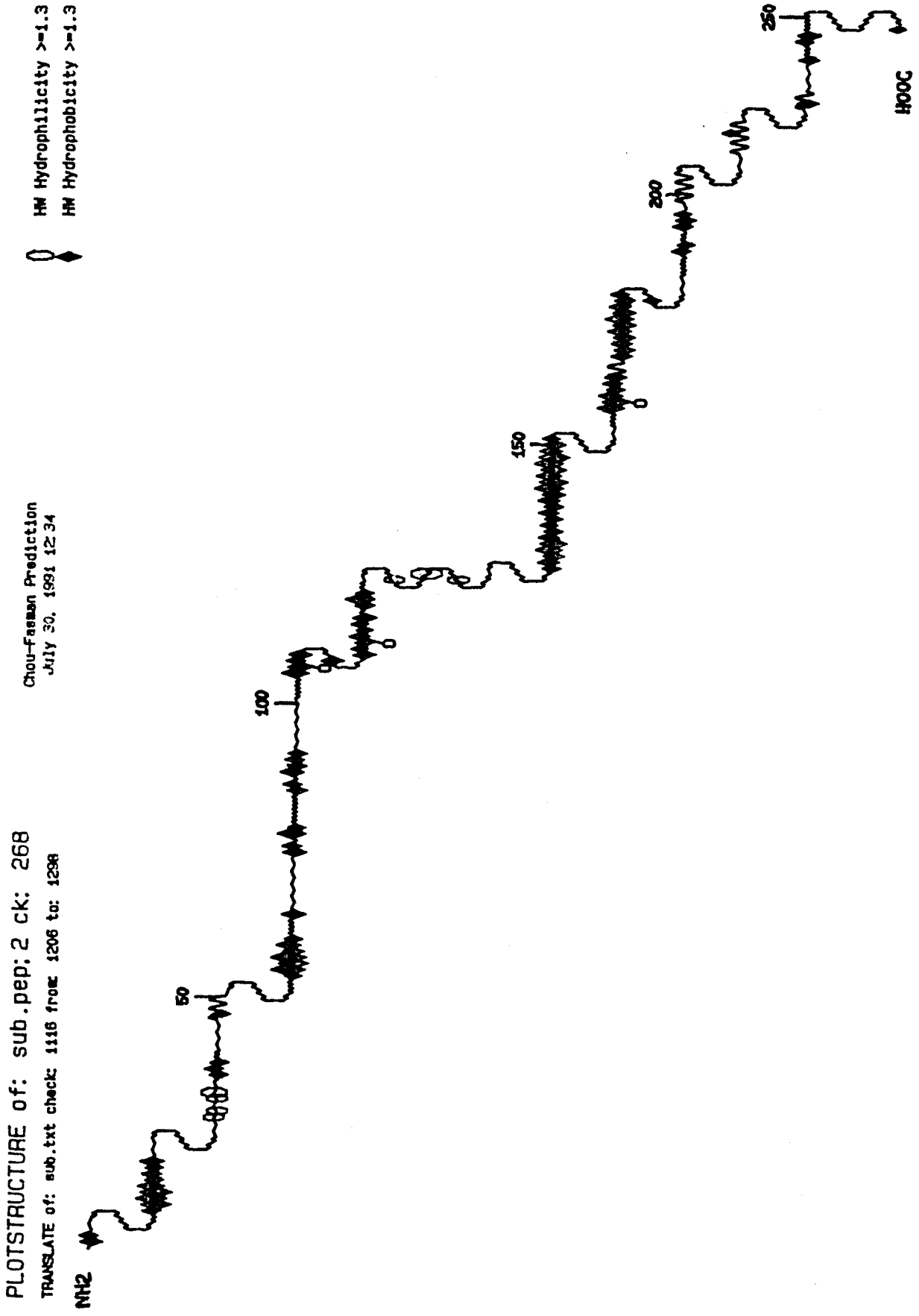
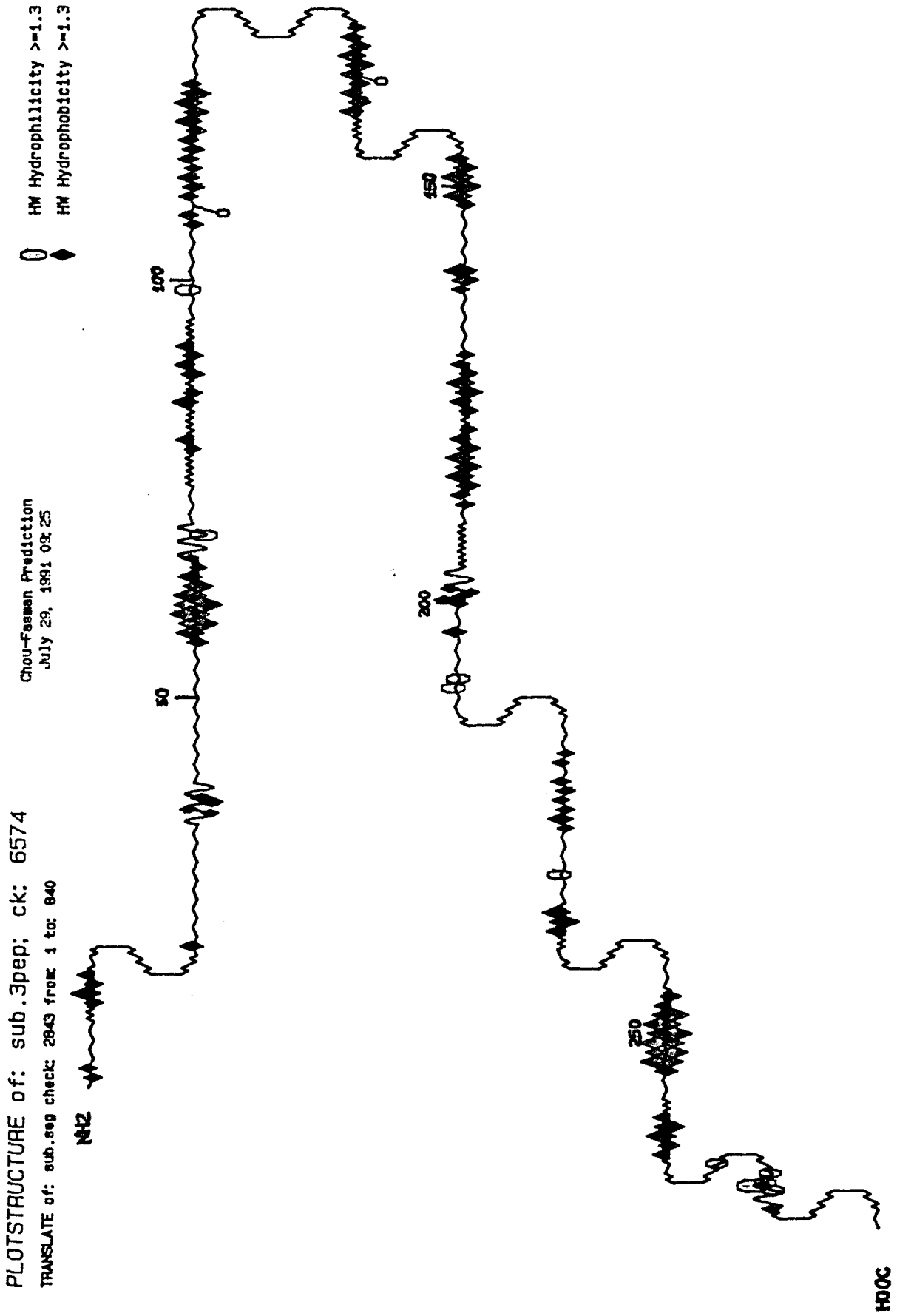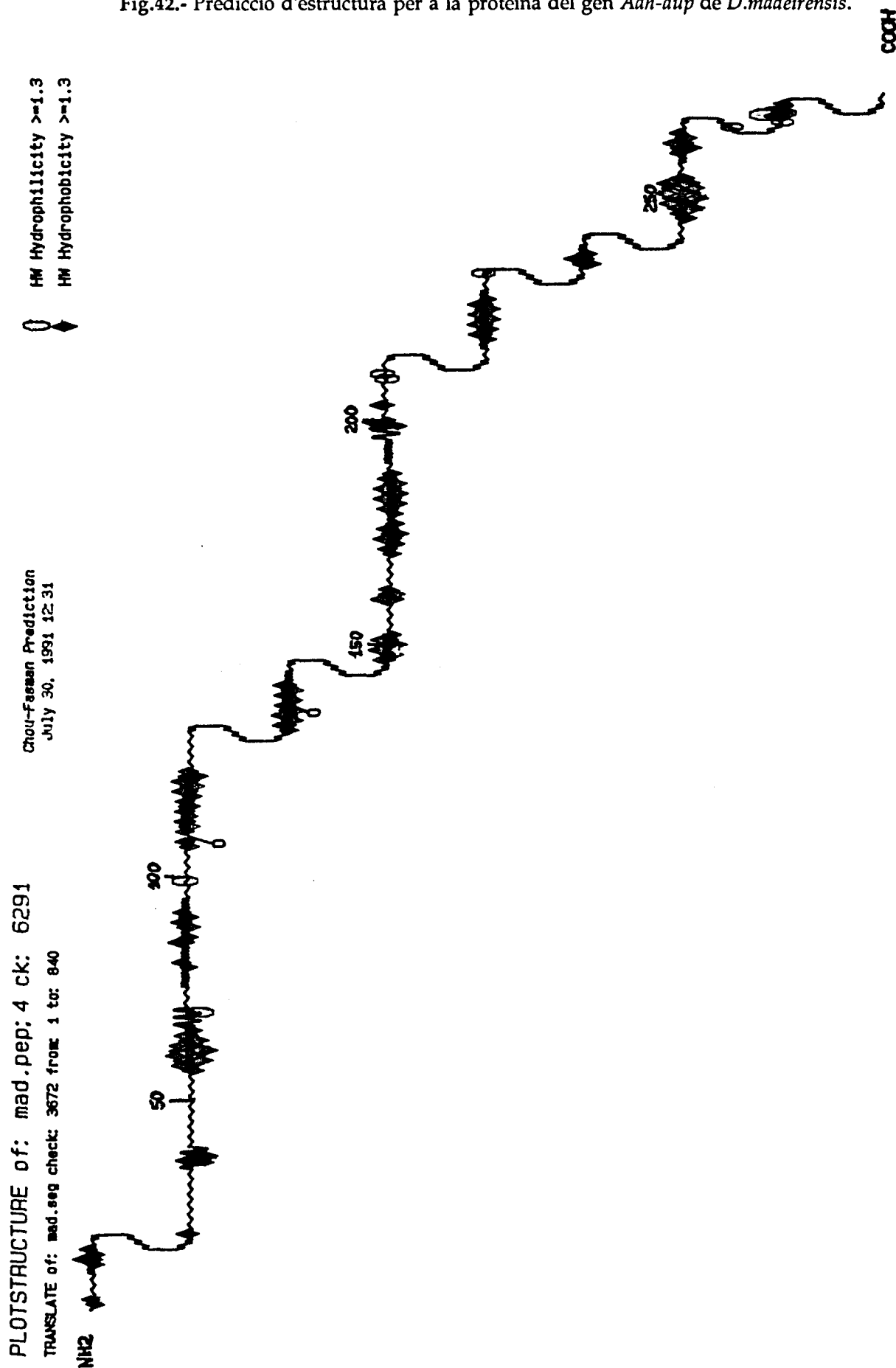Fig.36.- Predicció d'estructura per a la proteïna del gen *Adh* de *D.madeirensis*.
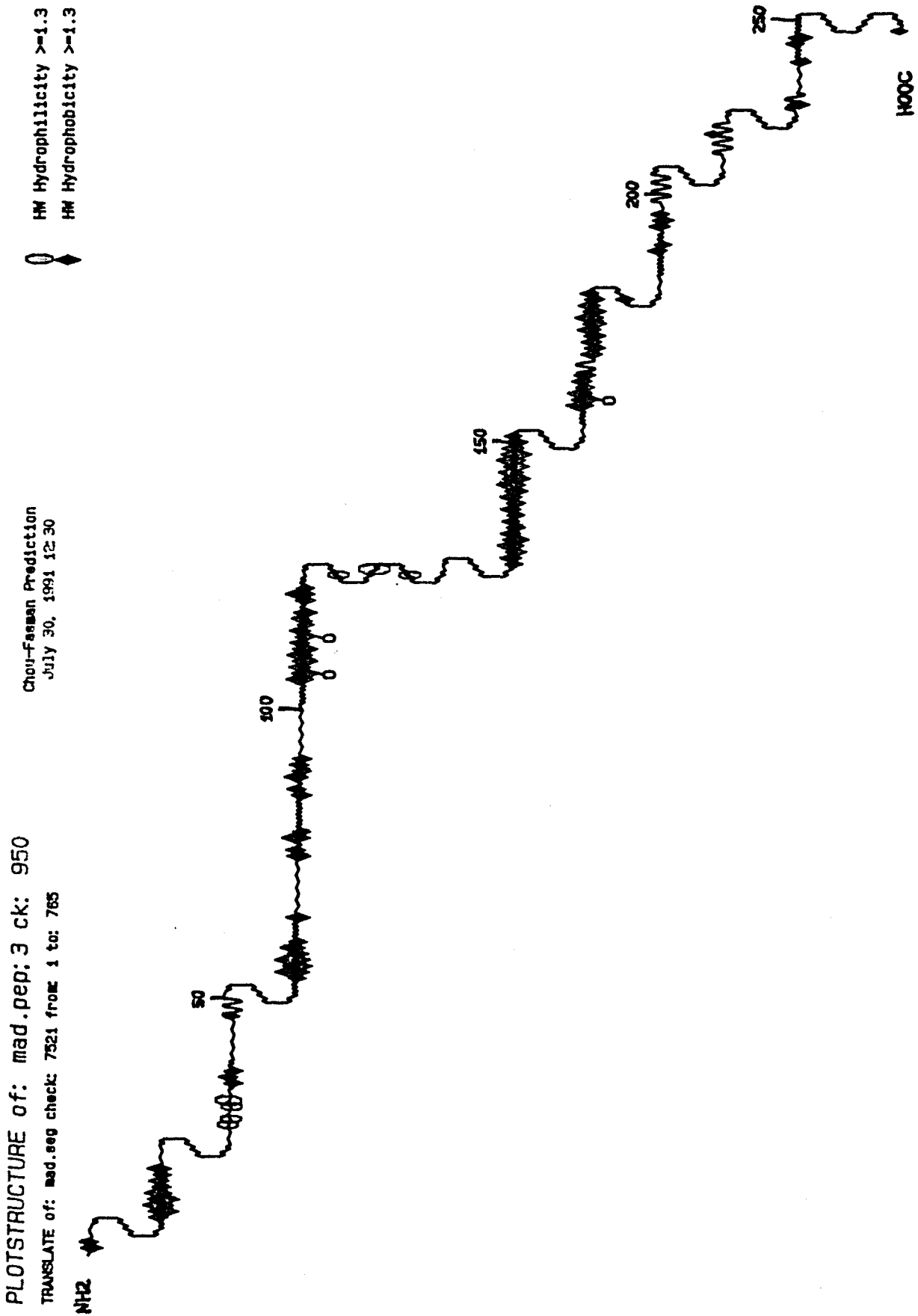
Fig.37.- Predicció d'estructura per a la proteïna del gen *Adh* de *D.guanche*.
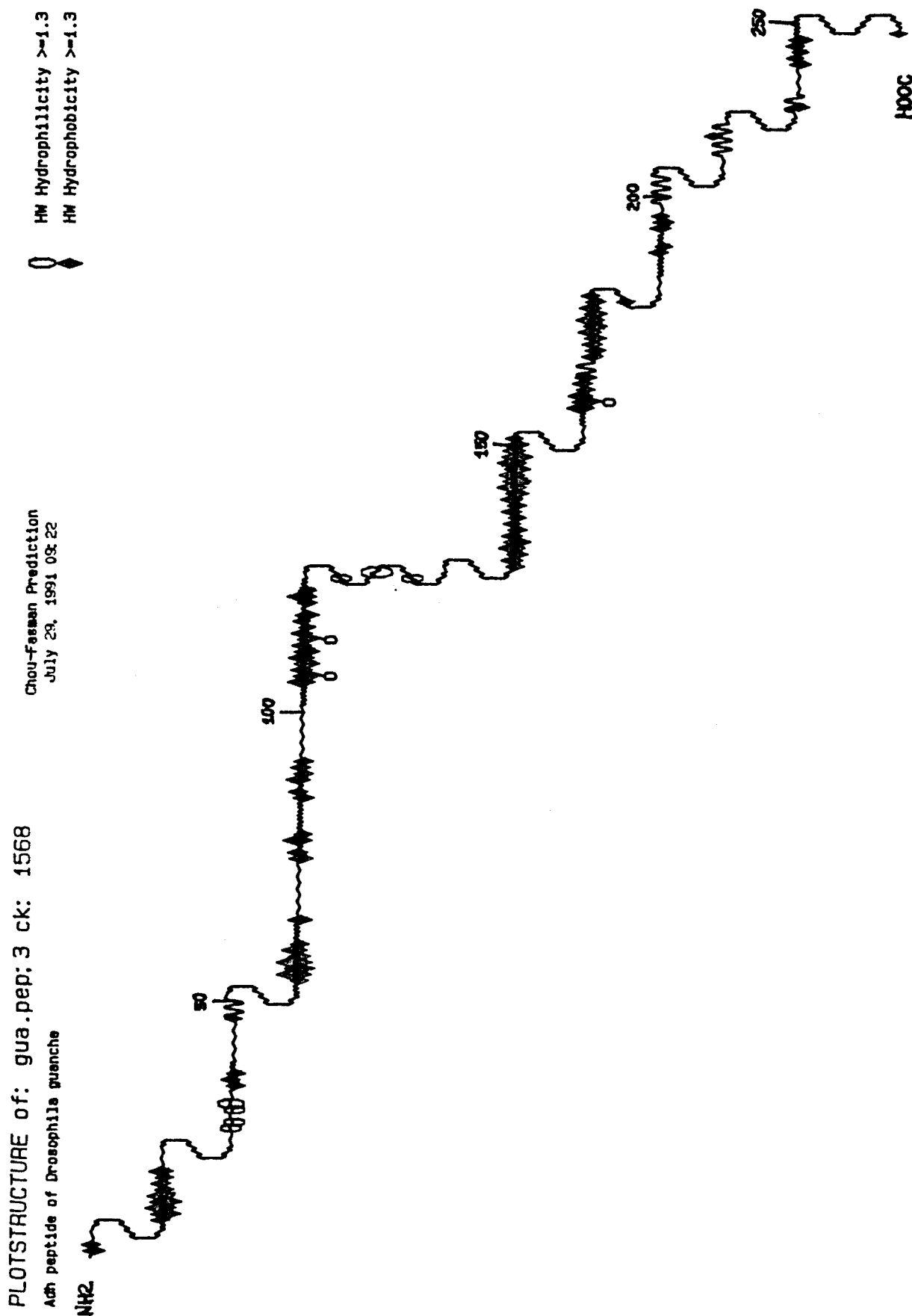
Fig.38.- Predicció d'estructura per a la proteïna del gen *Adh-dup* de *D.mauritiana*.
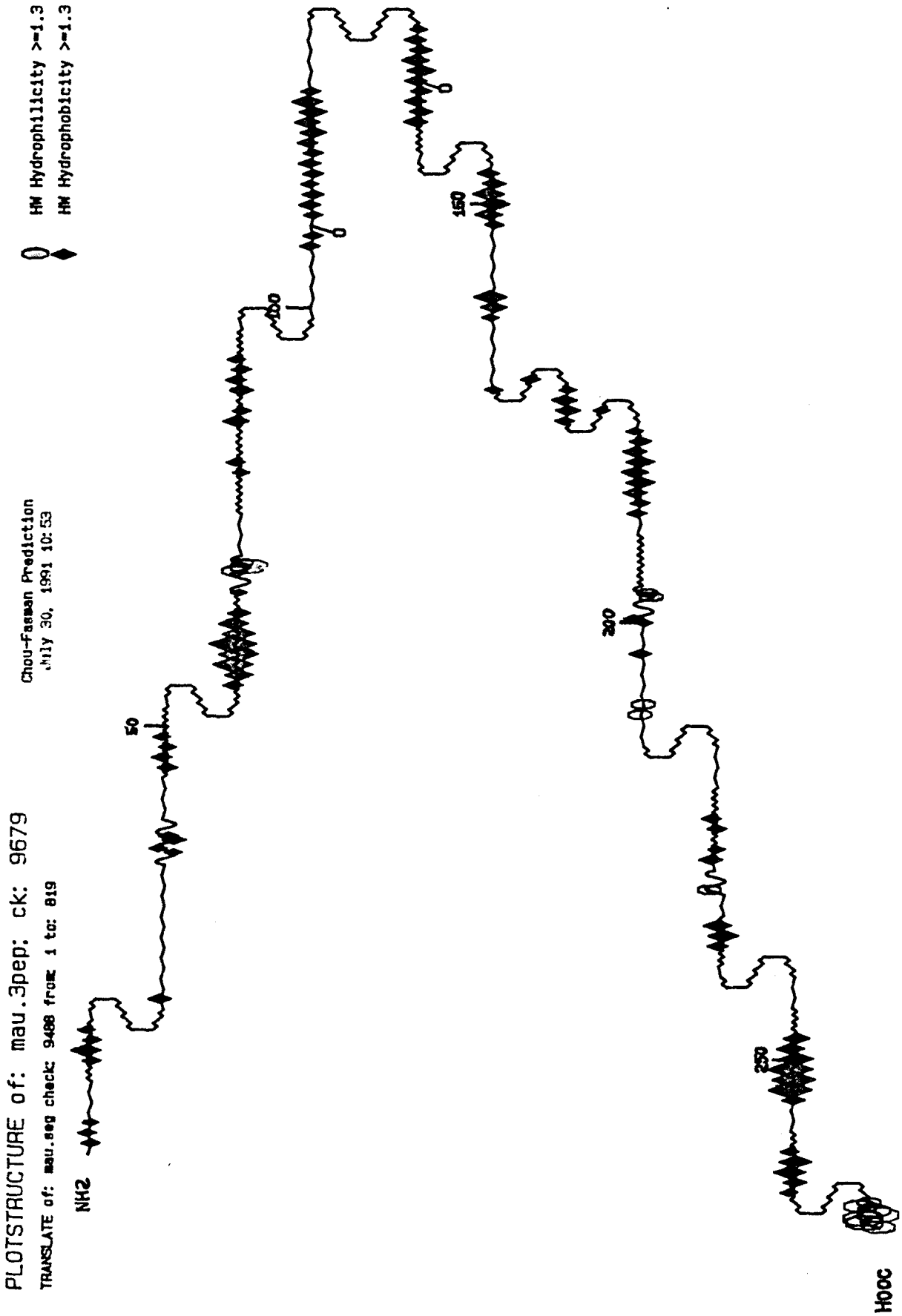
Fig.39.- Predicció d'estructura per a la proteïna del gen *Adh-dup* de *D.pseudoobscura*.
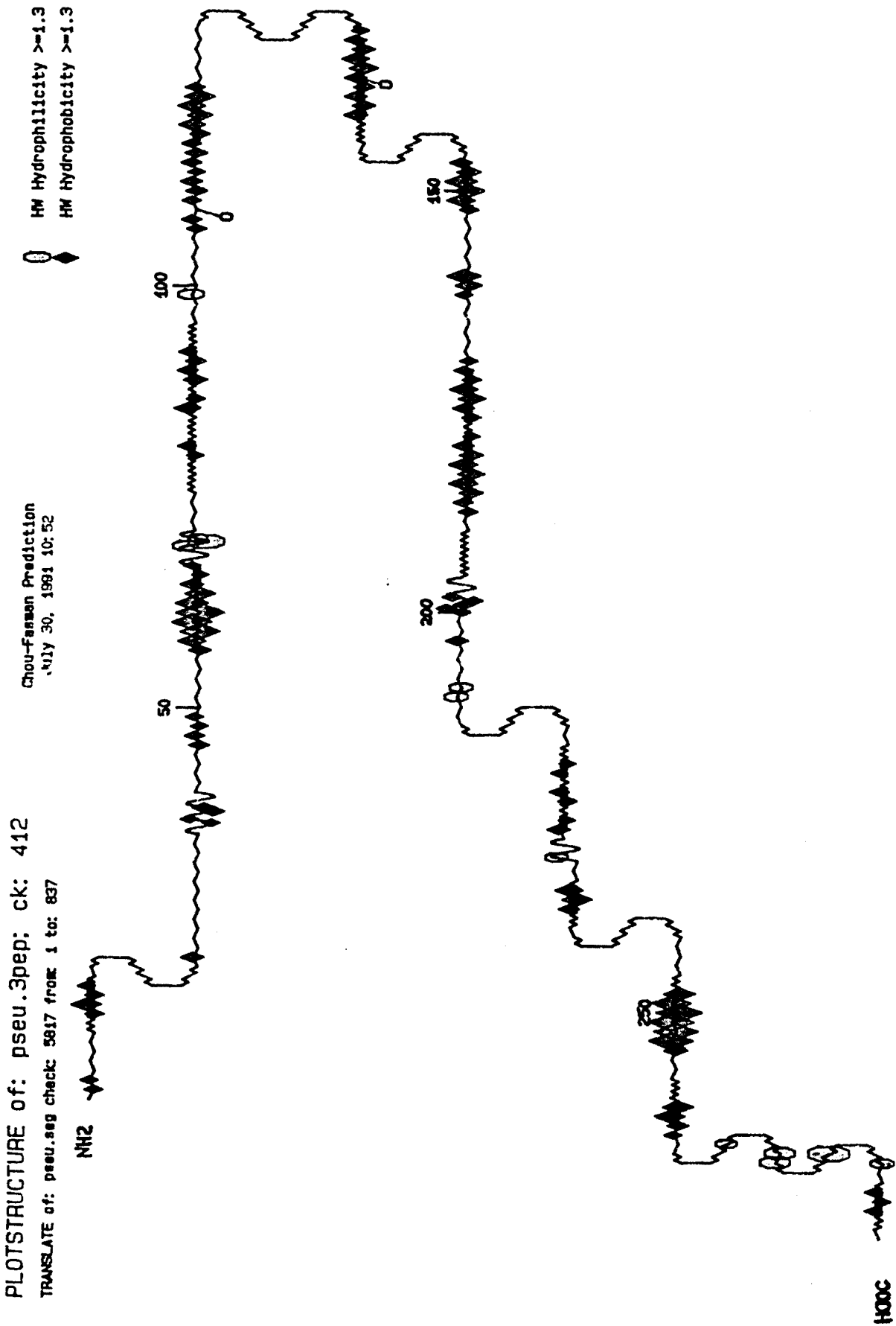
Fig.40.- Predicció d'estructura per a la proteïna del gen *Adh-dup* de *D.ambigua*.
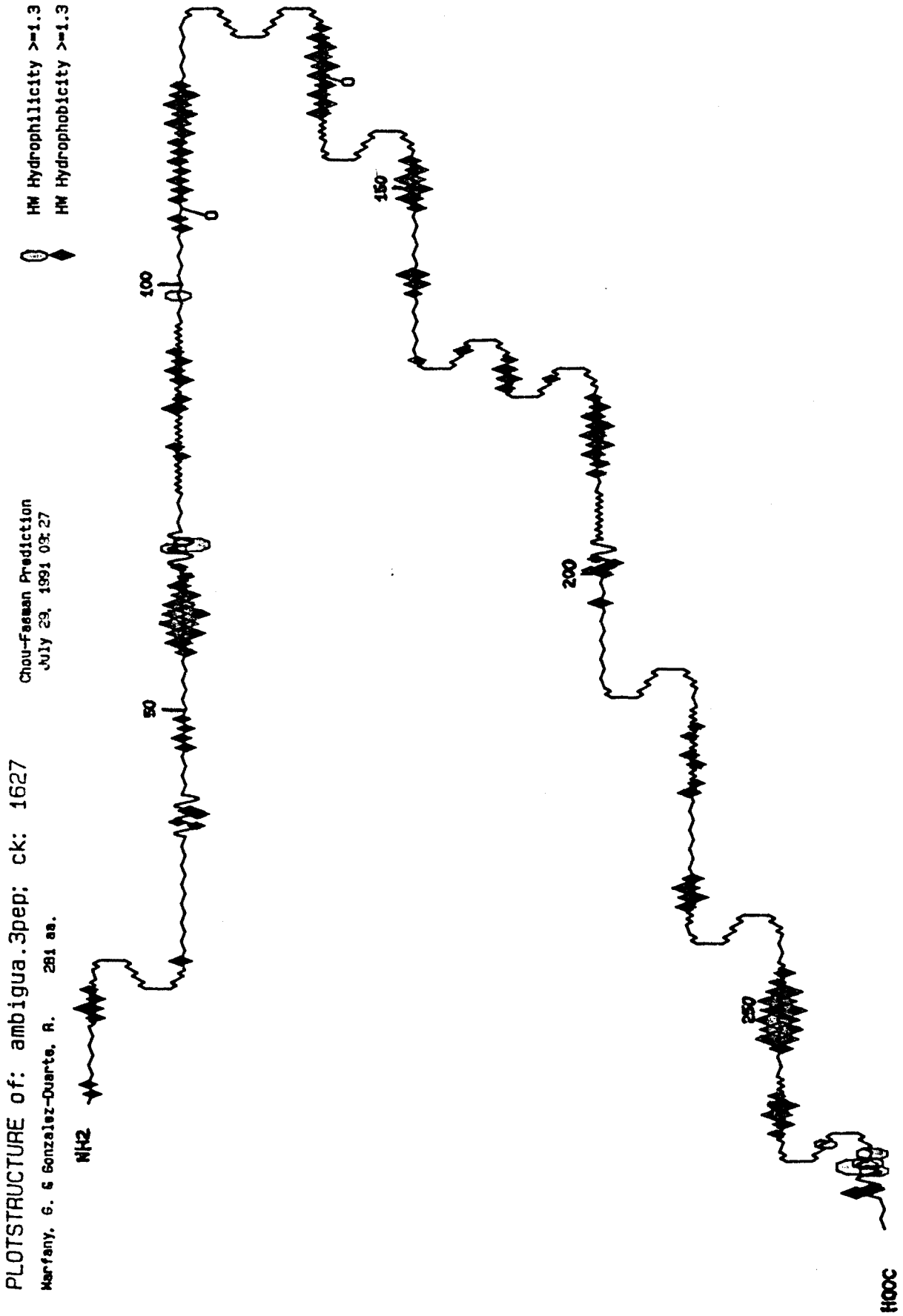
Fig.41.- Predicció d'estructura per a la proteïna del gen *Adh-dup* de *D.subobscura*.

PLOTSTRUCTURE of: sub.3pep; ck: 6574

TRANSLATE of: sub.seg check: 2843 from: 1 to: 840

Chou-Fasman Prediction
July 29, 1991 09:25

HM Hydrophilicity >=1.3
HM Hydrophobicity >=1.3

Fig.42.- Predicció d'estructura per a la proteïna del gen *Adh-dup* de *D.madeirensis*.

Fig.43.- Predicció d'estructura per a la proteïna del gen *Adh-dup* de *D.guanche*.
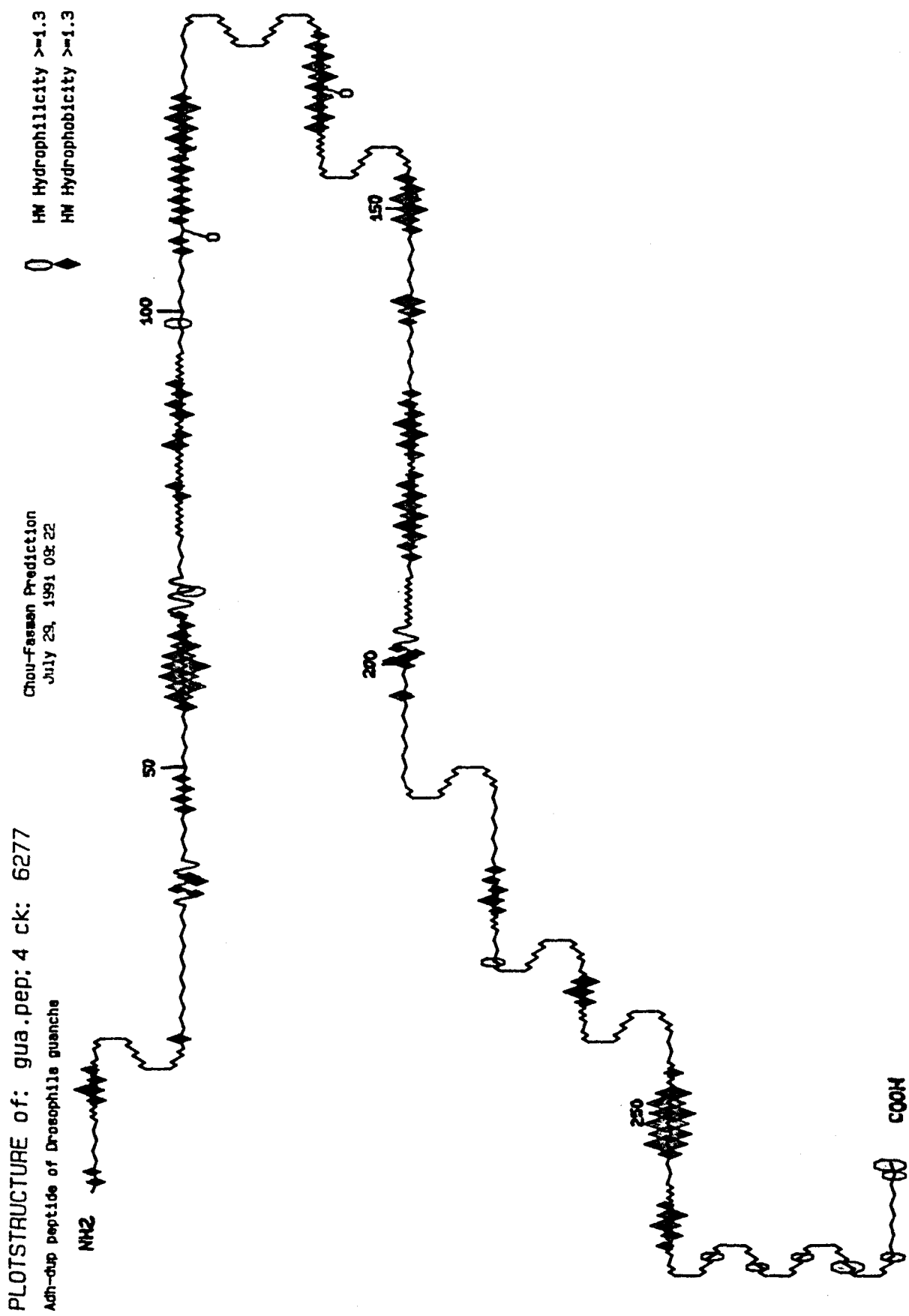
ESTUDI DE L'EXPRESSIÓ DEL GEN *Adh* MITJANÇANT TRANSFERÈNCIES *NORTHERN* A LES ESPÈCIES: *D.ambigua, D.guanche, D.madeirensis* I *D.subobscura* A DIFERENTS ESTADIS DEL DESENVOLUPAMENT.

Es va obtenir RNA total dels estadis del desenvolupament: larva III, pupa i adult de les espècies objecte del nostre estudi segons l'apartat 12 de Materials i Mètodes. S'han emprat diferents tipus de controls: com a control positiu, s'ha disposat de RNA de *D.melanogaster* de la soca Adh^F i com a control negatiu, de RNA de *D.melanogaster* de la soca Adh^{n4} (soca nul·la per a l'ADH en la qual no és detectable l'expressió del gen *Adh*). D'aquestes soques s'ha obtingut RNA dels mateixos estadis del desenvolupament que les espècies analitzades, així com també de l'estadi larva I. Com a control intern per quantificar la càrrega de les mostres, s'ha emprat una sonda d'un gen d'actina de *D.melanogaster, Hd19* (Fyrberg et al., 1979).

Donat que el gen *Adh* és un gen d'expressió elevada, no és necessari l'obtenció de RNA poliA$^+$ i cal poc temps d'exposició de les autorradiografies. Els resultats obtinguts es presenten a la Fig.44. Es detecta un únic trànscrit del gen *Adh* d'una mida similar al descrit per a *D.melanogaster*. A més, l'expressió durant els diversos estadis del desenvolupament correspon a un patró també similar al descrit per a*D.melanogaster* (Savakis et al., 1985), amb una disminució dràstica a l'estadi de pupa que el fa pràcticament indetectable a *D.melanogaster*, però que almenys, a l'espècie *D.ambigua*, és clarament visible. En els nostres resultats, s'observa expressió del gen *Adh* a l'estadi de pupa a *D.melanogaster*, però és degut a una sobrecàrrega de RNA, tal i com es pot detectar en el control d'actina.
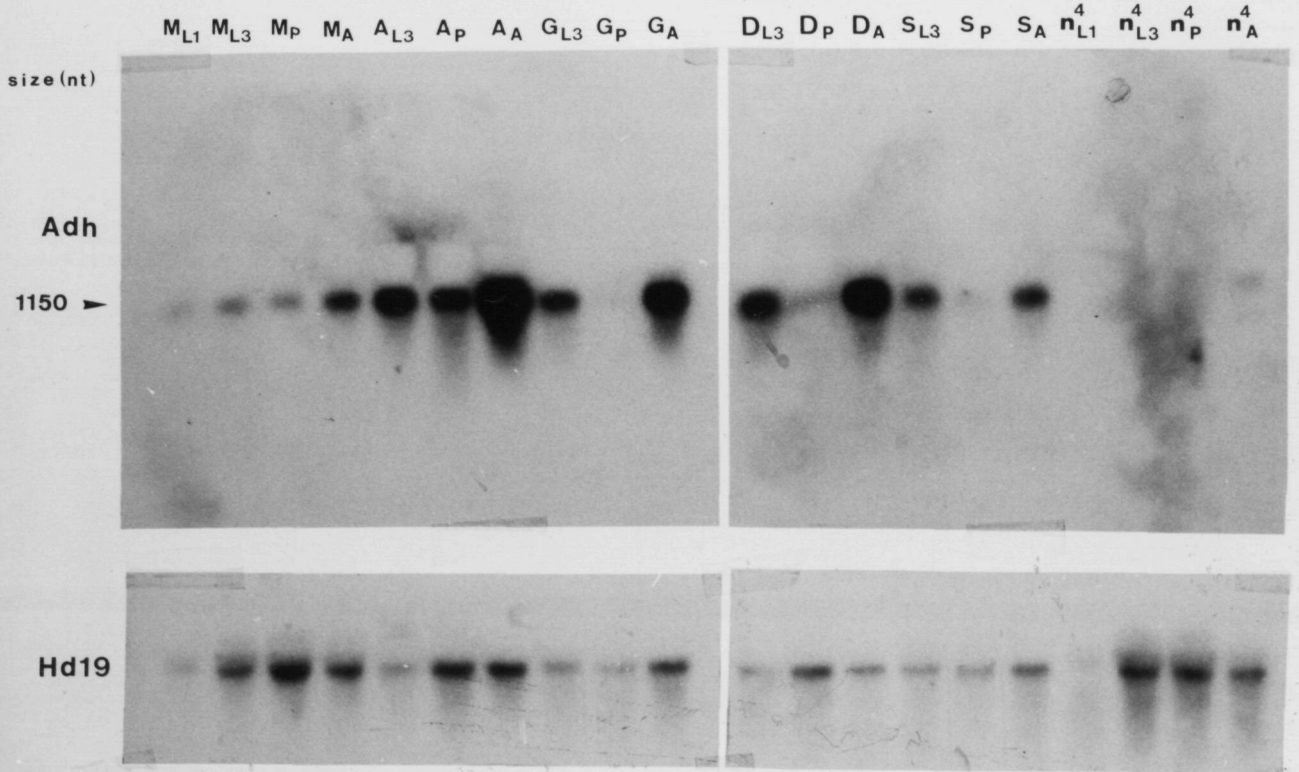
Fig.44- Northern de RNA total de les espècies: M-*D.melanogaster*, A-*D.ambigua*, G-*D.guanche*, D-*D.madeirensis* i S-*D.subobscura*, n4-*D.melanogaster* soca n4. Estadis del desenvolupament: L1-larva I, L3-larva III, P-pupa, A-adult.
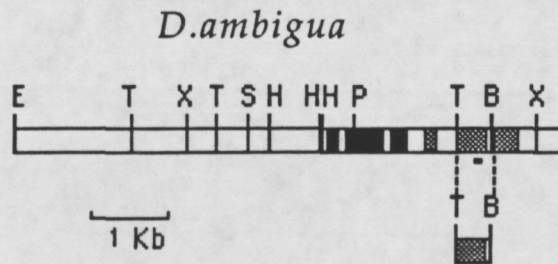
D.ambigua

Fig.45- Mapa de restricció de la regió genòmica del gen *Adh* a *D.ambigua*. La sonda emprada per a la detecció de transcrits del gen *Adh-dup* així com també la posició de l'oligonucleòtid, estan assenyalats. Les caixes negres i les ratllades representen, respectivament, la regió codificant del gen *Adh* i del gen *Adh-dup*. E-*Eco*RI, B-*Bam*HI, H-*Hind*III, P-*Pvu*II, T-*Pst*I, S-*Sal*I, X-*Xba*I.

ESTUDI DE L'EXPRESSIÓ DEL GEN *Adh-dup* MITJANÇANT TRANSFERÈNCIES *NORTHERN* I PCR A LES ESPÈCIES: *D.ambigua, D.guanche, D.madeirensis* I *D.subobscura* A DIFERENTS ESTADIS DEL DESENVOLUPAMENT.

La mateixa transferència *Northern* emprada en l'estudi d'expressió del gen *Adh* es va fer servir per a la detecció d'un possible transcrit del gen *Adh-dup*, ja que no havia estat descrit que aquest gen s'expressés, tot i que la gran conservació de seqüència a nivell inter-específic permetia suggerir que es tractava d'un gen realment funcional. Es van emprar dos tipus de sonda d'elevada especificitat:

a) un oligonucleòtid 17-mer, complementari al tránscrit que preteniem detectar i que correspon a una seqüència, conservada entre espècies, del segon exó de l'*Adh-dup* (posició especificada en la Fig.45). Aquest oligonucleòtid només compartia 4 posicions nucleotídiques amb el gen *Adh*, i encara que de mida petita, més del 75% de la seva composició són G+C, i per tant, permetia precisar millor les condicions d'hibridació, en tenir una temperatura de fusió relativament elevada. El marcatge es va realitzar per l'extrem 5' de l'oligonucleòtid mitjançant γ-^{32}P-dATP

b) un fragment de restricció que conté el segon exó del gen *Adh-dup* de *D.ambigua*. Aquest fragment *Pst*I-*Bam*HI conté prop de 430 pb i apareix subratllat a la Fig.45. El marcatge es va realitzar per *random hexamer priming* amb α-^{32}P-dCTP.

Els resultats d'aquest estudi van ser negatius. En el cas de l'oligonucleòtid, tot i no presentar homologia amb la seqüència del gen *Adh*, sí en presentava per alguna seqüència del RNA ribosomal, i encara que les condicions d'hibridació i rentat fóssin les de màxima astringència, els dos rRNA hibridaven amb l'oligonucleòtid. Pel que fa a la sonda de dsDNA, no s'aconseguí distingir cap banda d'hibridació, encara que el temps d'exposició es va allargar fins a un mes.

Ens vàrem plantejar incrementar la sensibilitat del mètode anterior, aïllant la fracció poliadenilada dels RNA totals a partir dels quals s'havia efectuat la transferència *Northern* (veure apartat 12 de Material i Mètodes). Aquest mètode tampoc va permetre distingir cap banda d'hibridació corresponent al possible tránscrit.

De l'anàlisi de l'expressió del gen *Adh-dup* es podria concloure que, a diferència del gen *Adh*, aquest si s'expressava potser era a nivells basals, quasi indetectables. Potser la seva expressió estaria restringida a cèl·lules o teixits de baixa representació en l'organisme sencer, o bé s'expressaria de forma puntual en algun estadi molt concret del desenvolupament.

És reconeguda l'alta eficiència de la tècnica de la PCR per a la detecció de quantitats ínfimes de motllo original. Per això es va decidir aplicar-la per detectar el possible tràit de l'*Adh-dup*. Es va emprar RNA total d'organismes adults de *D.melanogaster* i *D.ambigua* i es va procedir a una reacció de transcripció inversa i síntesi de la segona cadena de cDNA, prèvies a la reacció de PCR (apartat 13 de Material i Mètodes). Els encebadors per a la reacció van ésser dissenyats per tal de discriminar l'existència de DNA genòmic contaminant d'un possible tràit, de forma que els dos oligonucleòtids corresponien a cadenes complementàries de seqüències situades en dos exons diferents (Fig.26). Si l'amplificació corresponia a DNA genòmic, es trobaria una banda producte de mida superior, donat que conté un intró, mentre que si la banda amplificada procedia d'un tràit, la mida seria menor ja que l'intró estaria processat. El control positiu va ésser el producte amplificat a partir del tràit del gen *Adh*. Per ambdues reaccions d'amplificació es van emprar les mateixes condicions.

Els resultats van ser parcialment positius. Mentre que per al gen *Adh* s'obtenia una banda única d'amplificació corresponent al tràit processat, per al gen *Adh-dup* s'han obtingut repetidament dues bandes com a mínim, una banda corresponent al DNA genòmic contaminant o al tràit no processat i una altra banda de mida inferior que podria correspondre al tràit d'*Adh-dup* processat. Aquestes dues bandes apareixen fins i tot quan el RNA abans d'amplificar havia estat tractat amb DNasa, lliure de RNases, per tal d'eliminar DNA contaminant de la mostra.

En un treball recent (Kreitman i Hudson, 1991), es fa esment de l'existència de tràits del gen *Adh-dup*, malgrat aquesta informació no ha estat encara publicada i la seva accessibilitat és limitada (Undergraduate Thesis, University of Princeton).

MODEL MATEMÀTIC PROPOSAT PER A L'EVOLUCIÓ DE LES RETROSEQÜÈNCIES DEL GEN *Adh* A *D.subobscura.*

Hem adaptat el model proposat per Miyata i Tasunaga (1981) per explicar l'evolució d'una retroseqüència del gen de la α-globina de ratolí a l'evolució de les retroseqüències del gen *Adh* a *D.subobscura* (Fig.46). Aquest model ha necessitat d'una adaptació donat que disposàvem de dues retroseqüències en lloc d'una.
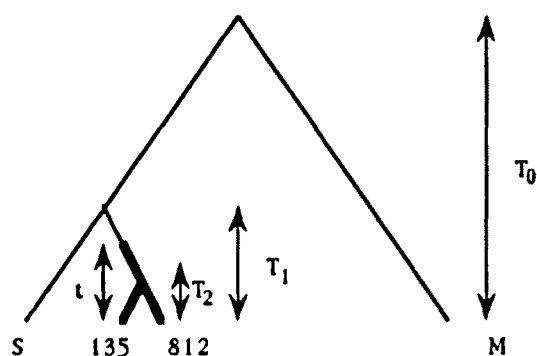
Fig.46.- Representació esquemàtica del model proposat per a l'origen de les retroseqüències a D.subobscura. S-D.subobscura; 812 i 135- retroseqüències S812 i S135; M-D.melanogaster.

Els axiomes d'aquest mètode es basen en una velocitat de mutació/fixació per a les posicions silencioses d'un gen (v_s) que és diferent de la de les posicions de reemplaçament (v_a) i diferent alhora de la velocitat de substitució en un pseudogen no funcional (v_0), tot i que aquestes es mantenen constants dins de la seva categoria al llarg de l'evolució gènica. No es disposa d'aquestes taxes de divergència, però en canvi, es poden calcular a partir de les substitucions corregides presentades entre dues seqüències homòlogues. Aquests valors han de donar-se en percentatges corregits per tal d'uniformitzar les inferències (Kcs i Kca, per a les substitucions silencioses i de reemplaçament, respectivament). Per últim, s'ha de disposar d'una espècie de referència, de la qual es conegui aproximadament la data de divergència respecte a una de les espècies analitzades. Amb aquest temps, que anomenem T_0 es poden estimar els següents paràmetres: T_1, o temps des de que es va originar la retroseqüència ancestre; T_2, o temps des de que la *retroseqüència* es va multiplicar; i t, temps des de que la *retroseqüència* va perdre la seva funcionalitat

El desenvolupament matemàtic és el següent:

1) $2v_sT_0 = Kcs \ (M/S)$

2) $2v_aT_0 = Kca \ (M/S)$

3) $2v_oT_2 = Kco \ (812/135)$

4) $v_sT_0 + v_s \ (T_0 - T_1) + v_s \ (T_1 - t) + v_o \ (t - T_2) + v_oT_2 = Kcs \ (M/812)$

5) $v_aT_0 + v_a \ (T_0 - T_1) + v_a \ (T_1 - t) + v_o \ (t - T_2) + v_oT_2 = Kca \ (M/812)$

6) $v_s \ T_1 + v_s \ (T_1 - t) + v_o \ (t - T_2) + v_oT_2 = Kcs \ (S/812)$

7) $v_a \ T_1 + v_a \ (T_1 - t) + v_o \ (t - T_2) + v_oT_2 = Kca \ (S/812)$

De les equacions 1) i 2) es dedueix v_s i v_a, que es substitueixen a les equacions de la 4) a la 7). Llavors es combinen les equacions 4) i 5), i també la 6) i la 7), obtenint-se:

8) $Kcs \ (M/S) - Kca \ (M/S) - \dfrac{t}{2T_0} \ [Kcs(M/S) + Kca \ (M/S)] = Kcs \ (M/812) - Kca \ (M/812)$

9) $\dfrac{T_1}{T_0} \ [Kcs \ (M/S) - Kca \ (M/S)] - \dfrac{t}{2T_0} \ [Kcs \ (M/S) - Kca \ (M/S)] = Kcs \ (S/812) - Kca \ (S/812)$

Si considerem $f(\ ) = Kcs(\ ) - Kca(\ )$, es simplifica a:

10) $f \ (M/S) - \dfrac{t}{2T_0} \ f \ (M/S) = f \ (M/812)$

11) $\dfrac{T_1}{T_0} \ f \ (M/S) - \dfrac{t}{2T_0} \ f \ (M/S) = f \ (S/812)$

De l'equació 10) es pot obtenir el valor de t, i llavors, de la 11) el valor de T_1. Per últim, el valor de v_0 es pot obtenir de les equacions 6) o 7) i el valor de T_2 de l'equació 3).

12) $t = 2T_0\{1 - f(M/812)/f(S/M)\}$

13) $T_1 = T_0\{1 - (f(M/812) - f(S/812))/f(S/M)\}$

14) $v_0 =$

15) $T_2 =$

Si la T_0 té un valor aproximat de 25 MY (milions d'anys), llavors:

$$v_s = 1.28 \times 10^{-8} \, nt/lloc/any \qquad\qquad T_1 = 3.05 \, MY$$

$$v_a = 1.24 \times 10^{-9} \, nt/lloc/any \qquad\qquad t = 1.46 \, MY$$

$$v_o = 2.75 \times 10^{-8} \, nt/lloc/any \qquad\qquad T_2 = 0.23 \, MY$$