

**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.



Memòria presentada per la Llicenciada en Ciències Biològiques GEMMA MARFANY I NADAL per a aspirar al grau de Doctora en Ciències Biològiques.

El present treball ha estat realitzat als laboratoris del grup de Genètica Molecular del Departament de Genètica de la Facultat de Biologia de la Universitat de Barcelona sota la direcció de la Doctora ROSER GONZÀLEZ I DUARTE.

Barcelona, Setembre de 1991

Vist i plau de:

DIRECTORA

Handwritten signature of Roser González i Duarte, the director.

AUTORA

Handwritten signature of Gemma Marfany, the author.

*"Aprèn on és l'enteniment, on és la força,
on és la prudència del conèixer,
allí és juntament la longevitat i la vida;
la llum dels ulls i la pau."*

(Baruc 3, 14)

*"Per acabar, fill meu, tingues present que fer llibres és una feina de no
acabar mai i que estudiar en excés danya la salut."*

(Eclesiastès 12, 12)

Als meus pares,

A en Jordi, per tot el que hem compartit,
compartim i compartirem

AGRAÏMENTS

En presentar aquest treball voldria expressar el més sincer agraïment a totes aquelles persones que han contribuït i ajudat, d'alguna forma, a la realització de la present Tesi Doctoral, i especialment:

A la Doctora Roser González i Duarte, per haver estat la Directora d'aquesta Tesi. La seva objectivitat i rigor científics han permès un diàleg distés que ha fructificat en el plantejament i discussió de noves idees. El suport i l'interés que he rebut sempre de part seva han suposat un constant estímul científic, i sobretot, el fonament d'una profunda i càlida relació humana.

Al Doctor Lluís Serra, per la seva excel·lent disposició en la discussió de resultats, en l'aprofundiment de conceptes evolutius i en el suggeriment d'idees que han estat força profitoses.

Al Dr. Albert Boronat i el seu grup, especialment a Carme Caelles, que no van dubtar en oferir coneixements i experiència sobre seqüenciació.

To Dr. David J. Finnegan and his group, for their teaching me how to microinject *Drosophila* eggs. I will always remember their warm and kind treatment when I stayed in Edimburgh.

Al Dr. Emili Saló, que em va permetre ajudar-lo mentre m'ensenyava els primers protocols de Biologia Molecular. Aquesta vella coneixença ha donat peu a una bona amistat i alguna que altra carinyosa desavinença.

Voldria agrair a tota la gent del departament de Genètica amb els que he tingut relació, especialment al Grup de Genètica Molecular (MOL) el haver-me donat suport i ajuda moral continuada, constituint un ambient ideal per al desenvolupament d'aquesta tasca científica. A la Dra. Lluïsa Vilageliu, per haver seguit amb interès, des del seu inici, els progressos d'aquesta Tesi. A la Dra. Sílvia Atrian, per ésser la meva contínua assessora lingüística i informàtica, i sobretot per la seva companyonia desinteressada, tant a Edimburg com a Barcelona. Als Drs. Susana Balcells i Daniel Grinberg, que tot i fer menys temps que ens coneixem, compartim "força" espai i xerrades profitoses al laboratori. Al Dr. Joan Fibla, amb el que barallant-nos

carinyosament, hem compartit dia a dia les vicissituds d'anar-se fent Doctor. Al Ricard Albalat per haver-me ensenyat que el bon humor i la tranquil·litat són bones eines de treball. A tota la gent nova que està entrant, que renova l'aire i les ganes del nostre grup, especialment a Teresa Luque, que continua part de la feina, tot just començada. I a la Dra. Neus Visa, que a més d'ensenyar-me les tècniques d'hibridació *in situ*, va compartir les millors i pitjors èpoques de l'inici d'aquesta Tesi, amb totes les petites penes i alegries quotidianes.

També voldria agrair al Grup d'Epigenètica (EPI) el haver sabut estar a l'altre costat del "mur de Berlí". Al Dr. Joan Collet, que sempre ha estat un bon assessor tècnic i un bon company. A la Dra. Marta Riutort, que pacientment m'ha ensenyat els intrínqulis de les filogènies per ordinador. A en Jordi Garcia, amb el que he compartit protocols, solucions, aparells, problemes i sobretot, experiments i noves idees. A tots els que hi són ara i als que hi han estat, gràcies per la vostra companyia.

Als meus pares, per haver estat sempre al meu costat i haver confiat en mí. Als meus germans, per donar-me sempre suport i companyia, i als meus nebots per ensenyar-me la joia de viure.

2.- OBJECTIUS.....	37
3.- RESULTATS	
3.1.- CAPÍTOL I: "The <i>Adh</i> in <i>Drosophila</i> : cytological localization and restriction analysis in several species"	41
3.2.- CAPÍTOL II: "The <i>Adh</i> genomic region of <i>Drosophila ambigua</i> : evolutionary trends in different species"	51
3.3.- CAPÍTOL III: "The <i>Drosophila subobscura Adh</i> genomic region contains valuable evolutionary markers"	63
3.4.- CAPÍTOL IV: "Evidence of retrotranscription of protein-coding genes in <i>Drosophila</i> "	85
3.5.- CAPÍTOL V: "Characterization and evolution of the <i>Adh</i> genomic region of <i>Drosophila madeirensis</i> and <i>Drosophila guanche</i> "	105
3.6.- CAPÍTOL VI: APÈNDIX	127
3.6.1.- Localització citològica del gen <i>Adh</i> a <i>Drosophila guanche</i>	130
3.6.2.- Detecció de seqüències homòlogues al gen <i>Adh</i> en <i>Southern</i> s de DNA genòmic total de les espècies: <i>D.melanogaster</i> , <i>D.ambigua</i> , <i>D.guanche</i> , <i>D.madeirensis</i> i <i>D.subobscura</i>	132
3.6.3.- Detecció de seqüències amb homologia a la repetició invertida situada a 3' de les retroseqüències de <i>D.subobscura</i> en <i>Southern</i> s de DNA genòmic total de les espècies: <i>D.melanogaster</i> , <i>D.ambigua</i> , <i>D.guanche</i> , <i>D.madeirensis</i> i <i>D.subobscura</i>	134
3.6.4.- Polimorfisme de restricció a la regió genòmica del gen <i>Adh</i> a <i>Drosophila subobscura</i>	137
3.6.5.- Caracterització estructural de diferents retroseqüències del gen <i>Adh</i> al genoma de <i>D.subobscura</i>	138
3.6.6.- Evidència de l'existència de retroseqüències del gen <i>Adh</i> als genomes de <i>D.madeirensis</i> i <i>D.guanche</i>	140
3.6.7.- Prediccions d'estructura secundària per a les proteïnes codificades pels gens <i>Adh</i> i <i>Adh-dup</i> de <i>D.mauritiana</i> , <i>D.ambigua</i> , <i>D.guanche</i> , <i>D.madeirensis</i> i <i>D.subobscura</i>	145

3.6.8.- Estudi de l'expressió del gen <i>Adh</i> mitjançant transferències <i>Northern</i> a les espècies: <i>D.ambigua</i> , <i>D.guanche</i> , <i>D.madeirensis</i> i <i>D.subobscura</i> a diferents estadis del desenvolupament.....	171
3.6.9.- Estudi de l'expressió del gen <i>Adh-dup</i> mitjançant transferències <i>Northern</i> i PCR a les espècies: <i>D.ambigua</i> , <i>D.guanche</i> , <i>D.madeirensis</i> i <i>D.subobscura</i> a diferents estadis del desenvolupament.....	173
3.6.10.- Model matemàtic proposat per a l'evolució de les retroseqüències del gen <i>Adh</i> a <i>D.subobscura</i> ,.....	175
4.- DISCUSSIÓ.....	179
4.1.- EL GEN <i>Adh</i> EN EL GENOMA DEL SUBGRUP <i>obscura</i>	182
4.2.- ESTRUCTURA I EXPRESSIÓ DEL GEN <i>Adh</i> A ESPÈCIES DEL SUBGRUP <i>obscura</i>	182
4.2.1.- EL GEN <i>Adh</i>	182
4.2.3.- EL GEN <i>Adh-dup</i>	183
4.3.- EVOLUCIÓ DE LA REGIÓ GENÒMICA DEL GEN <i>Adh</i> A ESPÈCIES DEL SUBGRUP <i>obscura</i>	185
4.3.1.- EVOLUCIÓ DE LES REGIONS CODIFICANTS DEL GENS <i>Adh</i> <i>Adh=dup</i>	185
4.3.2.- EVOLUCIÓ DE LES REGIONS NO CODIFICANTS.....	191
4.3.3.- ANÀLISI EVOLUTIVA DE LA SEQÜÈNCIA DE LA REGIÓ GENÒMICA DE L' <i>Adh</i>	194
4.4.- ALTRES SEQÜÈNCIES HOMÒLOGUES AL GEN <i>Adh</i> EN EL GENOMA DEL SUBGRUP <i>obscura</i> : RETROSEQÜÈNCIES.....	196
4.4.1.- ESTRUCTURA DE LES RETROSEQÜÈNCIES DE L' <i>Adh</i> A <i>D.subobscura</i>	196
4.4.2.- ESTRUCTURA DE LES RETROSEQÜÈNCIES DE L' <i>Adh</i> A <i>D.madeirensis</i> I <i>D.guanche</i>	197
4.4.3.- EVOLUCIÓ I ORIGEN DE LES RETROSEQÜÈNCIES DE L' <i>Adh</i>	198
4.5.- RELACIONS FILOGENÈTIQUES ENTRE LES ESPÈCIES DEL SUBGRUP <i>obscura</i>	201
4.4.1.- ESTRUCTURA DE LES RETROSEQÜÈNCIES DE L' <i>Adh</i> A <i>D.subobscura</i>	196

5.- CONCLUSIONS.....	211
6.- MATERIALS I MÈTODES GENERALS	219
6.1.- MATERIAL BIOLÒGIC.....	221
6.2.- OBTENCIÓ I PURIFICACIÓ DE DNA GENÒMIC DE <i>Drosophila</i>	221
6.2.1.- Aïllament a gran escala de DNA genòmic de <i>Drosophila</i>	221
6.2.2.- Mini-preparació de DNA genòmic de <i>Drosophila</i>	222
6.3.- CONSTRUCCIÓ DE LLIBRERIES GENÒMIQUES.....	223
6.3.1.- Obtenció de fag lambda a gran escala.....	223
6.3.2.- Preparació de fag vector EMBL4	224
6.3.3.- Preparació de DNA genòmic insert.....	226
6.3.4.- Lligació. Empaquetament <i>in vitro</i>	227
6.3.5.- Titolació i plaquejament de la llibreria.....	227
6.3.6.- Amplificació de la llibreria.....	228
6.4.- MARCATGE D'ÀCIDS NUCLEICS	228
6.4.1.- Marcatge per <i>nick-translation</i>	228
6.4.2.- Marcatge per <i>random-hexamer priming</i>	229
6.4.3.- Marcatge amb polinucleòtid kinasa	229
6.4.4.- Separació de la sonda marcada dels nucleòtids radioactius no incorporats.....	229
6.4.4.1.- Separació per cromatografia en Sephadex G-50.....	229
6.4.4.2.- Separació per precipitació i centrifugació diferencials.....	230
6.4.5.- Eficiència del marcatge	230
6.4.5.1.- Cerenkov	239
6.4.5.2.- Líquid de centelleig.....	231
6.5.- HIBRIDACIÓ D'ÀCIDS NUCLEICS.....	231
6.5.1.- Condicions de les hibridacions.....	231
6.5.2.- Condicions dels rentats.....	232
6.5.3.- Exposició.....	232
6.5.4.- Re-hibridació de filtres	233

6.6.- CRIVELLATGE DE LLIBRERIES GENÒMIQUES.....	233
6.6.1.- Benton and Davis	233
6.6.2.- Aïllament de fags recombinants positius. Re-crivellatges.....	233
6.6.3.- Mini-preparació de DNA de fags.lambda.....	233
6.6.3.1.- Mini-preparació de fags mitjançant bescanvi iònic.....	234
6.6.3.2.- Mini-preparació de fags mitjançant precipitació diferencial	234
6.7.- RESTRICCIÓ.....	235
6.7.1.- Enzims emprats.....	235
6.7.2.- Digestions parcials i digestions complertes de DNA genòmic.....	235
6.7.3.- Digestions per a una anàlisi de restricció.....	236
6.7.4.- Digestions preparatives.....	236
6.7.5. - Transferència <i>Southern</i>	236
6.8.- ELECTROFORESI D'ÀCIDS NUCLEICS.....	236
6.8.1. - Electroforesi en gels d'agarosa.....	237
6.8.1.1.- Electroforesi de DNA.....	237
6.8.1.2.- Electroforesi de RNA.....	237
6.8.2.- Electroforesi en gels de poliacrilamida.....	237
6.8.2.1.- Tractament dels vidres.....	237
6.8.2.2.- Tipus de gels	238
6.8.2.3.- Marcadors de pes molecular	238
6.9.- PURIFICACIÓ DE DNA A PARTIR DE GELS D'AGAROSA O D'ACRILAMIDA	238
6.9.1.- Purificació a partir d'electroforesi en gels d'agarosa.....	238
6.9.1.1.- Agarosa de baix punt de fusió.....	238
6.9.1.2.- Electroel·lució.....	239
6.9.1.3.- DEAE-80 (Schleicher and Schuell).....	239
6.9.1.4.- Mètode del nitrògen líquid/clorur de liti.....	240
6.9.2.- Purificació d'oligonucleòtids en gels de poliacrilamida.....	240
6.10.- CLONATGE DE DNA EN L'HOSTE <i>Escherichia coli</i>.....	241
6.10.1.- Preparació del vector i lligació.....	242
6.10.2.- Transformació de bactèries competents.....	242
6.10.3.- Detecció de les colònies transformades amb plasmidi recombinant.....	243

6.10.4.- Mini-preparació de DNA plasmídic.....	243
6.10.5.- Preparació a l'engrós de DNA plasmídic.....	244
6.11.- SEQÜENCIACIÓ DE DNA PEL MÈTODE DIDEOXI DE SANGER.....	244
6.11.1.- Seqüenciació en cadena senzilla	244
6.11.1.1.- Obtenció de cadena senzilla partir del fagsmídi BLUESCRIPT.....	245
6.11.1.2.- Reacció de seqüenciació.....	245
6.11.2.- Seqüenciació de DNA bicatenari.....	245
6.11.2.1.- Preparació dels plasmidis.....	245
6.11.2.2.- Reacció de seqüenciació.....	246
6.11.3.- Progressió de les seqüències.....	247
6.11.4.- Delecions direccionals. Mètode de l'ExoIII/S1	247
6.11.4.1.- Delecions bidireccionals, sense protecció d'extrems.....	250
6.11.4.2.- Protocols d'ExoIII/S1	250
6.11.4.2.1.- Delecions unidireccionals (amb protecció d'un extrem).....	250
6.11.4.2.2.- Delecions sense protecció d'extrems.....	252
6.11.5.- Fixació, secat i autoradiografia.....	252
6.12.- OBTENCIÓ DE RNA DE DIFERENTS ESTADIS DEL DESENVOLUPA- MENT	253
6.12.1.- Preparació de RNA a gran escala	253
6.12.2.- Mini-preparació de RNA.....	253
6.12.2.1.- Mètode de l'isotiocianat de guanidina.....	253
6.12.2.2.- Mètode del fenol calent amb clorur sòdic.....	254
6.12.3.- Purificació de la fracció de RNA poliA ⁺	254
6.12.4.- Transferència <i>Northern</i>	255
6.13.- PCR. REACCIÓ DE POLIMERITZACIÓ EN CADENA.....	256
6.13.1.- Disseny dels oligonucleòtids.....	256
6.13.2.- Condicions experimentals de la PCR amb DNA genòmic.....	256
6.13.3.- Condicions experimentals de la PCR amb RNA total.....	257
6.14.- HIBRIDACIONS <i>in situ</i> SOBRE CROMOSOMES POLITÈNICS.....	258
6.15.- SOFTWARE.....	259
7.- BIBLIOGRAFIA.....	261

ABREVIATURES:

aa:	aminoàcid.
AcO:	ió acetat.
<i>Adh</i>: gen	alcoholdehidrogenasa.
ADH:	enzim alcoholdehidrogenasa.
ATP:	trifosfat de adenosina.
cDNA:	DNA còpia.
cpm:	comptes per minut
ddNTP:	trifosfat de dideoxinucleòtid.
DEPC:	dietilpirocarbonat.
DNA:	àcid desoxirribonucleic.
DNApol:	polimeraasa de DNA.
DNAsa:	desoxirribonucleasa.
dNTP:	trifosfat de deoxinuclòtid.
DTT:	ditiotreitòl.
EDTA:	tetra-acetat sòdic de dietilendiamina.
IPTG:	isopropil-tiogalactòsid.
kb:	kilobase
kD:	kilodalton.
mRNA:	RNA missatger.
nt:	nucleòtid.
pb:	parell de bases.
RNA:	àcid ribonucleic.
RNAsa:	ribonucleasa.
r.p.m.:	revolucions per minut.
SDS:	dodecilsulfat sòdic.
SM;	tampó d'elució de fags (Maniatis et al., 1982).
TBE:	tampó Tris-Borat-EDTA (Maniatis et al., 1982).
TE:	tampó Tris-EDTA (Maniatis et al., 1982).
TEMED:	N,N,N',N',tetrametiletildiamina.
X-GAL:	5-brom-4-clor-3-indol- β -d-galactopirànòsid.

INTRODUCCIÓ

INTRODUCCIÓ

1.- EL SISTEMA ALCOHOLDESHIDROGENASA (*Adh*/ADH)

1.0.- GENERALITATS

El gen alcoholdehidrogenasa (*Adh*) i el seu producte d'expressió, l'enzim ADH constitueix un dels sistemes més estudiats a eucariotes superiors i dels quals es disposa més informació a *Drosophila*. A més, el gen *Adh* és un gen estructural que presenta l'interès addicional d'estar regulat al llarg del desenvolupament i d'ésser específic de teixit. El producte d'expressió d'aquest gen, l'enzim ADH té com funció primordial la utilització metabòlica dels alcohols.

Diverses disciplines de la Biologia tals com la biologia molecular, la genètica de poblacions, la citogenètica i la bioquímica l'han considerat com un model útil per verificar hipòtesis o aprofundir en determinats aspectes d'aquestes disciplines. Així, es troben nombroses revisions sobre el sistema, ja siguin de tipus general (Chambers et al., 1988; Chambers, en premsa) o centrades en aspectes més concrets, com ara: la genètica de poblacions (McDonald, 1983; Clegg i Epperson, 1985; Lewontin, 1986); el polimorfisme enzimàtic (Van Delden, 1982; Zera et al., 1985); el metabolisme de l'alcohol (Atrian i González-Duarte, 1986); l'expressió del gen *Adh* (Laurie-Ahlberg, 1985; Savakis i Ashburner, 1985; Savakis et al., 1986; Sofer i Martin, 1987); i l'evolució molecular del gen *Adh* (Sullivan et al., 1990).

1.1.- L'ENZIM ADH

1.1.1.- Propietats

El gen *Adh* codifica per a una proteïna relativament petita, de 255 o 257 aminoàcids, segons les espècies, de pes molecular aproximat de 27.5 kDa. En algunes espècies, com *D.melanogaster*, el producte gènic és processat post-traduccionament dins de la mateixa cèl·lula, de forma que la metionina N-terminal és delecionada, quedant un residu aminoacídic acetilat en l'extrem N-terminal. Es tracta d'una proteïna relativament abundant a *Drosophila*, per exemple, a *D.melanogaster* representa entre un 1% i un 2% de la proteïna total soluble.

L'enzim funcional ADH (EC 1.1.1.1.) és un homodímer, de forma que el

pes molecular total és de 54 a 55 kDa. La seva funció és la catàlisi de l'oxidació d'alcohols primaris i secundaris a aldehids i cetones, respectivament, emprant com a cofactor el NAD⁺. Cada monòmer presenta un domini d'unió al cofactor, de manera que co-existeixen tres conformèmers segons l'enzim estigui unit a cap, una o dos molècules de cofactor. Les propietats bioquímiques i cinètiques de l'enzim ADH han estat determinades per a una gran varietat d'espècies i és més eficient en la catàlisi dels alcohols secundaris que dels primaris. Aquest és un tret peculiar donat que els alcohols primaris són presents en quantitats més elevades que els secundaris en l'hàbitat de *Drosophila*, que sol ser ric en productes de fermentació alcohòlica produïda per llevats i, per tant, enriquit en etanol.

1.1.2.- Estructura

Existeixen dos dominis principals en l'enzim ADH, un d'unió al cofactor, i l'altre és el centre catalític, però no han estat situats topològicament. Segons prediccions de l'estructura secundària de la molècula (Benyajati et al., 1981), es creu que el domini d'unió al cofactor es trobaria en l'extrem N-terminal, ja que s'hi troben residus comuns a totes les deshidrogenases NAD⁺-depenents i que es suposa estan directament implicats en la unió al NAD⁺ (Duester et al., 1986; Ribas de Pouplana et al., 1991). Per altra part, aquest extrem N-terminal presenta una alternància d'hèlix α i làmines β o *Rossmann folds* (Rossmann et al., 1975) característica d'aquest tipus d'enzims. La regió C-terminal correspondria al domini catalític de l'enzim.

La localització dels introns del gen correspon fins cert punt a una separació de dominis proteics. Un dels introns separaria l'hèlix α de la làmina β a les *Rossmann folds*, una divisió que ha estat descrita a d'altres proteïnes que s'uneixen a mononucleòtids d'adenina (Lonberg i Gilberg, 1985). L'altre intró separaria els dominis catalític i d'unió al cofactor.

1.1.3.- Funció metabòlica. Selecció de mutants.

Les funcions principals de l'ADH són la metabolització i utilització energètica dels alcohols generats per la fermentació de microorganismes en l'hàbitat de larves i adults, juntament amb la capacitat detoxificadora front concentracions elevades d'aquests compostos en el medi. S'han proposat d'altres funcions alternatives o complementàries, com ara la intervenció en la síntesi de l'hormona juvenil o en el metabolisme de feromones (Madhavan et al., 1973; Winberg et al., 1982).

Concentracions baixes d'alcohol en el medi constitueixen un recurs d'explotació com a font d'energia, però hi ha un llindar de màxima concentració, a partir del qual els efectes de toxicitat sobre els individus es fan evidents, probablement degut a l'increment directe d'acetona i acetaldehid, els metabolits produïts per l'oxidació de l'isopropanol i de l'etanol.

Per altra part, aquesta funció oxidativa ha permès dissenyar sistemes de selecció de mutants (O'Donnell et al., 1977). La selecció de mutants nuls, és a dir, que no produeixen enzim funcional, s'ha basat en la capacitat que té l'ADH d'oxidar alcohols secundaris. Aquesta selecció és negativa per als individus que produeixen ADH, ja que els metabolits produïts en l'oxidació d'alguns alcohols secundaris (penté-2-ol) són fortament tòxics. El sistema de selecció positiva dels revertents, es basa en la capacitat de supervivència dels individus en medis que contenen més del 6% d'etanol, ja que els individus nuls per a l'ADH són molt sensibles a concentracions baixes d'etanol en el medi. Aquestes seleccions són molt útils, tant per aïllar nous mutants nuls, com per a seleccionar individus transformats mitjançant elements P sobre un fons ADH-nul.

1.1.4.- Variants genètiques

A *Drosophila melanogaster* s'han descrit nombrosos al·lels del gen *Adh*, alguns codifiquen per enzims actius, ja siguin d'incidència natural o induïts en el laboratori, i d'altres, per enzims no funcionals: una llista completa es troba a Ashburner (1985). Els al·lels naturals més freqüents i també els més estudiats són les variants *fast* (ADH^F) i *slow* (ADH^S) de l'enzim ADH, anomenades així per la seva diferent mobilitat electroforètica. Aquests dos al·lels presenten una única diferència en la proteïna per la qual codifiquen, i és la substitució del residu 192 de lisina (ADH^S) a treonina (ADH^F). De fet, es considera que l'al·lel *slow* és l'ancestral, ja que l'enzim ADH d'altres espècies del gènere presenten el mateix residu aminoacídic, i el polimorfisme ADH^F/ADH^S només es troba descrit a *D.melanogaster*.

Aquest polimorfisme ha estat molt analitzat des d'una perspectiva poblacional, ja que s'han descrit distribucions clínals que s'han relacionat amb les diferències en activitat ($ADH^F > ADH^S$) i la temperatura, per això se li ha atribuït un possible paper adaptatiu.

1.1.4.- Valor adaptatiu del polimorfisme genètic

Les espècies de *Drosophila* han anat aprofitant diferents nínxols ecològics i en aquesta adaptació al medi l'activitat alcohol-deshidrogenàsica ha desenvolupat un paper important, ja que la capacitat d'aprofitar els alcohols dels hàbitats naturals ha condicionat el manteniment de les espècies. *D.melanogaster* és més tolerant a l'alcohol que d'altres espècies cosmopolites. Les espècies més freqüents en llocs molt rics en alcohol, com ara bodegues i cellers, són *D.melanogaster* i *D.lebanonensis* (Monclús i Prevosti, 1978) i ambdues destaquen per la seva gran tolerància a l'etanol. Aquesta podria ser una possible explicació de la capacitat de *D.melanogaster*, originària de zones tropicals, de colonitzar nous ambients, i trobar-se actualment àmpliament distribuïda.

1.2.- EL GEN *Adh* AL GÈNERE *Drosophila*

1.2.1.- El locus *Adh*

La localització genètica i citològica del gen *Adh* es coneix per a poques espècies del gènere *Drosophila*. A *D.melanogaster*, l'*Adh* es va situar mitjançant lligament genètic al braç esquerra del cromosoma 2, posició 50:1 (Grell et al., 1965), i aquesta localització es va confirmar posteriorment per hibridació *in situ* sobre cromosomes politènics (Goldberg et al., 1983), concretament entre les bandes 35B1-35B3.

L'anàlisi genètica i molecular acurada de la regió cromosòmica on es situa el gen *Adh* ha mostrat que mutacions a 5' i a 3' d'aquest afecten a un mateix gen, *outspread* (*osp*) (Chia et al., 1985). Aquests resultats suggereixen que aquesta regió genòmica presenta una gran complexitat estructural. En espècies del mateix subgènere, com al grup *obscura*, el gen *Adh* es localitza en la divisió 36 del cromosoma U, element homòleg al braç 2L de *D.melanogaster* (Loukas et al., 1979; Schaeffer i Aquadro, 1987). Aquesta regió és molt propera al centròmer i per tant, a regions riques en heterocromatina.

1.2.2.- Estructura del gen *Adh* a *Drosophila*

El gen *Adh* va ser descrit per primera vegada a *D.melanogaster* (Goldberg, 1980; Benyajati et al., 1980). La seva estructura consisteix en tres exons de mides diferents (99, 405 i 267 pb, respectivament) separats per dos petits introns de 65 i 70 pb (Fig.1). L'expressió d'aquest gen de còpia única està dirigida per dos promotors diferents que, segons la seva posició

respecte el triplet d'iniciació, s'anomenen proximal i distal i que s'activen de forma diferencial i alternativa durant el desenvolupament (apartat 1.2.3). La regió codificant és la mateixa, però els transcrits produïts difereixen en els seus extrems 5' no traduïts (Benyajati et al., 1983) (Fig.1). Així, el promotor distal dóna lloc a un transcrit que conté un intró adicional de 654 pb que serà processat, de forma que els transcrits madurs presenten una mida similar (el mRNA distal és lleugerament més gran) i comparteixen 37 nt que precedeixen al primer codó. Aquest intró adicional en la regió 5' no traduïda s'anomena intró de l'adult (apartat 1.2.3). Els dos promotors consten de caixa TATA, que tot i no presentar la seqüència *standard*, s'adapten a la seqüència *consensus* descrita per a gens eucariotes (Breathnach i Chambon, 1981) (Taula I). El senyal de poliadenilació també correspon a la *consensus* i es troba a 121 pb del codó de terminació.

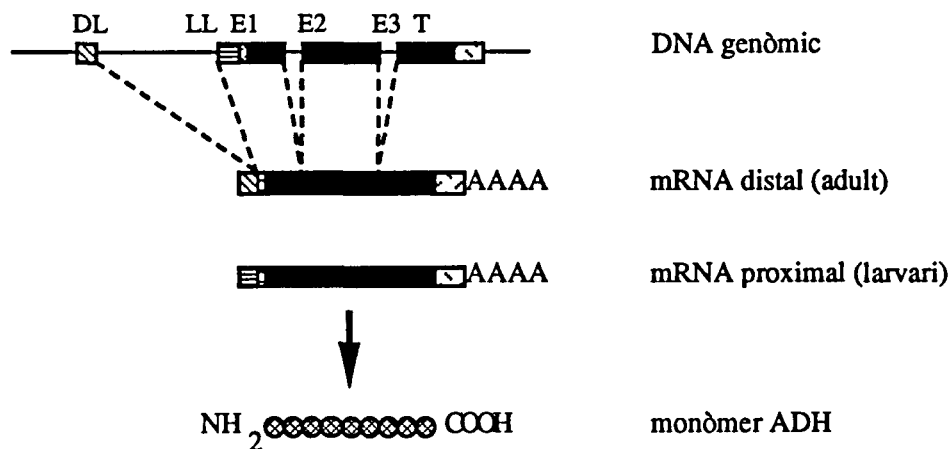


Fig.1.- Estructura del gen *Adh* i dels dos transcrits madurs, distal i proximal, a *Drosophila melanogaster*. El producte gènic és el mateix, ja que els dos transcrits només difereixen en el seu extrem 5' no traduït. DL - *leader* distal; LL- *leader* larvari; E1, E2 i E3- primer, segon i tercer exons, respectivament; T- *trailer*.

El gen *Adh* presenta aquesta mateixa estructura a d'altres espècies del mateix gènere (veure la revisió de Sullivan et al., 1990). Les diferències principals es troben en les mides dels introns, que en el cas dels que separen exons codificants només varien en pocs pb, mentre que la variació en la mida de l'intró de l'adult és més gran. Respecte la regió codificant, les longituds del segon i tercer exó estan conservades, mentre que la mida del primer pot variar en sis pb: a *D.melanogaster* i espècies properes mesura 99 pb, i a la resta d'espècies 93 pb (codificant per dos aminoàcids menys, el segon i el tercer). La seqüència de les caixes TATA es troba molt conservada, i només s'ha descrit una substitució a la caixa TATA distal de *D.lebanonensis* (Albalat i González-Duarte, 1990). A

D.pseudoobscura el senyal de poliadenilació també presenta una substitució respecte el senyal *standard* que es troba a la resta d'espècies analitzades.

TAULA I.- Comparació de la seqüència *consensus* deduïda de l'anàlisi de la composició nucleotídica en la caixa TATA de 60 gens eucariotes, segons Breathnach i Chambon (1981) amb les dues caixes TATA del gen *Adh* a *Drosophila melanogaster*.

Base Frequency	A	13	8	4	58	4	51	38	53	30	20	11
	T	16	6	49	1	56	6	22	6	20	7	9
	G	23	32	1	1	0	0	0	0	8	23	29
	C	8	14	6	0	0	3	0	1	2	10	11
<i>consensus</i>		G	G	T	A	T	A	A	A	A	G	G
		T	C					T		T	A	
Adult promoter		a	G	T	A	T	t	T	A	A	G	G
Larval promoter		T	a	T	A	a	A	T	A	c	G	G

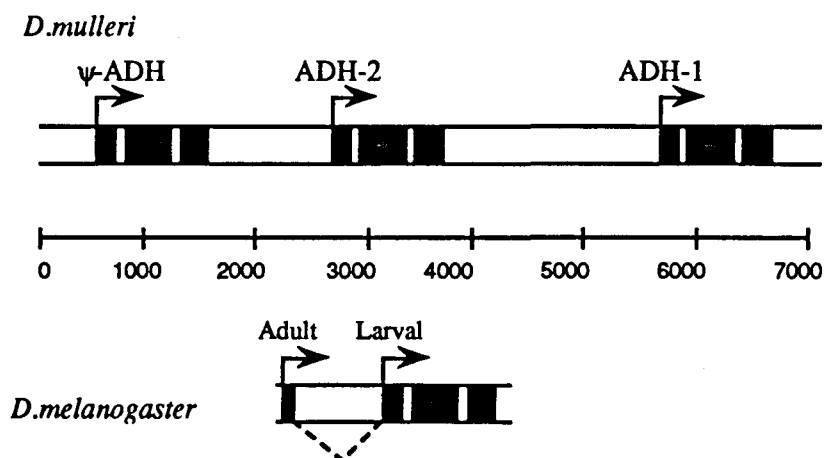


Fig.2.- Organització estructural de l'*Adh* a *D.melanogaster* (gen únic amb dos promotors) i *D.mulleri* (dos gens funcionals, cadascun amb un promotor i un pseudogen).

L'organització estructural del gen *Adh* a espècies del grup *repleta* és l'única excepció a aquest model general. En aquestes espècies el locus *Adh* ha sofert un fenomen de duplicació de forma que presenta dos gens funcionals, cadascun amb el seu promotor i a més, un pseudogen, ψ -*Adh* (Fig.2) (Batterham et al., 1984; Fisher i Maniatis, 1985; Atkinson et al., 1988; Menotti-Raymond et al.,

1990). Les caixes TATA d'aquests dos gens són similars entre elles i més semblants a la caixa TATA proximal del gen *Adh* de *D.melanogaster*, mentre que el pseudogen no presenta caixa TATA.

1.2.3.- Transcripció espacial i temporal

L'expressió del gen *Adh* a *D.melanogaster* i espècies amb organització similar, ve regulada per dos promotors diferents, que s'activen de forma alternativa durant el desenvolupament. Aquests dos promotors, s'anomenen segons la seva posició respecte la regió codificant: promotor distal, d'expressió majoritària en l'estadi adult, i el promotor proximal, d'expressió majoritària en l'estadi larvari, però que manté una expressió basal en l'estadi adult (Fig.3) (Savakis i Ashburner, 1985; Savakis et al., 1986). De fet, el patró d'expressió és força complex: en l'ou i de procedència materna, s'hi troben els dos tipus de transcrits, el proximal i el distal, en l'embrió es produeix una petita expressió distal, però sobretot, s'incrementa progressivament l'expressió a partir del promotor proximal fins a arribar a un màxim i decreix bruscament al final de la larva III coincidint amb un petit pic d'expressió distal. Durant l'estadi de pupa l'expressió és molt baixa, quasi nul·la, i torna a incrementar-se fins a nivells alts quan eclosiona l'adult, a partir principalment del promotor distal, i a nivells basals, a partir del promotor proximal. En els ovaris de la femella adulta, s'expressen ambdós transcrits en quantitats equivalents, els quals s'acumulen en els ous.

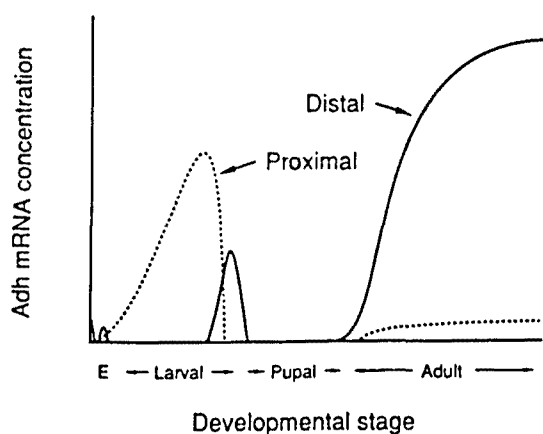


Fig.3.- Representació esquemàtica de l'expressió del gen *Adh* al llarg del desenvolupament a *D.melanogaster*. Les unitats de l'escala són relatives. La línia puntejada indica el transcrit del promotor proximal i la contínua, el del promotor distal (extret de Corbin i Maniatis, 1989a).

A les espècies del grup *repleta* on s'ha descrit la duplicació del gen *Adh*, el patró d'expressió és relativament similar al descrit per *D.melanogaster* (Fisher i Maniatis, 1985). L'ordenació dels gens, situats adjacentment, de 5' a 3' és la següent: ψ -*Adh*, *Adh-2* and *Adh-1* (Fig.2). El pseudogen es transcriu a nivells basals (malgrat no té caixa TATA) però no es tradueix. El gen *Adh-2* presenta

expressió a l'estadi adult, de forma comparable al promotor distal de *D.melanogaster*, però els nivells d'expressió no són tan elevats. El gen *Adh-1* presenta expressió específica d'estadi larvari.

L'expressió espacial del gen *Adh* és relativament consistent amb la funció i activitat detectada per l'enzim. Aquesta expressió és elevada en les cèl·lules del cos gras en totes les espècies analitzades però presenta variacions inter-específiques en l'expressió en d'altres teixits, com ara l'intestí mig, tubs de Malpighi i tractes genitals o reproductors (Ursprung et al., 1970; Maroni et al., 1982; Dickinson, 1980; Rabinow and Dickinson, 1981; Rabinow and Dickinson, 1986; Brennan i Dickinson, 1988; Visa, 1991).

1.2.4.- Regulació de l'expressió

La regulació de l'expressió de la majoria dels gens eucariotes ve determinada per la unió en trans de proteïnes a determinades regions de DNA, situades en cis. El gen *Adh* no és una excepció. S'han emprat tècniques diferents per analitzar els factors i regions implicats en la regulació transcripcional. Els primers treballs es van dirigir a l'obtenció de mutants per a l'expressió del gen, segons el sistema exposat al punt 1.1.3, però tots els mutants aïllats corresponien a mutacions de la regió codificant (Schwartz i Sofer, 1976).

D'altres mètodes han produït resultats d'interès. Per un cantó, l'obtenció d'híbrids entre espècies que presentin diferències en el seu patró d'expressió temporal i/o tissular (Dickinson i Carson, 1979; Rabinow i Dickinson, 1981; Rabinow i Dickinson, 1986; Dickinson et al., 1984) encara que una limitació important és que no totes les espècies poden fer híbrids viables.

També s'han emprat tècniques de genètica reversa, és a dir, construccions genètiques en què, o bé només hi ha part de les potencials regions promotores o reguladores de l'expressió, o bé alguna de les seqüències identificades com a reguladores ha estat mutagenitzada *in vitro*. Aquesta tècnica ha donat bons resultats gràcies al desenvolupament de la microinjecció amb elements P que permet la inserció estable de les construccions en el genoma i l'establiment de soques i que, per tant, complementa els resultats obtinguts amb diferents construccions d'expressió temporal (Goldberg et al., 1983; Benyajati i Dray, 1984; Posakony et al., 1985; Sofer i Martin, 1987; Fisher i Maniatis, 1988; Shen et al., 1989).

Els resultats obtinguts amb aquestes tècniques han permès establir l'existència de diferents regions en cis implicades en la regulació de l'expressió del gen *Adh* a *D.melanogaster* (Fig.4). Una de les regions es troba a unes 2 kb en direcció 5' del promotor distal i es tractaria d'un *enhancer* que actuaria sobre el promotor proximal. Per altra part, elements de l'especificitat tissular es trobarien situats més aprop del promotor distal, aproximadament a uns 660 pb i a més dos elements d'expressió temporal regularien l'activitat de les regions promotores. Un d'ells afectaria al promotor distal i es troba situat aprop de la regió implicada en l'especificitat tissular, i un altre element del promotor proximal es trobaria dins de l'intró de l'adult, a uns 386 pb del promotor proximal. També s'ha determinat la localització de dos caixes reguladores de l'expressió del gen *Adh-1* a *D.mulleri* (Fisher i Maniatis, 1988) (Fig.5) com exemple de la regulació de l'expressió a les espècies amb duplicació gènica: les caixes A i B. La caixa B actuaria com a *enhancer* específic de l'expressió en cos gras, l'òrgan on l'expressió d'aquest gen és majoritària, i la caixa A tindria com funció ampliar l'expressió tissular a d'altres teixits, com ara tubs de Malpighi i intestí mig. Un altre element situat a 3' del gen tindria una funció similar a la caixa B, inclús són intercanviables, però la caixa A, no té cap element anàleg. En aquesta espècie també s'ha descrit la presència d'un *enhancer* de l'expressió del gen *Adh-2*, situat en una regió a 5' del pseudogen i a una distància similar a l'element *enhancer* de *D.melanogaster*, a unes 2 kb del promotor distal.

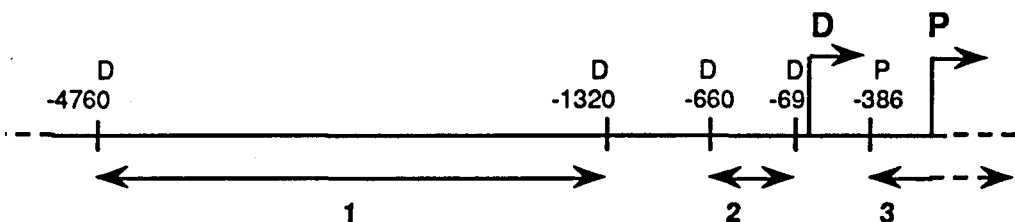


Fig. 4.- Regions flanquejants a 5' del gen *Adh* a *D.melanogaster*. Els punts d'inici de la transcripció distal i proximal estan representats per fletxes anomenades D i P, respectivament. Les xifres indiquen la distància en nombre de nucleòtids respecte els inicis distal (D) i proximal (P). Les dobles fletxes inferiors numerades 1, 2 i 3 representen la localització de les regions reguladores en cis (veure text). (Adaptat de Posakony et al., 1985).

Força interessants són els resultats de transformació inter-específica, és a dir, la microinjecció de construccions genètiques d'una espècie en una altra (Dickinson i Carson, 1979; Posakony et al., 1985; Fisher i Maniatis, 1986), que han permès deduir que els elements en cis de l'espècie donadora responen de forma

correcta als elements en trans de l'espècie receptora, la qual cosa permet generalitzar els resultats dins del gènere *Drosophila*.

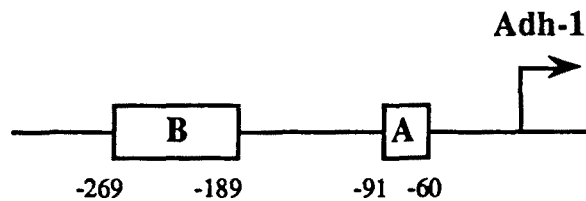


Fig.5.- Posicions dels *enhancers* Box A i Box B en la regió 5' del promotor del gen *Adh-1* de *D.mulleri* (Fisher i Maniatis, 1985).

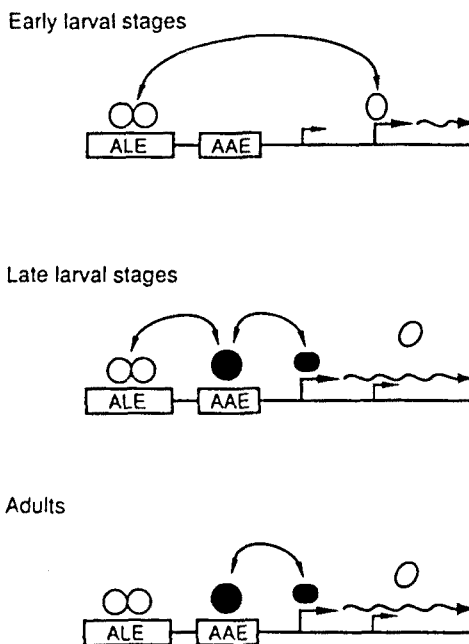


Fig.6.- Model de les interaccions *enhancer-promotor* en el gen *Adh* al llarg del desenvolupament a *D.melanogaster*. AAE-*enhancer* de l'adult, que estimula la transcripció des del promotor distal. ALE-*enhancer* larvari, que estimula la transcripció des del promotor proximal. Els cercles i ovals blancs són els factors d'unió a l'*enhancer* larvari i els factors d'unió al promotor proximal, respectivament. El cercle i l'oval negres són els factors d'unió a l'*enhancer* d'adult i el promotor distal, respectivament (extret de Corbin i Maniatis, 1989b).

L'alternància de promotors ha estat també objecte d'estudi. Els resultats obtinguts per Corbin i Maniatis (1989a) a *D.melanogaster* apunten cap a una interferència transcripcional: l'activació del promotor distal inactivaria o reprimiria l'expressió des del promotor proximal. S'han descrit dos *enhancers* principals: un d'expressió distal, AAE, localitzat entre -590 i -128 pb del promotor distal, i l'altre proximal, ALE, localitzat entre -5,000 i -660 pb del promotor distal. Aquest *enhancer* d'expressió proximal, a més, s'ha subdividit en tres segments que interactuarien de forma sinèrgica (Corbin i Maniatis, 1989a). S'ha proposat un model en el qual els factors d'unió a l'*enhancer* larvari i al promotor proximal estarien presents tant a l'estadi larvari com al d'adult, mentre que els factors d'unió a l'*enhancer* d'adult i el promotor distal serien específics de l'estadi adult,

de forma que la unió d'aquests últims a partir dels últims estadis larvaris iniciaria la transcripció des del promotor distal, desestabilitzant o interferint la transcripció des del promotor proximal, que quedaria només a nivells basals (Fig.6)(Corbin i Maniatis, 1989b). També s'ha apuntat la importància relativa de l'organització de la cromatina respecte l'activació d'un o altre promotor (Gasser i Laemmli, 1986), ja que la posició dels nucleosomes i l'empaquetament del DNA podria variar l'accessibilitat de les seqüències diana als factors de transcripció (Fig.7) (Benyajati et al., 1987; Ewel et al., 1990).

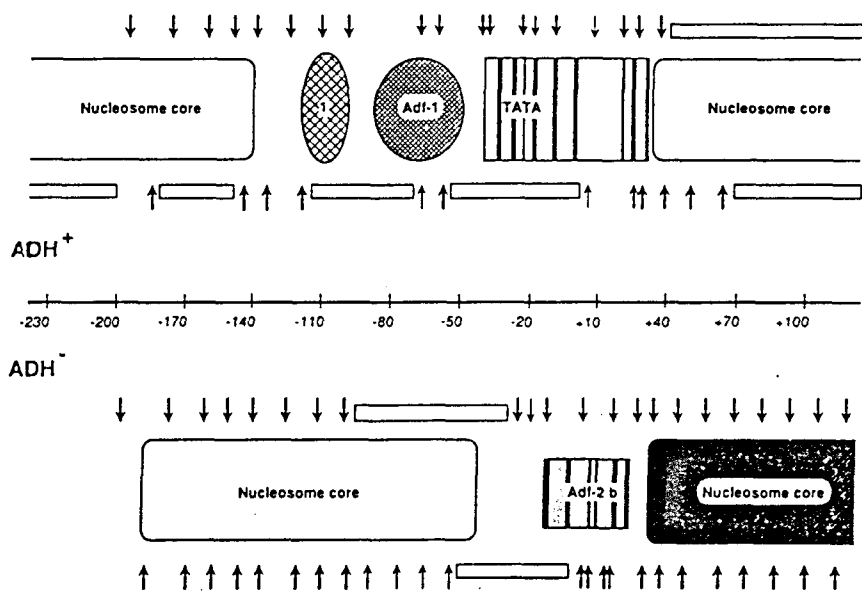


Fig.7.- Representació esquemàtica de les diferents interaccions DNA-proteïna a la cromatina del promotor distal del gen *Adh* a *D.melanogaster*. per a cèl·lules ADH⁺ i ADH⁻. Es presenten dades combinades d'accessibilitat de restricció, protecció front exonucleases, i talls específics per DNasaI (fletxes) i protecció front DNasaI (barres obertes). La distància indica el nombre de nucleòtids des de l'inici de transcripció distal (extret de Ewel et al., 1990).

L'estudi dels factors proteics implicats en l'expressió del gen *Adh* ha permès l'aïllament d'un factor d'expressió (Heberlein et al., 1985; Heberlein i Tjian, 1988) responsable de la regulació del promotor distal, l'Adf-1 (*Adh distal factor*) i s'ha determinat els llocs d'unió d'aquest factor (Fig.8), que concorden amb algunes de les regions reguladores en cis proposades (Fig.4). Sembla que aquest factor també regularia l'expressió d'altres gens, doncs també presenta punts d'unió a regions promotores dels gens *Antennapaedia* i *Dopadecarboxilasa*.

A més, l'expressió del gen *Adh* és induïble per l'etanol, encara que no pels metabolits produïts per la seva oxidació. S'ha determinat que l'element de

resposta a l'etanol, comprèn un màxim de 30 pb i es situa a 5' del promotor proximal, i sembla estar reprimint per l'enhancer ALE (enhancer d'expressió proximal). L'enhancer AAE sembla que no presenta resposta a l'etanol (Kapoun et al., 1990).

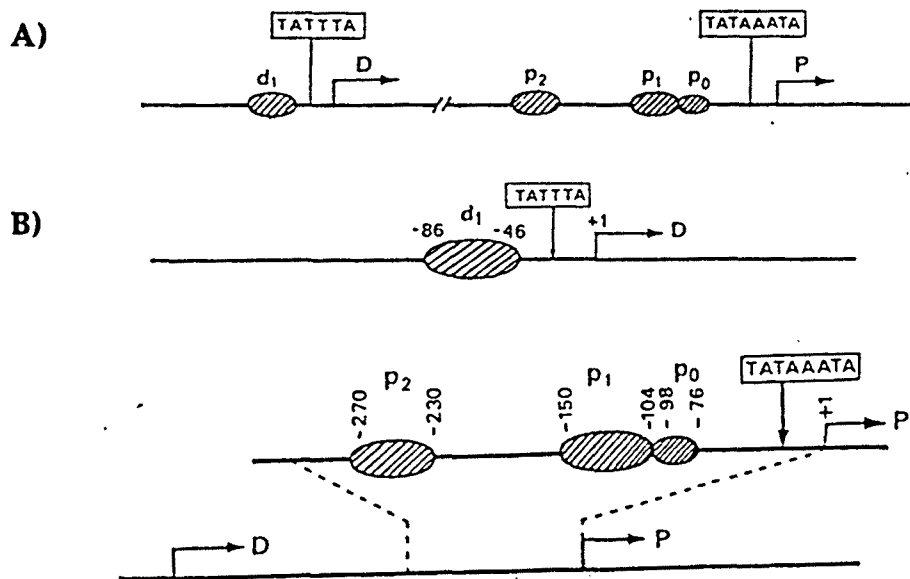


Fig.8- Llocs d'unió a proteïnes en la regió dels promotors del gen *Adh* a *D.melanogaster*. A) visió conjunta. B) unió al promotor distal i proximal en detall. D i P indiquen inicis de transcripció distal i proximal, respectivament. Adf-1 reconeix i es pot unir als llocs p1 i d1 (extret de Heberlein et al., 1985).

1.3.- EVOLUCIÓ MOLECULAR DEL GEN *Adh*

Les seqüències del gen *Adh* i del seu producte gènic de què es disposa per a establir comparacions són: *D.melanogaster* (Goldberg, 1980; Kreitman, 1983); *D.mauritiana* i *D.simulans* (Cohn et al., 1984; Cohn i Moore, 1988); *D.orena* (Bodmer i Ashburner, 1984); *D.mulleri* (Fisher i Maniatis, 1985); *D.sechellia* (Coyne i Kreitman, 1986); *D.pseudoobscura* (Schaeffer i Aquadro, 1987); *D.mojavensis* (Atkinson et al., 1988); *D.affinidisjuncta* (Rowan i Dickinson, 1988); *D.lebanonensis* (Vilarroya et al., 1989; Albalat i González-Duarte, 1990; Juan et al., 1990); *D.hydei* (Menotti-Raymond et al., 1991); *D.planitibia*, *D.heteroneura*, *D.differens*, *D.picticornis* and *D.silvestris* (Rowan i Hunt, 1991).

1.3.1.- Comparació de seqüències de proteïnes

S'ha dut a terme la comparació de diverses seqüències d'ADH, bé per seqüenciació directa de la proteïna (Vilarroya i Juan, 1991), bé per

traducció conceptual a partir de la seqüència de nucleòtids (Sullivan et al., 1990). Les seqüències analitzades no són representatives de tot el gènere *Drosophila*, perquè hi ha grups molt més estudiats que d'altres. Així, el subgrup *melanogaster* i el grup *repleta* han estat objecte de nombrosos estudis. El grup d'espècies de *Drosophila* que darrerament s'han analitzat són: *D.pseudoobscura* del subgrup *obscura*, l'espècie *D.lebanonensis* i el grup hawaiià (*picture-winged*).

La comparació de seqüències proteiques evidencia una gran conservació de seqüència primària, amb 170 identitats de 254 posicions, és a dir, una mitjana d'identitats del 67% de la molècula. És de destacar que la proteïna ADH del subgrup de *D.melanogaster* presenta una addició de dos aminoàcids en l'extrem N-terminal que han estat posicionats com a segon i tercer aminoàcids (Schaeffer i Aquadro, 1987) o com tercer i quart (Sullivan et al., 1990). Quan es considera la distribució d'identitats al llarg dels tres exons, sembla que el tercer exó és el més conservat, i el primer exó el que menys es conserva, però estadísticament aquesta distribució no es desvia significativament de la produïda per un efecte atzarós. L'única espècie de les esmentades que presenta una distribució no a l'atzar de substitucions és *D.pseudoobscura*, que en presenta una acumulació significativa al tercer exó.

Aquesta conservació de l'estructura primària també es correspon també a una elevada conservació en l'estructura secundària. Fins i tot canvis que en principi semblarien no conservatius, no varien l'estructura secundària de la molècula. Desafortunadament, no es disposa encara de l'estructura terciària.

En conjunt, la conclusió que sembla més clara és que l'alta conservació de la seqüència de la proteïna es deuria a una forta selecció purificadora (McDonald i Kreitman, 1991). De totes maneres, sembla que hi ha determinats residus (aproximadament uns 10 dins de les seqüències analitzades) que han sofert substitució diverses vegades al llarg de l'evolució i de forma independent. Aquestes posicions aminoacídiques podrien correspondre a llocs de selecció neutra.

1.3.2.- Comparació de seqüències de DNA de regions codificants

La comparació de seqüències és molt valuosa per establir relacions filogenètiques, donat que no es disposa d'un registre fòssil adient dels insectes, en general, i de *Drosophila*, en particular. En aquest sentit i com ja ha estat esmentat, l'*Adh* és el gen més estudiat a *Drosophila*, i per tant, les seves

dades de comparació de seqüència són molt informatives a l'hora de proposar arbres filogenètics.

L'alineament de seqüències genòmiques complertes només és possible quan les espècies comparades són molt properes, ja que les regions no codificants, no sotmeses, en principi, a constrenyiment funcional, acumulen mutacions, especialment de tipus inserció/deleció, que dificulten en gran mesura un alineament correcte. És per això que les dades de tipus evolutiu, per establir relacions filogenètiques en un grup d'espècies, es basen en la comparació de seqüències codificants.

Les substitucions silencioses són molt més nombroses que les de reemplaçament, com és d'esperar, degut al constrenyiment funcional. Aproximadament són superiors en un ordre de magnitud. Una conclusió que sembla deduir-se de l'evolució de les seqüències codificants als insectes, és que el seu DNA (*Drosophila*) evoluciona amb una taxa molt més ràpida que el DNA de mamífers (Sharp i Li, 1989), de forma que les taxes i temps evolutius calculats per als gens d'aquests darrers, no es poden aplicar en l'evolució dels primers. A més, el gen *Adh* sembla evolucionar més lentament dins del DNA de *Drosophila* i encara que aquest aspecte no permeti deduir taxes o temps evolutius fiables de la comparació de les seqüències d'*Adh*, sí permet, en canvi, proposar arbres filogenètics.

Un possible problema a tenir en compte és el polimorfisme intra-específic: per a les espècies en les quals s'ha seqüenciat més d'un clon (Kreitman, 1983), existeixen variacions puntuals, que habitualment són substitucions d'un únic nucleòtid en posicions silencioses, però que de fet demostren que cada individu pot posseir una seqüència única. Un exemple excepcional en aquest sentit, és el polimorfisme més estès a *D.melanogaster*, ADH^F/ADH^S , originat per una única substitució en la regió codificant del gen *Adh* en una posició de reemplaçament. El fet que aquest polimorfisme presenti un desequilibri de lligament elevat amb d'altres polimorfismes silenciosos a zones no codificants, fa pensar o bé en un origen molt antic del polimorfisme i un posterior aïllament geogràfic o genètic dels dos al·lels de forma que no hi hagi hagut recombinació entre ells, o bé en algun tipus de selecció conjunta (Kreitman i Hudson, 1991). Com a conseqüència paral·lela, el problema sorgeix sobre quina és la seqüència *standard* que s'ha de considerar per a la comparació inter-específica.

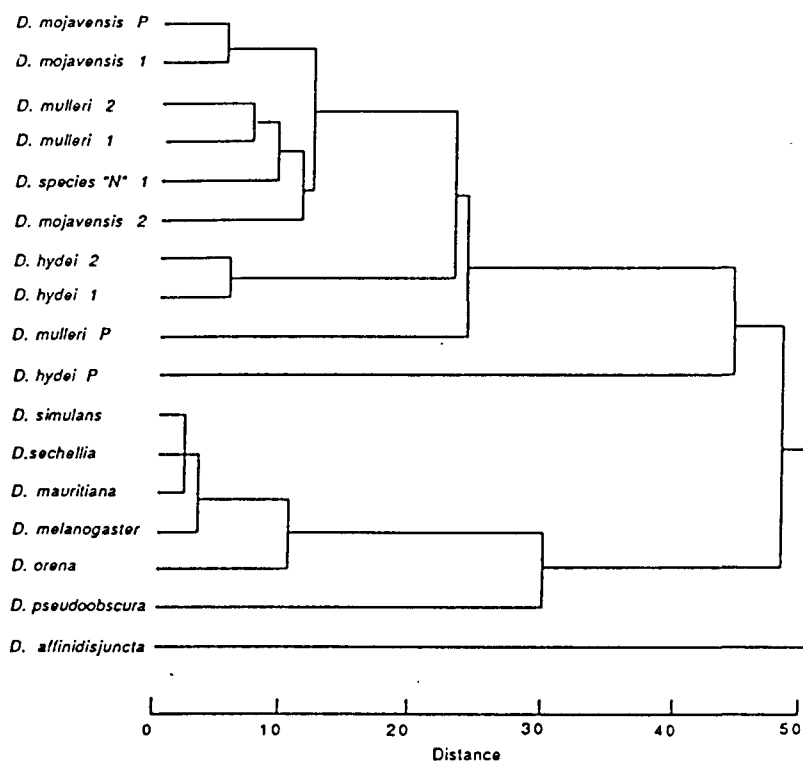


Fig.9.- Arbre filogenètic sense arrel de 17 seqüències del gen *Adh* de diverses espècies del gènere *Drosophila*, basat en les distàncies de la seqüència sencera de DNA (unitats arbitràries). "N" és *D.navojoa*. El programa emprat és el FITCH (mètode de Fitch-Margoliash) del paquet PHYLIP (J.Felsenstein) (extret de Sullivan et al., 1990).

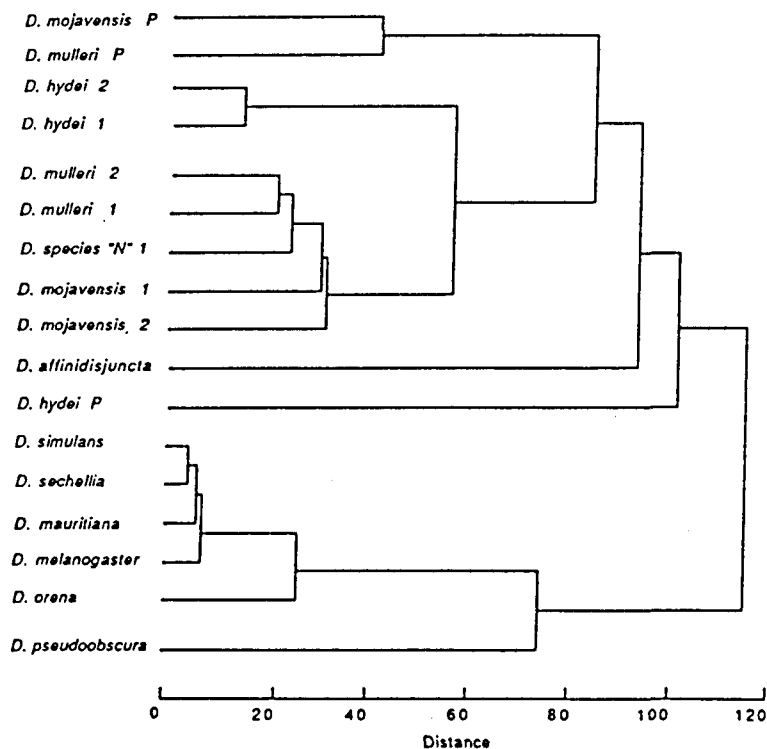


Fig.10.- Arbre filogenètic sense arrel de 17 seqüències del gen *Adh* de diverses espècies del gènere *Drosophila*, basat en les distàncies a partir de les posicions silencioses (unitats arbitràries). "N" és *D.navojoa*. El programa emprat és el FITCH (mètode de Fitch-Margoliash) del paquet PHYLIP (J.Felsenstein) (extret de Sullivan et al., 1990).

Les dades emprades en la construcció d'arbres filogenètics condicionen l'arbre o arbres resultants. Es pot emprar tota la seqüència codificant o bé, només tenir en compte parts informatives d'aquesta, com ara són: les posicions silencioses o les de reemplaçament. Per exemple, l'arbre filogenètic deduït emprant les dades de similitud en tota la regió codificant del gen *Adh* es troba a la Fig.9 (Sullivan et al., 1990), mentre que l'obtingut amb les substitucions silencioses únicament, es troba a la Fig.10 (Sullivan et al., 1990). Com es pot veure la posició de *D.affinidisjuncta* varia considerablement, així com també la posició relativa de les espècies del grup *repleta*.

1.3.3.- Evolució de seqüències reguladores

Últimament s'ha donat més importància com a responsables de l'especiació a les diferències en el patró de regulació gènica que a les diferències en la seqüència primària dels gens. Conseqüentment, s'han analitzat amb cert detall les regions potencialment implicades en la regulació de l'expressió de gens amb diversitat fenotípica inter-específica, com ara el gen *Adh* des d'un punt de vista evolutiu. Els experiments realitzats fins ara (apartat 1.2.4) permeten deduir que les diferències inter-específiques en el patró d'expressió d'aquest gen es deuen a diferències en els elements cis-reguladors.

La duplicació gènica en les espècies del grup *repleta* permet un estudi d'aquest tipus. S'ha trobat que la seqüència del gen *Adh-2* de *D.mulleri* (d'expressió majoritària en adult) és homòloga a la del promotor proximal (d'expressió màxima en l'estadi larvari) però de funció anàloga a la del promotor distal de *D.melanogaster*. En canvi, a *D.hydei*, l'expressió del gen *Adh-2* comença a l'estadi larvari i continua al llarg de tot l'estadi adult com si tingués dos promotors quan en realitat només en té un (homòleg al promotor proximal de *D.melanogaster*) i a més, presenta una organització en dos gens i un pseudogen similar a la que presenta *D.mulleri*. Aquests resultats indiquen que hi ha regions reguladores en cis, que no són pròpiament les promotores, responsables de l'especificitat d'expressió.

El gen *Adh-1* de *D.mulleri* és actiu principalment durant l'estadi larvari, i l'expressió des del seu promotor ve regulada per un enhancer situat a 3' del gen. El patró d'expressió d'aquest gen respecte a la seva activació temporal i especificitat de teixit és molt similar al del promotor proximal de *D.melanogaster*. El fet que regions situades a 5' del promotor del gen *Adh-1* de *D.mulleri* funcionin quan són introduïdes en *D.melanogaster* indica que malgrat la

divergència de la seqüència i la independència de la seva organització gènica, els elements reguladors en cis de diferents espècies reconeixen al mateix tipus de factors, i per tant, hi ha una conservació dels elements reguladors, al menys respecte la seva funció. Aquests experiments han estat també realitzats amb regions presumptivament promotores de *D.mojavensis* i *D.affinidisjuncta* i els resultats obtinguts són plenament confirmatoris. Fins avui encara falta definir exactament aquestes seqüències implicades en la regulació de l'expressió i trobar els factors de transcripció equivalents a Adf-1 a d'altres espècies.

1.3.4.- Evolució de la regió genòmica del gen *Adh*

Aquesta regió genòmica ha sofert un gran nombre de duplicacions i reorganitzacions al llarg de la seva evolució. A les espècies del grup *repleta* s'ha descrit una duplicació gènica mitjançant la qual en lloc d'un únic gen funcional amb dos promotors alternatius, existien dos gens, cadascun amb el seu propi promotor. Com ja s'ha esmentat, el resultat de l'expressió conjunta dels dos gens és molt similar a la que presenta un sol gen amb dos promotors, tant pel que fa a l'expressió temporal com tissular. Una anàlisi extensa a diferents espècies del grup *repleta* va permetre determinar que aquesta duplicació es deuria haver produït en els orígens d'aquest grup (Batterham et al., 1984), ja que es presenta a dos dels quatre subgrups que el formen, en concret als subgrups *mulleri* i *hydei*.

S'ha seqüenciat la regió sencera que comprèn els gens *Adh* de tres espècies: *D.mulleri* (Fisher i Maniatis, 1985; *D.mojavensis* (Atkinson et al., 1988) and *D.hydei* (Menotti-Raymond i Sullivan, 1991) i el gen *Adh-1* de l'espècie *D.navojoa* (Sullivan et al., 1990). A partir de la comparació de les seqüències s'han proposat diferents models. D'una banda, hom suggereix un model per a l'evolució del gen *Adh* al subgrup *mulleri* (Atkinson et al., 1988) assumint que l'enzim ADH és necessari tant en l'estadi larvari com l'adult, i segons el qual l'organització ancestral seria la d'un únic gen *Adh* (com a *D.melanogaster*). Un model alternatiu parteix de dos gens ancestrals amb dos promotors i la pèrdua d'un dels dos explicaria l'organització actual del gen *Adh* a *D.melanogaster*, amb un únic gen i dos promotors. No obstant, aquest últim model està amb contradicció en la similitud dels dos gens *Adh-1* i *Adh-2* del grup *repleta*, semblança que reforçaria la hipòtesi d'una divergència relativament recent. Per altra part, el fet que aquesta duplicació gènica només es trobi en aquestes espècies concretes, que a més estan relacionades entre sí, implicaria que la pèrdua d'un gen estructural s'ha donat de forma convergent i paral·lela en diferents grups d'espècies allunyades, i la possibilitat d'un fenomen d'aquest tipus és altament

improbable.

Vist que dels dos gens i el pseudogen derivats del gen *Adh* al grup *repleta*, només el pseudogen presenta seqüències homòlogues al promotor distal de *D.melanogaster*, que els dos promotors dels gens funcionals presenten seqüències homòlogues al promotor proximal, i tenint també en compte que les seqüències dels gens *Adh-1* i *Adh-2* s'assemblen més entre sí que al pseudogen, Atkinson et al. (1988) i Sullivan et al. (1990) proposen un model de passos consecutius (Fig.11). El primer pas seria la duplicació del gen únic de forma que s'hauria perdut el promotor proximal en la duplicació i s'obtidrien dos gens, un situat més a 5' amb un promotor distal (que donaria lloc al pseudogen posteriorment) i el situat més a 3' amb el promotor proximal. Després es produiria una segona duplicació d'aquest gen situat més a 3', donant lloc als ancestres dels gens *Adh-1* i *Adh-2*. El tercer fenòmen seria la inactivació del gen situat més a 5' que derivaria cap al pseudogen actual. Sembla ser que aquesta inactivació va tenir lloc un temps després de la primera duplicació de forma que durant algun temps un cert tipus de pressió de selecció va actuar sobre aquesta seqüència en els seus orígens. La inactivació de la funcionalitat del pseudogen pot deure's a una delecció (d'uns 870 pb) que aproxima el promotor distal del pseudogen al codó d'iniciació, i més clarament a un canvi en la pauta de lectura ocasionat per la delecció d'un nucleòtid. Aquests fenòmens es van produir anteriorment a la separació dels subgrups *mulleri* i *hydei*.

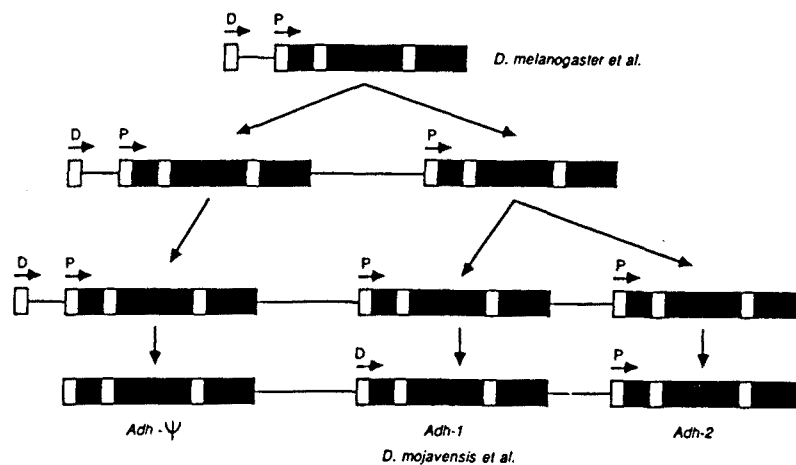


Fig.11.- Model de l'evolució del gen *Adh* per explicar l'organització actual a *D.melanogaster* i *D.mojavensis* (del subgrup *repleta*). P i D són els promotors proximal i distal, respectivament (extret d'Atkinson et al., 1988).

A més, també han tingut lloc fenòmens de conversió gènica com és d'esperar quan gens duplicats es troben tan propers en el genoma, malgrat aquesta no sigui una explicació plausible per explicar la semblança general dels gens *Adh-1* i *Adh-2*. En canvi, s'accepten dos fenòmens independents de conversió gènica. Un d'ells va ésser proposat dins dels gens de *D.mulleri* (Fisher i Maniatis, 1985) per explicar l'elevat grau de similitud entre exó 1, intró 1 i part de l'exó 2 dels gens *Adh-1* i *Adh-2* d'aquesta espècie. Una altra possible conversió gènica podria haver succeït a *D.hydei*, ja que els exons i introns dels dos gens són molt més similars entre ells que els seus homòlegs a *D.mojavensis* i *D.mulleri*, i suposar una duplicació de gens independent és menys parsimoniós que pressuposar una possible conversió gènica a *D.hydei*.

També s'ha descrit una duplicació gènica incompleta a *D.affinidisjuncta* (Rowan i Dickinson, 1988), espècie que pertany al grup hawaià o *picture-winged* de *Drosophila*. L'organització gènica que presenta és similar a la descrita per *D.melanogaster*, és a dir, un sol gen funcional amb dos promotors. Aquesta duplicació compendria part de la regió 5' del gen *Adh*.

Per últim, també mereix menció una duplicació encara més ancestral, que ha estat descrita per primera vegada a *D.pseudoobscura* (Schaeffer i Aquadro, 1987) i que també s'ha trobat a *D.mauritiana*, *D.simulans* (Cohn i Moore, 1988) i a *D.melanogaster* (Kreitman i Hudson, 1991) i que probablement es trobi en quasi totes les espècies del gènere. Aquesta duplicació ancestral va originar un nou gen situat a 400 pb cap a 3' del gen *Adh* i es fonamenta no només en la gran similitud de seqüència a nivell nucleotídic (més del 50%) i aminoacídic (més del 35%) sinó també en la mateixa disposició dels exons i introns. Malgrat encara no es coneix la seva possible funció, la gran conservació de seqüència i de totes les característiques que acompanyen a un gen que s'expressa suggereix que es tracta d'un gen funcional.

1.3.5.- L'origen evolutiu del gen *Adh* de *Drosophila*

L'enzim ADH de *Drosophila* és un enzim clarament diferent del seus anàlegs a llevat o mamífers, tant respecte les seves propietats catalítiques i estereoquímiques com estructurals. En primer lloc, s'ha de destacar el seu baix pes molecular i la no dependència d'ions metàl·lics, en concret del zenc, en clar contrast amb les característiques compartides per l'ADH d'aquells organismes. Tot apunta cap a un cas de convergència evolutiva.

Una similitud estructural més elevada es troba quan es compara l'ADH de *Drosophila* amb altres proteïnes, com ara la proteïna de 25 kD de *Sarcophaga peregrina*, un altre dípter, (Henikoff i Wallace, 1988) o d'altres deshidrogenases com la 15-OH-prostaglandina deshidrogenasa humana, o la ribitol deshidrogenasa de *Klebsiella*.

L'únic model per a explicar l'origen de l'ADH a *Drosophila* ha estat proposat per Jörnvall i col·laboradors (Jeffrey i Jörnvall, 1983; Jörnvall et al., 1983), basant-se en la diferenciació de dos tipus de deshidrogenases: de cadena llarga i de cadena curta. A partir d'una seqüència original han tingut lloc duplicacions successives, divergència de seqüències i convergència de funcions, produint finalment diverses alcohol i poliol deshidrogenases que es poden agrupar en dos grans grups: les deshidrogenases de cadena llarga, metal·loenzims, com ara les ADH de llevat i mamífer i les deshidrogenases de cadena curta, sense metalls, com l'ADH de *Drosophila* i la ribitol deshidrogenasa de *Klebsiella*. Aquest model es veu reforçat pel fet que els dos grups de deshidrogenases comparteixen molt pocs residus comuns fora dels possiblement implicats en la unió al cofactor i al substrat.

2.- EL SUBGRUP *obscura*

2.0.- GENERALITATS

D'ençà 1823, en que Fallèn va capturar a Suècia, una espècie de *Drosophila* a la qual va anomenar *D.obscura*, el grup *obscura* ha estat l'objecte d'intensa investigació, especialment a partir dels treballs de Dobzhansky i col·laboradors durant la primera meitat d'aquest segle.

Posteriorment, la informació sobre aquest grup s'ha anat incrementant procedint des de moltes disciplines de la Biologia (per revisions, Buzzatti-Traverso i Scossiroli, 1955; Lakovaara i Saura, 1982). Una dada il·lustrativa és que mentre que a 1955 Buzzati-Traverso i Scossiroli definien dinou espècies en aquest grup, a 1984, Krimbas i Loukas ja en comptabilitzaven trenta-dues.

La primera descripció diferenciadora de les espècies del grup *obscura*, va ésser proporcionada per Fallèn, en descriure *D.obscura* com una espècie fàcilment distingible d'altres espècies europees de *Drosophila*, a causa del color fosc del seu cos i a la presència de dues pintes sexuals en la part frontal dels tarses dels mascles. Malgrat la gran quantitat d'informació sobre molts aspectes d'aquest grup, existeix un punt crucial que continua éssent polèmic: les relacions filogenètiques.

El grup *obscura* es troba àmpliament distribuït pel món, tant a les regions geogràfiques Paleàrtica com la Neàrtica, i sembla estar ben establerta l'existència de dos subgrups: *obscura* i *affinis*. Tots dos subgrups tenen representació en els dos supracontinents, encara que la seva importància relativa varia: mentre al continent paleàrtic dominen les espècies del subgrup *obscura* (només hi ha un representant del subgrup *affinis*), al continent americà predominen en nombre les espècies del subgrup *affinis*. Les espècies més estudiades pertanyen al subgrup *obscura* i són aquelles que presenten una distribució més gran, *D.pseudoobscura* al continent nord-americà i *D.subobscura* a la regió paleàrtica, tot i que molt recentment, aquesta última espècie està colonitzant amb gran èxit el continent americà (revisat a Prevosti et al., 1989). Totes les espècies analitzades en el present treball pertanyen al subgrup *obscura* i són de distribució paleàrtica.

D.subobscura va ésser descrita per primer cop per Collin (1936) a Anglaterra. És l'espècie més present a l'Europa occidental, i es troba distribuïda des de Suècia fins al Nord d'Àfrica, des d'Iran a les Illes Canàries i des del nivell

del mar fins als boscos d'alta muntanya. Presenta com a particularitats, que necessita llum per pondre ous, i es tracta d'una espècie de clima temperat, tot i que presenta una gran adaptabilitat a diferents temperatures (segons races).

D.guanche (Monclús, 1976) és originària de les Illes Canàries (illa de Tenerife) i *D.madeirensis* (Monclús, 1984) és endèmica de les illes Madeira. Totes dues espècies poden viure en gran varietat d'hàbitats, però en els seus llocs d'origen es troben associades a la laurisilva.

D.ambigua (Pomini, 1940) també pot viure en una àmplia gama de condicions ecològiques, des de la Península Ibèrica, fins Holanda i Gran Bretanya, i també fins Austria. Malgrat no ésser cosmopolita presenta una distribució geogràfica molt àmplia.

Totes quatre espècies tenen cinc parells de cromosomes, bé acrocèntrics o metacèntrics i dos cromosomes puntiformes. Es creu que la disposició cromosòmica de *D.subobscura* és la bàsica per al gènere *Drosophila* (Sturtevant i Novitsky, 1940). Per altra part aquesta espècie es caracteritza per un elevadíssim grau de polimorfisme cromosòmic, degut principalment a un gran nombre d'inversions, que afecten a tots els cromosomes i que ha fet necessària una nomenclatura per a ordenar-les. La causa d'aquest elevat polimorfisme seria la presència de letals equilibrats. Existeixen clines en la distribució d'aquests polimorfismes cromosòmics, que semblen tenir un valor adaptatiu (Prevosti et al., 1988).

2.1.- ESTUDI DE LES RELACIONS FILOGENÈTIQUES

Com ja s'ha esmentat, l'establiment de filogènies és l'aspecte evolutiu més polèmic del grup *obscura*. S'han proposat nombroses filogènies emprant diferents tipus de dades, des de dades morfològiques i citològiques (Buzzati-Traverso i Scossiroli, 1955; Lakovaara i Saura, 1982; Krimbas i Loukas, 1984; Steineman et al., 1984), electroforètiques (Cabrera et al., 1983; Loukas et al., 1984; Loukas et al. 1986), i més recentment dades derivades d'estudis de variabilitat a nivell de DNA, ja sigui mitocondrial (Latorre et al., 1988; González et al., 1990), o bé nuclear (Goddard et al., 1990).

Les conclusions més generals, deduïdes inicialment de les comparacions a nivell morfològic (Buzzati-Traverso, 1955) i que més tard han estat ampliadades per altres tipus de dades són: a) els dos subgrups *-obscura* i *affinis-* es diferencien més

entre ells que les espècies d'un mateix subgrup que pertanyen a continents diferents (això confirma la consistència dels dos subgrups), b) les espècies del subgrup *obscura* han divergit més inter- que intra-continentalment, i per tant, l'origen dels dos subgrups i les radiacions dins de cada subgrup són esdeveniments antics, des d'un punt de vista evolutiu.

Els principals punts actualment conflictius són d'una banda, la ubicació de *D.ambigua*, una espècie paleàrtica que presenta més proximitat en alguns aspectes amb les espècies neàrtiques d'aquest subgrup, que amb les corresponents espècies europees, i d'altra banda, la situació de proximitat relativa de la tríade *D.subobscura*, *D.guanche* i *D.madeirensis*, espècies que tots els autors assenyalen com molt properes i que en termes evolutius s'han separat recentment.

La situació de *D.ambigua* varia segons el tipus de dades que es fa servir en l'establiment de filogènies. El fet que és l'única espècie europea que pot fer híbrids (encara que estèrils) amb el grup d'espècies americanes *D.pseudoobscura*-*D.persimilis*-*D.miranda*, quan no pot fer-ne amb cap espècie europea, ha fet que nombrosos autors la situessin a prop de les espècies neàrtiques. No obstant, quan es tenen en compte dades morfològiques (Buzzati-Traverso i Scossioli, 1955; Lakovaara i Saura, 1982) la similitud entre *D.ambigua* i aquest grup d'espècies americanes és menor que amb les espècies europees, i s'assembla més a *D.obscura* i *D.tristis*. Aquesta proximitat també es reflecteix en les comparacions a partir de dades electroforètiques (Fig.12) (Loukas et al., 1984) encara que hi ha autors (Cabrera et al., 1983) que situen *D.ambigua* més a prop de les espècies americanes (Fig.13). Les dades obtingudes de l'anàlisi del mtDNA (Fig.14), situen *D.ambigua* amb *D.obscura* i *D.subsilvestris* però no amb *D.tristis* i embrancada més a prop de *D.subobscura* i la resta d'espècies europees, que de *D.pseudoobscura* i espècies relacionades. Per últim, els resultats obtinguts per hibridació d'àcids nucleics (Fig.15) (Goddard et al., 1990) semblen contradir aquesta posició, ja que apunten cap a una més gran proximitat de *D.subobscura* amb *D.pseudoobscura* que cap d'elles amb *D.ambigua*. La majoria d'autors remarca, però, que la distància entre grups d'espècies relacionades europees i americanes és molt similar i suporta la hipòtesi que l'ancestre del grup va originar múltiples línies paleàrtiques i, una d'aquestes línies va colonitzar el continent americà molt poc després d'aquesta radiació.

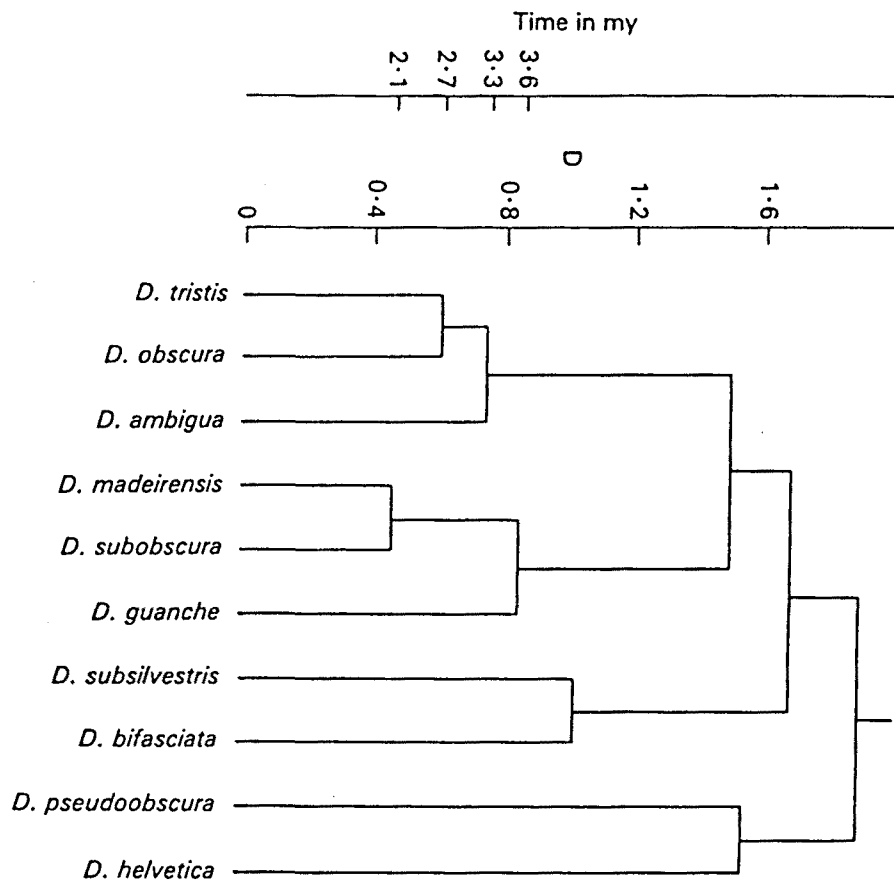


Fig.12.- Arbre filogenètic de 10 espècies de *Drosophila* del grup *obscura*. L'escala de temps, en milions d'anys, s'ha estimat a partir de dades electroforètiques mitjançant el mètode de UPGMA (extret de Loukas et al., 1984).

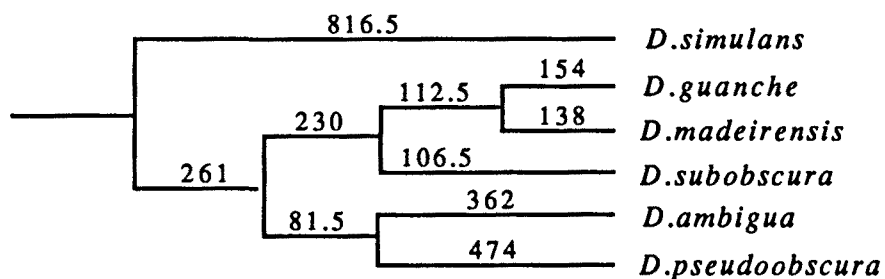


Fig.13.- Arbre filogenètic de 6 espècies del gènere *Drosophila* basat en el càlcul de distàncies (Nei, 1972) a partir de dades electroforètiques. Les xifres sobre cada branca són les distàncies patristiques multiplicades x 1000 (Farris, 1972) (adaptat de Cabrera et al., 1983).

Els treballs sobre la tríade d'espècies *D.subobscura*, *D.madeirensis* i *D.guanche* són més recents, ja que la descripció d'aquestes dos últimes espècies s'ha fet durant els últims 10 a 15 anys. Tots els autors confirmen la proximitat d'aquesta tríade que pot fer híbrids (*D.madeirensis* pot fer híbrids amb *D.subobscura*, fèrtils, i amb *D.guanche*, estèrils) encara que la posició relativa entre elles és discutida. Krimbas i Loukas (1984) i Loukas et al. (1984) fan un estudi comparatiu, ajuntant dades morfològiques, citològiques i electroforètiques, i

arriben a la conclusió que *D.guanche* es va separar primer de la branca que després donaria *D.subobscura* i *D.madeirensis* (Fig.12), en contraposició amb resultats previs (Cabrera et al.,1983) que suposen que *D.madeirensis* està més propera a *D.guanche* que a *D.subobscura* (Fig.13) però amb una clara concordància amb els resultats obtinguts de l'anàlisi del mtDNA (Fig.14) (González et al., 1990). Per tant, tot i que en un principi podia semblar que *D.madeirensis* estava més relacionada amb *D.guanche* (potser degut a la seva distribució reduïda i a un possible efecte fundador), com més dades es coneixen sembla confirmar-se la seva més gran proximitat a *D.subobscura*.

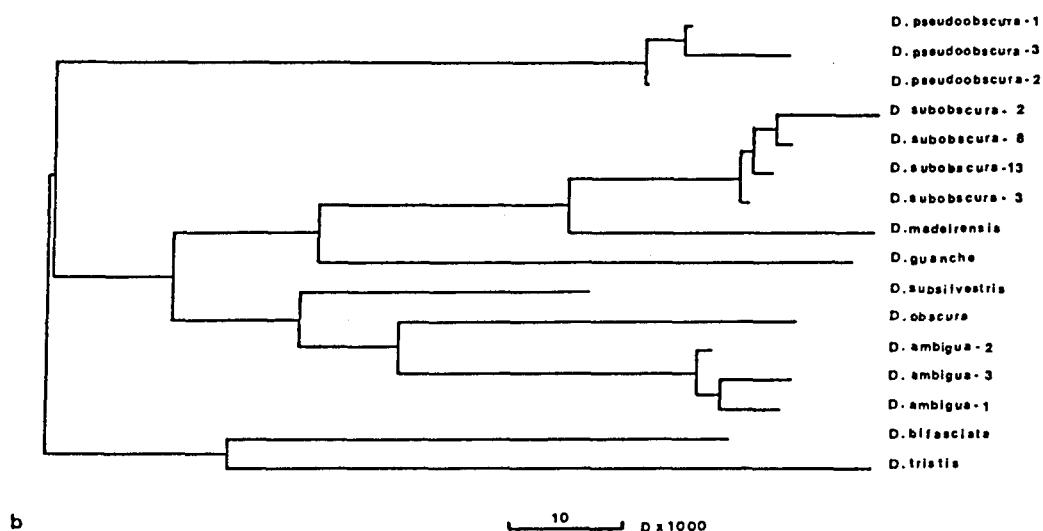


Fig.14.- Arbre filogenètic de 9 espècies de *Drosophila* pertanyents al subgrup *obscura* a partir de dades obtingudes de l'anàlisi del DNA mitocondrial, mitjançant el mètode de neighbour-joining (extret de González et al., 1990).

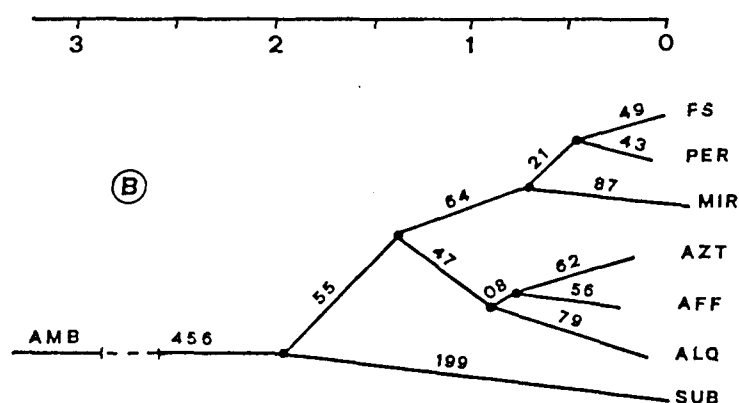


Fig.15.- Arbre filogenètic de 7 espècies de *Drosophila* pertanyents al grup *obscura* a partir de dades obtingudes de l'anàlisi de la reassociació del DNA, mitjançant el programa FITCH del paquet PHYLIP (J. Felsenstein). Les distàncies de les branques són 1/100 unitats de ΔT_m (extret de Goddard et al., 1990). AMB-*D.ambigua*; PS-*D.pseudoobscura*; PER-*D.persimilis*; MIR- *D.miranda*; AZT-*D.azteca*; AFF-*D.affinis*; ALQ-*D.algonquin*; SUB-*D.subobscura*.

3.- FENÒMENS DE RETROTRANSCRIPCIÓ I RETROPOSICIÓ

3.0.- GENERALITATS

Retrotranscripció és un terme general que s'aplica a diversos fenòmens que tenen com a base la còpia d'un transcrit (RNA) a DNA, mitjançant un mecanisme de transcripció inversa. Aquest terme es va crear per definir un pas essencial del mecanisme de replicació dels retrovirus i alguns virus de RNA i DNA i, posteriorment, es va ampliar a d'altres elements genètics, tals com retrotransposons, retroposons, etc.

L'aïllament i caracterització d'aquests elements ha permès conèixer els efectes que produeixen en els genomes. Així, se'ls relaciona amb reordenacions cromosòmiques de tots tipus, l'origen del DNA repetit, la transmissió horitzontal d'informació, l'aïllament genètic per disgènesi híbrida, l'especiació i processos de disrupció gènica, etc. Evolutivament són particularment interessants des del punt de vista de la complexitat i soroll de fons que produeixen sobre els genomes, éssent la causa de l'aparició de noves combinacions de seqüències de DNA que constitueixen material genètic en brut sobre el qual actua la selecció natural i la deriva genètica.

Els termes retrotransposició i retroposició s'apliquen a dos fenòmens: retrotranscripció i incorporació en el genoma, del DNA que s'ha copiat a partir del transcrit. Aquesta dualitat pot ésser inherent al tipus de DNA retrotranscrit, i per això, sota el nom de retrotransposó i retroposó s'hi han acollit molts tipus diferents de DNA que s'han retrotranscrit i posteriorment integrat en el genoma. Molts dels elements genètics que es retrotransposen, posseeixen les capacitats per retrotranscriure's i per incorporar-se en el genoma, perquè codifiquen per a una transcriptasa inversa i una integrasa. Aquest cas específic s'anomena retrotransposició i és el mecanisme habitual de replicació i proliferació d'aquests elements genètics. D'altres, no codifiquen per cap d'aquestes activitats o només per a una d'elles i s'han anomenat retroposons, i en aquest terme s'inclouen seqüències tipus LINEs (seqüències llargues disperses), SINEs (seqüències curtes disperses), els pseudogens de snRNA (RNA petit nuclear), els introns de gens ribosomals, pseudogens de gens estructurals, etc.

Existeixen grans revisions sobre el tema de la retroposició (Rogers, 1985; Vanin, 1985; Weiner et al., 1986; Wagner 1986) i darrerament s'ha elaborat i proposat una nomenclatura més específica i clarificadora de les característiques de

cada tipus de retrotranscrit (Taula II) (Hull i Will, 1989; Temin, 1989). Els elements genètics que aquí es comentaran són retroseudogens de gens estructurals que no tenen capacitat inherent per retrotranscriure's ni integrar-se en el genoma i per tant queden inclosos dins de l'apartat de les retroseqüències.

TAULA II.- Classificació de retroelements i retroseqüències (adaptat de Temin, 1989 i Hull i Will, 1989).

Element	Transcriptasa		Presència			Exemples
	inversa	Transposició	de LTRs	Virions		
Retró	sí	no	no	no	no	msDNA
Retroposó	sí	sí	no	no	no	SINEs, LINEs,
Retrotransposó	sí	sí	sí	no	no	Ty, còpia,...
Retrovirus	sí	sí	sí	sí (RNA)	sí	HIV, RSV,...
Pararetrovirus	sí	no	sí	sí (DNA)	sí	Hepatitis B
Retroseqüència	no	no	no	no	no	gens de cDNA

3.1.- RETROPOSICIÓ DE GENS ESTRUCTURALS: RETROSEQÜÈNCIES.

Com ja s'ha esmentat, els retroseudogens, o seqüències derivades de retrotranscrits de gens estructurals formen part de les retroseqüències, les quals no codifiquen per cap activitat transcriptasa inversa ni integrasa. Hi ha una sèrie de trets que caracteritzen aquests retroseudogens, éssent els més definitoris, l'absència de regions promotores reconeixibles, al menys, en aquells gens en què els promotors es troben fóra de la regió transcrita; la total o parcial absència d'introns, que estan perfectament escindits; la presència de la cua poliadenilada, típica dels transcrits madurs; i en alguns casos, una repetició de seqüències curtes de DNA en els extrems de les retroseqüències. Les tres primeres característiques denoten que s'han originat d'un transcrit, en principi, missatge madur, mentre que l'últim tret són les restes indicatius de la seva integració en el genoma.

En el genoma dels mamífers existeixen nombroses famílies multigèniques, molts membres dels quals no són funcionals (gens de tubulines, actines, immunoglobulines, citocroms, proteïnes ribosomals, enzims com la gliceraldehid-3-fosfat deshidrogenasa, etc). Apart dels gens estructurals amb funció, es poden detectar pseudogens genòmics (l'origen dels quals és una duplicació gènica seguida d'una posterior divergència i inactivació d'una de les còpies) i també retroseudogens (d'origen retrotranscripcional). Originalment, les

retroseqüències van ésser descrites en genomes de mamífers, i només es coneixia algun altre exemple dispers en aus. Es va hipotetitzar que aquest tipus d'elements genètics només es trobaria en el genoma de vertebrats (Vanin, 1985; Wagner, 1986).

Quatre punts es consideren rellevants respecte a l'origen de les retroseqüències: a) Com s'han format i integrat, b) Com han esdevingut heretables, és a dir, com s'han incorporat dins el genoma de la línia germinal, c) Quina és l'antiguitat d'aquests fenòmens i d) Si tenen o han tingut alguna funció.

La resposta al primer punt pressuposa que aquestes seqüències han passat per un RNA intermediari que s'ha copiat a cDNA (Fig.16). L'activitat transcriptasa inversa procediria d'altres elements genètics que codifiquen per un enzim amb aquestes característiques. La integració en el genoma podria donar-se en llocs de trencament a l'atzar en el cromosoma, principalment zones riques en polidA, que podrien unir-se d'alguna manera amb extrems del cDNA polidT. Una altra alternativa és que el mateix mRNA fos el que reconegués una zona rica en poli-dT amb la seva cua poliadenilada, i que allí mateix fos retrotranscrit i insertat.

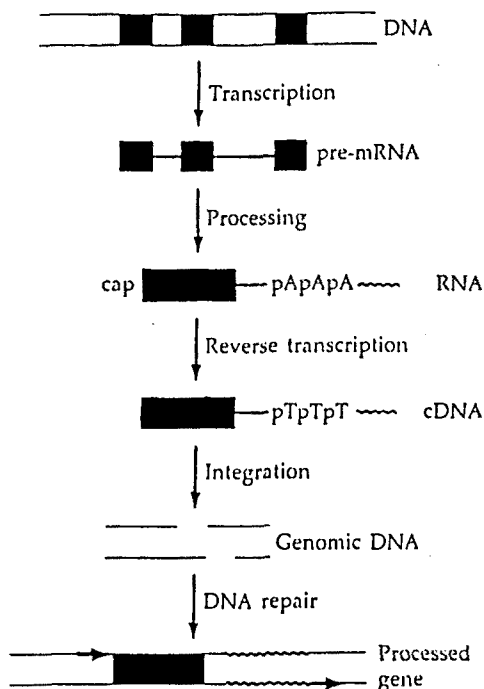


Fig.16.- Model per explicar l'origen d'una retroseqüència. Les caixes negres representen els exons d'un gen. La línia ondulada representa la cua de poli-dA, i la seva complementària, poli-dT en el cDNA. El DNA és transcrit en pre-mRNA, el qual és processat cap a mRNA. Aquest és transcrit inversament cap a cDNA per una transcriptasa inversa i s'integrarà en una localització diferent del DNA genòmic, que és reparat per un procés que duplica la diana d'inscripció (fletxes horitzontals) (extret de Li i Graur, 1990).

Si el gen del qual deriva la retroseqüència s'expressa en la línia germinal, pot explicar-se fàcilment com una còpia de cDNA d'aquest gen ha esdevingut heretable. Però, si no és aquest el cas, s'han proposat d'altres explicacions alternatives com ara: la transcripció basal d'alguns gens, o bé la presència de promotors vírics (deguts a la inserció de retrovirus) prop de gens inactius que els activaria en aquestes cèl·lules, o bé l'actuació d'una altra RNA polimerasa que originaria transcrits aberrants a partir d'aquests gens inactius en la línia germinal.

Es coneixen molt pocs exemples de funcionalitat d'aquestes retroseqüències, o que tant sols es transcriguin. Apart dels gens de la preproinsulina de pollastre i rata (Perler et al., 1980; Bento-Soares et al., 1985) i la prointerleukina 1 β humana (Clark et al., 1986), els pocs exemples dels quals es disposa solen ser gens de transcripció ubiqua -sense una regió promotora concreta- (Stein et al., 1983; Gruskin et al., 1987) o bé gens d'expressió específica en la línia germinal -com ara a testicles-(McCarrey i Thomas, 1987; Boer et al., 1987; Ashworth et al., 1990). La necessitat d'una regió promotora per a la transcripció es pot obviar sempre i quan hi hagi possibles substituïts, com ara, que hi hagi un promotor fort, d'origen víric o no, en les proximitats del lloc d'integració, que dins de la regió transcrita hi hagi seqüències presumiblement promotores, o bé que els transcrits derivin de gens d'expressió ubiqua que no empren seqüències promotores definides. Per altra part, la majoria d'autors que han establert comparacions entre retroseqüències i els gens precursors han trobat una gran similitud de seqüència que suggereix que són producte de fenòmens recents en termes de temps evolutiu. Curiosament, fins i tot en retroseqüències que no presenten transcripció ni funcionalitat coneguda, l'anàlisi evolutiva de la seva divergència amb el gen ancestre, evidencia tendències evolutives conservadores, és a dir, que durant algun període de temps, algun tipus de constrenyiment va actuar sobre la seva evolució (Miyata i Yasunaga, 1981; Li et al., 1981; Li i Johnson, 1989).

3.2.- RETROPOSICIÓ A *Drosophila*

A *Drosophila* s'han descrit elements genètics que es retrotranscriuen (per revisió Berg i Howe, 1989). Existeixen retrotransposons, que contenen repeticions terminals llargues (LTR) i codifiquen per a proteïnes similars a les dels retrovirus, com *copia*, *gypsy*, etc., i retroposons, elements mòbils sense LTR i que codifiquen per proteïnes menys semblants a les de retrovirus, classificables dins de les seqüències LINEs, com són entre d'altres, els elements F (DiNocera et al.,

1983), els I (Fawcett et al., 1986) i els elements R1 i R2 (inicialment, inserts de tipus I i II) que interrompen i inactiven els gens de RNA ribosomal. La majoria d'aquests últims elements similars als LINEs, presenten la particularitat de contenir les seqüències promotores necessàries per a l'inici de transcripció dins la regió que es transcriu, de forma que es poden retrotransposar sense pèrdua de la independència transcripcional.

No obstant, malgrat que s'han descrit activitats retro-transcriptases (Finnegan, 1989), encara no s'han descrit retroseqüències derivades de gens estructurals a *Drosophila*. En fort contrast amb el que s'esdevé a mamífers, on s'han trobat de 20 a 30 retroseqüències dels gens de les tubulines i fins a 200 retroseqüències de la gliceraldehid-3-fosfat deshidrogenasa, a *Drosophila* encara no se n'ha descrit cap, tot i que el nombre de gens que codifiquen per a les α - i β -tubulines, o per a la gliceraldehid-3-fosfat deshidrogenasa és similar al de mamífers. Aquest conjunt de dades havia fet suggerir que els retroseudogens de gens estructurals només es trobarien en genomes de mamífers, o com a molt, en genomes de vertebrats.

4.- DNA MITJANAMENT REPETIT EN EL GENOMA

4.0.- GENERALITATS: DNA REPETIT. POSSIBLES FUNCIONS.

Existeix una enorme varietat en la mida, nombre de còpies i organització d'elements de DNA repetit dels genomes eucariotes (per revisions, Jelinek i Schmid, 1982; Singer, 1982; Deininger i Daniels, 1986). Les seqüències de DNA repetit varien des d'una longitud menor de 10 pb fins a mides superiors a 1 kb. El seu nombre també pot variar en 10^6 ordres de magnitud, i la seva distribució pot estar ordenada en grups o tàndems, o bé estar dispersa en el genoma. Totes aquestes diferències evidencien un origen i possible funció molt diferent. Una altra característica del DNA repetit força interessant és la seva variabilitat i ràpida evolució, de forma que cada espècie presenta variacions en el tipus i seqüència de DNA repetit. Aquest tret l'ha validat com a marcador molecular per establir relacions filogenètiques.

Segons el tipus de DNA repetit s'han proposat diverses funcions, fins i tot s'ha suposat l'absència de funció. S'ha parlat de DNA egoista o paràsit per il·lustrar que aquest DNA no aconsegueix cap funció excepte la de mantenir-se en el genoma i augmentar el seu nombre, emprant una gran diversitat de mètodes. Probablement aquesta visió és excessivament simplista. Segons el tipus de DNA repetit se li ha assignat una sèrie de funcions entre les que sobresurten: contribuir al manteniment de l'estructura, el correcte aparellament i la recombinació legítima dels cromosomes durant la mitosi i la meiosi; afavorir reordenacions genòmiques per recombinació no homòloga o il·legítima, originant, per tant, l'aparició de noves combinacions gèniques; contenir regions d'unió amb la matriu nuclear, zona rica en enzims de tot tipus que tenen el DNA com substrat, permetent una millor funcionalitat; tamponar l'actuació d'agents mutagènics o destructors del DNA poc específics; intervenir en la regulació de l'expressió gènica, bé passivament (empaquetament de l'heterocromatina) o activament (per contenir elements reguladors o promotors); contribuir a l'especiació, etc.

4.1.- DNA MITJANAMENT REPETIT. TIPUS I ORIGEN.

Es coneixen dos tipus principals de DNA mitjanament repetits, els LINES i els SINES. És evident que existeix més d'un origen de les repeticions llargues i curtes (LINES i SINES, respectivament) donada la seva dissimilitud estructural. S'han proposat diversos mecanismes sobre el seu origen: transposició, duplicacions gèniques, reordenacions cromosòmiques, integracions de genomes vírics, o generació independent. Darrerament, s'han aportat dades que

involucren processos implicats en la mobilitat del DNA, com ara transposicions i DNAs transponibles, com possibles agents causants de bona part de les famílies de DNA repetit (revisions a Rogers, 1985; Sun et al., 1984; Weiner et al., 1986).

Algunes de les famílies de SINEs més ben caracteritzades a genomes de mamífer (Alu, Alu-like, B2,...) deriven de gens de RNA ribosomals o de RNA de transferència, i s'han proposat mecanismes de retroposició per a explicar la seva amplificació en el genoma. Algunes hipòtesis pressuposen que, a partir d'un gen "progenitor", es formaria per successives mutacions un pseudogen amb capacitat latent per esdevenir amplificable, incorporant noves alteracions. De moment, no es coneix quin tipus de canvi seria necessari ni com s'activaria aquest potencial ancestre de les repeticions. Però, en canvi, proporciona una explicació plausible de l'existència de diferents famílies relacionades de SINEs en espècies més o menys properes.

Els elements LINEs es troben àmpliament distribuïts en el genoma dels eucariotes, i es caracteritzen per contenir pautes de lectura oberta. Tot i que alguns membres de les famílies presenten una o més pautes truncades, la seqüència de les pautes de lectura sembla evolucionar sota constrenyiment funcional. No se sabia com aquests elements es podien transcriure i transposar, però cada vegada es coneixen més casos d'elements retrotransponibles amb promotors críptics de posició interna a la regió que es transcriu.

Un dels aspectes més interessants d'aquestes famílies de DNA és la conservació de les seqüències dels membres d'una mateixa família en el genoma. S'han proposat diversos mecanismes com ara la recombinació no homòloga i la conversió gènica, però no s'han aportat evidències clares. Per altra part, sembla que la majoria de famílies de DNA mitjanament repetit s'han originat recentment en el temps evolutiu, i que probablement la seva dispersió en el genoma emmascara restes d'altres famílies anteriors.

4.2.- DNA MITJANAMENT REPETIT EN EL GENOMA DE *Drosophila*.

Com a promig, només el 9% del genoma de *Drosophila* està constituït per DNA mitjanament repetit. Existeixen fins un centenar de famílies, cadascuna de pocs membres. Moltes d'aquestes famílies estan constituïdes per elements mòbils, amb una longitud superior a una kb. El patró de dispersió que presenta el DNA repetit de *Drosophila* rep el nom de dispersió de llarg abastament (*long-range*) ja que els DNA repetits d'una mida considerable es troben separats per fragments

llargs de DNA de còpia única o baixa. Les zones de DNA repetit poden estar constituïdes per agrupacions d'elements pertanyents a diferents famílies (Young, 1979; Wensink et al., 1979; Singer i Berg, 1991).

Segons Spradling i Rubin (1981), el DNA mitjanament repetit de *Drosophila melanogaster*, es subdivideix en tres grups: a) un està constituït per gens de còpia múltiple, que solen trobar-se agrupats, b) la major part està dispers en el genoma, dins d'aquest grup es troben els elements transponibles i c) seqüències que formen part de l'heterocromatina dels cromosomes. Malgrat que es té poca informació sobre l'organització del DNA repetit a d'altres espècies de *Drosophila* (Martin et al., 1983) s'han trobat diferències en la distribució cromosòmica dels elements repetits que confirmarien també que una proporció important del DNA repetit està constituït per famílies d'elements mòbils (Felger i Sperlich, 1989).

OBJECTIUS

OBJECTIUS

Les espècies de *Drosophila*: *D.guanche* i *D.madeirensis* han estat descrites recentment. Són espècies endèmiques de les illes Canàries i Madeira, respectivament, són bessones de *D.subobscura* i han estat situades entre les espècies del subgrup *obscura* que ocupen el Vell Mon. L'anàlisi de les relacions filogenètiques entre elles i respecte el subgrup *melanogaster* és de gran interès. *D.ambigua* és la quarta espècie triada per a aquest tipus d'anàlisi. Les raons d'aquesta elecció es basen en el fet que també és una espècie paleàrtica del mateix subgrup, però que al mateix temps presenta semblances importants amb algunes espècies americanes que estan incloses també en el mateix subgrup. La posició filogenètica d'aquesta espècie dins del subgrup *obscura* és, encara avui, objecte de debat.

L'objecte d'aquest treball ha estat l'anàlisi de les relacions filogenètiques entre aquestes espècies a partir de l'evolució d'una regió genòmica que conté un gen estructural funcional (*Adh*) i en posició adjacent, una còpia addicional d'aquest gen (*Adh-dup*) que conté les regions reguladores necessàries per a la seva expressió, malgrat la seva funció és desconeguda. El gen *Adh* i el seu producte estructural, en canvi, han estat àmpliament analitzats i són ben coneguts pel nostre grup de treball.

Així, doncs, els objectius concrets d'aquest treball són:

- 1.- Anàlisi de l'estructura i organització del gen *Adh* a les espècies *D.ambigua*, *D.subobscura*, *D.madeirensis* i *D.guanche*, mitjançant la construcció de genoteques, aïllament de clons amb homologia al gen *Adh*, i posterior anàlisi de restricció i seqüenciació d'aquests clons positius.
- 2.- Comparació de l'estructura del gen *Adh* entre aquestes espècies, i amb d'altres representants dels subgrups *obscura* i *melanogaster*.
- 3.- Aportació de noves dades moleculars per a la clarificació de les relacions filogenètiques de les espècies analitzades dins el subgrup *obscura* de *Drosophila*.

CAPÍTOL I: "The *Adh* in *Drosophila*: Chromosomal location and restriction analysis in species with different phylogenetic relationships"

Publicat a *Cromosoma* 100: 315-322 (1991)

The *Adh* in *Drosophila*: Chromosomal location and restriction analysis in species with different phylogenetic relationships

N. Visa, G. Marfany, L. Vilageliu, R. Albalat, S. Atrian, and R. González-Duarte

Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, Diagonal 645, E-08071 Barcelona, Spain

Received August 9, 1990

Accepted January 11, 1991 by H. Jäckle

Abstract. Restriction analysis of the genomic region containing the *Adh* gene and in situ hybridization assays were performed in six *Drosophila* species belonging to three different subgenera: *D. ambigua*, *D. subobscura*, *D. madeirensis* and *D. guanche* (sg. *Sophophora*); *D. immigrans* (sg. *Drosophila*); and *D. lebanonensis* (sg. *Pholidoris*). In agreement with previous observations, comparison of restriction maps of the *Adh* region shows that *D. subobscura* and *D. madeirensis* are very closely related. Partial homology is also observed with the rest of the *obscura* group species. Nevertheless, no resemblance at the restriction map level is detected when more distantly related species are compared. In *D. ambigua*, *D. immigrans* and *D. lebanonensis* in situ hybridization assays reveal a single chromosomal location for *Adh*, which in *D. lebanonensis* appears to be sex linked. In contrast, in *D. subobscura*, *D. madeirensis* and *D. guanche* multiple sites of hybridization with homologous and heterologous probes are observed. For example, in *D. subobscura* and *D. madeirensis* the functional *Adh* gene is located on the *U* chromosome and additional homologous retrosequences are found on the *E* chromosome.

been decisive (Ruiz and Fontdevila 1982). Such comparisons are much more informative when species are closely related and form hybrids, as the homology between chromosome elements is apparent as a result of chromosome pairing in polytene salivary gland cells. When species are distantly related, however, comparison of chromosome elements is difficult and indirect methods have to be used.

The original haploid karyotype of the genus *Drosophila* consists of five large chromosomes (elements *A–E*) and a dot chromosome (*F*), which have remained quite intact through the evolution of the genus. These earlier observations (Muller 1940; Sturtevant and Novitski 1941) have been confirmed by several comparative studies (Loukas et al. 1979; Pinsker and Sperlich 1984; Krimbas and Loukas 1984; Steinemann et al. 1984). Although the chromosomal elements have retained the essential identity, some structural rearrangements have been described. Paracentric inversions and fusions, which are obvious mechanisms for altering gene order within each chromosomal element, have been reported to occur relatively frequently in *Drosophila*, but translocations and pericentric inversions are thought to be rare (Clayton and Guest 1986). However, there is evidence for the establishment of both translocations and pericentric inversions in natural populations. The location of some pteridine loci appears to reflect the occurrence of a pericentric inversion (Gregg and Smucker 1965). In addition, the location of the 5S RNA gene cluster is different in *D. hydei* and *D. melanogaster* (Alonso and Berendes 1975) and the 18/28S RNA genes of some Hawaiian *Drosophila* species are autosomal rather than *X* linked (Stuart et al. 1981).

In situ hybridization with DNA probes is a powerful method for cytological mapping of specific genes and tracing karyotypic changes during evolution (Steinemann 1982; Steinemann et al. 1984; Loukas and Kafatos 1986; Whiting et al. 1989). The aim of the present study is to determine the chromosomal location and structural organization of the genomic region containing the alcohol dehydrogenase gene (*Adh*) in several *Drosophila* spe-

Introduction

The analysis of related genes and proteins from different species has proved to be of crucial relevance in evaluating evolutionary changes. Phylogenies of some *Drosophila* species have been proposed through the analysis of DNA structure of homologous genes, DNA cross hybridization assays and, at the protein level, following allozyme polymorphisms or analysis of immunological features. The polytene chromosomes of *Drosophila* and some other *Diptera* also allow comparative evolutionary analysis and, in cases where external morphologies are similar, the cytogenetic relationship between species has

Offprint requests to: R. González-Duarte

cies, in order to evaluate the evolutionary changes undergone by this structural region in a group of differently related species. Four species, *D. subobscura*, *D. madeirensis*, *D. guanche* and *D. ambigua*, belong to the *obscura* group (sg. *Sophophora*). The other two belong to different subgenera: *D. immigrans* (sg. *Drosophila*) and *D. lebanonensis* (sg. *Pholadoris*).

In *D. melanogaster* as well as in other *Drosophila* species the *Adh* gene has three exons interrupted by two introns and its expression is under the control of two different age-specific promoters (Goldberg 1980; Benyajati et al. 1983). Different organization of this genomic region has been described in species of the *repleta* group (Fischer and Maniatis 1985). In addition, at the 3' flanking region of *Adh* an open reading frame (3'ORF) showing a considerable degree of homology with *Adh* has been reported (Schaeffer and Aquadro 1987).

We determined the restriction maps of the region containing the *Adh* locus for each of the aforementioned species and recombinant clones carrying part of this region were used as homologous or heterologous probes for in situ hybridization assays.

Materials and methods

Fly stocks. Flies from natural populations caught in Spain were maintained in our laboratory for several years. The following species were used in our study: *D. ambigua* (Vidrà, Barcelona, 1985), *D. guanche* (Canary Islands, 1976), *D. madeirensis* (Madeira Islands, 1984), *D. subobscura* (strain H27, with standard chromosomal arrangements, provided by R. de Frutos, Dept. of Genetics, University of Valencia, Spain), *D. immigrans* (Bordils, Girona, 1979) and *D. lebanonensis* (Gandesa, Tarragona). A second strain of *D. lebanonensis* provided by the Umeå Stock Center (Sweden) was used for chromosome preparations. An isogenic strain of *D. melanogaster* (*Adh*^F, provided by Brian Clark, Dept. of Genetics, University of Nottingham, UK) was used as a control.

Preparation of genomic DNA and construction of libraries. Total genomic DNA from each species was isolated using the guanidine isothiocyanate method initially described for RNA extraction (Chirgwin et al. 1979) with minor modifications. High molecular weight total DNA was partially digested with MboI, then fractionated in a sucrose gradient to obtain DNA fragments of 15–20 kb. The libraries were constructed by ligating this prepared DNA into the BamHI site of an EMBL 4 phage vector (Frischauf et al. 1983) for *D. subobscura*, *D. madeirensis*, *D. guanche* and *D. ambigua*, or into λ -Charon 35 (Wilhelmine and Blattner 1983) for *D. immigrans* and *D. lebanonensis*. Between 150000 and 200000 recombinant plaques from each species were screened for sequences homologous to the *D. melanogaster Adh* gene. A 2.7 kb HindIII-EcoRI restriction fragment of sAC1 (Goldberg 1980) that contains the complete *Adh* gene and the 3' adjacent region was used as a probe. It was labelled either by nick-translation (Rigby et al. 1977) or by random-hexamer priming (Feinberg and Vogelstein 1983) with [α -³²P]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% sodium dextran sulphate. Nonspecifically bound probe was removed with one wash in 2 × SSC, 0.1% SDS at room temperature for 10 min, then two washes in 2 × SSC, 0.1% SDS at 65° C for 10 min each, and two further washes in 1 × SSC, 0.1% SDS at 65° C for 10 min each (1 × SSC is 0.15 M NaCl, 0.015 M sodium citrate, pH 7.5). After the wash steps, the filters were autoradiographed and DNA from positive recombinant phages was isolated according to Maniatis et al. (1982).

Restriction analysis of positive clones. Fragments from positive clones of each species were subcloned into pUC18 or Bluescript vectors (Table 2) and characterized by mapping with the following restriction enzymes: BamHI, EcoRI, HindIII, PvuII, PstI, Sall and XbaI (Boehringer Mannheim). Single and double digestions were performed (Maniatis et al. 1982). *Adh* was located on the restriction map by transferring digested and electrophoresed DNA to nitrocellulose membranes (Hybond-C, Amersham) using the transfer method of Southern. Labelling of the *Adh* probe, hybridization, washing of the filters and autoradiography, were the same as for plaque hybridization. Sequencing was performed by the dideoxy method of Sanger et al. (1977) and according to conventional techniques (Maniatis et al. 1982).

In situ hybridization. Salivary glands dissected from third instar larvae were fixed and squashed in 45% acetic acid. Coverslips were removed after freezing the preparations in liquid nitrogen and slides were immediately dehydrated in ethanol and air dried. Subsequently, they were incubated in 2 × SSC at 65° C for 30 min, dehydrated in ethanol and air dried. Just before hybridization the slides were immersed in 0.07 N NaOH for 2 or 3 min to denature chromosomal DNA, washed in 2 × SSC (3 × 5 min), dehydrated and air dried. DNA probes were labelled by nick-translation (Rigby et al. 1977) using Bio-11-dUTP or Bio-14-dATP (Enzo and BRL). For ten slides, 1 µg of biotinylated probe was dissolved in 200 µl solution containing 50% formamide, 10% dextran sulphate, 0.4 µg salmon sperm DNA in 2 × SSC. This hybridization mixture was incubated at 90° C for 10 min and applied to the squashes. The slides were then coverslipped, sealed and incubated in a moist chamber at 80° C for 10 min. After hybridization at 37° C for 16–20 h the slides were washed in 2 × SSC (5 × 10 min). The first wash was at room temperature, the following two at 42° C and the last two at room temperature. Then the slides were incubated for 5 min in PBS containing 0.1% Triton X-100 and washed in PBS (2 × 5 min). Detection of biotin was carried out by incubating the slides with a streptavidin-peroxidase complex (Detek 1-hrp Signal Generating System, Enzo) for 90 min. The slides were then washed in 2 × SSC (2 × 5 min), then in PBS containing 0.1% Triton X-100 (5 min) and finally in PBS (5 min). Peroxidase activity was detected using diaminobenzidine as a chromogen. The preparations were stained with Giemsa, air dried and mounted in DPX.

Chromosome analysis. Chromosomes were examined and photographed in Zeiss photomicroscope. Chromosome nomenclature and identification of bands were based on the following salivary gland chromosome maps: *D. immigrans*, Le Calvez (1953); *D. lebanonensis*, Berendes and Thijssen (1971); *D. subobscura*, Kunze-Mühl and Müller (1958); *D. ambigua*, Mainx et al. (1953); *D. guanche*, Moltó et al. (1987); *D. madeirensis*, Papaceit (1988).

Results

In situ hybridization

Preliminary in situ hybridizations were carried out using the 2.7 kb restriction fragment sAC_{H/E} from plasmid sAC1 as a probe. This fragment, containing the *D. melanogaster Adh* gene, was also used to screen genomic libraries of the six species under study. DNA fragments from positive clones were subcloned in different plasmids and used as homologous probes for in situ hybridization assays (Fig. 1). Sequence analysis revealed that some of the probes contained the functional *Adh* gene whereas others had homologous non-functional sequences (Table 1 and Fig. 2). Results obtained for each species with homologous as well as heterologous probes are summarized in Table 2.

Table 1. Comparison of positions and sequences of the *Adh* genomic region of the *Drosophila* species under study

Region	Parameter	Species					
		<i>D. lebanonensis</i>	<i>D. immigrans</i>	<i>D. ambigua</i>	<i>D. subobscura</i>	<i>D. guanche</i>	<i>D. madeirensis</i>
<i>Adh</i>							
Adult TATA box	Sequence Position	TATTTAA -1152	PS	TATTTAA -1026	TATTTAA -836	TATTTAA -831	TATTTAA PS
Larval TATA box	Sequence Position	CATAAATA -24	TATAAATA -25	TATAAATA -34	TATAAATA -32	TATAAATA -32	TATAAATA -32
First exon	Nt length Position	93 +45-+137	93 +59-+151	93 +67-+159	93 +65-+157	93 +65-+157	93 +64-+156
Second exon	Nt length Position	405 +199-+603	405 +215-+619	405 +226-+630	405 +226-+630	405 +227-+631	405 +224-+628
Third exon	Nt length Position	267 +670-+936	267 +684-+950	267 +697-+963	267 +699-+965	267 +695-+961	267 +696-+962
Polyadenylation signal	Position	+1064	PS	+1133	+1106	+1108	+1103
3'ORF							
CCAAT box	Sequence Position	PS	PS	GGCAATAAGG +1074	GGCAATAAGG +1053	GGCAATAAGG +1056	GGCAATAAGG +1050
TATA box	Sequence Position	PS	PS	TTAATTAATA +1207	TTAATTAATA +1162	TTAATTAATA +1170	TTAATTAATA +1159
First exon	Nt length Position	PS	PS	96 +1321-+1416	96 +1289-+1384	96 +1303-+1398	PS
Second exon	Nt length Position	PS	PS	405 +1709-+2113	405 +1670-+2074	405 +1522-+1926	PS
Third exon	Nt length Position	PS	PS	345 +2181-+2525	339 +2137-+2475	PS	PS
Polyadenylation signal	Position	PS	PS	+2536	+2498	PS	PS

PS, partially sequenced; Nt, nucleotide

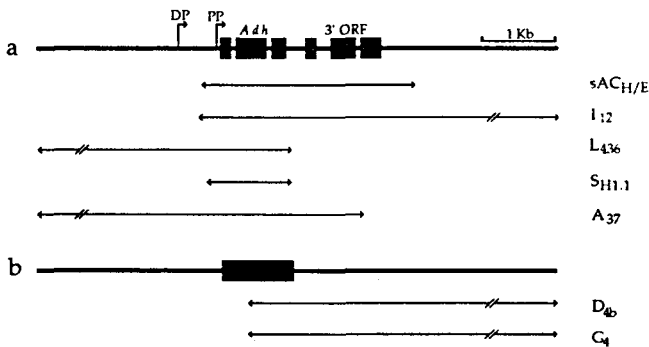


Fig. 1 a, b. Probes used for in situ hybridization. Subcloned genomic fragments are represented under the general diagram. **a** The following probes contain the functional *Adh* gene and contiguous sequences: *Drosophila melanogaster* (sAC_{H/E}), *D. immigrans* (I₁₂), *D. lebanonensis* (L₄₃₆), *D. subobscura* (S_{H1.1}) and *D. ambigua* (A₃₇). Solid boxes represent the coding regions of the *Adh* gene, and hatched boxes represent the putative coding regions of the 3' ORF. DP and PP indicate the position of the distal promoter and proximal promoter respectively. **b** The probes from *D. madeirensis* (D_{4b}) and *D. guanche* (G₄) contain an *Adh* pseudogene that is represented as a solid box

The sAC_{H/E} probe failed to give detectable hybrids in *D. guanche*, *D. immigrans* and *D. lebanonensis* but yielded positive results with *D. ambigua*, *D. subobscura* and *D. madeirensis*. In these three species hybridization was observed in the *B* chromosome element, homologous to the 2*L* chromosome arm of *D. melanogaster* (Table 3). In addition, a second hybridization site was found in division 64 of the *E* chromosome of *D. subobscura*. The two *Adh* hybridization sites found in this species with probe sAC_{H/E} were in agreement with those found using the homologous probe S_{H1.1}. However, in this latter case, the hybridization signal observed in the *U* chromosome was very faint, whereas the band in the *E* chromosome (Fig. 3c) was much more intense.

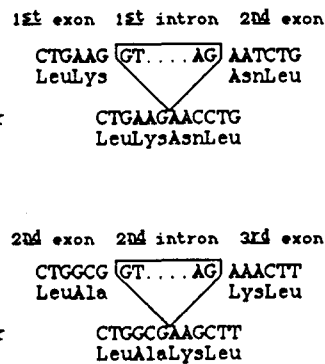


Fig. 2. Comparison of genomic sequences at some splicing junctions between the functional *Adh* gene of *Drosophila madeirensis* and *D. guanche* and their respective retrosequences. MAD, functional *Adh* gene of *D. madeirensis*. MAD-r, *Adh* retrosequence of *D. madeirensis*. GUA, functional *Adh* gene of *D. guanche*. GUA-r, *Adh*, retrosequence of *D. guanche*

The 9 kb genomic insert of *D. madeirensis* in plasmid D_{4b} did not contain the functional *Adh* gene. DNA sequence analysis of this genomic clone revealed the features of a retrosequence (Fig. 2). When this probe was used to localize its homologous region in *D. madeirensis* chromosomes it hybridized to division 64 in chromosome *E* (Fig. 3f). In contrast, probe sAC_{H/E} was found to hybridize at division 36 near the centromeric region of the *U* chromosome (Fig. 3c).

Three hybridization sites on *D. guanche* were observed when probe G₄ was used (Fig. 3d). Two of them were at division 98 of chromosome *O* and the third at division 62 of chromosome *E*.

D. ambigua showed only one *Adh* homologous region, at division 36 of chromosome *E*, when probes A₃₇ and sAC_{H/E} were used (Fig. 3a). A genomic clone containing this region was sequenced and agrees with the sequence expected for the functional gene (Table 1).

Table 2. Recombinant *Adh* probes and hybridization sites in salivary gland chromosomes of several *Drosophila* species

Probe			Hybridization site		
Species	Plasmid	Insert size (kb)	Species	Chromosome	Site
<i>D. melanogaster</i>	sAC _{H/E} ^a	2.7	<i>D. ambigua</i>	<i>E_R</i>	36
			<i>D. madeirensis</i>	<i>U</i>	36
			<i>D. subobscura</i>	<i>U</i>	36
			<i>E</i>	64	
<i>D. subobscura</i>	S _{H1.1} [Bluescript] ^b	1.1	<i>D. subobscura</i>	<i>U</i>	36
				<i>E</i>	64
<i>D. madeirensis</i>	D _{4b} [pUC ₁₈]	9.0	<i>D. madeirensis</i>	<i>E</i>	64
<i>D. guanche</i>	G ₄ [pUC ₁₈]	9.0	<i>D. guanche</i>	<i>O</i>	98 ^c
				<i>E</i>	62
<i>D. ambigua</i>	A ₃₇ [pUC ₁₈]	4.5	<i>D. ambigua</i>	<i>E_R</i>	36
			<i>D. subobscura</i>	<i>E</i>	64
				<i>U</i>	36
<i>D. immigrans</i>	I ₁₂ [pUC ₁₈]	5.0	<i>D. immigrans</i>	4	G ₄
<i>D. lebanonensis</i>	L ₄₃₆ [Bluescript]	3.6	<i>D. lebanonensis</i>	<i>X</i>	1

Plasmid vectors containing the inserts are indicated in brackets

^a sAC_{H/E} is a HindIII-EcoRI restriction fragment from plasmid sAC1

^b Plasmid S_{H1.1} contains two copies of the 1.1 kb insert

^c Two hybridization sites in the same division

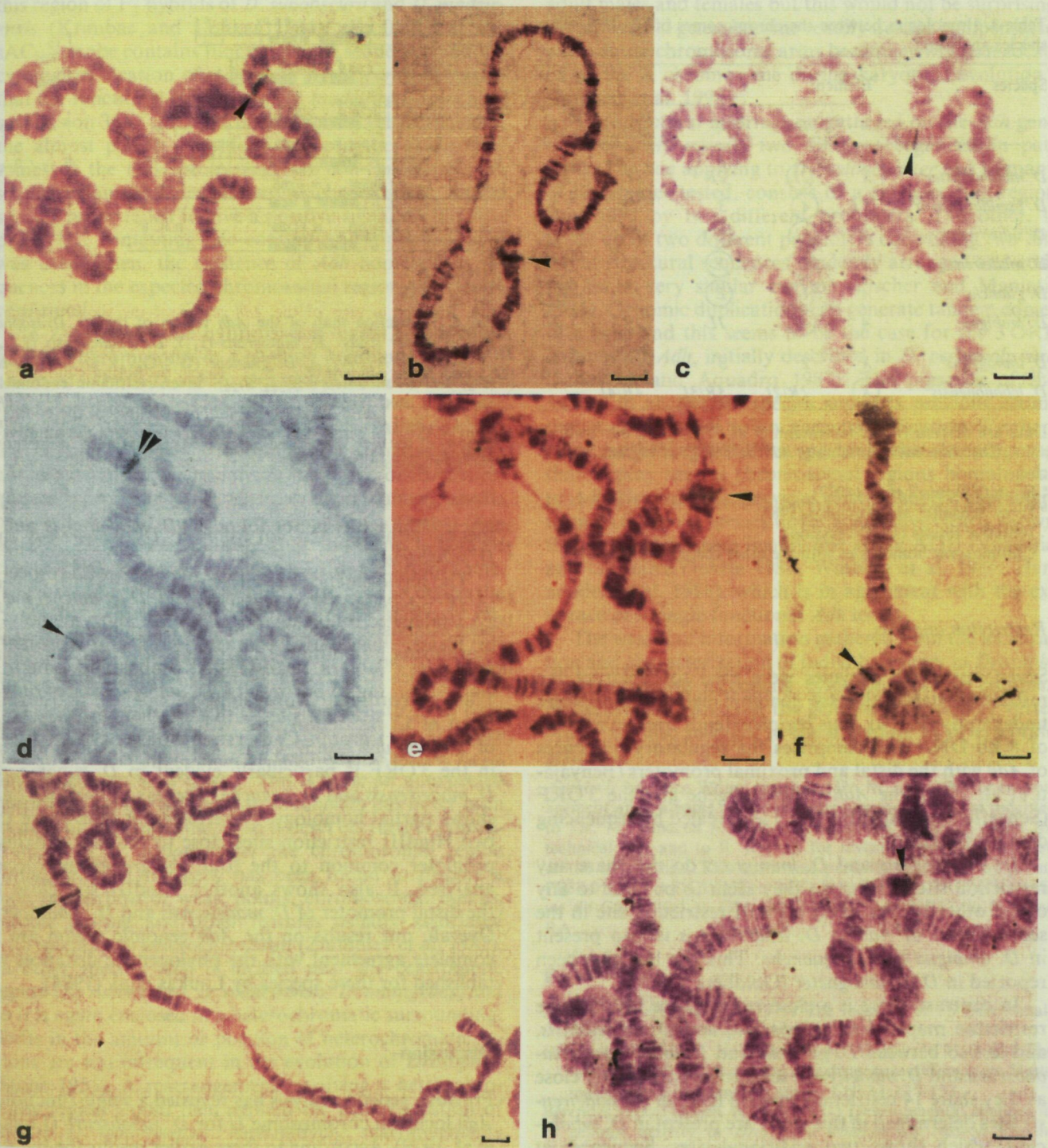


Fig. 3 a-h. In situ hybridizations to *Drosophila* salivary gland polytene chromosomes. **a** Hybridization of probe A_{37} to a single site on the E_R chromosome arm of *D. ambigua*. **b** Probe $sACH/E$ hybridizes to the U chromosome of *D. subobscura*. **c** Additional hybridization site on the E chromosome of *D. subobscura* with probe $S_{H1.1}$. **d** Probe G_4 hybridizes to three chromosomal sites in *D.*

guanche. **e** Hybridization of probe $S_{H1.1}$ to *D. madeirensis* chromosome U . **f** Probe D_{4b} hybridizes to the E chromosome of *D. madeirensis*. **g** Chromosome four of *D. immigrans* showing the single hybridization site obtained with the homologous probe I_{12} . **h** Probe L_{436} hybridizes to the X chromosome of *D. lebanonensis*. Arrowheads indicate hybridization sites. Bars represent 10 μ m

Chromosome homologies between *D. immigrans*, *D. lebanonensis* and the other species have not yet been determined. Homologous genomic probes, I_{12} and L_{436} respectively, were used to localize the *Adh* gene in these

two species. All the results obtained were consistent and showed that the *Adh* gene is at the G_4 site of *D. immigrans* chromosome 4 (Fig. 3g) and at division 1 of the X chromosome of *D. lebanonensis* (Fig. 3h).

Table 3. Homologies between the chromosomes of the *Drosophila* species analysed

Species	Element				
	A	B	C	D	E
<i>D. melanogaster</i> ^a	X	2L	2R	3L	3R
<i>D. pseudoobscura</i> ^b	XL	4	3	XR	2
<i>D. subobscura</i>	A	U	E	J	O
<i>D. madeirensis</i>	A	U	E	J	O
<i>D. guanche</i>	A	U ^c	E	J	O
<i>D. ambigua</i>	A	E	JL	JR	U
<i>D. immigrans</i> ^d	X	2	3	4	5
<i>D. lebanonensis</i> ^d	X	2R/2L	3R/3L	4R/4L	

Data are derived from Steinemann et al. (1984) and our own results. Chromosomes containing *Adh* homologous regions are circled

^a From O'Donnell et al. (1977)

^b From Schaeffer and Aquadro (1987)

^c Tentative results (see Discussion)

^d Chromosome homologies with the other species are unknown

Restriction analysis

Seven restriction enzymes were used to construct the restriction map of the *Adh* genomic region in the species under study (Fig. 4). This region extends over an average of 8 kb DNA and comprises the regulatory sequences of *Adh* with the distal and proximal promoters (Benyajati et al. 1983), the *Adh* structural gene and the 3'ORF (Schaeffer and Aquadro 1987) as revealed by sequencing data (Table 1).

D. lebanonensis and *D. immigrans* do not share any restriction site, neither do they seem to be close to any of the other species. Only a PvuII restriction site in the second exon of *Adh* in *D. lebanonensis* is also present in *D. ambigua* and *D. guanche*. This site has also been reported in *D. melanogaster* (Goldberg 1980).

In contrast, there is a close resemblance between the restriction maps of *D. subobscura*, and *D. madeirensis*, and less so between *D. guanche* and *D. ambigua*. *D. subobscura* and *D. madeirensis* seem to be particularly close as they share 14 restriction sites. In fact, all *D. madeirensis* sites determined in this study are present in *D. subobscura*, which shows two additional sites. Sequences at the flanking regions are also conserved, as three out of four distant sites are present in both species. Close proximity between these two species at the phenotypic and cytogenetic levels has been reported (Krimbas and Loukas 1984). Our results support this resemblance at the DNA level.

D. guanche shares with *D. madeirensis* and *D. subobscura* the four 3'ORF sites (PstI, XbaI, BamHI, BamHI), the two HindIII sites close to the proximal *Adh* promoter and the Sall site located between the distal and proximal promoters. This species also contains the PvuII site in the second *Adh* exon which is present in other *Drosophila*

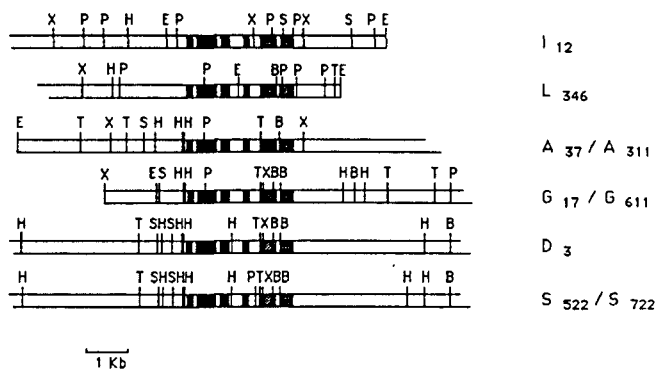


Fig. 4. Restriction map of the *Adh* region of the six *Drosophila* species under study: *D. immigrans* (I₁₂), *D. lebanonensis* (L₃₄₆), *D. ambigua* (A₃₆/A₃₁₁), *D. guanche* (G₁₇/G₆₁₁), *D. madeirensis* (D₃) and *D. subobscura* (S₅₂₂/S₇₂₂). Seven restriction endonucleases were used: BamHI (B), HindIII (H), EcoRI (E), PvuII (P), PstI (T), Sall (S), and XbaI (X). Solid boxes represent the coding regions of the *Adh* genes, and hatched boxes represent the putative coding regions of the 3'ORFs

species, although it is not found in *D. madeirensis* and *D. subobscura*.

The phylogenetic position of *D. ambigua* with respect to *D. subobscura*, *D. madeirensis* and *D. guanche* has been established following the analysis of allozyme polymorphisms, quantitative characters and chromosomal homologies (Cabrera et al. 1983; Krimbas and Loukas 1984). According to our results, *D. ambigua* shares only 5 out of 12 restriction sites with the other three species of the *obscura* group. Conservation of restriction sites in the 3'ORF is particularly evident in *D. subobscura*, *D. madeirensis* and *D. guanche*, whereas *D. ambigua* only shows partial homology. Nevertheless it does share the two HindIII restriction sites close to the *Adh* proximal promoter common to the four *obscura* group species analysed. It also shows another HindIII site close to the distal promoter of *D. madeirensis* and *D. subobscura*. Overall, our results on the *Adh* genomic region are in complete agreement with the phylogenetic distances established for these species by Loukas et al. (1984).

Discussion

Previous genetic analysis has revealed a single functional *Adh* locus on chromosome U for *D. subobscura* (Loukas et al. 1979). According to these earlier results the hybridization site on chromosome U would correspond to the functional *Adh* gene and consequently, the homologous *Adh* region on chromosome E might be due to non-functional *Adh* sequences (*Adh* pseudogenes). Of all the species analysed, *D. madeirensis* and *D. subobscura* are the most closely related. Both previous analyses and the present data strongly support this relationship. Accordingly, most restriction sites in the *Adh* gene and flanking regions are conserved. When a heterologous probe (sAC_{H/E}) was hybridized to the *D. madeirensis* chromosomes a positive band was observed at division 36 of the U chromosome. The strong homology described for

this region of F₁ hybrids of *D. subobscura* and *D. madeirensis* (Krimbas and Loukas 1984), the fact that the sAC_{H/E} probe contains functional *Adh* sequences and the cytological location of *Adh* in *D. subobscura* all suggest that the functional *Adh* gene of *D. madeirensis* is located on division 36 of the *U* chromosome. On the other hand, the almost perfect homology between the *E* chromosomes of the two species suggests non-functional *Adh* sequences on the *D. madeirensis* *E* chromosome. As the sAC_{H/E} probe failed to give a positive signal on this site, probe D_{4b} containing one copy of an *Adh* pseudogene was used. Then, the existence of *Adh* homologous sequences in the expected chromosomal region was clearly confirmed.

In *D. guanche* one hybridization signal was located on the *E* chromosome in a position homologous to that observed in both *D. subobscura* and *D. madeirensis*. Also two more very close sites appeared near the telomeric end of the *O* chromosome. In this case we cannot discard the possibility that the hybridization signal is due to homology with other regions of the probe or to retrosequence-type elements. Furthermore, preliminary results obtained with a probe containing the functional *Adh* gene of *D. guanche* indicate that this gene is located close to the centromeric region of the *U* chromosome. It is most unfortunate that many attempts to define the *U* region involved in hybridization were seriously impaired by the permanently distorted morphology of this chromosomal region, which moreover is close to a break point (Moltó et al. 1987).

The chromosome maps of *D. lebanonensis* have been described by Berendes and Thijssen (1971). The *Adh* gene has been mapped with an homologous probe, L₄₃₆, and appears to be *X* linked. This is quite an unexpected result because the *Adh* gene in *Drosophila* has never been located on chromosome *X*, so it deserves further consideration. In *D. melanogaster* the chromosomal region containing the *Adh* gene has been found to display a high proportion of weak points, chromosomal breaks, ectopic pairing and other features usually associated with intercalary heterochromatin (Zhimulev et al. 1982). In other species, *Adh* is located near the centromeric region (*D. subobscura*, *D. madeirensis*, *D. immigrans*) and so it is again enclosed in a heterochromatic surrounding. Shifts in the amount or position of heterochromatic regions are not infrequent in the evolution of *Drosophila* (Stone 1955). A rearrangement involving a heterochromatic region would then explain the actual location of *Adh* in *D. lebanonensis*. Alternatively, rearrangements of larger chromosomal portions could be involved. In *Drosophila* several examples of centric fusions between the *X* chromosome and an autosome have been illustrated (Stone 1955). For example, in *D. americana* an *X*-autosome fusion involving the *B* element has been reported (Lucchesi 1978) and such a rearrangement might have occurred in *D. lebanonensis* as well. However, taking into consideration the fact that the *X* chromosome of *D. lebanonensis* is a telomeric element, an additional pericentric inversion would have to be assumed to account for its actual configuration. Preliminary results do not support any difference in ADH activity between

adult males and females but this would not be surprising if "autosomal genes become readily dosage compensated when entire chromosome arms become newly translocated to the *X* chromosome during karyotype evolution" (Birchler et al. 1990).

The structural information gathered on the *Adh* gene in *Drosophila* reveals two different organization patterns. The one applying to *D. melanogaster* and to many of the species tested, consists of a single coding region controlled by two different promoters. The other, is made up of two different promoters controlling two different structural sequences tandemly arranged and coding for a very similar enzyme (Fischer and Maniatis 1985). Genomic duplications can generate tandem copies of a gene and this seems to be the case for the 3'ORF adjacent to *Adh*, initially described in *D. pseudoobscura* (Schaeffer and Aquadro 1987). The presence of the 3'ORF in all our six species supports previous claims that it corresponds to an ancient duplication. Furthermore, we here report three species of the *obscura* group that display several chromosomal regions homologous to *Adh*. Nevertheless, previous electrophoretic characterizations of the six species under study have clearly shown that in all cases there is only one active ADH (Vilageliu and González-Duarte 1984; Winberg et al. 1986; Hernández et al. 1988), which is in agreement with the existence of a single functional *Adh* locus.

The wealth of information gathered from the localization and characterization of individual genes in *Drosophila* is providing increasing evidence that the complex mechanisms involved in the genome evolution of vertebrates are also relevant to the dynamics of invertebrate genomes.

Acknowledgements. We are obliged to Dr. M. Izquierdo for encouragement and valuable advice regarding methods. We are indebted to Dr. M. Papacit for helpful discussions, to M.M. Duro for technical help and to R. Rycroft for revising the English version of the manuscript. This work was partially supported by grants PB88-0188 from DGICYT to RGD, AR-87 from CIRIT (Generalitat de Catalunya) to GM and AR-89 from CIRIT to NV and RA.

References

- Alonso C, Berendes HD (1975) The location of 5S (ribosomal) RNA genes in *Drosophila hydei*. *Chromosoma* 51:347-356
- Benyajati C, Spoerel N, Haymerle H, Ashburner M (1983) The messenger RNA for alcohol dehydrogenase differs in its 5' end in different developmental stages. *Cell* 33:125-133
- Berendes HD, Thijssen WThM (1971) Developmental changes in genome activity in *Drosophila lebanonensis casteeli* Pipkin. *Chromosoma* 33:345-360
- Birchler JA, Hiebert JC, Paigen K (1990) Analysis of autosomal dosage compensation involving the alcohol dehydrogenase locus in *Drosophila melanogaster*. *Genetics* 124:677-686
- Cabrera VM, González AM, Larruga JM, Gullon A (1983) Genetic distance and evolutionary relationships in the *Drosophila obscura* group. *Evolution* 37:675-689
- Chirgwin JM, Przybyla AE, MacDonald RJ, Rutter WJ (1979) Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* 18:5294-5299
- Clayton FE, Guest WC (1986) Overview of chromosomal evolution in the family Drosophilidae. In: Ashburner M, Carson HL, Thompson JN (eds) *The genetics and biology of Drosophila*. 3E. Academic Press, New York, pp 1-58

- Feinberg AP, Vogelstein B (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal Biochem* 132:6-13
- Fischer JA, Maniatis T (1985) Structure and transcription of the *Drosophila mulleri* alcohol dehydrogenase genes. *Nucleic Acids Res* 13:6899-6917
- Frischauf A-M, Lehrach H, Poustka A, Murray N (1983) Lambda replacement vectors carrying polylinker sequences. *J Mol Biol* 170:827-842
- Gregg TG, Smucker LA (1965) Pteridines and gene homologies in eye colour mutants of *Drosophila hydei* and *D. melanogaster*. *Genetics* 52:1023-1034
- Goldberg DA (1980) Isolation and partial characterization of the *Drosophila* alcohol dehydrogenase gene. *Proc Natl Acad Sci USA* 77:5794-5798
- Hernández JJ, Vilageliu LI, González-Duarte R (1988) Functional and biochemical features of alcohol dehydrogenase in four species of the *obscura* group of *Drosophila*. *Genetica* 77:15-24
- Krimbas CB, Loukas M (1984) Evolution of the *obscura* group *Drosophila* species. I. Salivary gland chromosomes and quantitative characters in *D. subobscura* and two closely related species. *Heredity* 53:469-482
- Kunze-Mühl E, Müller E (1958) Weitere Untersuchungen über die chromosomale Struktur und die natürlichen Strukturtypen von *Drosophila subobscura* Coll. *Chromosoma* 9:559-570
- Le Calvez J (1953) Carte cytologique des chromosomes géants des glandes salivaires de *Drosophila immigrans* Sturt. *Chromosoma* 6:170-174
- Loukas M, Kafatos FC (1986) The actin loci in the genus *Drosophila*: establishment of chromosomal homologies among distantly related species by in situ hybridization. *Chromosoma* 94:297-308
- Loukas M, Krimbas CB, Mavragani-Tsipidou P, Kastritsis CD (1979) Genetics of *Drosophila subobscura* populations. VIII. Allozyme loci and their chromosome maps. *J Hered* 70:17-26
- Loukas M, Krimbas CB, Vergini Y (1984) Evolution of the *obscura* group *Drosophila* species. II. Phylogeny of ten species based on electrophoretic data. *Heredity* 53:483-493
- Lucchesi JC (1978) Gene dosage compensation and the evolution of sex chromosomes. *Science* 202:711-716
- Mainx F, Koske Th, Smital E (1953) Untersuchungen über die chromosomale Struktur Europäischer Vertreter der *Drosophila obscura* Gruppe. *Z Indukt Abstammungs Vererbungslehre* 85:354-372
- Maniatis T, Fritsch EF, Sambrook J (1982) *Molecular cloning: A laboratory manual*, 1st edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
- Moltó MD, De Frutos R, Martínez-Sebastián MJ (1987) The banding pattern of polytene chromosomes of *Drosophila guanche* compared with that of *D. subobscura*. *Genetica* 75:55-70
- Muller HJ (1940) Bearings of the *Drosophila* work on systematics. In: Huxley J (ed) *The systematics*. Clarendon Press, Oxford, pp 185-268
- O'Donnell J, Mandel HC, Krauss M, Sofer W (1977) Genetic and cytogenetic analysis of the ADH region in *Drosophila melanogaster*. *Genetics* 86:553-566
- Papacit M (1988) Estudio de las diferencias genéticas entre *D. subobscura* Collins y *D. madeirensis* Monclús. PhD Thesis, University of Barcelona
- Pinsker W, Sperlich D (1984) Cytogenetic mapping of enzyme loci on chromosomes J and U of *Drosophila subobscura*. *Genetics* 108:913-926
- Rigby PWJ, Dieckmann M, Rhodes C, Berg P (1977) Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. *J Mol Biol* 113:237-251
- Ruiz A, Fondevila A (1982) The evolutionary history of *Drosophila buzzatii*. III. Cytogenetic relationships between two sibling species of the *buzzatii* cluster. *Genetics* 101:503-518
- Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 74:5463-5467
- Schaeffer SW, Aquadro CF (1987) Nucleotide sequence of the Adh gene region of *Drosophila pseudoobscura*: evolutionary changes and evidence for an ancient gene duplication. *Genetics* 117:61-73
- Steinemann M (1982) Analysis of chromosomal homologies between two species of the subgenus *Sophophora*: *D. miranda* and *D. melanogaster* using cloned DNA segments. *Chromosoma* 87:77-88
- Steinemann M, Pinsker W, Sperlich D (1984) Chromosome homologies within the *Drosophila obscura* group probed by in situ hybridization. *Chromosoma* 91:46-53
- Stone WS (1955) Genetic and chromosomal variability in *Drosophila*. *Cold Spring Harbor Symp Quant Biol* 20:256-269
- Stuart WD, Bishop JC, Carson HL, Frank MB (1981) Location of the 18/28S ribosomal RNA genes in two Hawaiian *Drosophila* species by monoclonal immunological identification of the RNA-DNA hybrids in situ. *Proc Natl Acad Sci USA* 78:3751-3754
- Sturtevant AH, Novitski E (1941) The homologies of the chromosome elements in the genus *Drosophila*. *Genetics* 26:517-541
- Vilageliu LI, González-Duarte R (1984) Alcohol dehydrogenase from *Drosophila funebris* and *Drosophila immigrans*: molecular and evolutionary aspects. *Biochem Genet* 22:797-815
- Whiting JH Jr, Pilely MD, Farmer JL, Jeffery DE (1989) In situ hybridization analysis of chromosomal homologies in *Drosophila melanogaster* and *Drosophila virilis*. *Genetics* 122:99-109
- Wilhelmine AML, Blattner FR (1983) Lambda Charon vectors (Ch 32, 33, 34 and 35) adapted for DNA cloning in recombination-deficient hosts. *Gene* 26:171-179
- Winberg J-O, Hovik R, McKinley-McKee JS, Juan E, González-Duarte R (1986) Biochemical properties of alcohol dehydrogenase from *Drosophila lebanonensis*. *Biochem J* 235:481-490
- Zhimulev IF, Semeshin VF, Kulichkov VA, Belyaeva ES (1982) Intercalary heterochromatin in *Drosophila*. I. Localization and general characteristics. *Chromosoma* 87:197-227

Note added in proof. After this manuscript was completed, the location of *D. guanche Adh* on the centromeric region of the U chromosome was confirmed.

**CAPÍTOL II: "The *Adh* genomic region of *Drosophila*
ambigua: evolutionary trends in different species"**

Publicat a Journal of Molecular Evolution 32: 454-462 (1991)

The *Adh* Genomic Region of *Drosophila ambigua*: Evolutionary Trends in Different Species

Gemma Marfany and Roser González-Duarte

Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, Avda. Diagonal, 645 (08028) Barcelona, Spain

Summary. The study of individual genes is essential to a comprehensive understanding of genome evolution. The wealth of information on alcohol dehydrogenase (*Adh*) in *Drosophila* makes this gene particularly suitable for such analysis. We have characterized more than 4 kb of the genomic *Adh* region in *Drosophila ambigua* and compared this region to *Drosophila mauritiana* and *Drosophila pseudoobscura*. The presence of two genes, *Adh* and 3'ORF (open reading frame), has been confirmed and some of their essential features have been inferred from primary structural analysis. Inter- and intraspecific comparisons have led us to support that both genes may have diverged from an ancient precursor. They appear to be evolving independently, and show a species-specific pattern. The *Adh* in the *obscura* group species lacks amino acids three and four when compared to the species of the *melanogaster* group and has accumulated most of its amino acid replacements in the third exon. Neither characteristic is observed when any other group species are compared, which suggests that these may be particular features of the evolution of the *obscura* group. The 3'ORF is highly conserved among the three species analyzed, although variability in the length of the third exon and the nucleotide substitution rate, which is much higher than in *Adh*, are worth noting. According to our data, both mutation/fixation rates and the distribution of mutations vary over time, which makes it difficult to predict the evolutionary dynamics of specific genome regions.

Key words: Alcohol dehydrogenase gene — *Drosophila* — Gene structure — Evolutionary trends — Nucleotide substitution rate

Introduction

Genetic changes that underlie evolutionary forces have been analyzed at different levels. DNA sequence comparisons have been crucial to defining inter- and intraspecific variability. The information gathered at this level, although relevant, bears no direct relation to the phenotype. Instead, the same type of analysis focused on individual genes can produce the necessary information to overcome this limitation. The alcohol dehydrogenase gene-enzyme system (*Adh*-ADH) has been studied thoroughly in the genus *Drosophila* and has proved to be an extremely useful model to illustrate the evolutionary forces that operate on structural genes [for recent reviews see Sofer and Martin (1987) and Chambers (1988)].

The genus *Drosophila* comprises many different phylogenetic groups, but only some of them have been analyzed in detail. For instance, there is a wealth of information concerning the *melanogaster* group, but distant species should be analyzed more extensively in order to answer relevant questions concerning the potential variability of gene structure, the formation process of complex regulatory patterns, and their effects on the phenotype. Recent data support that *Adh* is a gene that is evolving slowly (Sharp and Li 1989; Caccone and Powell 1990). Most of this information has been gathered through the analysis of *Adh* sequences of a group of sibling species of *Drosophila melanogaster* (Kreitman 1983; Bodmer and Ashburner 1984; Coyne and Kreitman 1986) and from single representatives of other groups (Fischer and Maniatis 1985; Schaeffer and Aquadro 1987; Atkinson et al. 1988; Rowan and Dickinson 1988). Further, when trying to establish a more comprehensive evolutionary history

of this specific genomic region, it becomes increasingly evident that each species or group of species ought to be considered on its own.

The *obscura* group of *Drosophila* comprises many species that have been the subject of prominent evolutionary studies (Lakovaara and Saura 1982; Cabrera et al. 1983; Loukas et al. 1984; Steinemann et al. 1984). The ADH protein of some European species has been characterized (Hernández et al. 1988). The aim of our work is to unravel the evolution of the *Adh* gene in a cluster of closely related European species of this group. We report the results obtained with the species *Drosophila ambigua*, the first that we have analyzed in detail. The *Adh* region has also been compared to one representative of the *melanogaster* group, *Drosophila mauritiana* (Cohn and Moore 1988) and to the only member of the *obscura* group that has been characterized to date, *Drosophila pseudoobscura* (Schaeffer and Aquadro 1987).

Recent data have also revealed that a number of complex events have occurred during the evolution of the *Adh* locus. Genome rearrangements such as gene duplications and the generation of pseudogenes have been fully documented in a reduced number of species (Oakshott et al. 1982; Batterham et al. 1984; Fisher and Maniatis 1985; Atkinson et al. 1988). Particularly interesting is the 3'ORF (open reading frame), a highly conserved sequence adjacent to *Adh*, described in several species of the subgenus *Sophophora*, as are the flanking regulatory regions potentially involved in the expression of both genes. The exploitation of previous information together with our present data can shed some light on the dynamics of the evolution of structural regions of the eukaryotic genome.

Materials and Methods

Fly Stock. DNA was obtained from a stock of flies of *D. ambigua* bred from a natural population caught in Vidrà, Barcelona, Spain, in 1985. Previous biochemical studies on *D. ambigua* ADH showed that this species was monomorphic for *Adh* (Hernández et al. 1988).

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. 1979), with minor modifications. High molecular weight total DNA was partially digested with *Mbo*I, then fractionated by sucrose gradient centrifugation to obtain DNA fragments of 15–20 kb. The library was constructed by cloning this DNA into the *Bam*HI site of an EMBL4 phage vector (Frischauf et al. 1983). Between 150,000 and 200,000 recombinant plaques were screened (Benton and Davis 1977) for sequences homologous to the *D. melanogaster Adh* gene. A 2.7-kb *Hind*III–*Eco*RI restriction fragment of *sACI* (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 (Rigby et al. 1977) or by

random-hexamer priming (Feinberg and Vogelstein 1983) with [α - 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% dextran sulphate. Non-specifically hybridized probe was removed with one wash in 2× SSC/0.1% SDS at room temperature for 10 min, then two washes in 2× SSC/0.1% SDS at 65°C for 10 min each, and finally two further washes in 1× SSC/0.1% SDS at 65°C for 10 min each (1× SSC is 0.15 M NaCl/0.015 M 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).

Restriction Analysis of Positive Clones. Positive clones were characterized by restriction mapping with the following restriction enzymes: *Bam*HI, *Eco*RI, *Hind*III, *Pvu*II, *Pst*I, *Sal*I, and *Xba*I (Boehringer Mannheim), using single and double digests (Maniatis et al. 1982). *Adh* was located on the restriction map by transferring digested and electrophoresed DNA to nitrocellulose membranes using the method of Southern (1975). Labeling of *Adh* probe, hybridization conditions, washes, and autoradiography were performed as for plaque hybridization.

DNA Sequence Analysis. Clones for sequencing were obtained by restriction digestion with suitable enzymes. These restriction fragments were subcloned into the Bluescript +KS, +SK phasmids (Stratagene). Several series of nested deletions were generated using the method of Henikoff (1984) to sequence large 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 et al. 1977), with [α - 35 S]dATP (NEN-DuPont) and the 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 h. Sequences on both strands were read and aligned with the 3.535-kb sequence of *D. pseudoobscura* (Schaeffer and Aquadro 1987) and the 4.596-kb sequence of *D. mauritiana* chosen as a representative member of the *D. melanogaster* group (Cohn and Moore 1988) as more sequence was available 3' to *Adh* for this species than for *D. melanogaster*. Alignments were determined by the Sequence Analysis Software Package of the Genetics Computer Group of the University of Wisconsin (Devereux et al. 1984).

Results

Clone Characteristics

The library screen yielded six positive clones that defined the same genomic region. Two of them, λ A3 and λ A311, were studied in detail. A restriction map of the *Adh* genomic region of *D. ambigua* is given in Fig. 1.

A restriction fragment of λ A3, containing the *Adh* gene and approximately half the length of the 3'ORF structural sequence was used to perform a cytological location on the polytene chromosomes. Hybridization at division 36 of chromosome E has been observed (Visa et al. 1991). This chromosome in *D. ambigua* is the homologue to chromosome 2L of *D. melanogaster* (Steinemann et al. 1984) where the *Adh* had previously been located (O'Donnell et al. 1977).

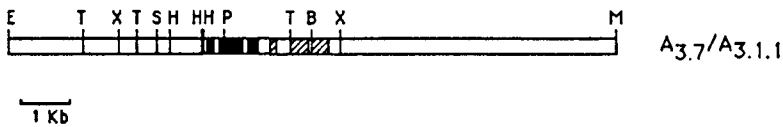
D. ambigua

Fig. 1. Restriction map of the *Adh* region of *D. ambigua*. Seven restriction endonucleases were used: B, *Bam*HI; E, *Eco*RI; H, *Hind*III; P, *Pvu*II; T, *Pst*I; S, *Sal*I; and X, *Xba*I. Solid boxes represent the coding regions of the *Adh* gene and hatched boxes represent the putative coding regions of the 3'ORF.

Sequence Comparison of *Adh* Coding Regions

A total of 4.185 kb of the genomic region containing the *Adh* of *D. ambigua* was sequenced (Fig. 2) and aligned with the homologous region of *D. mauritiana* (Cohn and Moore 1988) and that of *D. pseudoobscura* (Schaeffer and Aquadro 1987).

The structural features of the *D. ambigua Adh* gene, such as the adult and larval promoters, the proposed leaders, the transcriptional start sites and the polyadenylation signal were determined by homology with the compared sequences. Length differences were observed in the adult and larval leaders of the three species. The variability in the length of the introns is not unexpected, but consensus donor and acceptor splice sites are highly conserved [donor GTAAG, acceptor AG (Nevins 1983)]. One difference in the coding region of the *obscura* group species analyzed (*D. ambigua* and *D. pseudoobscura*) with respect to the *melanogaster* group species (*D. mauritiana* and others) is a deletion comprising 6 nucleotides (nt) corresponding to the third and fourth amino acids (Fig. 2).

The *Adh* coding region has accumulated 105 nt changes since *D. ambigua* and *D. mauritiana* diverged and 63 since *D. ambigua* and *D. pseudoobscura*. These substitutions are not randomly distributed in any codon position, as it is evident that most of them have occurred in the third nucleotide (*D. ambigua* vs *D. mauritiana* $\chi^2 = 83.45$ and *D. ambigua* vs *D. pseudoobscura* $\chi^2 = 40.66$; both cases $df = 2$; $P < 0.0001$). The observed 1:1 ratio of transitions to transversions deviates significantly from the expected ratio 1:2 (*D. ambigua* vs *D. mauritiana* $\chi^2 = 8.4$ and *D. ambigua* vs *D. pseudoobscura* $\chi^2 = 10.29$; both cases $df = 1$; $P < 0.01$). In addition, within the transition class, CT substitutions are more frequent than AG, also showing a significant deviation from the expected ratio of 1:1 (*D. ambigua* vs *D. mauritiana* $\chi^2 = 27.94$; $df = 1$; $P < 0.0001$; *D. ambigua* vs *D. pseudoobscura* $\chi^2 = 6.82$; $df = 1$; $P < 0.01$). Within the transversion class, GC substitutions are overrepresented whatever the sequence compared, although AC transversions are not infrequent between *D. ambigua* and *D. mauritiana*. The frequency of the four transversion types is clearly not random (*D. ambigua* vs *D. mauritiana* $\chi^2 = 38.0$; $df = 3$; $P < 0.0001$; *D. ambigua* vs *D. pseudoobscura*, $\chi^2 = 17.7$; $df = 3$; $P < 0.001$).

When considering the *Adh* coding sequence of *D.*

ambigua as a whole (Table 1), 27.6% of all sites are effectively silent (that is, 211.7 out of 765 nt sites). But according to our results 35.4% (75 nt) of these effectively silent sites have changed between *D. ambigua* and *D. mauritiana* and 24.1% (51 nt) between *D. ambigua* and *D. pseudoobscura Adh* sequences. Furthermore, these substitutions are randomly distributed throughout the three exons (*D. ambigua* vs *D. mauritiana* $\chi^2 = 1.04$; $df = 2$; $P > 0.5$; *D. ambigua* vs *D. pseudoobscura* $\chi^2 = 1.78$; $df = 2$; $P > 0.3$).

When considering the potential replacement sites (553.3) 30 changes (5.4%) have been detected between *D. ambigua* and *D. mauritiana*, and 12 changes (2.2%) between *D. ambigua* and *D. pseudoobscura*. If we compare *D. ambigua* vs. *D. mauritiana*, these replacement changes are not randomly distributed across the three exons but clustered in the third exon ($\chi^2 = 15.27$; $df = 2$; $P < 0.001$), although there is no deviation from random distribution along the whole sequence ($\chi^2 = 0.74$; $df = 2$; $P > 0.5$) when *D. ambigua* is compared with *D. pseudoobscura*. When considering the type of amino acid changes, 16 out of 20 are conservative between *D. ambigua* and *D. mauritiana*, as are 8 out of 10 between *D. ambigua* and *D. pseudoobscura*. Conservative changes are considered in the range of 5 to 4 on the scale of Feng et al. (1985). None of the essential amino acids predicted for alcohol dehydrogenase function (four glycines and one aspartic acid) has been replaced. Nevertheless, amino acid substitutions in *D. ambigua* alter the net charge of the enzyme.

3'ORF Sequence Analysis

Adjacent to the 3' end of *Adh*, there is a sequence that appears to be highly conserved in all the species studied and whose structure suggests a clear open reading frame [3'ORF (Schaeffer and Aquadro 1987)]. The main features of the 3'ORF of *D. ambigua* have been inferred from the comparison of the same structural region with both *D. pseudoobscura* and *D. mauritiana*. Three highly conserved sequences disrupted by two more variable regions may reveal the structure of a gene made up of three exons (Fig. 2). A detailed analysis of this sequence shows all the transcriptional and translational regulation signals of an expressing gene: a CCAAT and a TATA box, splicing consensus sequences, and a


```

CCCAAAGCCCTAAGAACGATAAGGCCGAGTAGCCTCGTCATTTAATTAALAACTAGATAAATGCCAGTGGCCAGTGGG 2240
CAGTGGCCATCGATTGCTAGCACAAGAACTGAACCTTTAGTAATAGAAAAAGAACTTGAGAAAGCAGCAACAAGATGTAC 2320
MetTyr
T A * * T * * G T * A A * *
GATCTGGGGGCAAGCAGTGTCTGCTATGTGGCTGACTGGCGTGGCATTGCACTGGAGACCAGCAAGGTCTCATGACCAA 2400
AspLeuAlaGlyLysHisValCysTyrValAlaAspCysGlyGlyIleAlaLeuGluThrSerLysValLeuMetThrLys
C
GAATATAGCCGTGAGTGTGTGGAGTGTGTAGAGTGGAGCCCGTTTGGGTTCTGGTTCCTGAAACAAGTCTCTAGGGTGC 2480
sAsnIleAla
TGCTGCTGCTGCTGGCGGCTTGTAGCTGCGAACTAATTAAGTTAATGACTTTTTCATGTTATTTTGTGTGCTGAGT 2560
TGCTACAATTAAGTTTATGTTTGTGGTGTITGATCTACTCTTGCTACAGGGAACTACTTTCGATTATGGATTAAGGGCAT 2640
G T A
A C G * T C T *
ACATAAAAAAAGCCCTTCAATATTGGTCTATTCTAGTAAAAATTTCTCTGATATCCAGAAATAGCCATTCTGCAG 2720
LysLeuAlaIleLeuGln
CAC T C C A * * G * A C A T T A A T * * A C *
C * C C * * T A * * T * * * * * * * * *
AGTGTGAAAAACCAACCGCCATCGCTCAGCTACGTTAAGCAGCACACAGATCTTCTTGGACCTTCGATGT 2800
SerValGluAsnGlnProAlaIleAlaGlnLeuGlnSerIleLysHisSerThrGlnIlePhePheTrpPheAspVa
A A A A * * G * T * G * * * C G *
* * G * G * T * * * * * * * * *
GACCATGGCCCGGAGGAAATGAAGAAATCTTCGACGAGGTCAATGGTCCAAATGGACTACATCGATGTGTTATCAATG 2880
lThrMetAlaArgGluGluMetLysLysTyrPheAspGluValMetValGlnMetAspTyrIleAspValLeuIleAsnG
A T * * A A T * T A A G A * * T * G *
* * T * * * * * * * * *
GGGCCACCTGTGGATGAGCGGAACATCGATGCCACCATCAACAGAACTTCGACCGCATGAAACAGCTGGCCAC 2960
lyAlaThrLeuLysAspGluArgAsnIleAspAlaThrIleAsnThrAsnLeuThrGlyMetMetAsnThrValAlaThr
T A * * A A A A A T * G C A * C T * C * T *
* T * * T G * A C * * * C C A * * T *
GTGTCGCCATATGGACCGAAGATGGCGGCTCGGTTGGATTGTGAATGTCACCTCGGTGATTGGACTGGACCC 3040
ValLeuProTyrMetAspArgLysMetGlyGlySerGlyLeuIleValAsnValThrSerValIleGlyLeuAspPr
T G T C C C T A C C * * A T C A * *
* * * * * * * * *
GTCGCCAGTCTTTTGGCGTACAGTCCCTCAAAGTTTGGTGTGATTGGCTTCACAGAACTCTAGCCGTGAGTGTGGGG 3120
oSerProValPheCysAlaTyrSerAlaSerLysPheGlyValIleGlyPheThrArgSerLeuAla
C C T * T T C C *
* * C * T * * C C *
GGGAGATCTTACCACATAAAATAACTATTCTCTGATATCTTTATCCTGCGCAGGATCCTCTGACTACAAATCAAATGG 3200
AspProLeuTyrTyrAsnGlnAsnGln
A C * * G T * * T A G G C * G C G * A G T A A A *
C * C * C * C * C * A
TGTGGCTGTGATGGCTGTCTGTGTGGTCCCAAGAGTGTGTGCGATCGGCAACTGAATGCCCTTCTGGAGTACGGCC 3280
yValAlaValMetAlaValCysCysGlyProThrLysValPheValAspArgGluLeuAsnAlaPheLeuGluTyrGlyG
T C * * G A T * * A T G G T * * C * * C G *
G T C * * T G T * * * C * * A * C * C G
AAACGTTTGGCGATCGCTGGCGCTGCCAATCGACGGCCCTCTGTGGCCAGAATATTGTAATGCCATTGAA 3360
lThrPheAlaAspArgLeuArgLeuAlaProCysGlnSerThrAlaValCysGlyGlnAsnIleValAsnAlaIleGlu
G * T C * C T * T A * T C * A A T * T C * C *
* * * * * * * * *
AGATCGCAGAACGACAGATATGGATTGGCCACAGGGCGCCCTGGAGCTGGTACCCCTGCACTGGTATTGGCATATGGC 3440
ArgSerGlnAsnGlyGlnIleTrpIleAlaAspLysGlyGlyLeuGluLeuValThrLeuHisTrpTyrTrpHisMetAl
C G C C A * * A T A * A T A *
T G A * * * C T * * G * T T * T G G C C A C G A *
CGATCAGTTATCAGCTATATGACAGACAGGATGACGAGGATCAGGAATGCTTTCTATCGGCGGGCTATCGGAAATAAA 3520
aAspGlnPheIleSerTyrMetGlnSerThrAspArgGluAspGlnGluCysPheLeuSerAlaAlaTyrArgLysEND
A T G C C A A T G A T A A A G T T T A A A G G T G A C A A A A A G T A G G C C A A T T T G A T T T G C G G T A T G G A A A T A G A T T T G A A A T A T T 3600
T G T G G G T C A T A T G A A T A A T G T C A T A T A C C G A G G C C C C T G A A A G T A T G C C C T G G T T T G G A T G G C A C T A C G A G T G C T A G 3680
A A G A C A A T A G A T T A T G A T A G T A G A G G G T G G A A T G G G A A G C T T T A C G C C T A G G A A T A C G T G G G T A T G C T C T A G A 3760

```

Fig. 2. Continued from page 457

polyadenylation signal, all of which are located in the expected positions. The sequence evolution of this potential gene has been analyzed in a similar way to that of *Adh*.

There is a remarkable similarity in the first two exons but there are length differences in the third exon that would account for the different number of amino acids in the protein: 281 for *D. ambigua*, 278 for *D. pseudoobscura*, and 273 for *D. mauritiana*.

The pattern of nucleotide substitution resembles that of the *Adh* gene. The putative coding region has accumulated a total of 144 changes since the divergence of *D. ambigua* and *D. mauritiana*, and 87

changes between *D. ambigua* and *D. pseudoobscura*. These substitutions are not randomly distributed among the three positions of a codon, but rather they accumulate in the third position (*D. ambigua* vs *D. mauritiana* $\chi^2 = 124.55$ and *D. ambigua* vs *D. pseudoobscura* $\chi^2 = 59.86$; both cases $df = 2$; $P < 0.0001$). The ratio of transitions to transversions is close to 1:1 as is found when analyzing the *Adh* gene, and this deviates significantly from the expected random ratio of 1:2 (*D. ambigua* vs *D. mauritiana* $\chi^2 = 18.0$; $df = 1$; $P < 0.0001$ and *D. ambigua* vs *D. pseudoobscura* $\chi^2 = 13.24$; $df = 1$; $P < 0.001$). When considering the type of transitions in the 3'ORF there is no relevant difference in CT or AG

Table 1. *Adh* coding sequences

Region	<i>D. ambigua</i> Length (bp)	<i>D. ambigua</i> vs <i>D. pseudoobscura</i>			<i>D. ambigua</i> vs <i>D. mauritiana</i>		
		No. of changes ^a	Percent differ- ences	Cor- rected percent diver- gence ^b	No. of changes ^c	Percent differ- ences	Cor- rected percent diver- gence ^b
ADH							
5' noncoding	32.0	3	9.4	10.0	12	37.5	52.0
Adult leader	76.0	11	14.7	16.1	16	21.1	24.7
Adult intron	953.0	225	23.6	28.3	263	27.6	34.4
Adult and larval leader	66.0	12	18.2	20.8	29	43.9	66.1
Exon 1 Silent	29.3	6	20.5	23.9	11	37.5	52.1
Replacement	63.7	2	3.1	3.2	2	3.1	3.2
Intron 1	66.0	22	33.3	44.1	25	37.9	52.7
Exon 2 Silent	113.7	24	21.1	24.8	36	31.7	41.1
Replacement	291.3	7	2.4	2.4	7	2.4	2.4
Intron 2	67.0	24	35.8	48.7	23	34.3	45.9
Exon 3 Silent	68.7	21	30.6	39.3	28	40.8	58.8
Replacement	198.3	3	1.5	1.5	21	10.6	11.4
Intergenic region	243.0	38	15.6	17.5	77	31.7	41.2
3' gene							
5' noncoding	20.0	2	10.0	10.7	5	25.0	30.4
Leader	93.0	15	16.1	18.2	30	32.3	42.2
Exon 1 Silent	24.0	1	4.2	4.3	6	25.0	30.4
Replacement	72.0	1	1.4	1.4	2	2.8	2.8
Intron 1	292.0	54	18.5	21.2	81	27.2	33.8
Exon 2 Silent	107.0	32	29.9	38.2	64	59.8	>100.0
Replacement	298.0	2	0.7	0.7	9	3.0	3.1
Intron 2	67.0	18	26.9	33.3	23	34.3	45.9
Exon 3 ^c Silent	86.0	32	38.2	53.5	46	59.2	>100.0
Replacement	259.0	19	7.5	7.9	17	7.1	7.4
Summary							
Noncoding	1975.0	424	21.4	25.2	584	29.5	37.5
Adh Silent	211.7	51	24.1	29.1	75	35.4	48.0
Replacement	553.3	12	2.2	2.2	30	5.4	5.6
Total	765.0	63	8.2	8.7	105	13.7	15.1
3' gene ^c Silent	217.0	65	30.3	38.8	116	55.6	>100.0
Replacement	629.0	22	3.5	3.6	28	4.6	4.7
Total	846.0	87	10.4	11.2	144	17.6	20.0

^a The effective number of silent sites for a gene depends on its amino acid composition. It is calculated as the percentage of the potentially silent sites (possible substitutions that would not lead to an amino acid replacement) out of the total coding nucleotides (Holmquist et al. 1972). Stop codons are included in these figures. Insertions and deletions are oversimplified to one single event in noncoding regions

^b Corrected percent divergence is estimated as $d = \frac{3}{4} \ln(1 - \frac{1}{3}p)$, where p is the proportion of nucleotide sites that differ between two sequences (Jukes and Cantor 1969)

^c Due to the variable length of the 3' ORF third exon we have not used the total figures for silent and replacement sites of this table. Instead we have considered the following figures when comparing this third exon of *D. ambigua* vs *D. pseudoobscura*: 83.7 for silent and 252.3 for replacement sites; when comparing *D. ambigua* vs *D. mauritiana*: 77.7 for silent and 240.3 for replacement sites

substitutions (*D. ambigua* vs *D. mauritiana* $\chi^2 = 2.72$ and *D. ambigua* vs *D. pseudoobscura* $\chi^2 = 1.8$; both cases $df = 1$; $P > 0.1$). A significant deviation from random frequency is obtained when the type of transversion is analyzed: AC substitutions are overrepresented when *D. ambigua* and *D. mauritiana* are compared ($\chi^2 = 20$; $df = 3$; $P < 0.001$) whereas GC transversions are more frequent when considering *D. ambigua* vs *D. pseudoobscura* ($\chi^2 = 12.1$; $df = 3$ and $P < 0.01$).

With respect to the 3'ORF in *D. ambigua* (Table

1) 25.7% (217.0 nt) of the 846 coding nucleotides are effectively silent. Due to the variable length of the third exon, the number of coding nucleotides changes with species. So, the number of effectively silent sites is 208.7 out of 819 coding nucleotides (25.5%) when comparing *D. ambigua* to *D. mauritiana*, and 214.7 out of 837 coding nucleotides (25.7%) when comparing *D. ambigua* to *D. pseudoobscura*. The number of effectively silent site substitutions is high and the proportion to total effectively silent sites is 55.6% (116 nt) for *D. ambigua* and *D. mauri-*

tiana, and 30.3% (65 nt) for *D. ambigua* and *D. pseudoobscura*. Neither silent ($\chi^2 = 4.57$; $df = 2$; $P > 0.1$) nor replacement substitutions ($\chi^2 = 5.35$; $df = 2$; $P > 0.05$) significantly deviate from random distribution among the three exons when comparing *D. ambigua* vs *D. mauritiana* but both types of substitution do show a significant deviation when considering *D. ambigua* vs *D. pseudoobscura* ($\chi^2 = 7.16$; $df = 2$; $P < 0.05$ for silent substitutions and $\chi^2 = 19.25$; $df = 2$; $P < 0.0001$ for replacement substitutions). Particularly, the second exon seems to present a considerable constraint for replacement substitutions whereas the third exon shows a great accumulation of both types of substitutions. Even without taking into account the last 24 nt of the third exon (the region that shows differences in the position of the termination codon), this trend is significant ($\chi^2 = 9.16$; $df = 2$; $P < 0.02$).

Alignment of the potential protein sequences of this 3' gene shows 22 replaced amino acids between *D. ambigua* and *D. mauritiana* and 16 between *D. ambigua* and *D. pseudoobscura*. In both cases, most of these changes are conservative.

Discussion

Sequence analysis of specific genes highlights some aspects of molecular evolution. Our data concern a genomic region that involves about 4 kb and contains two genes: the *Adh* and the 3'ORF, whose structural features strongly suggest that it is expressed. Both genes are tandemly arranged and appear to have evolved quite differently.

Nucleotide substitutions in exon and intron sites among the *Adh* genes of *D. ambigua*, *D. pseudoobscura* and *D. mauritiana* appear to resemble those of the *melanogaster* group (Ashburner et al. 1984). The number of amino acid changes is higher between *D. ambigua* and *D. mauritiana* than between *D. ambigua* and *D. pseudoobscura*, as could be expected given that these latter belong to different groups of the genus and thus have had more time to diverge. Nevertheless, the total number of replaced amino acids is not high and this could be explained by the functional constraint imposed on *Adh*, a gene coding for an enzyme that plays an important role in individual adaptation to habitat. Then, a strong purifying selection may have reduced replacement substitutions.

The accumulation of nucleotide substitutions in the third codon position and the deviation of the ratio of transitions to transversions is consistent with the high frequency of silent substitutions. Due to the degeneracy of the genetic code, substitutions in the third position of a codon, particularly transitions, are frequently silent. However, this does not

explain why certain types of transitions or transversions are more frequent than others.

Replacement substitutions in the *Adh* gene are not randomly distributed among the exons: some regions appear to be more selectively constrained than others, due probably to their contribution to enzyme function. Thus, when comparing *D. ambigua* vs *D. mauritiana*, most of the amino acid changes accumulate in the third exon, which is not thought to be involved in the substrate or coenzyme binding domains (Thatcher 1980; Benyajati et al. 1983; Duester et al. 1986). However, this accumulation is not evident when comparing any combination of any of the *obscura* species analyzed: *D. ambigua* vs *D. pseudoobscura*, and three other species, *D. subobscura*, *D. guanche*, and *D. madeirensis* (data not shown), where substitutions appear to be randomly distributed across the three exons. The different distribution of substitutions could be explained by the variation of the selection pressure on a particular DNA sequence over time. On the other hand, there are some common amino acid changes (mainly located in the third exon), which are shared by the *obscura* group species but which are absent in *D. mauritiana* and its close relatives. It could thus be assumed that these replacements took place after the divergence of the *melanogaster* and *obscura* groups and constitute a feature of the evolution of the *obscura* group (Sullivan et al. 1990).

Sequences downstream of the *Adh* gene were found to be more highly conserved than would have been predicted for intron and intergenic sequences (Kreitman 1983; Schaeffer and Aquadro 1987). We also found a 3'ORF in *D. ambigua*, which shows all the features of a functional gene. The pattern of nucleotide substitution of this 3'ORF does not differ from that of the *Adh* gene, i.e., there is a high level of silent nucleotide substitutions, a feature that defines functional constraint. This, together with other structural features, suggests that rather than being a pseudogene it is a fully functional gene, although neither a transcript nor any protein product has been reported to date.

A characteristic of this 3'ORF is its high similarity to *Adh* (Table 2), which is reflected at the nucleotide level (51.6% of homology, which significantly deviates from the expected random value of 25%; $\chi^2 = 287.1$; $df = 1$; $P < 0.0001$) and also at the amino acid level (36.47% of identities, strongly deviating from the expected random homology of 6.1%, $\chi^2 = 409.0$; $df = 1$; $P < 0.0001$). The two genes may have diverged from an ancient precursor (Schaeffer and Aquadro 1987). Estimates for this duplication event give values ranging from 60.5 million years (Myr) to 121 Myr, assuming 0.8% base substitution per million years in the *melanogaster* group or 0.4% base substitution per million years

Table 2. Nucleotide sequence comparison between the *Adh* gene and the 3' ORF of *D. ambigua*

Region	Length (bp)	No. of changes ^a	Percent differences	Corrected percent divergence ^b
ADH vs 3' ORF				
Exon 1				
Silent	29.3	15	51.2	86.1
Replacement	63.7	26	40.8	58.9
Exon 2				
Silent	113.7	70	61.6	>100.0
Replacement	291.3	127	43.6	65.3
Exon 3				
Silent	68.0	34	50	82.4
Replacement	196.0	97	49.5	80.9
Total exons				
Silent	211.0	119	56.4	>100.0
Replacement	551.0	250	45.4	69.7
Total	762.0	369	48.4	77.8

^a The effective number of silent sites for a gene depends on its amino acid composition. It is calculated as the percentage of the potentially silent sites (possible substitutions that would not lead to an amino acid replacement) out of the total coding nucleotides (Holmquist et al. 1972). Insertions and deletions are oversimplified to one single event

^b Corrected percent divergence is estimated as $d = \frac{1}{3} \ln(1 - \frac{1}{3}p)$, where p is the proportion of nucleotide sites that differ between two sequences (Jukes and Cantor 1969)

in the *obscura* group (Caccone et al. 1988; Caccone and Powell 1990).

Data obtained from interspecific comparison of the distribution of nucleotide substitutions in the 3'ORF differ substantially from those concerning the *Adh* gene. When comparing the *Adh* of *D. ambigua* vs *D. mauritiana*, silent substitutions are distributed randomly among the three exons, whereas replacement substitutions are highly significant. But, when considering the 3'ORF of these two species, there is no significant deviation from random distribution for either silent or replacement substitutions. Conversely, when comparing *D. ambigua* vs *D. pseudoobscura*, the 3'ORF does show a highly deviated distribution of both types of substitution. On the other hand, in the case of *D. ambigua* vs *D. mauritiana* it is striking that although all silent sites are saturated in the 3'ORF, this gene presents fewer replacement substitutions than *Adh*. This trend is not observed in *D. ambigua* vs *D. pseudoobscura*, where both types of substitutions are more frequent in 3'ORF than in *Adh*.

In summary, the 3'ORF has accumulated more nucleotide substitutions than *Adh*, which, according to other authors and our present data, is evolving relatively slowly (Sharp and Li 1989; Caccone and Powell 1990). Furthermore, some evolutionary

trends that could be inferred from our results are worth emphasizing: first, *Adh* and the 3'ORF are both evolving independently and second, their evolution follows a species-specific pattern. Consequently, considerable differences in the estimation of divergence for the three species analyzed would be obtained depending on the gene that is used for comparison. Moreover, if we consider that considerable variation in nucleotide substitution rates has been reported in species belonging to different *Drosophila* groups (Caccone and Powell 1990), an accurate estimation of divergence would require additional information. Divergence rate comparisons between vertebrate organisms and *Drosophila* are also unreliable, as it has been suggested that insect DNA is evolving at a higher rate (Caccone and Powell 1990).

The genomic region analyzed has proved to be useful in the elucidation of some aspects of genome evolution: two tandemly arranged genes that have diverged from a common ancestor are clearly evolving with a different evolutionary pattern. This illustrates a range of variability of single-copy DNA evolution, which is crucial to an understanding of evolutionary trends.

Acknowledgments. We are grateful to Dr. L. Serra for helpful comments on the manuscript and to R. Rycroft for revising the English version. This research was supported by grants from the Dirección General de Investigación Científica y Técnica (DGI-CYT) PB88-0188 and CIRIT (Generalitat de Catalunya) AR 87 and AR 88. G. Marfany is in receipt of an FPI fellowship.

References

- Ashburner M, Bodmer M, Lemeunier F (1984) On the evolutionary relationships of *Drosophila melanogaster*. *Dev Genet* 4:295-312
- Atkinson P, Mills W, Starmer LE, Sullivan DT (1988) Structure and evolution of the *Adh* genes of *Drosophila mojavensis*. *Genetics* 120:713-723
- Batterham P, Chambers GK, Starmer WT, Sullivan DT (1984) Origin and expression of an alcohol dehydrogenase duplication in the genus *Drosophila*. *Evolution* 38:644-657
- Benton WD, Davis RW (1977) Screening lambda-gt recombinant clones by hybridization to single plaques in situ. *Science* 196:180-182
- Benyajati C, Place A, Powers D, Sofer W (1981) Alcohol dehydrogenase gene of *D. melanogaster*: relationship to intervening sequences to functional domains in the protein. *Proc Natl Acad Sci USA* 78:2717-2721
- Benyajati C, Spoerel N, Haymerle H, Ashburner M (1983) The messenger RNA for alcohol dehydrogenase in *Drosophila melanogaster* differs in its 5' end in different developmental stages. *Cell* 33:125-133
- Bodmer M, Ashburner M (1984) Conservation and change in the DNA sequences coding for *Adh* in sibling species of *Drosophila*. *Nature* 309:425-430
- Cabrera VM, González AM, Larruga JM, Gullon A (1983) Genetic distance and evolutionary relationships in the *Drosophila obscura* group. *Evolution* 37:675-689

- Caccone A, Powell JR (1990) Extreme rates and heterogeneity in insect DNA evolution. *J Mol Evol* 30:273–280
- Caccone A, Amato GD, Powell JR (1988) Rates and patterns of scnDNA and mtDNA divergence within the *Drosophila melanogaster* subgroup. *Genetics* 118:671–683
- Chambers GK (1988) The *Drosophila* alcohol dehydrogenase gene–enzyme system. In: Caspari EW, Scandalios JG (ed) *Advances in genetics*, vol 25. Academic Press, New York, pp 39–107
- Chirgwin JM, Przybyla AE, MacDonald RJ, Rutter WJ (1979) Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* 18:5294–5299
- Cohn VH, Moore GP (1988) Organization and evolution of the alcohol dehydrogenase gene in *Drosophila*. *Mol Biol Evol* 5:154–166
- Coyne JA, Kreitman M (1986) Evolutionary genetics of two sibling species, *Drosophila simulans* and *Drosophila sechellia*. *Evolution* 40:673–691
- Devereux J, Haerberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res* 12:387–395
- Duester G, Jornvall H, Hatfield GW (1986) Intron-dependent evolution of the nucleotide-binding domains within alcohol dehydrogenase and related enzymes. *Nucleic Acids Res* 14:1931–1941
- Feinberg AP, Vogelstein B (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal Biochem* 132:6–13
- Feng DF, Johnson MS, Doolittle RF (1985) Aligning amino acid sequence: comparison of commonly used methods. *J Mol Evol* 21:112–125
- Fisher JA, Maniatis T (1985) Structure and transcription of the *Drosophila mulleri* alcohol dehydrogenase genes. *Nucleic Acids Res* 13:6899–6917
- Fischer JA, Maniatis T (1988) *Drosophila Adh*: a promoter element expands the tissue specificity of an enhancer. *Cell* 53:451–461
- Frischauf A-M, Lehrach H, Poustka A, Murray N (1983) Lambda replacement vectors carrying polylinker sequences. *J Mol Biol* 170:827–842
- Goldberg DA (1980) Isolation and partial characterization of the *Drosophila* alcohol dehydrogenase gene. *Proc Natl Acad Sci USA* 77:5794–5798
- Heberlein U, England B, Tjian R (1985) Characterization of *Drosophila* transcription factors that activate the tandem promoters of the alcohol dehydrogenase gene. *Cell* 41:965–977
- Henikoff S (1984) Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* 28:351–359
- Hernández JJ, Vilageliu LL, González-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, Jukes T (1972) Improved procedures for comparing homologous sequences in molecules of protein and nucleic acids. *J Mol Biol* 64:145–161
- Hopp TP, Woods KR (1981) Prediction of protein antigenic determinants from amino acid sequences. *Proc Natl Acad Sci USA* 78:3824–3828
- Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN (ed) *mammalian protein metabolism*, vol. 3. Academic Press, New York, pp 21–132
- Kreitman M (1983) Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. *Nature* 304:412–417
- Lakovaara S, Saura A (1982) Evolution and speciation in the *Drosophila obscura* group. In: Ashburner M, Carson HL, Thompson JNS (eds) *The genetics and biology of Drosophila*, vol 3b. Academic Press, London, pp 2–9
- Loukas M, Krimbas CB, Vergini Y (1984) Evolution of the *obscura* group *Drosophila* species. II. Phylogeny of ten species based on electrophoretic data. *Heredity* 53:483–493
- Maniatis T, Fritsch EF, Sambrook J (1982) *Molecular cloning: a laboratory manual*, ed 1. Cold Spring Harbor Laboratory, Cold Spring Harbor NY
- Miyata T (1982) Evolutionary changes and functional constraints in DNA sequences. In: Kimura M (ed) *Molecular evolution and the neutral theory*. Japan Scientific Societies Press, Tokyo, pp 233–266
- Nevens JR (1983) The pathway of eucaryotic mRNA formation. *Annu Rev Biochem* 52:441–466
- Oakeshott J, Chambers G, East P, Gibson J, Barker J (1982) Evidence for a genetic duplication involving *Adh* in *D. buzzatii* and related species. *Aust J Biol Sci* 35:73–84
- O'Donnell J, Mandel HC, Krauss M, Sofer W (1977) Genetic and cytogenetic analysis of the ADH region in *Drosophila melanogaster*. *Genetics* 86:553–566
- Rigby PWJ, Dieckmann M, Rhodes C, Berg P (1977) Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase. I. *J Mol Biol* 113:237–251
- Rowan R, Dickinson WJ (1988) Nucleotide sequence of the genomic region coding for alcohol dehydrogenase in *Drosophila affinisdisjuncta*. *J Mol Evol* 28:43–54
- Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 74:5463–5467
- Schaeffer SW, Aquadro CF (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 PM, Li W-H (1989) On the rate of DNA sequence evolution in *Drosophila*. *J Mol Evol* 28:398–402
- Sofer W, Martin PF (1987) Analysis of alcohol dehydrogenase gene expression in *Drosophila*. *Annu Rev Genet* 21:203–225
- Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* 98:503–517
- Steinemann M, Pinsker W, Sperlich D (1984) Chromosome homologies within the *Drosophila obscura* group probed by in situ hybridization. *Chromosoma* 91:46–53
- Sullivan DT, Atkinson PW, Starmer WT (1990) Molecular evolution of the alcohol dehydrogenase genes in the genus *Drosophila*. In: Hecht MK, Wallace B, MacIntyre RJ (ed) *Evolutionary biology*. Plenum, New York, pp 107–147
- Thatcher DR (1980) The complete amino acid sequence of three alcohol dehydrogenase alleloenzymes (Adh^{n-11} , Adh^s and Adh^{of}) from the fruitfly *D. melanogaster*. *Biochem J* 187:875–886
- 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
- Visa N, Marfany G, Vilageliu LL, Albalat R, Atrian S, González-Duarte R (1991) The *Adh* locus in *Drosophila*: chromosomal location and restriction analysis in species with different phylogenetic relationships. *Chromosoma* (in press)