

**ORIGEN DEL FENOTIPO *zeste* EN MACHOS DE LA CEPA
M115 DE *D. melanogaster* Y CAUSAS DE SU REVERSIÓN:
EL ELEMENTO TRANSPONIBLE *FB-NOF*.**

TESIS DOCTORAL

Martí Badal Soler

2007

UNIVERSITAT AUTÒNOMA DE BARCELONA

Grup de Mutagènesi

Unitat de Genètica

Departament de Genètica i de Microbiologia

Facultat de Biociències



Universitat Autònoma de Barcelona

Facultat de Biociències

Departament de Genètica i de Microbiologia

Unitat de Genètica

**ORIGEN DEL FENOTIPO *zeste* EN MACHOS DE LA CEPA
M115 DE *D. melanogaster* Y CAUSAS DE SU REVERSIÓN:
EL ELEMENTO TRANSPONIBLE *FB-NOF*.**

Memoria presentada por MARTÍ BADAL SOLER, en el marco del programa de Genética de tercer ciclo de la Universitat Autònoma de Barcelona, para optar al grado de Doctor en Genética por la Universitat Autònoma de Barcelona.

El trabajo se ha realizado en los laboratorios del Grup de Mutagènesi bajo la codirección de los Doctores ORIOL CABRÉ FABRÉ y NOEL XAMENA LÓPEZ.

Director de Tesis

Director de Tesis

Autor

Dr. Oriol Cabré Fabré

Dr. Noel Xamena López

Martí Badal Soler

El presente trabajo trata sobre dos temas coyunturalmente relacionados: el fenotipo de las cepas M115 y RM115 de *Drosophila melanogaster* y el poco conocido elemento transponible *FB-NOF*.

La cepa M115 de *D. melanogaster* presenta el fenotipo *zeste*¹ ampliado a los machos. Este fenotipo se caracteriza por los ojos de color amarillo claro y, aunque normalmente se presenta en hembras pues precisa del apareamiento de dos copias del gen *white*, puede observarse en machos portadores de duplicaciones de dicho gen. Es la conocida interacción *zeste-white*. M115 es una cepa inestable y se pueden encontrar individuos revertientes de forma regular. Un macho revertiente dio lugar a la cepa RM115. Este par de cepas son fenotípicamente parecidas a las descritas por la Dra. Rasmuson-Lestander, w^{+UZ} y w^{+UR} . Éstas son mutantes insercionales y la presencia de un elemento transponible *FB-NOF* en w^{+UZ} parece ser la causa de la ampliación del fenotipo *zeste* en los machos.

Siguiendo los pasos de la caracterización de w^{+UZ} y w^{+UR} , identificamos la inserción de un elemento transponible *FB-NOF* en nuestras cepas M115 y RM115, pocas kilobases a 3' del gen *white*. Aunque el punto de inserción es exactamente el mismo que en las cepas de Rasmuson-Lestander, nuestros experimentos de *Southern Blot* demuestran que se trata de inserciones distintas, lo cual plantea si esa región podría ser un *hot spot* de integración para *FB-NOF*. Al llevar el análisis un paso más adelante, encontramos que las cepas de ojos amarillos, M115 y w^{+UZ} , eran portadoras de una duplicación del gen *white*, que explica el fenotipo ampliado a los machos. La reversión en RM115, por su parte, consiste en la delección de una de las dos copias del gen. Así pues, debemos descartar las hipótesis de Rasmuson-Lestander que apuntaban a la interferencia de *FB-NOF* en la interacción *zeste-white*.

El elemento transponible *FB-NOF* es uno de los menos conocidos de *D. melanogaster*. Dio origen al grupo de transposones denominados *foldback* por su estructura modular repetitiva y se desconocen los detalles de su biología. En este trabajo proponemos una forma más sencilla de considerar la estructura del elemento que facilita su estudio, sobretudo a partir de secuencias genómicas. Además, analizamos su distribución en el genoma de *D. melanogaster*, la relación entre las secuencias que lo integran y la expresión de su secuencia codificante. Los resultados de dicho análisis revelan que *FB-NOF* funciona como un solo transposón con un sesgo en sus preferencias de inserción y que se mantiene activo en la actualidad.

ABREVIATURAS

aa	aminoácido	MOR	Moira
ATP	adenosín trifosfato	mRNA	<i>messenger RNA</i>
BAC	<i>bacteria artificial chromosome</i>	NBT	<i>nitroblue tetrazolium</i>
BCIP	5-bromo-4-cloro-3-indolyl- fosfato	NOF	<i>Nofretette</i>
bp	<i>base pair</i>	ORF	<i>open reading frame</i>
BRM	Brahma	PBS ^a	<i>phosphate buffered saline</i>
cM	centi Morgan	PBS ^b	<i>primer binding site</i>
CDS	<i>Coding sequence</i>	PcG	<i>polycomb group</i>
DAPI	4',6-diamidino-2-phenylindole	PCR	<i>polymerase chain reaction</i>
DNA	ácido desoxiribonucleico	PEV	<i>position effect variegation</i>
EDTA	<i>ethylenediaminetetraacetic acid</i>	PTGS	<i>post transcriptional gene silencing</i>
EGTA	<i>ethylene glycol tetraacetic acid</i>	RISC	<i>RNA-induced silencing complex</i>
EMS	metanosulfonato de etilo	RNA	ácido ribonucleico
ENU	N-etil-N-nitrosourea	RPM	revoluciones por minuto
ET	elemento transponible	SDS	sodio dodecilo sulfato
FB	<i>foldback</i>	SINE	<i>short interspersed nuclear elements</i>
FBS	<i>fetal bovine serum</i>	siRNA	<i>small interference RNA</i>
FITC	<i>Fluorescein Isothiocyanate</i>	SNP	<i>single nucleotide polymorphism</i>
FISH	<i>fluorescent in situ hybridisation</i>	STS	<i>sequence tagged site</i>
GFP	<i>green fluorescent protein</i>	TGS	<i>transcriptional gene silencing</i>
HTH	<i>helix-turn-helix</i>	TIR	<i>terminal inverted repeats</i>
IR	<i>inverted repeat</i>	tRNA	<i>transfer RNA</i>
IS	<i>insertion sequence</i>	TrxG	<i>trithorax group</i>
kb	kilobase	UZ	<i>unstable zeste</i>
kDa	kilo Dalton	UR	<i>unstable red</i>
LB	Luria-Bertrani	UTR	<i>untranslated region</i>
LINE	<i>long interspersed nuclear element</i>	UV	ultra violeta
LTR	<i>long terminal repeat</i>	VLP	<i>virus-like particles</i>
Mb	megabase	ZBS	<i>Zeste binding site</i>
MCS	<i>Multiple cloning site</i>		
MITEs	<i>miniature inverted transposable elements</i>		

1. RESUMEN Y LISTADO DE ABREVIATURAS	
2. ÍNDICE GENERAL	
3. INTRODUCCIÓN	
3.1. Antecedentes	13
3.2. Los elementos transponibles	16
El descubrimiento y redescubrimiento	
Los elementos de clase-I.1 se parecen a retrovirus	
Los elementos de clase-I.2 parecen mRNA	
Los elementos de clase-II <i>saltan</i> literalmente	
¿Existe la clase-III de elementos transponibles?	
(Casi) Todos los organismos tienen ETs	
Existe un patrón general en la distribución de ETs	
Los transposones pueden alterar gravemente los genomas huésped	
Sin embargo, también pueden suponer algunas ventajas...	
Los genomas tienen mecanismos de inactivación de ETs	
El elemento <i>FB</i> de <i>D. melanogaster</i> es uno de los transposones más desconocidos	
<i>NOF</i> se encuentra íntimamente asociado a <i>FB</i>	
3.3. El gen <i>white</i>	33
3.4. El gen <i>zeste</i>	36
Dos de las mutaciones más estudiadas de <i>zeste</i> son z^1 y z^{op6}	
La interacción <i>zeste-white</i> depende del número de copias de <i>white</i>	
<i>Zeste</i> ¹ podría inactivar por falta de activación	
3.5. El genoma de <i>Drosophila melanogaster</i>	41
La secuenciación <i>shotgun</i> necesita grandes números	
Los pares de secuencia permiten secuenciar grandes genomas	
La llave está en el <i>software</i>	
Finalizar la secuencia consumió algunos años extra	
4. OBJETIVOS	51
5. RESULTADOS Y DISCUSIÓN	
5.1. Análisis de las cepas M115 y RM115	55
Descubrimos la inserción de un elemento <i>FB</i>	
La inserción en M115 y RM115 es distinta de la de w^{+UZ}	
Aparecen los primeros indicios de una reordenación	
<i>white</i> está duplicado, por lo menos en las cepas de ojos amarillos	
Discusión general del análisis de las cepas M115 y RM115	
5.2. Análisis del Elemento Transponible <i>FB-NOF</i>	70
Clonamos los IR de un elemento <i>FB</i>	
La secuenciación de <i>FB</i> proporciona una nueva interpretación de su estructura	
Recorremos el genoma en busca de <i>FB-NOF</i>	
La inserción de <i>FB</i> se produce en contextos de fácil acceso	
La posición de <i>NOF</i> dentro de <i>FB</i> sí es específica de secuencia	
Los <i>NOF</i> en las bases de datos están truncados. ¿Lo están en M115?	
<i>NOF</i> tiene entre 1 y 3 ORFs que codifican proteínas únicas	
<i>NOF</i> podría generar un mRNA dicistrónico	
La expresión de los ORF respalda las predicciones	
Discusión general del análisis del elemento transponible <i>FB-NOF</i>	

6. CONCLUSIONES	105
7. MATERIALES Y MÉTODOS	
7.1. Técnicas <i>in vivo</i> e <i>in vitro</i>	
7.1.1. Cepas de <i>Drosophila melanogaster</i>	109
7.1.2. Extracción de DNA	
7.1.2.1. DNA genómico de <i>Drosophila melanogaster</i>	110
7.1.2.2. DNA plasmídico de <i>Escherichia coli</i> (miniprep)	110
7.1.3. Extracción de RNA	110
7.1.4. PCR (Reacción en Cadena de la Polimerasa)	111
7.1.4.1. <i>Real Time</i> PCR (PCR a Tiempo Real)	112
7.1.4.2. RT-PCR	114
7.1.4.3. Cebadores utilizados	114
7.1.5. Clonación de DNA	115
7.1.6. Marcado no radioactivo de sondas	118
7.1.7. Técnica de <i>Southern Blot</i>	118
7.1.7.1. Sondas utilizadas	120
7.1.8. FISH: Hibridación <i>in situ</i> Fluorescente	122
7.1.9. Secuenciación de DNA	122
7.1.9.1. Generación y secuenciación de clones solapados obtenidos mediante la exonucleasa <i>Bal31</i>	123
7.1.10. Cultivo celular de la línea Schneider-2 (S2) de <i>Drosophila melanogaster</i>	
7.1.10.1. Transfección transitoria	124
7.1.10.2. Localización subcelular de proteínas de fusión con GFP	125
7.1.10.3. Vectores de transfección	125
7.2. Técnicas <i>in silico</i>	
7.2.1. Búsqueda y alineamiento de secuencias de DNA	127
7.2.2. Predicción de pautas de lectura	127
7.2.3. Generación e interpretación de logos de secuencia	128
7.2.4. Creación de un programa simple de búsqueda automatizada.....	128
8. BIBLIOGRAFÍA.....	133
9. ANEXOS.....	143

3. INTRODUCCIÓN

3.1. ANTECEDENTES

El presente trabajo de tesis doctoral es consecuencia de los estudios de la Dra. Eva Baldrich sobre la movilización del elemento transponible *copia*, en una cepa w^a , mediante tratamientos con ENU, EMS o aplicación de choque térmico (Baldrich, 1998; Baldrich *et al.*, 2003). En sus experimentos, la Dra. Baldrich obtuvo un gran número de cepas mutantes con ojos color casi blanco o marrón, cuyo genotipo no se correspondía con una reversión total o parcial. Un macho con un fenotipo de ojos amarillos dio lugar a una de estas cepas, bautizada M115. Es una cepa inestable en la que aparecen regularmente machos de fenotipo salvaje. A partir de uno de estos individuos se estableció la cepa RM115, llamada revertiente, aunque técnicamente su fenotipo no revierte al original w^a , sino al fenotipo salvaje. Los primeros análisis de M115 y RM115 revelaron que *copia* no estaba presente en el gen *white*, pero detectaron lo que parecieron reordenaciones en la región 3', entre 3 y 5Kb detrás del gen (*downstream*) (ver Fig. 3.1).

En este punto, tomamos el reto de indagar y explicar el fenotipo de ambas cepas mutantes, con mayor precisión.

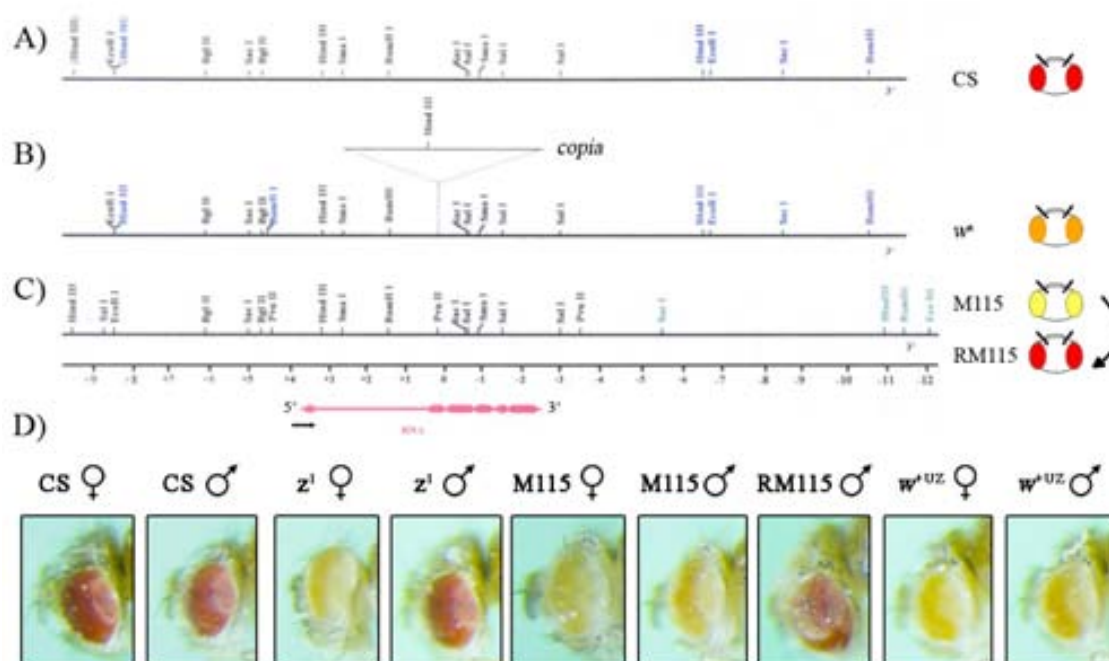


Fig.3.1. Los resultados de la Dra. Baldrich sobre las cepas M115 y RM115. Mapas de restricción de las distintas cepas usadas por la Dra. Eva Baldrich. **A)** Cepa de referencia CS. **B)** Cepa w^a , origen de los experimentos con ENU, EMS y choque térmico. La inserción del elemento transponible *copia* da lugar al origen de las coordenadas de Levis. **C)** M115 y RM115, pese a no mostrar diferencias entre ellas, difieren considerablemente con los controles en la región 3' del gen *white*. La escala muestra las coordenadas de Levis. En rojo se muestra el esquema del transcrito primario de *white* (Baldrich, 1998). **D)** Fotografías de cabezas de las cepas de estudio y de las de referencia, usadas en el trabajo.

En 1996, la Dra. Asa Rasmuson-Lestander publicó el análisis molecular de las cepas mutantes w^{+UZ} , con machos de ojos amarillos, y w^{+UR} , con machos de ojos rojos (llamadas *unstable zeste*, UZ y *unstable red*, UR, respectivamente) (Rasmuson-Lestander y Ekstrom, 1996)(ver Fig. 3.1D). El fenotipo de estas cepas es notablemente similar al de nuestras M115 y RM115, con la diferencia de que en este caso la cepa de ojos amarillos (UZ), deriva de la de ojos rojos (UR). Rasmuson-Lestander determinó que sus cepas eran portadoras del alelo mutante *zeste*¹, que se observa fenotípicamente como hembras de ojos amarillos debido al apareamiento de copias del gen *white*. También encontró una reordenación en la región 3' de *white*, que identificó como la inserción de un elemento transponible de la familia *foldback*. La inserción no es igual en las dos cepas, pues UR parece contener solamente secuencias *FB*, mientras que UZ cuenta con un elemento *FB-NOF* completo (ver Fig. 3.2). En su opinión, la inserción del elemento completo en UZ es la causa del fenotipo *zeste*¹ ampliado a los machos, de modo que Rasmuson-Lestander propuso que las secuencias *NOF* podrían modificar, de alguna forma desconocida, la actuación de la proteína mutante *Zeste*¹ sobre la expresión de *white*.

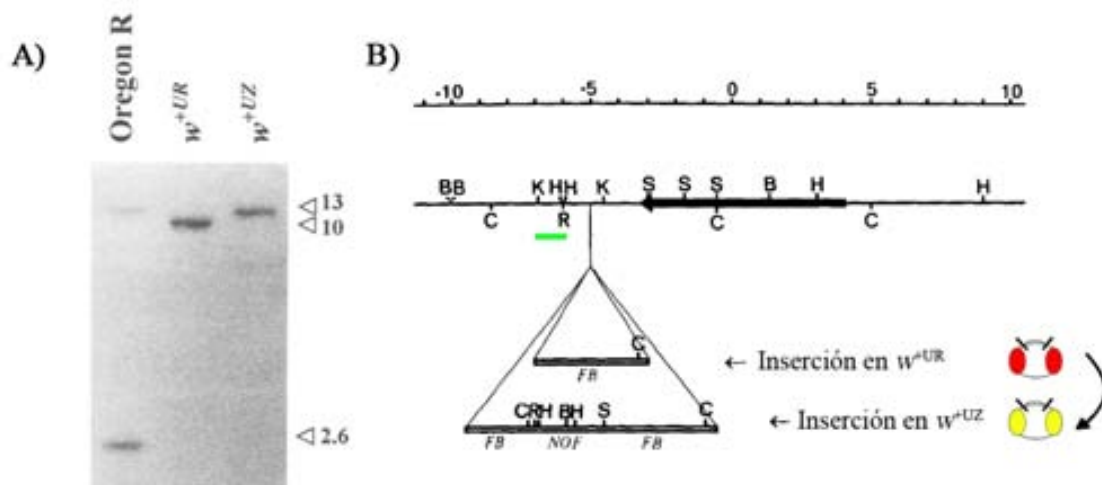


Fig.3.2. Descripción de w^{+UZ} , según Rasmuson-Lestander. A) *Southern blot* sobre las cepas Oregon R (referencia), w^{+UR} y w^{+UZ} , digeridas con *KpnI*, hibridado con la sonda pRK1.5 (barra verde en B). Se observa el tamaño de la inserción. B) Esquema de la inserción en w^{+UR} y w^{+UZ} . La flecha negra simboliza el gen *white* y su sentido de transcripción (nótese que la región 3' se encuentra a la izquierda de la imagen). En la parte superior se muestran las coordenadas de Levis. B:*Bam*HI, C:*Sac*I, H:*Hind*III, K:*Kpn*I, R:*Eco*RI, S:*Sal*I. (Rasmuson-Lestander y Ekstrom, 1996).

Tomando como modelo las cepas de Rasmuson-Lestander, pudimos encontrar y caracterizar inserciones similares, pero de tamaños distintos, en nuestras cepas M115 y RM115. Sin embargo, nuestros resultados pronto entraron en conflicto con las

conclusiones del grupo sueco, revelando reordenaciones más complejas de lo esperado. Nuestros experimentos, forzosamente diseñados a capricho de los resultados, nos ofrecen una nueva visión de la cepa UZ, utilizada para ensayos de mutagénesis, descubren las bases del fenotipo de las cepas M115 y RM115 y revelan nuevas características del elemento transponible *FB-NOF*, uno de los más desconocidos del genoma de *D. melanogaster*.

3.2. LOS ELEMENTOS TRANSPONIBLES

Los elementos transponibles (ETs) son entidades genéticas capaces de movilizarse dentro del genoma de su huésped. Sin duda alguna, los elementos transponibles parasitan los genomas donde se hospedan, pues todos ellos aprovechan la maquinaria celular para construir sus propias proteínas. Sin embargo, su actividad no es siempre dañina para la célula/organismo/especie y, como se verá más adelante, muchos son los ejemplos en los que la evolución ha favorecido los organismos que han podido aprovechar las habilidades de los elementos transponibles para sus propias necesidades.

El descubrimiento y redescubrimiento

Los elementos transponibles fueron descubiertos por la Dra. Barbara McClintock (1902-1992) (Fig.3.3) en sus investigaciones sobre el material genético del maíz (*Zea mays*) (McClintock, 1950). Durante su estudio de la relación entre diferencias de pigmentación del grano e inestabilidad génica, identificó lo que llamó elementos reguladores, que se insertaban en puntos del genoma y podían reprimir la expresión de los genes adyacentes. La escisión espontánea de esos elementos resultaba en la reactivación de los genes afectados. Su primera publicación al respecto es del 8 de Abril de 1950 y fue largamente ignorada por la comunidad científica por suponer un punto de vista radicalmente nuevo para la genética de la época. Sin embargo, en los años 70, el descubrimiento de las secuencias de inserción (IS) en bacterias y en *Drosophila melanogaster* (Shapiro y Adhya, 1969; Golubovsky *et al.*, 1977; respectivamente) significó el redescubrimiento de las hipótesis de McClintock, por lo que fue galardonada con el premio Nobel de Medicina en 1983.



Fig.3.3. Barbara McClintock

La llamada explosión de la biología molecular en los años 70 y 80, también se reflejó en el estudio de los ETs. Parece lógico, pues su pequeño tamaño y gran abundancia permitía una buena clonación y secuenciación, algo inusual en los inicios de la genómica. Quizás el punto culminante de la “fiebre de los ETs” de los años 80 fue la publicación, en 1989, de la compilación de David J. Finnegan titulada “*Eukaryotic transposable elements and genome evolution*” (Finnegan, 1989). En ella, el autor presenta la que fuera la primera clasificación formal de los ETs, basándose en su mecanismo de transposición (ver Fig.3.4).

En general, se admiten dos grandes clases de ETs:

Los elementos de clase-I: son aquellos que usan un intermediario de RNA en su proceso de transposición. Se dividen en dos subclases, los retrotransposones y los retroposones.

Los elementos de clase-II: se movilizan mediante un sistema de escisión y reinserción.

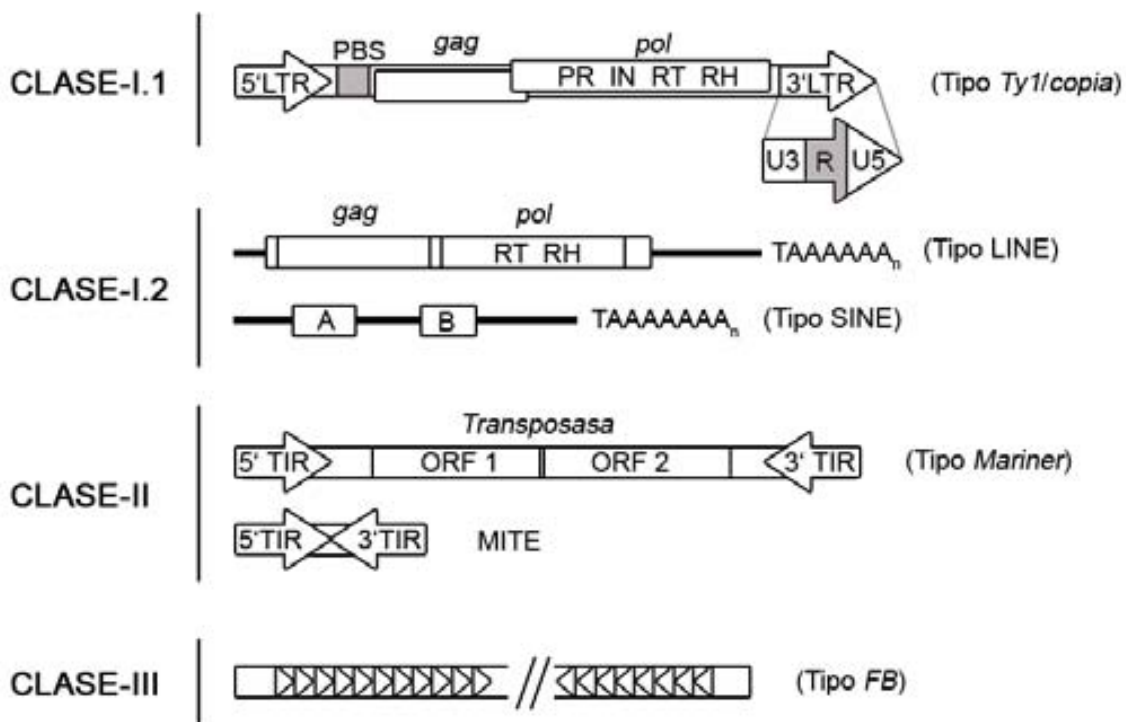


Fig.3.4. Clases de ETs. Representaciones esquemáticas de las distintas clases de elementos transponibles. Se muestran los componentes estructurales más representativos de cada clase. **LTR:** *long terminal repeat* **PBS:** *primer binding site* **PR:** proteasa **IN:** integrasa **RT:** retrotranscriptasa **RH:** RNAasaH **TIR:** *terminal inverted repeat*

Los elementos de Clase-I.1 se parecen a los retrovirus

La clase-I.1 de ETs está formada por aquellos elementos que cuentan con largas repeticiones terminales (LTR, del inglés *long terminal repeats*). Estas LTR son idénticas y su orientación es directa la una respecto a la otra. En ellas se encuentran elementos reguladores de la expresión del elemento, como el promotor de transcripción y la señal de poliadenilación. Los ORF tienen un gran parecido a los de retrovirus *gag*, *pol* y, algunas veces, incluso se puede encontrar un homólogo de *env*, como en los ETs *Tom* y *gypsy* de *D. melanogaster* (Arkhipova *et al.*, 1995). De hecho, actualmente *gypsy* es considerado un retrovirus.

El gen *gag* codifica una poliproteína que es procesada en tres proteínas distintas: las de la matriz, la cápside y la nucleocápside, todas ellas con gran parecido a sus homólogos

víricos. La proteína de la nucleocápside es la encargada de unir el RNA del retrotransposón, mediante un dominio dedos de zinc, al interior de la cápside pseudovírica.

El gen *pol* codifica todas las proteínas necesarias para la transposición del elemento: proteasa, transcriptasa reversa, RNAasaH e integrasa. Al igual que *gag*, el gen *pol* codifica un mRNA policistrónico que se traduce como una poliproteína y es procesado por la proteasa madura para dar lugar a las distintas proteínas funcionales.

El mecanismo general de transposición de los retrotransposones empieza con la síntesis del intermediario de RNA, iniciándose en el extremo 5' de la región R del LTR 5' y terminando en el extremo 3' de la región R del LTR 3'. Este intermediario es empaquetado en forma de partículas parecidas a virus (VLPs, del inglés *virus-like particles*) mediante el concurso de las proteínas *gag*. La transcriptasa reversa, codificada en uno de los cistrones de *pol*, usa un tRNA unido a la región PBS (de *primer binding site*) como cebador para iniciar la síntesis de la primera cadena de cDNA. Una serie ordenada de eventos se encargan de conseguir la molécula de dsDNA que será integrada en un nuevo punto del genoma, terminando así con la transposición del retrotransposón (Finnegan, 1992; Arkhipova *et al.*, 1995; Capy *et al.*, 1998).

Este mecanismo es llamado replicativo, pues el elemento *original* no se elimina de su posición. Así pues, la actividad de los retrotransposones es, *de facto*, la colonización del genoma huésped.

Algunos ejemplos de ETs de clase-I.1 son *copia* y *Tom* en *D. melanogaster*, *Ty* en *Saccharomyces cerevisiae*, *IAP* y *UL30* en rata y ratón, *THE* en humanos o *BS1* en maíz (Finnegan, 1989).

Los elementos de clase-I.2 parecen mRNA

La segunda clase de los ETs que transponen mediante intermediario de RNA son los llamados *retroposones*. La principal diferencia con los de clase-I.1 es la ausencia de LTR, y su principal característica es una región rica en adenina, en la región terminal 3', de gran parecido con la cola poli(A)⁺ de un mRNA. Aunque no existan regiones LTR, la inserción del retroposón genera duplicaciones directas de 6-15bp en el genoma huésped. Existen dos clases de retroposones, los de tipo LINE (LINE-*like*) y los de tipo SINE (SINE-*like*). La unidad transcripcional del tipo LINE-*like* está formada por dos ORF muy parecidos a *gag* y *pol* de retrovirus y retrotransposones, aunque normalmente no están solapados y parecen codificar menos funciones. La transcripción del intermediario empieza en las primeras bases de la mencionada unidad transcripcional del elemento, en la región 5'UTR, a partir

de promotores internos para la RNA polimerasa II. El RNA resultante es empaquetado en forma de ribonucleoproteína (RNP) junto con las proteínas *gag* y *pol*, formando VLPs. La RNP viaja de nuevo al núcleo, y el RNA del retroposón es retrotranscrito a dsDNA gracias a la actividad retrotranscriptasa de la proteína *pol*. La retrotranscripción se efectúa desde la cola poli(A)⁺ terminal, lo cual conlleva el riesgo de no conseguir llegar hasta el inicio del elemento. Este riesgo es real, y se pueden encontrar gran cantidad de copias truncadas en el extremo 5'. Estos no pueden movilizarse de forma autónoma, pues carecen de los promotores de inicio de transcripción e, incluso, de parte del ORF *gag*. Sin embargo, sí pueden ser movilizados en *trans* por las proteínas de otros elementos completos, si se consigue la transcripción a partir de algún promotor cercano propio del huésped (Capy *et al.*, 1998).

Los elementos SINE-like son más bien cortos, de alrededor de 500bp. Tienen una región de gran similitud con los genes de tRNAs (excepto la familia *Alu* de primates y *BI* de roedores, que provienen del 7SL RNA), una segunda región no relacionada con tRNA y una cola poli(A)⁺. La zona derivada del tRNA contiene un promotor de entrada de la RNA polIII con dos cajas (A y B) bien conservadas. En estos elementos no se aprecia ninguna región codificante, pero se ha demostrado que en mamíferos, salmónidos, tortugas y algunas plantas, el extremo 3' se corresponde con el 3' de elementos LINE-like, de modo que es probable que tomen prestadas sus enzimas para movilizarse (Capy *et al.*, 1998).

Son retroposones los elementos *LINE-1* y *SINE* humanos, los *I*, *F*, *G* y *jockey* de *D. melanogaster* o los *CRE* de *Trypanosoma brucei*.

Los elementos de clase-II saltan literalmente

La clase-II de ETs, según Finnegan, son aquellos que no usan un intermediario durante su movilización, sino que el mismo elemento *original* se escinde de su posición en el genoma, mediante unas enzimas endonucleasa propias, y se reinserta en otra posición. Estructuralmente, su principal característica son las repeticiones terminales invertidas (TIR o IR) presentes en cada extremo del elemento, e idénticas entre sí. Su región interna normalmente codifica una o dos proteínas: la endonucleasa que cataliza la escisión/inserción, llamada transposasa, y una enzima resolvasa que ayuda durante el proceso, aunque no todos los elementos de clase-II la poseen.

Puesto que el transposón se escinde y se vuelve a insertar en el genoma, el mecanismo de transposición de los elementos de clase-II es llamado conservativo. De todos modos, hay formas mediante las cuales un elemento de esta clase puede amplificarse. Sucede, por

ejemplo, cuando un transposón se moviliza y el corte producido en la cadena de DNA durante su escisión es reparado usando la cromátide hermana o bien el cromosoma homólogo como molde. Dicha reparación supondría el restablecimiento de una copia del transposón en la misma posición donde estaba, y al mismo tiempo la generación de una nueva inserción resultado de la movilización.

Las repeticiones invertidas de los elementos de clase-II son muy recombinogénicas. Esto es peligroso para el huésped, pero también para el transposón ya que la recombinación de los dos TIRs puede “vaciar” el elemento. En los años 90 se identificó una nueva familia de transposones que constaban, solamente, de repeticiones terminales invertidas: los MITEs (del inglés *miniature inverted-repeat transposable elements*) (para una revisión, ver Capy *et al.*, 1998). En principio parecían versiones en miniatura de elementos tipo *foldback*, pero finalmente se descubrió que esos TIRs compartían gran similitud con los de elementos ya conocidos de clase-II, como ocurre con las superfamilias de MITEs *Stowaway* y *Tourist-like*, muy extendidas en el reino vegetal (y la segunda también en el reino animal). Para la familia *Stowaway*, por ejemplo, se ha demostrado que sus TIRs pueden ser reconocidos por transposasas de la familia *mariner-like* del arroz (*Osmars*, una contracción de *oryza sativa mariners*), lo que constituye la primera evidencia de comunicación entre distintas familias de transposones de clase-II (Feschotte *et al.*, 2005).

Son elementos transponibles de clase-II: *P* y *Mariner* de *D. melanogaster*, *TCl* de *Caenorabditis elegans* y *Ac* y *Spm* de *Zea mays*.

¿Existe la clase-III de elementos transponibles?

Algunos transposones no acaban de encajar en la clasificación de Finnegan, pues su mecanismo de transposición es, hoy en día, desconocido. Probablemente la mayor clase dentro de este limbo de los transposones sea la de los llamados *foldback transposons*. Se caracterizan por la presencia de unos TIR inusualmente grandes, de hasta dos kilobases, formados por diversos grados de amplificación de repeticiones directas de pocas decenas de pares de bases. Muchos de ellos no parecen contener ninguna región codificante propia entre estos TIR, aunque a menudo se pueden encontrar fragmentos de DNA genómico del huésped o incluso otros transposones en su interior (Paro *et al.*, 1983; Brierley y Potter, 1985, respectivamente).

Ejemplos de este grupo de ETs son *foldback (FB)* de *D. melanogaster*, *galileo* de *D. buzatii*, *SoFT1* en el tomate (*Lycopersicon esculentum*), *FARE* en *Arabidopsis thaliana* y *Tnr8* en el arroz africano (*Oryza glaberrima*).

(Casi) Todos los organismos tienen ETs

La era de la genómica nos permite analizar la composición de los genomas hasta un punto inconcebible en los primeros tiempos del estudio de los ETs. Quizás, una de las grandes sorpresas de este análisis es la constatación de que prácticamente ninguna especie escapa a la invasión de un tipo u otro de ET. Hasta la fecha, sólo están libres de ETs la cepa secuenciada de *Bacillus subtilis*, sin rastro de las secuencias de inserción (*IS-like*) tan frecuentes en procariotas, y el parásito *Plasmodium falciparum* (Hua-Van *et al.*, 2005).

Sin embargo, el grado de colonización, entendido como el porcentaje del genoma ocupado por ETs, no es el mismo en todas las especies en las que se ha estudiado. Existe una relación lineal entre el tamaño del genoma huésped y la cantidad relativa de transposones, de modo que a mayor genoma, mayor colonización. No sólo aumenta el número neto de ETs, sino que las especies con genomas mayores también muestran un mayor número de familias distintas. Hua-Van y colaboradores recopilaron la información disponible en el primer trimestre de 2004, en la tabla que se muestra a continuación (Tabla 3.1). Puesto que los datos son para las porciones eucromáticas de los genomas, en algún caso la tendencia no se mantiene.

Especie	Tamaño del genoma Mb	Nº copias de ETs	% del genoma	Nº total de familias	Clase-I			Clase-II	Referencia
					LTRs	LINEs	SINEs	DNA	
<i>Saccharomyces cerevisiae</i>	12	331	3.1	5	5	0	0	0	Kim <i>et al.</i> (1998)
<i>Caenorhabditis elegans</i>	97	3718	6.5	34	19	3	5	7	Duret <i>et al.</i> (2000), The <i>C. elegans</i> Sequencing Consortium (1998) Ganko <i>et al.</i> (2001)
<i>Drosophila melanogaster</i>	137	1572	3.1	~100	49	27	0	19	Kaminker <i>et al.</i> (2002)
<i>Anopheles gambiae</i>	278	?	16?	~80	8	2	1	0	Holt <i>et al.</i> (2002)
<i>Mus musculus</i>	2700	2.9 × 10 ⁸	37.9	?	10%	19%	8%	0.9%	Waterston <i>et al.</i> (2002)
<i>Homo sapiens sapiens</i>	2900	3.16 × 10 ⁸	45	263 ^a	8%	21%	13%	3%	Lander <i>et al.</i> (2001)
<i>Arabidopsis thaliana</i>	125	5500	14	160	70 ^b	10 ^b		80 ^b	The Arabidopsis Genome Initiative (2000) ^a Kidwell (2002)

Tabla 3.1. Abundancia de ETs en genomas de distintas especies estudiadas. Estos resultados no tienen en cuenta los elementos MITE ni las regiones heterocromáticas. Se considera que el porcentaje total de ETs en el genoma de *D. melanogaster* ronda el 10%. Adaptado de Hua-Van *et al.*, 2005

Por clases, los elementos de clase-I predominan en la mayoría de los genomas estudiados, excepto en *A. thaliana*, donde existe un empate técnico debido a la presencia de un gran número de elementos de clase-II. En realidad, si contásemos los MITEs como elementos transponibles de clase-II, estos serían los más abundantes en *Oryza sativa*

(70%), *A. thaliana* (50%) y parece que también en *C. elegans*, aunque no hay datos definitivos.

Existe un patrón general en la distribución de ETs

Se asume que la movilización de la mayoría de ETs es un fenómeno al azar. Sin embargo, los análisis de distribución de ETs en genomas muestran que los transposones, sin importar la clase o la familia a la que pertenezcan, se encuentran mayoritariamente en las regiones heterocromáticas de los cromosomas, en particular en los centrómeros y los telómeros. Esta tendencia es excepcionalmente clara en dípteros, pues hasta el 50% de la heterocromatina de *D. melanogaster* y el 60% de la de *Anopheles gambiae* podría estar constituida por ETs (Holt *et al.*, 2002; Hoskins *et al.*, 2002).

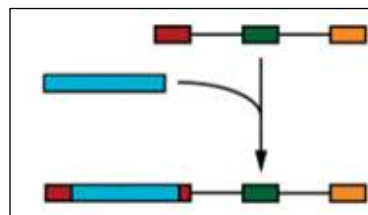
No está muy claro el porqué de esta coincidencia, pero hay consenso sobre el hecho que la selección natural podría jugar un papel importante. Existen tres vías por las que los ETs podrían acumularse en heterocromatina. La primera, se basa en la capacidad de los transposones para interrumpir genes. La selección natural actuaría contra estos fenómenos, restringiendo las inserciones a lugares pobres en genes, como las regiones heterocromáticas. La segunda, asume que las proteínas codificadas por los ETs son perjudiciales para los organismos, en tanto que modifican el DNA, de modo que transposones en regiones activas del genoma estarían bajo la estricta criba de la selección natural. En cambio, las inserciones en regiones silentes, como las heterocromáticas, pasarían fácilmente inadvertidas. La tercera, observa que la elevada frecuencia de recombinación de los ETs es un peligro para la estabilidad de los genomas huéspedes, por lo cual las inserciones en regiones con una tasa elevada de recombinación estarían severamente restringidas. Justamente, las regiones heterocromáticas tienen una baja tasa de recombinación.

A falta de estudios exhaustivos, parece que la distribución preferente de los ETs en las regiones heterocromáticas de los genomas huésped es el reflejo de la interacción compleja entre las tres vías antes mencionadas bajo el efecto de la selección natural.

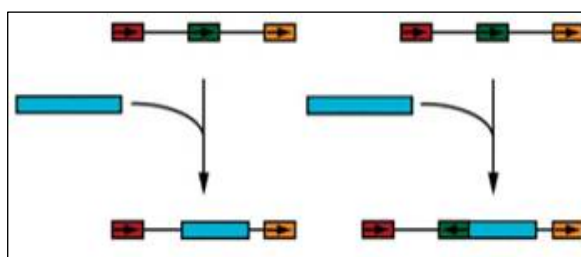
Los transposones pueden alterar gravemente los genomas huésped

La caprichosa movilización de los ETs en el genoma compromete el funcionamiento celular de múltiples formas. A continuación se describen las más conocidas, con algunos ejemplos asociados (Los esquemas son adaptaciones de Han y Boeke, 2005):

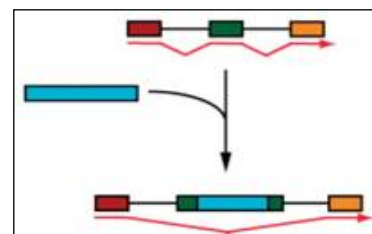
- **Disrupción de genes.** Las inserciones en exones o en regiones promotoras son las formas más evidentes de mutación ocasionadas por ETs. Ocurre, por ejemplo, en el mutante w^1 de *D. melanogaster*, debido a la inserción de un elemento *Doc* de 4,9Kb cerca del punto de inicio de transcripción del gen *white* (Driver *et al.*, 1989).



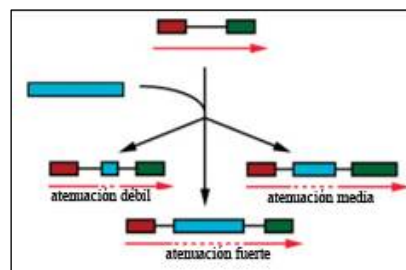
- **Delección o Reordenamiento.** La inserción de un ET puede significar la pérdida o inversión de la región genómica previa, durante la resolución de los productos intermedios de la inserción. En los trabajos de Gilbert y colaboradores, analizaron 37 inserciones distintas del retroelemento LINE-1 humano y detectaron un caso de delección de más de 71Kb resultado de la movilización (Gilbert *et al.*, 2002). Más tarde, en siguientes experimentos, el análisis de 100 inserciones reveló 16 nuevos casos de delección o inversión (Gilbert *et al.*, 2002)



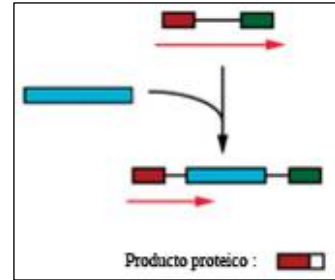
- **Alteración del procesamiento del mensajero.** Un ET introducido en un intrón, puede modificar las formas de procesamiento alternativo de un gen (*alternative splicing*) sirviendo de donador o aceptor de puntos de *splicing* o bien enmascarando exones. En un trabajo publicado en 1993, se demuestra que la inserción de un elemento L1, causó la pérdida del exón 44 en el mensajero del gen de la distrofina en un paciente de distrofia muscular de Duchenne (Narita *et al.*, 1993).



- **Atenuación de la expresión por exceso de elongación.** Los ETs pueden introducir varias kilobases de secuencia en un intrón, lo que puede suponer un descenso del número de transcritos primarios que consiguen finalizarse, atenuando la expresión del gen afectado. Este es un mecanismo que depende del tamaño del transposón, de modo que existen distintos grados de atenuación. Han definió este mecanismo como “reóstato molecular”, al describir el efecto de las secuencias del retroposón L1 insertadas en intrones (Han *et al.*, 2004).

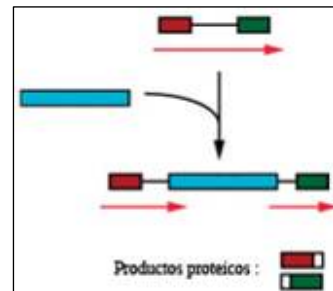


- **Poliadenilación prematura.** Los ETs completos contienen sus propias señales de poliadenilación. Aunque un gen con un ET insertado en uno de sus intrones consiga transcribirse correctamente, éstas pueden confundir la maquinaria de maduración del mRNA y cortar el transcrito prematuramente. Es el caso del alelo

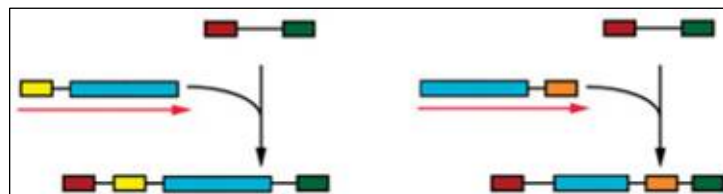


w^a del gen *white* de *D. melanogaster*, en el que la inserción de un retrotransposón *copia* en el segundo intrón resulta en la generación de transcritos truncados en la señal de poliadenilación del ET. La disminución de la cantidad de mensajeros completos de *white* da un fenotipo de ojos anaranjados.

- **Rotura de genes.** La inserción en un intrón puede romper un gen en dos, cuando ésta se produce en sentido inverso. Se generan dos mensajeros, uno con la primera parte del gen que termina en la señal de poliadenilación del ET; el segundo iniciándose en el promotor antisentido de un LTR o extremo de un retroposón, y continuando con el resto de exones. Recientemente, Wheelan y colaboradores presentaban un análisis de genes rotos por el elemento LINE-1 en humanos e identificaban 3 genes aparentemente “rotos” y 12 candidatos más (Wheelan *et al.*, 2005), algunos con interesantes perspectivas evolutivas.

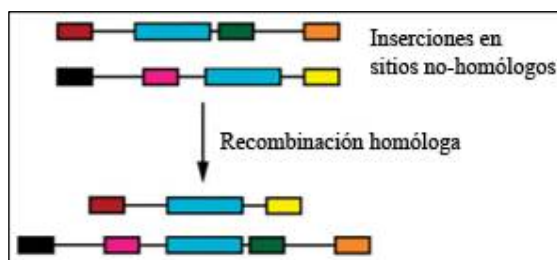


- **Transducciones (*exon shuffling*).** La escisión de los ETs puede ser poco precisa. A menudo arrastran parte del contexto genómico inicial cuando se movilizan, por ejemplo exones. La inserción de este material dentro de otro gen supone la adición de un nuevo exón y la creación de genes quiméricos. Un buen ejemplo es el caso descrito por Moschetti, en el que un elemento *FB* movilizó el extremo 3' del gen *white* dentro de los genes *outspread* (*osp*), *CG6487* y *CG3973*. En el trabajo, detectan y analizan estructuralmente los transcritos quiméricos (Moschetti *et al.*, 2004).

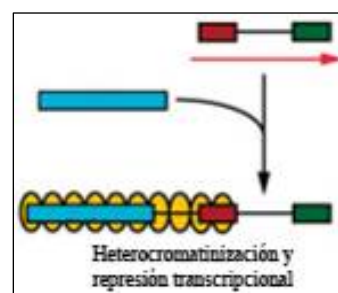


- **Recombinación homóloga no alélica.** Los ETs pueden producir grandes reordenaciones, estén dentro o fuera de genes. Puesto que existen diversas copias casi idénticas en el genoma, son un punto habitual de recombinación homóloga en sitios no

homólogos, generando translocaciones, inversiones, deleciones o roturas cromosómicas. Caso ejemplar, el descrito por Cáceres en 1999 en el cual múltiples copias del elemento *Galileo* (de tipo *foldback*) son las causantes de la inversión 2j de *D. buzzatii* (Caceres *et al.*, 1999).



- Regulación epigenética.** Un mecanismo ampliamente usado para el silenciamiento de transposones es la heterocromatinización. Sin embargo, si un ET se inserta cerca de un gen y es heterocromatinizado, la expresión del gen vecino puede resultar afectada. Quizás uno de los primeros ejemplos de represión transcripcional por heterocromatinización debida a un ET sea el caso descrito por Druker en 2004: la inserción del retrotransposón *IAP* en el sexto intrón del gen murino *Cabp* da lugar a ratones con diversos grados de expresión del gen debido a la naturaleza estocástica de la regulación epigenética en el promotor bidireccional presente en su LTR (Druker *et al.*, 2004).



Sin embargo, también pueden suponer algunas ventajas...

La misma existencia de los ETs acabó con la concepción estática de los genomas, pues éstos introducen una gran variabilidad genética, como se ha visto en el apartado anterior. De todos modos, es tanto el tiempo de convivencia que hoy en día no hay duda de que existe una coevolución de los genomas junto con los ETs que contienen. Es sorprendente ver hasta que punto la célula eucariota ha sido capaz de “domesticar” algunos transposones e integrarlos *de facto* en algunos de los procesos más íntimos de su propia biología. No faltan ejemplos de transposones involucrados en la esencia de la estructura cromosómica, cómo las repeticiones centroméricas en *A. thaliana*, *D. virilis* o *Z. mays* (Wong y Choo, 2004), los telómeros de *D. melanogaster*, formados exclusivamente por los retroposones *Het-A* y *TART* (Pardue y DeBaryshe, 2003), o el papel del aislador de *gypsy* en la creación de dominios funcionales de la cromatina y la unión a la matriz nuclear (Byrd y Corces, 2003).

Más aún, algunas especies han integrado componentes originales de transposones para su propio uso, más allá de funciones estructurales. Lander y colaboradores fijaron en 47 los genes humanos derivados de ETs hasta la fecha. Curiosamente, la mayor parte derivan de

antiguas proteínas de elementos de clase-II, aunque la clase-I sea mayoritaria en mamíferos (Para un listado exhaustivo ver Lander *et al.*, 2001). Entre estos genes destacan, por su trascendencia, RAG1 y RAG2 (del inglés *recombination-activating genes*) que codifican las proteínas encargadas de la recombinación de las regiones V(D)J de los receptores de antígeno de las células T del sistema inmune (TCRs). Ya en 1998 no cabía la menor duda de que tanto RAG1 como RAG2 son formas adaptadas de transposasa e, incluso, los trabajos de Agrawal y colaboradores demostraron que estas proteínas pueden movilizar fragmentos de DNA *in vitro*, de forma parecida al bacteriófago Tn10 (Agrawal *et al.*, 1998).

Los genomas tienen mecanismos de inactivación de ETs

Aunque algunos ETs hayan sido “domesticados” por los genomas donde se hospedan, su innegable potencial mutagénico ha resultado en la evolución de mecanismos de silenciamiento. Éste se lleva a cabo, básicamente, por dos métodos: Silenciamiento Transcripcional (TGS, del inglés *transcriptional gene silencing*) y Silenciamiento Post-Transcripcional (PTGS, del inglés *post-transcriptional gene silencing*).

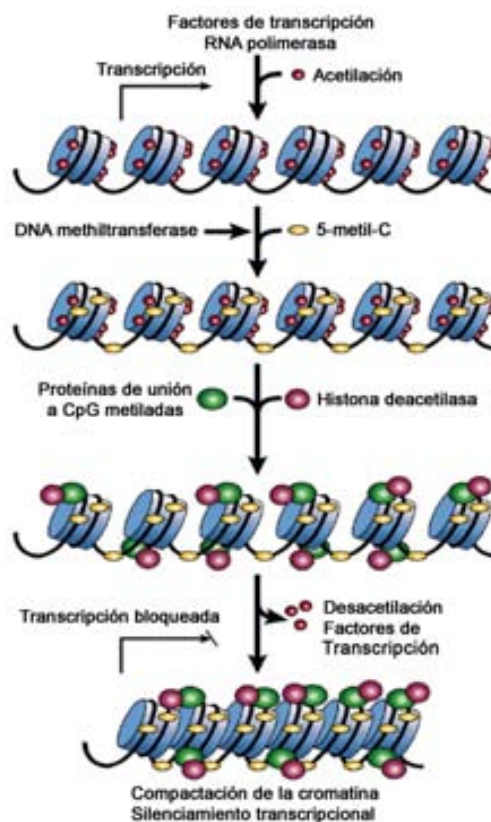


Fig.3.5. Silenciamiento Transcripcional. La metilación del DNA y la modificación de las colas de las histonas promueven el silenciamiento génico.

El TGS es uno de los mecanismos universales de silenciamiento génico. Se basa en la heterocromatinización de los genes a reprimir mediante la metilación de su secuencia. La metilación del DNA provoca la unión de la proteína MeCP2, que recluta un complejo multiproteico que contiene los desacetiladores de histonas HDAC1 y HDAC2. La desacetilación de las histonas H3 y H4 conlleva la formación de una estructura cerrada de la cromatina que impide la unión de factores de transcripción, de modo que la expresión génica queda silenciada (ver Fig.3.5) (Razin, 1998). La mayoría de las especies eucariotas usan el TGS para configurar patrones de expresión, reprimiendo genes específicos en algunos tejidos y activándolos en otros. Dicho silenciamiento puede ser transitorio o estable, pues el estado de silenciamiento conseguido mediante TGS puede transmitirse a las células hijas en cada paso de replicación. Se ha demostrado, en varias especies, la metilación y consiguiente silenciamiento de ETs como un método común de control de su actividad (Okamoto y Hirochika, 2001; Cheng *et al.*, 2006). Parece que las plantas usan TGS preferentemente para la inactivación de ETs.

Hasta hace relativamente poco tiempo, se creía que *Drosophila* no tenía sistemas de metilación del DNA. Sin embargo, Salzberg y colaboradores identificaron, en 2004, secuencias metiladas en DNA genómico de adultos, lo que dio pie a especular sobre la implicación del TGS en el control de ETs (Salzberg *et al.*, 2004). En 2006, Mandrioli y Borsati analizaron las secuencias metiladas encontradas por Salzberg y concluyeron que el 14% correspondían a transposones conocidos. Si el 15-20% del genoma de *Drosophila* está constituido por ETs y si su mecanismo de silenciamiento preferido fuese TGS, entonces la proporción de ETs entre las secuencias metiladas debería ser mayor (Mandrioli y Borsatti, 2006). Su conclusión final fue que *Drosophila* controla sus ETs mediante otro mecanismo: el silenciamiento post-transcripcional.

El PTGS es uno de los últimos grandes descubrimientos en el control de la expresión génica. Algunos autores lo denominan “el sistema inmune del genoma”, por analogía al sistema inmune de los vertebrados (Plasterk, 2002). El PTGS se basa en la formación de pequeñas moléculas de RNA (siRNA del inglés *small interfering RNA*) que, mediante la asociación con proteínas específicas, consiguen inhibir RNA mensajeros complementarios (ver Fig.3.6). Se ha demostrado su existencia en plantas y animales, y es una herramienta ampliamente usada en los laboratorios de investigación. Tal ha sido su impacto, que sus descubridores, Fire y Mello, fueron galardonados con el premio Nobel de Medicina del año 2006.

El mecanismo de PTGS se inicia con la presencia de una molécula de RNA de doble cadena (dsRNA). Ésta puede tener diversos orígenes: retrovirus, retrotransposones, el apareamiento de dos TIR de un ET de clase-II transcrito por extensión de un promotor cercano, o bien la transformación a doble cadena de un RNA aberrante de cadena simple, mediante una RNA polimerasa RNA dependiente (RdRP). En cualquier caso, el mecanismo más común es que la nucleasa Dicer, de la familia RNAasaIII, procese el dsRNA primario en pequeños fragmentos siRNA de 19-27nt. Estos se unen al complejo silenciador de RNA interferente (RISC, del inglés RNA interference silencing complex) y lo conducen al mRNA diana por complementariedad de secuencia. La proteína más relevante del complejo RISC es la Argonaute 2 (Ago-2), con un dominio PAZ de unión al extremo 3' de la molécula siRNA (PAZ también está presente en Dicer) y un dominio PIWI, estructuralmente muy parecido a la familia de nucleasas RNAasaH, que es el encargado de degradar el mRNA diana, silenciando la expresión del gen del que procede. Los complejos RISC pueden reutilizarse, amplificando la reacción de degradación de forma parecida a las cascadas del sistema inmune adaptativo de los vertebrados (ver las revisiones de Plasterk, 2002; Stram y Kuzntzova, 2006).

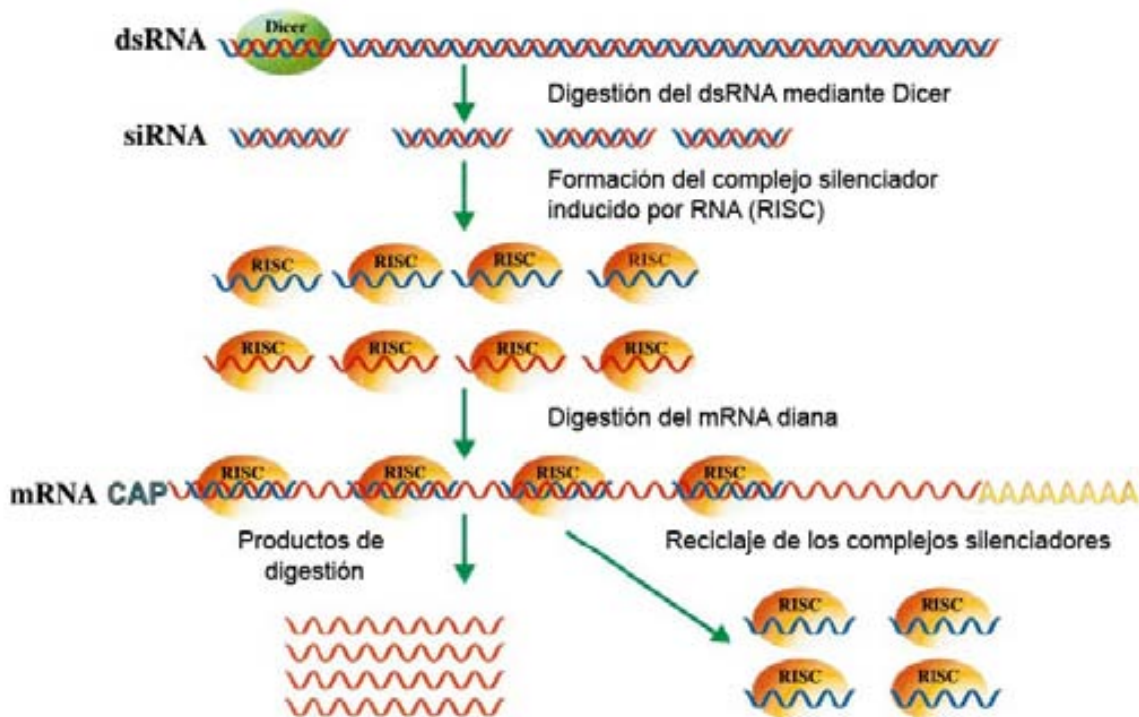


Fig. 3.6. Silenciamiento post-transcripcional (PTGS). Una molécula de dsRNA es procesada por Dicer en pequeños fragmentos de siRNA que, junto al complejo endonucleolítico RISC degradan selectivamente cualquier mRNA complementario. Adaptado de Stram y Kuzntzova, 2006.

El elemento *FB* de *D. melanogaster* es uno de los transposones más desconocidos

Una de las primeras evidencias de ETs con enormes repeticiones terminales invertidas es el trabajo de Potter y colaboradores en 1980, en el que buscaba nuevas familias de transposones en el genoma de *D. melanogaster* (ver B en figura 3.7). Describieron 5 fragmentos distintos de *foldback* DNA y, para algunos de ellos, demostraron que habían sido movilizados en el pasado. Una característica especialmente turbadora era la heterogeneidad de tamaño entre los TIR de las distintas secuencias, que variaba entre 0,19Kb y 1,31Kb, así como la caprichosa presencia o ausencia de secuencias de distinto tamaño en su interior (Potter *et al.*, 1980). En 1981, la continuación del análisis demostró que la estructura de lo que llamaron la familia *FB* de ETs era particularmente rara, pues los mapas de restricción mostraban que esos TIR tenían un patrón altamente repetitivo (ver C en figura 3.7). El estudio de la secuencia parcial del extremo de la familia Dm_FB3 llevó a la primera descripción de la estructura repetitiva del elemento *FB*, que se ha mantenido vigente hasta nuestros días (ver A en figura 3.7): una región inicial de unos 300bp, rica en repeticiones de una secuencia de 10bp, da lugar a una estructura ordenada en la que dicha secuencia se ve amplificada hasta los 31bp. Cinco de estas repeticiones imperfectas forman las grandes repeticiones de 153bp que se aprecian en los mapas de restricción (Truett *et al.*, 1981).

Aunque las secuencias de los dos TIR son muy parecidas, no son idénticas, pudiéndose encontrar gran cantidad de sustituciones. Además, el tamaño de los dos TIR de un elemento tampoco acostumbra a ser el mismo de modo que, en algunos casos, las micrografías electrónicas predecían una secuencia interna inexistente pues el lazo se debía al DNA sobrante del TIR más largo.

El elemento *FB*, entonces, debía describirse como un elemento puramente estructural, pues en ningún caso se encontraron secuencias codificantes propias ni en sus TIR ni en las regiones internas. La falta de cualquier maquinaria enzimática propia que catalizara su movilización es la principal causa de su asignación a la discutida clase-III.

El entusiasmo inicial en el estudio del elemento *FB* dio como resultado la identificación de distintos alelos debidos a la inserción de este nuevo ET, cómo w^c o w^{DZL} , así como su implicación en las grandes reordenaciones descritas por Ising a mediados de los años 70 (Ising y Block, 1984). Justamente, el estudio de uno de los grandes transposones de Ising, *TE28*, llevó a la identificación de una secuencia de 4,5Kb que

bautizaron como *NOF* debido a la similitud del lazo que formaba entre los TIR de *FB* en las micrografías, con el perfil clásico de los bustos de la reina egipcia Nefertiti (*Nofretete* en alemán) (Goldberg *et al.*, 1982).

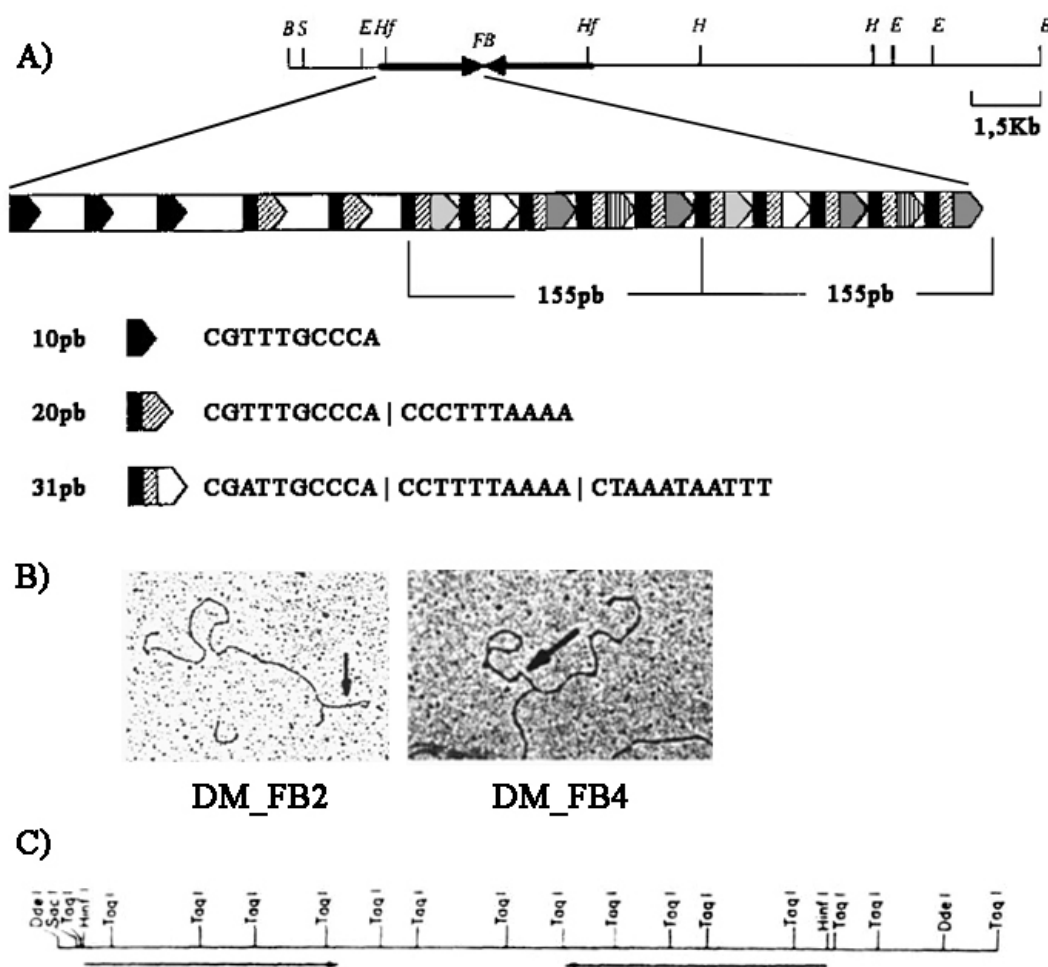


Fig.3.7. El elemento transponible *FB*. A) Esquema del elemento propuesto por Nicholas Harden y Michael Ashburner en 1990. La amplificación de una repetición inicial de 10bp constituye la base de los IR. B) Micrografías electrónicas de fragmentos de DNA con un elemento *FB*. La flecha indica el apareamiento de los IR. También se observa el lazo de secuencia no apareada. C) Mapa de restricción de un elemento *FB*. Se aprecia la estructura altamente repetitiva gracias al patrón de dianas *TaqI*. Adaptado de Truett *et al.*, 1981; Harden y Ashburner, 1990;

En la última década, con el incremento de datos genómicos de distintas especies y el aumento de la potencia de cálculo en los sistemas informáticos, se han descubierto elementos *foldback* en muchas otras especies: *galileo* (*D. buzatii*), *SoFT1* (tomate), *FARE* (*Arabidopsis thaliana*) o *Tnr8* (*Oryza glaberrima*). Sin embargo, aunque todos ellos son internamente repetitivos, ninguno alcanza el tamaño ni la estructura perfectamente ordenada de *FB*.

NOF* se encuentra íntimamente asociado a *FB

El estudio de la distribución de *NOF* en el genoma de *D. melanogaster* dio resultados sorprendentes, pues si bien se pueden encontrar alrededor de 5 elementos por genoma, nunca nadie encontró secuencias homólogas a *NOF* que no estuvieran dentro de un elemento *FB*. Esta asociación tan íntima disparó las especulaciones acerca del papel de *NOF* como transposasa de *FB* (Brierley y Potter, 1985) y, en consecuencia, se consideró a la asociación *FB-NOF* el elemento completo y a las secuencias *FB* versiones truncadas (Templeton y Potter, 1989; Harden y Ashburner, 1990). Como estimación general, se puede considerar que un 10% del total de secuencias *FB* son elementos *FB-NOF* completos.

La secuencia de *NOF* es muy parecida a la de un ET de clase-II. Está flanqueada por dos grandes TIR de 308bp, no repetitivos e idénticos entre sí, y contiene entre una y tres pautas posibles de lectura, según la secuencia que se analice: dos ORF en orientación directa (ORF1, 633aa y ORF2, 403aa) y uno solapado en reversa (ORF3, 182aa) en la secuencia de Templeton y Potter, o bien un gran ORF solitario (1056aa) en la secuencia de Harden y Ashburner (Templeton y Potter, 1989; Harden y Ashburner, 1990). La diferencia entre las dos secuencias se debe a una pequeña región de 5bp en la posición 2610 del elemento, justo al final del ORF1 de Templeton y Potter: TTTTTT pasa a 6bp, AAACAC, en Harden y Ashburner. Más allá del cambio de tres aminoácidos, la introducción de una sexta base produce un corrimiento de la pauta de lectura que evita un codón sin sentido *UGA* en el mensajero y eso, según Harden, representa la formación de un sólo ORF de gran tamaño. En las proximidades de los ORF1 y 2 de Templeton, así como en el ORF de Harden, se encuentran regiones que las predicciones informáticas indican como promotores y señales de poliadenilación. El pequeño ORF3 identificado por Templeton y Potter, sin embargo, además de encontrarse inusualmente en orientación reversa y solapado respecto a los otros dos ORF, no cuenta con dichas predicciones, con lo cual es poco probable que sea realmente funcional.

Hasta la fecha, se sabe muy poco sobre las posibilidades reales de que se expresen algunas de estas proteínas. El único estudio existente es el realizado por Templeton y Potter sobre su ORF1, del cual consiguieron generar un anticuerpo contra la región hidrofílica entre los residuos Gly26-Val35. Con dicho anticuerpo pudieron localizar la proteína en tejido adulto, en las cámaras de huevos (*egg chambers*). El tipo celular con mayor expresión parecía ser las células adyuvantes (*nurse cells*). Posteriores estudios de

Western Blot sobre extractos de proteínas de embriones de 6-18h identificaron una sola proteína de 71kDa, valor muy aproximado al predicho para el producto del ORF1.

La demostración de que, por lo menos, uno de los ORF de *NOF* es funcional, reforzó enormemente la opinión de que en *NOF* se encuentra la clave de la movilización del elemento completo *FB-NOF*.

3.3. EL GEN *white*

Si existe un decano entre los genes de *D. melanogaster*, este es el gen *white*. Descrito en 1910 como un factor hereditario que determinaba color blanco en los ojos, sirvió al célebre investigador Thomas Hunt Morgan (1866-1945) (Fig.3.8) para confirmar las leyes mendelianas de la herencia, sobre las cuales él mismo se mostraba escéptico. El patrón de herencia ligada al sexo del factor *white*, junto con los últimos descubrimientos de la época sobre las diferencias cromosómicas entre hembras (XX) y machos (XY) de *D. melanogaster*, permitieron que Morgan diera las primeras hipótesis sobre la localización de los factores hereditarios en los cromosomas (el término *gen* sería adoptado algo más tarde, después de ser introducido por Johanssen en 1909). Sus trabajos con *white* significaron el inicio de la genética como ciencia experimental.



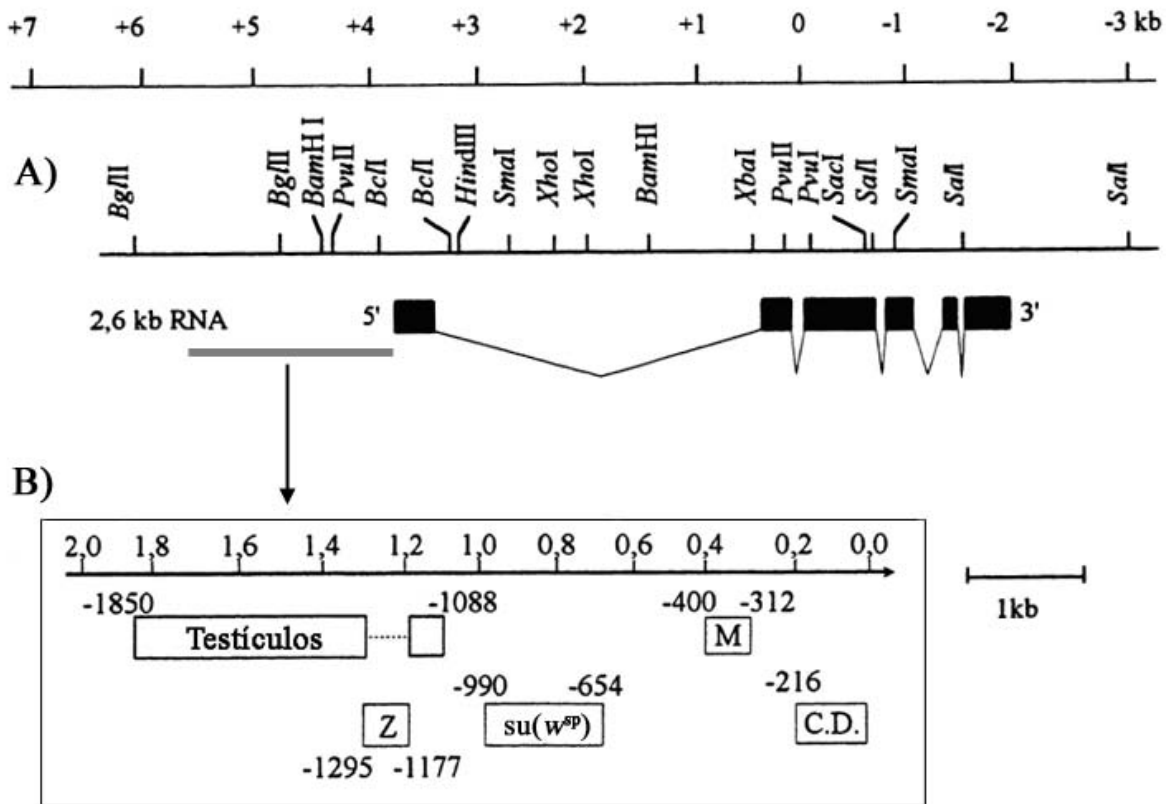
Fig.3.8. Thomas H. Morgan.

Fue galardonado con el premio Nobel de Medicina en 1933 por sus hallazgos sobre el papel de los cromosomas en la herencia.

Durante los años 70 se localizó en la posición 1-1.5 del cromosoma X mediante mapas genéticos. En los 80, el uso generalizado de técnicas moleculares, como la hibridación *in situ*, refinaron la posición de *white* en la banda 3C2 según el mapa de cromosomas politénicos. En 1982 Levis y colaboradores presentaron el primer mapa físico del gen: un mapa de restricción muy detallado de un clon de 48Kb obtenido a partir de una cepa con la mutación w^a (Levis *et al.*, 1982). En el mismo trabajo, restringieron a una región máxima de 14Kb todas las secuencias necesarias para la correcta expresión de un fenotipo salvaje w^+ y tomaron como origen de coordenadas la posición del retrotransposón *copia* en el alelo w^a , estableciendo lo que más tarde se conocería como las “coordenadas de Levis”. Los estudios funcionales sobre *white* identificaron un mRNA mayoritario de 2,7Kb con el que se acotaron los fragmentos de restricción correspondientes a las regiones exónicas del gen (O'Hare *et al.*, 1983). Sin embargo, la predicción más precisa de la estructura del gen (y de algunos de sus alelos mutantes) no se alcanzó hasta 1984, con la publicación de la secuencia completa de la región de 14Kb en la que se encuentra (O'Hare *et al.*, 1984).

El gen *white* consta de seis exones, el primero de los cuáles se encuentra separado del resto por un gran intrón de 2,9Kb (ver Fig.3.9). Las dos kilobases inmediatamente anteriores al inicio de transcripción contienen las principales regiones reguladoras de la

expresión génica de *white*: los potenciadores (*enhancers*) de expresión en testículos y en túbulos de Malpighi, el control de compensación de dosis y las zonas de interacción de las proteínas Zeste y Su(*w^{sp}*) (ver Fig.3.9) (Hazelrigg, 1987).



La expresión de *white* se da durante todo el desarrollo de *D. melanogaster*, aunque de

Fig.3.9. El gen *white*. A) Se muestra la región genómica del locus *white* con las principales dianas de enzimas de restricción. Debajo, representación esquemática del mRNA de 2,6Kb. Las cajas oscuras representan los exones. La escala superior, en kilobases, usa el sistema de coordenadas de Levis. B) Vista esquemática de las cajas potenciadoras de expresión específica de tejido y demás regiones de importancia para la correcta expresión de *white*. La escala superior, en kilobases, muestra la distancia al punto de inicio de transcripción. Z: punto de unión de Zeste. Su(*w^{sp}*): supresor de *white spotted*. M: túbulos de Malpighi. C.D.: compensación de dosis. Adaptado de Hazelrigg, 1987).

modo variable. Sin embargo, parece que su representación, en mRNA, es muy baja: según las estimaciones de Fjose y colaboradores, los transcritos de *white* constituyen entre el 0,0003% (basal) y el 0,003% (máximo) del total de RNA poli(A)⁺, tanto en embriones y larvas como en adultos (Fjose *et al.*, 1984). En nuestra opinión, estos datos deberían actualizarse, pues la metodología usada en 1984 para dicha cuantificación presenta algunas dudas.

Sea como fuere, los trabajos de Fjose demostraron que el gen *white* tiene dos picos principales de expresión. El primero, entre las 5-6 horas de la embriogénesis, se relaciona con la diferenciación de las células que formarán los túbulos de Malpighi. El segundo,

durante el tercer estadio larval, refleja el incremento de células de la porción ocular del disco imaginal óculo-antenal. La relación de los picos de expresión con los distintos tipos celulares se demostró mediante hibridaciones *in situ* sobre cortes histológicos. Estas hibridaciones demostraron, también, que *white* se expresa en una porción ventral de la larva de tercer estadio, muy cerca de la faringe, identificada como el órgano fotorreceptor larval. Aunque Fjose y colaboradores no consiguen detectarlo, del análisis de mutantes *w⁻* se desprende que también está implicado en la pigmentación de los conductos testiculares de los machos adultos.

En conjunto, pues, parece clara la relación del producto del gen *white* con algunos de los tejidos pigmentados de *Drosophila*. Durante años se especuló con la posibilidad de que la proteína White estuviera involucrada en la biosíntesis de los pigmentos omocromos (marrones) y pteridinas (rojos), mayoritarios en estos tejidos. Sin embargo, Tulle Hazelrigg, en 1987, mostraba la similitud de secuencia entre *white* y proteínas de transporte transmembranal ATP dependientes de *E.coli* (Hazelrigg, 1987). Tearle y colaboradores, en 1989, plantearon la posibilidad de que los genes *white*, *brown* y *scarlett* fueran las permeasas encargadas de la captación de los precursores de los pigmentos (Tearle *et al.*, 1989). En la actualidad, poco más se sabe de la función de la proteína White.

Esta carencia no ha impedido que el gen *white* sea, quizás, el más usado para el estudio de la genética de *Drosophila*. Su estudio, y el de sus mutantes, ha dado gran cantidad de información acerca de las consecuencias de la inserción de elementos transponibles, efectos de la posición en el genoma sobre la expresión, el fenómeno de la transvección, la compensación de dosis y la regulación específica de tejido mediante secuencias reguladoras en *cis*. La base de datos FlyBase (www.flybase.org), que centraliza la mayor parte de la información pública sobre distintas especies de *Drosophila*, tiene descritos 1587 alelos, 3153 referencias publicadas y 13305 fuentes de cepas disponibles. Muchas de éstas, fruto del uso de *white* como marcador de inserción de transgenes, después del desarrollo de los vectores pCasper, portadores de un *miniwhite*, que confiere un color anaranjado a moscas de fenotipo *w⁻*.

3.4- EL GEN *zeste*

La descripción de Gans, en 1949, de un fenotipo mutante caracterizado por el color amarillo claro de los ojos, de herencia recesiva ligada al sexo, fue la primera referencia pública, y la que le dio el nombre, sobre el gen *zeste* de *D. melanogaster* (*zeste* es la piel del limón, o de la naranja, en francés). Dicho fenotipo fue llamado *zeste*¹ (*z*¹) (Gans, 1949).

El gen *zeste* se encuentra en la posición 3A3 del cromosoma X, a una distancia genética de 1,5cM del gen *white*. Se puede considerar un gen compacto, pues su región codificante, de 2,6Kb, contiene sólo dos pequeños intrones de 121bp y 62bp (ver Fig.3.10). El transcrito de *zeste*, de 2,4Kb, es escaso: las primeras estimaciones indicaron que podía representar menos del 0,01% del total de RNA poli(A)⁺ presente en la célula (Pirrotta *et al.*, 1987; Mansukhani *et al.*, 1988). Sin embargo, no se han efectuado estudios más detallados en éste sentido.

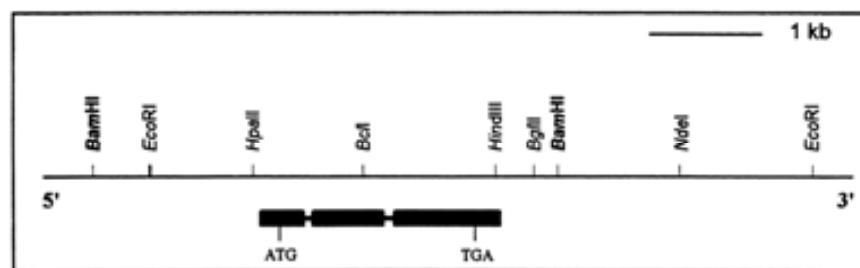


Fig.3.10. Esquema del gen *zeste*. Las cajas representan los tres exones de *zeste*. Las barras que las unen simbolizan los dos pequeños intrones. Se muestran los puntos de inicio y finalización de traducción. La barra superior simboliza el DNA genómico, con las dianas para enzimas de restricción más representativas.

La proteína codificada por *zeste* cuenta con 574aa y un peso molecular de 80kDa, aproximadamente. Se puede separar en tres grandes regiones funcionales (ver Fig.3.11):

La primera, en el extremo N-terminal, es una estructura en forma de hélice-giro-hélice (HTH o *helix-turn-helix*, en inglés) que permite la unión al DNA sin especificidad de secuencia. Una variación del HTH es el homeodominio, presente en varias proteínas relacionadas con la formación del patrón de segmentación.

La segunda, en el extremo C-terminal, es una zona hidrofóbica que se estructura en forma de cremallera de leucina de 74aa, que interviene en la interacción con otras unidades de Zeste (centenares, según Chen *et al.*, 1992), u otras proteínas, para formar agregados. El domino en cremallera de leucina también se encarga del reconocimiento de la secuencia de

DNA específica de sus elementos de unión, los ZBS (del inglés *Zeste binding site*), cuya secuencia consenso es YGAGYG.

La tercera, en la zona central, contiene tres dominios *opa*, ricos en residuos de glutamina y alanina, también presentes en otros genes reguladores, como *ultrabitorax*, *antennapedia*, *dorsal*, *hairy*, *deformed*, *kruppel*, *hunchback*, *rough* o *E74* (Chen *et al.*, 1992). Una región adyacente, rica en prolina, parece importante para la correcta interacción proteína-proteína de Zeste, pues sustituciones hidrofóbicas en esta región crean mutantes hiperagregantes como z^1 o z^{op6} (Rosen *et al.*, 1998).



Fig.3.11. Esquema de la proteína Zeste. Se muestran las principales zonas funcionales de Zeste y sus dominios asociados. Debajo se muestran las posiciones de las mutaciones z^1 y z^{op6} . DNA: dominio HTH de interacción al DNA. P: región rica en prolina. H: zonas hidrofóbicas que configuran el dominio de cremallera de leucina.

La proteína Zeste forma parte del grupo *trithorax* (trxG, en adelante), un vasto conjunto de proteínas encargadas de mantener el patrón transcripcional activo de los genes homeóticos. El trxG antagoniza, funcionalmente, el grupo *polycomb* (PcG, en adelante), encargado del mantenimiento del estado transcripcional inactivo de dichos genes. El desarrollo correcto del embrión de *Drosophila* necesita un patrón de expresión homeótico muy estricto, espacial y temporalmente, de modo que el balance correcto entre los grupos trxG y PcG es imprescindible.

Zeste forma agregados en dos fases. En primera instancia, el contacto entre cremalleras de leucina forma homodímeros. Luego, la región rica en prolina anterior a la cremallera, moderadamente hidrofóbica, agrupa los dímeros para formar agregados multiméricos. El proceso de agregación es muy ordenado y mutaciones en la cremallera de leucina que impiden la dimerización, también impiden la formación de agregados. Sin embargo, las mutaciones que impiden la multimerización, no necesariamente impiden la dimerización (Chen y Pirrotta, 1993).

En el año 2000, experimentos *in vitro* demostraron que Zeste, Brahma (BRM) y ATP eran necesarios para iniciar la transcripción a partir de los ZBS (Kal *et al.*, 2000). La formación de agregados de Zeste en el promotor, sirve de ancla para el complejo remodelador de la estructura de la cromatina BRM, un grupo multiproteico perteneciente,

también, al *trxG*. BRM, de la familia SWI/SNF, modifica el posicionamiento de los nucleosomas para conseguir una conformación abierta de la cromatina que permita la unión de factores de transcripción, o incluso de la misma maquinaria de transcripción, en la región promotora. El reclutamiento del complejo BRM por parte de Zeste depende de la interacción de su cremallera de leucina con la proteína Moira (MOR), principalmente, componente central del grupo BRM (ver Fig.3.12). Sin embargo, la falta de actividad de una fusión entre la región de unión al DNA y la cremallera de leucina, demostró que la región central de Zeste es necesaria para el reclutamiento del complejo BRM, aunque su unión al DNA fuera correcta (Kal *et al.*, 2000).

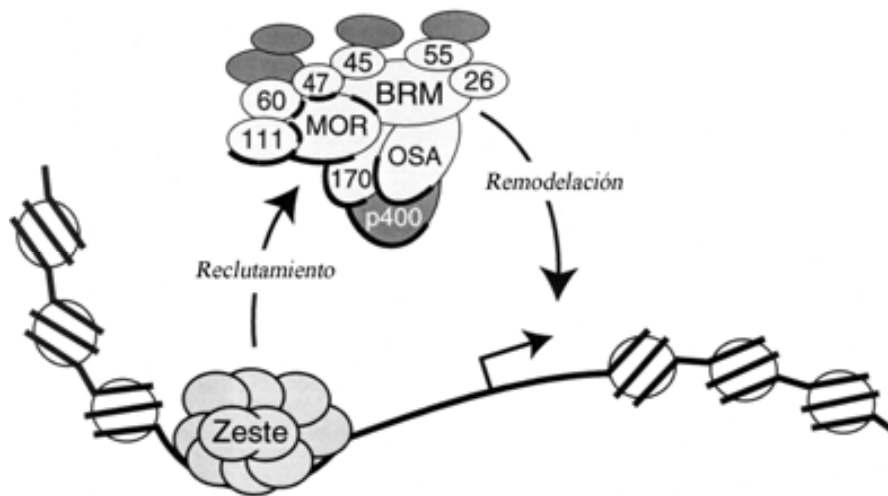


Fig.3.12. Zeste recluta el complejo BRM. La interacción de Zeste con diversas proteínas del complejo BRM, principalmente MOR, aunque también 111 y 170, permite el reclutamiento de BRM en los ZBS de las regiones promotoras de algunos genes. La reestructuración local de la cromatina por parte de BRM permite el acceso a la maquinaria de transcripción.

Existen mutantes por delección de *zeste*. Estos mutantes son viables y fértiles, de modo que *zeste* no es un gen esencial en el desarrollo de *Drosophila*. Sin embargo, sí que es un gen importante, pues en cepas homocigotas por alelos nulos de *zeste*, a menudo aparecen individuos con transformaciones homeóticas que mueren en poco tiempo. También aparece una elevada frecuencia de segregación cromosómica anormal, con una tendencia a la no disyunción de los cromosomas X maternos, con la consiguiente aparición de individuos XXY en gran proporción (Pirrotta, 1991).

Dos de las mutaciones más estudiadas de *zeste* son z^1 y z^{op6} .

El alelo z^1 consiste en un cambio de base en la posición 2360 que implica la sustitución K425M. La lisina 425 es el único aminoácido cargado en una región de 166aa en la parte central de la proteína que, como se ha visto, es importante para la formación de agregados.

Así pues, el cambio de la lisina 425 por una metionina crea un gran dominio hidrofóbico en una zona especialmente sensible, lo cual confiere propiedades hiperagregantes a la proteína codificada por el alelo z^1 .

El alelo z^{op6} se consiguió a partir de una cepa z^1 , irradiada con rayos X (Lifschytz y Green, 1984). Se trata de una segunda sustitución, P454L, que incrementa el efecto hiperagregante de z^1 .

Chen y Pirrotta, en 1993, experimentaron con sustituciones nucleotídicas en las regiones z^1 , z^{op6} y circundantes, y vieron que aminoácidos hidrofóbicos (como valina, leucina o metionina) en las regiones 425/428 (z^1) o $z^1+454/456$ (z^{op6}), daban lugar a proteínas hiperagregantes. Contrariamente, sustituciones que dieran lugar a aminoácidos cargados producían proteínas con un nivel de agregación parecido al salvaje, aún en presencia de las sustituciones hidrofóbicas. En conclusión, pues, el carácter hidrofóbico de las sustituciones es lo que confiere a los productos de z^1 y z^{op6} sus propiedades hiperagregantes y no las bases concretas mutadas o las sustituciones exactas producidas.

La interacción *zeste-white* depende del número de copias de *white*

Se conoce como interacción *zeste-white* a la represión de la expresión del gen *white* por el alelo z^1 . La represión afecta solamente a los ojos, que presentan un color amarillo, en individuos con dos o más copias apareadas de *white*. En general, pues, el fenotipo z^1 sólo es visible en hembras, pues el gen *white* se encuentra en el cromosoma X. Aún así, machos z^1 portadores de una duplicación de *white* también sufren la represión. Mediante experimentos de transgenia se pudo determinar que la región mínima necesaria para la interacción *zeste-white* corresponde a la región reguladora de *white*, portadora de las cajas ZBS, de modo que una duplicación parcial del gen también puede dar lugar a la interacción.

La interacción *zeste-white* mantiene cierta similitud con la variegación por efecto de posición (PEV, del inglés *position effect variegation*). La coloración amarillo pálido del fenotipo z^1 puede tomar distintas intensidades o presentar un fenotipo punteado con ocelos oscuros, en un fondo amarillo, en las células dónde la represión de *white* no ha sido eficaz durante las fases de determinación de la expresión. Sin embargo, la inhibición de *white* disminuye a bajas temperaturas (19°C), con la aparición de más ocelos oscuros, mientras que el efecto de la inhibición por PEV aumenta. Esto indica que, aunque el efecto sea parecido, la base molecular de la interacción *zeste-white* no es la misma que la del PEV (Rasmuson-Lestander *et al.*, 1993).

Zeste¹ podría inactivar por falta de activación

Durante mucho tiempo se pensó que la inhibición de la proteína Zeste¹ sobre *white* se debía a su incapacidad para aproximar, mediante un lazo de DNA, alguna región potenciadora distante al promotor de *white*, debido a su naturaleza hiperagregante. Sin embargo, en 1995 Judd propuso que Zeste⁺ podría actuar, junto con otras proteínas, estabilizando dominios cromosómicos (Judd, 1995). Puesto que existe una gran redundancia en estos mecanismos, la ausencia de Zeste no sería letal para el organismo, aunque sí problemática en algunos aspectos. La proteína Zeste¹, entonces, inhibiría *white* por una mala unión a sus ZBS o por un mal contacto con MOR u otras proteínas del complejo remodelador BRM. La deficiencia de Zeste¹ para activar la transcripción de *white*, permitiría a las proteínas del grupo PcG inactivar el gen desde las primeras etapas del desarrollo embrionario. El silenciamiento de PcG, entonces, sería mantenido epigenéticamente durante toda la vida del individuo mediante la heterocromatinización del gen. Éste modelo encaja bien con las observaciones, pues permite la expresión de un color rojizo de ojos en aquellas células en las que Zeste¹ consigue activar *white*, dando lugar a un fenotipo variegado similar al PEV. La disminución de la temperatura, en éste caso, favorecería la variegación pues los agregados de Zeste¹ disminuyen con las bajas temperaturas, pareciéndose a los agregados de la forma salvaje Zeste⁺.

3.5. EL GENOMA DE *Drosophila melanogaster*

Desde los albores de la genética, la comunidad científica ha tenido la necesidad, intelectual y experimental, de conocer la ordenación de los genes y demás entidades genéticas, en el genoma. Las primeras aproximaciones fueron los mapas genéticos de recombinación y los estudios de ligamiento. Más tarde, la aparición de las técnicas de hibridación permitió la generación de los primeros mapas físicos de baja resolución. Aunque la sofisticación de los protocolos llevó a la disminución gradual de la distancia mínima entre dos marcadores, el mapa físico de alta resolución de un genoma implica la obtención de la secuencia completa del mismo.

El genoma de *D. melanogaster* cuenta con, aproximadamente, 180Mb, de las cuales 120Mb corresponden a su porción euromática. Si consideramos que una reacción de secuenciación da una media de unos 500bp consecutivos, uno de los principales obstáculos de la secuenciación de genomas se hace evidente. Sin embargo, hay otras dificultades: zonas del genoma resistentes a la clonación, regiones heterocromáticas altamente repetitivas, muy inestables y la presencia de varias copias de elementos transponibles que confunden y dificultan las tareas de ensamblaje de los fragmentos de secuencia generados.

La limitación de longitud en las reacciones de secuenciación se resuelve mediante distintas técnicas de generación de fragmentos al azar, llamadas *shotgun sequencing*.

La secuenciación *shotgun* necesita grandes números

Dado un fragmento de DNA, la secuenciación *shotgun* se basa en la generación al azar de un número tal de subclones de pequeño tamaño que cubran en exceso la totalidad del fragmento original. El exceso en la cobertura conlleva el solapamiento de los fragmentos pequeños de modo que, idealmente, los subclones pueden ser ordenados mediante las identidades de sus extremos, para conseguir la secuencia completa del fragmento de DNA original. A medida que aumenta el tamaño del fragmento, debe incrementarse el grado de cobertura, pues en caso contrario aumentan las zonas no representadas en la muestra de subclones. Similarmente, con el incremento del tamaño del fragmento a secuenciar también aumenta el número de copias de elementos repetitivos.

Las lagunas de secuencia impiden el ensamblaje final, dividiendo la secuencia en un conjunto de fragmentos de los cuáles no conocemos la posición en el genoma ni la orientación entre ellos. Estos fragmentos aislados de secuencia son los *contigs*. A su vez, la

presencia de copias idénticas o casi idénticas de elementos transposibles u otras secuencias repetitivas, confunde el proceso de ensamblaje, como se muestra en la figura 3.13, resultando también en el aislamiento de fragmentos de la secuencia original.

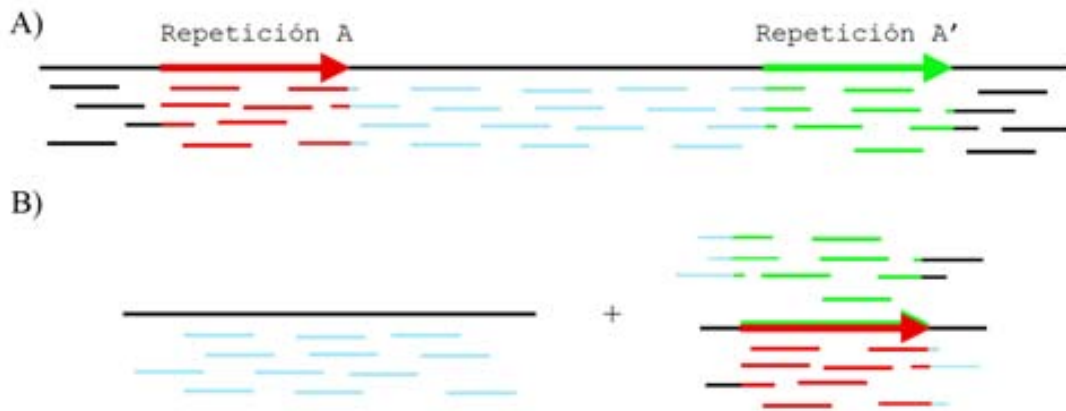


Fig.3.13. Las secuencias repetidas confunden el ensamblaje. A) secuencia de DNA con dos copias de un elemento repetitivo (A y A') y composición de un conjunto de subclones hipotético generado para secuenciación *shotgun*. B) El ensamblaje de los subclones según el apareamiento de bases genera dos fragmentos aislados debido a la confusión de los subclones de las dos repeticiones.

Por los motivos mencionados, inicialmente las técnicas de secuenciación *shotgun* sólo se aplicaron a pequeños genomas víricos y fragmentos de 30-40Kb clonados en cósmidos, considerados el límite de la técnica. En consecuencia, las estrategias de secuenciación de genomas mayores pasaban por la generación de un mapa físico, con el cual poder seleccionar un conjunto de cósmidos que cubrieran la totalidad del genoma. Una vez conseguido el conjunto, cada uno de los cósmidos era secuenciado independientemente mediante una estrategia *shotgun*. Las lagunas de secuencia y los problemas derivados de secuencias repetidas eran resueltos caso por caso en una fase posterior de finalización, con un elevado coste de tiempo y dinero.

Los pares de secuencia permiten secuenciar grandes genomas

En 1994, se secuenció el genoma de *Haemophilus influenzae* con un protocolo íntegramente *shotgun*. Se consiguió mediante el ensamblaje de un conjunto de clones de 2Kb y 16Kb, representativos de toda su secuencia genómica. Éste, de 1,8Mb, estaba considerado más allá de las posibilidades de la técnica, pero la clave fue el desarrollo de un programa informático capaz de usar la información de los denominados “pares de secuencia” (del inglés *mate pairs*, literalmente “pares de colegas”). Los pares de secuencia son los primeros 500bp de cada uno de los extremos de un mismo clon. De modo que, para cada par de secuencia existe una relación de orden y distancia que nosotros conocemos. En

el caso de *H. influenzae* la distancia entre pares de secuencia es de aproximadamente 1Kb y 15Kb, según el clon de procedencia, aunque se pueden generar pares de secuencia a partir de clones mayores. Gracias a esta información, podemos posicionar distintos *contigs*, aunque no estén solapados, como se muestra en la figura 3.14. Ésta es la auténtica revolución de los pares de secuencia, pues dada una suficiente cobertura del genoma, siempre podremos encontrar un clon con uno de los pares en un *contig* y el otro en un *contig* distinto, lo que implica poder ensamblar un genoma sin la necesidad de leerlo secuencialmente.

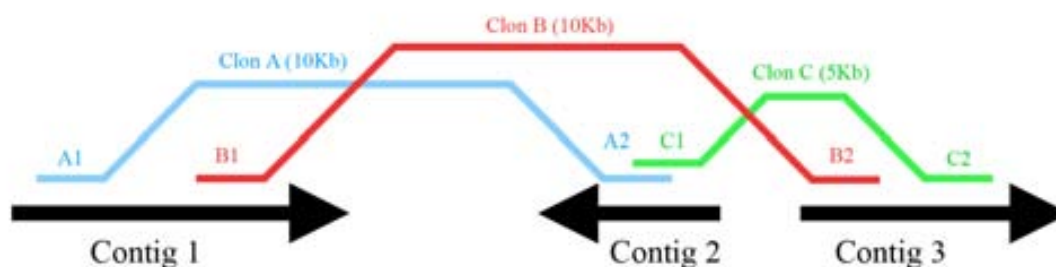


Fig.3.14. Ensamblaje mediante pares de secuencia. Tres *contigs* no solapados, pueden ser ordenados y alineados gracias a la información de largo alcance proporcionada por los pares de secuencia. Las flechas negras simbolizan los *contigs*. Las barras de colores simbolizan los clones con pares de secuencia en distintos *contigs*. Para cada clon se muestra el tamaño de inserto que restringe las distancias a las que se encuentran los *contigs*.

En realidad, para un elevado grado de cobertura, la secuencia proporcionada por los pares de secuencia es suficiente para ensamblar completamente un genoma, sin necesidad de secuenciar completamente cada uno de los subclones. De este modo, el límite de la técnica *shotgun* pasó de los cósmidos de 30-40Kb a los grandes clones BAC y P1, que pueden albergar hasta 100Kb de secuencia.

La secuenciación del genoma de *D. melanogaster* empezó en Mayo de 1999 como una colaboración entre la corporación Celera genomics Inc. y el proyecto público liderado por la universidad de Berkeley, el *Berkeley Drosophila Genome Project* (BDGP). Esta iniciativa formaba parte de un plan mayor, debía probar la técnica para una empresa más ambiciosa: la secuenciación del genoma humano, de 3.000Mb (Myers *et al.*, 2000).

La aplicación del protocolo *shotgun* al genoma de *D. melanogaster* generó una colección de clones BAC con una cobertura 13X del genoma y una colección de clones plasmídicos de 2Kb y 10Kb, con una cobertura de 14,6X. La secuenciación de los extremos de cada clon dio como resultado 654.000 pares de secuencia procedentes de los clones de 2Kb, 497.000 pares procedentes de los de 10Kb y 12.152 pares procedentes de

los BAC. En total, 1.163.152 pares de secuencia. Puesto que la longitud media de cada reacción de secuenciación fue de 551bp, y teniendo en cuenta que cada par de secuencia son dos fragmentos secuenciados, consiguieron un total aproximado de 1.200Mb secuenciadas, lo que significa diez veces la secuencia eucromática del genoma de *D. melanogaster*.

La llave está en el software

El ensamblaje de todos los fragmentos requiere un proceso informático muy bien diseñado, pues la cantidad de datos imposibilita un proceso manual. Con este fin, Celera genomics Inc. desarrolló el denominado Ensamblador (*assembler*) (ver figura 3.15) (Myers *et al.*, 2000), un *software* compuesto por distintos pasos que selecciona y junta los fragmentos de secuencia fruto de las fases iniciales del protocolo, para conseguir una secuencia consenso lo más completa posible del genoma.

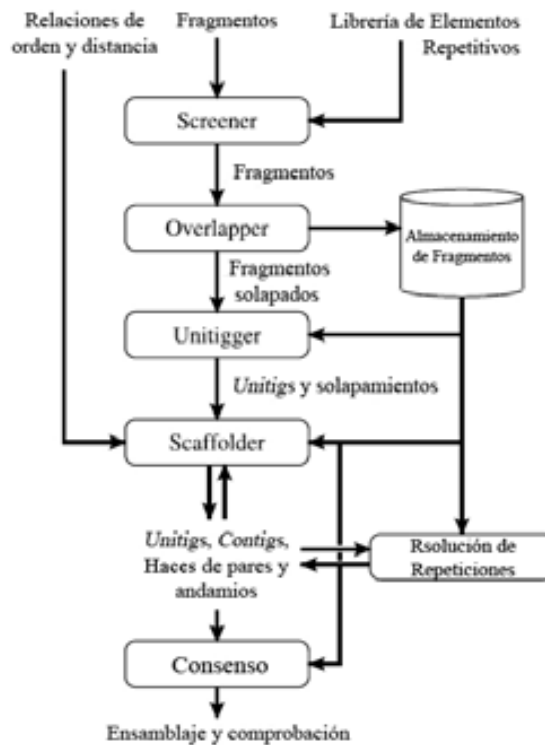


Fig.3.15. El programa ensamblador. Diagrama de flujo del programa ensamblador desarrollado por Celera genomics Inc. Adaptado de Myers *et al.*, 2000.

SCREENER: cada una de las secuencias iniciales fue revisada en búsqueda de elementos repetitivos tales como DNA ribosómico, DNA de genes de histonas, heterocromatina y retrotransposones conocidos. El *screener* puede usar dos filtros, el blando y el duro. El filtro blando marca la secuencia como repetitiva, pero no la excluye del ensamblaje. El filtro duro elimina la secuencia y ya no se toma en consideración en los

siguientes pasos. Durante el proceso de criba, la mayor parte de las secuencias repetitivas fueron filtradas según el filtro duro.

OVERLAPPER: compara cada una de las secuencias con todas las demás para determinar los solapamientos existentes. Usa un algoritmo parecido al del programa de búsqueda por similitud BLAST, que considera un solapamiento a una región mínima de 40bp seguidos con una identidad superior al 94%. Se realizaron 212 millones de solapamientos, con una media de 33,7 solapamientos por fragmento. Las secuencias con un elevado número de solapamientos pueden contener algún elemento repetitivo.

UNITIGGER: una colección de fragmentos claramente solapados forma un *unitig*. Típicamente, los *unitigs* se encuentran flanqueados por elementos repetidos o lagunas de secuencia, con lo cual cada uno de ellos deberá relacionarse con otros *unitigs* mediante la información de pares de secuencia.

Los elementos repetitivos son un gran elemento de confusión en este punto, pues los excelentes solapamientos que producen sus diversas copias dan lugar a *unitigs* que no se corresponden con la realidad. Afortunadamente, los *unitigs* provenientes de secuencias repetitivas pueden identificarse por el elevado número de solapamientos, considerándose *unitigs* sobrecolapsados (del inglés *overcollapsed repeats*). Estos artefactos del proceso de ensamblaje pueden ser eliminados, y seleccionar solamente los *unitigs* de secuencia única, llamados U-*unitigs*.

Se pasó de 3,15 millones de fragmentos de secuencia, a 54.000 *unitigs*, de los cuales 9.413 fueron U-*unitigs* con un tamaño medio de 12,2Kb. El conjunto de los U-*unitigs* cubre 115,4Mb, prácticamente la totalidad de la secuencia eucromática de *D. melanogaster*.

SCAFFOLDER: la información de los pares de secuencia permite unir un conjunto de U-*unitigs* en un *contig*. Pero puesto que la información de un par de secuencia no es siempre fiable, sólo se tuvieron en cuenta aquellas relaciones entre U-*unitigs* que contaran con dos o más pares de secuencia entre ellos. El conjunto de pares de secuencia que relaciona dos U-*unitigs* es denominado un haz de pares de secuencia (*bundle of mate pairs*). Se consiguieron unos 20.000 haces de pares de secuencia, con una media de 10,6 pares por haz. De forma parecida al *unitigger*, el *scaffolder* junta y ordena los *contigs* para formar un andamio (*scaffold*). A su vez, los andamios se ordenaron mediante pares de secuencia de largo alcance, obtenidos de clones BAC. El resultado del *scaffolder* fue un conjunto de 3.736 *contigs* de un tamaño medio de 30,6Kb, relacionados entre ellos por

5.973 haces de pares de secuencia. Esto fue, esencialmente, un primer ensamblaje de la secuencia eucromática no repetitiva del genoma.

RESOLUCIÓN DE REPETICIONES: las lagunas de secuencia entre los U-unitigs, muchas de ellas debidas a la presencia de elementos repetitivos, se rellenaron mediante una estrategia con tres grados crecientes de agresividad, usando *unitigs* y fragmentos de secuencia con presencia de repeticiones (ver figura 3.16). **Rocas:** en una primera fase se usaron *unitigs* relacionados con los U-unitigs con, por lo menos, dos pares de secuencia. **Piedras:** la segunda fase usó *unitigs* relacionados mediante un par de secuencia y un solapamiento. **Grava:** la última fase intentó cubrir el resto de las lagunas mediante fragmentos solapados, aunque sin relaciones de pares de secuencia.

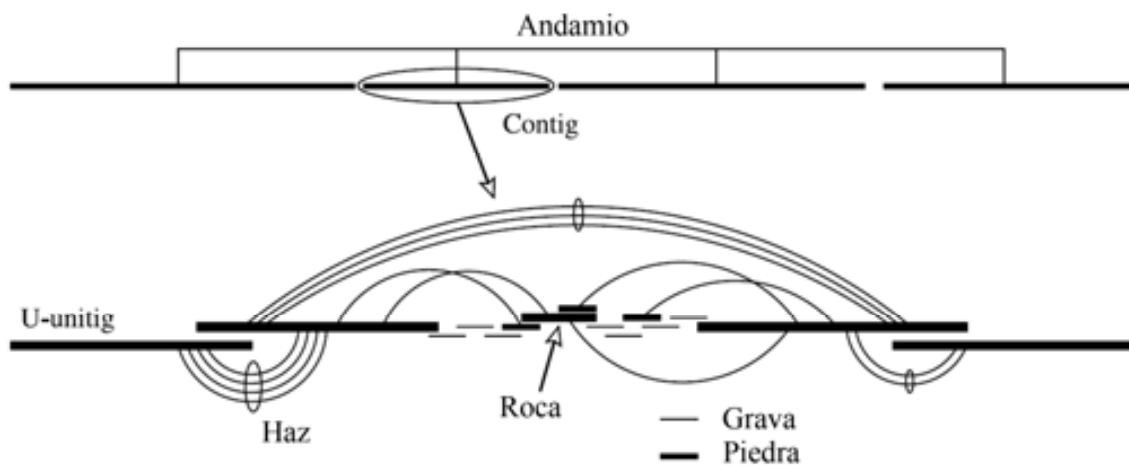


Fig.3.16. Estructura de un andamio. Los andamios son grupos ordenados de *contigs*. Un conjunto de U-*unitigs* relacionados mediante haces de pares de secuencia forman la base de un *contig*. Las lagunas entre ellos se rellenan con rocas, piedras y grava, de forma cada vez más arriesgada, para intentar completar la secuencia. Adaptado de Myers *et al.*, 2000.

CONSENSO: la fase final del proceso, y la más trivial, unifica todas las secuencias de una misma región en una sola secuencia consenso. Para ello, se usaron unas estimaciones de calidad de secuencia obtenidas durante la secuenciación de cada uno de los fragmentos, para priorizar las zonas de elevada calidad y minimizar la presencia de errores de secuenciación.

Finalizar la secuencia consumió algunos años extra

La secuencia de salida del ensamblador, presentada en Mayo del 2000, constituyó el primer lanzamiento (*Release 1*) de una secuencia genómica incompleta, con lagunas de secuencia y *contigs* aislados o desordenados, y con algunas regiones de baja calidad. La corrección de errores se encargó al *Lawrence Berkeley National Laboratory* (LBNL) y al

Human Genome Sequencing Center, Baylor College of Medicine (HGSC). Gracias a la generación de dos mapas físicos del genoma de *D. melanogaster*, uno de STS y el otro sobre un alineamiento mínimo de clones BAC, pudieron determinar la posición de los *contigs* aislados y corregir errores de posicionamiento del proceso inicial. Algunas de las lagunas de secuencia se cubrieron mediante la generación de nuevas reacciones de secuenciación, a partir de cebadores diseñados para la ocasión, otras gracias a un nuevo ensamblaje con pares de secuencia nuevos de mayor calidad.

En octubre de 2000 se presentó el segundo lanzamiento de la secuencia genómica. Cubrió 330 lagunas de secuencia y actualizó cerca de 3000 anotaciones. Quedaron alrededor de 1300 lagunas.

El tercer lanzamiento, en junio de 2002, cubrió casi todas las lagunas existentes y fue especialmente cuidadoso con las secuencias repetitivas, pues se hizo un estudio exhaustivo de todos los transposones presentes en la porción eucromática del genoma. El tercer lanzamiento también contribuyó al ensamblaje de gran parte de secuencias heterocromáticas (Celniker *et al.*, 2002).

Después de dos actualizaciones del tercer lanzamiento, en noviembre de 2004 se presentó el cuarto lanzamiento. Se revisó la anotación de genes y demás elementos constitutivos del genoma, se amplió la cobertura de secuencia heterocromática (gracias al programa *Drosophila Heterochromatin Genome Project*) y se incorporó la secuencia del genoma mitocondrial de *Drosophila*. Aunque aún permanecen algunas lagunas de secuencia y algunos *unitigs* sobrecolapsados, se considera que su secuencia es la de mayor calidad hasta el momento. Todas las secuencias no alineadas u ordenadas, se anotaron como pertenecientes al cromosoma U, para poder estar disponibles para el estudio aunque su localización u ordenación sea desconocida.

En el momento de redacción del presente trabajo de tesis doctoral, la secuencia genómica de *D. melanogaster* se encontraba en la tercera revisión del cuarto lanzamiento (*release 4.3*), aunque estaba prevista la publicación formal del quinto lanzamiento para finales de 2006 o principios de 2007, según la información publicada en el sitio *web* de *FlyBase* (<http://flybase.bio.indiana.edu/annot/dmel-release4-notes.html>).

4. OBJETIVOS

Como se describe en el apartado “Antecedentes”, el presente trabajo se inició como la descripción molecular de una cepa mutante (M115) y su revertiente fenotípico (RM115), aparecidos en el laboratorio a raíz de unos experimentos de mutagénesis con ENU, EMS y choque térmico. Puesto que los efectos de estos mutágenos son diversos y de carácter aleatorio, los objetivos iniciales fueron muy generales:

OBJETIVO 1º: Caracterización molecular de la ampliación del fenotipo *zeste* a los machos de la cepa M115.

OBJETIVO 2º: Estudio de las causas de la reversión fenotípica en los machos de M115 que originaron la cepa RM115.

A medida que el trabajo avanzaba en el cumplimiento de los objetivos 1º y 2º, los resultados abrían nuevas líneas de estudio que marcaron sus propios objetivos. Sin embargo, estos no fueron más específicos, pues el desconocimiento del elemento transponible *FB-NOF* no lo permitió:

OBJETIVO 3º: Revisión de la estructura y de la distribución genómica del elemento transponible *FB-NOF* de *D. melanogaster*.

OBJETIVO 4º: Análisis de la región codificante del elemento transponible *FB-NOF*. Expresión y localización subcelular.

5. RESULTADOS Y DISCUSIÓN

5.1 ANÁLISIS DE LAS CEPAS M115 Y RM115

La prioridad en el inicio del estudio de las cepas M115 y RM115, fue profundizar en los paralelismos existentes entre estas cepas y las descritas por la Dra. Rasmuson-Lestander, w^{+UR} y w^{+UZ} . A partir de los trabajos de la Dra. Baldrich, conocíamos la similitud fenotípica de los individuos M115 y RM115 con los w^{+UR} y w^{+UZ} , respectivamente: la ampliación de la interacción *zeste-white* a los machos y la existencia de una alteración en el mapa de restricción, justo después del gen *white*. Sin embargo, no disponíamos de datos suficientes para comparar ambos pares de cepas.

Descubrimos la inserción de un elemento *FB*

Desde un punto de vista clásico, los pasos a seguir para determinar la similitud de las reordenaciones detectadas en M115 y RM115 con las descritas por Rasmuson-Lestander, pasarían por la elaboración de un mapa de restricción completo de la región y la hibridación con sondas específicas de *FB* y *NOF*. Sin embargo, la publicación del primer lanzamiento de la secuencia genómica de *D. melanogaster*, en enero del año 2000, nos permitió trazar una estrategia algo más atrevida.

Tomamos como hipótesis de trabajo que las reordenaciones detectadas en nuestras cepas se debían a la inserción de un elemento transponible *FB-NOF* en M115 y sólo *FB* en RM115. Puesto que, en cualquier caso, siempre existirían secuencias *FB* en el punto de inserción, diseñamos un experimento de PCR con un cebador en el extremo del IR de *FB* (*FB*) y otro en un punto de la región genómica cercana a la supuesta inserción (*G5*) (ver Fig. 5.1 1). La determinación de esta región genómica se hizo con los datos de *Southern blot* de Baldrich y Rasmuson-Lestander que, aunque aproximados, cuentan con precisión suficiente para poder encontrar la inserción en uno o dos intentos (ver Figs. 3.1 y 3.2, en la sección Antecedentes).

Según el mapa de restricción de Baldrich, la reordenación se encuentra por debajo de la diana *PvuII* en -3,3Kb (coordenadas de Levis), pero antes de la diana *HindIII* en -6Kb. El mapa de Rasmuson-Lestander sitúa su inserción una kilobase por encima de las dianas *HindIII* en -6Kb, cerca de la *KpnI* de -4,2Kb. En consecuencia, para empezar la búsqueda de un posible elemento *FB* en las cepas M115 y RM115, elegimos un punto situado alrededor de -5Kb. El cebador *G5* se encuentra en -5233bp, según las coordenadas de Levis para el gen *white*.

Elegir un punto concreto en la secuencia supone un compromiso, pues dada la precisión del *Southern blot*, puede suceder que la inserción real se encuentre por arriba o por debajo del punto elegido. Para tratar de cubrir ambas posibilidades, el cebador en secuencia genómica, G5, fue usado también en su forma reversa y complementaria, G5r. Puesto que el cebador FB, se encuentra en el IR, hibrida en los dos extremos del elemento y en las dos orientaciones (ver Fig. 5.1).

El experimento consta de dos reacciones de PCR, usando en cada una el cebador de *FB* y uno de los cebadores en secuencia genómica. Ambas reacciones son excluyentes. Entre las dos, podemos rastrear un límite aproximado de 4Kb alrededor del cebador genómico, en busca de elementos *FB* (ver Fig 5.1). Las cepas analizadas fueron CS como control negativo, pues no está descrita la presencia de elementos *FB* en la región. Como control positivo, w^{+UZ} proporcionada por Rasmuson-Lestander. Las cepas problema fueron, obviamente, M115 y RM115.

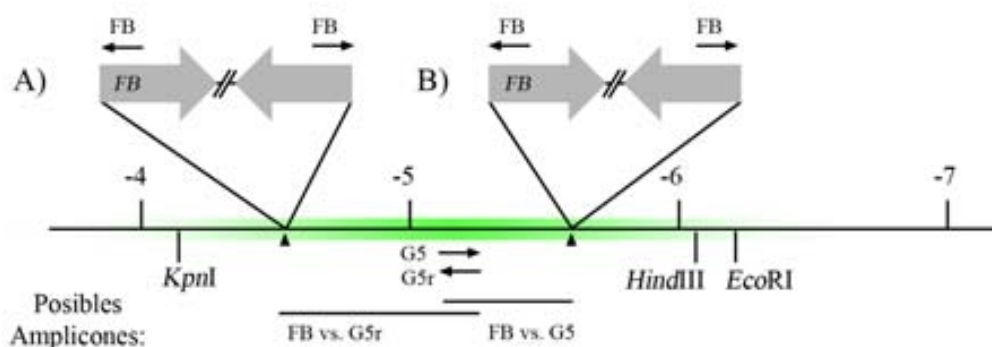


Fig.5.1. Estrategia de PCR. Si la reordenación descubierta por Baldrich entre *KpnI* y *HindIII* se debe a un elemento *FB*, existen dos posibilidades, **A)** La inserción se encuentra antes del punto elegido para el cebador *G5*, -5233, de modo que sólo puede amplificarse el producto entre los cebadores *FB* y *G5r*, o **B)** la inserción se encuentra después de -5233 y el único producto de PCR posible es entre *FB* y *G5*. La región degradada simboliza la zona de fiabilidad alrededor de *G5/G5r*. La numeración muestra las coordenadas de Levis.

El resultado del experimento mostró amplificación con la pareja *G5* y *FB* en w^{+UZ} , M115 y RM115, pero no en CS. Esto sitúa un elemento *FB* en estas cepas, por debajo del punto elegido. El fragmento amplificado contiene el extremo 5' de la inserción. La reacción *G5r* vs. *FB* no muestra amplificación, pues sólo una puede darla (ver Fig. 5.2A).

El tamaño del fragmento amplificado es de entre 800bp y 850bp. Considerando que el cebador *FB* se encuentra a 478bp del inicio del elemento, deducimos que existe una inserción de *FB* a una distancia de aproximadamente 400bp por debajo del cebador *G5*. Sorprendentemente, la posición de la inserción en nuestras cepas y en w^{+UZ} parece ser casi idéntica. Para esclarecer esta posible coincidencia y con el fin de precisar el punto exacto

de la inserción, los productos de PCR fruto del experimento fueron secuenciados mediante un cebador interno en *FB* (FBs), situado a 123bp del inicio del elemento (ver Fig. 5.2C).

Los resultados de la secuenciación desvelaron el punto de inserción del elemento *FB* en las cepas M115, RM115 y w^{+UZ} , con la máxima precisión posible. La inserción se produce en la posición -5655, según las coordenadas de Levis, en la región posterior del gen *white* (ver Fig. 5.2D).

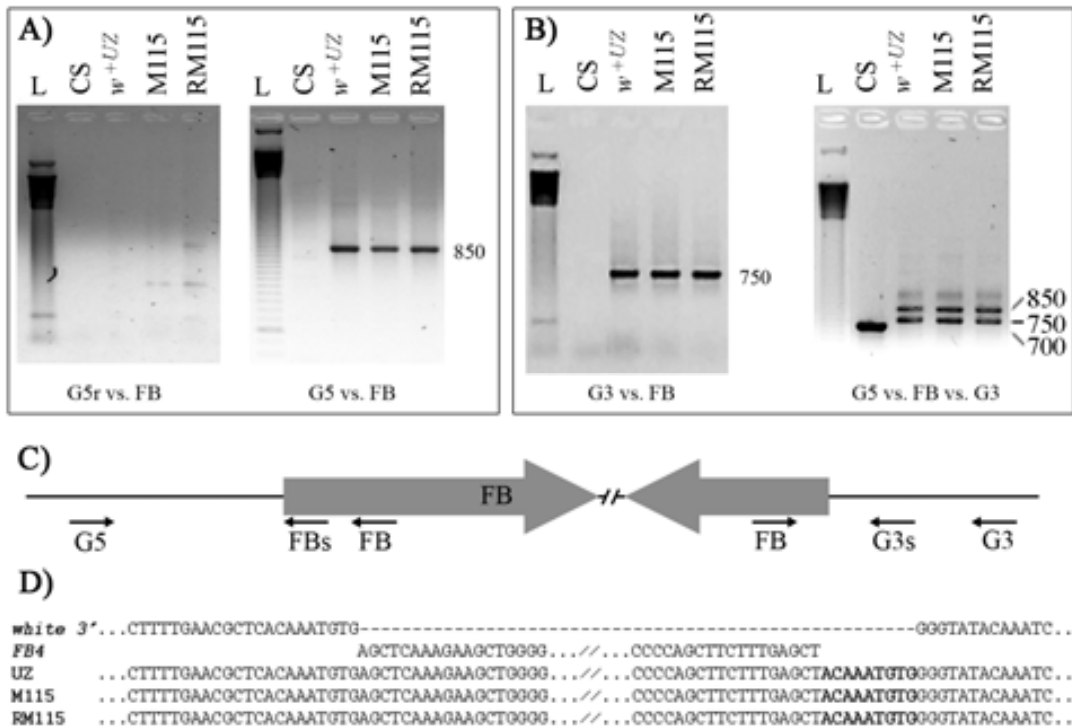


Fig.5.2. Localización de un elemento *FB*. **A)** Resultados de las reacciones de PCR G5r vs. *FB* y G5 vs. *FB*. Solamente se observa amplificación al usar esta última pareja, lo cual indica la presencia de un elemento *FB* por debajo de -5233. **B)** Resultados de las reacciones G3 vs. *FB*, extremo 3' de la inserción, y reacción dúplex usando los tres cebadores G5, *FB* y G3. Las cepas con inserción de *FB* muestran las dos bandas correspondientes a los extremos de la inserción, mientras que CS muestra el fragmento de 700bp correspondiente a la región entre G5 y G3. **C)** Esquema de la inserción del elemento *FB*. Se muestra la posición aproximada de todos los cebadores implicados en las PCR y la secuenciación. El esquema no está a escala. **D)** Alineamiento de secuencias entre la región genómica salvaje de la región 3' del gen *white*, la secuencia publicada del elemento transponible *FB4* y las secuencias obtenidas de las cepas w^{+UZ} , M115 y RM115. En negrita se muestra la duplicación directa de 9bp, longitud típica de los elementos *foldback*.

El conocimiento del extremo 5' de la inserción nos permitió diseñar el cebador G3, para amplificar un fragmento de DNA que contuviera el extremo 3' de la inserción (ver. Fig. 5.2C). La posición de G3 es -5963, a una distancia de 730bp de G5 y 308bp del punto de inserción de *FB*. El fragmento de DNA entre los cebadores *FB* y G3 se muestra en la Fig. 5.2B. Como se esperaba, sólo amplifica en las cepas en las que se había encontrado la inserción del elemento *FB*: w^{+UZ} , M115 y RM115.

Puesto que en los dos experimentos de PCR, hasta el momento, la cepa CS siempre había dado resultados de no amplificación, buscamos la forma de demostrar la ausencia de *FB* mediante un resultado positivo. Así pues, se diseñó un experimento de PCR dúplex con los cebadores G5, FB y G3, que amplifica los dos extremos de la inserción, más la región entre G5 y G3 en ausencia del transposón. Debido al tamaño de los elementos *FB* y a la fuerza de sus estructuras secundarias, la amplificación G5 vs. G3 en una cepa con la inserción es altamente improbable. Los resultados de la PCR dúplex se muestran en la Fig. 5.2B, y demuestran positivamente la ausencia de *FB* en la cepa CS.

El extremo 3' de la inserción se secuenció a partir del cebador G3s, situado 161bp por debajo de la inserción de *FB*, sobre el producto de amplificación G3 vs. FB (ver Fig. 5.2C). La secuencia confirma el punto de inserción revelado por la secuenciación del extremo 5', y muestra la duplicación directa de 9bp, longitud típica de la transposición de los elementos *FB* (ver Fig. 5.2D).

La coincidencia del punto de inserción entre las cepas M115, RM115 y w^{+UZ} supuso un resultado desconcertante. Podría explicarse de tres formas: 1) Nos encontraríamos delante del mismo alelo de *white* descrito por Rasmuson-Lestander. 2) La región podría ser un punto caliente (*hot spot*) de inserción para el elemento *FB*. 3) Se trataría de una curiosa casualidad.

La inserción en M115 y RM115 es distinta de la de w^{+UZ}

El hecho de que la cepa w^{+UZ} hubiera sido usada en el pasado por otros miembros del grupo hizo pensar en la posibilidad que M115 fuera el resultado de una contaminación y, por lo tanto, que el cromosoma X analizado en el presente trabajo fuera el mismo analizado por Rasmuson-Lestander. Sin embargo, la cepa w^{+UZ} es portadora del marcador *singed* (*sn*) en el cromosoma X, mientras que M115 y RM115 son portadoras de *forked* (*f*). Además, los datos de estabulación de la unidad confirmaron que la cepa w^{+UZ} dejó de utilizarse antes del inicio del trabajo de la Dra. Baldrich, y no estaba en mantenimiento durante el mismo. Puesto que los marcadores del cromosoma X no coinciden entre w^{+UZ} y M115, y ya que no había una población permanente de w^{+UZ} en el laboratorio, la hipótesis de la contaminación se considera poco probable.

Si se trata de un punto caliente para la inserción del elemento *FB* o, por el contrario, fue casualidad, probablemente nunca se sabrá. El objetivo del trabajo fue analizar, con la mayor precisión posible, las causas y los efectos de la mutación ocurrida en M115 y la

reversión en RM115, de modo que continuamos con el diseño de nuevos experimentos para conseguirlo.

Conocíamos el punto exacto de la inserción. Nos faltaba por conocer el tamaño y la composición del material insertado, pues aunque ya hubiéramos demostrado la presencia de *FB* en M115 y RM115, faltaba saber si alguna de ellas contaba con la presencia de un elemento completo *FB-NOF*.

Para determinar el tamaño, lo mejor es la técnica de *Southern blot*, de modo que usamos la sonda pWKK, generada *ad hoc*, para comparar el tamaño del fragmento de restricción *KpnI-KpnI*, donde se encuentra la inserción, entre las cepas w^{+UZ} , M115, RM115 y el control CS (ver materiales y métodos). Es importante destacar que el extremo 3' del clon pWKK contiene el fragmento *EcoRI-KpnI* que utilizó Rasmuson-Lestander en la descripción de las inserciones de w^{+UZ} y w^{+UR} .

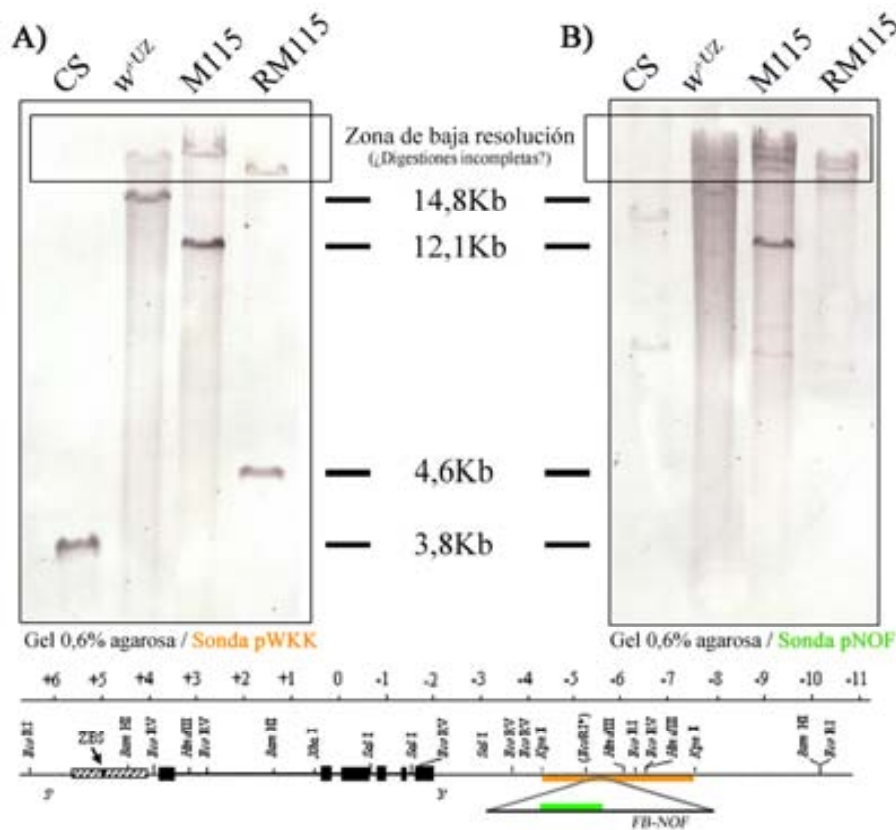


Fig.5.3. Tamaño y composición de las inserciones. DNA genómico de CS, w^{+UZ} , M115 y RM115 digerido con la enzima de restricción *KpnI* e hibridado con, **A)** sonda pWKK y **B)** sonda pNOF. Todos los carriles corrieron en el mismo gel, al 0,6% de agarosa, en las mismas condiciones y se transfirieron a la vez. El filtro fue cortado e hibridado por separado con cada sonda. Se muestran los tamaños de las bandas más relevantes y la región de baja resolución del gel. Entre paréntesis, el tamaño de las inserciones en cada banda.

Los resultados del experimento mostraron una banda *KpnI-KpnI* de 12,1Kb en M115 y 4,6Kb en RM115 (ver Fig. 5.3A), muy distintas de las descritas por Rasmuson-Lestander

para las cepas w^{+UZ} y w^{+UR} . Las cepas control dieron bandas dentro de lo esperado: 3,8Kb para CS y 14,8Kb para w^{+UZ} . Es de esperar encontrar algunas discrepancias con los tamaños descritos en otros trabajos, pues las características de cada electroforesis y el uso de distintos métodos para calcular los tamaños, resultan en valores diferentes. Las cepas w^{+UZ} , M115 y RM115 presentaron una acumulación de material de elevado peso molecular en la zona superior de cada carril, que interpretamos como el resultado de digestiones incompletas, cosa común para la enzima de restricción *KpnI*.

Para determinar la composición del material insertado, las mismas digestiones se hibridaron con la sonda pNOF (ver materiales y métodos). Aunque las cepas w^{+UZ} y M115 mostraron diversas bandas de elevado peso molecular en la zona de baja resolución del gel, se aprecian con claridad señales de hibridación en los mismos fragmentos de 14,8Kb y 12,1Kb que hibridaron con la sonda de *white* (ver Fig. 5.3B). La cepa salvaje CS y el revertiente RM115, sin embargo, no mostraron coincidencia entre los fragmentos que hibridaron con *white* y los que lo hicieron con *NOF*.

De estos resultados se deducen unas inserciones de 11Kb en la cepa w^{+UZ} , 8,3Kb en M115 y 0,8Kb en RM115. Estos valores demuestran, más allá de toda duda, que el origen de M115 y RM115 es distinto de w^{+UZ} y w^{+UR} , como ya apuntamos en el apartado anterior. En particular, mientras que parece segura la inserción de un elemento *FB-NOF* en M115, al igual que en w^{+UZ} , la reversión en RM115 parece producirse por la escisión de la mayor parte del elemento y no contiene *NOF*. No podemos trazar un paralelismo con w^{+UR} , pues los eventos descritos por Rasmuson-Lestander implican que una cepa con un elemento *FB* (w^{+UR}), pasa a contener un elemento completo *FB-NOF* (w^{+UZ}). Justo lo contrario que M115 y RM115.

La reversión detectada en RM115 es tan extrema, que casi entra en conflicto con los resultados de PCR que detectaron *FB* (el cebador *FB* se encuentra a 478bp del inicio del elemento, de modo que, entre los dos extremos, amplificamos por lo menos 956bp de *FB* en todas las cepas mutantes). Si bien los resultados de *Southern blot* son siempre aproximados, lo cierto es que el gel tiene resolución suficiente para asegurar que el incremento real de la banda *KpnI-KpnI* en RM115, respecto de CS, no superaría por mucho 1Kb. Este tamaño entra dentro del alcance de una reacción de PCR, de modo que se intentaron varias combinaciones de cebadores flanqueantes al elemento *FB*, para conseguir un fragmento que abarcara toda la inserción. Ninguna de las combinaciones dio resultado, aunque usáramos polimerasas especiales para reacciones difíciles o largas. Aunque pueda parecer lógico, pues los IR de *FB* pueden producir estructuras secundarias muy fuertes en

cadena simple, no lo es si se tiene en cuenta que un elemento *FB* de alrededor de 1Kb está formado, en su mayor parte, por las secuencias iniciales no internamente repetitivas.

Aparecen los primeros indicios de una reordenación

Para superar el aislamiento aparente entre los dos lados de la inserción, tanto en M115 como en RM115 y conseguir más indicios sobre la composición de las secuencias insertadas, elaboramos un mapa de restricción completo de la región. Obviamente, la sonda pWKK no pudo usarse por contener el punto de inserción, de modo que se utilizaron dos sondas menores, una de la región inmediatamente anterior a *FB*, pWKE* y otra inmediatamente posterior, pWEK. Con el uso de dos sondas pudimos detectar un mayor número de dianas de restricción y confirmar la posición de algunas de ellas (ver Sección 7.1.7.1 en Materiales y métodos).

Los resultados del mapa de restricción muestran algunas características sorprendentes, que no se habían considerado hasta el momento (ver Fig. 5.4). La posición y distancia de las dianas de restricción confirma la presencia de un elemento *NOF* en la cepa M115. Aunque la diana *Pst*I aparece entre *Sal*I y *Bam*HI en la hibridación con pWKE*, su posición es correcta en la hibridación con pWEK y, puesto que el primer experimento de *Southern blot* ya indicaba la presencia del elemento, no se buscó la posición de más dianas en esta cepa. Cabe mencionar, pero, que a juzgar por el orden *Sal*I→*Eco*RI, la orientación del elemento es inversa respecto a la expresión del gen *white*, lo cual podría interferir en su expresión y ampliar el fenotipo *zeste*¹.

Lo inesperado fue, sin embargo, descubrir que la cepa RM115 también cuenta con la presencia de un elemento *NOF*. Se mapearon hasta siete posiciones internas del elemento y todas ellas se encuentran en el orden y distancias relativas en las que se encuentran en la secuencia publicada de *NOF*. Este elemento también se encuentra en orientación inversa respecto de *white*, aunque RM115 no presenta el fenotipo *zeste*¹ ampliado. Otra consecuencia del uso de dos sondas fue la detección de la siguiente diana *Kpn*I a una distancia de aproximadamente 30Kb y no 4Kb como mostraba el primer experimento de *Southern blot* con la sonda pWKK.

El patrón de restricción de *NOF* en RM115 se observó, solamente, al rastrear las dianas con la sonda pWKE*. Al tratar de confirmar las posiciones con la sonda pWEK no obtuvimos el mismo patrón, sino uno vagamente similar al de la región 3' del gen *white*, aunque todas las dianas de restricción estarían algo desplazadas de su posición esperada y aparece un sitio *Pst*I no existente en la secuencia salvaje (ver Fig. 5.4).

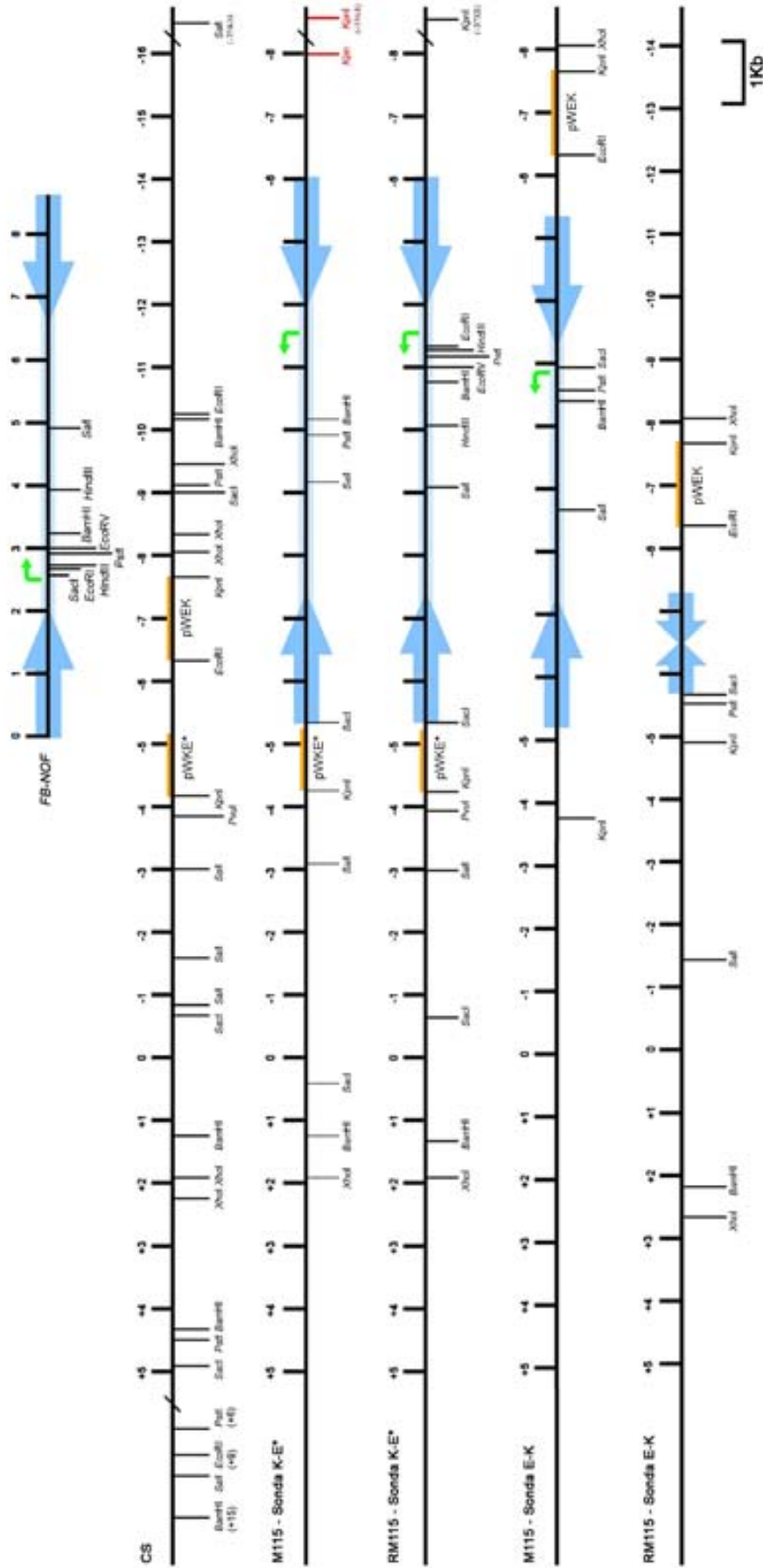


Fig.5.4. Mapas de restricción de las cepas CS, M115 y RM115. En la parte superior se muestra la posición de las dianas de restricción en un elemento *FB-NOF* hipotético. La flecha verde indica el punto de inicio y la dirección de transcripción del primer ORF de *NOF*. Las barras naranjas marcan la posición y área de hibridación de las sondas pWKE* y pWEK. El tamaño de los elementos insertados es aproximado.

Los indicios revelados por el mapa de restricción supusieron un punto de inflexión en el estudio. En particular, las discrepancias encontradas para la posición de la diana *KpnI* en RM115 al usar una sonda u otra. Encontrar una sola banda *KpnI-KpnI* de gran tamaño con la sonda pWKE*, nos hizo pensar que, tal vez, las señales de hibridación en la zona de baja resolución observadas en el experimento inicial de *Southern blot*, pudieran no ser digestiones incompletas. Para comprobar esta hipótesis, repetimos el experimento de *Southern blot* corriendo las digestiones *KpnI* en un gel de agarosa al 0,3%, para obtener mayor resolución para las bandas de elevado peso molecular. Se hibridaron los fragmentos con las sondas pWKE*, pWEK y pNOF.

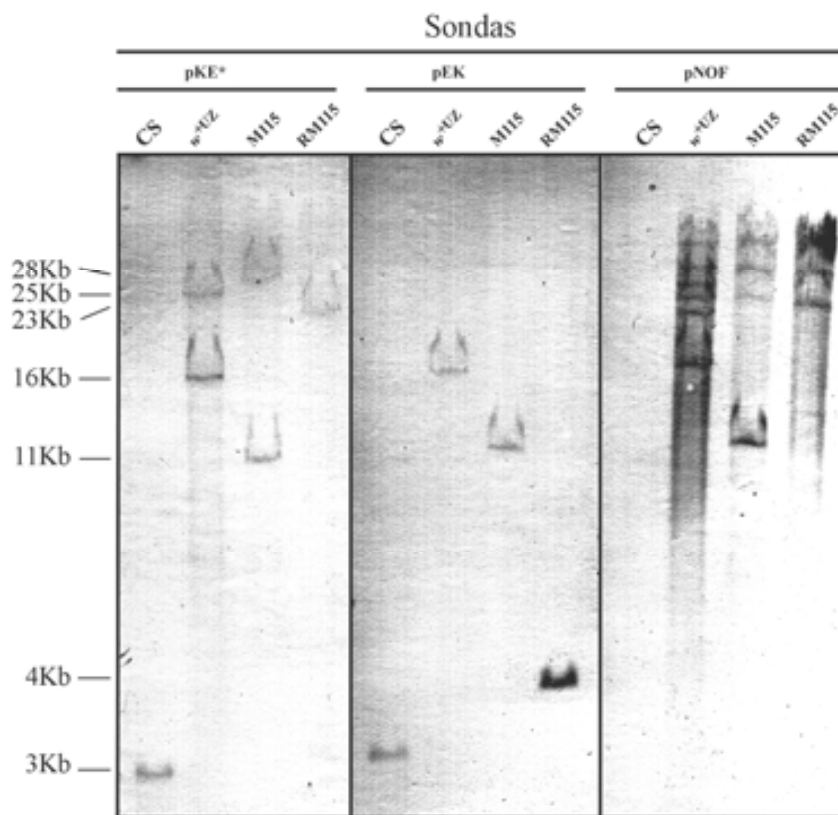


Fig.5.5. Tamaño y composición de las inserciones, segundo experimento. DNA genómico de CS, *w*^{+UZ}, M115 y RM115 digerido con la enzima de restricción *KpnI* e hibridado con las sondas pWKE*, pWEK y pNOF. Todos los carriles corrieron en el mismo gel, al 0,3% de agarosa, en las mismas condiciones y se transfirieron a la vez. El filtro fue cortado e hibridado por separado con cada sonda.

El resultado de este nuevo experimento confirmó nuestras sospechas sobre las señales de hibridación en la zona de elevado peso molecular (ver Fig. 5.5), pues sin duda corresponden a fragmentos de restricción *KpnI-KpnI*. La sonda pWKE* hibridó con dos fragmentos a la vez en las cepas de ojos amarillos, *w*^{+UZ} (16Kb y 25Kb) y M115 (11Kb y 28Kb), por sólo uno de 23Kb en RM115. La sonda pWEK hibridó con los mismos fragmentos de 16Kb en *w*^{+UZ}, 11Kb en M115 y 4Kb en RM115, que ya detectamos en el

primer experimento de *Southern blot* con la sonda pWKK (ver Figs. 5.3 y 5.5). La sonda pNOF hibrida con diversos fragmentos, pero parece clara la coincidencia con los de 16Kb y 11Kb de w^{+UZ} y M115, respectivamente. Debido a las diferencias entre los geles de los dos experimentos de *Southern blot*, el tamaño calculado de los fragmentos varía un poco. Puesto que este segundo experimento está pensado para mejorar la resolución de bandas de elevado peso molecular, confiamos más en estos tamaños.

En conjunto, los datos de los mapas de restricción y del segundo experimento de *Southern blot* nos llevan a una conclusión muy alejada de las hipótesis iniciales. La hibridación de la sonda pWKE* con dos fragmentos *KpnI-KpnI* distintos en las cepas w^{+UZ} y M115 indica que, probablemente, la región anterior a la inserción se encuentra duplicada. Si el fragmento presuntamente duplicado alcanzara la región promotora del gen *white*, el fenotipo *zeste*¹ ampliado a machos quedaría perfectamente explicado en las dos cepas, lo cual supondría abandonar las hipótesis de Rasmuson-Lestander que proponían un efecto de la actividad de *NOF* sobre el gen *white*. La cepa revertiente, por su parte, aunque no muestra bandas dobles, sí muestra bandas distintas con las dos sondas. Esto podría significar que las secuencias con las que hibridan no se encuentran adyacentes en el mismo fragmento *KpnI-KpnI* y que la reversión, de algún modo, separó físicamente los dos extremos de la inserción de M115, haciendo que cada sonda reconozca fragmentos *KpnI-KpnI* distintos en RM115.

***white* está duplicado, por lo menos en las cepas de ojos amarillos**

Los resultados del *Southern blot* son claros, pero otros fenómenos pueden producir bandas dobles sin necesidad de tratarse de duplicaciones. Por ejemplo, podría ser que la diana *KpnI*, en las cepas w^{+UZ} y M115, estuviera desplazada respecto de CS y la sonda reconociera los dos fragmentos *KpnI-KpnI* adyacentes. Improbable, pero posible.

Para determinar si el gen *white* está duplicado en estas cepas, diseñamos un experimento de PCR cuantitativa a tiempo real. Amplificamos y cuantificamos cuatro regiones genómicas: la primera, wZB, en el ZBS del gen *white*, imprescindible en la interacción *zeste-white*. La segunda, wBc, en el primer exón de *white*. La tercera, wKE, justo antes del punto de inserción, en el fragmento *KpnI-KpnI*. La cuarta, CG, más allá de la diana *KpnI* posterior al punto de inserción. Una región del gen *zeste* sirve de control de copia única (ver Fig. 5.6).

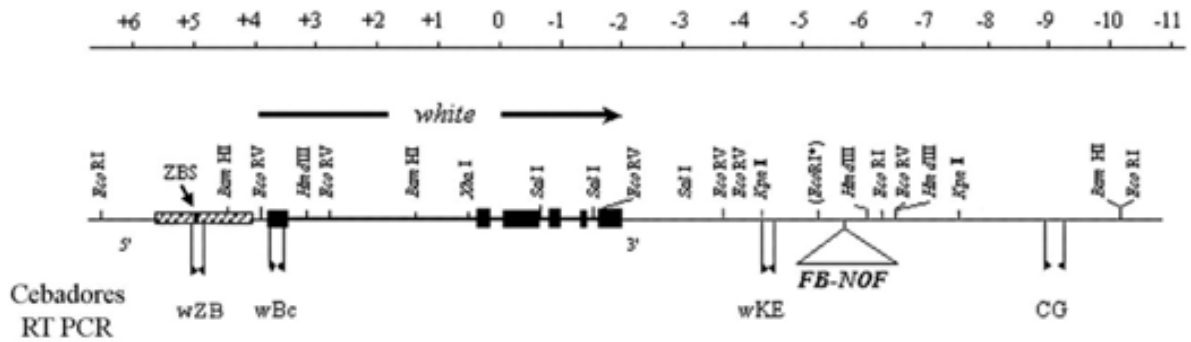


Fig.5.6. Cebadores para PCR a tiempo real. Esquema de la región genómica del gen *white*. La caja rayada indica la zona promotora, con el ZBS. Las cajas oscuras simbolizan los exones. Debajo, se muestran las posiciones aproximadas de los cebadores utilizados en el experimento de PCR a tiempo real.

El primer paso para cuantificar la dosis génica consiste en determinar la relación de cada reacción respecto al control, dividiendo sus concentraciones estimadas. El valor obtenido, sin embargo, no puede tomarse como una cuantificación, pues cada reacción de PCR tiene su propia eficiencia y no se pueden comparar dos reacciones distintas sin normalizar. Correríamos el riesgo de asignar mayor dosis a una región que simplemente amplifique con más facilidad. Asumiendo que la eficiencia de cada reacción se mantiene entre cepas, normalizamos cada una de las relaciones obtenidas en las cepas problema con el valor obtenido para la cepa salvaje CS, la única que garantiza una sola copia de cada región (evidentemente, el experimento cuenta con DNA exclusivamente de machos). El resultado de la normalización sí puede considerarse una cuantificación relativa de la dosis génica de cada región, respecto a la región control.

En la tabla 5.1 se muestran las relaciones de eficiencia de cada reacción respecto del control *zeste* para cada cepa y los valores normalizados que indican la dosis génica de cada región. La comparación de los valores de cada reacción ejemplifica el porqué de la normalización: en RM115, un valor de 0,51 en la relación wBc/z y un valor de 2,16 en la relación wKE/z, determinan la misma dosis génica ($1,2 = 1$ copia), debido a que la reacción wKE es un 78% más eficiente que la de *zeste* (*ratio* 1,78), mientras que la reacción wBc es un 68% menos eficiente (*ratio* 0,42).

Cepa	RELACIÓN CON <i>zeste</i>				NÚMERO DE COPIAS			
	wZB/z	wBc/z	wKE/z	CG/z	wZB	wBc	wKE	CG
CS	0,91 ^a	0,42 ^a	1,78 ^a	0,33 ^a	1,0	1,0	1,0	1,0
w ^{+UZ}	2,06	n.a.	n.a.	0,41	2,2	n.a.	n.a.	1,2
M115	1,90	0,75	3,73	0,27	2,1	1,8	2,1	0,8
RM115	0,90	0,51	2,16	0,27	1,0	1,2	1,2	0,8

(a) usado como relación de eficiencia. (n.a.) no aplicable.

Tabla 5.1. Cuantificación relativa de la dosis génica del gen *white* y alrededores. Se muestran, en fondo blanco, las relaciones de concentración estimada de cada reacción respecto a *zeste*, en cada cepa. Las celdas con fondo gris muestran los valores normalizados, que sí pueden considerarse una cuantificación relativa.

En el análisis de los resultados, tuvimos en cuenta el hecho de que la PCR cuantitativa a tiempo real y su *software* asociado, está diseñada para mediciones con diferencias mayores que las que estábamos determinando. Por eso, los resultados obtenidos tienen una precisión sin sentido en nuestro experimento. No se puede tener 2,1 copias de una zona duplicada. Así, hay que entender que los valores “1,8”, “2,1” y “2,2” indican una dosis génica de “2”. Igualmente, los valores “1,2” y “0,8” indican una dosis génica de “1”.

Con estas consideraciones, pues, los resultados demuestran que todas las regiones cuantificadas anteriores al punto de inserción están duplicadas en las cepas w^{+UZ} y M115, mientras que la región posterior no lo está. La cepa RM115, sin embargo, cuenta con una sola copia de cada región.

Puesto que el primer individuo M115 se mantuvo en cultivo con hembras de cromosomas X unidos, solamente el cromosoma X de este macho se transmite a sus descendientes. El mantenimiento del fenotipo entre éstos indica que la duplicación detectada en los experimentos anteriores está contenida en el cromosoma X. Sin embargo, la segunda copia de *white* podría estar adyacente, en tándem, en el locus *white* o, por el contrario, podría haber translocado a un punto lejano en el cromosoma. Para tratar de identificar la posición de las dos copias recurrimos a la hibridación *in situ* sobre cromosomas politénicos. Preparaciones de cromosomas politénicos de las cepas CS, M115 y RM115 se hibridaron con sondas para *white* (pWXX), detectada mediante anticuerpos conjugados con Alexa 488 (color verde) y *NOF* (pNOF), detectada mediante anticuerpos conjugados con rodamina (color rojo) (ver Fig. 5.7).

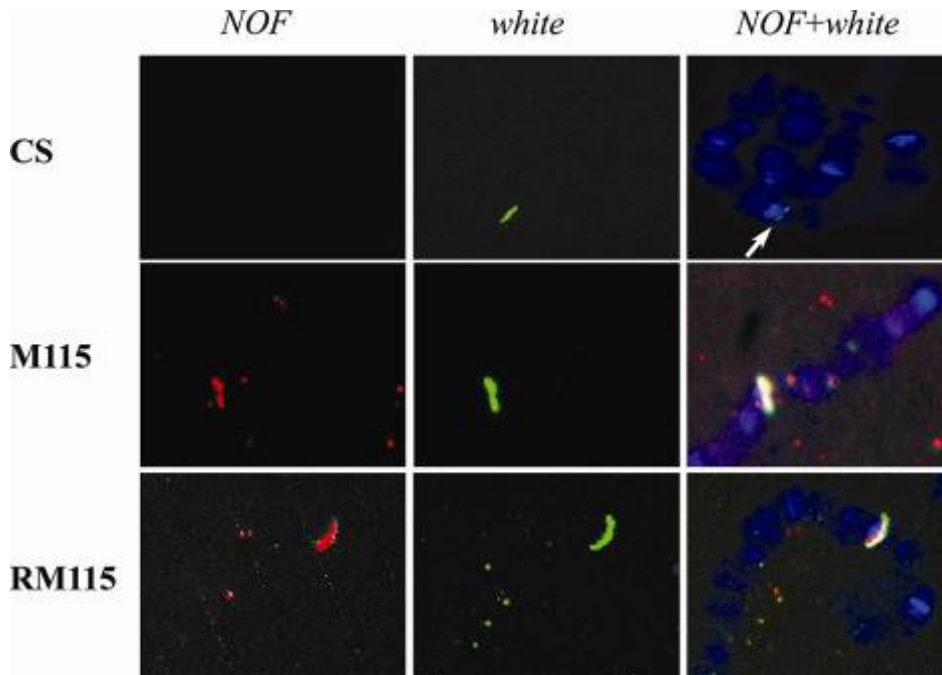


Fig.5.7. Detección de *white* y *NOF* mediante FISH. Se muestran, ampliadas, las zonas de hibridación de las sondas pNOF (rojo) y pWXX (verde) en las cepas CS, M115 y RM115. La imagen de la derecha muestra la superposición de las dos imágenes anteriores, sobre los cromosomas politénicos teñidos con DAPI. La flecha señala la posición del locus *white* en CS, pues la señal es muy débil.

En todos los casos se observa una única señal de hibridación de la sonda pWXX, en la banda X3C1, lo cual indica que no existe translocación asociada a la duplicación de *white*, en la cepa M115 (ver Fig. 5.7). Por lo menos, no en nuestro nivel de resolución. Debemos concluir, por lo tanto, que la región duplicada se encuentra en tándem o, en todo caso, suficientemente cercana como para hibridar en la misma banda del cromosoma politénico.

El resultado de la FISH es coherente con la posible presencia de un elemento *FB-NOF* completo en la inserción de la cepa RM115, como se había sugerido a raíz del mapa de restricción, dada la hibridación de la sonda pNOF en la misma banda del locus *white*.

Discusión general del análisis de las cepas M115 y RM115

La descripción final de las cepas M115, RM115 y colateralmente w^{+UZ} , dista bastante de las hipótesis iniciales. El fenotipo de ojos amarillo claro en M115 y w^{+UZ} es consecuencia de la duplicación del gen *white* en un fondo genético *zeste*¹. La reversión, que da lugar a la cepa RM115, se debe a la pérdida de una de las copias. Estos resultados nos permiten prescindir de las hipótesis de Rasmuson-Lestander, no respaldadas por pruebas, sobre la posible interferencia del elemento *FB-NOF* en la expresión de *white*. Esto no significa que no pueda existir tal interferencia pero, de existir, no sería la causante del fenotipo *zeste*¹ en M115 ni en w^{+UZ} .

La razón que llevó a error al grupo de Rasmuson-Lestander fue, sin duda, la elección de la sonda en su experimento de *Southern blot*. Su sonda es equivalente a nuestra pWEK, que da un resultado muy similar al presentado en su artículo, sin las bandas dobles que evidencian la duplicación. Evidentemente, sus explicaciones fueron las más coherentes con los resultados de que disponían. Sin embargo, puesto que la reversión de la cepa w^{+UZ} se estableció y usó como un ensayo de mutagénesis (Rasmuson, 1984; Xamena *et al.*, 1984; Fujikawa y Kondo, 1986; Vogel, 1989; Batiste-Alentorn *et al.*, 1991; Rasmuson y Larsson, 1992), es importante conocer con precisión el fundamento de la mutación. En este sentido, los resultados obtenidos mediante el uso de dicho ensayo quizás deberían revisarse y ajustarse a eventos de recombinación y no de escisión de un transposón, como se creía hasta el momento.

Respecto a los mecanismos moleculares de la mutación y reversión, lo más probable es que se deban a fenómenos de recombinación homóloga entre elementos *FB* o *FB-NOF* no homólogos, es decir, situados en sitios no homólogos del genoma. Si eso fuera cierto, los dos extremos de la duplicación en M115 contendrían secuencias *FB* (por lo menos) (ver Fig. 5.8). La inserción detectada es, sin duda, uno de los extremos de la reordenación. Encontrar el otro extremo fue, sin embargo, imposible. Experimentos de *Southern blot* cubriendo las 30Kb anteriores al gen *white* y reacciones de PCR con cebadores en +20Kb y +25Kb (según coordenadas de Levis), en combinación con el cebador *FB*, no dieron ningún indicio de la presencia de otro elemento *FB* o *FB-NOF* cercano (los datos no se muestran, pues no hay diferencias entre las cepas).

La reversión, por su parte, parece ser el resultado de una delección por recombinación. No sería nada inusual, sino todo lo contrario, pues la cepa M115 es inestable y, probablemente, todos los individuos revertientes que genera se deban a delección por recombinación de una de las dos copias de *white*, ¿pero cuál? Es interesante recordar que, en RM115, las dos sondas flanqueantes al punto de inserción detectado hibridan con fragmentos *KpnI-KpnI* distintos. En M115, la banda de 11Kb se mantiene en las hibridaciones de las dos sondas, lo cual significa que el fragmento *KpnI-KpnI* está intacto a pesar de la inserción de *FB-NOF*. Esto no sucede en RM115, con lo cual nos inclinamos por la posibilidad de que la delección afecte al locus *white*, quedando intacta la copia fruto de la duplicación (ver Fig. 5.8).

Finalmente, puesto que el fragmento detectado por la sonda pWKE* en RM115 tiene un tamaño 6Kb menor que el detectado en M115, creemos que la diana *KpnI* en 3' no es la misma en las dos cepas. Igual que el fragmento detectado por la sonda pWEK, la

integridad de la secuencia en el extremo 3' de la duplicación tampoco se mantiene entre M115 y RM115. Este dato, junto con la relativa cercanía de la copia duplicada al locus *white*, sugiere que toda la región genómica entre la duplicación y el locus *white*, incluyendo el gen, podría haber sido reemplazada por un fragmento de secuencia flanqueado por elementos *FB-NOF*, al estilo de los transposones gigantes de Ising (Ising y Block, 1984).

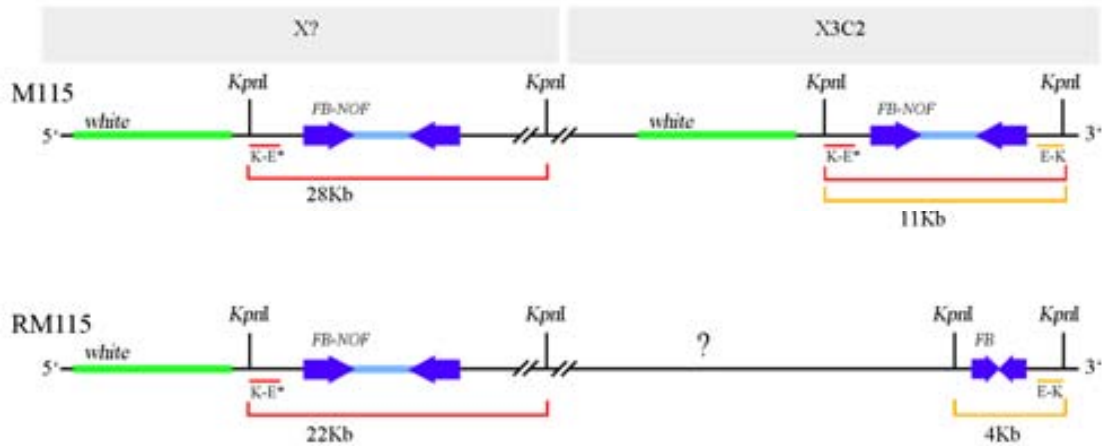


Fig.5.8. Propuesta sobre las reordenaciones en M115 y RM115. Esquema de la explicación más probable de las reordenaciones ocurridas en M115 y RM115, según los datos de nuestros experimentos. No está dibujado a escala.

5.2 ANÁLISIS DEL ELEMENTO TRANSPONIBLE *FB-NOF*

En las primeras fases del estudio de las cepas M115 y RM115, con todas las hipótesis aún abiertas, empezamos a estudiar las posibilidades de que un elemento transponible *FB-NOF* pudiera alterar la expresión de un gen cercano. Para nuestra sorpresa, la información sobre el transposón era escasa y antigua, pues la mayoría de referencias son de los años 80 y principios de los 90 (para un resumen de la información previa, ver la sección de *FB-NOF* en la introducción).

La modernización de las técnicas clásicas de biología molecular, así como la aparición de nuevas metodologías y de las bases de datos genómicas, nos permitió retomar el estudio de *FB-NOF* para aportar nuevos datos sobre su estructura y, en la medida de lo posible, de su función.

Clonamos los IR de un elemento *FB*

Los únicos IR secuenciados de un elemento *FB* son los de *FB4*, en 1982 (Potter *et al.* 1982). Pero, justamente, *FB4* es un mal representante de la familia *foldback*, pues la inserción casual de un elemento *HB* en su interior oculta la mayor parte de *FB*. Por eso decidimos obtener nuestra propia secuencia completa de los IR de un elemento *FB*, a saber, el presente en la región 3' del gen *white* en M115. Contar con un elemento completo *FB-NOF* fue una ventaja, pues no solamente podíamos asegurar que los IR estarían enteros, sino que nos permitiría estudiar la transición entre *FB* y *NOF*.

En primer lugar tratamos de conseguir dos clones, uno para cada IR completo del elemento. La forma más rápida de conseguirlos es mediante PCR y clonación, pero puesto que la secuencia de los IR de *FB* es internamente repetitiva, era imprescindible cubrir todo el IR en una sola reacción. El tamaño y la composición de dichos IRs no ayudan en la tarea, pues cada uno mide algo más de 2Kb y, aunque el porcentaje total de C/G en las zonas repetitivas ronda el 30%, estas no son nada homogéneas y existen largas tiras formadas exclusivamente por A/T, a intervalos regulares. Estas regiones pueden alterar localmente la topología de la hebra de DNA, dificultando el proceso de amplificación.

Los dos cebadores de la región 3' del gen *white* (G5 y G3) utilizados para la detección inicial de la inserción (ver Fig. 5.2), se combinaron con diversos cebadores en *NOF* para tratar de conseguir fragmentos compuestos por secuencia de *NOF*, algo de secuencia

genómica y un IR de *FB* entero entre ellas (ver cebadores en Materiales y métodos). Con el uso del sistema *Fail Safe*™ PCR System, con su batería de *Fail Safe*™ PCR Premix Selection, pudimos amplificar dos fragmentos de DNA prometedores entre los cebadores G5/NOF5 y G3/NOF3 (ver Fig. 5.9). Puesto que la amplificación se consiguió con la premix *D*, los fragmentos fueron nombrados *bandaD5'* (4,5Kb) y *bandaD3'* (2,4Kb), respectivamente. Los dos fragmentos fueron clonados en vectores plasmídicos para su posterior análisis de restricción y secuenciación, estableciendo los clones pCRD5' y pBSD3' (ver Fig. 5.9 y la sección de sondas en Materiales y métodos). Cabe destacar que las dificultades que preveíamos con las repeticiones internas de *FB* realmente aparecieron. Las reacciones fallaban a menudo o resultaban en escaso producto y, como se observa en la Fig. 5.9, en los procesos de clonación aparecían plásmidos con insertos de distintos tamaños, generalmente menores a los calculados para las *bandasD*.

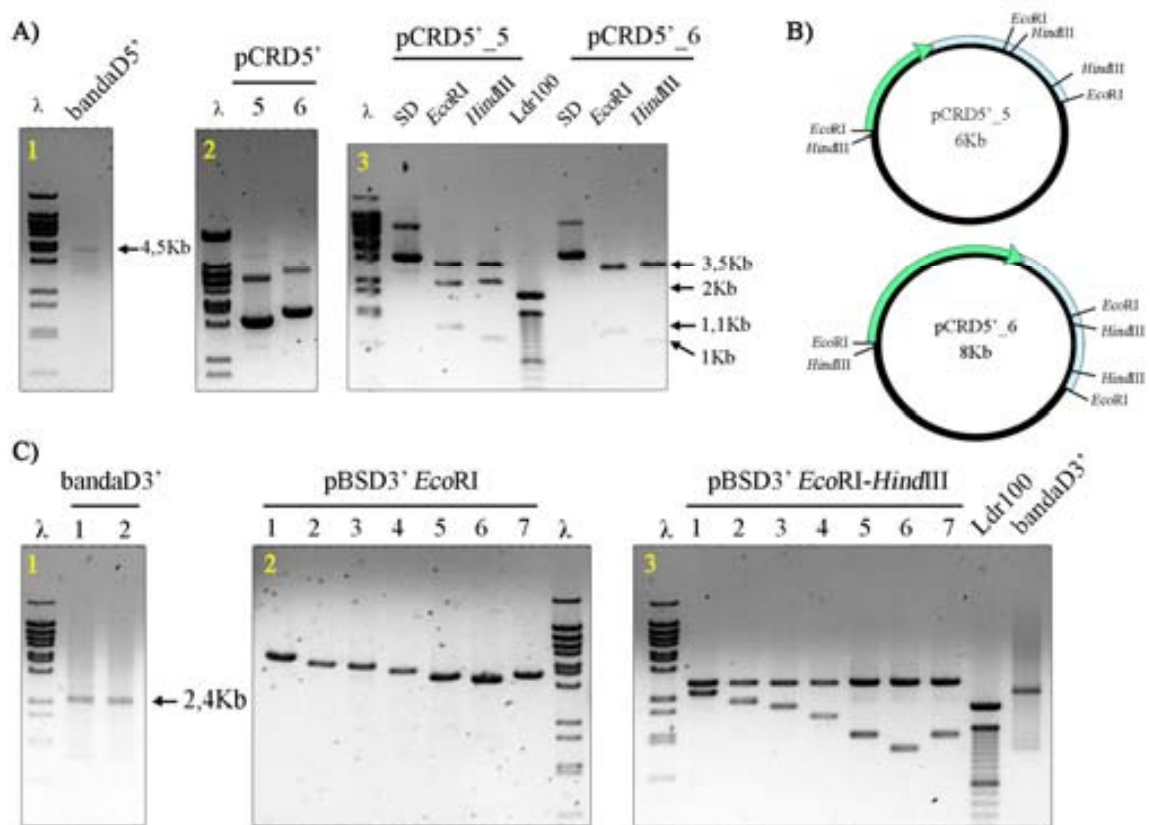


Fig.5.9. Los clones pCRD5' (A) y pBSD3' (C). Se muestran los momentos clave durante el proceso de amplificación (1), clonación (2) y selección (3) de los dos clones que contienen los IR enteros del elemento *FB* presente en la cepa M115. En (B) se muestran los esquemas de dos clones pCRD5', que ejemplifican las deleciones internas en los IR de *FB*. SD: Sin Digerir. λ : lambda *Bst*EII. Ld100: marcador 100bp

Un sencillo mapa de restricción de dos clones pCRD5' sirve de ejemplo. pCRD5'_5 y pCRD5'_6 difieren en casi dos kilobases, aún partiendo del mismo molde. ¿En qué parte

del clon se encuentra la diferencia? Las digestiones *EcoRI* y *HindIII* nos lo muestran claramente: ambos clones liberan un fragmento *EcoRI* de 1,1Kb, lo esperado entre la diana *EcoRI* de *NOF* y la presente en el vector pCR4-TOPO[®]. También liberan el fragmento *EcoRI* correspondiente al vector plasmídico, de 3,5Kb. Sin embargo, pCRD5' _5 muestra otro fragmento, de 2Kb, que debe contener el IR de *FB* y una kilobase de *NOF*, hasta su primera diana *EcoRI*. Esta banda no se observa en pCRD5' _6 porque solapa con la banda de 3,5Kb del vector. Un patrón similar se observa con las digestiones *HindIII*.

La conclusión es que la kilobase y media de menos en pCRD5' _5 se encuentra en el fragmento correspondiente al IR de *FB*, algo lógico debido a su patrón repetitivo. Puesto que dichas repeticiones parecen demasiado grandes para expandirse por causa de errores en la PCR, creemos que lo más probable es la aparición de deleciones, de modo que tomamos los clones de mayor tamaño como los más fiables a la cadena molde original. Se establecieron, pues, como clones de trabajo para los IR de *FB*, los plásmidos pCRD5' _6 y pBSD3' _1. Posteriormente, pCRD5' _6 se separó en dos fragmentos *EcoRI*: pBSD5' _6A, usado para la secuenciación del IR (contiene el IR de *FB*, la primera kilobase de *NOF* y un trozo de DNA genómico) y pBSD5' _6B, usado de sonda en algunas ocasiones (contiene la segunda kilobase de *NOF*, hasta el cebador NOF5). En adelante, en las referencias a dichos plásmidos no se mencionará el número de clon.

La secuenciación de *FB* proporciona una nueva interpretación de su estructura

El proceso de secuenciación de los clones pBSD5' y pBSD3' se planeó con sumo cuidado, pues las repeticiones internas de *FB* dejaban muy poco margen de maniobra a las técnicas usadas comúnmente. El tamaño de los IR, unas dos kilobases, imposibilita la secuenciación en una sola reacción. En estos casos, lo normal es recurrir a la estrategia de *primer walking*, pero en una región formada por repeticiones es inútil pues los cebadores hibridan en múltiples posiciones y es imposible distinguir los fragmentos adyacentes. Una estrategia *shotgun* es igualmente inapropiada, pues no hay forma de ordenar las secuencias correctamente, ni de determinar el número de repeticiones reales. Sin embargo, existe una forma de conseguir secuencias adyacentes completamente fiables: generar un conjunto de subclones anidados y ordenados, mediante la digestión controlada de uno de los extremos del IR con la exonucleasa *Bal31* (ver Materiales y métodos).

Generamos dos colecciones de subclones con deleciones terminales, una para pBSD5' y otra para pBSD3' (Fig. 5.10A y B respectivamente). En todo momento se tuvo especial

cuidado en que las deleciones no superaran los 500bp, la media de una reacción de secuenciación, para asegurar el solapamiento de las secuencias de clones adyacentes (con excepción de pBSD5', en el que empezamos a secuenciar por la porción inicial no repetitiva del IR de *FB*).

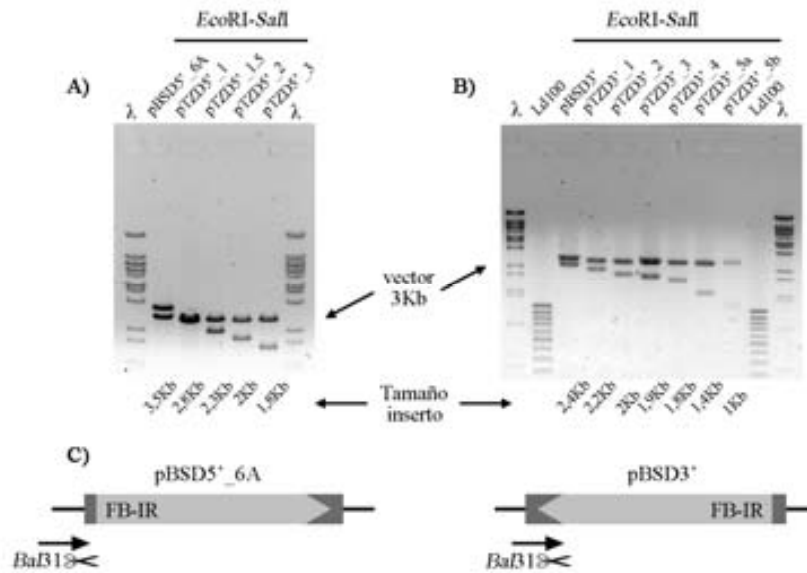


Fig.5.10. Generación de subclones con deleciones anidadas. Los clones de partida, pBSD5' y pBSD3' sufrieron rondas de digestión con la exonucleasa *Bal31*. Se muestran los subclones con los insertos liberados mediante digestión *EcoRI-SalI*. A) Colección D5' y B) Colección D3'. La banda de 3Kb corresponde al vector plasmídico. Se muestran los tamaños de los insertos de cada subclon. C) Esquema de los plásmidos pBSD5'_6A y pBSD3' con el punto de entrada de la reacción *Bal31*. No están a escala.

Pero esto no basta cuando se trata de *FB*. Puesto que el tamaño de los subclones se calcula en gel de agarosa, no sabemos con suficiente precisión el alcance de la digestión *Bal31* y fácilmente podríamos saltar alguna de las repeticiones que forman parte del IR del elemento (según los datos previos, las repeticiones miden 31bp y se organizan en conjuntos de 5 para formar bloques de 154bp) (Harden y Ashburner, 1990). Para encontrar la forma de alinear las secuencias de los subclones, minimizando las posibilidades de saltar alguna de las repeticiones, analizamos la estructura de las mismas en profundidad en busca de características distintivas que pudieran orientarnos durante el alineamiento.

En las primeras descripciones de la secuencia repetitiva de *FB*, se creyó que cada una de las repeticiones era prácticamente idéntica a las adyacentes, sin ningún tipo de patrón (Truett *et al.*, 1981; Potter, 1982). Sin embargo, eso no es correcto, pues sí existe un orden en la forma en que *FB* se organiza:

- La región no repetitiva consta de 470bp, con algunas repeticiones directas cortas que no suponen problemas de alineamiento. Los últimos 150bp contienen una secuencia muy parecida a la que forma las unidades repetitivas del resto del elemento, encabezada por el trinucleótido CTC, aunque no puede considerarse parte de la región repetitiva.
- La unidad básica de la porción repetitiva de *FB* es lo que llamamos el “bloque”. Su tamaño es de 154bp (Fig. 5.11) y no se conoce si hay un número fijo por IR. El inicio de un bloque siempre contiene el trinucleótido CTC, que no se encuentra en ninguna otra parte de la porción repetitiva de *FB*. Por lo general, los bloques sí son idénticos.
- Los bloques, a su vez, están formados por cinco unidades de lo que llamamos “Repeticiones”. Su tamaño es de 31bp y no son idénticas, sino que pueden distinguirse perfectamente y respetan un orden muy estricto de posición (Fig. 5.11). La quinta repetición es más corta, mide 27bp, y conecta con el siguiente bloque de repeticiones. La primera y la segunda repetición se distinguen por el cuarteto final (CCTT y TCCT, respectivamente). La tercera, la cuarta y la quinta se distinguen en su secuencia intermedia y el final truncado de la quinta.

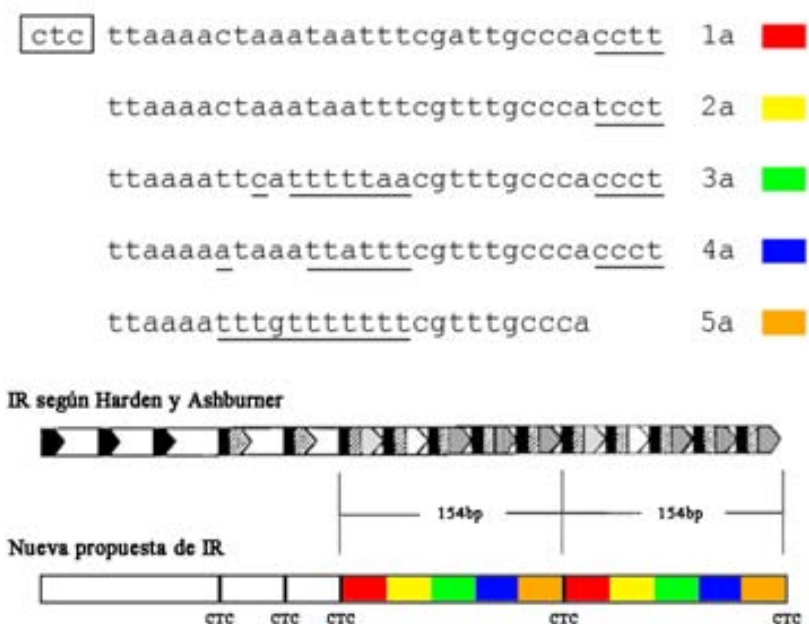


Fig.5.11. Un bloque de la región repetitiva de *FB*. Se muestra la secuencia entera de un bloque, formateada de forma que se aprecien sus cinco repeticiones y las características distintivas entre ellas. Debajo, esquema original de Harden y Ashburner (1990) comparado con nuestra propuesta de estructura de bloques.

Esta nueva visión de la estructura de la zona repetitiva de *FB*, nos mostró una suerte de mapa con el que orientarnos durante el proceso de secuenciación. En particular, el trinucleótido CTC, que se erige como una marca cada 154bp, es algo muy conveniente pues los geles de agarosa sí tienen esa precisión. Así pues, en cada una de las secuencias obtenidas a partir de los subclones, se identificaron las marcas CTC de inicio de bloque. En un segundo paso, se decidía cuáles de los bloques solapaban entre secuencias adyacentes dependiendo de la longitud degradada por *Bal31* y de la secuenciada. Un ejemplo puede ser más explicativo:

Entre los clones pTZD5’_1 y pTZD5’_1.5 existe una diferencia de 500bp (Fig. 5.10 y 5.12). La secuenciación de pTZD5’_1 resultó en 710bp de secuencia de buena calidad hasta el último CTC correctamente identificable. La secuencia solapada entre los dos clones es, por lo tanto, alrededor de 200bp. En total, pTZD5’_1 cuenta con 400bp de secuencia de la región no repetitiva y dos bloques enteros (308bp), más retazos del tercero. Los inicios CTC de estos bloques se encuentran en 436bp, 560bp y 706bp.

En el clon pTZD5’_1.5, la secuencia empieza en medio de un bloque, en una tercera repetición. El primer inicio de bloque (CTC) se encuentra a 70bp, de modo que debe coincidir con uno alrededor de 570bp en pTZD5’_1, según el cálculo aproximado de la digestión *Bal31*, de unos 500bp. En conclusión y dado el margen de error de los geles de agarosa, el primer CTC que encontramos en pTZD5’_1.5 corresponde al segundo bloque en pTZD5’_1 (560bp) (Fig. 5.12).

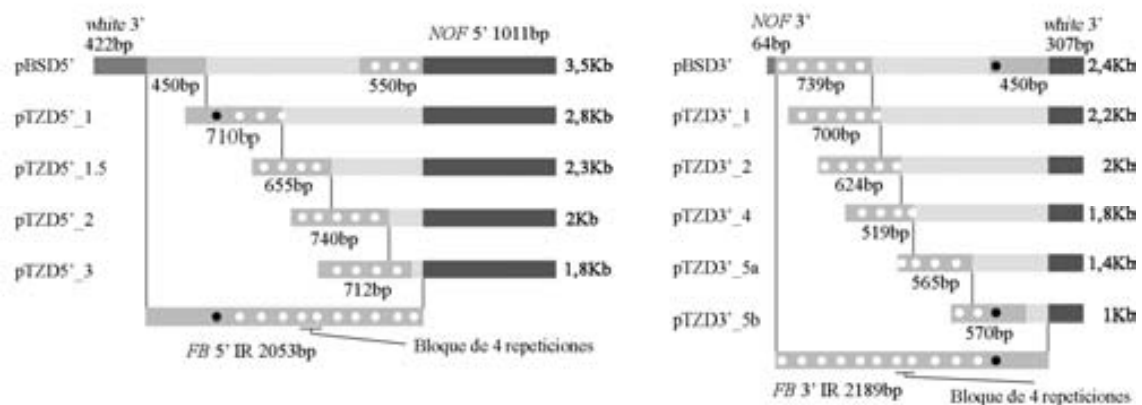


Fig.5.12. Esquema del proceso de alineamiento de los subclones *Bal31*. Se muestran los clones originales pBSD5’ y pBSD3’, con sus colecciones de subclones generados por digestión con *Bal31* y sus tamaños calculados en gel de agarosa. Se indican las partes correspondientes a *white*, *NOF* y el alcance de las reacciones de secuenciación. Los círculos indican la posición aproximada de los trinucleótidos CTC de inicio de bloque. El círculo negro indica el comienzo de la región previa a las repeticiones de bloques. Debajo de cada colección, se muestra el resultado del ensamblaje de los fragmentos de secuencia.

Siguiendo el proceso de alineamiento y ensamblaje descrito, pudimos obtener la secuencia completa de los dos IR de la porción *FB* del elemento *FB-NOF* presente en la cepa M115. Ambas secuencias fueron alojadas en los servidores públicos de GenBank, con los números de acceso DQ005444 (5') y DQ005445 (3').

Esta es la primera vez que se secuencian los dos IR completos de un mismo elemento *FB-NOF*, de forma fiable. Quizás lo más destacable, en este punto del análisis, fue la posibilidad de observar la transición entre *FB* y *NOF* bajo el prisma de nuestra estructura para las repeticiones de *FB*. Anteriormente, solamente se había secuenciado esta transición en dos ocasiones (Templeton y Potter, 1989; Harden y Ashburner, 1990) y en ningún caso establecieron un criterio claro de separación entre *FB* y *NOF*. Sin embargo, cualquier conclusión que dedujéramos a partir de tan pocas secuencias sería bastante débil, por lo cual nos propusimos recopilar y analizar el mayor número de elementos *FB* o *FB-NOF* que pudiéramos identificar en las secuencias genómicas de *D. melanogaster*, presentes en las bases de datos públicas.

Recorremos el genoma en busca de *FB-NOF*

La búsqueda de *FB-NOF* en las secuencias genómicas, fruto del consorcio internacional para la secuenciación del genoma de *D. melanogaster*, suponía algunas ventajas y un inconveniente. Las ventajas serían disponer de un buen número de secuencias con las que poner a prueba la estructura de bloques y repeticiones definida en el anterior apartado y la posibilidad de encontrar características comunes que ayudaran a definir mejor la secuencia y la estructura de *FB*, así como su relación con *NOF*. El inconveniente: no podríamos confiar plenamente en que las secuencias estuvieran correctamente alineadas, dada la estrategia de secuenciación *shotgun* empleada con el genoma de *D. melanogaster* (ver Introducción).

Este inconveniente es insalvable. Los elementos *FB-NOF* forman *unitigs* sobrecolapsados, de modo que el paso por la fase de *scraper* del programa ensamblador los elimina y marca la región para un análisis más detallado en la fase de finalización. No hay forma de determinar si dichos esfuerzos de finalización tuvieron éxito en las lagunas de secuencia producidas por elementos *FB* (una parte importante del total, según la Dra. Susan Celniker, comunicación personal). Habría que volver a secuenciar esos fragmentos por un sistema más fiable, como el empleado en este trabajo. Sin embargo, quizás el problema del alineamiento no fuera tan importante en nuestro objetivo. Lo más probable es

que un alineamiento incorrecto resultara en un elemento *FB* con un número erróneo de bloques de repeticiones, según nuestro esquema, pero la estructura del mismo debería mantenerse. Así pues, la comparación de elementos *FB-NOF* de secuencias genómicas con nuestro elemento secuenciado, u otros más antiguos, sería legítima siempre que se restringiera a estructura y no se tuviera en cuenta el tamaño total o el número de bloques de repeticiones.

La búsqueda de elementos *FB-NOF* en las secuencias fruto del proyecto de secuenciación del genoma de *D. melanogaster* se llevó a cabo con las diversas versiones del script `cazaFB.pl`, diseñado *ad hoc* para este propósito (ver Materiales y métodos para una descripción detallada del mismo). Los resultados se muestran en la tabla 5.2.

Cromosoma	<i>FB</i>	<i>NOF</i> enteros	<i>NOF</i> trunc. 3'	<i>NOF</i> trunc. 5'	Tamaño (bp)
X	8	0	1	1	22.626.939
Xh	0	0	0	0	
2L	2	0	0	0	47.815.885
2Lh	0	0	0	0	
2R	1	0	0	0	
2Rh	0	0	0	0	
3L	4	1	0	1	57.521.608
3Lh	2	0	0	0	
3R	3	0	1	0	
3Rh	0	0	0	0	
4	0	0	0	0	1.351.857
Yh	0	0	0	0	347.038
U	1	0	0	1	10.049.037
	21	1	2	3	139.712.364

Tabla. 5.2. Relación de los elementos *FB* y *FB-NOF* en el genoma de *D. melanogaster*. Según el script `cazaFB.pl`. Tamaño de cada cromosoma según lanzamiento 5.1.

El programa identificó 21 elementos *FB*, 6 de los cuáles contenían secuencias *NOF*. Notablemente, ninguno de estos elementos *NOF* está completo. Aunque la secuencia de uno de los presentes en 3L muestre el inicio de sus dos IR, y sea considerado “completo” en la Tabla 5.2, la parte central del elemento se encuentra ausente. El análisis de todos los elementos encontrados (ver las secuencias completas en material suplementario) confirma la validez de nuestra estructura consenso para *FB*.

Para un elemento tan grande y heterogéneo como *FB*, no es posible deducir una secuencia consenso, más allá de la región no repetitiva y la idealización de un bloque de

repeticiones. Aún así, se alinearon estas regiones no repetitivas de los IR 5' de todos los elementos encontrados, para tener una estimación de la variabilidad de su secuencia y se observaron frecuentes SNPs (polimorfismos de un solo nucleótido, del inglés *Single Nucleotide Polymorphism*), pequeñas deleciones e inserciones, que nos sugirieron que para *FB*, quizás es más importante la estructura que la precisión en su secuencia (al fin y al cabo, no se ha podido detectar ninguna pauta de lectura en sus IR) (ver Fig. 5.13, aunque está ampliada con elementos no procedentes de *cazaFB.pl*). El alineamiento de las regiones no repetitivas de los IR 3' no se llevó a cabo, pues nuestro objetivo no era la cuantificación de la variabilidad. Una consecuencia de las diferencias de secuencia observadas entre elementos, es que *cazaFB.pl* podría no haber reconocido todos los elementos presentes en el genoma, pues el programa es muy estricto con la secuencia de referencia (los primeros y los últimos 27bp de la secuencia del elemento *FB* en M115, ver Materiales y métodos).

Afortunadamente, en diciembre de 2002 Joshua S. Kaminker, del grupo de la Dra. Susan Celniker, publicó un estudio sobre la presencia de elementos transponibles en el tercer lanzamiento de la secuencia genómica de *D. melanogaster* (Kaminker *et al.*, 2002). Su análisis nos sirvió para probar la fiabilidad de *cazaFB.pl*, enfrentado a las más modernas herramientas bioinformáticas.

Kaminker encuentra 32 elementos *FB*, 11 más que *cazaFB.pl*. Sin embargo, la revisión manual de sus secuencias descubre algunos errores, según nuestro criterio. Algunas de ellas no pueden considerarse más que restos de antiguas inserciones, puesto que guardan solo alguna similitud con la estructura ordenada de *FB*. Otras son los dos IR de un mismo elemento, equivocadamente identificados como dos elementos distintos. El error se deduce de la coincidencia en la localización citológica y la secuencia de la duplicación directa de 9bp generada por la inserción del elemento. Finalmente, sólo 6 secuencias de las descritas por Kaminker fueron incorporadas en nuestro análisis por tener polimorfismos en los primeros 27bp e incumplir, así, con los criterios de búsqueda de *cazaFB.pl*. La tabla 5.3 muestra la relación de todos los elementos encontrados.

Con este conjunto de 27 elementos *FB*, 6 de ellos *FB-NOF*, pudimos realizar una comparación de los contextos genómicos, las secuencias de inserción y estudiar la integración de *NOF* dentro de *FB*.

RESULTADOS Y DISCUSIÓN

INSERCIONES		indel ^a y/o cambios de base	NOF	Localización ^b				Gen más cercano	Cito.
cazaFB.pl	Kaminker FB			5'	i	3'	0		
AE003423	FBti0019542		NO		x			ph-p	X
AE003426	FBti0019540		NO	x				CG4116	
AE003428	FBti0019562		NO				x	Rala	
AE003445	FBti0019607		NO		x			rdgA	
AE003486	FBti0019629		NO				x	Ptp10D	
AE003512	FBti0019643+FBti0019645		SÍ (163bp)		x			kek5	
AE003569	FBti0019675		NO				x	CG32511/CG14579	
AE003574	FBti0019725+FBti0019727		SÍ (163bp)		x			CG17600	
AE003669	FBti0019826		NO				x	betaInt-nu	2L
AE003781	FBti0019846		NO		x			CG31619	
AE003782	FBti0019853	IR3' G13A	NO	x				CG3651	
AE003783	FBti0019855	3' cortado IR5' G18T	NO	x				CG5929	
AE003787	FBti0019903	IR5' G18T	NO	x				Ogt	2R
AE003839	FBti0019976		NO	x				CG14762	
AE003481	FBti0020034		NO		x			cip4	3L
AE003557_1	FBti0020064		NO		x			nmo	
AE003557_2	FBti0020065+FBti0020067		SÍ (526bp)		x			nmo	
AE003592	FBti0020180+FBti0020182		SÍ (992bp)		x			Pka-R1 (5'UTR)	
AE003598	FBti0020208	3' cortado del IR5' 17-22	NO		x			CG7369	
AE003599	FBti0020219	ins IR3' 12T IR5' G18T IR3' C19A	NO		x			nrm	
FB3Lh1*	-		NO					-	3Lh
FB3Lh2*	-		NO					-	
AE003735	FBti0020382+FBti0020384		SÍ (163bp)				x	E2f	3R
AE003739	FBti0020385		NO		x			CG31160	
AE003768	FBti0020392		NO				x	cul5 / CG11873	
AE003843	FBti0020425	IR5' G18T	NO		x			Pur-alpha	4
FBU1*	-		SÍ (234bp)					-	Un

Tabla 5.3. Los elementos *FB* y *FB-NOF* en el genoma de *D. melanogaster*.

(a) Indel: inserciones y/o deleciones

(b) Localización: 5' : menos de 5Kb del extremo 5' del gen más cercano o UTR 5'

i : en un intrón

3' : menos de 5Kb del extremo 3' del gen más cercano o UTR 3'

0 : aislado. Más de 5Kb del gen más cercano

IR5'/IR3' Puesto que *FB* no codifica ninguna proteína, la notación 5' y 3' solamente indica izquierda y derecha, según se muestra en la secuencia.

(*) Elementos detectados en secuencias heterocromáticas o sin posición cromosómica asignada. No se dispone de número de acceso u otra información.

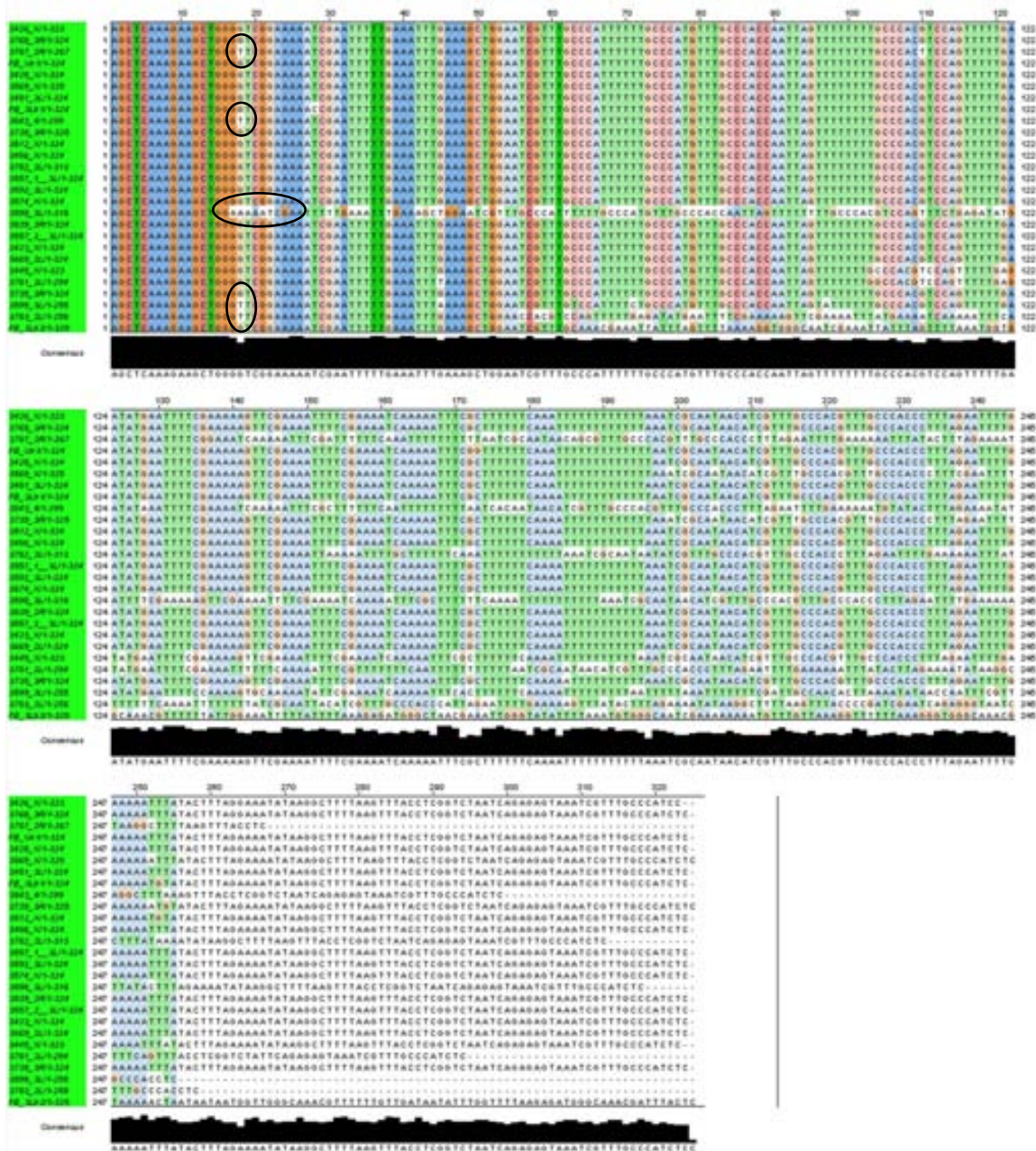


Fig. 5.13. Las regiones no repetitivas 5'. Alineamiento de las regiones no repetitivas de los IR 5' de los elementos *FB* genómicos. La intensidad de color indica el grado de conservación de la secuencia. El gráfico inferior muestra la secuencia consenso calculada por ClustalW. Los círculos negros muestran algunos de los cambios de base en los primeros 27bp, los utilizados por *cazaFB.pl*.

La inserción de *FB* se produce en contextos de fácil acceso

Puesto que disponíamos de la localización precisa de los elementos *FB* y *FB-NOF* presentes en las secuencias eucromáticas, investigamos su posición en relación a los genes cercanos para determinar si había alguna preferencia de inserción. El 54,2% de los elementos se encuentran en secuencias intrónicas, el 20,8% cerca del extremo 5' de genes

y el 25% restante se encuentran aislados, lejos del área de influencia de secuencias codificantes o sus regiones reguladoras. Notablemente, no se encontraron inserciones cercanas al extremo 3' de ningún gen (ver Tabla 5.3). En este trabajo definimos como “área de influencia” de un gen a la región de 5Kb inmediatamente anterior y posterior a su inicio de transcripción y señal de poliadenilación, respectivamente. Esta distribución coincide con la de otros elementos transponibles y retrovirus, como *P*, *MLV*, *AAVr* y *HIV-1* (Wu y Burgess, 2004).

En cuanto a la abundancia de elementos *FB-NOF* por cromosoma, esta no se corresponde con la descrita para la mayoría de elementos transponibles. Según Bartolomé y colaboradores (Bartolome *et al.*, 2002) existe un déficit de inserciones en el cromosoma X, sobretodo de transposones de clase-II, pero sus resultados ya indicaban que la densidad de elementos *FB-NOF* es mayor en el cromosoma X. Sin embargo, su trabajo se hizo con los datos genómicos del segundo lanzamiento, que contaba un total de 35 elementos *FB(-NOF)*, algo sobreestimado según nuestro presente análisis (de hecho, el tercer lanzamiento ya solamente contaba 32 elementos, ver Kaminker *et al.*, 2002). Nuestros resultados sobre un total de 27 elementos respaldan este sesgo (ver Tabla 5.4), con una densidad de 0,35 elementos por megabase en el cromosoma X, 2,9 veces mayor que en el cromosoma 2 y 1,8 veces mayor que en el cromosoma 3. En este análisis no se ha tenido en cuenta los cromosomas 4 e Y, pues su número de inserciones no permite calcular una densidad. Así pues, a diferencia de la mayoría de transposones, parece que *FB-NOF* tiene preferencia por insertarse en el cromosoma X.

Cromosoma	Nº Inserciones	Inserciones/Mb	Tamaño (bp)	Porcentaje genoma
X	8	0.35	22.626.939	16%
2	6	0.12	47.815.885	34%
3	11	0.19	57.521.608	41%

Tabla 5.4. Distribución de elementos *FB-NOF* en el genoma de *D. melanogaster*.

Otro dato de interés biológico en el análisis de un elemento transponible es la secuencia de inserción. Definir una secuencia consenso para la inserción de *FB* sería, sin duda, un gran descubrimiento. Sin ningún dato que pudiera servir de pista, asumimos que la

duplicación directa de 9bp que genera la inserción de *FB* es la secuencia de reconocimiento del transposón, de modo que las comparamos entre todos los elementos genómicos. El cálculo de frecuencias de cada base en cada posición reveló que no existen preferencias (ver Fig. 5.14B), de modo que concluimos que *FB* no posee un mecanismo de inserción específico de secuencia. Sin embargo, al tomar las propiedades de los nucleótidos (Fig. 5.14A), se observa una fuerte tendencia a las interacciones débiles (A/T), excepto en la primera posición de la secuencia de reconocimiento que es C/G en el 67% de los casos (en particular G, en el 48,1% de los casos analizados).

Para reforzar el análisis de la secuencia de inserción, se buscaron posiciones conservadas en los alrededores mediante la elaboración de un logo de secuencia (Fig. 5.14C). Los logos de secuencia proporcionan el valor informativo de cada posición de una secuencia de un modo más preciso que una secuencia consenso o una tabla de frecuencias como la de la Fig. 5.14B, pues muestra todos los nucleótidos presentes en cada posición y su relevancia en bits de información. Cada posición en una secuencia de DNA contiene 2 bits (para una explicación más detallada de los logos de secuencia y su interpretación, ver Shaner *et al.*, 1993). En el logo se incorporaron las 20 bases adyacentes a la secuencia de 9bp que constituye la posible secuencia de reconocimiento de *FB*.

El resultado confirma que no existen posiciones especialmente conservadas en los puntos de inserción de *FB*, aunque también se aprecia un sesgo a favor de una composición rica en A/T (ver Fig. 5.14C). Las posiciones en las que el logo no muestra ninguna base son las que tienen un equilibrio entre las cuatro, lo cual resulta en cero bits de información para cada nucleótido.

La conclusión que se desprende del análisis de las secuencias del punto de inserción de los *FB* y *FB-NOF* en secuencias genómicas, es que éstos tienden a insertarse en regiones potencialmente activas del genoma y, por lo tanto, con una conformación de la cromatina accesible durante las fases de movilización del transposón. Adicionalmente, prefieren secuencias “débiles” donde abrir la hebra de DNA y eventualmente cortarla, con mayor facilidad e insertarse.

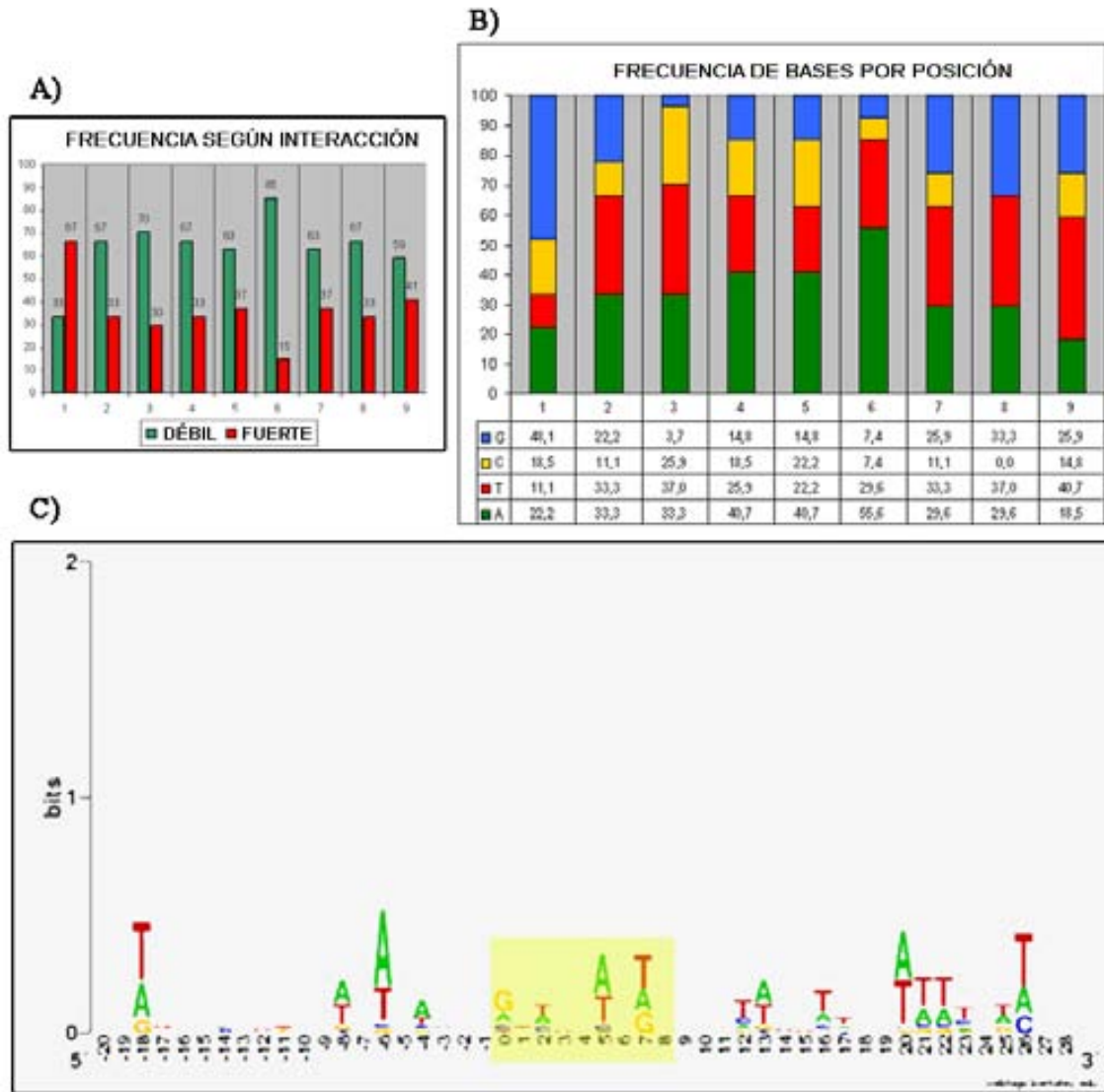


Fig. 5.14. Análisis de la secuencia de inserción de *FB*. Se muestran las frecuencias nucleotídicas en cada posición según **A)** interacción y **B)** base. **C)** Logo de secuencia a partir de todas las zonas de inserción y sus alrededores. La caja resalta los 9bp que forman la duplicación directa.

La posición de *NOF* dentro de *FB* es específica de secuencia

Otro de los aspectos no resueltos en la biología de *FB-NOF* es la relación que guardan entre ellos las dos partes del elemento: *FB* y *NOF*. Aunque todos los encontrados en las secuencias genómicas tienen la parte *NOF* truncada, podemos comparar la posición que ocupa dentro de *FB*, pues disponemos de sus extremos secuenciados. Esta información está libre de errores por parte de los algoritmos de ensamblaje, siempre que nos limitemos a observar la posición de *NOF* dentro de un bloque de repeticiones.

Disponemos de 9 secuencias en las que se observa la transición *FB-NOF*, los 6 elementos encontrados en secuencias genómicas (una de las cuales, AE003557_2, cuenta con las transiciones en los dos IR de *FB*) y nuestra secuencia perteneciente al elemento

estudiado en la cepa M115, también con las dos transiciones secuenciadas. Estas secuencias se alinearon en dos bloques: las directas, llamadas 5' a falta de una nomenclatura más adecuada, y las reversas, llamadas 3'. El resultado del alineamiento se muestra en la figura 5.15.

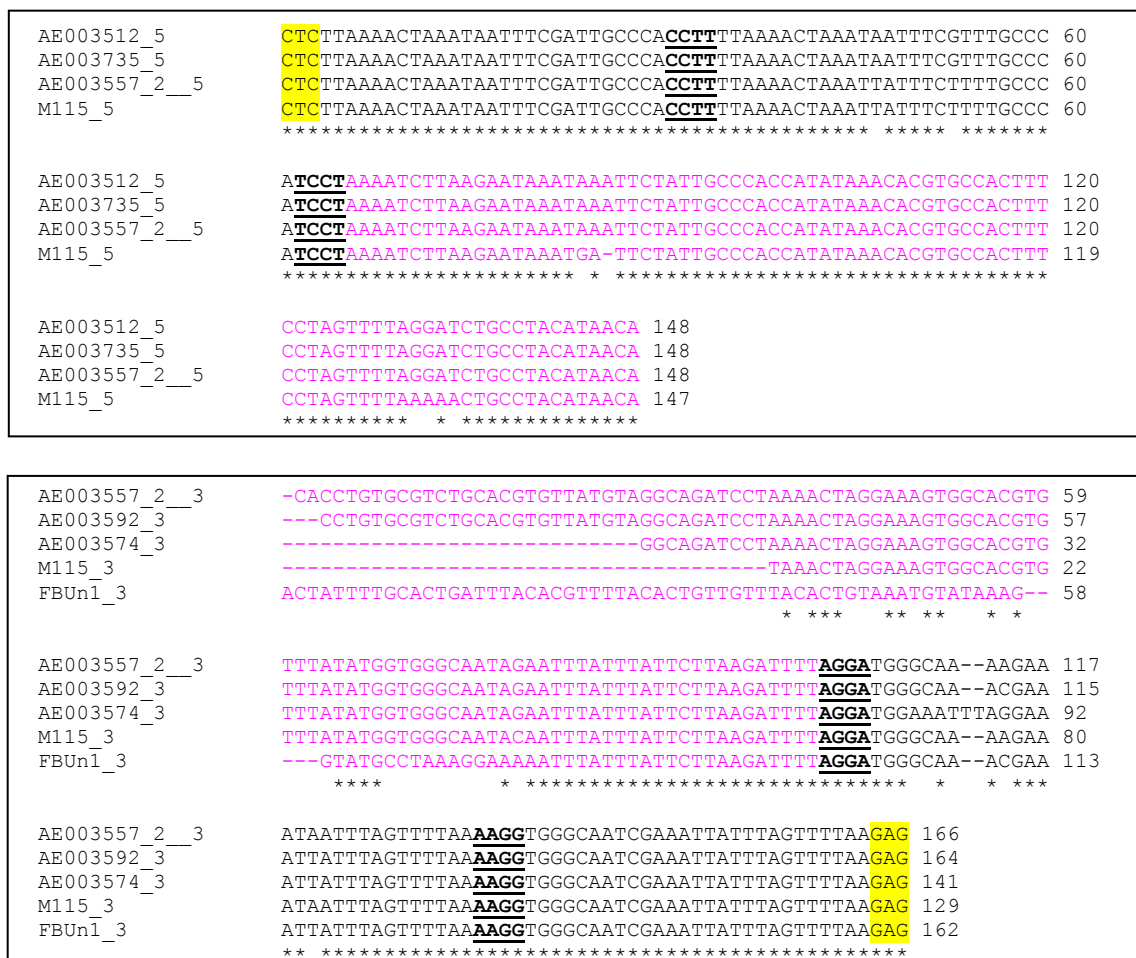


Fig. 5.15. Alineamiento de las transiciones *FB-NOF*. Las 4 transiciones 5' y las 5 transiciones 3' se alinearon con la herramienta *online* ClustalW. Resaltado en amarillo se muestran los inicios de bloque. Las secuencias subrayadas muestran los finales de repetición (ver Fig. 5.11). Los fragmentos fucsia son secuencias pertenecientes a *NOF*.

Sorprendentemente, pues *a priori* no hubiéramos apostado por ello, todos los elementos analizados se encontraron perfectamente alineados, posicionados después del final de una segunda repetición (TCCT) de un bloque, ya sea por el extremo 5' como el 3'. De este resultado se desprenden algunas conclusiones muy interesantes.

La posición de *NOF* dentro de *FB* es específica de secuencia, lo cual sugiere dos alternativas:

1. *FB* y *NOF* tienen un mecanismo enzimático compartido. Este hipotético sistema enzimático catalizaría la introducción de *NOF* entre los IR de *FB* mediante el reconocimiento de la secuencia distintiva de una segunda repetición en cada IR y efectuando dos cortes de doble cadena. Así conseguiría que la inserción de las secuencias *NOF* coincidieran, en los dos IR, con un final de segunda repetición. Según esta hipótesis, cualquier fragmento interno de *FB*, entre los dos puntos elegidos para el corte, sería eliminado. Esta hipótesis puede explicar un fenómeno propio y distintivo de *FB-NOF*: la asimetría. Puesto que los bloques de repeticiones de los IR de *FB* son esencialmente idénticos, la inserción de *NOF* podría darse en cualquier final de segunda repetición, en cualquier bloque. Si los dos procesos de corte y empalme fueran independientes, uno para cada extremo de *NOF*, el número de bloques restantes desde el inicio del IR en 5' y en 3' no tendría porqué coincidir. Esto implicaría que, a la práctica, *FB* y *NOF* tendrían una relación de simbiosis en la que *NOF* proveería el sistema enzimático y *FB* un sitio de inserción "seguro". Un punto débil de esta hipótesis es la falta de duplicación directa, u otro rastro, en el punto de inserción de *NOF* dentro de *FB*. Otro punto en contra es la total ausencia de secuencias ancestrales de *FB* y *NOF* autónomos en el genoma, que podrían demostrar que ambos elementos fueron independientes en el pasado. Además, nunca se ha descrito la transposición o escisión de secuencias *NOF* entre elementos *FB*.
2. *FB-NOF* es un solo elemento transponible con unos grandes IR en tres fases. Una primera región no repetitiva externa, de 470bp. Una segunda región formada por bloques de repeticiones. Y finalmente, una tercera región no repetitiva interna, de 326bp. Entre los dos IR se encuentra la zona codificante del elemento. Esta hipótesis no requiere de la inserción de *NOF* dentro de *FB*, explicando porqué no se observa ningún rastro de la misma en el punto de transición. Además, la unificación de las dos secuencias en un solo elemento resolvería el misterio de la transposasa de *FB*, que estaría codificada en lo que actualmente llamamos *NOF* y sería la responsable de la duplicación directa de 9bp característica de la inserción del elemento. Respalda esta hipótesis que la única movilización descrita de un elemento de esta clase, fuera de un elemento *FB-NOF* completo y por escisión precisa en la duplicación de 9bp (Collins y Rubin, 1983). Sin embargo, las diferencias en el número de bloques en la porción repetitiva del IR debería explicarse mediante deleciones por recombinación. Bajo esta hipótesis, los elementos *FB-NOF* sin secuencias *NOF* serían elementos

estrictamente defectivos. Esta organización sería parecida a la descrita en 1997 para el elemento SoFT del tomate (*Lycopersicon esculentum* cv. VFNT Cherry) (Rebatchouk y Narita, 1997).

La relación descrita entre *FB* y *NOF* es aplicable a cualquier elemento que quiera analizarse en el futuro. En particular, se observa perfectamente en las secuencias de Templeton y Potter (1989) y Harden y Ashburner (1990), excluidas del análisis general expresamente porque la transición *FB-NOF* es una porción muy marginal de la secuencia y dudábamos de su fiabilidad. Sin embargo, su análisis *a posteriori*, muestra total coincidencia con el modelo propuesto (ver Fig. 5.16).



Fig. 5.16. Predicción de las transiciones *FB-NOF*. Las transiciones 5' y 3' de los elementos *FB-NOF* secuenciados por Templeton (1989) y Harden (1990) alineadas con la secuencia de *FB* utilizada como modelo en el trabajo (ver Fig. 5.11) y la de *NOF* (en fucsia) según el elemento en M115. Los fragmentos subrayados corresponden a finales de repeticiones.

Los *NOF* en las bases de datos están truncados. ¿Lo están en M115?

La constatación de que *FB* y *NOF* están íntimamente relacionados nos llevó a la siguiente fase de estudio del elemento: la comparación de las secuencias genómicas de *NOF*. Sin embargo, como se ha mencionado anteriormente, todos los elementos *FB-NOF* encontrados en las secuencias genómicas están truncados. Quizás se trate de un artefacto del proceso de ensamblaje o bien a un defecto de representatividad en el conjunto de fragmentos que se secuenciaron. Al fin y al cabo, las secuencias *NOF* se encuentran aisladas entre los IR de *FB*, propensos a formar estructuras secundarias que pueden dificultar su clonación. Por supuesto, es perfectamente posible que la cepa usada para la secuenciación genómica tenga todos los elementos *NOF* truncados. Harden y Ashburner (1990) ya observaron la variabilidad de *FB* y *NOF* entre cepas, clasificándolas entre las que no muestran hibridación con la sonda de *NOF* (la cepa salvaje Oregon-R y las marcadas como [*net pr pk cn sp/CyO*] y [*Sp b pr Coi Pin/CyO*]), cepas con alguna hibridación débil (CS, [*ve se cp e*]) y cepas con diversas señales de hibridación (*Basc*, entre

8 y 14 sitios *NOF* según procedencia del *stock*). Sea como fuere, puesto que no disponemos de la cepa utilizada para la secuenciación del genoma de *D. melanogaster*, no podemos apoyar ninguna de las dos alternativas. Sin embargo, decidimos analizar los elementos *FB-NOF* en el genoma de nuestra cepa mutante M115 para aportar más datos y, adicionalmente, determinar su grado de integridad.

Primero, un sencillo experimento de *Southern blot* nos sirvió para determinar el número de elementos *FB* y *FB-NOF* presentes en M115. DNA genómico digerido con la enzima *SaII*, que no corta en *FB* y una sola vez *NOF*, fue hibridado con las sondas pFB y pNOF (ver Fig. 5.17A y C). Se esperaban bandas de distinto tamaño para cada elemento, según la distancia de las dianas *SaII* externas. El resultado muestra entre 25 y 30 señales de hibridación con la sonda pFB, aunque algunas podrían representar más de una banda del mismo tamaño, a juzgar por las diferencias de intensidad de las mismas. Tomando en cuenta esta consideración, creemos que una previsión de alrededor de 40 copias sería más realista. La hibridación con la sonda pNOF, por su parte, muestra 6 bandas claras, más algunas sombras débiles de las que no podemos estar muy seguros. Esto sitúa la cepa M115 en el grupo de contenido medio-alto tanto en *FB* como en *NOF*.

Para determinar la integridad de estos 6 elementos, se llevó a cabo otro experimento de *Southern blot*. En esta ocasión, DNA genómico de M115 fue digerido con la enzima de restricción *BmgBI*, que corta en los dos extremos de *FB*, aislando cualquier elemento *FB* o *FB-NOF* de su contexto genómico (ver Fig. 5.17C). Así, ordenamos cada uno de los elementos presentes en el genoma estrictamente por su tamaño. Evidentemente, aquellos que tengan el mismo tamaño no se separarán en la electroforesis pero el objetivo del experimento, en esta ocasión, no es obtener un recuento. Se utilizaron tres sondas: la hibridación con la sonda pFB muestra los elementos *FB* o *FB-NOF* según su tamaño, la sonda pNOF5 muestra todos los elementos *FB-NOF* que contengan la parte 5' de *NOF* (entre la primera y segunda Kb) y, finalmente, la hibridación con la sonda pNOF3 detecta todos los elementos *FB-NOF* que contengan la parte 3' de *NOF* (de la segunda a la cuarta Kb). Se espera que aquellos elementos que estén truncados solamente hibriden con la sonda pFB y una de las pNOF, dependiendo de la región ausente.

Los resultados de la hibridación con pFB evidencian la naturaleza heterogénea de este transposón, aunque se pueden distinguir dos grandes grupos: los mayores de 8Kb y los menores de 6Kb (ver Fig. 5.17B). Todos los del primer grupo hibridan con las dos sondas de *NOF*, de modo que son elementos *FB-NOF* completos. En el segundo grupo, sin

embargo, la mayoría no contiene secuencias *NOF* detectables, excepto dos que hibridan con la sonda pNOF5 pero no con la pNOF3, indicando que la región 3' de *NOF* está ausente.

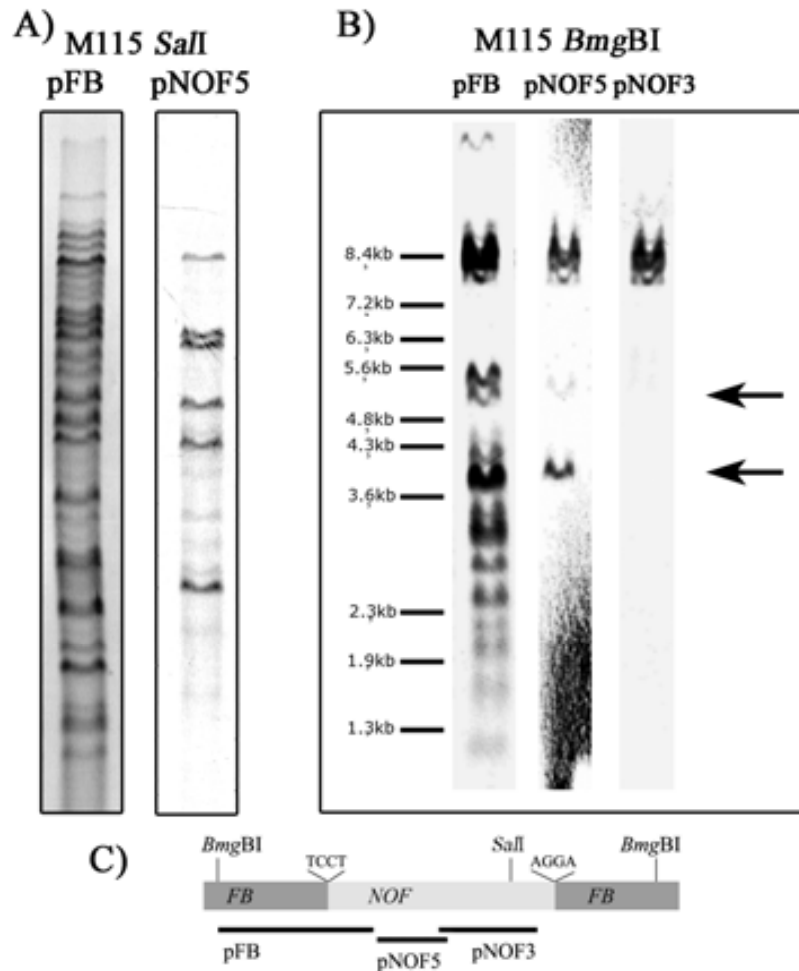


Fig. 5.17. La integridad de los elementos *FB-NOF* en M115. A) Número de elementos *FB* y *FB-NOF* en la cepa M115. Los dos experimentos no guardan la misma escala B) Integridad de los *FB-NOF* mediante la discriminación de los elementos según su hibridación con las sondas pFB, pNOF5' y/o pNOF3'. Las flechas muestran la posición de los elementos truncados. C) Esquema de *FB-NOF* con las posiciones de las sondas y las dianas de restricción utilizadas. El esquema no está a escala.

El resultado de estos experimentos muestra que M115 pertenece al grupo de cepas con un número elevado de elementos *FB-NOF* y que la mayoría, por lo menos dos terceras partes, están completos. La cepa utilizada por el Consorcio para la Secuenciación del Genoma de *Drosophila* contiene 27 elementos *FB*, 6 de los cuales son *FB-NOF* truncados. Estos datos concuerdan con los obtenidos por Harden y Ashburner (1990) para cepas con escasa presencia de *NOF* (20-30 copias de *FB*), pues cepas con 8-14 elementos completos suelen tener entre 40 y 50 copias de *FB*. Puesto que en los últimos esfuerzos de finalización de secuencias genómicas, en especial en el cuarto y quinto lanzamientos, las

secuencias internas de *FB* han mejorado notablemente, creemos que la ausencia de elementos *NOF* enteros podría no ser debida a artefactos de la clonación y/o secuenciación *shotgun*, sino a la propia naturaleza de la cepa.

***NOF* tiene entre 1 y 3 ORFs que codifican proteínas únicas**

La presencia de diversas copias enteras de *FB-NOF* en algunas cepas sugiere que el elemento podría estar activo en la actualidad. Con el fin de comprobarlo, diseñamos algunos experimentos para evaluar la funcionalidad de *NOF*, como ya hicieron Templeton y Potter en 1989. La evaluación funcional que realizamos constó de dos fases, una bioinformática y otra experimental.

Como se describe en la introducción, existe cierta discrepancia sobre el número de ORFs que contiene *NOF*. Esto es debido, en gran medida, a que sólo existen dos secuencias (casi) completas del elemento, la de Templeton y Potter (1989) y la de Harden y Ashburner (1990) y discrepan en una pequeña región de 5-6bp. Antes de empezar con cualquier predicción bioinformática con herramientas modernas, decidimos esclarecer esta discrepancia mediante un sencillo experimento de PCR y secuenciación.

Una región de 700bp alrededor del punto de discrepancia fue amplificada a partir de DNA genómico de machos de las cepas CS, M115 y RM115 (ver Fig. 5.18). Como ya pasó en otras ocasiones, la cepa CS no amplificó ningún fragmento de *NOF*, aunque justamente Harden y Ashburner (1990) describieron que su genoma contenía una copia del elemento. El fragmento de PCR, en las cepas que sí amplificaron, fue purificado y secuenciado directamente, sin clonar, para tener una visión general representativa de todas las copias de *NOF* presentes en el genoma de cada cepa. Así, en el caso que hubiera distintas familias de elementos, podríamos observar la mezcla reflejada en el cromatograma resultante de la secuenciación (Fig. 5.18).

Aunque el cromatograma muestre algunos picos poco definidos, pues la zona a secuenciar se encuentra muy próxima al cebador NOF2740 utilizado para la secuenciación, los picos puros demuestran que no existen discrepancias de secuencia en nuestras cepas portadoras de *NOF*. Todos los elementos, o en todo caso la gran mayoría, contienen la secuencia descrita por Harden y Ashburner (1990): AAACAC.

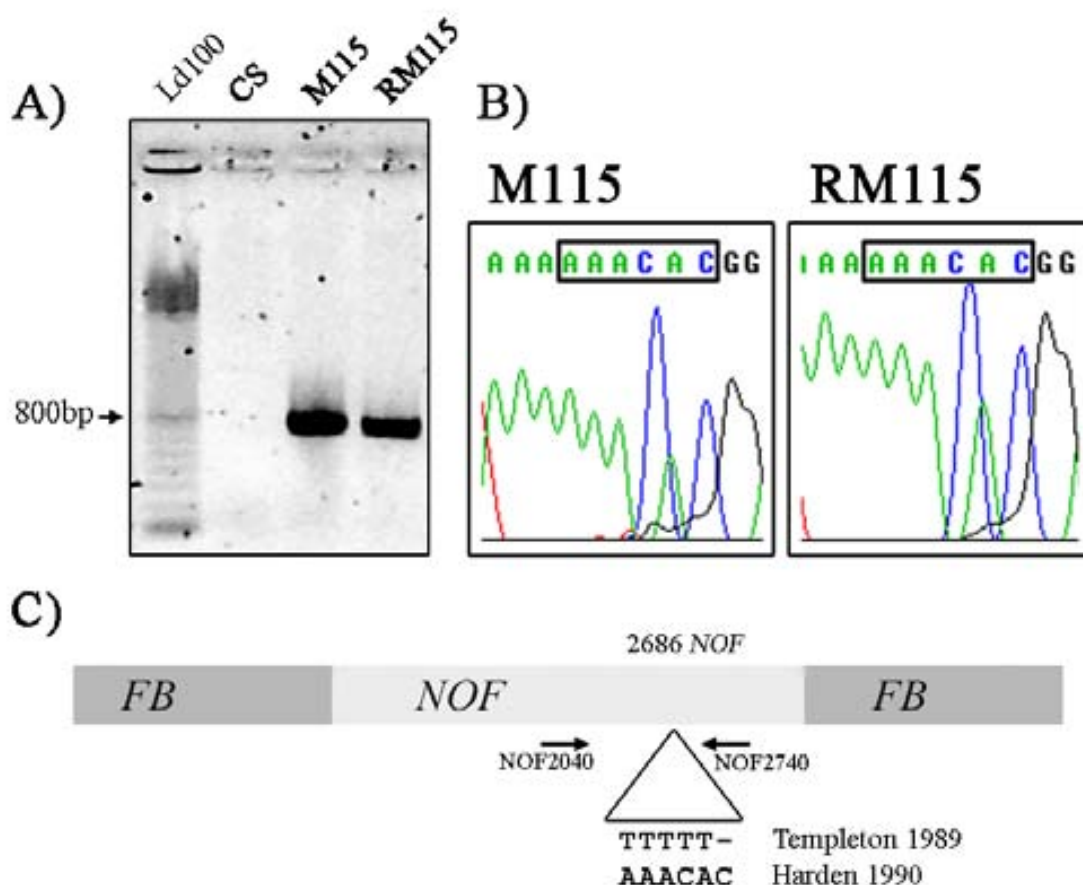
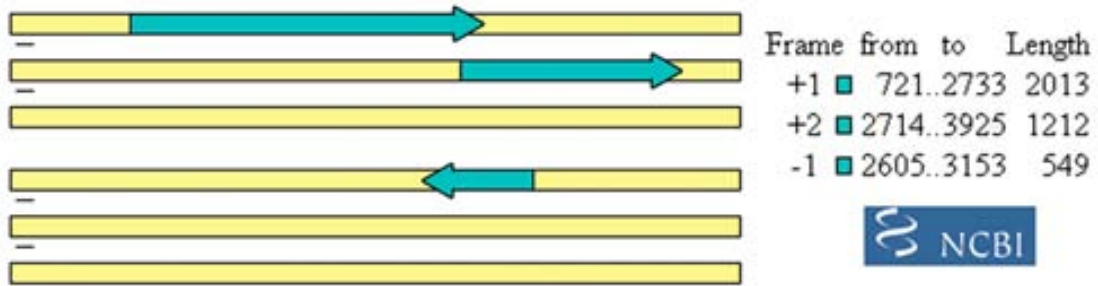


Fig. 5.18. Secuenciación del punto de discrepancia en la secuencia de *NOF*. **A)** El fragmento de 700bp que contiene el punto de discrepancia, en las cepas CS, M115 y RM115. Se muestra la banda de 800bp en el marcador de peso molecular **B)** Detalle de los cromatogramas de secuenciación de M115 y RM115 en el punto de discrepancia. **C)** Esquema de *FB-NOF* con los cebadores utilizados en el experimento y las alternativas en el punto 2686 de *NOF*, según la bibliografía.

Una vez tuvimos una secuencia firme con la que trabajar, utilizamos las más modernas herramientas bioinformáticas para el análisis de secuencias nucleotídicas y la predicción funcional de proteínas para conseguir extraer la mayor cantidad de información posible de la secuencia antes de planear experimentos concretos con los que probar las predicciones. En particular, nos interesaba poder descartar una de las dos hipótesis de partida sobre el número de ORFs de *NOF*. Sin embargo, como suele ocurrir, obtuvimos resultados contradictorios en función del *software* utilizado (ver Fig. 5.19).

PREDICCIÓN ORF FINDER



PREDICCIÓN GENSCAN

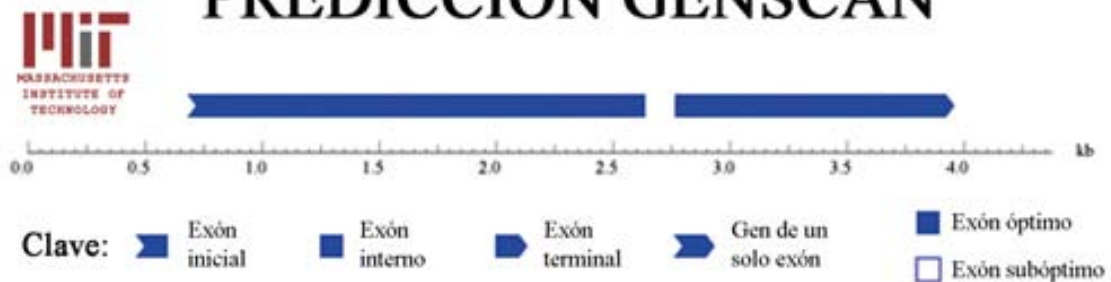


Fig. 5.19. Predicción de ORFs en la secuencia de *NOF*. Se muestra el resultado del uso de dos programas de predicción de pautas de lectura, *ORF finder* y *GenScan*. En el primero, la secuencia de *NOF*, en las 6 fases posibles de lectura, se muestra como barras amarillas. Las zonas azules corresponden a los ORF detectados. Se muestran los puntos de inicio/final y longitud de cada uno. En el segundo, la regla corresponde a *NOF* y la flecha al ORF detectado, con un intrón.

Aún analizando los mismos datos, *ORF finder* predice tres ORFs, mientras que *GenScan* apunta a la existencia de solo un gran ORF. Las tres pautas abiertas de lectura propuestas por *ORF finder* guardarían un gran parecido con las predicciones de Templeton y Potter en 1989, con la diferencia que ORF-1 es 110bp más largo debido al corrimiento de lectura introducido en la corrección de la secuencia. En este sentido, es interesante notar que ORF-1 y 2 tendrían un solapamiento de 20bp, algo inusual pero sin efectos, teóricamente, pues se encuentran en distintas fases. Los ORF-2 y 3 serían exactamente iguales a los propuestos en 1989, con el ORF-3 en sentido inverso entre los ORF-2 y 1 (ver Fig. 5.19). *GenScan*, por su parte, predice un gran ORF de 3078bp, muy parecido al de Harden y Ashburner (1990), pero con un intrón de 128bp en la zona de transición entre los ORF-1 y 2 de las predicciones de *ORF finder*. Para mantener la homogeneidad de los nombres, al ORF predicho por *GenScan* lo nombramos ORF-0.

Puesto que no teníamos base para descartar ninguna de las dos alternativas de predicción, decidimos pasar al análisis funcional de todos los ORF predichos. La tabla 5.5

presenta el resumen de los resultados obtenidos de los distintos programas usados. En el apartado Anexos se encuentran los informes completos de cada uno de los ORF analizados.

	Tamaño bp/aa/KDa	Mayor similitud ^a (Id, Po, P)	Dominios conocidos <i>PROSITE</i>	Localización Celular ^b	Dominios transmembran- ales ^c	Soluble/ Transmembra- na ^d
ORF-0	3078bp/1025aa/117KDa	PMS1_HUMA N (Id25%, Po45%, P=0,14)	0	Anclada/ Intracelular	3	Soluble
ORF-1	2013bp/670aa/76KDa	N4BP3_MOUS E (Id37%, Po48%, P=0,95)	0	Anclada/ Extracelular	2	Soluble
ORF-2	1212bp/403aa/46KDa	Q7S6R0_NEU CR (Id34%, Po51%, P=0,033)	0	Intracelular/ Anclada	1	Soluble
ORF-3	549bp/182aa/20KDa	Q04VW2_LEP BJ (Id27%, Po51%, P=0,76)	0	Transmembra na/ Nuclear	2	Soluble

a) *WU-Blast2* (a.a. seq). Id=identidades. Po=positivos. P=probabilidad. b) *ProtLoc*. c) *TransMem*. d) *SOSUI*.

Tabla 5.5. Resumen de los resultados del análisis de los ORFs de *NOF*. Se muestran los datos más relevantes sobre cada ORF. La búsqueda en *Prosit* no tuvo en cuenta los motivos más frecuentes. En el resultado de *ProtLoc*, solamente se muestran las dos primeras posiciones. En la predicción transmembranal se indica solamente el resultado de *TransMem*.

El análisis funcional de los ORF fue, a lo sumo, poco revelador. La comparación de las secuencias de aminoácidos con la base de datos *Swiss Prot* no muestra ninguna similitud relevante con proteínas conocidas. Los mejores resultados para cada ORF se muestran en la Tabla 5.5. Aunque los porcentajes de positivos alcanzan valores de hasta el 51%, las identidades son bajas y el valor P de los alineamientos es muy pobre en todos los casos ya que *Blast* solo consigue alinear pequeños fragmentos. La búsqueda de motivos proteicos conocidos en las secuencias, almacenados en la base de datos *ExPasy-Prosit*, dio el mismo resultado, no pudiéndose detectar más que algunos puntos de fosforilación y glicosilación con poco valor predictivo al analizar un péptido completamente desconocido.

Puesto que las comparaciones de secuencia no dieron información útil, pasamos a usar programas más específicos, aunque también más arriesgados pues usan modelos matemáticos para predecir las propiedades de los péptidos. Los programas desarrollados por el grupo del Dr. Querol en la Universitat Autònoma de Barcelona, *ProtLoc* y *TransMem*, fueron los primeros en predecir datos interesantes. *ProtLoc*, un software de predicción de localización subcelular mediante el análisis de los porcentajes de cada aminoácido en el péptido, indicó que ORF-0, 1 y 2 podrían dar proteínas citosólicas, ancladas a membrana o, incluso, extracelulares (ver Tabla 5.5). ORF-3, por su parte, podría

producir una proteína integral de membrana o bien nuclear. Se tuvieron en cuenta las dos primeras predicciones pues, como describen en su artículo (Cedano *et al.*, 1997), entre las dos aciertan en el 91% de los casos. El programa *TransMem* confirmó dos grandes dominios potencialmente transmembranales en ORF-3, que reforzaron la idea que pudiera ser integral de membrana. El resto de ORFs también contienen algunos dominios potencialmente transmembranales, pero son demasiado cortos (menos de 20aa) para constituir una hélice transmembrana (ver Anexos). Para confirmar la validez de los dominios predichos por *TransMem*, usamos otra herramienta de predicción: *SOSUI*. Éste analizó la hidrofobicidad de cada uno de los péptidos, en conjunto, y también de sus dominios transmembranales y les asignó un valor en la escala de Kyte y Doolittle (Kyte y Doolittle, 1982). Según esta escala, hélices con valores superiores a 1,75 se consideran transmembranales. Sin embargo, la mayoría de los dominios potencialmente transmembranales identificados por *TransMem* no superaron este umbral, sobretodo porque *SOSUI* no tiene en consideración dominios menores de 20aa. El único que sí lo hizo fue el dominio corto (25aa) de ORF-3, con un valor de 1,86. Aún así, al analizar el péptido en conjunto *SOSUI* cambia su predicción y ya no lo considera transmembranal.

Puesto que no se encontraron proteínas parecidas a nivel de secuencia ni dominios funcionales, para finalizar el estudio *in silico* se buscaron proteínas estructuralmente parecidas con el *software* *PHYRE*. Éste predice la probabilidad, para cada región de un péptido, de estructurarse como hélice alfa, lámina beta o giro y busca parecidos con otras proteínas de estructura y función conocidas. Obviamente, *PHYRE* también tiene en cuenta la similitud de secuencia como un elemento más. El supuesto es que la coincidencia estructural puede significar coincidencia funcional. Adicionalmente, *PHYRE* genera una predicción tridimensional del péptido en el formato *Protein Data Bank* (.pdb). El resultado de las comparaciones puede consultarse en el CD que acompaña el trabajo. En resumen, los ORF-0 y 1 tienen parecidos leves con proteínas estructurales de virus y también con algunas enzimas hidrolasa. ORF-2, por su parte, consigue una puntuación de $E=0,58$ y una precisión del 80%, al compararla con la endonucleasa II (EcoO109I) de *Escherichia coli*. Ambas predicciones encajan bastante bien con péptidos generados por un elemento transponible. El producto de ORF-3, sin embargo, resulta en valores muy bajos de similitud, de modo que consideramos que no existe ninguna proteína similar, a nivel de secuencia ni estructura, en las bases de datos.

En conclusión, después del análisis bioinformático de la secuencia del elemento *FB-NOF*, una vez despejada la incertidumbre en las secuencias iniciales, tenemos dos conclusiones mutuamente excluyentes. En la primera, existiría una sola región codificante de 3078bp con un pequeño intrón, que codifica una proteína de 1025aa, soluble, citosólica y con un ligero parecido a proteínas virales y a nucleasas. En la segunda, la región codificante de *NOF* contendría tres ORFs, dos de los cuales darían lugar a péptidos solubles, localizados en el citoplasma o anclados a membrana, con cierta similitud a enzimas hidrolasa y endonucleasa. El tercer ORF, en sentido inverso, codificaría una pequeña proteína soluble que podría ser nuclear y que no muestra ninguna similitud con otras proteínas conocidas.

***NOF* podría generar un mRNA dicistrónico**

La fase experimental de la evaluación funcional de *NOF* tuvo como objetivos comprobar la expresión de los ORF de *NOF* y determinar cual de las hipótesis basadas en las predicciones bioinformáticas se ajusta más a los datos biológicos recopilados. En primer lugar investigamos la expresión de *NOF* a nivel de RNA. En estos experimentos usamos las cepas M115 y RM115 indistintamente, pues ya comprobamos que ambas contienen elementos *NOF* completos (ver Fig. 5.17 para integridad de *FB-NOF* en M115 y tomar en consideración que RM115 proviene de M115).

RNA total de machos, tratado con DNAasa para eliminar falsos positivos, fue retrotranscrito a cDNA con cebadores específicos para cada ORF, sobre el cual tratamos de amplificar un fragmento que demostrara su expresión. En la tabla 5.6 se describen los cebadores, los cDNA teóricos y los ORF que pueden amplificarse.

cDNA	PCR/Tamaño		ORF
NORF1r	NORF1f vs. NORF1r	603bp	ORF-1
NORF0.2.3r	NORF0f vs. NORF0.2.3r	775bp/647bp (int)	ORF-0
	NORF2.3f vs. NORF0.2.3r	421bp	ORF-2
NORF2.3f	NORF2.3f vs. NORF0.2.3r	431bp	ORF-3

Tabla 5.6. Cebadores usados para RT-PCR. Se muestran las combinaciones de cebadores que generaron los cDNAs y las combinaciones utilizadas para las PCR. El tamaño de la banda NORF0f vs. NORF0.2.3r puede variar según la presencia o ausencia del intrón predicho por GenScan. La columna de la derecha muestra el ORF correspondiente a cada reacción.

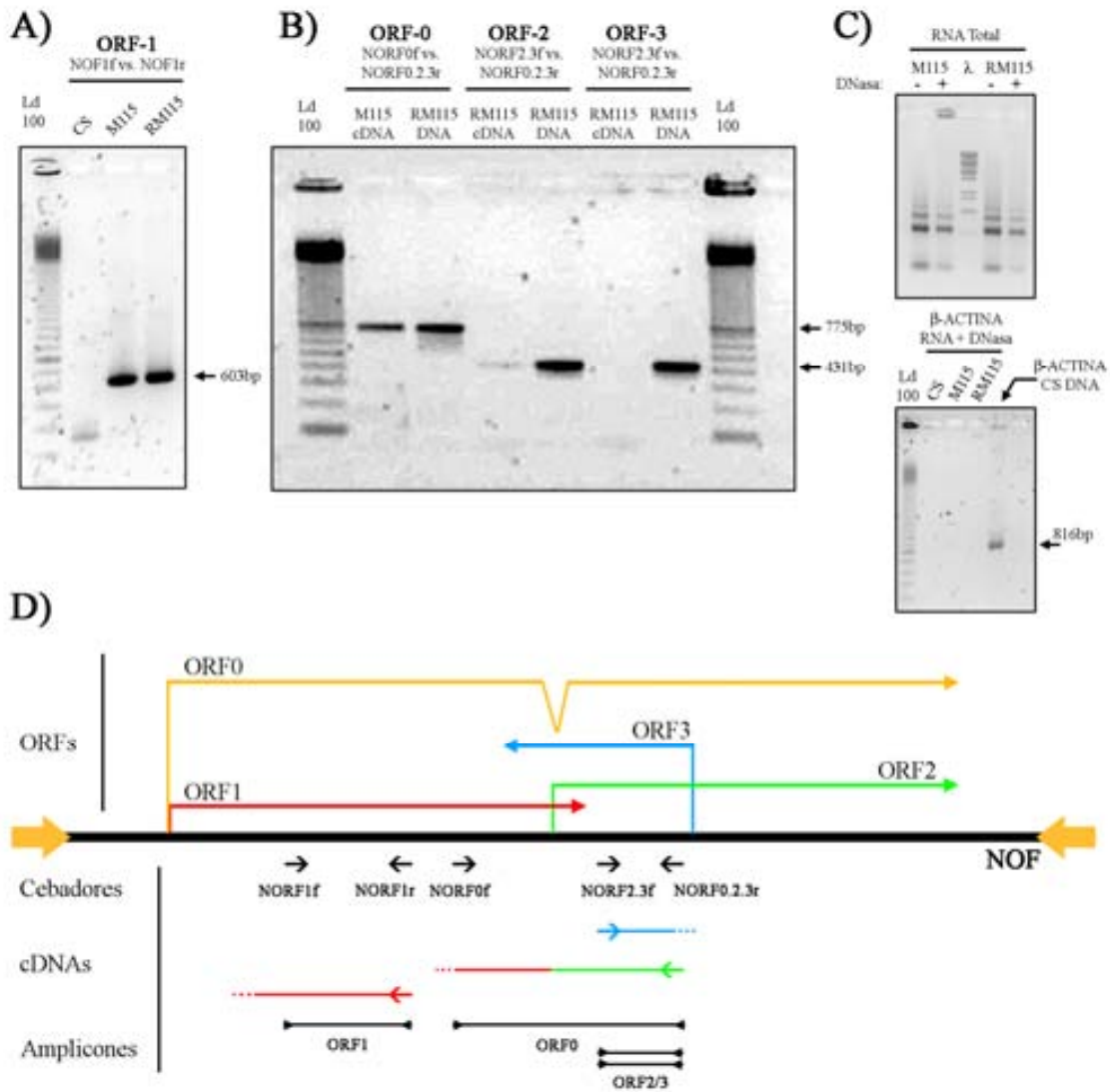


Fig. 5.20. Expresión de los ORF de *NOF*. A) RT-PCR específica de ORF-1 sobre cDNA de CS, M115 y RM115. B) RT-PCR específicas de ORF-0, 2 y 3. Cada reacción se hizo sobre cDNA y DNA genómico, como control positivo. C) Electroforesis de comprobación de la integridad del RNA usado y de la ausencia de contaminación de DNA (ver Materiales y métodos). D) Esquema de *NOF* con las predicciones de cada ORF (flechas de colores en zona superior), los cebadores utilizados, tanto para la retrotranscripción como la PCR (flechas negras), los cDNA teóricamente generados (barras de colores inferiores) y sus respectivos amplicones (barras negras). El esquema no está dibujado a escala.

Como puede observarse en la figura 5.20, los resultados muestran amplificación a partir de los cDNAs de ORF-1, ORF-0 y ORF-2. No conseguimos detectar la presencia de cDNA correspondiente al ORF-3.

Este resultado parece respaldar la hipótesis que la región codificante de *NOF* está formada por un solo ORF grande, pues el cDNA generado desde ORF-2 se extiende hasta ORF-1, con lo cual el inicio de transcripción debe de ser compartido. Sin embargo, la predicción de *GenScan* incluía un intrón estratégico (128bp) en ORF-0, cuyo procesado

eliminaría la región que contiene el codón *stop* y la señal de poliadenilación correspondiente al ORF-1, así como la metionina inicial de ORF-2. Por lo tanto, sin ese intrón la región codificante de *NOF* tiene que estar partida en dos ORF solapados y en fases de lectura diferentes. En nuestros resultados no se observa rastro del intrón en la amplificación NORF0f vs. NORF0.2.3r, contando con resolución para ello en el gel de agarosa (Fig. 5.20), de modo que nos encontramos en una situación intermedia entre las distintas predicciones bioinformáticas. Este es un resultado interesante, pues podría implicar la generación de un mRNA con dos ORF distintos, es decir, un mRNA dicistrónico. Dada la longitud de ORF-1, la traducción de ORF-2 debería iniciarse a partir de un IRES (*Internal Ribosome Entry Site*), presuntamente situado al inicio de ORF-2.

Los mRNA policistrónicos son característicos de los procariotas (como el operon *lac*, de *E. coli*), pero no lo son en eucariotas (Blumenthal, 2004). En los últimos años se han identificado una serie de posibles candidatos en especies tan alejadas como *Caenorhabditis elegans* (Spieth *et al.*, 1993), *D. melanogaster* (Misra *et al.*, 2002) y *Homo sapiens* (Lee, 1991), aunque algunos autores consideran que las evidencias no son concluyentes y que los métodos de identificación de este tipo de mRNA podrían no ser adecuados (Kozak, 2005). Sea como fuere, nuestros experimentos apuntan a la existencia de un RNA portador de los dos ORF predichos. Evidentemente nuestros datos no pueden distinguir si este RNA es una forma intermedia no procesada de ORF-0, o un producto no maduro (pre-poliadenilación) de ORF-1, o bien una expresión residual frente a una mayoría de mRNA monocistrónico de ORF-1 y 2. La resolución de la hipótesis del mRNA dicistrónico requeriría el análisis de todas las formas posibles de mRNA derivadas de *NOF*, para descartar la presencia de mRNAs monocistrónicos. Adicionalmente, requeriría acotar la presunta secuencia IRES de ORF-2 y demostrar que puede expresar el CDS 3' en un plásmido dicistrónico, sin tener capacidad promotora intrínseca. Sería, sin duda, una comprobación muy interesante, pero cae fuera del objetivo del presente trabajo.

La expresión de los ORF respalda las predicciones

Paralelamente a los análisis bioinformáticos y de transcripción, se construyeron tres vectores con distintas combinaciones de ORF, para comprobar experimentalmente la localización subcelular de las proteínas de *NOF*. En todos los casos se usó pPL18 como vector de expresión (ver sección Materiales y métodos para una descripción detallada de la creación de los vectores). Los tres vectores son:

- pP0G: Portador de los ORF-1 y 2 en la forma nativa en la que se encuentran en *NOF*, fusionados en fase con la proteína fluorescente eGFP (verde). Amplificados a partir de los cebadores ORF1f y ORF2r.
- pP1Y: Portador del ORF-1, fusionado en fase con la proteína fluorescente eYFP (amarillo). Amplificado a partir de los cebadores ORF1f y ORF1r.
- pP2C: Portador del ORF-2, fusionado en fase con la proteína fluorescente eCFP (cian). Amplificado a partir de los cebadores ORF2f y ORF2r.

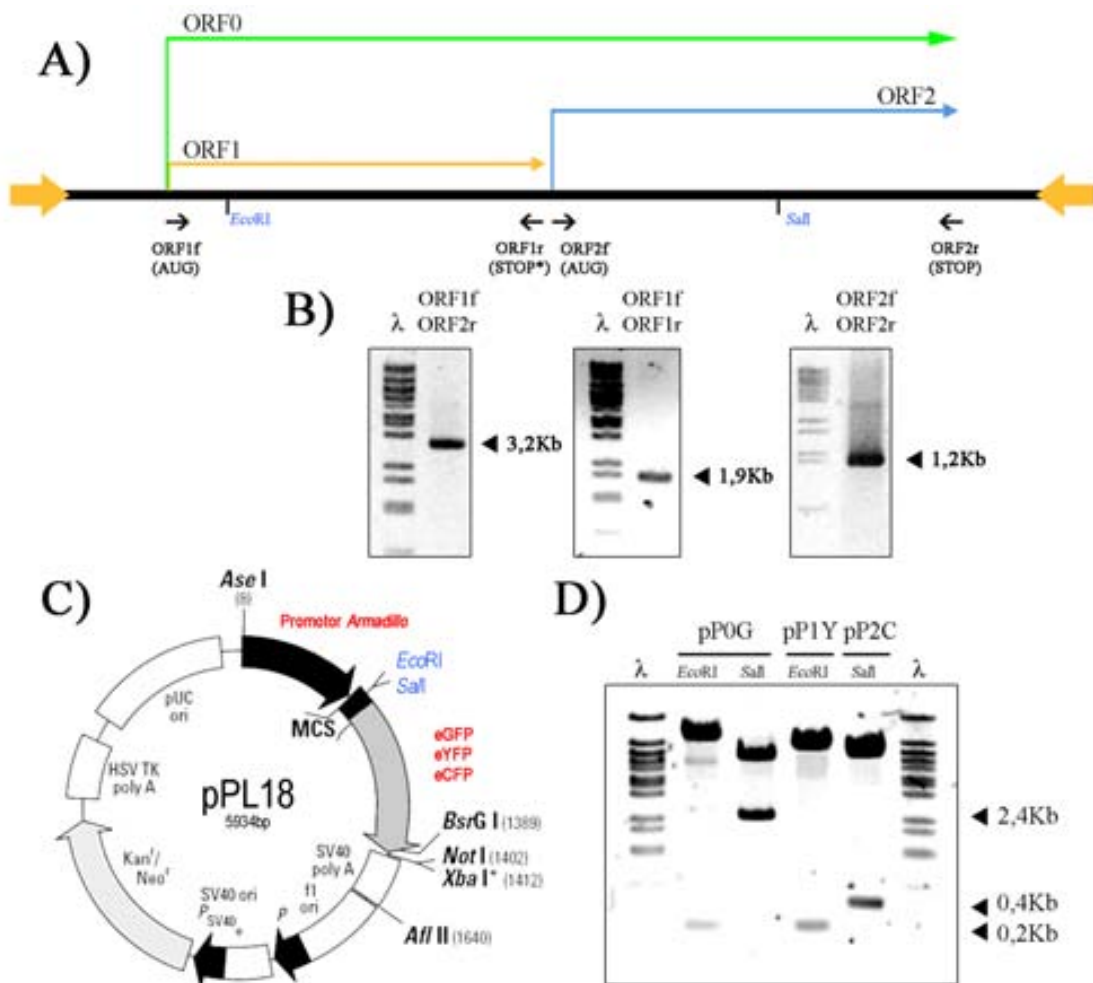


Fig. 5.21. Los vectores de transfección. A) Esquema de las reacciones de PCR sobre *NOF* que se clonaron en el vector de expresión. No está a escala. B) Comprobación de los amplicones en gel de agarosa. C) Esquema del vector de expresión pPL18. D) comprobación de los clones pP1Y, pP2C y pP0G en gel de agarosa. El tamaño de las bandas *EcoRI* y *Sall* confirman la orientación.

Los pasos intermedios en la generación de los vectores se muestran gráficamente en la figura 5.21. Puesto que estos vectores se diseñaron y generaron en paralelo con los análisis bioinformáticos, pP1Y contiene la secuencia ORF-1 descrita por Templeton y Potter

(1989) y no la versión corregida derivada de nuestros análisis. Esta secuencia es 113bp más corta debido a un error de secuenciación en el trabajo original, pero puesto que en sus experimentos demostraron que este fragmento es capaz de generar una proteína parecida a la nativa y dado que las regiones importantes para el destino de una proteína en la célula se encuentran en la zona N-terminal, creímos innecesario repetir los experimentos de localización subcelular con una versión actualizada de ORF-1.

Los tres vectores se introdujeron por transfección transitoria en células de la línea Schneider-2 (S2) de *D. melanogaster*. La localización subcelular de las proteínas se comprobó a partir de las imágenes de microscopía de fluorescencia de distintos experimentos. Aunque cada uno de los vectores exprese una versión distinta de proteína fluorescente, con máximos de emisión ligeramente distintos, observamos que el filtro FITC usado en nuestras instalaciones detectaba porciones significativas del espectro de cada una de ellas, de modo que todas las imágenes se presentan bajo este filtro.

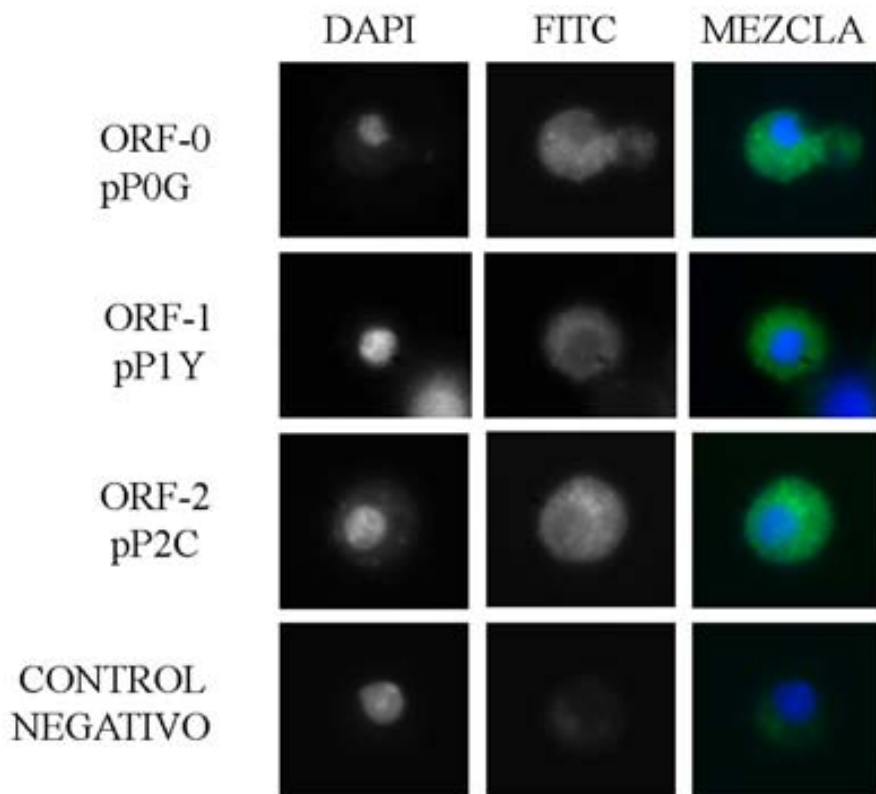


Fig. 5.22. Transfección de los vectores con los ORFs de NOF en células S2. Células S2 transfectadas transitoriamente con los vectores de expresión pP0G, pP1Y y pP2C. Se incluye un control negativo, sin DNA externo, para estimar la fluorescencia propia de la célula. DAPI muestra la cromatina (núcleo) y FITC el canal verde. El color de la mezcla de las dos imágenes no es real.

Los resultados de las transfecciones muestran que tanto ORF-1 como ORF-2 localizan mayoritariamente en el citoplasma, como ya indicaban las predicciones informáticas (ver Fig. 5.22). Las células transfectadas con ORF-0, la versión más parecida a la realidad de *NOF*, solo pueden mostrar la localización del péptido generado por ORF-2, pues es el que genera la fusión con eGFP en este plásmido. Los resultados de la transfección con ORF-0 coinciden con los de ORF-2, de modo que no parece que la presencia de los dos péptidos en la célula modifique la localización de la proteína ORF-2.

Discusión general sobre el elemento transponible *FB-NOF*

Por primera vez se han secuenciado los dos IR de un elemento completo *FB-NOF*, de forma fiable. Esto nos ha permitido renovar el esquema general de la estructura del elemento establecido a finales de los años 80. Aunque su estructura repetitiva y ordenada ya se conocía, nuestra propuesta de bloques y repeticiones con conectores CTC es más sencilla y esquemática. Esto facilita el análisis rápido de cualquier elemento, más allá de las pequeñas variaciones de secuencia frecuentemente observadas. A falta de conocer si estos IR, inusualmente largos y modulares, funcionan como diana de alguna proteína, propia o celular, parece que el objetivo de la secuencia de *FB* sea el mantenimiento de una estructura secundaria específica. En este caso, las pequeñas diferencias de secuencia observadas entre IRs no serían de gran importancia funcional.

La posibilidad de un análisis estructural facilita la búsqueda de otros elementos de su clase en secuencias genómicas, pues las herramientas basadas únicamente en la secuencia tienden a cometer errores de identificación. Una simple revisión manual de los resultados de un programa tan sencillo como `cazaFB.pl`, nos ha permitido corregir el listado publicado de elementos en el genoma de la cepa secuenciada de *D. melanogaster*, reduciendo el número de elementos anotados de 32 a 27 en el tercer lanzamiento.

En lo referente a *FB-NOF* como un todo, cabe destacar el hecho de que la parte *NOF* se encuentre en una posición predeterminada dentro de *FB*, como es un final de segunda repetición. Tradicionalmente las dos partes del elemento se describen por separado (ver, por ejemplo, listados de transposones en Kaminker *et al.*, 2002; Bartolome *et al.*, 2002), pero nuestros análisis demuestran su vinculación y proponemos dos hipótesis que pueden explicar la organización de *FB-NOF*, ambas con pruebas a favor y en contra. Ya tengan una relación simbiótica o sean un solo elemento transponible, proponemos que en adelante *FB-NOF* sea tratado como una única entidad.

La mayoría de los elementos analizados en este trabajo se encuentran en zonas eucromáticas, donde se sitúan en las proximidades 5' de genes, o en intrones. Esto es notable, pues las teorías más aceptadas sobre distribución de transposones (modelo de intercambio ectópico y modelo de inserción mutacional) predicen que éstos deberían encontrarse mayoritariamente en heterocromatina y en regiones pobres en genes (Bartolome *et al.*, 2002). Así, se minimizarían las posibilidades de producir efectos deletéreos sobre el individuo que pudieran ser objeto de selección. Sin embargo, se sabe que otros transposones y retrotransposones también tienen preferencia para la inserción en las proximidades de secuencias codificantes (Wu y Burgess, 2004), de modo que la selección natural es un factor importante en el análisis de la distribución de transposones. El hecho de que todos los elementos *FB-NOF* encontrados en el genoma de la cepa secuenciada estén incompletos, puede indicar que se trata de inserciones antiguas que ya sufrieron un proceso de selección natural (Bartolome *et al.*, 2002). Sin embargo, es chocante que las secuencias *NOF* sean más sensibles a la selección que las secuencias *FB*, pues éstas se mantienen en el punto de inserción aún teniendo una demostrada capacidad recombinogénica. Posiblemente la capacidad de expresión de *NOF* y su interferencia sobre genes cercanos, por siRNA, tenga un papel importante en su selección negativa.

Contrasta con lo encontrado en las secuencias genómicas, el hecho de que existan cepas con un elevado número de elementos completos, como nuestra cepa M115. Significaría que estas movilizaciones han sido recientes, en términos evolutivos, de modo que el elemento *FB-NOF* podría continuar activo en un subconjunto de cepas de *D. melanogaster*. Al fin y al cabo, *FB-NOF* expresa sus ORF en estas cepas, como demuestran nuestros experimentos de RT-PCR. El hecho de que muchas cepas con elementos completos muestren inestabilidad (Harden y Ashburner, 1990) refuerza la hipótesis que *FB-NOF* continúa activo en la actualidad. De nuevo, nuestra cepa M115 y su revertiente fenotípica RM115 son un buen ejemplo de la relación entre un elevado número de elementos *FB-NOF* completos e inestabilidad.

La biología de *FB-NOF* no será completa hasta que sepamos qué expresa su secuencia codificante. En nuestro trabajo hemos detectado la expresión en adultos de dos de los tres posibles ORFs y demostrado que generan péptidos con localización citoplasmática. No parece muy lógico, pues en general deberían ejercer su función en el núcleo si fueran enzimas con actividad transposasa y/o resolvasa, como uno esperaría en un elemento transponible. Además, el parecido más próximo con otras proteínas es con endonucleasas,

que normalmente son nucleares. Existe la posibilidad de que este resultado sea un artefacto experimental, pues podría requerirse la interacción con algún factor, ausente en la fase adulta, para su translocación efectiva al núcleo. Igualmente, la sobreexpresión de estas proteínas podría alterar su patrón de distribución y estar observando, en las imágenes de microscopía, el efecto de la acumulación de proteína en el citoplasma que podría ocultar una fracción correctamente localizada.

Nunca fue objetivo de nuestro trabajo responder a todas las preguntas abiertas sobre el desconocido elemento *FB-NOF* de *D. melanogaster*, pero estamos convencidos de que nuestros descubrimientos significan un avance importante en su completa caracterización.

6. CONCLUSIONES

Teniendo en cuenta los objetivos fijados para el presente trabajo de investigación y a la luz de los resultados obtenidos en las diversas estrategias experimentales, concluimos que:

1. El fenotipo *zeste* ampliado en los machos de las cepas M115 y w^{+UZ} es consecuencia de la duplicación completa del gen *white* en estas cepas, portadoras del alelo *zeste*¹.
2. Tanto la cepa M115 como RM115 son mutantes insercionales de un elemento transponible *FB-NOF* en la posición -5655, según las coordenadas de Levis. La inserción descrita anteriormente en la cepa w^{+UZ} de Rasmuson-Lestander y por extensión la de w^{+UR} , se encuentra exactamente en el mismo punto.
3. La inserción en M115 y la reversión en RM115 son distintas a las descritas en w^{+UZ} y w^{+UR} , pese a coincidir el transposón implicado y el punto exacto de inserción.
4. La reversión fenotípica en RM115 se debe a la delección de una de las copias del gen *white* duplicadas en M115.
5. La estructuración de los IR de *FB* en bloques formados por cinco repeticiones no idénticas y delimitados por el trinucleótido CTC, facilita el análisis bioinformático y molecular del elemento transponible *FB-NOF*.
6. La posición de las secuencias *NOF* entre los IR de *FB* es específica de secuencia. Se encuentran siempre después del final de una segunda repetición, por los dos extremos.
7. La distribución de *FB-NOF* en el genoma no es homogénea. Se aprecia un sesgo a favor de las inserciones en el cromosoma X.
8. La inserción del elemento transponible *FB-NOF* no es específica de secuencia. Sin embargo, tiene preferencia por secuencias ricas en A/T situadas en las proximidades 5' de genes, o en intrones.

9. La integridad de los elementos *FB-NOF* en M115 es muy elevada, en comparación con la integridad de los elementos encontrados en la cepa utilizada para la secuenciación del genoma de *D. melanogaster*.
10. La secuencia de los elementos *FB-NOF* presentes en las cepas M115 y RM115 en la posición +2686 de *NOF*, coincide con la descrita por Harden y Ashburner en 1990: AAACAC.
11. La secuencia codificante de *FB-NOF* contiene dos ORFs solapados por 20bp que podrían expresarse en un mismo mRNA dicistrónico.
12. Las proteínas codificadas en los dos ORFs tienen localización citoplasmática en las condiciones del estudio.

7. MATERIALES Y MÉTODOS

7.1- TÉCNICAS in vivo e in vitro

7.1.1- CEPAS DE *Drosophila melanogaster*

En el presente trabajo se han usado distintas cepas de *Drosophila melanogaster*. Para la estabulación general de las cepas, se ha mantenido una temperatura constante de 25°C, con una humedad relativa alrededor del 60% y con un ciclo luz/oscuridad de 12h/12h. Las poblaciones desarrollan su ciclo vital en frascos de cristal con medio nutritivo estándar suplementado con ampicilina, cambiándose a frascos con medio fresco cada 15 días.

- **CEPAS CONTROL:**

- Canton-S (CS): Cepa silvestre de referencia.
- Zeste-1 (z^1): Cepa de referencia de la mutación z^1 .
- Inestable Zeste (UZ): Cepa de referencia de la mutación w^{+UZ} amablemente cedida por la Dra. Åsa Rasmuson-Lestander (Rasmuson-Lestander y Ekstrom, 1996).

- **CEPAS DE MANTENIMIENTO:**

- BASC (In(1) $sc^{SIL} sc^{8R} +S, sc^{CI} sc^{\delta} w^a B$): Cepa portadora de una inversión en el cromosoma X que impide la recombinación con su homólogo. Usada para generar cepas homocigotas para el cromosoma X a partir de un solo macho.
- X unidos (C(1) DX, y f/Y): Los cromosomas X de estas hembras siempre segregan juntos. Puesto que individuos X^X/X no eclosionan o son, si lo consiguen, estériles, las hembras de la cepa son genotípicamente X^X/Y . Con estas hembras se pueden conseguir líneas en las que un cromosoma X mutante se transmita de padres a hijos machos continuamente, ahorrándose embarazosos eventos de recombinación en el cromosoma X.

- **CEPAS MUTANTES:**

- M115 (sc, z^1, f): Cepa de machos mutantes que presentan el fenotipo $zeste^1$. Se mantienen rutinariamente con hembras de la cepa de X unidos debido a la inestabilidad de las hembras homocigotas para el cromosoma X mutante.

- RM115 (*sc*, z^1 , *f*): Cepa establecida a raíz de un macho M115 cuyo fenotipo revirtió a una tonalidad de ojos silvestre.

7.1.2- EXTRACCIÓN DE DNA

7.1.2.1- Extracción de DNA genómico de *Drosophila melanogaster*

El procedimiento usado es una modificación del descrito por el Dr. Jaume Piñol (Piñol *et al.*, 1988).

Homogeneizar en Tampón A (100mM NaCl; 10mM EDTA; 0,5% Triton X-100; 20mM Tris·HCl pH 7,5) alrededor de 0,3g de moscas previamente anestesiadas. Separar los núcleos del resto de cubiertas celulares mediante filtración a través de una malla de nailon (150 mesh). Realizar dos lavados mediante centrifugación a 4000RPM durante 10 minutos y resuspender en Tampón B (100mM NaCl; 10mM EDTA; 20mM Tris·HCl pH 7,5). Para la lisis nuclear, añadir NaClO_4 4M y 4% SDS. Estas condiciones favorecen la disociación de las proteínas unidas al DNA y la eliminación de pigmentos. Desproteínizar con un paso de fenol y dos de cloroformo/alcohol isoamílico (24:1 v/v). Limpiar de sales la fase acuosa, con el DNA purificado y desproteínizado, mediante precipitación del DNA con etanol absoluto. Lavar e hidratar con etanol 70% y disolver en agua destilada.

7.1.2.2- Extracción de DNA plasmídico de *Escherichia coli*: *Miniprep*.

Las extracciones de DNA de plásmidos clonados en cepas de *Escherichia coli* se han realizado usando el kit *QIAprep*® *Spin Miniprep Kit* (QIAGEN, EE.UU.) según el protocolo propuesto en el manual. Se toma como material de partida 3mL de cultivo crecido en medio LB durante aproximadamente 16 horas, a 37°C y agitación constante a 300 RPM. La base del procedimiento es la lisis alcalina de las células y la precipitación del DNA genómico junto con los restos celulares. El DNA plasmídico, que no sedimenta junto con el genómico, se captura en una membrana por afinidad de carga, se limpia con una solución a base de etanol y se eluye con agua destilada a una concentración alrededor de los 200ng/μL.

7.1.3- EXTRACCIÓN DE RNA

Para la extracción de RNA de *D. melanogaster* se ha usado el sistema basado en TRIzol® (Invitrogen, EE.UU.). Este es una solución monofásica de fenol e isotiocianato de

guanidina que estabiliza las cadenas de RNA durante el homogeneizado de la muestra. La adición de cloroformo separa la mezcla en una fase orgánica, donde se encuentran las proteínas, una interfase que contiene el DNA genómico y una fase acuosa con el RNA. De la fase acuosa se puede obtener el RNA precipitándolo con isopropanol. Adicionalmente, se puede recuperar el DNA de la interfase por precipitación con etanol y las proteínas de la fase orgánica mediante precipitación con isopropanol. El RNA se disuelve en agua libre de RNAasas y se trata con DNA-free™ (Ambion, EE.UU.) para eliminar cualquier contaminación de DNA genómico.

El RNA es muy sensible a la degradación, ya que las enzimas RNAasa son ubicuas y muy persistentes. Por eso, todo el trabajo con RNA se ha llevado a cabo con material especialmente tratado para estar libre de RNAasas: los recipientes de vidrio se sometieron a 180°C de calor seco. El agua y el material de plástico (puntas de pipeta y tubos *ependorf*) se autoclavaron durante 1h a 120°C de calor húmedo, en recipientes de vidrio previamente tratados. Todas las soluciones se prepararon con agua tratada y sales especialmente reservadas para el uso con RNA. El material de electroforesis fue tratado con NaOH 2M durante un mínimo de 30' antes de entrar en contacto con las muestras y aclarado con agua libre de RNAasas.

7.1.4- PCR: REACCIÓN EN CADENA DE LA POLIMERASA

Los experimentos de PCR se han realizado en distintos termocicladores:

- PTC-100™ *Programable Thermal Controller* (MJ Research Inc. / BIO-RAD, EE.UU.) De efecto Peltier con tapa termoregurable. Uso con tubos de 0,6mL.
- *iCycler™ Thermal Cycler* (BIO-RAD Laboratories Inc., EE.UU.). De efecto Peltier con tapa termoregurable. Uso con tubos de 0,2mL.
- *Light Cycler* (Roche Diagnostics GmbH, Alemania). Termociclador diseñado específicamente para experimentos de PCR a tiempo real. Uso con capilares de 20μL

Todas las reacciones de PCR se preparan en campana de flujo laminar, previamente irradiada con luz UV durante 10-15 minutos, con el objetivo de degradar posibles contaminantes de DNA que podrían servir de molde alternativo en las reacciones. Asimismo, durante la mezcla de los componentes siempre se han usado puntas de pipeta equipadas con filtro anti-aerosoles.

A lo largo del trabajo se han usado distintas enzimas DNA polimerasa termoestables adaptadas a estrategias de PCR particulares:

- *Taq DNA Polymerase* (Invitrogen Corp., EE.UU.): Enzima DNA polimerasa clásica de *Thermus aquaticus*, usada en reacciones de productos cortos de secuencia no compleja.
- *ELONGASE® Enzyme Mix* (Invitrogen Corp., EE.UU.): Mezcla de *Taq* DNA pol. y *Pyrococcus sp.* GB-D DNA pol. con actividad correctora 3'-5' exonucleasa (*proofreading*). Capaz de amplificar fragmentos de hasta 20Kb, se ha usado principalmente para experimentos de PCR de largo alcance.
- *Expand Long Template PCR System* (Roche Molecular Systems Inc., Alemania): Mezcla de *Taq* DNA pol. y *Tgo* DNA pol. Con capacidad correctora 3'-5' exonucleasa. Se ha usado para reacciones de PCR de largo alcance.
- *Fail Safe™ PCR System; Fail Safe™ PCR Premix Selection Kit* (Epicentre® Biotechnologies, EE.UU.): Sistema integral de PCR que proporciona una mezcla de enzimas, una de ellas con actividad correctora para mayor fidelidad y procesividad, junto con una batería de tampones con distintas concentraciones de sales (en particular $MgCl_2$) y cosolventes (en particular betaína) que comprenden un amplio abanico de posibilidades para conseguir u optimizar el resultado de reacciones de PCR de fragmentos ricos en GC o con estructuras secundarias.

7.1.4.1- Real Time PCR (PCR a Tiempo Real)

En los experimentos de PCR a tiempo real se ha usado el termociclador *LightCycler*. Este aparato está especialmente diseñado para PCR a tiempo real, con lo cual incorpora un sensor óptico para la precisa cuantificación de la fluorescencia en las muestras. La reacción se efectúa en capilares de cristal de 20 μ L de capacidad, sometidos a las diferentes temperaturas programadas en el ciclo de PCR mediante un sistema de flujos de aire caliente/frío (ver Fig. 7.1). A diferencia de los termocicladores estándar de PCR, *LightCycler* no mantiene la temperatura fija en cada paso del programa, sino que esta es regulada de forma dinámica. Esto se debe a la brevedad de cada uno de los pasos de los ciclos de PCR, de apenas unos segundos de duración, lo cual es posible por el reducido volumen de los capilares y su gran superficie de contacto con la muestra, que facilita un intercambio casi instantáneo de temperatura.

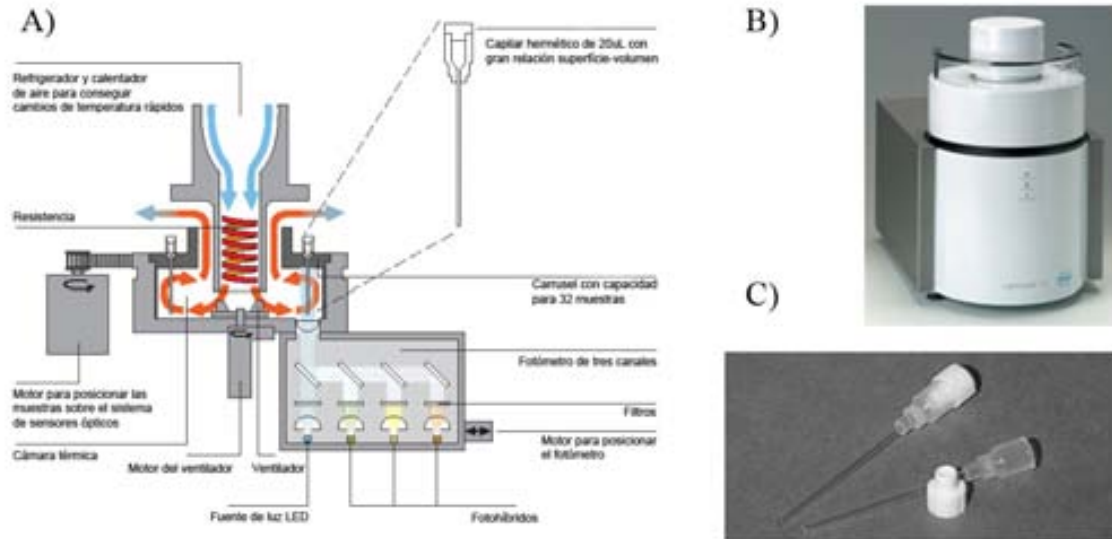


Fig. 7.1. Termociclador *LightCycler* para PCR a tiempo real. A) Esquema del funcionamiento de *LightCycler*. Se muestran los componentes básicos y el flujo de aire que controla la temperatura. B) Diseño exterior del aparato. C) Detalle de los capilares usados en las reacciones.

Para la reacción de PCR se ha usado el kit *LightCycler DNA Master SYBR Green I* (Roche, Alemania), consistente en una mezcla de PCR ya preparada a la que sólo debe añadirse el DNA molde y los cebadores. Uno de los componentes de la mezcla es el *SYBR Green I*, un compuesto fluorescente con gran afinidad para el DNA de doble cadena (ver Fig. 7.2). La asociación DNA-*SYBR Green I* incrementa significativamente la emisión fluorescente de esta última molécula, con lo cual puede cuantificarse, con gran exactitud, el incremento de amplicones después de cada ciclo de PCR. Para esta tarea se emplea el programa proporcionado con el termociclador *LightCycler* según se indica en el manual del usuario.

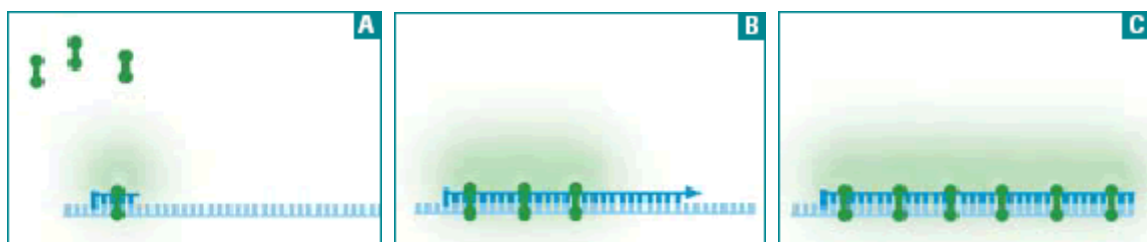


Fig. 7.2. Incorporación de *SYBR Green I*. Durante la reacción de PCR, el *SYBR Green I* se incorpora al DNA de doble cadena después de cada ciclo de renaturalización.

El programa de PCR usado en este trabajo es básicamente el propuesto en el manual del *kit*, modificado únicamente por las temperaturas de apareamiento de los cebadores específicos de nuestras reacciones.

7.1.4.2- RT-PCR

La RT-PCR es una reacción de amplificación a partir de un molde de cDNA. Para eso, hace falta retrotranscribir el RNA diana con una enzima transcriptasa reversa y un solo cebador de DNA complementario a la cadena codificante. Las reacciones de retrotranscripción se hicieron con la enzima Superscript™ II RNase H⁻ RT (Invitrogen, EE.UU.), siguiendo los pasos especificados en el manual del fabricante. Obviamente, todo el material de plástico y el agua para las diluciones de los cebadores y dNTPs fue tratado para estar libre de RNAasas.

Una vez conseguido el molde de cDNA, ya no es necesario tomar las exigentes precauciones del trabajo con RNA, de modo que se usaron los protocolos estándar de PCR.

7.1.4.3- Cebadores utilizados

La siguiente tabla muestra el nombre, la secuencia 5'-3', la temperatura de fusión (T_m) y el porcentaje de G/C de todos los cebadores utilizados a lo largo del trabajo. La T_m de los cebadores con colas 5' (sombreadas en la secuencia del cebador), usados para el clonaje de secuencias *NOF* en los vectores de expresión (sección 7.1.10.3), no es válida para los primeros ciclos de amplificación, pues las “colas” no hibridan con la secuencia molde.

NOMBRE/USO	SECUENCIA 5'-3'	T _m °C	%G/C
Inserción elemento <i>FB</i>			
FB	TTT TGA AGA GGT GGG CTA ACG A	58,4	45,5
G5	CAG TGG GAC AAA GTG CGC TAA	59,8	52,4
G5r	TTA GCG CAC TTT GTC CCA CTG	59,8	52,4
G3	CGC AGT GTA GGA CCT ATC AGG T	62,1	54,5
FBs	CAC AAA AAC TGG ACG TGG GC	59,4	55,0
G3s	CGA CAC AAG GGG CAG AAC TTC	61,8	57,1
PCR tiempo real			
Zf	AAC ACA TCA ACA GCC TCC T	54,5	47,4
Zr	TTC ATT GTC AGC TCC TCG A	54,5	47,4
wZBf	GAC AGG CGA GTG ACA ATA A	54,5	47,4

MATERIALES Y MÉTODOS

wZBr	GAT CTA GCG ACA GAC AAC AA	55,3	45,0
wBCf	GAT CAG GAG CTA TTA ATT CGC	55,9	42,9
wBCr	GCT ACC AGA ATA ATC TGT TTC G	56,5	40,9
wKEf	CGT AAT CAG TTC TTA TCG CG	55,3	45,0
wKEr	CCT TTT TAT GTA CCT ACG CGT	55,9	42,9
wCGf	AAT TCT GCA ACT GCA AGA TG	51,5	40,0
wCGr	ATG CCA GGC TAA TCA GAA TC	52,4	45,0
Clones IR de <i>FB</i>			
NOF5	AGG TGG GCG ATG TCA AGT CG	62,1	60,0
NOF3	GGC ACT AGT ATC CAG CTG TTG	55,4	52,4
Zona de discrepancia de <i>NOF</i>			
NOF2040	GCT TTA ATG AGT GTA GGA TT	47,3	35,0
NOF2740	AGC GGA TAA ATT CGT GGT GC	55,8	50,0
RT-PCR			
NORF1f	CCC AGA ACC GAG TTA CAT AC	51,4	50,0
NORF1r	AGG TGG GCG ATG TCA AGT CG	62,1	60,0
NORF0f	GGA CCA GGA GGA AAT TGA CG	55,6	55,0
NORF2.3f	AAC TGA GCC GCG ATG AAA CGC	60,9	57,1
NORF0.2.3r	CCG TTT GAC TCT CTA GTG GA	53,5	50,0
Vectores de expresión			
ORF1f	AAG GTA CCA AGT GTT ACT GTC AGG ATG CC	66,7*	48,3
ORF1r	TAC CCG GGT GTC CAA AAA TTT AAA TCC TTG	64,0*	40,0
ORF2f	AAG GTA CCA ACT GAG CCG CGA TGA AAC GC	69,5*	55,2
ORF2r	TAC CCG GGC ACT TTT AAA CAA GGC TTA	63,4*	44,4

7.1.5- CLONACIÓN DE DNA

A lo largo de este trabajo, se han clonado gran cantidad de fragmentos de DNA en vectores plasmídicos, bien para secuenciación, bien para su uso como sondas. La

procedencia de dichos fragmentos es diversa, si bien en general provienen de amplicones de PCR o de fragmentos de restricción de otros clones. A modo de vectores se han usado, mayoritariamente, tres tipos de plásmidos:

- pBluescript SK+ (Promega, EE.UU.): Vector de elevado número de copias por célula, incorpora el *cassette* de resistencia a ampicilina y la posibilidad de identificar clones recombinantes por diferencia de color azul/blanco gracias a la α -complementación del cistrón Z en el operón *Lac*.
- pTZ18U: Muy similar al pBluescript SK+, por conferir resistencia a ampicilina y la selección por color. Tiene la ventaja de ser algo más pequeño e incorporar las secuencias de los cebadores universales T7 y M13 -20 muy cerca del sitio de clonación, lo cual hace del pTZ18U un excelente vector para secuenciación.
- pCR-TOPO (Invitrogen Corp., EE.UU.): El sistema *TOPO Cloning*® de Invitrogen ha sido usado extensamente a lo largo del trabajo para la clonación directa de fragmentos generados por PCR, ya sea mediante *Taq* DNA polimerasa (*TOPO-TA Cloning System*) como por mezclas de enzimas con actividad correctora (*Zero Blunt TOPO Cloning*, pues estas corrigen automáticamente las adeninas adicionales que la *Taq* incorpora al disociarse de la cadena molde).

Los vectores pCR-TOPO se adquieren linealizados y con los extremos asociados covalentemente a la topoisomerasa I del virus *vaccinia*. Esta enzima captura los extremos OH libres del producto de PCR, los une al vector y se libera, dejando el plásmido recircularizado, sin necesidad de un aporte externo de energía, lo cual permite la ligación vector-inserto a temperatura ambiente y en escasos minutos.

Confiere resistencia a ampicilina y/o kanamicina y la selección de los clones recombinantes se realiza mediante la disrupción del gen letal *ccdB* de *E. coli*.

DESFOSFORILACIÓN DE VECTORES: Cuando se debe clonar en pBSK o pTZ18U y se dispone de poca cantidad de inserto, es muy recomendable desfosforilar el vector para evitar la recircularización del mismo sin incorporar el fragmento de DNA que nos interesa.

A partir de 2µg de DNA del vector linealizado (máximo 44µL), añadir 5µL de tampón de desfosforilación 10X, 2µL de fosfatasa alcalina y H₂O hasta un volumen de reacción de 50µL. Incubar 30 minutos a 37°C. Detener la reacción con 5µL de EDTA 0,2M pH 8 e inactivar la enzima a 75°C durante 10 minutos. Para evitar interferencias con los procesos subsiguientes de

clonación, purificar el DNA desfosforilado con el kit *Nucleospin Extracts* (Macherey-Nagel GmbH, Alemania).

LIGACIÓN: Excepto cuándo se usa el sistema *TOPO Cloning*[®], es necesario ligar el vector digerido y desfosforilado con el inserto de DNA, digerido con la misma enzima que el vector (o una que dé extremos compatibles), usando la Ligasa del bacteriofago T4.

Mezclar el DNA inserto con el vector en una relación 3:1 para favorecer el encuentro entre los dos. Mantener 5 minutos a 65°C para romper las posibles uniones débiles en regiones complementarias y añadir 1µL de Tampón de ligación 10X, 1µL de T4 DNA Ligasa (Roche, Alemania) y H₂O hasta 10µL. Mantener durante toda la noche a una temperatura de 15-16°C.

OBTENCIÓN DE CÉLULAS COMPETENTES: Salvo en los casos en los que se ha usado un kit de clonación, como el *PCR TOPO Cloning System* que incorpora sus propias células competentes. Estas se han preparado mediante la permeabilización con CaCl₂. La cepa de *E. coli* usada para este fin ha sido DH5α (*SupE44 Δlac UI69 (Φ80 lacZΔM15) hsdR17 recA1 endA1 gyrA96 thy-1 relA1*).

Partir de 0,2mL de un cultivo de noche en LB, e inocular 20mL de medio de cultivo 2xYT. Dejar crecer el cultivo hasta una OD₅₅₀ de 0,6 (fase exponencial de crecimiento) y repartir en dos tubos de 10mL. Centrifugar a 4°C (5000 RPM, 5 minutos) y resuspender las células en 5mL de CaCl₂ 50mM frío y estéril. Incubar las células durante 45-60 minutos en hielo, centrifugar de nuevo y resuspender en 2mL de CaCl₂ 50mM. Incubar una hora más en hielo, añadir 15% de glicerol estéril y alícuotar en criotubos a razón de 100-150µL por tubo. Congelar inmediatamente por inmersión en nitrógeno líquido.

TRANSFORMACIÓN: Todas las transformaciones se han realizado por el procedimiento del choque térmico.

Mezclar suavemente el DNA resultante de la ligación (de 1 a 5µL) con una alícuota, recién descongelada en hielo, de células competentes. Incubar 20 minutos en hielo. Durante este tiempo, el DNA forma complejos hidroxilo-calcio fosfato con el CaCl₂ y forma un precipitado fino que se adhiere a las paredes celulares de las células competentes. Someter las células a un choque térmico de un minuto a 42°C y añadir inmediatamente 0,5mL de LB precalentado a 37°C. Este proceso permite la entrada del DNA en la célula. Incubar durante una hora a 37°C con agitación (200 RPM) y sembrar en placas con LB-agar suplementadas con el antibiótico apropiado y X-Gal si se van a identificar los clones recombinantes por la α-complementación en el operón *lac*.

7.1.6- MARCADO NO RADIOACTIVO DE SONDAS

Las sondas usadas en el trabajo, exceptuando las destinadas a FISH, han sido marcadas no radioactivamente con digoxigenina mediante el método de *Random Primer* usando el *DIG DNA Labeling and Detection Kit* (Roche, Alemania), según las especificaciones del fabricante.

Después de la reacción de marcado, el DNA se purifica con *Nucleospin Extracts* (Macherey-Nagel GmbH, Alemania) para limpiar la sonda de nucleótidos no incorporados, con el fin de evitar que saturen la membrana y evitar ruido de fondo.

7.1.7- TÉCNICA DE SOUTHERN BLOT

ELECTROFORESIS: De 2 a 5µg de DNA, digerido con la enzima de restricción apropiada, se sometieron a electroforesis durante aproximadamente 15 horas a bajo voltaje (~35V). En estas condiciones se consigue una migración electroforética uniforme y una muy buena resolución de los fragmentos de DNA. Para conseguir una mayor precisión en el cálculo de los tamaños de los fragmentos de restricción, en la mayoría de los casos se han usado geles de 18cm de longitud. Según la conveniencia de cada experimento, se han usado concentraciones de agarosa entre 0,6% (*Agarose D-1 Low EEO*, Pronadisa, España) y 0,3% (*SeaKem® Gold Agarose*, Cambrex, EE.UU.).

TRANSFERENCIA: Los fragmentos de restricción separados electroforéticamente se transfirieron a membranas de nailon cargadas positivamente (*Nylon Membranes positively charged*, Roche, Alemania) mediante transferencia por vacío usando el dispositivo *VacuGene XL* (GE Healthcare, Reino Unido), especialmente diseñado para tal uso. Este tipo de transferencia supone una ventaja respecto al método tradicional de capilaridad, pues permite acortar el tiempo de todos los pasos del proceso. Esto es especialmente importante durante la neutralización y la transferencia, pues se disminuye la posibilidad de renaturalización espontánea del DNA, un factor muy dependiente del tiempo, enriqueciendo la fracción de cadenas desnaturalizadas que se transfieren a la membrana de nailon. El *VacuGene XL* se ha utilizado según las recomendaciones del fabricante, aunque aumentando ligeramente los tiempos de cada paso, pues temíamos que los fragmentos de muy elevado peso molecular con los que lidiábamos pudieran tardar algo más de tiempo en transferir. El proceso entero se resume como: 10 minutos de **Despurinización** (0,25M

HCl), 15 minutos **Desnaturalización** (1,5M NaCl; 0,5M NaOH; pH > 11), 15 minutos **Neutralización** (1M Tris-HCl pH 8; 1,5M NaCl), 1 hora **Transferencia** (20xSSC pH 7).

Al finalizar la transferencia, la membrana se seca entre dos hojas de papel absorbente y el DNA se fija covalentemente al nailon mediante la exposición de la membrana a luz ultravioleta durante 3 minutos.

HIBRIDACIÓN: La hibridación de la sonda con el DNA fijado en la membrana se realiza en una solución tampón a base de Na₂HPO₄ 0,25M pH 7,5; agente bloqueador 1% (puede usarse leche desnatada en polvo); EDTA 1mM; SDS 10%. Este método no requiere prehibridación y se efectúa a una temperatura de entre 65 y 68°C, según la restrictividad que deseemos en la hibridación, durante 15 horas, aproximadamente.

Al término de la incubación, se realizan dos lavados de 15 minutos a la misma temperatura de hibridación con una solución de baja fuerza iónica (Na₂HPO₄ 20mM pH 7,5; EDTA 1mM; SDS 1%) para lavar los restos de sonda de la membrana y restringir la hibridación a los fragmentos específicos.

Tanto la hibridación como los lavados de restricción posteriores se llevaron a cabo en el interior de tubos de hibridación herméticos, en un horno de hibridación (Ecogen Hybridation Oven, Ecogen, España).

DETECCIÓN: La detección de la sonda hibridada se realiza según los métodos no radioactivos de precipitación de color o de quimioluminiscencia. Ambos comparten un procedimiento común aunque cambia el sustrato final y el método de revelado.

Incubar la membrana recién hibridada durante 15 minutos en solución de lavado (ácido maleico 0,1M; NaCl 0,15M; Tween-20 0,3% v/v; pH 7,5). Bloquear durante 30 minutos en una solución con agente bloqueante (ácido maleico 0,1M; NaCl 0,15M; leche desnatada en polvo 1%; Tween 20 0,3% v/v; pH 7,5). Incubar 30 minutos con una dilución 1:10⁴ o 2:10⁴ de un anticuerpo contra la digoxigenina, conjugado a la enzima fosfatasa alcalina (*Anti-Digoxigenin-AP FAB Fragments*, Roche, Alemania) en solución con agente bloqueante. Realizar dos lavados de 15 minutos con solución de lavado. Equilibrar la membrana a pH 9,5 mediante la inmersión en solución de equilibrado (Tris-HCl pH 9,5; NaCl 0,1M; MgCl₂ 50mM) durante 10 minutos, para conseguir unas condiciones de funcionamiento óptimas para la fosfatasa alcalina. Finalmente, incubar con un sustrato adecuado.

En este trabajo se ha usado bien CDP-Star (Roche, Alemania), bien una mezcla de NBT/BCIP (*NBT/BCIP Stock Solution*, Roche, Alemania). El **CDP-Star** es un sustrato quimioluminiscente que, en su reacción de catálisis por la fosfatasa alcalina, libera un

fotón. La exposición de una película fotográfica a la membrana empapada en solución de CDP-Star permite detectar las zonas de hibridación de la sonda. En el método de **precipitación** de color, sin embargo, la membrana se incubaba a oscuras sumergida en la solución con NBT/BCIP. La reacción del sustrato con la fosfatasa alcalina forma un fino precipitado marrón-azulado sobre la membrana en las zonas de hibridación.

7.1.7.1- Sondas Utilizadas

pBSWKK: Región 3' adyacente al gen *white*. Comprende el fragmento de restricción entre *KpnI* [-4282] - *KpnI* [-7502], según las coordenadas de Levis.

VECTOR: pBluescript SK+

INSERTO: 3,2Kb

CLONADO EN: *KpnI*

OBTENIDA POR: Anna Portela

pBSD5'_6B/pNOF5: Sonda de *NOF* 5', entre *EcoRI* [1011] y el cebador de PCR MTINOF5 [2151]. No contiene secuencias *FB*. Contiene 6bp correspondientes al vector pCR-BluntII-TOPO, entre MTINOF5 y *EcoRI*.

VECTOR: pBluescript SK+

INSERTO: 1,1Kb

CLONADO EN: *EcoRI*

OBTENIDA POR: Martí Badal

pTZD5'_1: Sonda de *FB* entre [227] y la diana *EcoRI* [1011] de *NOF*. No contiene DNA genómico adyacente a la inserción del elemento. La interferencia de la secuencias *NOF* en la sonda es mínima, pues no existe ningún elemento *NOF* no asociado a *FB*.

VECTOR: pTZ18U

INSERTO: 3Kb

CLONADO EN: *EcoRI*-Romo

OBTENIDA POR: Martí Badal

pNOF3: Sonda *NOF3'*, entre cebadores de PCR situados en [2040] y [4041] del elemento transponible. No contiene secuencias de *FB* ni de DNA genómico adyacente.

VECTOR: pCR4-TOPO

INSERTO: 2Kb

CLONADO EN: TOPO (liberar con *EcoRI*)

OBTENIDA POR: Anna Portela

pBSWKE*: Sonda de la región 3' adyacente al gen *white*, inmediatamente anterior a la inserción del elemento *FB-NOF* en nuestras cepas mutante y revertiente. Comprende el fragmento de restricción *KpnI* [-4282] hasta la diana *EcoRI* artificial, presente en la secuencia del plásmido pWP2 (Walter Gerhing, Universidad de Basilea, Suiza), correspondiente a [-5348] según las coordenadas de Levis. Esta sonda está incluida en la región de reconocimiento de la sonda pBSWKK.

VECTOR: pBluescript SK+

INSERTO: 1,1Kb

CLONADO EN: *KpnI* – *EcoRI*

OBTENIDA POR: Martí Badal

pTZWEK: Sonda de la región genómica inmediatamente posterior a la inserción de *FB-NOF* en las cepas mutante y revertiente. Comprende el fragmento de restricción *EcoRI* [-6265] – *KpnI* [-7502], subclonado a partir de la sonda pBSWKK y, por lo tanto, incluido en su región de reconocimiento.

VECTOR: pTZ18U

INSERTO: 1,2Kb

CLONADO EN: *EcoRI* – *KpnI*

OBTENIDA POR: Martí Badal

pCw25: Sonda de la región 5' de *white*, moderadamente alejada del inicio del gen. Comprende la región entre los cebadores de PCR w20 [+20Kb] – w25 [+25Kb] según las coordenadas de Levis.

VECTOR: pCR4-TOPO

INSERTO: 5Kb

CLONADO EN: TOPO (Liberar con *EcoRI*)

OBTENIDA POR: Martí Badal

pWBaBa: Sonda que comprende el fragmento de restricción *BamHI* [4438] – *BamHI* [1383] del gen *white*, según las coordenadas de Levis. El fragmento fue subclonado a partir del plásmido pWP2.

VECTOR: pBluescript SK+

INSERTO: 3Kb

CLONADO EN: *BamHI*

OBTENIDA POR: Anna Portela

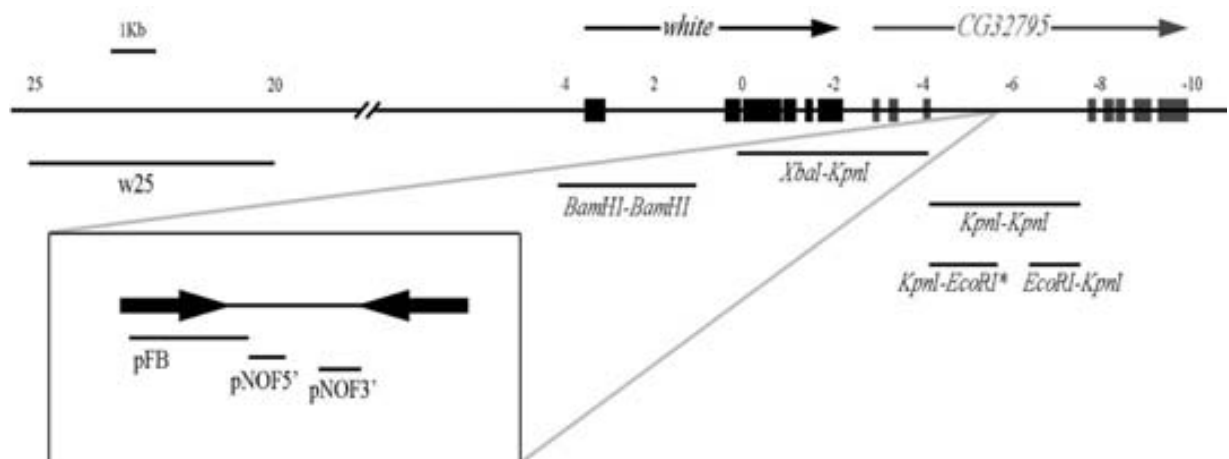
pWXK: Sonda que comprende el fragmento de restricción *XbaI* [442] – *KpnI* [-4282] del gen *white*, según las coordenadas de Levis. Subclonado a partir del plásmido pWP2.

VECTOR: pTZ18U

INSERTO: 4,4Kb

CLONADO EN: *XbaI* – *KpnI*

OBTENIDA POR: No registrado



7.1.8- FISH: Hibridación *in situ* fluorescente

La técnica de FISH se aplicó sobre preparaciones de cromosomas politénicos. Para ello, glándulas salivales de larvas de tercer estadio de las cepas a analizar fueron microdisecionadas y fijadas sobre portaobjetos previamente tratados con solución SSC Denhart (Montgomery *et al.*, 1987). Sólo las mejores preparaciones se usaron, posteriormente, para la FISH. La técnica se basa en la hibridación de sondas específicas con los cromosomas politénicos fijados en la preparación. Dichas sondas, marcadas con biotina o digoxigenina, se detectan mediante el uso de anticuerpos o estreptavidina, respectivamente, conjugados con moléculas fluorescentes que podremos visualizar con un microscopio de fluorescencia equipado con los filtros apropiados.

La sonda y la preparación se desnaturalizan por separado. Usar 30µL de sonda por cada porta e incubar durante un mínimo de 14 horas a 37°C, en una cámara húmeda. Lavar a 37°C con 2xSSC, equilibrar con 1xPBS, bloquear con solución de bloqueo al 4% y pasar al revelado de la hibridación. Añadir 100µL de una solución formada por una dilución 1:100 de estreptavidina conjugada con el fluorocromo Alexa 488 (verde) y 1:100 de un anticuerpo anti-digoxigenina conjugado con rodamina (rojo) e incubar 20 minutos a temperatura ambiente. Lavar 10 minutos con 4xSSC; 0,05% Tween-20 a 42°C y equilibrar con 1xPBS a temperatura ambiente. Montar las preparaciones con solución de montaje suplementada con DAPI y compuestos que eviten la pérdida de fluorescencia (*antifade*). Guardar las preparaciones en la oscuridad hasta su observación al microscopio.

En este trabajo se ha usado un microscopio Olympus BX-FLA equipado con los siguientes filtros:

HQ FITC → excitación HQ480/40; emisión HQ535/50; beamsplitter Q505LP

NARROW X HQ TRITC → excitación HQ545/30; emisión HQ610/75; beamsplitter Q565LP

DAPI U-MNU → excitación 360; emisión 420; beamsplitter 400

TRIPLE DAPI/FITC/TRITC → excitación D/F/R X; emisión D/F/R M; beamsplitter D/F/R BS

7.1.9- SECUENCIACIÓN DE DNA

En el transcurso del trabajo se han secuenciado gran cantidad de fragmentos de DNA, generalmente clonados en vectores plasmídicos, aunque en ocasiones se ha secuenciado los amplicones de PCR directamente. Todas las sondas usadas en las hibridaciones fueron

comprobadas mediante secuenciación de sus extremos y comparación con los datos genómicos, de disponibilidad pública.

La secuenciación de DNA se ha realizado en las instalaciones del *Servei de síntesi d'àcids nucleics i de microarrays de la Universitat Autònoma de Barcelona*, por técnicos especializados y usando un secuenciador automático Applied Biosystems 373 A, equipado con el *Upgrade Stretch Kit* y controlado por ordenador.

Todas las secuencias obtenidas fueron cuidadosamente analizadas y corregidas en caso de error evidente, o repetidas en caso de ambigüedad.

7.1.9.1- Generación y secuenciación de clones solapados obtenidos mediante exonucleasa *Bal31*.

La secuenciación de insertos internamente repetitivos puede conseguirse mediante la generación de una serie de clones con deleciones terminales solapadas, que permitan el uso de un cebador externo para la reacción de secuenciación. La mejor manera de conseguir dichos clones es mediante el uso de la enzima *Bal31* (Roche, Alemania).

Bal31 es una exonucleasa 5'-3' que degrada los dos extremos de hebras lineales de DNA de doble cadena. Este proceso de degradación es constante, con lo cual controlando el tiempo de reacción, se puede controlar, *grosso modo*, el alcance de la digestión. La cinética *Bal31* se ha llevado a cabo según lo descrito en Sambrook *et al.*, 1989, concretamente:

Linealizar 40µL (aprox. 15µg) del plásmido original con la enzima de restricción adecuada, en una reacción de 50µL. Usar 23,5µL de la digestión para la cinética *Bal31*: añadir 25µL de Buffer *Bal31* 2x y 1,5µL *Bal31* e incubar a 30°C durante el tiempo que se estime oportuno, normalmente 5, 10, 15 y 20 minutos en reacciones seriadas. Detener la reacción inmediatamente después del tiempo de digestión con 1µL EGTA 0,2M e incubar a 65°C durante 5 minutos.

Finalizada la cinética, los extremos digeridos del vector son imperfectamente romos, con lo cual se aconseja cortar con una enzima de restricción el extremo que no deseamos secuenciar, para contar con un extremo cohesivo que facilite la posterior clonación.

Cada uno de los clones conseguidos puede usarse, con la debida planificación, para nuevas rondas de digestión con *Bal31*, consiguiéndose, así, una colección de clones con deleciones anidadas de un fragmento de DNA original (ver Fig. 7.3).

La secuenciación de la colección de clones con el cebador M13 -20, presente en el vector pTZ18U, permite la obtención de secuencias solapadas y, por lo tanto, alineables, cada vez más internas en el fragmento de DNA original.

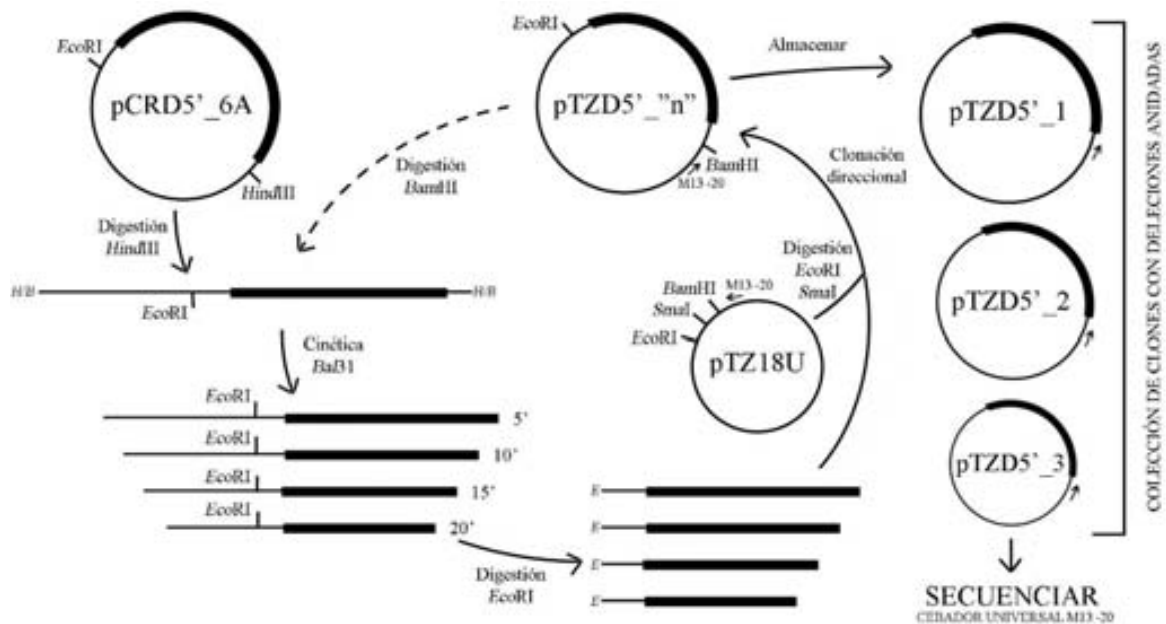


Fig. 7.3. La reacción de *Bal31*. Esquema general de la reacción de *Bal31* y los pasos de digestión y clonación asociados. En los plásmidos, la barra negra representa la región que deseamos secuenciar.

7.1.10- CULTIVO CELULAR DE LA LÍNEA S2 DE *Drosophila melanogaster*

A lo largo de los experimentos de cultivo celular, en todo momento se han usado células Schneider-2 (Invitrogen Corp., EE.UU.) procedentes del mismo stock congelado. Para la descongelación, mantenimiento y congelación de las células se han seguido los protocolos y precauciones sugeridas por Invitrogen.

El cultivo de las células se ha realizado en *Schneider's insect medium* (Sigma-Aldrich, EE.UU.), preparado según el manual de Sigma, suplementado con 10% Suero Fetal Bovino (FBS) inactivado, 50U Penicilina-G/mL y 50µg Estreptomicina/mL.

Los cultivos se mantienen a oscuras a 25°C en un incubador sin sistema de CO₂ (J.P. Selecta S.A., España).

7.1.10.1- Transfección Transitoria de S2

La transfección transitoria de células de la línea S2 se ha realizado con el kit comercial *FuGene® 6 Transfection Reagent* (Roche, Alemania) siguiendo las especificaciones del protocolo propuesto. Se ha usado una proporción 3:1 de *FuGene®*:DNA. Este kit se basa

en la formación de complejos entre el DNA que deseamos introducir en las células y una mezcla de lípidos. El resultado es la formación de liposomas: minúsculas burbujas lipídicas con un lumen interno rico en nuestro DNA recombinante. Estos liposomas son virtualmente inocuos para las células, lo que facilita la optimización de la técnica y permite trabajar con menos cantidad de material. Puesto que los lípidos son capaces de fusionarse con la célula sin casi intermediarios, la eficiencia de transfección resultante es muy elevada. Las células son capaces de captar y expresar este DNA durante un período corto de tiempo (2 o 3 días) en ausencia de presión selectiva. En nuestros experimentos, la expresión de las construcciones recombinantes se ha comprobado el segundo día después de la transfección.

7.1.10.2- Localización subcelular de proteínas de fusión con GFP

La localización subcelular de las proteínas de fusión expresadas en las células transfectadas se ha efectuado por observación directa de la fluorescencia de eGFP, eYFP y eCFP, en un microscopio de fluorescencia Olympus BX50 (Japón) equipado con una cámara digital Leica DC350F (Alemania). El protocolo abreviado es el siguiente:

TRATAMIENTO DE LOS CUBREOBJETOS: Cubrir una de las caras de un cubreobjetos con una fina lámina de Poli-L-lisina 0,1%. Dejar a temperatura ambiente durante 5 minutos. Enjuagar dos veces los cubreobjetos con agua destilada y secarlos 1h a 65°C.

ADHESIÓN DE LAS CÉLULAS: Centrifugar 5 minutos a 1000RPM, de 2 a 3mL de células de un cultivo a un 80-90% de confluencia. Lavar las células dos veces con 1mL de 1xPBS. Resuspenderlas en 400-600µL de 1xPBS. Cubrir con 100µL de células la cara tratada con Poli-L-lisina de un cubreobjetos. Dejar a temperatura ambiente durante 30-40 minutos.

MONTAJE: Succionar el PBS de los cubreobjetos (las células se mantienen adheridas). Añadir 25µL de solución de montaje con DAPI 0,2µg/mL y solución *antifade* para prevenir la pérdida de fluorescencia a corto plazo. Poner el cubreobjetos en un portaobjetos, retirar el exceso de solución de montaje y sellar. Observar al microscopio de fluorescencia.

7.1.10.3- Vectores de transfección

Los tres vectores de transfección usados en este trabajo se construyeron sobre el vector de expresión pPL18, creado en nuestro laboratorio como una variación de pPL17 (amablemente cedido por la Dra. Marie-Lou Pardue del *Massachusetts Institute of Technology* en Cambridge, MA, EE.UU.). pPL17 es, a su vez, una modificación del vector pEGFP-N1 (Invitrogen, EE.UU.), al que se le reemplazó el promotor vírico CMV por el del gen *armadillo* de *D. melanogaster* (datos no publicados). Adicionalmente, el cassette

eGFP fue reemplazado por eYFP y eCFP para dar lugar a tres versiones del vector pPL17 con tres proteínas fluorescentes distintas.

El clonaje direccional de los amplicones de cada ORF se llevó a cabo mediante el uso de cebadores portadores de colas 5' con dianas de restricción. Los cebadores 5' (ORF1f y ORF2f) tienen la secuencia de reconocimiento de la enzima *KpnI* (GGTACC), mientras que los cebadores 3' (ORF1r y ORF2r) tienen la secuencia de *XmaI* (CCCGGG). Por supuesto, ninguna de las dos enzimas tiene dianas de restricción en la secuencia de *NOF*. La digestión de los productos de amplificación y de los vectores con estas enzimas de restricción, da lugar a fragmentos compatibles *KpnI-XmaI* que solo pueden ligarse en una orientación. Puesto que *KpnI* ocupa una posición 5' respecto a *XmaI* en el MCS de pPL18, el resultado de la clonación es la inserción de cada ORF en dirección 5'-3' respecto al promotor *armadillo* y a la proteína fluorescente (ver Fig. 7.4). Aunque la planificación del experimento fue bastante cuidadosa, la fase de la zona de fusión entre cada ORF y su respectiva proteína fluorescente se aseguró mediante secuenciación desde el cebador EGFP_N propuesto en el manual del vector original pEGFP_N1 (Invitrogen, EE.UU.).

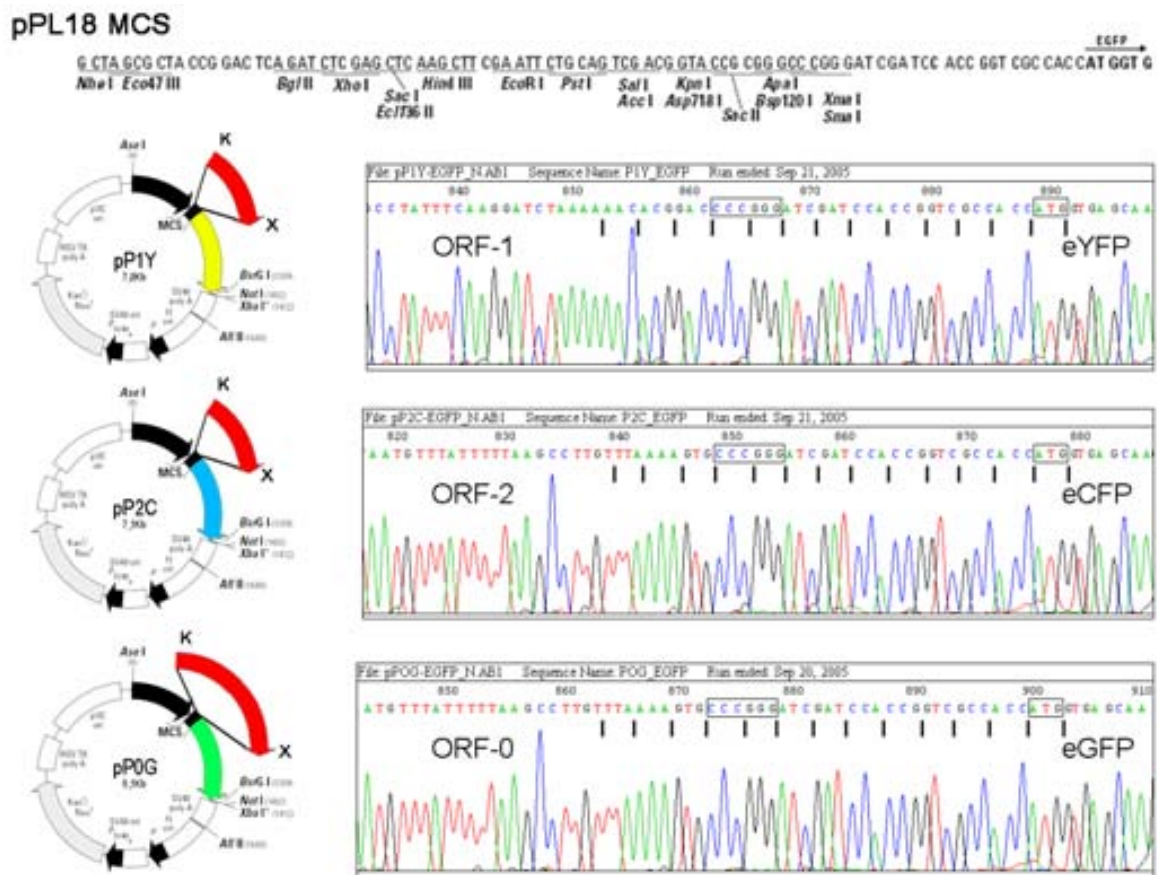


Fig. 7.4. Los vectores de Transfección. El esquema superior muestra el MCS de pPL18. Debajo, esquemas de los vectores de transfección con la secuencia de transición entre el ORF de *NOF* y la proteína fluorescente.

7.2- TÉCNICAS *in silico*

7.2.1- BÚSQUEDA Y ALINEAMIENTO DE SECUENCIAS DE DNA

Las secuencias de los genes *white* y *zeste*, y de sus productos de expresión, así como las anotaciones históricas de los elementos *FB* y *NOF*, fueron obtenidas mediante búsquedas por palabra clave en las bases de datos públicos *Medline* y *FlyBase* (<http://www.pubmed.gov> y <http://flybase.bio.indiana.edu> respectivamente). Las regiones genómicas adyacentes a los genes o elementos estudiados se han obtenido mediante alineamientos con la base de datos usando la herramienta *Blast* propia de *FlyBase* (<http://flybase.bio.indiana.edu/blast/>) y restringiendo la búsqueda a sólo las secuencias anotadas por el consorcio para la secuenciación del genoma de *D. melanogaster*. Para la comparación de secuencias se han producido alineamientos “a pares” (<http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi>) y alineamientos múltiples con la herramienta *ClustalW* (<http://www.ebi.ac.uk/clustalw/>), aunque sin desprender de los resultados ninguna relación evolutiva.

7.2.2- PREDICCIÓN DE PAUTAS DE LECTURA

La secuencia del elemento transponible *FB-NOF* fue analizada mediante los siguientes programas informáticos:

- Análisis de secuencias codificantes
 - *GenScan* (MIT) <http://genes.mit.edu/GENSCAN.html>
 - *ORF Finder* (NCBI) <http://www.ncbi.nlm.nih.gov/projects/gorf/>
- Predicción funcional de los péptidos predichos
 - *ProtLoc* (UAB) <http://bioinf.uab.es/cgi-bin/trsdb/protloc.cgi>
 - *TransMem* (UAB) <http://bioinf.uab.es/cgi-bin/trsdb/transmem.cgi>
 - *SOSUI* (Nagoya University) http://bp.nuap.nagoya-u.ac.jp/~sosui/sosui_submit.html
 - *PHYRE* (Imperial College London) <http://www.sbg.bio.ic.ac.uk/phyre/>

ProtLoc es una herramienta que predice la localización subcelular de un péptido, basándose en la composición de aminoácidos (Cedano *et al.*, 1997). *TransMem* busca dominios potencialmente transmembranales (Aloy *et al.*, 1997). *SOSUI* calcula la hidrofobicidad media de un péptido y de sus dominios potencialmente transmembranales. Predice si se trata de una proteína soluble (citosólica) o si, por el contrario, contiene hélices potencialmente transmembranales (Hirokawa *et al.*, 1998). Si las predice, muestra gráficamente la localización de las hélices y su estructuración en una membrana celular. *PHYRE* es un predictor de la estructura terciaria de un péptido (*PHYRE* es una evolución de 3D-PSSM, ver Kelley *et al.*, 2000). Mediante cálculos biofísicos, predice la forma tridimensional de una proteína a partir de su secuencia. Además, compara la homología estructural con otros péptidos de estructura conocida, en una especie de alineamiento funcional. *PHYRE* genera coordenadas en formato “.pdb” que pueden visualizarse con el programa de acceso libre *DeepView - Swiss-PdbViewer* (<http://expasy.org/spdbv/>) (Guex y Peitsch, 1997).

7.2.3- GENERACIÓN E INTERPRETACIÓN DE LOGOS DE SECUENCIA

Los logos de secuencia son formas gráficas de visualizar la información de una secuencia de DNA, RNA o proteína. Los programas que generan logos de secuencia usan, para ello, los conceptos básicos de la teoría de la información de Claude Shannon. El logo muestra la frecuencia y la conservación de cada base en cada posición de la secuencia analizada, así como su relevancia, reflejada en forma de los bits de información que contiene (Shaner *et al.*, 1993). En este trabajo se ha utilizado, para la generación de logos de secuencia, la herramienta en línea *Weblogo* (<http://weblogo.berkeley.edu>).

7.2.4- CREACIÓN DE UN PROGRAMA SIMPLE DE BÚSQUEDA AUTOMATIZADA

El lenguaje de programación *perl* es, posiblemente, el más utilizado por los biólogos para la creación de pequeños programas que, aunque simples, agilizan algunas tareas que pueden consumir gran cantidad de tiempo. Tales son el formateo de secuencias, la búsqueda de patrones o la descarga de información de las bases de datos públicas en base a criterios específicos. Cualquier entrada de base de datos, ya sea una secuencia de DNA, RNA o aminoacídica, es susceptible de poder ser procesada, pues la principal conveniencia

de *perl* para los biólogos moleculares es la facilidad con la que puede manejar archivos de texto.

En este trabajo se ha usado la versión de *perl* PXXPerl 5.8.7-6 para crear el programa `cazaFB.pl`, un sencillo *script* para encontrar elementos *FB* en secuencias genómicas.

Para empezar, el usuario de `cazaFB.pl` debía proporcionar un listado con los números de acceso a secuencias que puedan contener un elemento *FB*. Esta lista se elaboraba a partir de un alineamiento de los primeros 120bp de *FB* con la base de datos de *FlyBase*, usando la herramienta *Blast* y seleccionando la opción de buscar únicamente en secuencias derivadas del proyecto de secuenciación genómica. Conseguíamos, así, los accesos a las secuencias muy largas (~150Kb) de los *contigs* usados para el ensamblaje de la secuencia completa del genoma de *D. melanogaster*, que mostraban regiones de similitud con *FB*.

El *script* `cazaFB.pl` memorizaba los números de acceso de cada una de las secuencias, se conectaba con la base de datos *GenBank* y descargaba cada uno de los archivos de secuencia, con toda la información suplementaria proporcionada. A partir de cada uno de los archivos, `cazaFB.pl` separaba la secuencia completa del resto de información e identificaba como un elemento *FB* cualquier secuencia que empezase con los primeros 27bp de *FB* y terminara con los últimos 27bp de *FB* (iguales que los primeros, pero invertidos y complementarios).

Debido a los cambios de diseño en la base de datos *FlyBase*, `cazaFB.pl` debió ser modificado. A partir de septiembre de 2006, todas las secuencias genómicas de *D. melanogaster* se refieren al mismo número de secuencia, pues finalmente se unieron todos los fragmentos del genoma en un solo archivo para cada brazo cromosómico. En consecuencia, la estrategia basada en *Blast* de `cazaFB.pl` quedó inutilizada, de modo que decidimos descargar las secuencias genómicas enteras y adaptar el *script* para trabajar localmente.

En el método de reconocimiento de *FB* se ha tenido en cuenta la posibilidad de que hubiera más de un elemento *FB* en una sola secuencia. En este caso, `cazaFB.pl` identifica cada una por separado e informa al usuario del número de *FB* encontrados. El programa también analiza los 9bp anteriores y posteriores de cada *FB* e informa de si son idénticos, cómo cabría esperar pues *FB* genera duplicaciones directas de 9bp en sus sitios de inserción, o bien si son distintos, caso de recombinación no homóloga entre dos

elementos. Otras informaciones presentadas para cada elemento son su localización citológica, según lo indicado por *GenBank*, su longitud en pares de bases y los distintos identificadores de acceso a la secuencia.

El siguiente paso del programa consiste en analizar cada una de las secuencias de *FB* encontradas y buscar elementos *NOF* en su interior. Para eso, se ha usado un procedimiento análogo a la búsqueda del mismo *FB*, con la identificación como *NOF* de cualquier región interna de *FB* que se inicie con los primeros 17bp de *NOF* y termine con los mismos en reverso complementario. La última versión de `cazaFB.pl`, además, exige que la secuencia *NOF* esté situada en finales de segunda repetición de bloque de *FB* por cada extremo para evitar falsos positivos (ver descripción general de *FB-NOF* en el apartado 5.2). Frente a la posibilidad de encontrar elementos *NOF* truncados, en el caso de no encontrar un *NOF* completo `cazaFB.pl` comprueba si existe por lo menos uno de los extremos de *NOF* (primero el 5' y después el 3') informando, en caso afirmativo, de la presencia de un elemento *NOF* truncado en ese *FB*.

La salida del programa `cazaFB.pl` resulta en la creación de tres archivos de texto con toda la información extraída por el algoritmo:

FBout(Cr).txt	Localización citológica
	Número total de <i>FBs</i> en secuencia
	Secuencia entera de cada elemento <i>FB</i> con su duplicación directa y valoración de la misma
	Presencia o ausencia de <i>NOF</i> y su integridad
FBclust(Cr).txt	Formato FASTA de cada secuencia de <i>FB</i> encontrada. El encabezado está formado por la localización citológica, el número de <i>FB</i> para esa secuencia y el tamaño. Listo para ser procesado por el programa de alineamiento múltiple <i>ClustalW</i>
Informe(Cr).txt	Número total de secuencias analizadas
	Número total de <i>FB</i> encontrados
	Número total de <i>NOF</i> encontrados y su integridad
	Tamaños medios de <i>FB</i> según presencia o ausencia de elementos <i>NOF</i> en su interior

8. BIBLIOGRAFÍA

Agrawal, A., Eastman, Q. M. and Schatz, D. G. (1998). "Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system." Nature **394**(6695): 744-51.

Aloy, P., Cedano, J., Oliva, B., Aviles, F. X. and Querol, E. (1997). "'TransMem': a neural network implemented in Excel spreadsheets for predicting transmembrane domains of proteins." Comput Appl Biosci **13**(3): 231-4.

Arkhipova, I. R., Lyubomirskaya, N. V. and Ilyin, Y. V. (1995). Drosophila Retrotransposons. Heidelberg, R.G. Landes Company.

Baldrich, E. (1998). Análisis genético y molecular de los mutantes de *Drosophila melanogaster* obtenidos a partir de individuos *white-apricot* sometidos a choque térmico y a tratamiento con N-etil-N-nitrosurea. Departament de Genètica i de Microbiologia. Bellaterra, Universitat Autònoma de Barcelona.

Baldrich, E., Velazquez, A., Xamena, N. and Cabre, O. (2003). "Germline mutations induced by N-nitroso-N-ethylurea do not affect the inserted copia retrotransposon in a *Drosophila melanogaster* *w^a* mutant." Mutagenesis **18**(6): 527-31.

Bartolome, C., Maside, X. and Charlesworth, B. (2002). "On the abundance and distribution of transposable elements in the genome of *Drosophila melanogaster*." Mol Biol Evol **19**(6): 926-37.

Batiste-Alentorn, M., Xamena, N., Creus, A. and Marcos, R. (1991). "Genotoxicity studies with the unstable *zeste-white* (UZ) system of *Drosophila melanogaster*: results with ten carcinogenic compounds." Environ Mol Mutagen **18**(2): 120-5.

Blumenthal, T. (2004). "Operons in eukaryotes." Brief Funct Genomic Proteomic **3**(3): 199-211.

Brierley, H. L. and Potter, S. S. (1985). "Distinct characteristics of loop sequences of two *Drosophila foldback* transposable elements." Nucleic Acids Res **13**(2): 485-500.

Byrd, K. and Corces, V. G. (2003). "Visualization of chromatin domains created by the gypsy insulator of *Drosophila*." J Cell Biol **162**(4): 565-74.

Caceres, M., Ranz, J. M., Barbadilla, A., Long, M. and Ruiz, A. (1999). "Generation of a widespread *Drosophila* inversion by a transposable element." Science **285**(5426): 415-8.

Capy, P., Bazin, C., Higuët, D. and Langin, T. (1998). Dynamics and evolution of transposable elements. Heidelberg, Landes Bioscience.

Cedano, J., Aloy, P., Perez-Pons, J. A. and Querol, E. (1997). "Relation between amino acid composition and cellular location of proteins." J Mol Biol **266**(3): 594-600.

Celniker, S. E., Wheeler, D. A., Kronmiller, B., Carlson, J. W., Halpern, A., Patel, S., Adams, M., Champe, M., Dugan, S. P., Frise, E., Hodgson, A., George, R. A., Hoskins, R. A., Lavery, T., Muzny, D. M., Nelson, C. R., Pacleb, J. M., Park, S., Pfeiffer, B. D.,

Richards, S., Sodergren, E. J., Svirskas, R., Tabor, P. E., Wan, K., Stapleton, M., Sutton, G. G., Venter, C., Weinstock, G., Scherer, S. E., Myers, E. W., Gibbs, R. A. and Rubin, G. M. (2002). "Finishing a whole-genome shotgun: release 3 of the *Drosophila melanogaster* euchromatic genome sequence." Genome Biol **3**(12): RESEARCH0079.

Chen, J. D., Chan, C. S. and Pirrotta, V. (1992). "Conserved DNA binding and self-association domains of the *Drosophila zeste* protein." Mol Cell Biol **12**(2): 598-608.

Chen, J. D. and Pirrotta, V. (1993). "Stepwise assembly of hyperaggregated forms of *Drosophila zeste* mutant protein suppresses *white* gene expression *in vivo*." Embo J **12**(5): 2061-73.

Cheng, C., Daigen, M. and Hirochika, H. (2006). "Epigenetic regulation of the rice retrotransposon Tos17." Mol Genet Genomics **276**(4): 378-90.

Collins, M. and Rubin, G. M. (1983). "High-frequency precise excision of the *Drosophila foldback* transposable element." Nature **303**(5914): 259-60.

Driver, A., Lacey, S. F., Cullingford, T. E., Mitchelson, A. and O'Hare, K. (1989). "Structural analysis of *Doc* transposable elements associated with mutations at the *white* and *suppressor of forked* loci of *Drosophila melanogaster*." Mol Gen Genet **220**(1): 49-52.

Druker, R., Bruxner, T. J., Lehrbach, N. J. and Whitelaw, E. (2004). "Complex patterns of transcription at the insertion site of a retrotransposon in the mouse." Nucleic Acids Res **32**(19): 5800-8.

Feschotte, C., Osterlund, M. T., Peeler, R. and Wessler, S. R. (2005). "DNA-binding specificity of rice *mariner-like* transposases and interactions with *Stowaway* MITES." Nucleic Acids Res **33**(7): 2153-65.

Finnegan, D. J. (1989). "Eukaryotic transposable elements and genome evolution." Trends Genet **5**(4): 103-7.

Finnegan, D. J. (1992). "Transposable elements." Curr Opin Genet Dev **2**(6): 861-7.

Fjose, A., Polito, L. C., Weber, U. and Gehring, W. J. (1984). "Developmental expression of the *white* locus of *Drosophila melanogaster*." Embo J **3**(9): 2087-2094.

Fujikawa, K. and Kondo, S. (1986). "DNA repair dependence of somatic mutagenesis of transposon-caused *white* alleles in *Drosophila melanogaster* after treatment with alkylating agents." Genetics **112**(3): 505-22.

Gans, M. (1949). "*zeste* (z): Mutation liee au sexe et limitee par le sexe chez *Drosophila melanogaster*." Bulletin biologique de la France et de la Belgique **83**: 136-157.

Gilbert, N., Lutz-Prigge, S. and Moran, J. V. (2002). "Genomic deletions created upon LINE-1 retrotransposition." Cell **110**(3): 315-25.

- Goldberg, M. L., Paro, R. and Gehring, W. J. (1982). "Molecular cloning of the *white* locus region of *Drosophila melanogaster* using a large transposable element." Embo J **1**(1): 93-98.
- Golubovskiy, M. D., Ivano, Y. N. and Green, M. M. (1977). "Genetic instability in *Drosophila melanogaster*: putative multiple insertion mutants at the *singed bristle* locus." Proc Natl Acad Sci U S A **74**(7): 2973-5.
- Guex, N. and Peitsch, M. C. (1997). "SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling." Electrophoresis **18**(15): 2714-23.
- Han, J. S. and Boeke, J. D. (2005). "LINE-1 retrotransposons: modulators of quantity and quality of mammalian gene expression?" Bioessays **27**(8): 775-84.
- Han, J. S., Szak, S. T. and Boeke, J. D. (2004). "Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes." Nature **429**(6989): 268-74.
- Harden, N. and Ashburner, M. (1990). "Characterization of the *FB-NOF* transposable element of *Drosophila melanogaster*." Genetics **126**(2): 387-400.
- Hazelrigg, T. (1987). "The *Drosophila white* gene: a molecular update." Trends in Genetics **3**(2): 43-47.
- Hirokawa, T., Boon-Chieng, S. and Mitaku, S. (1998). "SOSUI: classification and secondary structure prediction system for membrane proteins." Bioinformatics **14**(4): 378-9.
- Holt, R. A., Subramanian, G. M., Halpern, A., Sutton, G. G., Charlab, R., Nusskern, D. R., Wincker, P., Clark, A. G., Ribeiro, J. M., Wides, R., Salzberg, S. L., Loftus, B., Yandell, M., Majoros, W. H., Rusch, D. B., Lai, Z., Kraft, C. L., Abril, J. F., Anthouard, V., Arensburger, P., Atkinson, P. W., Baden, H., de Berardinis, V., Baldwin, D., Benes, V., Biedler, J., Blass, C., Bolanos, R., Boscus, D., Barnstead, M., Cai, S., Center, A., Chaturverdi, K., Christophides, G. K., Chrystal, M. A., Clamp, M., Cravchik, A., Curwen, V., Dana, A., Delcher, A., Dew, I., Evans, C. A., Flanigan, M., Grundschober-Freimoser, A., Friedli, L., Gu, Z., Guan, P., Guigo, R., Hillenmeyer, M. E., Hladun, S. L., Hogan, J. R., Hong, Y. S., Hoover, J., Jaillon, O., Ke, Z., Kodira, C., Kokoza, E., Koutsos, A., Letunic, I., Levitsky, A., Liang, Y., Lin, J. J., Lobo, N. F., Lopez, J. R., Malek, J. A., McIntosh, T. C., Meister, S., Miller, J., Mobarry, C., Mongin, E., Murphy, S. D., O'Brochta, D. A., Pfannkoch, C., Qi, R., Regier, M. A., Remington, K., Shao, H., Sharakhova, M. V., Sitter, C. D., Shetty, J., Smith, T. J., Strong, R., Sun, J., Thomasova, D., Ton, L. Q., Topalis, P., Tu, Z., Unger, M. F., Walenz, B., Wang, A., Wang, J., Wang, M., Wang, X., Woodford, K. J., Wortman, J. R., Wu, M., Yao, A., Zdobnov, E. M., Zhang, H., Zhao, Q., Zhao, S., Zhu, S. C., Zhimulev, I., Coluzzi, M., della Torre, A., Roth, C. W., Louis, C., Kalush, F., Mural, R. J., Myers, E. W., Adams, M. D., Smith, H. O., Broder, S., Gardner, M. J., Fraser, C. M., Birney, E., Bork, P., Brey, P. T., Venter, J. C., Weissenbach, J., Kafatos, F. C., Collins, F. H. and Hoffman, S. L. (2002). "The genome sequence of the malaria mosquito *Anopheles gambiae*." Science **298**(5591): 129-49.

Hoskins, R. A., Smith, C. D., Carlson, J. W., Carvalho, A. B., Halpern, A., Kaminker, J. S., Kennedy, C., Mungall, C. J., Sullivan, B. A., Sutton, G. G., Yasuhara, J. C., Wakimoto, B. T., Myers, E. W., Celniker, S. E., Rubin, G. M. and Karpen, G. H. (2002).

"Heterochromatic sequences in a *Drosophila* whole-genome shotgun assembly." Genome Biol **3**(12): RESEARCH0085.

Hua-Van, A., Le Rouzic, A., Maisonhaute, C. and Capy, P. (2005). "Abundance, distribution and dynamics of retrotransposable elements and transposons: similarities and differences." Cytogenet Genome Res **110**(1-4): 426-40.

Ising, G. and Block, K. (1984). "A transposon as a cytogenetic marker in *Drosophila melanogaster*." Mol Gen Genet **196**(1): 6-16.

Judd, B. H. (1995). "Mutations of *zeste* that mediate transvection are recessive enhancers of position-effect variegation in *Drosophila melanogaster*." Genetics **141**(1): 245-53.

Kal, A. J., Mahmoudi, T., Zak, N. B. and Verrijzer, C. P. (2000). "The *Drosophila brahma* complex is an essential coactivator for the *trithorax* group protein *zeste*." Genes Dev **14**(9): 1058-71.

Kaminker, J. S., Bergman, C. M., Kronmiller, B., Carlson, J., Svirskas, R., Patel, S., Frise, E., Wheeler, D. A., Lewis, S. E., Rubin, G. M., Ashburner, M. and Celniker, S. E. (2002). "The transposable elements of the *Drosophila melanogaster* euchromatin: a genomics perspective." Genome Biol **3**(12): RESEARCH0084.

Kelley, L. A., MacCallum, R. M. and Sternberg, M. J. (2000). "Enhanced genome annotation using structural profiles in the program 3D-PSSM." J Mol Biol **299**(2): 499-520.

Kozak, M. (2005). "Regulation of translation via mRNA structure in prokaryotes and eukaryotes." Gene **361**: 13-37.

Kyte, J. and Doolittle, R. F. (1982). "A simple method for displaying the hydrophobic character of a protein." J Mol Biol **157**(1): 105-32.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., LeHoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E.,

Frazier, M., Gibbs, R. A., Muzny, D. M., Scherer, S. E., Bouck, J. B., Sodergren, E. J., Worley, K. C., Rives, C. M., Gorrell, J. H., Metzker, M. L., Naylor, S. L., Kucherlapati, R. S., Nelson, D. L., Weinstock, G. M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D. R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H. M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R. W., Federspiel, N. A., Abola, A. P., Proctor, M. J., Myers, R. M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D. R., Olson, M. V., Kaul, R., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G. A., Athanasiou, M., Schultz, R., Roe, B. A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W. R., de la Bastide, M., Dedhia, N., Blocker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J. A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D. G., Burge, C. B., Cerutti, L., Chen, H. C., Church, D., Clamp, M., Copley, R. R., Doerks, T., Eddy, S. R., Eichler, E. E., Furey, T. S., Galagan, J., Gilbert, J. G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L. S., Jones, T. A., Kasif, S., Kasprzyk, A., Kennedy, S., Kent, W. J., Kitts, P., Koonin, E. V., Korf, I., Kulp, D., Lancet, D., Lowe, T. M., McLysaght, A., Mikkelsen, T., Moran, J. V., Mulder, N., Pollara, V. J., Ponting, C. P., Schuler, G., Schultz, J., Slater, G., Smit, A. F., Stupka, E., Szustakowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y. I., Wolfe, K. H., Yang, S. P., Yeh, R. F., Collins, F., Guyer, M. S., Peterson, J., Felsenfeld, A., Wetterstrand, K. A., Patrinos, A., Morgan, M. J., de Jong, P., Catanese, J. J., Osoegawa, K., Shizuya, H., Choi, S. and Chen, Y. J. (2001). "Initial sequencing and analysis of the human genome." *Nature* **409**(6822): 860-921.

Lee, S. J. (1991). "Expression of growth/differentiation factor 1 in the nervous system: conservation of a bicistronic structure." *Proc Natl Acad Sci U S A* **88**(10): 4250-4.

Levis, R., Bingham, P. M. and Rubin, G. M. (1982). "Physical map of the white locus of *Drosophila melanogaster*." *Proc Natl Acad Sci U S A* **79**(2): 564-8.

Lifschytz, E. and Green, M. M. (1984). "The *zeste-white* interaction: induction and genetic analysis of a novel class of *zeste* alleles." *Embo J* **3**(5): 999-1002.

Mandrioli, M. and Borsatti, F. (2006). "DNA methylation of fly genes and transposons." *Cell Mol Life Sci* **63**(17): 1933-6.

Mansukhani, A., Gunaratne, P. H., Sherwood, P. W., Sneath, B. J. and Goldberg, M. L. (1988). "Nucleotide sequence and structural analysis of the *zeste* locus of *Drosophila melanogaster*." *Mol Gen Genet* **211**(1): 121-8.

McClintock, B. (1950). "The origin and behavior of mutable loci in maize." *Proc Natl Acad Sci U S A* **36**(6): 344-55.

Misra, S., Crosby, M. A., Mungall, C. J., Matthews, B. B., Campbell, K. S., Hradecky, P., Huang, Y., Kaminker, J. S., Millburn, G. H., Prochnik, S. E., Smith, C. D., Tupy, J. L., Whitfield, E. J., Bayraktaroglu, L., Berman, B. P., Bettencourt, B. R., Celniker, S. E., de Grey, A. D., Drysdale, R. A., Harris, N. L., Richter, J., Russo, S., Schroeder, A. J., Shu, S. Q., Stapleton, M., Yamada, C., Ashburner, M., Gelbart, W. M., Rubin, G. M. and Lewis, S.

E. (2002). "Annotation of the *Drosophila melanogaster* euchromatic genome: a systematic review." Genome Biol **3**(12): RESEARCH0083.

Montgomery, E., Charlesworth, B. and Langley, C. H. (1987). "A test for the role of natural selection in the stabilization of transposable element copy number in a population of *Drosophila melanogaster*." Genet Res **49**(1): 31-41.

Moschetti, R., Marsano, R. M., Barsanti, P., Caggese, C. and Caizzi, R. (2004). "*FB* elements can promote exon shuffling: a promoter-less *white* allele can be reactivated by *FB* mediated transposition in *Drosophila melanogaster*." Mol Genet Genomics **271**(4): 394-401.

Myers, E. W., Sutton, G. G., Delcher, A. L., Dew, I. M., Fasulo, D. P., Flanigan, M. J., Kravitz, S. A., Mobarry, C. M., Reinert, K. H., Remington, K. A., Anson, E. L., Bolanos, R. A., Chou, H. H., Jordan, C. M., Halpern, A. L., Lonardi, S., Beasley, E. M., Brandon, R. C., Chen, L., Dunn, P. J., Lai, Z., Liang, Y., Nusskern, D. R., Zhan, M., Zhang, Q., Zheng, X., Rubin, G. M., Adams, M. D. and Venter, J. C. (2000). "A whole-genome assembly of *Drosophila*." Science **287**(5461): 2196-204.

Narita, N., Nishio, H., Kitoh, Y., Ishikawa, Y., Minami, R., Nakamura, H. and Matsuo, M. (1993). "Insertion of a 5' truncated L1 element into the 3' end of exon 44 of the dystrophin gene resulted in skipping of the exon during splicing in a case of Duchenne muscular dystrophy." J Clin Invest **91**(5): 1862-7.

O'Hare, K., Levis, R. and Rubin, G. M. (1983). "Transcription of the *white* locus in *Drosophila melanogaster*." Proc Natl Acad Sci U S A **80**(22): 6917-6921.

O'Hare, K., Murphy, C., Levis, R. and Rubin, G. M. (1984). "DNA sequence of the *white* locus of *Drosophila melanogaster*." J Mol Biol **180**(3): 437-55.

Okamoto, H. and Hirochika, H. (2001). "Silencing of transposable elements in plants." Trends Plant Sci **6**(11): 527-34.

Pardue, M. L. and DeBaryshe, P. G. (2003). "Retrotransposons provide an evolutionarily robust non-telomerase mechanism to maintain telomeres." Annu Rev Genet **37**: 485-511.

Paro, R., Goldberg, M. L. and Gehring, W. J. (1983). "Molecular analysis of large transposable elements carrying the *white* locus of *Drosophila melanogaster*." Embo J **2**(6): 853-860.

Piñol, J., Francino, O., Fontdevila, A. and Cabré, O. (1988). "Rapid isolation of *Drosophila* high molecular weight DNA to obtain genomic libraries." Nucleic Acids Res **16**(6): 2736.

Pirrotta, V. (1991). "The genetics and molecular biology of *zeste* in *Drosophila melanogaster*." Adv Genet **29**: 301-48.

Pirrotta, V., Manet, E., Hardon, E., Bickel, S. E. and Benson, M. (1987). "Structure and sequence of the *Drosophila zeste* gene." Embo J **6**(3): 791-9.

- Plasterk, R. H. (2002). "RNA silencing: the genome's immune system." Science **296**(5571): 1263-5.
- Potter, S., Truett, M., Phillips, M. and Maher, A. (1980). "Eucaryotic transposable genetic elements with inverted terminal repeats." Cell **20**(3): 639-47.
- Potter, S. S. (1982). "DNA sequence of a *foldback* transposable element in *Drosophila*." Nature **297**(5863): 201-4.
- Rasmuson, A. (1984). "Effects of DNA-repair-deficient mutants on somatic and germ line mutagenesis in the UZ system in *Drosophila melanogaster*." Mutat Res **141**(1): 29-33.
- Rasmuson, A. and Larsson, J. (1992). "Somatic and germline mutagenesis assayed by the unstable *zeste-white* test in *Drosophila melanogaster*." Mutagenesis **7**(3): 219-23.
- Rasmuson-Lestander, A. and Ekstrom, K. (1996). "Genetic and molecular analysis of a set of unstable *white* mutants in *Drosophila melanogaster*." Genetica **98**(2): 179-92.
- Rasmuson-Lestander, A., Larsson, J. and Rasmuson, B. (1993). "Position-effect variegation and *z1* mediated *white* repression in the *In(1)wis* system in *Drosophila melanogaster*." Hereditas **119**(3): 209-18.
- Razin, A. (1998). "CpG methylation, chromatin structure and gene silencing-a three-way connection." Embo J **17**(17): 4905-8.
- Rebatchouk, D. and Narita, J. O. (1997). "Foldback transposable elements in plants." Plant Mol Biol **34**(5): 831-5.
- Rosen, C., Dorsett, D. and Jack, J. (1998). "A proline-rich region in the Zeste protein essential for transvection and *white* repression by Zeste." Genetics **148**(4): 1865-74.
- Salzberg, A., Fisher, O., Siman-Tov, R. and Ankri, S. (2004). "Identification of methylated sequences in genomic DNA of adult *Drosophila melanogaster*." Biochem Biophys Res Commun **322**(2): 465-9.
- Sambrook, J., Fritsch, E. and Maniatis, T. (1989). Molecular Cloning, Cold Spring Harbour Laboratory Press.
- Shaner, M., Blair, I. and Schneider, T. (1993). Sequence logos: a powerful, yet simple, tool. Proceedings of the twenty-sixth annual hawaii international conference on system sciences, IEEE Computer Society Press, Los Alamitos, CA.
- Shapiro, J. A. and Adhya, S. L. (1969). "The galactose operon of *E. coli* K-12. II. A deletion analysis of operon structure and polarity." Genetics **62**(2): 249-64.
- Spieth, J., Brooke, G., Kuersten, S., Lea, K. and Blumenthal, T. (1993). "Operons in *C. elegans*: polycistronic mRNA precursors are processed by trans-splicing of SL2 to downstream coding regions." Cell **73**(3): 521-32.

- Stram, Y. and Kuzntzova, L. (2006). "Inhibition of viruses by RNA interference." Virus Genes **32**(3): 299-306.
- Tearle, R. G., Belote, J. M., McKeown, M., Baker, B. S. and Howells, A. J. (1989). "Cloning and characterization of the *scarlet* gene of *Drosophila melanogaster*." Genetics **122**(3): 595-606.
- Templeton, N. S. and Potter, S. S. (1989). "Complete *foldback* transposable elements encode a novel protein found in *Drosophila melanogaster*." Embo J **8**(6): 1887-94.
- Truett, M. A., Jones, R. S. and Potter, S. S. (1981). "Unusual structure of the *FB* family of transposable elements in *Drosophila*." Cell **24**(3): 753-63.
- Vogel, E. W. (1989). "Somatic cell mutagenesis in *Drosophila*: recovery of genetic damage in relation to the types of DNA lesions induced in mutationally unstable and stable X-chromosomes." Mutat Res **211**(1): 153-70.
- Wheelan, S. J., Aizawa, Y., Han, J. S. and Boeke, J. D. (2005). "Gene-breaking: a new paradigm for human retrotransposon-mediated gene evolution." Genome Res **15**(8): 1073-8.
- Wong, L. H. and Choo, K. H. (2004). "Evolutionary dynamics of transposable elements at the centromere." Trends Genet **20**(12): 611-6.
- Wu, X. and Burgess, S. M. (2004). "Integration target site selection for retroviruses and transposable elements." Cell Mol Life Sci **61**(19-20): 2588-96.
- Xamena, N., Creus, A. and Marcos, R. (1984). "Mutagenic activity of some intercalating compounds in the *Drosophila zeste* somatic eye mutation test." Mutat Res **138**(2-3): 169-73.

9. ANEXOS

ÍNDICE DE ANEXOS

1. Secuencia completa del IR 5' del elemento <i>FB</i> estudiado en M115.....	144
2. Secuencia completa del IR 3' del elemento <i>FB</i> estudiado en M115.....	146
3. Secuencia de la parte <i>NOF</i> según Templeton y Potter en 1989.....	148
4. Secuencia de la parte <i>NOF</i> según Harden y Ashburner en 1990.....	150
5. Informe del análisis bioinformático del ORF-0.....	152
6. Informe del análisis bioinformático del ORF-1.....	155
7. Informe del análisis bioinformático del ORF-2.....	157
8. Informe del análisis bioinformático del ORF-3.....	159
9. Código <i>Perl</i> del programa <i>cazaFB.pl</i>	161
10. Artículo “An <i>FB-NOF</i> mediated duplication of the <i>white</i> gene is responsible for the <i>zeste</i> ¹ phenotype in some <i>Drosophila melanogaster</i> unstable strains”	
11. Artículo “Molecular and bioinformatic analysis of the <i>FB-NOF</i> transposable element”	

EXTREMO 5' *FB-NOF* M115 [DOWNSTREAM *white* → *NOF*]

CAGTGGGACAAAGTGCCTAAATTTGGATATATAATAAATCCAAAACCAAAAAGAAAACAA
 AAAAGGGTAAAATATTTTACTGTTGTTAAAATTCGATCATTATTACTGATATTAA
 TATAGTAGATATAATCGATTTAAATTCCTTGCACCTCATGTTTCATCGTTTTTCGATTTAAAT
 TATAATTACAATTATCTGTTGAATATCTCGGTTGAGCAGTGTGCAGGGTGGCGAATTGCTT
 TACTTCGCATGCAAAACACCGCCTCAACTTTGACCCACTGTGCGGCCAAGGCTTTTCGCTT
 TTCGCTTTTATACACTATACTATATACTTTATATTTGTTTTGAGCGCGGCTTCGCAGCGAG
 TGGAAGTGGAAGAGAAATCGCTTGCATTGGCGCGCTTTTGAACGCTCACAAATGTGAGCTC
 AAAGAAGCTGGGGTCGGAAAAATCGAATTTTGAATTTGAAAGCTGGAATCGTTTGCCCA
 TTTTTTGCCCATGTTTGCCCAACCAATTAGTTTTTTTTTGCCACGTCCAGTTTTTGAGATAT
 GAATTTTCGAAAAAGTTCGAAAATTTTCGAAAATCAAAAATTCGCTTTTTCCAAAATTTT
 TTTTTTTAAATCGCAATAACATCGTTTGCCACGTTTGCCACCCCTTAGAATTTTGAAAA
 AATTTATACTTTAGAAAATATAAGGCTTTTAAGTTTACCTCGGTCTAATCAGAGAGTAAAT
 CGTTTGCCCATCTCTTAAAACCAAATATTATCAACAAAAACGTTTGCCCAACCAATTATTA
 CTAGTTTTTATCGTTTGCCACCCCTTTAAAAACCTTTAACAAAAATTTTTTTTCGATTGC
 CCACATTTAAAATACACCCAATTTTCGTTAGCCACCTCTTAAAATAAAAATTTCCAATAA
 AAAACGTTTGCCACCAATTTAAAACCTAAATAATTTTCGTTTGCCCATCCTTTAAAATTCAT
 TTTAACGTTTGCCACCCCTTTAAAATAAATTTATTTTCGTTTGCCACCCCTTTAAAATTTGT
 TTTTTTCGTTTGCCCATCTTAAAACCTAAATAATTTTCGATTGCCACCTTTTAAAACCTAAA
 TAATTTTCGTTTGCCCATCCTTTAAAATTCATTTTTAACGTTTGCCACCCCTTTAAAATAA
 ATTATTTTCGTTTGCCACCCCTTTAAAATTTGTTTTTTTCGTTTGCCCATCTTAAAACCTAA
 ATAATTTTCGATTGCCACCTTTTAAAACCTAAATAATTTTCGTTTGCCCATCCTTTAAAATTC
 ATTTTTAACGTTTGCCACCCCTTTAAAATAAATTTATTTTCGTTTGCCACCCCTTTAAAATTT
 TGTTTTTTTCGTTTGCCCATCTTAAAACCTAAATAATTTTCGATTGCCACCTTTTAAAACCT
 AAATAATTTTCGTTTGCCCATCCTTTAAAATTCATTTTTAACGTTTGCCACCCCTTTAAAAA
 TAAATTATTTTCGTTTGCCACCCCTTTAAAATTTGTTTTTTTCGTTTGCCCATCTTAAAAC
 TAAATAATTTTCGATTGCCACCTTTTAAAACCTAAATAATTTTCGTTTGCCACCCCTTTAAAA
 ATAAATTATTTTCGTTTGCCACCCCTTTAAAATTTGTTTTTTTCGTTTGCCCATCTTAAA
 CTAAATAATTTTCGATTGCCACCTTTTAAAACCTAAATAATTTTCGTTTGCCCATCCTTTAAA
 ATTCATTTTTAACGTTTGCCACCCCTTTAAAATAAATTTATTTTCGTTTGCCACCCCTTTAA
 AATTTGTTTTTTTCGTTTGCCCATCTTAAAACCTAAATAATTTTCGATTGCCACCTTTTAA
 AACTAAATAATTTTCGTTTGCCCATCCTTTAAAATTCATTTTTAACGTTTGCCACCCCTTTA

AAAATAAATTATTTTCGTTTGCCACCCCTTTAAAATTTGTTTTTTTCGTTTGCCCACTCTTA
 AACTAAATAATTTTCGATTGCCACCCCTTTAAAATAAATAATTTTCGTTTGCCCATCCTTT
 AAAATTCATTTTAAACGTTTGCCACCCCTTTAAAATAAATAATTTTCGTTTGCCACCCCTT
 TAAAATTTGTTTTTTTCGTTTGCCCACTCTTAAAATAAATAATTTTCGATTGCCACCCCTT
 TAAAATAAATAATTTTCGTTTGCCCATCCTTTAAAATTCATTTTAAACGTTTGCCACCCCT
 TTAAAATAAATAATTTTCGTTTGCCACCCCTTTAAAATTTGTTTTTTTCGTTTGCCCACTCT
 TAAAATAAATAATTTTCGATTACCCACCCCTTTAAAATAAATAATTTTCGTTTGCCCATCCT
 TTAAAATTCATTTTAAACGTTTGCCACCCCTTTAAAATAAATAATTTTCGTTTGCCACCC
 TTTAAAATTTGTTTTTTTCGTTTGACCACTCTTAAAATAAATAATTTTCGATTGCCACCC
 TTTAAAATAAATAATTTCTTTTGCCCATCCTAAAATCTTAAGAATAAATGATTCTATTGC
 CCACCATATAAACACGTGCCACTTTCCTAGTTTTAAAATACTGCCTACATAACACGTGCAGA
 CGCACAGGTGTTTCTGGGTTTATATAGACCAAAAATTGGTTCCGATTGCCAATCTTGTTAT
 TTACAGTTTACC

Total: 2635 bp

1..422: Secuencia genómica *downstream*

423..2472: *FB 5' Inverted Repeat*

2472..2635: *NOF Inverted Repeat*

EXTREMO 3' FB-NOF M115 [NOF → DOWNSTREAM *white*]

TAAACTAGGAAAGTGGCACGTGTTTATATGGTGGGCAATACAATTTATTTATTCTTAAGAT
TTTAGGATGGGCAAAGAAATAATTTAGTTTTAAAAGGTGGGCAATCGAAATTATTTAGTT
 TTAA**GAG**TGGGCAAACGAAAAAACAAATTTTAAAGGGTGGGCAAACGAAATAATTTATTT
 TTAAAGGGTGGGCAAACGTTAAAAATGAATTTTAAAGGATGGGCAAACGAAATTATTTAGT
 TTTAAAGGTGGGCAATCGAAATTATTTAGTTTTAA**GAG**TGGGCAAACGAAAAAACAAAT
 TTTAAAGGGTGGGCAAACGAAATAATTTATTTTAAAGGGTGGGCAAACGTTAAAAATGAAT
 TTTAAAGGATGGGCAAACGAAATTATTTAGTTTTTAAAGGTGGGCAATCGAAATTATTTAG
 TTTTAA**GAG**TGGGCGAACGAAAAATCCAATTTTAAAGGGTGGGCAAACGAAATAATTTATT
 TTTAAAGGGTGGGCAAACGTTAAAAATGAATTTTAAAGGATGGGCAAACGAAATTATTTAG
 TTTTAAAGGTGGGCAATCGAAATTATTTAGTTTTTAA**GAG**TGGGCAAACGAAAAAACAAA
 TTTTAAAGGGTGGGCAAACGAAATAATTTATTTTAAAGGGTGGGCAAACGTTAAAAATGA
 ATTTTAAAGGATGGGCAAACGAAATTATTTAGTTTTTAAAGGTGGGCAATCGAAATTATTT
 AGTTTTTAA**GAG**TGGGCAAACGAAAAAACAAATTTTAAAGGGTGGGCAAACGAAATAATTT
 ATTTTAAAGGGTGGGCAAACGTTAAAAATGAATTTTAAAGGATGGGCAAACGAAATTATT
 TAGTTTTTAAAGGTGGGCAATCGAAATTATTTAGTTTTTAA**GAG**TGGGCAAACGAAAAAAC
 AAATTTTAAAGGGTGGGCAAACGAAATAATTTATTTTAAAGGGTGGGCAAACGTTAAAA
 TGAATTTTAAAGGATGGGCAAACGAAATTATTTAGTTTTTAAAGGTGGGCAATCGAAATTA
 TTTAGTTTTTAA**GAG**TGGGCAAACGAAAAAACAAATTTTAAAGGGTGGGCAAACGAAATAA
 TTTATTTTAAAGGGTGGGCAAACGAAATTATTTAGTTTTTAAAGGTGGGCAATCGAAATT
 ATTTAGTTTTTAA**GAG**TGGGCAAACGAAAAAACAAATTTTAAAGGGTGGGCAAACGAAATA
 ATTTATTTTAAAGGGTGGGCAAACGTTAAAAATGAATTTTAAAGGATGGGCAAACGAAAT
 TATTTAGTTTTTAAAGGTGGGCAATCGAAATTATTTAGTTTTTAA**GAG**TGGGCAAACGAAAA
 AACAAATTTTAAAGGGTGGGCAAACGAAATAATTTATTTTAAAGGGTGGGCAAACGTTA
 AAAATGAATTTTAAAGGATGGGCAAACGAAATTATTTAGTTTTTAAAGGTGGGCAATCGAA
 ATTATTTAGTTTTTAA**GAG**TGGGCAAACGAAAAAACAAATTTTAAAGGGTGGGCAAACGAA
 ATAATTTATTTTAAAGGGTGGGCAAACGTTAAAAATGAATTTTAAAGGATGGGCAAACGA
 AATTATTTAGTTTTTAAAGGTGGGCAATCGAAATTATTTAGTTTTTAA**GAG**TGGGCAAACGA
 AAAAAACAAATTTTAAAGGGTGGGCANACGAAATAATTTATTTTAAAGGGTGGGCAAACGT
 TAAAAATGAATTTTAAAGGATGGGCAAACGAAATTATTTAGTTTTTAAATGGTGGGCAAACG
 TTTTTTATTGGAAATTTTATTTTAA**GAG**GTGGGCTAACGAAATTGGGTGTATTTTAAAT
 GTGGGCAATCGAAAAAATTTTGTAAAGGTTTTTAAAGGGTGGGCAAACGATAGAAA

CTAATAATAATGGTTGGGCAAACGTTTTTTGTTGATAATATTTGGTTTTAAGAGATGGGCA
AACGATTTACTCTCTGATTAGACC **GAG**GTAACCTTAAAAGCCTTATATTTTCTAAAGTATA
AATTTTTTCAAATTCTAAAGGGTGGGCAAACGTGGGCAAACGATGTTATTGCGATTTAAA
AAAAAAAATTTTGAAAAAGCGAATTTTTTGATTTTCGAAAATTTTCGAACTTTTTCGAAA
ATTCATATCTCAAAAACCTGGACGTGGGCAAAAAAACTAATTGGTGGGCAAACATGGGCAA
AAAATGGGCAAACGGTCCAGCTTTCAAATTTCAAAAATTCGATTTTTCCGACCCCAGCTT
CTTTGAGCT **ACAAATGTGGGTATACAAATCACATAAATAGCAGGCTGCGCAATGATGAAA**
CATCAGAGATTCCCTAACACTTTTTGTTGAGACCTTTCGTTTGTGGATAGCTGGCCGGCTA
ATCTTCAATTCCTATTCATACTATCTATGTATGTTTGAAGTTCTGCCCTTGTGTGCGGT
GTTTGGCGTCACATTGCAACCGTTTGCTGAACCCTGCGAACCCGTGGAACGCGCATTC

Total: 2498 bp

1..64: *NOF Inverted Repeat*

65..2266: *FB 3' Inverted Repeat*

2268..2275: 9bp Duplicación directa

2276..2498: Secuencia genómica *downstream*

SECUENCIA NOF DE TEMPLETON Y POTTER, 1989

>gi|7962|emb|X15469.1|DMFBTRAN Drosophila melanogaster DNA
for complete foldback (FB) transposable element

TGTTTGCCCATCGCTAAAAC TAAATTAATTTTCGATTGCCCACCTTTTATTACTTTATTA
TTTCTTTTGCCCATCCTTTTATCTTTAGTATATATATATTCTATTGCCACCATATAAA
CACGTGCCACTTTCCTAGTTTTAGGATCTGCCTACATAACACGTGCAGACGCACAGGTG
TTTCTGGGTTTATATAGACCAAAAATTGGTTCCGATTGCCAATCTTGTAAATTTACAGTT
TACCAGGTAATTACATAATTTTCAAACCTCACTTTATGATAGGGTCCAATTTTTTACCT
GTGACAAAGTGTTAAATTTTTTAAGAATGGGTTTTTTCATGGCAGGTCAGAATCCTCTAT
AAAATCTAAAACACTTGTGCGGTATTTGAAAATCGCTCTCCTCCTTGATTCTCATATTAG
GTGTAAAAGATAAATCCGGAAC TCATAATTAATAATTTTTTTATGTGAAAAAGTTGTGC
GCGATTTTAACTACGCTTACCCAGTGCTGGAAAAGTTAAAGTTGTTTTGTTTTTCAAAG
AAAGTGAAAGTTGCTAAGCACGAAC TTAAGAAATCTGAGTGATTGTGTTAAATTTATTT
GAATCCTTGTGAATTTTGTGACAGTCTTTTTAAAGACTTGCAAAATTTTCATATTATT
CGGTTCTTGCTTTTATTTTTATAACAACGCGTTTTTCTTTAGGCATACCTTTATACATT
TACAGTGTAACAACAGTGTAACAACGTGTAATCAGTGCAAAATAGTTTTTTTTTATTTA
CTCCATAAAAAATAAGTGTTACTGTCAGGATGCCGGCCAAACCGCAAGTCGATGGTCAC
ACCTTAGTGGATGCATTTTGCTGCGCAATATTTTTACGGAGACTG GAGCTC TTAAGCC
AAGAAGCGATAAAGTTTGGATGGATATAAGCAACCAATTGAAAGGAGCGATCAGCGCGA
AGACGCTTAATTTCTACGCCAGAATCAATAGGAATAACATGATAACTGTGGTTAAAGAA
CGATGTG GAATTC AA CAGCTG GATACTAGTGCCAATTTAACTTTAAATAGCACATTTCC
TGATGATGACCCGGAGTTCAGATCACCGAAGCTT CAAAAATGGACCATTGCCTATTT
TGTACTTTAACTGGAGTTGGACCTGGAATTGTGGAGATCAATTGCCCCCAAAAAGGAT
CAAAAACTGAAAACTGCAACCTAACTGGACGGATACTATGGCAAAGTTGATATACAA
AAAAGTTCCTCTTCCGTGTGCATTTAATTTTAGAAAAGCTAAACTTTCCGACAAAGTGG
ATAATATTTGGCTACGAATTGAAGGCTATTGCAATGACTGCAGCTCAATTTTAAAGGGA
CATTGCCTTGTGAAACCCGATGAACAATGCGGCATAATGATATCTGTTTCAGTACCGGA
CACACGAGGTATACCTCATAATAAAAAACGACGGTGC ACTGGATCGAGACGACTTGAAA
TTGGGAACGAGTTGATTTTAAAAAAGCTGCATTGTGGAGGAAGGAAGCCACCGACAAC
ATGAATGATGACGACCCAGAACCGAGTTACATACCAAATTTACCAACCTTCGGAAACT
TCGTGAAGAGGCAACTAACAGACACCTAGGAATTACCAAGGATCG GGATCC AGTTTCAT
CATTATACCTTAAAAAGTATGAGGGTGAATTGGCTGGATGCATTCTTGACATTGGATTG
GATGAATTTTTCTGCATATACTGCACAGGAACCCAAGTAAAAACATATGCATCAAGGAT
AAAAACTATTAGAAAGATTTCTATTGACGCAACTGGAAGCGTGGTGT TACCCATCCAAA
AACCAAACGGTGACTCTAGTTATGTTTTTCTGTACCAAATTTGTAATGGAGGGTGACGAC
AGTATATTTCCAGTTTTTTCAGATGCTGTCGGCTAAACATGACACAGCCAGCATAACAGTT
TTGGTTAAGCAGATTTATATCAAAGTCGGGGCATTTTCCACTGGAGGTTGTATCTGATT
TTTCTTGGCATTGCTAAATGGAAT AAGCTTAAGCTT TAATGAGTGTAGGATTGCGACG
TATATAAAAAAATGTTTCCACAGCCTTTTGATGGAGGAACGGACGGATCTGCCACCCTG
CTATATTCGACTTGACATCGCCACCTAATTAATAATGATATGCCGGAAGAACGTCTTCA
AAAGTAAATTACCGAACCTCAAGGATTTTTATACTAGATGTATTGGTCTTGCAACAACG
TGTGAGACAAAGGACAGTTTTGCGGAATTAATTAATAATCAGTACTGATTGTGCGACTGAG
CCAATCCTCAGGGGAAGATGAAAAAGGAGACATTCTTCAAGTTACAGGAATGAAAAGT
ATCTGCTCGCCAGAATAGCTACATTTACTGCCCCGGATCACAAGGAGACCATTGAGGAC
AACTGCATACCAGAGGACCAGGAGGAAATTGACGAGGATGTTACGGACTTTATCTCTAA
TATTAATAATCGCTGCCGAAGAAGAAGCGTTAAATTTGCAATTCGGTCAACTGTGCGCCAA
ATCCGTATTTCTACCTGAGCTAATGCCACCATTAATTAAGTTGTGCAAAATTTTTGTT
TTATGGACAAACGTGATGAAGGAAAAGTTGTGTTCCAAATATGATGTGCGGCTCTTCGGC

TCTTGTGGAAGCCTATTTCAAGGATTTAAA**TTTTT**GGACATGAGCATATTCACCGACC
 AGTGAGAGCGGATAAATTCGTGGTGCAACATATCCGATGCATCGAAGCTGTTTGCTTGC
 TGGAAGCTGAGCCGCGATGAAACGCAAGACCGTTAAAACCTCCAGCTTTATAAAAAGAAAA
 CGCTCCTAAGAAAATGTGCAGTAAGGAAACCAAGGGATTTCTGGAGGAAATACTTGAAG
 AAAGCGAAGTGAATAACCTTTTACAAGAAGAAAACCTGGAAGGTGAAGAATAAAACAATA
 AAGCCCACGGAAGGAAATGATGCTGAAGACAACGACACTGATGATGAAAACAAGGAAAT
 GGATTTAAGTGAACAGCCCAAAGAAAAACCAAGGGGAAAATATCTCAAAAAATGCCCCA
 ATGTGGAGTTATTATACAATCGACCACATCGAAGGAAACAGGACGAAATTTTGCATAAT
 GGTGGATCAATGGGACCCGTCTGGATTGGCAAACAATTATTGCAATTCAAAAATACTTG
 TCCGTTTGACTCTCTAGTGGAAATATT**GTCGAC**CGCATACATAGACAATTTTTATTACA
 AAAGCCTATTGGATGATTTCTACACTGACAACCTTGACGATAGAATTGGTGAAAAAGTAT
 GCCGTGAGGGAGTTTCGTCCAGTCTCTACTGCGACAGAGGTCTGGTCTAAAAAGTTT
 TTTTGATGAAAAACACCAGATTATAAAATGCGACGCAAATATTGGGTCTTTTATTGAAA
 AAGCGCTGAATGGAGTACCCAGTGCCTCAAGTCATCGGACCCATATAAAAAACAACCAT
 GATTGCAGGAACCAAAAATATATCCACCATCGGCTGGAGGTTATAGATGTCGAAAAAGT
 TGGCCACCTCGACGTCCAGGAGGTAGTGATCCCCTTTATTGATGAGTTTTTTGCAAGAA
 CTGATGGAGAATGTAAAATATGCGGTGGACAACAGATCCTTGAAAGGCAGCCAGGACCG
 CATGTCATACTTGATATAGAATTTGCAATGGATGCTTTTCATCAAATTCATCATAACGG
 TTTACCAGGAACGACCCTTTACTTCAAGTGCCGGGAGGAAATTTTAATACAGGAAAAG
 AAATATATTTTAAGTGGTGCCATCGAATATGTTCTGCGATGGGAGGGGAAATTGGACA
 TTACATTGCATATTGCCGCAGAGTCATTGGATCTTGGGAAGTGCACAACGATATGTGCA
 GGCAATGGAAAAAGTTCTCAGCTCTAAATACCAAAATGACACTCCACATTTTGATATAC
 ACCCGGAAAAATTAATGTTTATTTTTTAAGCCTTGTTTAAAAGTGTA AAAAATATTTGTT
 GTTAAAAATTACAATCTTAAGTCCTTTGCAAACGTTGTTTAAAAATAAAATTAATTTAA
 TTATTTTACAAAACCTTAACCCTTTTTCACTTTTATACCTAATATAAAGAGGTCCGTA
 GTATCAAGGAGGAGAGCGATTTTCAAATACCGACAAGTGTTTTAGATTTTATAGAGGAT
 TCTGACCTGCCATGAAAAACCCATTCTTAAAAAATTTAACACTTTGTCCACAGGTA
 AAAAAATTGGACCTATCATAAAGTGAGGTTTAAAAATTATGTAATTACCTGGTAAACTGT
 AAAAAATTACAAGATTGGCAATCGGAACCAATTTTGGTCTATATAAACCCAGAAACACCTGT
 GCGTCTGCACGTGTTATGTAGGCAGATCCTAAAACCTAGGAAAGTGGCACGTGTTTATATGG
 TGGGCAATAGAATATAATTATTCTTAAGATTTT**AGGATGGG**CAAAGAAATAAATTTAGT
 TTTAAAAGGTGGGCAATCGAAAATTTTTTTTAGTTTTAA**GAG**TGGGCAAACGAAAAAACA
 AATTTTAAAGGGTGGGCAAACGAAATAATTTATTTTTTAAAGGGTGGGCAACGT

Colores de letra:

- - Turquesa ->FB IRs
- - Azul -> NOF IR 5'
- - Rojo ->NOF IR 3'

Resaltado (principales dianas de restricción)

- - **Amarillo** -> SalI GTCGAC
- - **Verde** -> PvuII CAGCTG
- - **Turquesa** -> EcoRI GAATTC
- - **Fucsia** -> HindIII AAGCTT
- - **Púrpura** -> BamHI GGATCC
- - **Rojo** -> SacI GAGCTC

SECUENCIA NOF HARDEN Y ASHBURNER, 1990

>gi|8297|emb|X51937.1|DMNOFFB Drosophila melanogaster NOF-FB transposable element

GAATTCAA CAGCTG GATACTAGTGCCAATTTAACTTTAAATAGCACATTTTCCTGATGAT
 GACCCGGAGTTCCAGATCACCGAAGCTTCAAAAAATGGACCATTGCCTATTTTGTACTT
 TAACCTGGAGTTGGACCTGGAATTGTGGAGATCAATTGCCCCAAAAAGGATCAAAAA
 CTGAAAACTGCAACCTAACTGGACGGATACTATGGCAAAGTTGATATACAAAAAGTT
 CCTCTTCGGTGTGCATTTAATTTTAGAAAAGCTAAACTTTCCGACAAAGTGGATAATAT
 TTGGCTACGAATTGAAGGCTATTGCAATGACTGCAGCTCAATTTTAAAGGGACATTGCC
 TTGTGAAACCCGATGAACAATGCGGCATAATGATATCTGTTTCAGTACCGGACACACGA
 GGTATACCTCATAATAAAAAACGACGGTGCCTGGATCGAGACGACTTGAAATTGGGAA
 CGAGTTGATTTTAAAAAAGCTGCATTGTGGAGGAAGGAAGCCACCGACAACATGAATG
 ATGACGACCCAGAACCGAGTTACATAACCAATTTACCAACCCTTCGAAACTTCGTGAA
 GAGGCAACTAACAGACACCTAGGAATTACCAAGGATCGGGATCCAGTTTCATCATTATA
 CCTTAAAAAGTATGAGGGTGAATTGGCTGGATGCATTCTTGACATTGGATTGGATGAAT
 TTTTCTGCATATACTGCACAGGAACCCAAGTAAAAACATATGCATCAAGGATAAAAACT
 ATTAGAAAGATTTCTATTGACGCAACTGGAAGCGTGGTGTACCCATCCAAAAACCAA
 CGGTGACTCTAGTTATGTTTTTCTGTACCAAATTGTAATGGAGGGTGACGACAGTATAT
 TTCCAGTTTTTTCAGATGCTGTGCGCTAAACATGACACAGCCAGCATAACAGTTTTGGTTA
 AGCAGATTTATATCAAAGTCGGGGCATTTCCTACTGGAGGTTGTATCTGATTTTTCTT
 GGCATTGCTAAATGGAAT AAGCTTAAGCTT TAATGAGTGTAGGATTGCGACGTATATAA
 AAAAATGTTTCCACAGCCTTTTGATGGAGGAACGGACGGATCTGCCACCCTGCTATATT
 CGACTTGACATCGCCACCTAATTAATGATATGCCGGAAGAAGCTCTTCAAAGTAA
 ATTACCGAACCTCAAGGATTTTTATACTAGATGTATTGGTCTTGCAACAACGTGTGAGA
 CAAAGGACAGTTTTGCGGAATTAATTAATCAGTACTGATTGTGCGACTGAGCCAATCC
 TCAGGGGAAGATGAAAAAGGAGACATTCTTCAAGTTACAGGAATGAAAAGTATCTGCT
 CGCCAGAATAGCTACATTTACTGCCCGGATCACAAGGAGACCATTGAGGACAACCTGCA
 TACCAGAGGACCAGGAGGAAATTGACGAGGATGTTACGGACTTTATCTCTAATATTA
 ATCGCTGCCGAAGAAGAAGCGTTAAATTGCAATTCGGTCAACTGTGCGCCAAATCCGTA
 TTTCTACCTGAGCTAATGCCACCATTAATTAAGTTGTGCAAATATTTTGTTTTATGGA
 CAAACGTGATGAAGGAAAAGTTCTGTTCCAAATATGATGTCGGCTCTTCGGCTCTTGTG
 GAAGCCTATTTCAAGGATTTAAA AAACAC GGACATGAGCATATTCCACCGACCAGTGAG
 AGCGGATAAATTCGTGGTGAACATATCCGATGCATCGAAGCTGTTTGCAAGCTGGAAC
 GAGCCGCGATGAAACGCAAGACCGTTAAACTCCCAGCTTTATAAAAGAAAACGCTCCT
 AAGAAAATGTGCAGTAAGGAAACCAAGGGATTTCTGGAGGAAATACTTGAAGAAAGCGA
 AGTGAATACCTTTTACAAGAAGAAAACCTGGAAGGTGAAGAATAAAACAATAAAGCCCA
 CGGAAGGAAATGATGCTGAAGACAACGACACTGATGATGAAAACAAGGAAATGGATTTA
 AGTGAACAGCCCAAAGAAAAACCAAGGGGAAAATATCTCAAAAAATGCCCAATGTGGA
 GTTATTATACAATCGACCACATCGAAGGAAACAGGACGAAATTTTGCATAATGGTGGAT
 CAATGGGACCCGTCTGGATTGGCAAACAATTATTGCAATTCAAAAATACTTGTCCGTTT
 GACTCTCTAGTGGAATATT GTCGAC CGCATAACATAGACAATTTTTATTACAAAAGCCT
 ATTGGATGATTTCTACACTGACAACCTTGACGATAGAATTGGTGAAAAAGTATGCCGTCG
 AGGGAGTTTTCGTCCAGTCTCTACTGCGACAGAGGTCTGGTCCTAAAAAGTTTTTTTGT
 GAAAAACACCAGATTATAAAATGCGACGCAAATATTGGGTCTTTTATTGAAAAGCGCT
 GAATGGAGTACCCAGTGCCTCAAGTCATCGGACCCATATAAAAAACAACCATGATTGCA
 GGAACCAAAAATATATCCACCATCGGCTGGAGGTTATAGATGTCGAAAAAGTTGGCCAC
 CTCGACGTCCAGGAGGTAGTGATCCCTTTTATTGATGAGTTTTTTGCAAGAACTGATGG
 AGAATGTAAAATATGCGGTGGACAACAGATCCTTGAAAGGCAGCCAGGACCGCATGTCA

TACTTGATATAGAATTTGCAATGGATGCTTTTCATCAAATTCATCATAACGGTTTACCA
 GGAACGACCACCTTTACTTCAAGTGCCGGAGGAAATTTTAATACAGGAAAAGAAATATAT
 TTTAAGTGGTGCCATCGAATATGTTCTGCGATGGGAGGGGAAATTGGACATTACATTG
 CATATTGCCGCAGAGTCATTGGATCTTGGGAAGTGCACAACGATATGTGCAGGCAATGG
 AAAAAGTTCTCAGCTCTAAATACCAAATGACACTCCACATTTTGATATACACCCGGAA
 AAATTAATGTTTATTTTAAAGCCTTGTTTAAAAGTGTAATAAATATTTGTTGTTAAAA
 TTACAATCTTAAGTCCTTTGCAAACGTTGTTTAAAATAAAATTAATTAATTTTAA
 CAAAACCTTAACCCTTTTTCACCTTTTATACTAATATAAAGAGGTCCGTAAAGTATCAAG
 GAGGAGAGCGATTTTCAAATACCGACAAGTGTTTTAGATTTTATAGAGGATTCTGACCT
 GCCATGAAAAACCCATTCTTAAAAAATTTAACACTTTGTCCAGGTAATAAATTGGACC
 CTATCATAAAGTGAGGTTTGAAAATTATGTAATTACCTGGTAAACTGTAAATTACAAGA
 TTGGCAATCGGAACCAATTTTGGTCTATATAAACCCAGAAACACCTGTGCGTCTGCAC
 GTGTTATGTAGGCAGATCCTAAACTAGGAAAGTGGCACGTGTTTATATGGTGGGCAAT
 AGAATTTATTTATTCTTAAGATTTTAGGATGGGCAAAAGAAATAATTTAGTTTTAAAA
 GGTGGGCAATCGAACTTTTCGAAATTATTTAGTTTTAAGAGTGGGCAAACGAAAAAACA
 AATTTTAAAGGGTGGGCAAACCAATAATTTATTTTAAAGGGTGGGCAAACCTTAAAA
 ATGAATTTTAAAGGATGGGCAAACGAAATTATTTAGTTTTAAAGGTGGGCAA

Colores de letra:

- - Turquesa ->FB IRs
- - Rojo ->NOF IR 3'
-

Resultado (principales dianas de restricción)

- - Amarillo -> SalI GTCGAC
- - Verde -> PvuII CAGCTG
- - Turquesa -> EcoRI GAATTC
- - Fucsia -> HindIII AAGCTT
- - Púrpura -> BamHI GGATCC

NOF ORF-0

Resultados GenScan

GENSCAN 1.0 Date run: 22-Feb-107 Time: 04:51:27
 Sequence NOF: 4383 bp : 37.99% C+G : Isochore 1 (0 - 43 C+G%)
 Parameter matrix: HumanIso.smat

Predicted genes/exons:

Gn.Ex	Type	S	.Begin	...End	.Len	Fr	Ph	I/Ac	Do/T	CodRg	P....	Tscr...
1.01	Init	+	721	2640	1920	0	0	84	44	1300	0.889	112.74
1.02	Term	+	2768	3925	1158	1	0	-29	42	644	0.824	38.98
1.03	PlyA	+	3981	3986	6							1.05

Explanation

Gn.Ex : gene number, exon number (for reference)
Type : Init = Initial exon (ATG to 5' splice site)
 Intr = Internal exon (3' splice site to 5' splice site)
 Term = Terminal exon (3' splice site to stop codon)
 Sngl = Single-exon gene (ATG to stop)
 Prom = Promoter (TATA box / initiation site)
 PlyA = poly-A signal (consensus: AATAAA)
S : DNA strand (+ = input strand; - = opposite strand)
Begin : beginning of exon or signal (numbered on input strand)
End : end point of exon or signal (numbered on input strand)
Len : length of exon or signal (bp)
Fr : reading frame (a forward strand codon ending at x has frame x mod 3)
Ph : net phase of exon (exon length modulo 3)
I/Ac : initiation signal or 3' splice site score (tenth bit units)
Do/T : 5' splice site or termination signal score (tenth bit units)
CodRg : coding region score (tenth bit units)
P : probability of exon (sum over all parses containing exon)
Tscr : exon score (depends on length, I/Ac, Do/T and CodRg scores)

ATGCCGGCCAAACCGCAAGTCGATGGTCACACCTTAGTGGATGCATTTTGCTGCGGAATATTTTACGGAGA
CTGGAGCTCTTAAGCCAAGAAGCGATAAAGTTTGGATGGATATAAGCAACCAATTGAAAGGAGCGATCAGCGC
GAAGACGCTTAATTTCTACGCCAGAATCAATAGGAATAACATGATAACTGTGGTTAAAGAACGATGTGGAATT
CAACAGCTGGATACTAGTGCCAATTTAACTTTAAATAGCACATTTTCTGATGATGACCCGGAGTTCCAGATCA
CCGAAGCTTCAAAAAATGGACCATTGCCTATTTTGTACTTTAACCTGGAGTTGGACCTGGAATTGTGGAGATC
AATTGCCCCCAAAAAGGATCAAAAACTGAAAACTGCAACCTAACTGGACGGATACTATGGCAAAGTTGATA
TACAAAAAAGTTCTCTTCCGTGTGCATTTAATTTTAGAAAAGCTAAACTTTCCGACAAAAGTGGATAATATTT
GGCTACGAATTGAAGGCTATTGCAATGACTGCAGCTCAATTTTAAAGGGACATTGCCTTGTGAAACCCGATGA
ACAATGCGGCATAATGATATCTGTTTCAGTACCGGACACACGAGGTATACCTCATAATAAAAAACGACGGTGC
ACTGGATCGAGACGACTTGAAATTGGGAACGAGTTGATTTTAAAAAAGCTGCATTGTGGAGGAAGGAAGCCA
CCGACAACATGAATGATGACGACCCAGAACCGAGTTACATACCAAATTTACCAACCCCTTCGGAAACTTCGTGA
AGAGGCAACTAACAGACACCTAGGAATTACCAAGGATCGGGATCCAGTTTCATCATTATACCTTAAAAAGTAT
GAGGGTGAATTGGCTGGATGCATTCTTGACATTGGATTGGATGAATTTTCTGCATATACTGCACAGGAACCC
AAGTAAAAACATATGCATCAAGGATAAAAACTATTAGAAAGATTTCTATTGACGCAACTGGAAGCGTGGTGT
ACCCATCCAAAAACCAACGGTGACTCTAGTTATGTTTTTCTGTACCAAATTTGTAATGGAGGGTGCACGACGT
ATATTTCCAGTTTTTTCAGATGCTGTGGCTAAACATGACACAGCCAGCATAACAGTTTGGTTAAGCAGATTTA
TATCAAAGTCGGGCATTTTCCACTGGAGGTTGTATCTGATTTTTCTTGGCATTGCTAAATGGAATAAGCTT
AAGCTTTAATGAGTGTAGGATTGCGACGTATATAAAAAAATGTTTCCACAGCCTTTTGATGGAGGAACGGACG
GATCTGCCACCCTGCTATATTCGACTTGACATCGCCACCTAATTAATGATATGCCGGAAGAACGTCTTCA

AAAGTAAATTACCGAACCTCAAGGATTTTTTATACTAGATGTATTGGTCTTGCAACAACGTGTGAGACAAAGGA
 CAGTTTTGCGGAATTAATTAATCAGTACTGATTGTGCGACTGAGCCAATCCTCAGGGGAAGATGAAAAAGGA
 GACATTCTTTCAAGTTACAGGAATGAAAAGTATCTGCTCGCCAGAATAGCTACATTTACTGCCCCGGATCACA
 AGGAGACCATTGAGGACAACATGCATACCAGAGGACCAGGAGGAAATGACGAGGATGTTACGGACTTTATCTC
 TAATATTAATAATCGCTGCCGAAAGAAGCGTTAAATTTGCAATTCGGTCAACTGTCGGCCAAATCCGTATTTT
 CTACCTGAGCTAATGCCACCATTAATTAAGTTGTGCAAATATTTTGTTTTATGGACAAACGTGATGAAGGAAA
 AGTTGTGTTCCAAATATGATGTCGGCTCTTCGGCTCTTGTGGAAGCCTATTTCAAGGATTTAAaAaacacGGA
 CATGAGCATATTCACCGACCAAGTGTGAGAGCGGATAAAATTCGTGGTGCAACATATCCGATGCATCGAAGCTGTT
 TGCTTGCTGGAACGTGAGCCGCGATGAAACGCAAGACCGTTAAAACTCCCAGCTTTATAAAAAGAAAACGCTCCT
 AAGAAAATGTGCAGTAAGGAAACCAAGGGATTTCTGGAGGAAAATACCTTGAAGAAAGCGAAGTGAATACCTTT
 TACAAGAAGAAAACCTGGAAGGTGAAGAATAAAACAATAAAGCCACGGAAGGAAATGATGCTGAAGACAACGA
 CACTGATGATGAAAAACAAGGAAATGGATTTAAGTGAACAGCCCAAAGAAAACCAAGGGGAAAATATCTCAA
 AAATGCCCCAATGTGGAGTTATTATACAATCGACCACATCGAAGGAAACAGGACGAAATTTTGATAATGGTG
 GATCAATGGGACCCGCTCTGGATTGGCAAACAATTTATGCAATTCAAAAATACTTGTCCGTTTACTCTCTAGT
 GGAAATATTGTCGACCGCATAACATAGACAATTTTTATTACAAAAGCCTATTGGATGATTTCTACACTGACAAC
 TTGACGATAGAATTTGGTGAAAAAGTATGCCGTCGAGGGAGTTTCGTCCAGTCTCTACTGCGACAGAGGTCTGG
 TCCTAAAAAGTTTTTTTATGATGAAAAACACCAGATTATAAAATGCGACGCAAATATTGGGTCTTTTATTGAAA
 AGCGCTGAATGGAGTACCCAGTGCCTCAAGTCATCGGACCCATATAAAAAACAACCATGATTGCAGGAACCAA
 AAATATATCCACCATCGGCTGGAGGTTATAGATGTCGAAAAAGTTGGCCACCTCGACGTCCAGGAGGTAGTGA
 TCCCCTTTTATTGATGAGTTTTTTTGAAGAACTGATGGAGAATGTAAAATATGCGGTGGACAACAGATCCTTGA
 AAGGCAGCCAGGACCGCATGTCATACTTGATATAGAATTTGCAATGGATGCTTTTCATCAAATTCATCATAAC
 GGTTTACCAGGAACGACCACTTTACTTCAAGTGCCGGGAGGAAATTTAATACAGGAAAAGAAATATATTTTA
 AGTGGTGCCATCGAATATGTTCTGCGATGGGAGGGGAAATTTGGACATTACATTGCATATTGCCGAGAGTCA
 TTGGATCTTGGGAAGTGCACAACGATATGTGCAGGCAATGAAAAAGTTCTCAGCTCTAAATACAAAATGAC
 ACTCCACATTTTGATATACACCCGAAAAATTAATGTTTATTTTTAAGCCTTGTTTAAAAGTGTAA

Predicted peptide sequence(s) :

>NOF|GENSCAN_predicted_peptide_1|1025_aa
 MPAKPVVDGHTLVDAFCCANIFTEGALKPRSDKVVMDISNQLKGAI SAKTLNIFYARINR
 NNMITVVKERCGIQQLDTSANLTLNSTFPDDDDPEFQITEASKNGPLPILYFNLELDLELW
 RSIAPKKDQKTEKLPNWDTMAKLIYKVVPLPCAFNFRKAKLSDKVDNIWLRIEGYCND
 CSSILKGHCLVKPDEQCGIMISVSVDPDRGIPHNKKRRCCTGSRRLIEGNEILKKAALWR
 KEATDNMNDDDPEPSYIPNLPLRKLREEATNRHLGITKDRDPVSSLYLKKYEGELAGCI
 LDIGLDEFFCIYCTGTQVKTYASRIKTIKISIDATGSVVLPIQKPNGDSSYVFLYQIVM
 EGDSSI FPFVQMLSAKHDTASIQFWLSRFISKSGHFPLEVVSDFSLALLNGISLSFNCR
 IATYIKKCFHSLMEERTDLPPCYIRLDIAHLIKMICRKNVFKSKLPNLKDFYTRCIGLA
 TTCETKDSFAELIKSVLIVALSQSSGEDEKGDILSSYRNEKYLLARIATFTAPDHKETIE
 DNCIPEDQEEIDEDVTDIFISNIKIAAEAAALNCNSVNCRPNPYFLPELMPPLIKLCKYFV
 LWTNVMKEKLCISKYDVGSSALVEAYFKDLKNTDMSIFHRPKMCSKETKGFLEEILEESEV
 EYLLQEENWKVKNKTIKPTEGNDAEDNDTDDENKEMDLSEQPKPRGKYLKKCPNVELL
 YNRPHRRKQDEILHNGGSMGPVWIGKQLLQFKNTCPFDLVEILSTAYIDNFYKSLDD
 FYTDNLTIELVKKYAVEGVSSSLYCDRGLVLKSFDFDEKHQI IKCDANIGSFIEKALGNVP
 SASSHRTHIKNNHDCRNQKYIHRLEVIDVEKVGHLDVQEVVIFIDEFFARTDGECKIC
 GGQQILERQPGPHVILDIEFAMDAFHQIHHNGLPGTTLLQVPGGNFNTGKEIYFKWCHR
 ICSCDGRGNWTLHCILPQSHWILGSAQRVQAMEKVLSSKYQNDTPHFDIHPEKLMFIFK
 PCLKV*

Resultados ProtLoc

Anchored => 5.70753098987889
Intracellular => 6.08770294179993
Extracellular => 6.51277820857436
Membrane => 8.09030603173045
Nuclear => 8.3970961994759

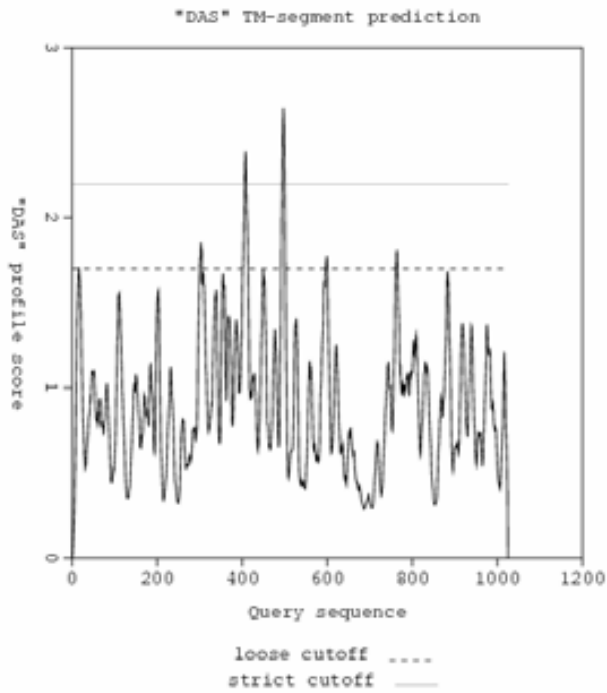
Resultados TransMem

TM : *EVVSDFLALLNGI* | 398 => 412

TM : *LIKSVLIVA* | 491 => 500

TM : *SLVEILSTA* | 758 => 767

Resultados DAS Transmembrane predictor



Resultados SOSUI

Total length : 1025 A. A.

Average of hydrophobicity : -0.380878

This amino acid sequence is of a SOLUBLE PROTEIN.

Hits for all PROSITE (release 20.6) motifs on sequence ORF-0 :

no hit!

NOF ORF-1

Resultados ORF Finder

ATGCCGGCCAAAACCGCAAGTCGATGGTCCACACCTTAGTGGATGCATTTTGTGCGCGAATATTTTTACGGAGA
 CTGGAGCTCTTAAGCCAAGAAGCGATAAAAGTTTGGATGGATATAAGCAACCAATTGAAAGGAGCGATCAGCGC
 GAAGACGCTTAATTTCTACGCCAGAATCAATAGGAATAACATGATAACTGTGGTTAAAGAACGATGTGGAATT
 CAACAGCTGGATACTAGTGCCAATTTAACTTTAAATAGCACATTTCTGATGATGACCCGGAGTTCAGATCA
 CCGAAGCTTCAAAAAATGGACCATTGCCTATTTGTACTTTAACCTGGAGTTGGACCTGGAATTTGGGAGATC
 AATTGCCCCCAAAAAGGATCAAAAAACTGAAAAACTGCAACCTAACTGGACGGATACATATGGCAAAGTTGATA
 TACAAAAAAGTTCCCTCTTCCGTGTGCATTTAATTTTAGAAAAGCTAAACTTTCCGACAAAAGTGGATAATATTT
 GGCTACGAATTGAAAGGCTATTGCAATGACTGCAGCTCAATTTTAAAGGGACATTGCCTTGTGAAACCCGATGA
 ACAATGCGGCATAATGATATCTGTTTCAGTACCGGACACACGAGGTATACCTCATAATAAAAAACGACGGTGC
 ACTGGATCGAGACGACTTGAAATTTGGGAACGAGTTGATTTTAAAAAAGCTGCATTTGTGGAGGAAGGAAGCCA
 CCGACAACATGAATGATGACGACCCAGAACCGAGTTACATACCAAATTTACCAACCTTCGGAAACTTCGTGA
 AGAGGCAACTAACAGACACCTAGGAATTACCAAGGATCGGGATCCAGTTTCATCATTATACCTTAAAAAGTAT
 GAGGGTGAATTGGCTGGATGCATTTCTTGACATTTGGATTTGGATGAATTTTTCTGCATATACTGCACAGGAACCC
 AAGTAAAAACATATGCATCAAGGATAAAAACTATTAGAAAAGATTTCTATTGACGCAACTGGAAGCGTGGTGT
 ACCCATCCAAAACCAAACGGTACTCTAGTTATGTTTTTCTGTACCAAATTTGTAATGGAGGGTGACGACAGT
 ATATTTCCAGTTTTTTCAGATGCTGTCGGCTAAACATGACACAGCCAGCATACAGTTTTTGGTTAAGCAGATTTA
 TATCAAAGTCGGGGCATTTTTCCACTGGAGGTTGTATCTGATTTTTTCCCTTGGCATTTGCTAAATGGAATAAGCTT
 AAGCTTTAATGAGTGTAGGATTTGCGACGTATATAAAAAAATGTTTTCCACAGCCTTTTGATGGAGGAACGGACG
 GATCTGCCACCCTGCTATATTCGACTTGACATCGCCACCTAATTTAAATGATATGCCGGAAGAACGTCTTCA
 AAAGTAAATTACCGAACCTCAAGGATTTTTTATACTAGATGTATTGGTCTTGCAACAACGTGTGAGACAAAGGA
 CAGTTTTGCGGAATTAATTAATCAGTACTGATTTGTCGCACAGCAATCCTCAGGGGAAGATGAAAAAGGA
 GACATTTCTTCAAGTTACAGGAATGAAAAGTATCTGCTCGCCAGAATAGCTACATTTACTGCCCCGGATCACA
 AGGAGACCTTTGAGGACAACCTGCATACCAGAGGACCCAGGAGGAAATTTGACGAGGATTTTACGGACTTTATCTC
 TAATATTAATAATCGCTGCCGAAGAAGGAGCGTTAAATTTGCAATTCGGTCAACTGTGCGCCAAATCCGTATTTT
 CTACCTGAGCTAATGCCACCATTAATTAAGTTGTGCAATATTTTTGTTTTATGGACAAACGTGATGAAGGAAA
 AGTTGTGTTCCAAAATATGATGTCGGCTCTTCGGCTCTTGTGGAAGCCTATTTCAAGGATTTAAaAaacGGA
 CATGAGCATATTCACCGACCAGTGAGAGCGGATAAAATTCGTGGTGAACATATCCGATGCATCGAAGCTGTT
 TGCTTGCTGGAACCTGAGCCGCGATGAAACGCAAGACCGTTAA

>NOF_ORF1

MPAKPQVDGHITLVDAFCCANIFTEETGALKPRSDKVWMDISNQLKGAISAKTLNIFYARINR
 NNMITVVKERCGIQQLDTSANLTLNSTFPDDDEPQITEASKNGPLPIILYFNLELDLELW
 RSIAPKKDQKTEKLQPNWDTMAKLIYKKVPLPCAFNFRKAKLSDKVDNIWLRIEGYCND
 CSSILKGHCLVKPDEQCGIMISVSVDPDRGIPHNNKRRCTGSRRLIEIGNELILKKAALWR
 KEATDNMDDDEPSYIPNLPTLRKLREEATNRHLGITKDRDPVSSLYLKKYEGELAGCI
 LDIGLDEFFCIYCTGTQVKTYASRIKTIKISIDATGSVVLPIQKPNGDSSYVFLYQIVM
 EGDDSIFFVFMQLSAKHDTASIQFWLSRFISKSGHFPLEVVSDFSLALLNGISLSFNECR
 IATYIKKCFHSLMEERTDLPPCYIRLDIAHLIKMICRKNVFKSKLPNLKDFYTRCIGLA
 TTCETKDSFAELIKSVLIVALSQSSGEDEKGDILSSYRNEKYLARIATFTAPDHKETIE
 DNCIPEDQEEIDEDVTDIFISNIKIAAEEELNCNSVNCRPNPYFLPELMPPLIKLCKYFV
 LWTNVMKEKLCISKYDVGSSALVEAYFKDLKNTDMSIFHRPVRADKFVVQHIREAVCLL
 ELSRDETQDR*

Resultados ProtLoc

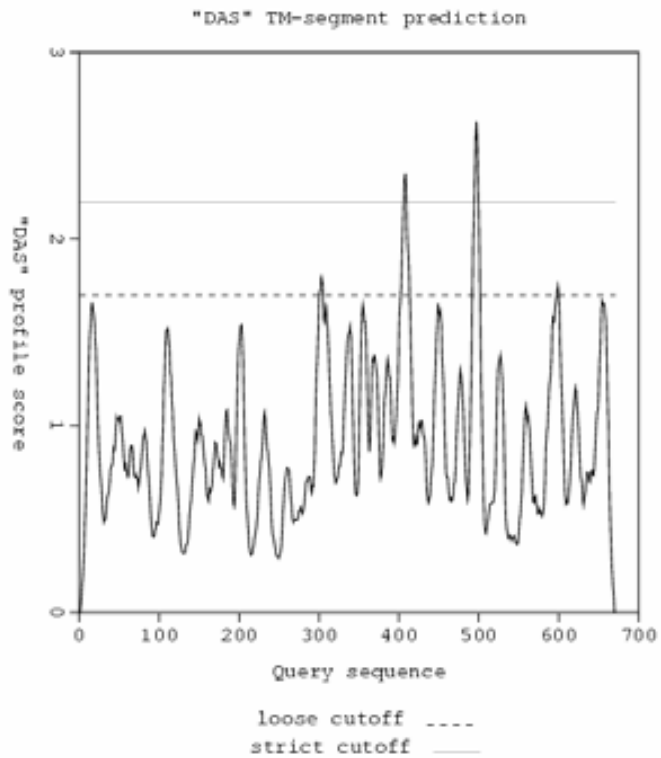
Anchored => 6.99728008732301
Extracellular => 7.79062709992656
Intracellular => 7.87563313870295
Membrane => 8.73210303489131
Nuclear => 9.54870086489336

Resultados TransMem

TM : *EVVSDFLALLNGI* | 398 => 412

TM : *LIKSVLIVA* | 491 => 500

Resultados DAS transmembrane prediction server



Resultados SOSUI

Query title : None

Total length : 670 A. A.

Average of hydrophobicity : -0.272835

This amino acid sequence is of a SOLUBLE PROTEIN.

Hits for all PROSITE (release 20.6) motifs on sequence ORF-1 :

no hit!

NOF ORF-2**Resultados ORFinder**

ATGAAACGCAAGACCGT**TAAA**AACTCCCAGCTTTATAAAAAGAAAACGCTCCTAAGAAAATGT
GCAGTAAGGAAACCAAGGGATTTCTGGAGGAAATACTTGAAGAAAGCGAAGTGGAATACCT
TTTACAAGAAGAAAACCTGGAAGGTGAAGAATAAAACAATAAAGCCCACGGAAGGAAATGAT
GCTGAAGACAACGACACTGATGATGAAAACAAGGAAATGGATTTAAGTGAACAGCCCAAAG
AAAAACCAAGGGGAAAATATCTCAAAAAATGCCCAATGTGGAGTTATTATACAATCGACC
ACATCGAAGGAAACAGGACGAAATTTTGCATAATGGTGGATCAATGGGACCCGTCTGGATT
GGCAAACAATTATTGCAATTCAAAAATACTTGTCCGTTTGACTCTCTAGTGGAAATATTGT
CGACCGCATACATAGACAATTTTTTATTACAAAAGCCTATTGGATGATTTCTACACTGACAA
CTTGACGATAGAATTGGTGA AAAAGTATGCCGTCGAGGGAGTTTCGTCCAGTCTCTACTGC
GACAGAGGTCTGGTCC**TAAAA**AGTTTTTTTTGATGAAAAACACCAGATTATAAAATGCGACG
CAAATATTGGGTCTTTTTATTGAAAAAGCGCTGAATGGAGTACCCAGTGCCTCAAGTCATCG
GACCCATATAAAAAACAACCATGATTGCAGGAACCAAAAATATATCCACCATCGGCTGGAG
GTTATAGATGTCGAAAAAGTTGGCCACCTCGACGTCCAGGAGGTAGTGATCCCCTTTATTG
ATGAGTTTTTTGCAAGAAGTATGGAGAATGTAATAATATGCGGTGGACAACAGATCCTTGA
AAGGCAGCCAGGACCGCATGTCATACTTGATATAGAATTTGCAATGGATGCTTTTCATCAA
ATTCATCATAACGGTTTACCAGGAACGACCACTTTACTTCAAGTGCCGGGAGGAAATTTTA
ATACAGGAAAAGAAATATATTTTAAGTGGTGCCATCGAATATGTTCCCTGCGATGGGAGGGG
AAATTGGACATTACATTGCATATTGCCGCAGAGTCATTGGATCTTGGGAAGTGCACAACGA
TATGTGCAGGCAATGGA AAAAGTTCTCAGCTCTAAATACCAAAATGACACTCCACATTTTG
ATATACACCCGGAAAAATTAATGTTTATTTTTAAGCCTTGTTTAAAAGT**TAA**

>NOF_ORF2

MKRKTVKTPSFIKENAPKKMCSKETKGFLEEILEESEVEYLLQEENWKVKNKTIKPTEGN
DAEDNDTDDENKEMDLSEQPKEKPRGKYLKCPNVELLYNRPHRRKQDEILHNGGSMGPV
WIGKQLLQFKNTCPFDLVEILSTAYIDNFYYKSLDDFYTDNLTIELVKKYAVEGVSS
LYCDRGLVLKSFDEKHQIIKCDANIGSFIEKALNGVPSASSHRTHIKNNHDCRNQKYIH
HRLEVIDVEKVGHLDVQEVVIFPIDEFFARTDGECKICGGQQILERPQPHVILDIEFAM
DAFHQIHHNGLPGTTTLLQVPGGNFNTGKEIYFKWCHRICSDGRGNWTLHCILPQSHWI
LGSAQRYVQAMEKVLSSKYQNDTPHFDIHPEKLMFIFKPCLKV*

Resultados ProtLoc

Intracellular => 8.76403568034248

Anchored => 9.54436905501773

Extracellular => 10.7493593002534

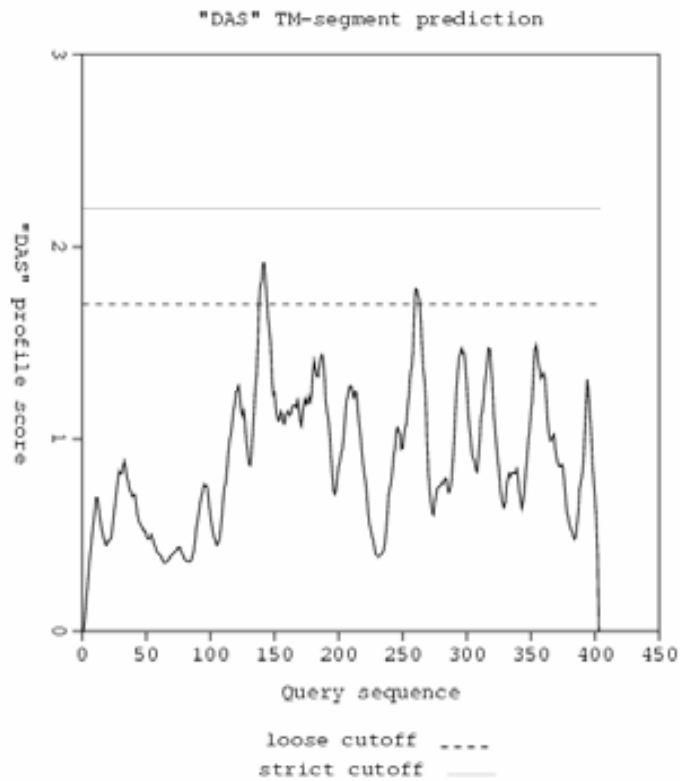
Nuclear => 12.4693550935535

Membrane => 13.1065868066115

Resultados TransMem

TM : *SLVEILSTA* | 136 => 145

Resultados DAS transmembrane prediction server



Resultados SOSUI

Query title: None

Total length: 403 A. A.

Average of hydrophobicity: -0.582630

This amino acid sequence is of a SOLUBLE PROTEIN.

Hits for all PROSITE (release 20.6) motifs on sequence ORF-2 :

no hit!

NOF ORF-3**Resultados ORFinder**

ATGTATGCGGTCGACAATATTTCCACTAGAGAGTCAAACGGACAAGTATTTTTGAATTGCAATAATTGTT
 TGCCAATCCAGACGGGTCCCATTTGATCCACCATTATGCAAAATTTTCGTCCGTGTTTCCCTTCGATGTGGTGC
 ATTGTATAATAACTCCACATTTGGGGCATTTTTGAGATATTTTCCCCTTGGTTTTTCTTTGGGCTGTTCA
 CTTAAATCCATTTCCCTTGTTTTTCATCATCAGTGTCTTTCAGCATCATTTCCCTCCGTGGGCTTTA
 TTGTTTTATTCTTCACCTTCCAGTTTTCTTCTTGTAAGGTATTTCCACTTCGCTTTCTTCAAGTATTTT
 CTCCAGAAATCCCTTGGTTTCCCTTACTGCACATTTTCTTAGGAGCGTTTTTCTTTTATAAAGCTGGGAGTT
 TTAACGGTCTTGGTTCATCGCGGCTCAGTTCCAGCAAGCAAACAGCTTCGATGCATCGGATATGTTGC
 ACCACGAATTTATCCGCTCTCACTGGTCGGTGGAATATGCTCATGTCCGTGTTTTTT**TAA**

>NOF_ORF_3

MYAVDNI STRESNGQVFLNCNNCLPIQTGPIDPPLCKISSCFLRGRLYNNSTLGHFLRYFPLGFSLGCS
 LKSI SLFSSSVSLSSASFPSVGFIVLFFTFQFSSCKRYSTSLSSISSRNPLVSLLLHIFLGAFSFIKLG
 LTVLRFIAAQFQQANSFDASDMLHHEFIRSHWSVEYAHVRVF*

Resultados ProtLoc

Membrane => 35.7967065703774

Nuclear => 38.4716796664314

Anchored => 38.7163065563758

Extracellular => 43.4930395121461

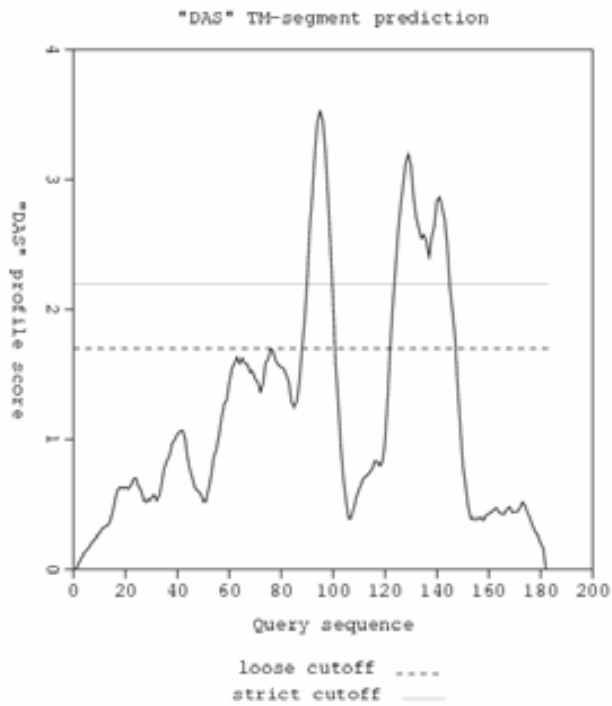
Intracellular => 46.1784341075808

Resultados TransMem

TM : GFSLGCSLKSI SLFSSSVSLSSASFPSVGFIVLFFTF | **63** => **100**

TM : LVSLLLHIFLGAFSFIKLG VLVLRV | **121** => **146**

Resultados DAS transmembrane prediction server



Resultados SOSUI

Query title : None

Total length : 182 A. A.

Average of hydrophobicity : 0.366484

This amino acid sequence is of a SOLUBLE PROTEIN.

Hits for all PROSITE (release 20.6) motifs on sequence ORF-3 :

no hit!

```

#!/usr/bin/perl
use LWP::Simple;
no warnings 'uninitialized';
print "*****\n";
print "*          * \n";
print "*          * \n";
print "*          * \n";
print "*          * \n";
print "*          * \n";
print "*          * \n";
print "*Badal,M.Portela,A.Xamena,N.Cabre,O.*\n";
print "*****\n";
sleep 1;

@cytos = ('X','Xh','2L','2Lh','2R','2Rh','3L','3Lh','3R','3Rh','4','Yh','U');
foreach $cromo (@cytos) {

$numFB=0; $sum=0; $seqid=""; $seq=""; $control=""; $seqline=""; $llargada=0; @FBtotal=(); $difer=0;
open FILE, "Sequence$cromo.html" or die "Can't open file\n" ;
while (<FILE>) {
    chomp;
    if (/^.<div class="header">>(.)<.div>$/) { #save ID numbers in a variable $seqid
        $seqid=$1;
        print "$seqid\n";
    }
    if (/^.<div class="sequence">(\w+)<.div>$/g) { #read each line of sequence in the file...
        $seqline=$1;
        $seq .= " ".$seqline; #...and append it into $seqline
    }
}
close FILE;
$seq =~ s/(\W|\d)//g;
while ($seq =~ /(.{9}AGCTCAAAGAAGCTGGGGTCGGAAAAA.+?TTTTTCCGACCCCAGCTTCTTTGAGCT.{9})/g) {
    push @FBtotal, $1; #find any FB and save each in an array
    $numFB ++; $numFBtotal ++; #count the FB in this particular sequence
}
open OUT, ">FBout$cromo.txt"; #Copy everything in a text file 'FBout.txt'
open CLUSTALW, ">FBclust$cromo.txt";

```

```

print OUT "Location Chromosome $cromo\n";
print OUT "Total FB in sequence: $numFB\n";

# Should any sequence have more than one FB, each would be in different fields in the @FBtotal array
# thus, we retrieve the FB ordered and numbered thanks to the 'conditional' variable $control{$sum};
# we seize this opportunity to calculate the sequence's length, and add it to its description,
# and also to compare the direct repeats flanking the FB insertion.
until ($sum == $numFB) {
    $control{$sum} = $FBtotal[$sum];
    if ($numFB != 0) {
        if ($control{$sum} =~ /^(.{9}).+({9})$/) {
            $dupl="$1 / $2";
            if ($1 eq $2) {
                $sigaltat= "IDENTICAL";
            } else {
                $sigaltat = "DIFFERENT !!!";
                $difer ++;
            }
        }
        $llargada = length $control{$sum};
        print OUT "FB number ", $sum+1, "; $llargada bp :\n$control{$sum} \nDDuplication: $dupl ... $sigaltat\n";
        print CLUSTALW "\n\n\n>$cromo(", $sum+1, ") , FB ", $sum+1, " of $numFB. $llargada bp.\n$control{$sum}\n";
    } else { print OUT "No FB found\n"; }

# The next step is to search for NOF elements into the identified FB. First we look for entire elements.
# If none is found, then we try to identify 5' or 3' fragments. Any NOF found is counted and their lengths
# are processed.
if ($control{$sum} =~ /TCCTAAAATCTTAAGAATAAA.+ATTTATTCTTAAGATTTTAGGA/) {
    print OUT "***one entire NOF!**\n"; $NOFsender ++;
} elsif ($control{$sum} =~ /TCCTAAAATCTTAAGAATAAA/) {
    print OUT "***one 3' truncated NOF!**\n"; $NOFtr3 ++; $longFBNOF3 = $longFBNOF3 + $llargada;
    $longFBNOF3Q = $longFBNOF3Q + $llargada**2;
} elsif ($control{$sum} =~ /ATTTATTCTTAAGATTTTAGGA/) {
    print OUT "***one 5' truncated NOF!**\n"; $NOFtr5 ++; $longFBNOF5 = $longFBNOF5 + $llargada;
    $longFBNOF5Q = $longFBNOF5Q + $llargada**2;
} else {
    print OUT "***no NOF found!**\n"; $capNOF ++; $longFBnoNOF = $longFBnoNOF + $llargada;
    $longFBnoNOFQ = $longFBnoNOFQ + $llargada**2;
}

```

```

    }
    $sum ++;
}
close OUT;
close CLUSTALW;
}
# Calculate the total number of elements of each type, its average length and standard deviation.
# this are the SD expressions. The program is aware that if no NOF is found, we should we should get an error.

unless ($NOFtr3 == 0) {
    $SDtrun3 = sqrt(($longFBNOF3Q-((($longFBNOF3**2)/$NOFtr3))/NOFtr3);
}
unless ($NOFtr5 == 0) {
    $SDtrun5 = sqrt(($longFBNOF5Q-((($longFBNOF5**2)/$NOFtr5))/NOFtr5);
}
unless ($capNOF == 0) {
    $SDFBnoNOF = sqrt(($longFBnoNOFQ-((($longFBnoNOF**2)/$capNOF))/capNOF);
}
open INFO, ">informe.txt";
print INFO "\nNum.of analysed sequences: $y\n";
print INFO "Num of FB found: $numFBtotal\nNum of NOF found: ", $NOFsencer+$NOFtr3+$NOFtr5 ,"\n\nentire NOFs:
$NOFsencer\n3' truncated NOFs: $NOFtr3\n5' truncated NOFs: $NOFtr5\nSequences without NOF: $capNOF\n\n";
print INFO "Unequal direct duplications: $difer\n\n";

unless ($NOFtr3 == 0) {
    print INFO "Average length of 3' truncated FB+NOF: ", int ($longFBNOF3 / $NOFtr3) ," bp +/- ", int $SDtrun3,"
bp\n";
}
unless ($NOFtr5 == 0) {
    print INFO "Average length of 5' truncated FB+NOF: ", int ($longFBNOF5 / $NOFtr5) ," bp +/- ", int $SDtrun5,"
bp\n";
}
unless ($capNOF == 0) {
    print INFO "Average length of FB with no NOF: ", int ($longFBnoNOF / $capNOF) ," bp +/- ", int $SDFBnoNOF," bp\n";
}
close INFO;

```

Como es obvio, un trabajo de esta envergadura no se realiza en solitario.

La dirección del Grup de Mutagènesi, mis directores de tesis, mis compañeros de laboratorio a lo largo de los años, mis amigos de aquí y de allí y, por supuesto, mi familia

A TODOS ELLOS, MI MÁS SINCERA GRATITUD.