

Tesis Doctoral

**Dinámica evolutiva de las reordenaciones cromosómicas y
coincidencia de los puntos de rotura: análisis molecular de las
inversiones fijadas en el cromosoma 2 de *Drosophila Buzzatii***

Oriol Calvete Torres

Juny 2010

Memoria presentada por el Licenciado
en Biología Oriol Calvete Torres para
la obtención del título de Doctor en
Genética.

Bellaterra, Mayo de 2010

El Doctor Alfredo Ruiz Panadero, Catedrático del Departament de Genètica i Microbiologia de la Facultat de Biociències de la Universitat Autònoma de Barcelona,

Certifica que: Oriol Calvete Torres ha realizado bajo su dirección el trabajo de investigación llevado a cabo en el Departament de Genètica i Microbiologia de la Facultat de Biociències de la Universitat Autònoma de Barcelona que ha conducido a la elaboración de la Tesis Doctoral titulada Dinámica evolutiva de las reordenaciones cromosómicas y coincidencia de los puntos de rotura: Análisis molecular de las inversiones fijadas en el cromosoma 2 de *Drosophila buzzatii*.

Dr. Alfredo Ruiz Panadero

Agraïments

Després de tot aquest temps convivint amb la tesi, queda només agrair a tota la gent que m'ha fet costat en aquest camí. Tot just es massa d'hora per reconèixer el valor just d'aquesta etapa, treure les conclusions adequades i assimilar-ne les experiències. Moltes coses i persones s'han quedat enrere i no sempre per pròpia voluntat. El temps dirà si m'he equivocat. Aquesta reflexió no arribarà a tothom qui voldria, però m'agradaria pensar que durant tot aquest temps s'han sentit apreciades com es mereixen. Tots aquells pensaments i sentiments que no he tingut en compte, temps d'ordenar o que no he sabut transmetre, espero fer-los arribar algun dia.

Tinc el privilegi d'haver gaudit del recolzament necessari en tot moment. Sóc molt conscient que aquesta tesi es un premi a compartir i que tinc la sort que hi hagi gent al meu voltant que s'alegri tant o més que jo de concloure aquesta etapa. N'estic molt orgullós d'haver sobreviscut, però afortunadament no en sóc l'únic. La meva família; els meus pares, el Rodri i l'Esther, la meva germana, la Marta, la meva àvia i la Maria han estat els millors amics que he pogut tenir en aquest viatge. M'han escoltat, entès, recolzat, ajudat i guiat. Aquesta tesi es tant meva com seva, perquè jo sòl no hi hagués arribat. No tots em pogut arribar al final del camí, però sense dubte l'hem iniciat i recorregut plegats. Us estic enormement agraït.

Un munt de persones han contribuït a que aquest treball s'hagi pogut realitzar. Molta gent desconeix el que han significat en aquest procés. Espero algun dia fer-vos arribar com n'heu estat d'importants. Gràcies a tota aquella gent que, per bé o per mal ja formen part d'aquest treball. Gràcies a l'Alfredo i a la resta de membres del meu grup (Ferran, Bàrbara, Marta, Ale, Olivia, als Fernandos, Deodoro i a l'Antonio, Raquel, Natalia i Sonia i, en especial a la Mar...ànims que també et queda poquet! Gràcies Mar per alleugerir-me la càrrega, escoltar-me i fer que les coses hagin estat mes fàcils). A la Pepi (i la resta del trio, Montse i

Tomàs) per l'atenció, dedicació y amistat en els primers passos de l'aventura. A la Carol, la Maria i la Liliana. Amics que he deixat enrere, als nouvinguts i als de tota la vida. A l'Emma i l'Anna. A la gent del pis de dalt pels cafès i dinars! Als qui han estat en tot el procés i aquells que s'han anat incorporant. A la Danae, el Villa i la Claudia. I, sobretot, a tota la colla de tarats que m'heu fet costat i que heu permès que això s'hagi acabat. Per la vostra companyia, col.laboració, ajuda i comprensió. A l'Elsa i al Nahuai. A arreglar el món amb l'Albano. A les previes amb el Marc. Gràcies als berenars a la Autònoma. Al Cristian i la Clara, que m'han aguantat, i molt, quan mes ho he necessitat. Espero viure molts dies de Reis amb vosaltres! A la banda sonora d'aquesta etapa, als cafès al Tribbiani i a les birres arreu. Al Dani i la Neus, al David i la Txell, al Jona i l'Anna i a l'Alessio i la Marta. Als sopars a Gràcia i als partits del Barça. Als qui heu ajudat activament (merci Jude i Adri!) i als que us ho mereixeu de tota manera (Nati, Margue, Nuria i Miqui). Als companys de viatge, a la Fanny i l'Angel, al Dani i de nou a la Maria per haver compartit Keylong. A la gent que m'ha suportat i ajudat als Estats Units. A l' Esther, al Xavi i a la Júlia. A la Mansi. A l'Helena pels Northern i al Clement...per tot!!. Per tota aquesta gent i molta més que no tinc espai per seguir enumerant, ha valgut la pena. Per tota aquesta gent tindria forces per repetir. Gràcies per fer-me apreciar tant aquesta etapa.

Finalment també agrair a l'AGAUR haver-me concedit la beca necessària per realitzar aquest treball y als membres del Tribunal per acceptar la invitació. Dra. Segarra, Dra. Ponsà, Dra. Casacuberta, Dra. Ruiz-Herrera i Dr. Casacuberta.

ÍNDICE

<u>Resumen</u>	1
<u>Abreviaturas</u>	3
<u>1. INTRODUCCIÓN</u>	5
1.1. Reorganizaciones cromosómicas	7
1.1.1. Tasas de fijación	8
1.1.2. Inversiones cromosómicas	9
1.1.3. Origen de las inversiones	10
1.1.3.1. Roturas simples	11
1.1.3.2. Recombinación ectópica	13
1.1.4. Reutilización de los puntos de rotura	14
1.2. Elementos transponibles	16
1.2.1 Clasificación	16
1.2.1.1. Elementos de clase I	18
1.2.1.2. Elementos de clase II	20
1.2.2. Abundancia	20
1.2.3. Impacto evolutivo en los genomas	22
1.2.3.1. Efectos de posición	22
1.2.3.2. Reordenaciones cromosómicas	23
1.2.3.3. Fuente de material codificante y no codificante	24
1.3. Evolución cromosómica en el género <i>Drosophila</i>	25
1.3.1. Cromosomas politénicos	25
1.3.2. Abundancia y distribución de las inversiones	26
1.3.3. Generación de las inversiones en <i>Drosophila</i>	27
1.3.4. Significado adaptativo de las inversiones en <i>Drosophila</i>	31

1.3.4.1. Efectos directos en los cromosomas	32
1.3.4.2. Mecanismos indirectos derivados de la supresión de la recombinación	33
1.3.5. Grupo <i>repleta</i>	35
1.3.5.1. <i>Drosophila buzzatii</i>	36
1.3.5.2. <i>Drosophila mojavensis</i>	37
1.3.5.3. Inversiones $2m$, $2n$ y $2z^7$	38
1.3.5.4. Genoteca y mapa físico de clones BAC del genoma de <i>D. buzzatii</i>	39
1.4. Objetivos de este trabajo	43
2. <u>MATERIALES Y MÉTODOS</u>	45
2.1. Cepas de <i>Drosophila</i>	47
2.2. Protocolos básicos	47
2.2.1. Diseño de cebadores	47
2.2.2. PCR	48
2.2.3. Medios de cultivo	49
2.2.4. Electroforesis	53
2.2.5. Digestión con enzimas de restricción	54
2.2.6. Clonación	54
2.2.7. Extracción de DNA plasmídico	57
2.2.8. Extracción de DNA genómico	58
2.2.9. Secuenciación	60
2.3. Clones BAC	60
2.4. Hibridación <i>in situ</i>	61
2.4.1. Preliminares de la técnica de hibridación <i>in situ</i>	61
2.4.2. Prehibridación	65
2.4.3. Hibridación	65

2.4.4. Posthibridación	66
2.4.5. Revelado	67
2.4.6. Visualización y localización de las señales de hibridación	68
2.5. Electroforesis de campo pulsante (PFGE)	68
2.6. Hibridación de Southern	69
2.6.1 Marcado y detección de sondas con DIG	69
2.6.2. Transferencia	69
2.6.3. Prehibridación e hibridación	71
2.6.4. Posthibridación y detección	71
2.7. Búsquedas de similitud y alineamientos	72
2.8. Secuenciación y anotación de los clones BAC 1N19 y 20O19	73
2.9. Extracción de mRNA	74
2.10. RT-PCR	76
2.11. RACE-PCR	77
2.11.1. 3'RACE	77
2.11.2. 5'RACE	79
2.12. Detección de dsRNA	81
2.13. Hibridación de Northern	84
2.13.1. Preliminares de la técnica de hibridación de Northern	86
2.13.2. Marcado de la sonda con P32	86
2.13.3. Electroforesis	87
2.13.4. Transferencia	88
2.13.5. Prehibridación e hibridación	90
2.13.6. Posthibridación y detección	90
3. <u>RESULTADOS</u>	93

3.1. Inversiones $2m$ y $2n$	95
3.1.1. Notación de los puntos de rotura	95
3.1.2. Cartografía de los puntos de rotura en <i>D. mojavensis</i>	96
3.1.3. Aislamiento de los puntos de rotura en <i>D. buzzatii</i>	105
3.1.4. Electroforesis de campo pulsante e hibridación de Southern	106
3.1.5. Secuenciación de los clones BAC 1N19 y 20O19	115
3.1.6. Anotación y análisis de los clones BAC 1N19 y 20O19	117
3.1.7. Análisis detallado de los puntos de rotura en <i>D. buzzatii</i>	128
3.1.8. ORF en los puntos de rotura AC y BE de <i>D. buzzatii</i>	131
3.1.9. Duplicación asociada a la inversión $2m$	134
3.1.10. Microinversión	135
3.1.11. Análisis detallado de los elementos transponibles en los puntos de rotura	135
3.2. Inversión $2z^7$	138
3.2.1. Notación de los puntos de rotura	138
3.2.2. Cartografía de los puntos de rotura en <i>D. mojavensis</i>	138
3.2.3. Aislamiento del punto de rotura distal en <i>D. buzzatii</i>	140
3.2.4. Análisis de los puntos de rotura	140
3.3. Comparación de las regiones de los puntos de rotura en el género <i>Drosophila</i>	145
3.3.1. Inversiones $2m$ y $2n$	147
3.3.2. Inversión $2z^7$	148
3.4. Gen <i>CG4673</i>	149
3.4.1. El gen <i>CG4673</i> de <i>D. buzzatii</i>	149
3.4.2. Dominios conservados del gen <i>CG4673</i>	154
3.4.3. <i>NPL4</i> en el género <i>Drosophila</i>	156
3.4.4. Región reguladora y promotora del gen <i>CG4673</i>	161

3.4.5. Repeticiones invertidas	167
3.4.6. Expresión del gen <i>CG4673</i> de <i>D. buzzatii</i>	174
3.4.6.1. Amplificación específica de las copias del gen <i>CG4673</i> en <i>D. buzzatii</i>	174
3.4.6.2. RT-PCR del mRNA del gen <i>CG4673</i> en <i>D. buzzatii</i>	178
3.4.6.3. Detección de dsRNA del gen <i>CG4673</i> en <i>D. buzzatii</i>	182
3.4.6.4. Detección de los transcritos del gen <i>CG4673</i> en <i>D. buzzatii</i> mediante hibridación de Northern	186
3.4.6.5. RACE-PCR del <i>CG4673</i> en <i>D. buzzatii</i>	190
4. <u>DISCUSIÓN</u>	197
4.1. Inversiones $2m$ y $2n$	199
4.1.1. Cartografía de los puntos de rotura	199
4.1.2. Aislamiento y secuenciación de los puntos de rotura	200
4.1.2.1. Hibridación de Southern	200
4.1.2.2. Anotación de los clones BAC 1N19 y 20019	202
4.1.3. Generación de las inversiones $2m$ y $2n$	210
4.1.3.1. Duplicación asociada a la inversión $2m$	211
4.1.3.2. Inversión $2n$	218
4.1.4. Relación entre las inversiones $2m$ y $2n$	220
4.1.4.1. Reutilización del punto de rotura	221
4.1.4.2. Relación entre los mecanismos propuestos	223
4.2. Inversión $2z^7$	227
4.2.1. Elementos transponibles anotados en los puntos de rotura	227
4.2.2. Mecanismo de generación de la inversión $2z^7$	229
4.2.3. Posición y consecuencias de los puntos de rotura	231
4.3. Comparación de los puntos de rotura de las inversiones $2m$, $2n$ y $2z^7$ en el	

género <i>Drosophila</i>	232
4.4. Comparación de las inversiones caracterizadas en <i>D. buzzatii</i>	238
4.4.1. Relación entre el mecanismo de generación y el éxito evolutivo de las inversiones	239
4.4.2. TE en los puntos de rotura de las inversiones de <i>D. buzzatii</i>	241
4.4.2.1. Acumulación de TE en el punto de rotura proximal	242
4.4.2.2. Elemento <i>BuT5</i>	244
4.5. NPL4	247
4.5.1. Dominios y función	248
4.5.2. Datos de expresión	251
4.5.3. Región reguladora	253
4.5.4. Transcritos funcionales de <i>CG4673</i> en <i>D. buzzatii</i>	254
4.5.4.1. Transcrito del pseudogen Ψ <i>CG4673</i> en <i>D. buzzatii</i>	255
4.5.4.2. Transcrito antisentido del gen <i>CG4673</i> en <i>D. buzzatii</i>	256
4.5.5. Relación entre los diferentes transcritos del gen <i>CG4673</i>	259
5. <u>CONCLUSIONES</u>	263
6. <u>REFERENCIAS</u>	269
7. <u>ANEXOS</u>	
7.1. Secuencias anotadas obtenidas en este trabajo	CD adjunto

Índice de Tablas

1.1. Inversiones analizadas molecularmente	28
2.1. Secuencias de cebadores	50
2.2. Secuencias de cebadores	51
2.3. Secuencias de cebadores	51
2.4. Secuencias de cebadores	52
2.5. Especificaciones del proceso de secuenciación de los clones BAC 1N19 y 20O19	74
2.6. Sondas usadas en la hibridación de Northern	85
2.7. Composición de los geles usados en la hibridación de Northern	85
2.8. Muestras usadas en la hibridación de Northern	88
3.1. Clones BAC que contienen los puntos de rotura de las inversiones $2m$ y $2n$	98
3.2. Localización de los extremos de los clones BAC en el genoma de <i>D. mojavensis</i>	99
3.3. PCR e hibridaciones <i>in situ</i> del paseo cromosómico	103
3.4. Hibridación de Southern	109
3.5. Localización de los <i>contigs</i> de los clones BAC secuenciados	117
3.6. Anotación del clon BAC 1N19	122
3.7. Anotación del clon BAC 20O19	124
3.8. Contenido de los clones BAC 1N19 y 20O19	127
3.9. Bloques de similitud entre el gen <i>CG4673</i> y la región AC de <i>D. buzzatii</i>	133
3.10. TE anotados en los puntos de rotura AC, BE y EF de <i>D. buzzatii</i>	137
3.11. TE anotados en los puntos de rotura PQ y RS de las especies estudiadas	141
3.12. Ordenación de los genes en las regiones de los puntos de rotura de las inversiones $2m$, $2n$ y $2z^7$ en las 12 especies secuenciadas de <i>Drosophila</i>	146
3.13. Estructura y comparación del gen <i>CG4673</i> de <i>D. buzzatii</i> con <i>D. mojavensis</i>	

y <i>D. melanogaster</i>	151
3.14. Nombre de los genes ortólogos a <i>CG4673</i> , <i>CG5079</i> y <i>CG5071</i> en las 12 especies secuenciadas de <i>Drosophila</i>	157
3.15. Identidad entre las repeticiones invertidas de <i>D. buzzatii</i> y <i>D. koepferae</i> con <i>D. mojavensis</i>	171
3.16. Número de sustituciones nucleotídicas y tiempos de divergencia entre las repeticiones invertidas de <i>D. buzzatii</i> y <i>D. koepferae</i> con <i>D. mojavensis</i>	173
3.17. Bloques de similitud entre los puntos de rotura AC y BE de <i>D. buzzatii</i>	175
3.18. Número de sustituciones nucleotídicas y tiempos de divergencia entre las regiones AC y BE de <i>D. buzzatii</i>	175
3.19. Tamaños de los productos de RT-PCR y cebadores utilizados	181
3.20. Detección de dsRNA	184
4.1. Contenido de TE en los clones BAC 1N19 y 20O19	204
4.2. ANOVA de la densidad de TE anotados en los clones BAC 1N19 y 20O19	206
4.3. Roturas esperadas y observadas en las regiones de los puntos de rotura de las inversiones $2m$, $2n$ y $2z^7$ en las 12 especies secuenciadas de <i>Drosophila</i>	235
4.4. Expresión binomial de la probabilidad de encontrar n roturas en las 12 especies secuenciadas de <i>Drosophila</i>	236

Índice de Figuras

1.1. Mecanismos de inversión	13
1.2. Coincidencia de roturas	15
1.3. Árbol filogenético del complejo <i>buzzatii</i>	38
1.4. Hibridación <i>in situ</i> de los clones BAC de <i>D. buzzatii</i> en <i>D. repleta</i>	41
1.5. Bloques sinténicos entre <i>D. buzzatii</i> y <i>D. repleta</i>	42
2.1. Técnicas de 3' RACE y 5' RACE	78
3.1. Inversiones en el cromosoma 2 de <i>D. buzzatii</i>	95
3.2. Paseo cromosómico	101
3.3. Regiones de los puntos de rotura de <i>D. mojavensis</i>	107
3.4. Hibridación de Southern	111
3.5. Comparación de los clones BAC secuenciados con <i>D. mojavensis</i>	119
3.6. Contenido de los clones BAC 1N19 y 20O19	126
3.7. Anotación de los puntos de rotura de las inversiones $2m$ y $2n$ en <i>D. buzzatii</i>	129
3.8. Estrategia utilizada para amplificar el punto de rotura RS en <i>D. buzzatii</i>	139
3.9. Anotación de los puntos de rotura de la inversión $2z^7$ en las especies estudiadas	143
3.10. Conservación de los puntos de rotura en el género <i>Drosophila</i>	145
3.11. Alineamiento de las traducciones del transcrito RA del gen <i>CG4673</i>	153
3.12. Alineamiento de las traducciones del transcrito RB del gen <i>CG4673</i>	154
3.13. Dominios conservados de la proteína PA del gen <i>CG4673</i> de <i>D. buzzatii</i>	155
3.14. Comparación de las proteínas codificadas por el gen <i>CG4673</i> de las 12 especies secuenciadas de <i>Drosophila</i>	159
3.15. Dominios conservados de los transcritos del gen <i>CG4673</i> de <i>D. persimilis</i> y <i>D. sechellia</i>	160

3.16. Alineamiento nucleotídico de la región del gen <i>CG4673</i> con el navegador VISTA	163
3.17. Predicción bioinformática de promotores	166
3.18. Regiones de los puntos de rotura BE y AC de <i>D. buzzatii</i>	169
3.19. Comparación de las secuencias anotadas como repetición invertida	170
3.20. Alineamiento de los puntos de rotura AC y BE de <i>D. buzzatii</i>	177
3.21. Experimentos de RT-PCR	179
3.22. Experimentos para la detección de dsRNA	185
3.23. Dot Blot e hibridación de Northern	189
3.24. Alineamiento de la región 3' del gen <i>CG4673</i>	191
3.25. Alineamiento de la región 5' del gen <i>CG4673</i>	195
4.1. Posibles mecanismos de inversión del fragmento B-C con duplicación del gen <i>CG4673</i>	215
4.2. Posibles relaciones entre las inversiones $2m$ y $2n$	221
4.3. Mecanismo propuesto de generación del cromosoma $2mn$ de <i>D. buzzatii</i>	224
4.4. Esquema de la localización de las inversiones descritas en el cromosoma 2 de <i>D. buzzatii</i>	245
4.5. Datos de expresión de la transcripción RA del gen <i>CG4673</i> en <i>D. melanogaster</i>	252

Resumen

El interés por los mecanismos que generan las reordenaciones cromosómicas se ha reavivado recientemente gracias al enorme avance que supone la comparación de genomas secuenciados. En *Drosophila*, las inversiones cromosómicas son abundantes como polimorfismos intraespecíficos y como cambios fijados entre especies. Sin embargo, aun sabemos poco acerca de la generación de estas inversiones naturales y sus consecuencias moleculares. Se conocen diversos mecanismos moleculares capaces de generar inversiones cromosómicas. En el mecanismo clásico, un fragmento cromosómico generado por dos roturas puede repararse erróneamente y resultar invertido (NHEJ). Por otro lado, los elementos transponibles (TE) pueden actuar como sustrato de la recombinación ectópica entre copias insertadas en orientación inversa en sitios cromosómicos alejados e inducir inversiones. Los TE también pueden generar inversiones por otros procesos como la transposición aberrante o como simples endonucleasas. De cualquier modo, ambos mecanismos son capaces de producir duplicaciones asociadas a la inversión. En diversos análisis moleculares de puntos de rotura de inversiones en *Drosophila* se han descrito duplicaciones invertidas que la flanquean. Sin embargo, un problema aun no resuelto es si el mecanismo *per se* puede producir las coincidencias y no aleatoriedad observada de los puntos de rotura o, si la selección modula su distribución independientemente de cómo se produzcan.

Este trabajo pretende arrojar luz a este problema y completa el estudio de nuestro grupo de investigación que ha demostrado anteriormente que tres inversiones polimórficas del cromosoma 2 de *Drosophila buzzatii* ($2j$, $2q^7$, $2z^3$) se generaron por recombinación ectópica entre copias del elemento transponible *Galileo*. En la única inversión fijada estudiada previamente en *D. buzzatii* (inversión $5g$) no se ha demostrado la implicación de TE. Con el fin de analizar un posible perfil temporal, hemos llevado a cabo el estudio de los puntos de

rotura de las otras tres inversiones fijadas en el cromosoma 2 de *D. buzzatii* desde su divergencia del ancestro del grupo repleta: $2m$, $2n$ y $2z^7$. Por otro lado, las inversiones $2m$ y $2n$ están dispuestas en tándem, compartiendo un punto de rotura a nivel citológico. En ambos puntos de rotura de la inversión $2z^7$ se han encontrado fragmentos degenerados de un elemento de la familia *Galileo* que sugieren que esta inversión podría haberse originado, igual que las inversiones polimórficas anteriormente estudiadas, por el mecanismo de recombinación ectópica. Por otro lado, el análisis molecular de las inversiones $2m$ y $2n$, ha confirmado la coincidencia citológica de uno de los puntos de rotura. La inversión $2m$ está asociada a una duplicación génica que incluye el gen *CG4673* y dos genes anidados en éste (*CG5071* y *CG5079*). En la posición original (punto de rotura AC), los genes *CG4673* y *CG5079* se han pseudogeneizado. En la duplicación (punto de rotura BE), el gen *CG4673* es funcional pero no hay rastro de los genes anidados debido a la pérdida del intrón 6 (que incluye ambos genes anidados). Esta duplicación parece además, haber sufrido una posterior micro-inversión. Existen tres posibles mecanismos para explicar el origen de esta compleja reorganización y la reutilización observada. (i) primero se originaría la inversión $2m$ mediante NHEJ y posteriormente la inversión $2n$ por recombinación ectópica entre copias del elemento *BuT5*. (ii) primero se originaría la inversión $2n$ mediante NHEJ y posteriormente la inversión $2m$ por inserción híbrida. (iii) que ambas inversiones hubieran sido simultáneas. (1 rotura escalonada + 2 roturas DSB más reparación NHEJ de los dos fragmentos generados).

Finalmente, se habrían localizado hasta tres transcritos relacionados con el gen *CG4673* en *D. buzzatii* por uno sólo en el genoma no invertido de *D. mojavensis*. Uno procedería de la copia degenerada del gen (punto de rotura AC) y otros dos de la copia del gen funcional (punto de rotura BE). Estos transcritos se relacionan formando moléculas de dsRNA que podrían estar regulando la expresión del gen funcional.

Abreviaturas

aa: Aminoácidos

BLAST (*Basic Local Alignment Search Tool*): Herramienta de búsqueda y alineamiento

BAC (*Bacterial Artificial Chromosome*): Cromosoma artificial bacteriano

CDS (*Coding Sequence*): Secuencia codificante

CIP (*Calf Intestine Alkaline Phosphatase*): Fosfatasa alcalina de intestino de ternera

D. buzzatii: *Drosophila buzzatii*

DNA (*DeoxyriboNucleic Acid*): Ácido desoxirribonucleico

DPDB (*Drosophila Polymorphism Database*): Base de datos de polimorfismo de *Drosophila*

DSB (*Double Strand Break*): Rotura de doble cadena

EDTA: Ácido etilen diamino tetraacético

g/mg/µg/ng: Gramos/miligramos/microgramos/nanogramos

GRB (*Genomic Regulatory Block*): Bloque de regulación genómica

L/mL/µL: Litros/militros/microlitros

LCR (*Low Copy Repeat*): Repeticiones de bajo número de copias

LINE (*Long Interspersed Nuclear Elements*): Elementos nucleares dispersos largos

LTR (*Long Terminal Repeats*): Repeticiones terminales largas

M/mM/µM: molar/milimolar/micromolar

ma: Millones de años

NAHR (*Non-Alelic Homologous Recombination*): Recombinación homóloga no alélica

NCBI (*National Center for Biotechnology Information*): Centro Nacional de información biotecnológica

NHEJ (*Non-Homologous End Joining*): Unión de extremos no homólogos

NMD (*Nonsense-Mediated mRNA Decay*): Degradación del mRNA mediada por transcritos

no codificantes (sin sentido)

NPL (*Nuclear Pore Localization*): Localización de poro nuclear

ORF (*Open Reading Frame*): Marco abierto de lectura

pb/Kb/Mb: Pares de bases/Kilobases/Megabases

PCR (*Polymerase Chain Reaction*): Reacción de polimerización en cadena

PEV (*Position-Effect Variegation*): Variegación por efecto de posición

PFGE (*Pulsed Field Gel Electrophoresis*): Electroforesis de campo pulsante

pmoles: Picomoles

UTR (*UnTranslated Region*): Región no traducida

RACE (*Rapid Amplification of Complementary Ends*): Amplificación de los extremos

RM: RepeatMaker

RNA (*RiboNucleic Acid*): Ácido ribonucleico

RP (*Random Primer*): Cebadores al azar

rpm: Revoluciones por minuto

RT: Retrotranscripción

SDS: DodecilSulfato Sódico

SINE (*Short Interspersed Nuclear Elements*): Elementos nucleares dispersos cortos

SSB (*Single Strand Break*): Rotura de una única hebra

Tampón TAE: Tris base, ácido acético y EDTA

Tampón TBE: Tris base, ácido bórico y EDTA

TAP (*Tobacco Acid Pyrophosphatase*): Pirofosfatasa ácida del tabaco

TE (*Transposable Element*): Elemento transponible

TPRT (*Target Primed Reverse Transcription*): Transcripción reversa cebada desde el objetivo

TSD (*Target Site Duplications*): Duplicaciones del sitio de inserción

V: Voltios

1. INTRODUCCIÓN

La variabilidad genética constituye la materia prima del proceso evolutivo. Dicha variabilidad se compone desde un cambio nucleotídico a un cambio estructural en los cromosomas. Las reorganizaciones cromosómicas juegan un importante papel en la dinámica evolutiva de los genomas. Dichas alteraciones en la arquitectura de los cromosomas alteran de forma brusca el orden, orientación y contexto de los genes. Las nuevas funciones y combinaciones resultantes juegan un papel crucial en la adaptación de los individuos al ambiente y en la aparición de nuevas especies (White 1978).

Existen dos grandes tipos de reorganizaciones cromosómicas, los cambios cromosómicos numéricos y los estructurales. Los primeros incluyen variación en el número de cromosomas. Se clasifican en aneuploidías (cambios en el número de copias de un cromosoma), fusiones y fisiones cromosómicas y poliploidías (presencia de tres o más juegos completos de cromosomas o genomas). Los cambios cromosómicos estructurales afectan a la arquitectura. Se pueden clasificar en inversiones, translocaciones (intercambio recíproco de segmentos entre dos cromosomas no homólogos), transposiciones (movilizan una región sin intercambio recíproco), deleciones (pérdida de una porción de un cromosoma) y duplicaciones. Las duplicaciones son segmentos de DNA casi idéntico (segmentales). En este trabajo, se pretende aportar nuevos datos moleculares a la controversia suscitada respecto a los mecanismos que generan dichas reorganizaciones así como respecto a la distribución de sus puntos de rotura.

1.1. Reorganizaciones cromosómicas

Tradicionalmente, las reorganizaciones cromosómicas grandes se han detectado a nivel citológico mediante microscopía óptica. Recientemente, las técnicas moleculares y de secuenciación han permitido detectar a escala más fina dichas reordenaciones, aumentando las

estimaciones iniciales de las frecuencias de estos eventos.

Los segmentos conservados entre especies son regiones genómicas que contienen un conjunto de dos o más genes en el mismo orden y orientación sin interrupción por reorganizaciones (Nadeau y Taylor 1984; Ehrlich *et al.* 1997). Sin embargo, Mural *et al.* (2002), observaron múltiples microrreorganizaciones que interrumpían la mayoría de estos segmentos conservados. Los bloques sinténicos (Pevzner y Tesler 2003a) no representan necesariamente áreas de similitud entre dos genomas sino regiones de similitud que pueden ser interrumpidas por pequeñas áreas desiguales (microrreorganizaciones). De este modo, los bloques sinténicos pueden convertirse en segmentos conservados con microrreorganizaciones. Las regiones de los puntos de rotura se consideran los intervalos entre dos bloques sinténicos consecutivos (Pevzner y Tesler 2003a). La cartografía comparativa entre genomas, permite detectar estos bloques sinténicos conservados entre ambas especies y, por tanto, las reorganizaciones cromosómicas que las diferencian. El número de bloques esperados entre dos genomas que han divergido recientemente será menor y de mayor tamaño que los que resultan de comparar dos genomas con un ancestro común lejano.

1.1.1. Tasas de fijación

Una reorganización cromosómica puede ser polimórfica (si no todos los individuos de una especie la tienen) o fijada (si es común en todos ellos). Para estimar la tasa de fijación de reorganizaciones cromosómicas entre dos especies relacionadas, podemos dividir el número de reorganizaciones entre el tiempo de separación de las especies (dos veces el tiempo de divergencia). Sin embargo, esta estimación no sería comparable entre linajes debido a las diferencias de tamaño de los genomas. De este modo, Ranz *et al.* (2001), estimaron la tasa de reorganización como el número de roturas cromosómicas causadas por las reorganizaciones

por megabase (Mb) por millón de años (ma). Se establecen dos puntos de rotura en una translocación recíproca, los mismos en una inversión y tres en una transposición (Sankoff 1999). Por tanto, si se confunde el evento de reorganización, se contabiliza un número distinto de puntos de rotura, lo que conduce a errores en el cálculo de las tasas. Calcular las tasas a partir de los bloques sinténicos conservados reduce el error (Blanchette *et al.* 1996).

Ranz *et al.* (2001), establecieron una tasa de reorganización para *Drosophila* de entre 0,05 y 0,07 puntos de rotura por megabase por millón de años, lo que supone el doble que la estimada en roedores, 0,03 (Bourque *et al.* 2004). Estas diferencias se pueden explicar, en parte, por parámetros como el tamaño poblacional (10^6 de *Drosophila* contra 10^5 de roedores) y el tiempo generacional (10 generaciones por año contra 2, respectivamente) (Eyre-Walker *et al.* 2002). Por otro lado, el hecho de que las reorganizaciones en nematodos sean menos deletéreas que en *Drosophila* debido a los cromosomas holocéntricos de los primeros, explicaría la elevada tasa observada en nematodos de 0,5-0,7 puntos de rotura por megabase por millón de años (Dernburg 2001).

Aunque las reorganizaciones se han dado durante toda la historia evolutiva, ni el tipo ni la frecuencia coinciden en los distintos eucariotas. En este trabajo, nos hemos centrado en el estudio del origen y consecuencias genómicas y estructurales del tipo más común, las inversiones.

1.1.2. Inversiones cromosómicas

La presencia de segmentos invertidos en los cromosomas fue postulada por primera vez en *Drosophila* por Sturtevant (1917), como explicación a la supresión de la recombinación en determinadas regiones de los cromosomas en el cruzamiento entre algunas cepas de *D. melanogaster*. No fue hasta 1921 (Sturtevant), cuando la comparación de los mapas de

ligamiento confirmó la presencia en el cromosoma 3 de una inversión fijada entre *D. melanogaster* y *D. simulans*. Las inversiones se observaron y describieron por primera vez en los cromosomas politénicos de las glándulas salivares de *Drosophila* (Painter 1934). En estudios posteriores (Sturtevant y Dobzhansky 1936) la observación citológica de los cromosomas politénicos permitió descubrir numerosas inversiones polimórficas en *D. pseudoobscura*. Una inversión es un bloque sinténico en orientación opuesta entre especies (si está fijada) o dentro de especies (si es polimórfica.). Las inversiones están claramente relacionadas con la distancia filogenética, siendo las especies más distantes evolutivamente las que muestran una mayor reorganización en el orden de los genes (Bhutkar *et al.* 2008). Este tipo de alteración estructural puede denominarse pericéntrica o paracéntrica según se incluya o no (respectivamente) el centrómero en el segmento invertido.

1.1.3. Origen de las inversiones

En el modelo clásico (Krimbas y Powell 1992), las inversiones se producen debido a dos roturas aleatorias y simultáneas en un mismo cromosoma y la consiguiente reparación errónea del fragmento (*Non Homologous End Joining*, NHEJ). También pueden generarse inversiones cromosómicas por recombinación ectópica entre secuencias repetidas. Los elementos transponibles (TE) han sido descritos como agentes clásicos de la recombinación ectópica, aunque también pueden generar inversiones por otros mecanismos distintos (véase sección 1.2.3.2).

1.1.3.1. Roturas simples

Las roturas en los cromosomas y los procesos de detección y reparación son comunes en las células eucariotas (Rouse *et al.* 2002). Las roturas pueden ser de una cadena (*Single Strand Break*, SSB) o ambas cadenas (*Double Strand Break*, DSB). Las roturas se producen por simple tensión física (acción propiamente mecánica) e inducidas por agentes clastogénicos como ciertas sustancias químicas o radiaciones ionizantes. Las SSB, por lo general, se producen de forma natural en la horquilla de replicación, mientras que los diferentes agentes como endonucleasas o topoisomerasas y la tensión física en la meiosis pueden producir DSB (Pfeiffer *et al.* 2000). Las DSB producen extremos romos mientras que dos SSB cercanas producen una rotura escalonada de extremos cohesivos. Si una inversión se genera por al menos un extremo escalonado, cuando se reparan los extremos cohesivos se producen duplicaciones en orientación opuesta flanqueando la inversión (Ranz *et al.* 2007) (Figura 1.1A).

Se ha descrito que también los elementos transponibles son una fuente de inestabilidad promoviendo la expansión de repeticiones (Zhou *et al.* 2001) y por sí mismos pueden causar roturas simples de SSB (debido a transposiciones incompletas) y DSB en calidad de endonucleasas (Craig 1997). Las repeticiones terminales invertidas también tienen la capacidad de formar estructuras secundarias, que estimulan la producción de roturas de doble cadena (Lobachev *et al.* 1998). Finalmente, los TE y otras repeticiones pueden actuar como agentes directos de la inversión por recombinación ectópica (Petes y Hill 1988).

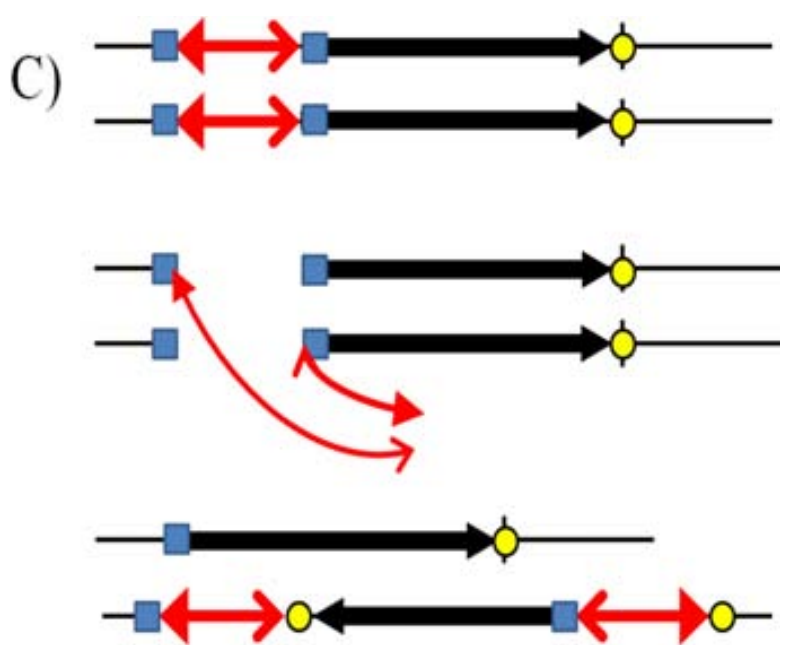
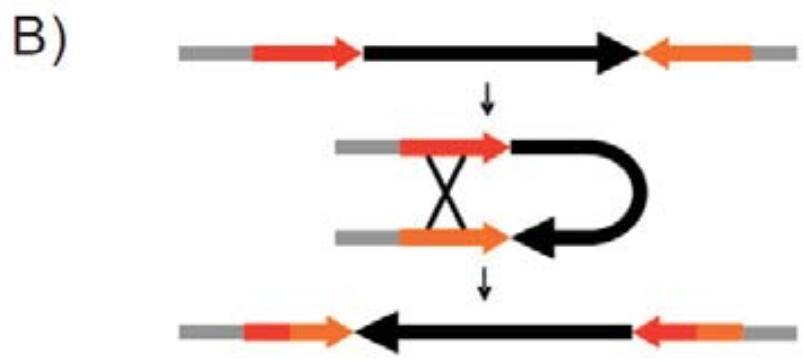
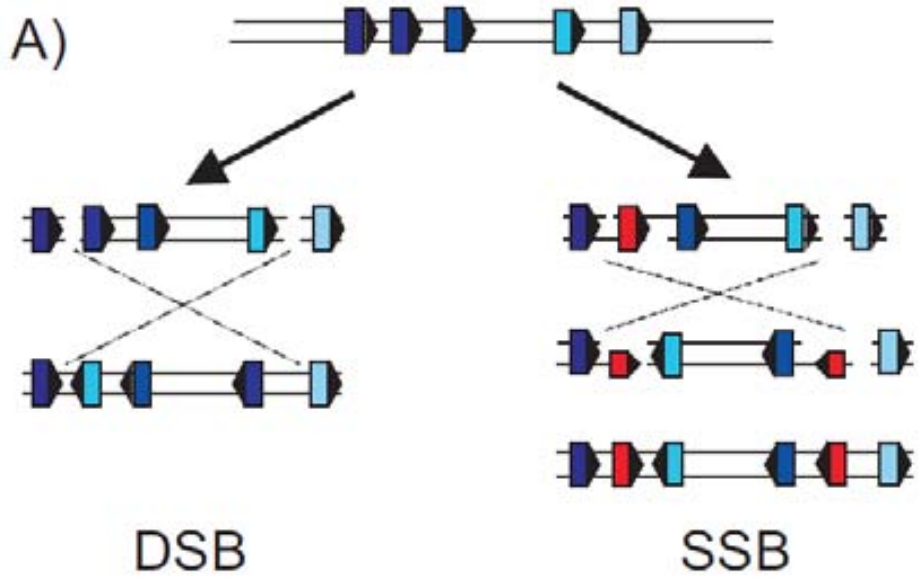


Figura 1.1: (Página anterior). Mecanismos de inversión. A) Rotura simple con reparación errónea (NHEJ). Izquierda, rotura de doble cadena (DSB). En la derecha, rotura de cadena sencilla (SSB). La reparación del extremo cohesivo (rojo) en la SSB, genera una duplicación invertida flanqueando la inversión. B) Recombinación ectópica entre dos fragmentos con homología (flechas naranja y roja). C) Inserción híbrida. Sinapsis bimolecular entre dos extremos de distintos TE debido a la escisión errónea. Los cuadrados azules indican las TSD del TE. El círculo amarillo indica la diana de inserción del elemento híbrido.

1.1.3.2. Recombinación ectópica

Dos secuencias homólogas en diferentes sitios del mismo cromosoma pueden recombinar (Yoon y Richardson 1978). A diferencia de la recombinación alélica entre cromosomas homólogos durante la meiosis, la recombinación ectópica se produce entre regiones no alélicas del genoma (*Non Alelic Homologous Recombination*, NAHR). Si las dos secuencias se encuentran en la misma orientación, el resultado es un fragmento acéntrico y una deleción en el cromosoma. Si las dos secuencias se encuentran en orientación invertida, resulta una inversión (Lim y Simmons 1994). En cualquier caso, en estas recombinaciones, el resultado es un par de elementos quiméricos flanqueando la inversión (Figura 1.1B). Necesariamente, los TE se observan en las juntas de los puntos de rotura. La naturaleza de las secuencias homólogas es dispar (genes de familias génicas, genes de tRNA y rRNA, repeticiones LCR...), aunque algunos trabajos de caracterización molecular de puntos de rotura apuntan a los elementos transponibles como causantes más frecuentes de la recombinación ectópica (Schneuwly *et al.* 1987; Eggleston *et al.* 1996; Schwartz *et al.* 1998; Cáceres *et al.* 1999; Casals *et al.* 2003; Delprat *et al.* 2009). De éste modo, su lugar de inserción en el genoma determinaría las regiones sensibles a la rotura, causando la reestructuración del genoma y conduciendo la evolución de éste (Kazazian 2004).

Los puntos de rotura de los mamíferos tienden a coincidir con regiones enriquecidas con secuencias repetitivas (Dehal *et al.* 2001). Las duplicaciones segmentales también se han descrito como regiones *hotspot* o sitios preferenciales de NAHR o intercambio desigual provocando inversiones. En vertebrados, el análisis comparativo entre humano y ratón localiza entre el 25% y el 53% de los puntos de rotura en regiones con duplicaciones segmentales (Armengol *et al.* 2003). Existe la misma relación en la comparación humano-gorila (Nickerson y Nelson 1998). Bailey *et al.* (2004) reportaron abundantes casos con duplicaciones segmentales primate-específicas relacionadas con las reorganizaciones encontradas en la comparación con el genoma humano. En levaduras existen numerosas regiones repetidas homólogas entre sí que, continuamente se han asociado a las reorganizaciones (Goffeau *et al.* 1996). Las reorganizaciones en *Saccharomyces* se asocian frecuentemente a los transposones *Ty* y otras secuencias multicopia como rRNA y tRNA (Fischer *et al.* 2000; Dunham *et al.* 2002; Yoshida *et al.* 2003).

1.1.4. Reutilización de los puntos de rotura

Se habla de reutilización de dos o más puntos de rotura de reorganizaciones independientes cuando éstos se producen en la misma región entre dos bloques sinténicos en un mismo genoma (González *et al.* 2007). El término reutilización ha sido adoptado erróneamente para definir cuando dos puntos de rotura de reorganizaciones independientes en distintos linajes son descritos en una región ortóloga entre ambos (Figura 1.2) (Sankoff 2009). Esta coincidencia no es secuencial. Estas regiones son recurrentes y susceptibles a la rotura y explican el origen polifilético de algunos puntos de rotura y reorganizaciones. Estas recurrencias no son reutilizaciones y se consideran puntos de rotura evolutivos (Ruiz-Herrera *et al.* 2006). Estrictamente, el término reutilización sólo debería describir una región donde se

acumulen puntos de rotura que generen reorganizaciones distintas secuencialmente en un mismo linaje (Gu *et al.* 2008). Los términos reutilización y recurrencia en este trabajo se utilizan en este sentido.

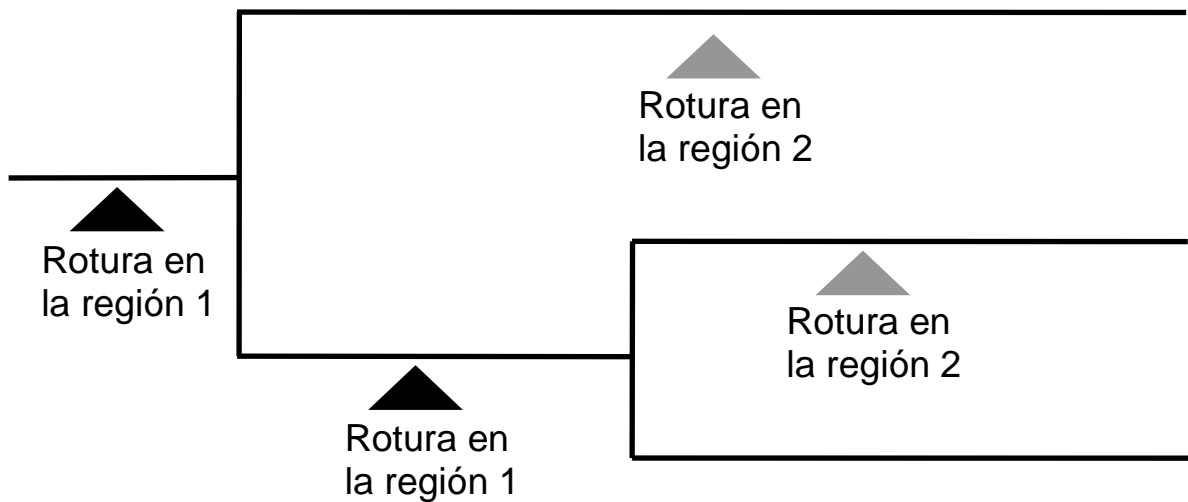


Figura 1.2: Tipos de coincidencia de roturas. Los triángulos indican en que rama del árbol se produce cada rotura. En negro (región 1), cuando se generan de forma secuencial y en gris (región 2), cuando se producen en dos linajes distintos en una región ortóloga. En la región 1 hay reutilización mientras que en la región 2 hay recurrencia.

En el modelo de rotura al azar (Nadeau y Taylor 1984), la distribución esperada de la longitud de los bloques sinténicos es exponencial. Sin embargo, los datos apoyan la distribución no aleatoria de los puntos de rotura en mamíferos (Pevzner y Tesler 2003b; Bourque *et al.* 2004) y en *Drosophila* (Bhutkar *et al.* 2008) ya que algunas regiones acumulan más roturas y tienden a reutilizarse. En el género *Drosophila*, cada punto de rotura observado se usa un promedio de 1,5 veces (Bhutkar *et al.* 2008). En mamíferos, el 20% de los puntos de rotura muestran reutilización (Murphy *et al.* 2005).

El modelo de distribución no al azar de los puntos de rotura propone que las microinversiones y la reutilización son la razón del sesgo en el primer modelo de distribución al azar descrito por Ohno (1973). La mejora en las técnicas de comparación y en las secuencias genómicas permitieron el descubrimiento de nuevos bloques sinténicos en la comparación humano-ratón, que habían permanecido escondidos en el pasado (Mural *et al.* 2002). El desarrollo del programa GRIMM-*synteny algorithm* (Pevzner y Tesler 2003a), permitió postular los dos escenarios más parsimoniosos para transformar el orden de los bloques sinténicos del cromosoma X de ratón a humano. La comparación de los genomas humano y de ratón, ha revelado un exceso de pequeñas reorganizaciones y una extraordinaria densidad de puntos de rotura en regiones particulares del genoma (Bailey *et al.* 2004).

Sin embargo, la razón por la que las roturas no se distribuyen aleatoriamente es todavía controvertida. Se ha propuesto que las causas de esta distribución observada puedan deberse al propio mecanismo de generación o, a procesos de supervivencia diferencial de las inversiones. La hipótesis mutacional (Richards *et al.* 2005) propone la existencia de regiones frágiles o *hotspots* que son preferenciales o susceptibles a las roturas y donde éstas tienden a acumularse (Pevzner y Tesler 2003b; Peng *et al.* 2006). La presencia de regiones frágiles habría sido sugerida en estudios de cáncer e infertilidad (Cohen *et al.* 1996). Clásicamente, los *hotspots* se han descrito como motivos específicos repetidos (repeticiones de CCG y regiones ricas en AT) que confieren una gran flexibilidad y baja estabilidad (Schwartz *et al.* 2006). Por otro lado, los elementos transponibles (TE) tienden a acumularse en regiones de baja recombinación debido a una baja tasa de eliminación (Bartolomé *et al.* 2002). Se han descrito bajas tasas de recombinación dentro de las regiones invertidas en los heterocariotipos (Sturtevant 1917; Armengol *et al.* 2003) que puede llegar a extenderse hasta ~1 Mb de los puntos de rotura fuera de la inversión (Andolfatto *et al.* 2001). Los productos viables del entrecruzamiento serán no recombinantes ya que si éstos incluyen un fragmento de la inversión resultan desequilibrados.

De este modo, se esperan TE de colonización secundaria en los fragmentos invertidos de las inversiones polimórficas no implicados en el origen de la inversión en los puntos de rotura (Langley *et al.* 1988; Ashburner 1989; Montgomery *et al.* 1991; Charlesworth 1996). Sin embargo, la recombinación se reduce únicamente cuando las inversiones son polimórficas y aumenta cuando la inversión aumenta de frecuencia (se reduce el número de heterocariotipos). Por otro lado, después de algún tiempo, los TE acumulados en las inversiones se delecionan ya que la selección actúa en contra de las anomalías cromosómicas causadas por recombinación ectópica. Por lo tanto, la recombinación ectópica es un mecanismo selectivo contra las inserciones de TE (Petrov *et al.* 2003). Las regiones de los puntos de rotura pueden convertirse en regiones *hotspot* de inserción de TE (Cáceres *et al.* 2001, Casals *et al.* 2006). Estos nuevos TE acumulados en los puntos de rotura podrían actuar en nuevos eventos de reorganización por recombinación ectópica reutilizando el punto de rotura (véase sección 1.1.3.2).

Singh y Petrov (2004) calcularon una tasa de desintegración cercana a la mitad de la fracción no codificante en 12 ma. De este modo, después de algún tiempo, es posible que sólo se encuentren trazas de los TE en los puntos de rotura de las inversiones más antiguas.

Por otro lado, la distribución no aleatoria observada podría explicarse por la fijación diferencial (Charlesworth 1974, Kirkpatrick y Barton 2006). Las diferencias en la frecuencia y la observación no al azar de las inversiones se debería al mecanismo, región del cromosoma o impacto en el genoma de cada inversión. La hipótesis selectiva de Becker y Lenhard (2007) propone que la selección actúa contra las roturas que interrumpen sistemas de regulación (*Genomic Regulatory Blocks*, GRB), eliminando estas inversiones por la selección purificadora. En este caso también se observaría una distribución no al azar porque no todas las regiones podrían interrumpirse con puntos de rotura.

1.2. Elementos transponibles

Los elementos transponibles (TE) o móviles fueron descubiertos en el maíz (McClintock 1984), aunque actualmente se han encontrado en todos los organismos desde procariotas a eucariotas superiores (Capy 1998). Debido a su capacidad para replicarse y propagarse por el genoma, están presentes en un elevado número de copias constituyendo una fracción importante de éstos. Además tienen un impacto considerable en la evolución de los genomas que los albergan. Los TE pueden ser autónomos (si conservan la maquinaria necesaria para la transposición) o no autónomos. Mayoritariamente, los TE no autónomos pueden ser movilizados *en trans* mediante la maquinaria de transposición de un elemento autónomo (Craig *et al.* 2002).

1.2.1. Clasificación

Los TE se clasifican según su mecanismo de transposición: elementos de clase I o retrotransposones y elementos de clase II o transposones de DNA (Finnegan 1989).

1.2.1.1. Elementos de clase I

Los elementos de clase I transponen mediante el mecanismo de copiar y pegar (*copy-and-paste*). Este mecanismo implica un intermediario de RNA y la posterior transcripción reversa (retrotranscripción) a DNA. De este modo, a diferencia de los elementos de clase II (ver más adelante), los elementos de clase I aumentan en número tras la transposición. Debido a este mecanismo de inserción, estos elementos generan duplicaciones del sitio de inserción (TSD) flanqueando los extremos. Estas TSD se generan por la reparación de la cadena sencilla

de DNA de las roturas escalonadas de dichos sitios de inserción.

Este grupo se subdivide en función de la presencia o no de repeticiones terminales largas (LTR) en los elementos. Los elementos tipo *Penelope* forman un grupo distinto aunque aparentemente ausente en los mamíferos, y se encuentran en las primeras etapas de caracterización (Gladyshev y Arkhipova 2007). Los elementos tipo retrovirales o retroelementos (ERV) con repeticiones terminales largas (LTR) incluyen los retrovirus endógenos. Estos elementos se retrotranscriben de forma similar a los virus. Los retrotransposones con LTR codifican los genes *gag*, *pol* y *env* (Capy *et al.* 1998), que permiten su movilización. Los retrotransposones sin LTR se subdividen en tipos *LINE* y *SINE*. Los primeros tienen la maquinaria para movilizarse (autónomos) y codifican dos ORF. El ORF1 codifica una proteína de unión a RNA y el ORF2 una proteína multifuncional con actividad endonucleasa y transcriptasa reversa. Sin embargo, sus abundantes copias suelen encontrarse sin las regiones 5' y pierden la capacidad de retrotranscribirse aun en presencia de otros elementos. Los elementos *SINE* no son autónomos, y necesitan de la maquinaria que codifican los *LINE*.

La transcripción reversa de los retrotransposones sin LTR se inicia desde del propio extremo 3'OH del sitio de inserción (mecanismo de *Target Primed Reverse Transcription*, TPRT) (Ostertag y Kazazian 2001). Según este modelo, primero se transcribe el mRNA del elemento desde el promotor *polIII*. Las proteínas necesarias se sintetizan desde los ORF1 y ORF2. Dichas proteínas se asocian al mRNA y el conjunto ribonucleoproteico retorna al núcleo. La endonucleasa (ORF2) hace un corte en la cadena complementaria en el DNA en el sitio de inserción para exponer un 3'-hidroxilo desde donde se cebará la retrotranscripción del RNA. El mRNA sirve de molde, retrotranscribiendo desde la poli-A (que queda insertada). Las inserciones en nuevas ubicaciones pueden ser completas o truncadas del extremo 5' si no se finaliza la retrotranscripción (Eickbush y Jamburuthugoda 2008).

1.2.1.2. Elementos de clase II

Estos transposones se mueven utilizando un intermediario de DNA de cadena única o doble y la mayoría no tienen mecanismos replicativos autónomos (Craig *et al.* 2002). En eucariotas, se subdividen en tres clases en función de su mecanismo de transposición: (i) Elementos tipo cortar y pegar (*cut-and-paste*) (Plasterk 1995), que se transponen por escisión directa del DNA de doble cadena y vuelve a insertarse en otro lugar del genoma. (ii) Elementos tipo *Helitron* (Kapitonov y Jurka 2001; Feschotte y Wessler 2001), que utilizan un mecanismo de replicación tipo círculo rodante (*rolling circle*). (iii) Elementos tipo *Maverick* (Pritham *et al.* 2007) que aparentemente replican mediante el uso de una DNA polimerasa propia. Los elementos tipo cortar-y-pegar y *Maverick* tienen repeticiones terminales invertidas (TIR) aunque los *Helitron* también tienen motivos cortos conservados. Los elementos autónomos de los transposones tipo cortar-y-pegar codifican un ORF (transposasa), aunque algunos codifican una segunda proteína necesaria para la transposición (Pritham y Feschotte 2007). Además, los elementos *Helitron* codifican helicasas (Kapitonov y Jurka 2006) y los *Maverick* algunas proteínas relacionadas con las codificadas en los virus de doble cadena (Pritham *et al.* 2007).

1.2.2. Abundancia

La abundancia y la contribución relativa de los diferentes tipos de elementos transponibles es muy variable en los distintos organismos. La capacidad de replicación y movilización de los TE hace que se encuentren en alto número de copias en los genomas pero también que exista alta variabilidad entre individuos dentro de una especie. En procariotas sólo se han descrito TE de DNA, mientras que los eucariotas contienen ambos tipos, si bien

los más abundantes en organismos superiores son los de tipo I. Las especies eucariotas muestran enormes variaciones en la cantidad de TE en sus genomas. Estas variaciones explican en gran medida las grandes diferencias en el tamaño los genomas e incluso entre especies estrechamente relacionadas (Kidwell 2002).

Los TE *SINE* y *LINE* son los más abundantes (33,4%) en el genoma humano (Lander *et al.* 2001). Sin embargo, los transposones de DNA (300.000 copias) representan cerca de 15 veces más que en *C. elegans* y 40 veces más que en *D. melanogaster* (Feschotte y Pritham 2007).

Respecto a la contribución relativa, existe una gran variación entre la abundancia de elementos de tipo I y tipo II desde la práctica ausencia de elementos de DNA en *Saccharomyces* hasta práctica ausencia de retroelementos en *Trichomonas* (Pritham y Feschotte 2007). El porcentaje relativo del número total de TE de tipo I llega hasta el ~90% en humano y ratón y algo menos en *D. melanogaster* (~80%). En cambio, sólo el ~15% de la cantidad relativa de TE son de tipo I en *O. sativa* y *C. elegans*.

El genoma de *Trichomonas vaginalis* está compuesto por un 37% de TE (Feschotte y Pritham 2007). En *C. elegans* los más abundantes son los de tipo II, que suponen en 5,3% del genoma. En *S. cerevisiae*, el 3,1% de la fracción genómica son TE tipo LTR (Kim *et al.* 1998). En *D. melanogaster*, se han contabilizado 1.572 TE (completos y parciales) en la fracción eucromática clasificados en 96 familias distintas (Adams *et al.* 2000). El número de copias de cada familia varía de 1 a 146 y representan el 3,86% de la fracción eucromática, aunque considerando también la heterocromatina podría llegar hasta el ~9% (Spradling y Rubin 1981). Los LTR son los más abundantes (682 copias) representando un 2,65%, mientras que los *LINE* (486), los TIR (372) y los *Foldback* (32), representan el 0,87%, el 0,31% y el 0,04%, respectivamente (Kaminker *et al.* 2002).

1.2.3. Impacto evolutivo en los genomas

Históricamente se han considerado parásitos de los genomas (Orgel y Crick 1980) y potentes múgatenos que contribuyen a la inestabilidad genómica (Hedges y Deininger 2006). Sin embargo, se han descrito como una fuerza poderosa que contribuye a la alteración de la arquitectura de los genomas. El impacto de los elementos transponibles ha sido ampliamente estudiado (Pritham y Feschotte 2007; Goodier y Kazazian 2008). Los TE contribuyen a la generación de la diversidad alélica en las poblaciones naturales y modelizan el contexto genómico y epigenético de los huéspedes de tres modos distintos: (i) afectando la función génica o reguladora a través de su inserción, (ii) promoviendo reordenamientos cromosómicos y, (iii) como una fuente de material codificante y no codificante que permite la aparición de novedades genéticas (tal como nuevos genes y secuencias reguladoras) (Kidwell y Lish 2001).

1.2.3.1. Efectos de posición

Las inserciones de TE pueden ir acompañadas por deleciones en los sitios de inserción y escisión. La mayoría de los elementos de DNA parecen tener preferencia de inserción en regiones ricas en genes (Pritham y Feschotte 2007). De este modo, la posibilidad de que la función génica se altere aumenta considerablemente. La presencia de TE puede influenciar la epigenética de los genes cercanos (Slotkin y Martienssen 2007). Las inserciones de estos TE en regiones reguladoras (Girard y Freeling 1999) o regiones codificantes pueden alterar el marco de lectura (codones stop prematuros, procesamientos alternativos...) y producir nuevos puntos de inicio de transcripción de los genes (Goodier y Kazazian 2008). Las inserciones de TE también pueden promover la expansión del número de microsatélites alterando el marco de lectura y errores en la replicación (Arcot *et al.* 1995).

Algunos transposones de DNA desempeñan un papel importante en la atracción de la maquinaria responsable de la formación y el mantenimiento de la heterocromatina debido a su naturaleza repetitiva (Lippman *et al.* 2004). La inserción en regiones promotoras de elementos silenciados en la heterocromatina, produce un efecto de heterocromatinización de la región y silenciamiento de los genes implicados (Grewall *et al.* 2007).

Finalmente, los elementos de DNA (Lisch 2002) y en algún caso de retroelementos (Xing *et al.* 2006) pueden capturar genes y movilizarlos a otro lugar del genoma, modificando su contexto genético. Estos elementos de DNA (denominados *MULE*) descritos mayormente en plantas (Jiang *et al.* 2004) actúan como transportadores pudiendo también barajar y duplicar genes (Juretic *et al.* 2005). El mecanismo de círculo rodante de transposición de los elementos tipo *Helitron* sugiere que también puedan actuar como los elementos tipo *MULE*.

1.2.3.2. Reordenaciones cromosómicas

Los TE pueden generar inversiones de muchas formas. Las escisiones incompletas y las huellas de inserción de los elementos de tipo II producen pequeñas reorganizaciones (Brookfield 2004). Los elementos de DNA han sido relacionados con la generación de roturas de DSB y SSB (véase sección 1.1.3.1). Mediante recombinación ectópica entre copias de elementos transponibles se pueden producir deleciones, duplicaciones e inversiones (véase sección 1.1.3.2). Se ha demostrado en algunas inversiones naturales en *D. buzzatii* el mecanismo de recombinación ectópica entre distintas copias del elemento *Galileo* (véase sección 1.3.3).

Finalmente, los elementos transponibles de tipo II parece que también podrían generar reorganizaciones cromosómicas mediante transposición aberrante o alternativa (Lim y Simmons 1994; Gray 2000). A pesar de que no se cumplan las condiciones necesarias para la

recombinación ectópica, una transposición inusual de éstos (Gray *et al.* 1996) puede producir inversiones, duplicaciones y deleciones. En primer lugar, se produciría la sinapsis de los extremos de elementos transponibles complementarios dando como resultado un elemento híbrido (sinapsis bimolecular) (Chalmers y Kleckner 1996). Después de la escisión, el elemento híbrido se insertaría en una nueva diana. El mecanismo de reparación de roturas de doble cadena restablecería las cadenas aunque reorganizadas como en el mecanismo tradicional de recombinación homóloga. Este mecanismo de reparación es homólogo al de recombinación V(D)J que tiene lugar en células somáticas del sistema inmunitario en el proceso de producción de anticuerpos (Zhou *et al.* 2004). El tipo de reorganización resultante de esta transposición aberrante dependerá de la localización cromosómica de la diana de inserción (Zhang *et al.* 2009). En el modelo clásico, una inversión se producirá si el elemento híbrido se inserta en el brazo cromosómico implicado en la formación de dicho híbrido (Figura 1.1C). Las copias de TE implicadas resultan idénticas y no quiméricas como en el modelo de recombinación ectópica.

1.2.3.3. Fuente de material codificante y no codificante

Los retrotransposones parecen haber contribuido en mayor grado al incremento del tamaño de los genomas eucariotas (Piegu *et al.* 2006), aunque recientemente, también se han descrito aumentos de tamaño con transposones de DNA (Pritham *et al.* 2007). A menor escala, el proceso de transcripción de los retroelementos puede incorporar un fragmento de la región a 3' o 5' del DNA del huésped provocando pequeñas duplicaciones (Goodier y Kazazian 2008). Por otro lado, la domesticación de TE puede contribuir de forma directa en el montaje de nuevos genes y funciones reguladoras (Volff 2006). El proceso más habitual de domesticación consiste en proporcionar la transposasa a antiguos genes o genes en formación.

1.3. Evolución cromosómica en el género Drosophila

El género *Drosophila* es uno de los organismos más utilizados en estudios de biología en general y de evolución en particular. Una de las razones es la presencia de cromosomas politénicos en sus glándulas salivares que permitió descubrir y analizar las inversiones cromosómicas. Durante más de 70 años, el estudio de las inversiones cromosómicas de *Drosophila* ha sido un paradigma de los estudios de evolución de las poblaciones.

1.3.1. Cromosomas politénicos

Los cromosomas politénicos se encuentran en las glándulas salivares del tercer estadio larval (Sorsa 1988) de muchas especies de dípteros (Lacadena 1996). Este tipo de cromosomas interfásicos son el resultado de un proceso de endoreduplicación y se producen por el apareamiento de dos moléculas de DNA homólogas seguido de repetidos ciclos de replicación sin división celular. En las glándulas salivares de *Drosophila*, estos cromosomas se forman debido a 10 rondas de replicación sucesivas que dan lugar a la yuxtaposición de 1.024 moléculas de DNA. Su observación a través del microscopio óptico permite distinguir fácilmente un patrón constante de bandas oscuras e interbandas claras. El patrón de bandas dentro de los segmentos conservados entre dos especies también se conserva. La comparación de los patrones de bandas entre cromosomas permite detectar las reorganizaciones con simples técnicas citológicas. Por tanto, estos cromosomas hacen que *Drosophila* sea el organismo ideal para desarrollar estudios sobre reorganizaciones cromosómicas.

1.3.2. Abundancia y distribución de las inversiones

Las inversiones son, sin duda, la reorganización más abundante en *Drosophila*. Los estudios de la variación cromosómica existente en poblaciones naturales en distintas especies del género *Drosophila* desvelaron un sorprendente grado de polimorfismo cromosómico para inversiones paracéntricas. Se han estimado entre 18.000 y 28.000 inversiones polimórficas y entre 22.000 y 56.000 inversiones fijadas durante la evolución del género (Stone *et al.* 1960). Aproximadamente dos tercios de las especies del género *Drosophila* presentan inversiones que segregan en las poblaciones naturales (Sperlich y Pfriem 1986; Powell 1997).

Existe una gran variedad en el número de inversiones entre especies. Este patrón parece ser distinto a lo esperado de acuerdo con una distribución al azar de las inversiones. Mientras se han descrito especies completamente monomórficas, dentro de las especies que acumulan mayor número de inversiones destacan *D. willistoni* y *D. paulistorum* (grupo *willistoni*) con 50 y 63 inversiones respectivamente (Dobzhansky y Powell 1975) y *D. subobscura* (grupo *obscura*) con 66 inversiones (Krimbas 1992). Ambos grupos pertenecen al subgénero *Sophophora*, que parece contener una proporción muy superior de especies polimórficas para inversiones que los demás subgéneros, *Drosophila* y *Idiomya* (Carson 1992). Por el contrario, la distribución de las inversiones en los cromosomas, tanto en *D. subobscura* como en *D. willistoni*, parece ser aleatoria y se observa un elevado grado de polimorfismo en todos los cromosomas (Sperlich y Pfriem 1986). En *D. pseudobscura* han sido descritas 34 inversiones polimórficas en el cromosoma 3, tres en el cromosoma X, pero ninguna en el resto (Powell 1992). Por el contrario, en este caso aislado, la distribución de las 34 inversiones del cromosoma 3 no difiere significativamente de una distribución uniforme (Olvera *et al.* 1979). Aun así, la distribución de los puntos de rotura de las inversiones dentro de los cromosomas no parece al azar (véase sección 1.1.4) y tienden a acumularse en regiones

cercanas al centrómero (Tonzetich *et al.* 1988; Krimbas y Powell 1992). En *D. melanogaster* se contabilizaron los puntos de rotura por subdivisión cromosómica y su distribución fue comparada con la distribución esperada por azar según una distribución de Poisson (Lemeunier y Aulard 1992). A excepción del cromosoma X, que presenta muy pocas inversiones, se demostró que la distribución observada discernía significativamente de la esperada por azar.

Tampoco existe distribución al azar en los elementos cromosómicos básicos A, B, C, D, E, F de Muller (1940). Estos elementos se han conservado en estructura y contenido a lo largo de la evolución del género debido a que la mayoría de cambios cromosómicos han consistido en inversiones paracéntricas y fusiones céntricas (Clayton y Guest 1986).

1.3.3. Generación de las inversiones en *Drosophila*

Los TE han sido relacionados con reorganizaciones a gran escala y con inversiones naturales. La técnica de hibridación *in situ* en cromosomas politénicos de *D. melanogaster* logró relacionar patrones de distribución del elemento *hobo* con los puntos de rotura de cuatro inversiones endémicas (Lyttle y Haymer 1992). En *D. willistoni* se han relacionado los lugares de inserción de los elementos *P* y la localización de puntos de rotura de algunas inversiones naturales (Regner *et al.* 1996). Evgen'ev *et al.* (2000) reportaron que la localización genómica de los transposones *Penelope* y *Ulysses* no tiene una distribución al azar en *D. virilis*. Los sitios de inserción de ambos elementos estarían relacionados con los puntos de rotura de las inversiones encontradas en otras especies del grupo *virilis*.

Los análisis moleculares de los puntos de rotura y la discusión de los mecanismos de origen no coinciden en los distintos estudios de inversiones entre especies de *Drosophila* (Tabla 1.1). Tampoco parece existir un patrón o mecanismo especie-específico.

Tabla 1.1: Inversiones analizadas molecularmente en los géneros *Drosophila* y *Anopheles*. Para cada inversión se muestra la especie, su frecuencia, y si hay repeticiones y/o duplicaciones anotadas en los puntos de rotura. Los TE se indican en que punto de rotura se localizan, distal (D) o proximal (P) y si están anotados en las juntas de los puntos de rotura (*). Las duplicaciones pueden ser de ambos puntos de rotura (^α), de uno (^β) o en tándem en el mismo punto de rotura (^γ). También se indica el mecanismo más plausible descrito para cada inversión. Un interrogante indica cuando el trabajo no ha resuelto el mecanismo que ha originado la inversión, aunque se propone el mecanismo más consistente.

Inversión						
(E. de Muller)	Especie	Frecuencia	Duplicaciones	Repeticiones	Mecanismo más plausible	Referencia
<i>In(3L)Payne</i> (D)	<i>Dmel</i>	Polimórfica	No	No	?NHEJ x DSB	Wesley y Eanes 1994
<i>D.mel/D.sub</i>	<i>Dmel</i>	Fijada	No	No	?NHEJ x DSB	Cirera <i>et al.</i> 1995
<i>2Rd'(E)</i>	<i>Agam</i>	Polimórfica	No	<i>Odysseus*</i> (distal)	?	Mathiopoulos <i>et al.</i> 1998
<i>2j(E)</i>	<i>Dbuz</i>	Polimórfica	No	Homólogas (<i>Galileo*</i>) + otros	Rec. ectópica	Cáceres <i>et al.</i> 1999
<i>In(2L)t</i> (B)	<i>Dmel</i>	Polimórfica	No	<i>LINE*</i> (D)	?	Andolfatto <i>et al.</i> 1999
<i>2q⁷(E)</i>	<i>Dbuz</i>	Polimórfica	No	Homólogas (<i>Galileo*</i>) + otros	Rec. ectópica	Casals <i>et al.</i> 2003
<i>In(3R)Payne</i> (E)	<i>Dmel</i>	Polimórfica	Invertidas ^α	<i>G*</i> (proximal)	NHEJ x SSB?	Matzkin <i>et al.</i> 2005
<i>Arrowhead</i> (C)	<i>Dpse</i>	Polimórfica	No	Homólogas ≠ TE*	Rec. ectópica ≠ TE	Richards <i>et al.</i> 2005
<i>2La</i>	<i>Agam</i>	Polimórfica	Invertidas ^γ	Complejo de TE*	?Inserción híbrida o NHEJ(SSB)	Sharakhov <i>et al.</i> 2006
<i>3R(8)</i> (E)	<i>Dyak</i>	Fijada	Invertidas ^α	<i>BS</i> y <i>CR1a</i> (distal)	NHEJ x SSB	Ranz <i>et al.</i> 2007
<i>3R(7)</i> (E)	<i>Dmel</i>	Fijada	Invertidas ^α	No	NHEJ x SSB	Ranz <i>et al.</i> 2007
<i>Xe</i> (A)	<i>Dmoj</i>	Fijada	Invertidas ^β	No	?	Cirulli y Noor 2007
<i>In(4)a</i> (B)	<i>Dame</i>	Polimórfica	No	Homólogas (<i>MITE*</i>)	Rec. ectópica	Evans <i>et al.</i> 2007
<i>5g</i> (C)	<i>Dbuz</i>	Fijada	No	<i>Helitron</i> (D) + <i>TART</i> (P)	NHEJ (DSB)	Prazeres <i>et al.</i> 2009
<i>2z³</i> (E)	<i>Dbuz</i>	Polimórfica	No	Homólogas (<i>Galileo*</i>) + otros	Rec. ectópica	Delprat <i>et al.</i> 2009

Recientemente, Evans *et al.* (2007) concluyeron que la inversión polimórfica *In(4)a* de *D. americana* se habría originado por recombinación ectópica entre copias de un elemento *MITE*. Las tres inversiones polimórficas estudiadas en *D. buzzatii*: $2z^3$ (Delprat *et al.* 2009), $2q^7$ (Casals *et al.* 2003) y $2j$ (Cáceres *et al.* 1999) se originaron mediante recombinación ectópica entre dos copias del elemento de clase II *Galileo*. Estas inversiones pudieron ser relacionadas con la recombinación ectópica al describir los TE exactamente en las juntas de los bloques conservados. Por el contrario, aunque se encontraron trazas de TE en las regiones de los puntos de rotura de la inversión $5g$ (Prazeres *et al.* 2009), fijada en el genoma de *D. buzzatii*, no se relacionaron con la recombinación ectópica. En *D. melanogaster*, en el análisis molecular de las inversiones *In(2L)t* (Andolfatto *et al.* 1999) y *In(3R)Payne* (Matzkin *et al.* 2005), se encontraron trazas de TE en los puntos de rotura. Estos TE están aparentemente anotados en las juntas según la interpretación de las figuras de estos trabajos pero el análisis molecular no lo especifica y la localización no es tan precisa como en los trabajos descritos anteriormente. Igualmente, estos TE tampoco pudieron ser relacionados con la recombinación ectópica. En la inversión *In(3R)Payne* se identificaron complejas reorganizaciones en sus puntos de rotura, incluyendo la duplicación y fragmentación de los genes implicados en los puntos de rotura. La inversión quedaría flanqueada externamente por pseudogenes de los genes que flanquean internamente los puntos de rotura. Las roturas escalonadas en ambos puntos de rotura podrían ser la causa de esta inversión.

No se encontraron TE en la inversión *Arrowhead* en *D. pseudoobscura* (Richards *et al.* 2005) pero se describieron repeticiones homólogas en ambos puntos de rotura. Esta inversión podría haberse originado por recombinación ectópica entre estas repeticiones de origen desconocido. Tampoco se encontraron TE en los puntos de rotura de la inversión *In(3L)Payne* en *D. melanogaster* (Wesley y Eanes 1994). Esta inversión parece el resultado de un proceso limpio de cortar-pegar, por lo que el mecanismo de inversión por rotura de doble cadena

(DSB) con reparación errónea (NHEJ) podría explicarla. En los estudios de los puntos de rotura de una inversión fijada desde la divergencia de los linajes de *D. melanogaster* y *D. subobscura* (Cirera *et al.* 1995) y otra en el cromosoma X entre las especies *D. mojavensis* y *D. arizonae* (Cirulli y Noor 2007) tampoco se encontraron TE. Por el contrario, en la primera, se encontró una nueva unidad de transcripción en uno de los puntos de rotura. En la segunda se encontró una duplicación invertida en uno de los puntos de rotura, aunque también fue observada en los cromosomas parentales no invertidos (Runcie y Noor 2009).

En los demás invertebrados los elementos repetitivos tendrían un papel fundamental ya que las reorganizaciones suceden en regiones ricas en repeticiones tanto en Nemátodos (Coghlan y Wolfe 2002) como en Anopheles. En la inversión polimórfica *2La* (Sharakhov *et al.* 2006) se describió la duplicación parcial de los genes implicados en el punto de rotura proximal y la presencia de DNA repetitivo y elementos transponibles degradados en las juntas de los puntos de rotura. El elemento *Odysseus* fue descrito en la junta del punto de rotura distal aunque no pudo ser relacionado inequívocamente con el origen de la inversión *2Rd'* (Mathiopoulos *et al.* 1998).

Ranz *et al.* (2007) realizaron una comparación de los genomas de algunas especies del grupo *melanogaster* concluyendo que el 59% de las inversiones fijadas estudiadas están asociadas a duplicaciones invertidas de genes. En el genoma de *D. yakuba*, el rango de estas repeticiones varía ente 46 pb y 1,9 Kb (en un promedio de 321 pb). Estos resultados sugieren que el mecanismo predominante en estas especies es el NHEJ con roturas escalonadas. Recientemente, se realizó una comparación a gran escala analizando y comparando los 12 genomas secuenciados de *Drosophila* (Bhutkar *et al.* 2008). Este estudio compara las tasas de reorganizamiento de las especies y dentro de los brazos cromosómicos pero no resuelve los mecanismos de origen ni el contenido de los puntos de rotura.

1.3.4. Significado adaptativo de las inversiones en *Drosophila*

Existen diferentes explicaciones para la expansión y adaptación de las inversiones (Hoffmann y Rieseberg 2008). La deriva genética es la fuerza que conduce la frecuencia de las inversiones con efecto neutro en una población (Hartl 1970; Liberman 1976; White 1978).

Sin embargo, la frecuencia de las inversiones varía de forma regular en el espacio y el tiempo proporcionando una fuerte evidencia que los polimorfismos de inversión se mantienen por selección (Mettler *et al.* 1977; Stalker 1980, Krimbas y Powell 1992). El significado adaptativo de las inversiones determina su frecuencia en una población.

Las frecuencias de las inversiones en las poblaciones pueden variar en forma de clinas geográficas en función del ambiente (Prevosti *et al.* 1988). Incluso si la ventaja supone adaptación local que varía con el ambiente, se pueden establecer frecuencias cíclicas estacionales (Dobzhansky 1971; Lee *et al.* 2002).

La selección actúa sobre los distintos efectos derivados de las inversiones. Por un lado, en condiciones de sobre dominancia de las inversiones, pueden establecerse como polimorfismos equilibrados debido al exceso de heterocariotipos (Rieseberg 2001). En condiciones de baja dominancia de las inversiones, pueden expandirse hasta la fijación o desaparecer. La sobre dominancia (heterosis) se produce cuando los heterocigotos para una inversión tienen mayor eficacia respecto los homocigotos invertidos o no invertidos. La baja dominancia se refiere a la situación en la que los heterocigotos tienen menor eficacia.

Por otro lado, la frecuencia de una inversión varía debido a la selección local ejercida en los efectos directos de los puntos de rotura y los efectos derivados de la supresión de la recombinación (Kirckpatrick y Barton 2006). La selección actúa sobre las consecuencias funcionales de los efectos de posición de los puntos de rotura de las inversiones en los cromosomas. Finalmente, la supresión de la recombinación en las inversiones retiene grupos

de alelos favorables (desequilibrio de ligamiento) porque el producto del apareamiento con cromosomas no invertidos resulta desequilibrado durante la meiosis (Navarro *et al.* 1997). La supresión de la recombinación se extiende más allá de los puntos de rotura (Andolfatto *et al.* 2001) aunque en las regiones centrales de las inversiones grandes ocurre normalmente (Schaeffer *et al.* 2003; Kennington *et al.* 2006). Las combinaciones alélicas retenidas con mayor eficacia que de forma independiente, están coadaptadas (Dobzhansky 1970; Schaeffer *et al.* 2003). Las inversiones también pueden verse seleccionadas por llevar *loci* adaptados aun sin condiciones de epistasia (Kirckpatrick y Barton 2006).

1.3.4.1. Efectos directos en los cromosomas

Algunos puntos de rotura tienen un claro efecto directo. En función de donde se produzcan, pueden interrumpir sistemas de regulación o incluso genes, modificando el patrón de expresión. Una inversión también puede afectar la regulación de un gen al cambiarlo de su contexto genómico (Sperlich 1963). Apenas se conoce el efecto mutacional de los puntos de rotura y su posible efecto en el transcriptoma ya que existen pocos estudios moleculares de los genes cercanos a los puntos de rotura de inversiones polimórficas de *Drosophila* y ninguno de inversiones fijadas. Un posible cambio en la expresión génica fue documentada en el estudio de la inversión *In(3L)Payne* de *D. melanogaster*, aunque nunca fue confirmado (Wesley y Eanes 1994). El estudio de la inversión polimórfica *2j* de *D. buzzatii* (Puig *et al.* 2004) muestra el silenciamiento de un gen cercano al punto de rotura proximal en los embriones de las líneas fijadas con la inversión *2j*. La frecuencia en las poblaciones de la inversión *2j* es clinal. Los individuos con la inversión *2j* se caracterizan por un mayor tamaño del cuerpo adulto y un mayor tiempo de desarrollo que los individuos con la ordenación *2st*.

Por otro lado, pueden existir efectos de posición distintos a los efectos derivados de la localización de los puntos de rotura. Las duplicaciones asociadas a los mecanismos de roturas escalonadas pueden incluir genes. Estos genes duplicados pueden adquirir nuevas funciones o subfuncionalizar las antiguas.

Las inversiones también pueden recolocar un gen en un nuevo contexto regulador. Ocasionalmente, se puede trasladar un gen cerca de la fracción heterocromática y quedar parcialmente inactivo. Este efecto de posición, conocido con el nombre de PEV (*Position-Effect Variegation*), fue descubierto en el estudio de las inversiones *Bm2* (Belgovsky y Muller 1938) y *w[m4]* (Sitko 1938) de *Drosophila*. El efecto de posición por variegación causa el silenciamiento de un gen por su yuxtaposición con la heterocromatina (Spofford 1976; Eissenberg 1989; Henikoff 1990; Weiler y Wakimoto 1995). Tolchkov *et al.* (1984), describieron que la inversión pericéntrica *In(1LR)pn2a* en el cromosoma X de *D. melanogaster* supone la variegación de algunos genes vitales. Mas tarde, Tolchkov *et al.* (1997) describieron una segunda inversión polimórfica (*In(1)r30*) que separa estos genes de la heterocromatina.

1.3.4.2. Mecanismos indirectos derivados de la supresión de la recombinación

La hipótesis de la coadaptación explica que, debido a la reducción de la recombinación en los heterocariotipos, las inversiones pueden mantener combinaciones alélicas favorables (Dobzhansky 1970; Charlesworth 1974) y ser seleccionadas positivamente. La coadaptación fue demostrada en inversiones naturales polimórficas de *D. melanogaster* (Hartl 1977). La inversión es sólo portadora, su aumento de frecuencia depende de que los alelos capturados sean o no favorables. Sin embargo, en las regiones centrales de las inversiones no existen estas fuertes restricciones donde el apareamiento homólogo permite el intercambio de genes por

entrecruzamiento y conversión génica (Hasson y Eanes 1996; Andolfatto *et al.* 2001). No obstante, a pesar del potencial para el intercambio de genes, muchos marcadores genéticos ubicados dentro o cerca de inversiones muestran asociaciones significativas con la propia inversión o entre ellos (Prakash y Lewontin 1968; Stalker 1976; Weeks *et al.* 2002; Munté *et al.* 2005). Estas asociaciones sugieren que es la selección quien mantiene unidos las combinaciones favorables de los alelos. Schaeffer *et al.* (2003) demostraron que en *D. pseudoobscura* se mantienen altos niveles de disequilibrio de ligamiento dentro de las regiones invertidas y los *loci* adyacentes. Incluso si la inversión captura alelos deletéreos recesivos como en el caso de *D. persimilis* (Coyne *et al.* 1992), puede generar sobredominancia asociativa que estabiliza la frecuencia de la inversión. En *D. melanogaster*, la inversión *(3R)Payne* aumenta de frecuencia cuando disminuye la latitud y ha sido supuestamente asociada con la variación en el tamaño y la resistencia térmica, rasgos que también varían clinalmente. Kennington *et al.* (2006) observaron en *(3R)Payne* una selección epistática fuerte dentro de los genes de la inversión siendo éstos el objetivo de la selección. La inversión *In(4)a* (Evans *et al.* 2007) se ha asociado con la distribución del cromosoma Neo-X de *D. americana* (fusión entre el cromosoma X y 4). Esta inversión también se distribuye en clinas latitudinales y se ha descrito efecto de coadaptación con un gen cercano al punto de rotura proximal. La inversión *In(2L)t* en *D. melanogaster* (Lemeunier y Aulard 1992; Andolfatto *et al.* 1999) y en *D. subobscura* (Balanyà *et al.* 2003) se observa distribuida en clinas geográficas paralelas aun en continentes distintos indicando adaptación local.

Alternativamente, en el mecanismo de adaptación local (Kirkpatrick y Barton 2006) los *loci* están individualmente adaptados al medio sin necesidad de epistasia ni coadaptación entre ellos. La adaptación local opera bajo condiciones demográficas generales.

1.3.5. Grupo *repleta*

El grupo *repleta* está incluido en el subgénero *Drosophila*, divergido del subgénero *Sophophora* hace 40-62 millones de años (Powell 1997). Incluye más de 100 especies repartidas en 6 subgrupos; *mulleri*, *hydei*, *mercatorum*, *repleta*, *fascicola* e *inca* (Wasserman 1982; Rafael y Arcos 1989; Wasserman 1992). Se han descrito un total de 296 inversiones en el grupo (177 fijadas y 119 polimórficas). El subgrupo *mulleri* acumula 214 (129 intraespecíficas) de las inversiones descritas en el grupo. Existe una evidente acumulación de inversiones en el cromosoma 2 de las especies del grupo puesto que 208 de las 296 inversiones en el grupo (70%), estarían localizadas en dicho cromosoma, que representa sólo el 23% de la eucromatina (Wasserman 1992). Este cromosoma es el más dinámico y, por tanto, el más informativo filogenéticamente.

Las observaciones y descripciones citológicas de Wasserman (1992) sugieren además una abundante coincidencia de muchos puntos de rotura. Sólo pudieron ser descritos 323 puntos de rotura en el cromosoma 2 de los 416 esperados (208 inversiones x dos puntos de rotura por inversión). La coincidencia de puntos de rotura se presenta también dentro de los subgrupos o incluso dentro de los complejos, lo que indica que no se debe simplemente a la saturación de un número finito de bandas distinguibles en el cromosoma 2. Aunque la observación citológica tiene una resolución limitada, estos datos sugieren que existe una coincidencia real entre los puntos de rotura de distintas inversiones. No existe información a nivel molecular sobre la naturaleza y origen de estas coincidencias observadas citológicamente.

1.3.5.1. *Drosophila buzzatii*

Nuestro grupo de investigación está centrado en el estudio de la variabilidad genética estructural, así como su origen y consecuencias evolutivas. El grupo ha tratado de entender la generación de las inversiones caracterizando sus puntos de rotura en *Drosophila buzzatii* como organismo y modelo de estudio.

D. buzzatii tiene una distribución subcosmopolita. Originalmente de Sudamérica, se ha extendido por el sur de Europa, el norte de África y Australia (Fondevila *et al.* 1981). El cariotipo de la especie *D. buzzatii*, está formado por cinco pares de cromosomas telocéntricos y un par de cromosomas puntiformes. El mapa citológico del patrón de bandas de los cromosomas politénicos de *D. buzzatii* se describe en Ruiz y Fontdevila (1981), Wasserman (1982) y Ruiz y Wasserman (1993). *D. buzzatii*, descrita por primera vez por Patterson y Wheeler (1942), pertenece al complejo *buzzatii* del subgrupo *mulleri*, incluido en el grupo *repleta*. El complejo *buzzatii* está compuesto por un conjunto monofilético de 14 especies cactófilas estrechamente relacionadas (Ruiz y Wasserman 1993; Ruiz *et al.* 1997; Tidon-Sklorz y Sene 2001) clasificadas en tres *clusters*: *stalker* (*D. richardsoni* y *D. stalker*), *martensis* (*D. martensis*, *D. uniseta*, *D. venezolana* y *D. starmeri*) y *buzzatii* (*D. buzzatii*, *D. koepferae*, *D. borborema*, *D. serido*, *D. seriema*, *D. antonietae* y *D. gouveae*).

En el complejo *buzzatii* se han descrito 86 inversiones paracéntricas (Ruiz y Wasserman 1993), siendo el cromosoma 2, con 61 inversiones el más dinámico, como en todo el grupo *repleta* (Wasserman 1992). También se ha observado una distribución claramente no aleatoria de los puntos de rotura en las inversiones polimórficas y fijadas (Cáceres *et al.* 1997).

Desde la ordenación ancestral del grupo (Primitivo I) se han fijado en *D. buzzatii* 4 inversiones paracéntricas y se han descrito 8 inversiones polimórficas así como un cierto número de reordenaciones endémicas raras (Ruiz y Wasserman 1993). La mayoría de estas

inversiones también se encuentran en el cromosoma 2. En nuestro grupo se han caracterizado los puntos de rotura de las tres inversiones polimórficas más frecuentes: $2j$ (Cáceres *et al.* 1999), $2q^7$ (Casals *et al.* 2003) y $2z^3$ (Delprat *et al.* 2009). Todas ellas pertenecen al cromosoma 2 (elemento E de Muller) y sus análisis concluyeron que el origen de dichas reorganizaciones es la recombinación ectópica entre copias del elemento de clase II *Galileo*. Además, los puntos de rotura se han convertido en puntos calientes (*hotspots*) para la inserción de otros elementos transponibles (Cáceres *et al.* 2001; Casals *et al.* 2003).

Por otro lado, sólo una de las inversiones fijadas ha sido estudiada molecularmente en *D. buzzatii* (Prazeres *et al.* 2009). A diferencia de las anteriores, la inversión $5g$ se localiza en el cromosoma 5 (elemento C de Muller) y, en este caso, la recombinación ectópica no fue el origen de esta inversión. Las tres inversiones polimórficas no tienen más de 1 millón de años (ma) de antigüedad, mientras que la fijación de la inversión $5g$ se ha estimado en unos 4 ma (Figura 1.3). En este trabajo, nos hemos propuesto completar el análisis molecular de los puntos de rotura de las tres inversiones fijadas de *D. buzzatii* desde su divergencia del ancestro del grupo (Ruiz y Wasserman 1993): $2m$, $2n$ y $2z^7$.

1.3.5.2. *Drosophila mojavensis*

En la filogenia de la Figura 1.3 también se muestra *D. mojavensis*. Esta especie pertenece al complejo *mulleri*, incluido en el subgrupo *mulleri* del grupo *repleta*. El cromosoma 2 de *D. mojavensis* difiere en 7 inversiones del cromosoma 2 del ancestro, Primitivo I, y en 10 inversiones del de *D. buzzatii* (González *et al.* 2007). *D. mojavensis* carece de las inversiones $2m$, $2n$ y $2z^7$ por lo que representa la ordenación ancestral (estándar) en lo que se refiere a las regiones de los puntos de rotura de estas inversiones. *D. mojavensis* es la especie secuenciada más cercana a *D. buzzatii* (Drosophila 12 Genomes Consortium 2007),

por lo que su genoma se ha tomado como referencia para este trabajo. Esta información nos ha permitido cartografiar con precisión los puntos de rotura de las inversiones $2m$, $2n$ y $2z^7$ en el cromosoma ancestral (no invertido para estas inversiones). El tiempo de divergencia entre *D. mojavensis* y *D. buzzatii* se ha estimado entre 10 ma (Tamura *et al.* 2004) y 15 ma (Russo *et al.* 1995).

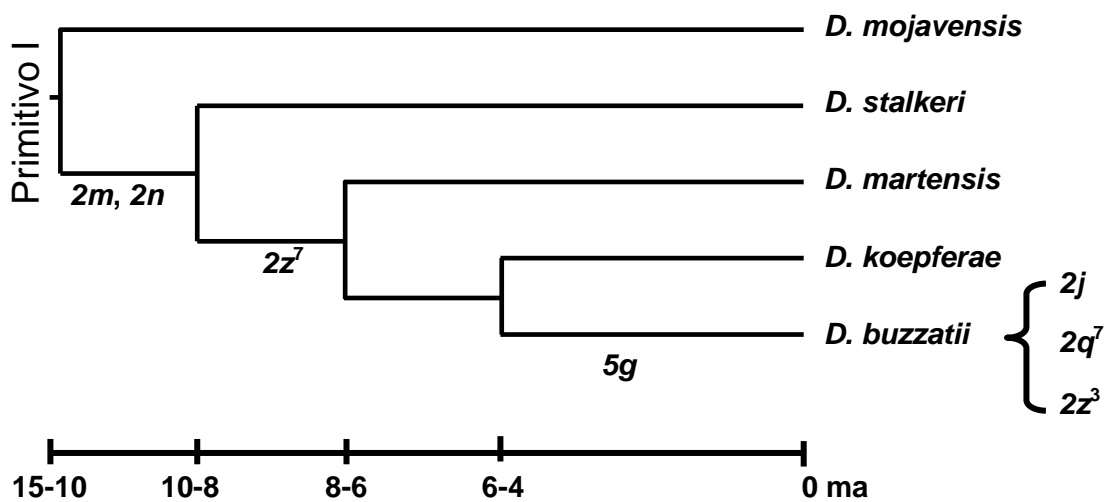


Figura 1.3: Árbol filogenético del complejo *buzzatii* desde el Primitivo I. La especie secuenciada más cercana al complejo es *D. mojavensis*. Las inversiones de este trabajo ($2m$, $2n$ y $2z^7$) y las analizadas en trabajos anteriores $5g$ (fijada), $2j$, $2q^7$ y $2z^3$ (polimórficas) se muestran en la rama correspondiente. Los tiempos de divergencia consideran las estimas de Russo *et al.* (1995) y Tamura *et al.* (2004) en millones de años (ma).

1.3.5.3. Inversiones $2m$, $2n$ y $2z^7$

Las inversiones $2m$, $2n$ y $2z^7$ están fijadas en *D. buzzatii* y ocurrieron en el cromosoma 2 (elemento E de Muller, Schaeffer *et al.* 2008) del Primitivo I ancestro del grupo repleta

(González *et al.* 2007). Las inversiones $2m$ y $2n$ son comunes en todo el complejo *buzzatii*, mientras que la inversión $2z^7$ sólo se encuentra en los *clusters buzzatii* y *martensis* (Figura 1.3). El origen de las inversiones fijadas $2m$ y $2n$ corresponde por lo tanto al tiempo de divergencia del complejo *buzzatii* del resto del grupo (~8-10 ma). La inversión $2z^7$ se fijó posteriormente a la divergencia del *cluster stalker* y anterior a la divergencia del *cluster martensis* (~6-8 ma).

El estudio citológico de los cromosomas de las glándulas salivales y los análisis de los mapas comparativos indican que las inversiones $2m$ y $2n$ están dispuestas en tándem y comparten un punto de rotura (Ruiz y Wasserman 1993, González *et al.* 2007). Por lo menos, bajo un punto de vista citológico existe coincidencia entre el punto de rotura proximal de la inversión $2m$ y distal de la inversión $2n$.

1.3.5.4. Genoteca y mapa físico de clones BAC

Para seguir estudiando la dinámica evolutiva del genoma de *D. buzzatii*, se construyó una genoteca de clones BAC usando el vector *pTARBAC2.1*. La genoteca se compone de 18.353 clones con un tamaño medio del inserto de 152 Kb y una representación esperada de aproximadamente 18x (González *et al.* 2005). Además, se elaboró un mapa físico del genoma de *D. buzzatii* que muestra la localización cromosómica de muchos de los clones BAC. Para construir el mapa, una muestra representativa de 9.555 clones fueron digeridos con enzimas de restricción y, mediante la comparación de sus patrones de *fingerprint* en geles de agarosa, se ensamblaron en *contigs* (estructura compuesta de distintos clones solapantes) usando el programa FingerPrint Contigs (FPC). Después de la revisión manual, se contabilizaron 345 *contigs* (una media de 26 clones por *contig*) y 670 *singletons* (clones aislados no incluidos en ningún *contig*). Para la construcción del mapa físico, se hibridaron *in situ* 427 clones

representativos de los *contigs* en preparaciones de cromosomas politénicos de *D. buzzatii*. El mapa físico se compone de 181 *contigs* anclados (que contienen 7.788 clones).

La genoteca de clones BAC y el mapa físico del genoma de *Drosophila buzzatii* permitieron llevar a cabo una comparación detallada de los cromosomas X y 2 entre *D. buzzatii* y *D. repleta*. La comparación de los bloques sinténicos conservados y las regiones de los puntos de rotura entre ambas especies se realizó con la hibridación *in situ* heteróloga de clones BAC de *D. buzzatii* en los cromosomas de *D. repleta*. Un clon BAC de la genoteca que contiene un punto de rotura híbrida con una única señal en el genoma de *D. buzzatii*. Cuando este clon es hibridado en el genoma de *D. repleta* se producen dos señales (Figura 1.4). En el caso del cromosoma 2, se observaron 17 segmentos sinténicos conservados (Figura 1.5). El programa GRIMM (Pevzner y Tesler 2003a) permitió inferir un mínimo de 9 inversiones fijadas y dos reutilizaciones entre las dos especies. Una de las reutilizaciones acontecería entre las inversiones $2m$ y $2n$. Seis inversiones se fijaron en el linaje que va de la especie ancestral del grupo, Primitivo I, a *D. repleta* mientras que tres inversiones se fijaron en el linaje que va de Primitivo I a *D. buzzatii* ($2m$, $2n$ y $2z^7$). Como *D. mojavensis*, *D. repleta* carece de las inversiones $2m$, $2n$ y $2z^7$ por lo que ambas representan la ordenación ancestral (estándar) en lo que a estas inversiones se refiere. Las regiones de los puntos de rotura de la inversión $2z^7$ en *D. buzzatii* están contenidas en los clones BAC 26D17 y 22N23 (distal y proximal, respectivamente). El clon 22N23 es solapante con el clon BAC 40C11, que ha sido totalmente secuenciado y anotado (Negre *et al.* 2005). El punto de rotura distal de la inversión $2m$ de *D. buzzatii* está contenido en el clon BAC 14B19. El punto de rotura proximal de la inversión $2m$ y distal de la inversión $2m$ está contenido en el clon BAC 22B03. Finalmente, el punto de rotura proximal de la inversión $2n$ está contenido en el clon BAC 16H04 (véase Figura 1.5).

La observación citológica nos permite descartar solapamiento o independencia entre ambas inversiones que impliquen cierta cantidad de material genético. Los clones BAC 1N19,

14B19 y 16H04 (González *et al.* 2007) producen dos señales de hibridación en el genoma no invertido de *D. repleta*. Si existiera solapamiento o imbricación, alguno de los tres clones hubiera producido tres señales de hibridación. Por tanto, la región que separa el punto de rotura proximal $2m$ del punto de rotura distal $2n$, no puede ser mayor que el tamaño de una banda del patrón de bandas de los cromosomas politénicos (~100 Kb). La información citológica no nos permite distinguir casos de solapamiento o independencia de una porción pequeña de material genético. Por otro lado, la información citológica tampoco puede resolver la cronología entre ambas inversiones. La cuestión de la cronología tampoco puede ser resuelta con la filogenia puesto que no hay especies que sólo tengan una de las dos inversiones.

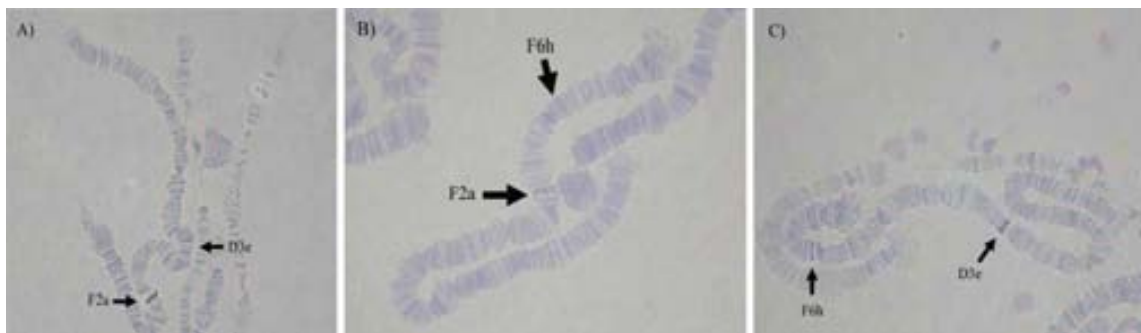


Figura 1.4: Resultados de la hibridación de clones BAC de *D. buzzatii* en preparaciones cromosómicas de *D. repleta* que contienen los puntos de rotura de las inversiones fijadas entre ambas especies. En la Figura se observan las hibridaciones heterólogas de los clones BAC que contienen los puntos de rotura de las inversiones $2m$ y $2n$: A) BAC 1N19; B) BAC 16H04; C) BAC 20019. Se indica la posición citológica de las señales de hibridación. Tomado de González *et al.* (2007).

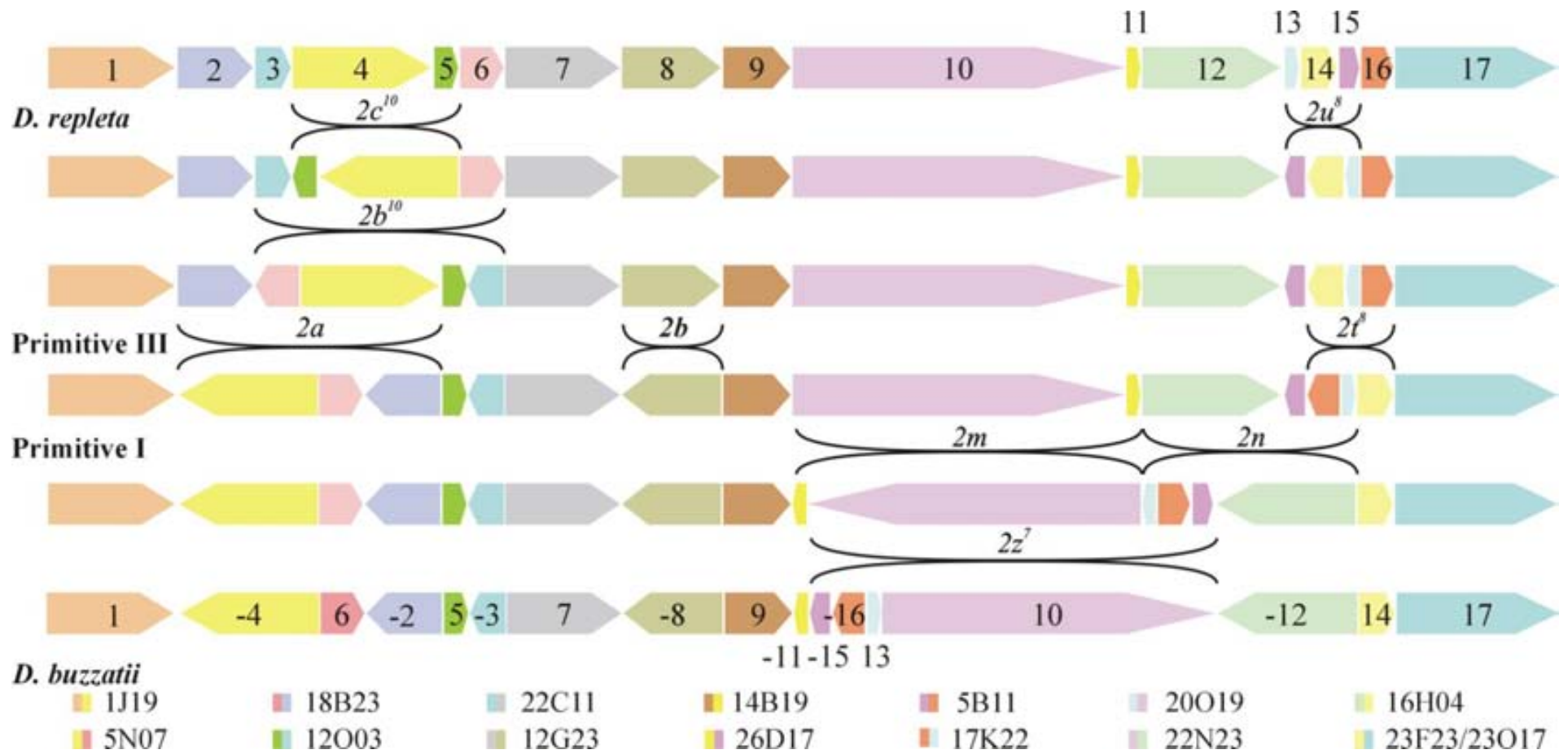


Figura 1.5: Vista esquemática de los bloques sinténicos del cromosoma 2 entre los genomas de *D. buzzatii* y *D. repleta*. Primitivo I representa la ordenación ancestral del grupo. Primitivo III representa la ordenación ancestral de los subgrupos *repleta* y *mercatorum*. Los clones BAC que contienen los puntos de rotura entre los bloques se muestran en la parte inferior. Tomado de González *et al.* (2007).

1.4. Objetivos de este trabajo

El estudio de los mecanismos de generación y las consecuencias funcionales de las reordenaciones cromosómicas nos proporciona información sobre la base genómica de la evolución, la adaptación y la adquisición de nuevas funciones genéticas. Este trabajo trata de contribuir a esclarecer el origen y la significación evolutiva de las inversiones cromosómicas así como las causas de la reutilización de los puntos de rotura.

En nuestro grupo de trabajo se han caracterizado y analizado las inversiones polimórficas $2z^3$, $2j$ y $2q^7$ y la inversión fijada $5g$ en *D. buzzatii* (véase Introducción). Con el estudio de las inversiones de este trabajo, $2m$, $2n$ y $2z^7$, completamos el análisis de las inversiones fijadas en *D. buzzatii* desde su divergencia del Primitivo I del grupo. La completa cobertura de las inversiones de *D. buzzatii*, nos permitirá analizar globalmente las reorganizaciones acontecidas en su genoma en distintos momentos de su evolución. Sólo la inversión fijada en el cromosoma 5 no se ha generado por recombinación ectópica, por lo que podría existir algún motivo diferencial o relación entre el mecanismo y el éxito de la fijación de una inversión. Por otro lado, la cobertura temporal desde la divergencia del grupo (12,5 ma) hasta las inversiones más recientes y polimórficas (~1 ma), nos debe permitir discutir sobre si existe un mecanismo preferencial en los distintos cromosomas o en función de la edad de inversiones. La presencia y frecuencia de *Galileo* (TE causante de las tres inversiones polimórficas) podría intervenir directamente en la frecuencia de las inversiones y su mecanismo de origen.

Finalmente, el estudio de las inversiones $2m$ y $2n$ debe aportar luz a la problemática de la reutilización de los puntos de rotura. El análisis molecular de los puntos de rotura de las inversiones $2m$ y $2n$ nos permitirá contrastar el grado real de la coincidencia observada a nivel citológico.

Este trabajo tiene los siguientes objetivos principales:

1. Localizar y cartografiar las regiones de los puntos de rotura de las inversiones $2m$, $2n$ y $2z$ ⁷ en el genoma de *D. mojavensis* (genoma no invertido) que representa la ordenación ancestral.

2. Aislar los puntos de rotura de estas inversiones en el genoma de *D. buzzatii* para analizar el contenido molecular de estos y, de este modo, conocer molecularmente el contenido de los puntos de rotura que nos permitirá:

(i) Determinar si los puntos de rotura han alterado de alguna manera los genes presentes en las correspondientes regiones interrumpiendo alguna unidad de transcripción o de regulación, así como su impacto en el genoma y posible relación con el motivo de la fijación de estas inversiones.

(ii) Obtener información sobre el mecanismo por el que se generaron estas inversiones y contrastar la posible implicación de los elementos transponibles en su generación.

(iii) Desvelar la relación histórica y cronológica entre ambas inversiones.

3. El estudio y análisis molecular del punto de rotura compartido entre las inversiones $2m$ y $2n$ nos permitirá conocer el grado real de la coincidencia observada citológicamente y aportar nuevos datos al estudio de las causas de la reutilización de los puntos de rotura y su distribución no aleatoria.

4. La comparación del análisis molecular de estas inversiones con las previamente descritas, nos permitirá conocer:

(i) Si existe algún mecanismo preferencial específico de cromosoma o especie.

(ii) Si existe algún mecanismo preferencial en función de la edad de la inversión.

(iii) Si existe alguna relación entre el éxito de la inversión (fijación) y el mecanismo que las genera.

2. MATERIALES Y MÉTODOS

2.1. Cepas de Drosophila

La cepa st-1 de *D. buzzatii* ha sido utilizada en todos los análisis realizados con esta especie. Esta cepa es también la que se usó para la construcción de la genoteca de clones BAC y el mapa físico del genoma de *D. buzzatii* (González *et al.* 2005). Otras especies se han usado a lo largo de este trabajo: la cepa MA-4 de *D. martensis* (Guaca, Venezuela), la cepa KO-2 de *D. koepferae* (Sierra de San Luis, Argentina) y la cepa 15081-1352.22 de *D. mojavensis* (Isla Catalina, California, EEUU). Esta última cepa es la misma que se utilizó en la secuenciación del genoma completo de *D. mojavensis* (Drosophila 12 Genomes Consortium 2007) y que hemos utilizado como referencia en los análisis comparativos. Los genomas secuenciados de *D. melanogaster* y *D. virilis* (Drosophila 12 Genomes Consortium 2007) también se han tomado como grupo externo a la divergencia entre *D. buzzatii* y *D. mojavensis*, permitiendo localizar las regiones más conservadas en la secuencia de ésta última.

2.2. Protocolos básicos

2.2.1. Diseño de cebadores

Los cebadores de este trabajo han sido diseñados tomando como referencia el genoma de la especie secuenciada más cercana. En el caso de *D. buzzatii*, se ha usado *D. mojavensis*. Para el diseño de cebadores en las especies *D. koepferae* y *D. martensis* se han usado las propias secuencias de *D. buzzatii*. En las reacciones de PCR heterólogas, los cebadores se han diseñado en las regiones más conservadas de los genomas más cercanos. El diseño se ha realizado con el programa PRIMER de MS-DOS. Generalmente se han seleccionado secuencias de 20 pares de bases con un mínimo del 45% del contenido en guanina-citosina y

una temperatura de hibridación entre 60 y 68°C. Se ha evitado en la medida de lo posible, que formaran estructuras secundarias por palindromías, dímeros de más de 4 pb entre los cebadores directos y reversos y sin cadenas de tres nucleótidos iguales. Siempre que el cebador se ha diseñado en una región codificante, se ha evitado hacer coincidir el último nucleótido 3' con la tercera base de un codón (más susceptible a no ser homóloga por la degeneración del código). La definición de los tripletes de la secuencia se ha determinado mediante el alineamiento con el programa MUSCLE (Edgar 2004) disponible en (<http://www.ebi.ac.uk/Tools/muscle/index.html>) usando las traducciones conceptuales proteicas. Se ha priorizado que el extremo 3' coincida en estos casos con una primera base de codón. Los cebadores utilizados en este trabajo se muestran en las Tablas 2.1, 2.2, 2.3 y 2.4.

2.2.2. PCR

Las reacciones de PCR se llevan a cabo en 50 µL de volumen final con la siguiente composición: 50-200 ng de DNA, 40 pmoles de cada cebador, 200 µM de dNTPs, 1,5 mM Mg₂Cl y 1 unidad de Taq Polimerasa (*Platinum[®] Taq DNA Polymerase High Fidelity* de InvitrogenTM). Los tiempos determinados en el termociclador varían considerablemente en función de la longitud que se quiere amplificar y en función de si los cebadores se han diseñado en la misma especie que amplifican o no. Cuando la especie molde no coincide con la especie usada para el diseño de los cebadores, la temperatura de hibridación de la PCR se estableció 5 grados por debajo de los parámetros indicados para dar cuenta de los desajustes de nucleótidos.

1. 3 minutos a 94°C.
2. 30 segundos a 94°C.
3. 30 segundos a la temperatura de hibridación propia de los cebadores (-5°C si los se han

diseñado en otra especie a la que se está usando de molde).

4. X minutos a 68°C (a razón de un minuto por 1.000 nucleótidos que se quiera amplificar).
5. Repetir 2, 3, 4 de 25 a 30 veces (en función de la cantidad de producto que se necesite para los siguientes pasos).
6. 10 minutos a 68°C.
8. Fin del programa. Guardar a 4°C.

2.2.3. Medios de cultivo

A los medios de cultivo (líquido y sólidos) se les añade el correspondiente antibiótico en función de la resistencia del plásmido para evitar crecimientos inespecíficos (Chloramfenicol para el crecimiento de clones BAC-*pTARBAC2.1*- o Ampicilina para el crecimiento de células transformadas con los plásmidos *pBSK*-clonaciones de digestiones- o *pGEM*-clonaciones de productos de PCR-). Se han usado dos tipos de cultivo líquido: El medio TB (*Terrific Broth*), permite mayor crecimiento al ser más rico en nutrientes se ha usado para el crecimiento de células de la genoteca de clones BAC (sólo una copia por célula). El medio LB (*Lysogeny broth*) ofrece un crecimiento más pausado y se ha usado para el crecimiento de células competentes transformadas con vectores plasmídicos (~30 copias por célula).

El medio LB-agar en placas se ha usado para sembrar las transformaciones. Al añadirle X-gal permite la diferenciación de células transformadas. Se disuelve LB-agar en polvo en agua destilada a razón de 35 gramos por litro y se esteriliza en el autoclave. Antes de añadir X-gal (20 mg/mL) y el antibiótico (25 mg/mL) debe enfriarse hasta 50°C al ser sustancias termolábiles. Se necesitan aproximadamente ~25 µL por placa de medio. Una vez el medio se

solidifica, las placas se guardan a 4°C en posición invertida. Deben sacarse de la nevera 1 hora antes de usarlas con el fin de evitar la condensación de la humedad.

Tabla 2.1: Secuencia de los cebadores usados en el paseo cromosómico y para cerrar los huecos en el ensamblaje de los BAC secuenciados.

PCR	Directo (5' → 3')	Reverso (5' → 3')
A5	GCCGTACGCGTTCTCTATAA	TGCCGAAC TTGTTAGACGTA
A4	CCGATTGACTAACGTTAAGA	GTCTCACGTTTCGCATACCA
A3	TGGTATGCGAAACGTGAGAC	TTGACTGGAGCAGCATGTAC
A2	GTTGTTGTCAGGCAGTTGCA	TGGTCCAAGTAAAACACAC
A1	ATATTGCAGGTTTCAGACAC	TTGTCATTATCGCCGATTAC
B1	GCGAGGATCCAGATTATGAA	GCCATGCAGACCATCATAAC
B2	GCTCGAGCAATTCATCTACA	TTATCACGAACCAGAGCCAT
B3	CCAGCACTTCGATACCATCA	GCGATACTCCACAGCTATTG
B4	GTGGAGTATCTGCTAGTTGA	TATAGAGAACGCGTACGGCA
C2	GCTTGCATGCTAACGAGTTG	AAAAAGAGTTCCTCGAGGGT
C1	TATGACAACGCGGAAATTGT	CAGGTGGAATTCGTGGACAA
D1	GGCATGGCCATCTACGATAT	GGATCCGGGAAGTATTCCTC
D2	GACAAGGCCAGAAGCATAAT	GCTCTAATTAGGCGCACATA
E3	TATTGTGCTAATCTGGCAAG	TTACGTTTCATCGCTAACAGA
E2	TACTATTACGTTGGCTGCTA	GAGGAGAGGTCATCAGCTGA
E1	CTTACTCAATCTCATGTCCA	TTGAAGTAGGTGTGCTCGAA
F1	TACGACACACATCGGAACTC	GCGCCAATACGAGTAGAGTA
F2	GCTGATGAAGTGAAAGTCAA	GATAGACACGCCTTGTAAGT
F3	AAGGATAAACGTTGCCGAAG	GACTTTTGGTTGGCTTGTC
F4	CGAATATGTCGTTCTTGCGA	TATGGAACCGTGCTCGACTA
PCR	Ensamblaje	
6CL-7CR	CTACGCAGATAAGCAGGCTT	AACTGTCAGCAGCAACGTGT
4CL-3CR	TGTCGACTATAGTTAAGCGT	GGCAGTAGTCGTCGCATTAT
3CL-5CR	GTGAGGCAATGCGTAACATT	CTTCTTGCTACGCATAATCT

Tabla 2.2: Secuencias de los cebadores para amplificar los puntos de rotura de las inversiones $2m$ y $2n$ (AC, BE y DF) en *D. buzzatii*.

Región	Cebador	Secuencia (5' → 3')	Secuencia y especie donde fueron diseñados
AC	A	GCAACAACAAGTCACAATGA	Intergénico (<i>msi-CG4673</i>) (<i>D. mojavensis</i>)
	C	TGAATTAGAAACCACCGTCA	<i>scrib</i> (<i>D. mojavensis</i>)
BE	B	GGCATACGTCAGCTGATGAC	<i>Ssadh</i> (<i>D. mojavensis</i>)
	E	ATCAATATGGCAACGAGGTG	<i>CG31127</i> (<i>D. mojavensis</i>)
DF	D	TGTTCGAGCAGCACTACATA	Región conservada aguas abajo del gen <i>Or98b</i> en <i>D. mojavensis</i>
	F	AATAAGCAGCAAGTGCACAG	<i>CG8147</i> (<i>D. mojavensis</i>)

Tabla 2.3: Secuencia de los cebadores usados en la amplificación de los puntos de rotura de la inversión $2z^7$ (PQ y RS).

Región	Cebador	Secuencia (5' → 3')	Región y especie donde han sido diseñados
PR	P1	GTCATCCCAGTTAACGCAGT	<i>CG14645</i> (<i>D. mojavensis</i>)
	P2	TTTCCAAGGAGACTTAGCTC	P1 (punto de rotura PR de <i>D. buzzatii</i>)
	R1	AGCAGTGCCAGAGCCGAGAT	<i>CSN5</i> (<i>D. mojavensis</i>)
	R2	GGAGATCAACGAGAATTACA	R1 (punto de rotura PR de <i>D. buzzatii</i>)
QS	Q	TCTTACCGAAGGTCATTCTG	<i>CG14290</i> (BAC 40C11 de <i>D. buzzatii</i>)
	S	GTGTGCAAGGATACGGTGGA	<i>Dlh</i> (BAC 40C11 de <i>D. buzzatii</i>)

Tabla 2.4: Secuencia de los cebadores diseñados en la región del gen *CG4673* de *D. buzzatii* (BE). *cebadores del protocolo del kit RLM-RACE de Ambion (AM1700).

Cebador	Secuencia (5'-3')
firstL	GAAGAGACGATGTTCGCACAA
lastR	CGAAACTAAAGATTCAATCA
WschR	TACCTGCCGATGCTTGAGAC
SsadhL	GACGACTATGTGGATATCAA
IR5'L	CATGAAACCGCGCATTTAAC
IR5'R	TATTAACAAACCTCTTCAAC
0R	ATTTTTTAGCAATGTCTGA
0L	TAGATATACCATGTCGCCGT
2R	TTACAGCGTACAAAATGTGT
2L	GAGTTCTGCCACCAGCTCA
3R	GCACGAGTACCACAGCCAAA
3L	CCTGATTACTTCAATGTATT
7R	GCCGCTGCAGCCGCACATTC
7L	GATAGATGGCTGATCAACAC
3'UTR-R	TCCAAGTGCAGCTTGACAC
3'UTR-L	CATGCTTATTGAATCTTTAT
5R	CTGACCGCTCAGGAGTGCAT
6L	GGTGCATCCTTGGTAGGTAT
8R	GACGACCAGCTATCCCAGTG
8Rint	TGGACCTGCAACCACTGCAC
1L	GCAGACTGCACGCGTATTAG
1Lint	GAAGAGACGATGTTCGCACAA
7Rint	ACGAAAGACGCCAGCCATTG
2Lint	CTTTAGCTCCGTGGCCAAGC
5'RACE-Adapter*	GCUGAUGGCGAUGAAUGAACACUGCG... ...UUUGCUGGCUUUGAUGAAA
5'RACE-Outer*	GCTGATGGCGATGAATGAACACTG
5'RACE-Inner*	CGCGGATCCGAACACTGCGTTTGCTGGCTTTGATG
3'RACE-Adapter*	GCGAGCACAGAATTAATACGACTCACTATAGGT12VN

3'RACE-Outer*	GCGAGCACAGAATTAATACGACT
3'RACE-Inner*	CGCGGATCCGAATTAATACGACTCACTATAGG
H1	ATGTCGAAGATTGGTATTAATGG
H2	GTTCGACACGACCTTCATGT

2.2.4. Electroforesis

Los geles se han realizado entre el 0,7% y el 1% de concentración de agarosa en función del tamaño de las bandas. Habitualmente se han realizado a una intensidad de voltaje de 65V, controlando la migración de las bandas mediante los colorantes azul de bromo fenol y cianol-xileno añadidos al tampón de carga. La tinción del DNA separado por peso molecular en los geles de agarosa se realiza sumergiendo los geles durante 20 minutos en una solución de bromuro de etidio en H₂O destilada (0,5 µg/mL). El tamaño de los fragmentos del gel se estima por comparación con un DNA digerido de tamaño conocido que se añade en todos los geles como marcador de peso molecular. Para recuperar fragmentos específicos del gel se ha utilizado el kit *QUIAquick*[®] de QIAGEN.

1. Recortar de un gel de agarosa el trozo de menor tamaño que contenga el DNA de peso molecular que queremos recuperar.
2. Añadir tres volúmenes de tampón QG del kit a un volumen de gel (100 mg ~100 µL).
3. Incubar a 50°C durante 10 minutos.
4. Colocar la mezcla en una columna del kit y centrifugar durante 1 minuto para capturar el DNA en el filtro.
5. Vaciar la columna y limpiar el filtro con 0,5 mL de tampón QG del kit. Centrifugar 1 minuto.

6. Vaciar de nuevo la columna y añadir 0,75 mL de tampón PE del kit. Dejar reposar de 2 a 5 minutos y centrifugar 1 minuto.
7. Para eluir el DNA del filtro, añadir 30 μ L de H₂O destilada. Dejar reposar 1 minuto y centrifugar 1 minuto.

2.2.5. Digestión con enzimas de restricción

Preparación de la siguiente mezcla de reacción en un tubo *Eppendorf* estéril:

X μ L de DNA (X en función de la concentración de la muestra).

1,5 μ L de tampón concentrado 10x.

0,5 μ L enzima de restricción.

H₂O destilada hasta llegar a 15 μ L de mezcla.

Antes de incubar un mínimo de 3 horas a 37°C, se realiza un breve centrifugado para bajar todos los productos al fondo del tubo *Eppendorf*.

2.2.6. Clonación

El proceso de clonación consiste en la amplificación de un fragmento de DNA insertado en un plásmido mediante su replicación en células competentes transformadas (cepa DH5 α de *E. coli*). El plásmido contiene un sitio de clonación múltiple en el interior del gen *lacZ* y otro de resistencia a un antibiótico. El gen *lacZ* permite la selección de colonias recombinantes por α -complementación (Sambrook *et al.* 1989). Este sistema de selección consiste en la complementación del operón *lacZ* truncado de las células competentes mediante el segmento proximal proporcionado por el vector. Se siembra el producto transformado en placas de LB+antibiótico+X-gal. El antibiótico sólo permite el crecimiento de colonias

transformadas dado que la resistencia a dicho antibiótico se la proporciona el plásmido. El X-gal es un análogo de la lactosa que permite diferenciar las colonias transformadas con un plásmido sin inserto (colonias azules) de las que hayan incorporado plásmido con inserto (colonias blancas), debido a que las células tendrán inactivado el gen *lacZ* y no podrán metabolizar el X-gal. Los clones se pueden conservar a -70°C añadiendo a 750 μL , 250 μL de glicerol.

a) Obtención de células competentes

1. Se prepara un cultivo nocturno de células de la cepa DH5 α de *E. coli* en medio LB.
2. Inocular 20 mL de medio de cultivo LB con 0,2 mL del cultivo nocturno. Se deja en agitación a 37°C hasta que la densidad óptica a longitud de onda 550 nm sea igual a 0,4 (aproximadamente corresponde a 3 horas de crecimiento).
3. Enfriar el cultivo en hielo durante 20 minutos. Repartir en dos tubos de 10 mL.
4. Centrifugar 10 minutos a 4.000 rpm a 4°C .
5. Decantar el sobrenadante. Resuspender las células en 4,5 mL de CaCl_2 50 mM frío y estéril.
6. Dejar 30-45 minutos en hielo.
7. Centrifugar 10 minutos a 4.000 rpm a 4°C . Decantar el sobrenadante.
8. Resuspender muy suavemente en 1 mL de CaCl_2 50 mM frío y estéril.
9. Dejar una hora en hielo.
10. Añadir un 15% de glicerol y hacer alícuotas de 50 a 100 μL que se guardarán a -80°C .

b) Ligación

Existen variaciones sustanciales en función del origen del inserto. De este modo, debemos tener en cuenta una serie de consideraciones en función de si hemos clonado un

fragmento de restricción o un producto de PCR. Si el inserto procede de una digestión con una enzima de restricción hemos usado el vector *pBSK*, que habremos linearizado mediante la misma enzima de restricción con el que se ha obtenido el fragmento de interés. De este modo coincidirán los extremos cohesivos entre dicho fragmento y el vector. El sitio de clonación múltiple dentro del gen *lacZ* contiene una batería de enzimas de restricción para este efecto. Si el fragmento a clonar procede de una PCR, hemos usado el vector *pGEM* (*pGEM-T*[®] y *pGEM TEasy*[®] de Promega). Esta ligación también es cohesiva. Las Taq polimerasas añaden por defecto al final de las cadenas amplificadas una adenina adicional. El vector *pGEM* ya está linearizado y con los extremos cohesivos complementarios (Timina).

1. Preparación de la siguiente mezcla de reacción en un tubo *Eppendorf* estéril:

X μ L de DNA del inserto (X en función de la concentración de la muestra).

Y μ L de DNA del vector digerido (Y en función de la concentración de la muestra).

H₂O destilada hasta llegar a 18 μ L de mezcla.

(Generalmente las cantidades de inserto y de vector se ponen en una relación 3:1).

2. Se centrifuga brevemente para juntar el contenido antes de dejar 5 minutos a 65°C.

3. Se deja 10 minutos en hielo.

4. Se añade:

2 μ L de tampón concentrado 10x.

1 μ L de ligasa.

5. Se deja toda la noche a 4°C.

a) Transformación

1. Dejar descongelar una alícuota (50-100 μ L) de células competentes en hielo.

2. Mezclar suavemente y en frío el DNA de la reacción de ligación con la alícuota de células competentes (el volumen del DNA no debe de superar el 10% de la total para evitar

problemas de ósmosis).

3. Se deja 20 minutos en hielo.
4. Se incuba 2 minutos a 42°C en el baño.
5. Se añade 0,5 mL de medio LB, previamente calentado a 37°C y se mezcla suavemente.
6. Incubar 1 hora a 37°C.
7. Extender, con una asa de Digralesky, en placas de LB-agar+Ampicilina+X-gal a razón de 0,2 mL por placa.
8. Dejar toda la noche a 37°C.
9. Las colonias blancas están transformadas con el vector que contiene el inserto.

2.2.7. Extracción de DNA plasmídico

Este proceso de extracción de DNA plasmídico de las células de un cultivo nos permite recuperar el vector con el inserto amplificado por clonación. Se ha usado tanto para recuperar insertos clonados en plásmidos *pGEM* y *pBSK* como para recuperar clones BAC de la genoteca de DNA genómico.

1. Inocular una colonia con el asa de Kohl en 3 mL de medio de cultivo TB al que se ha añadido previamente el volumen suficiente de solución de antibiótico. Dejar agitando toda la noche a 37°C.
2. Poner 1,5 mL de cultivo en un tubo *Eppendorf* centrifugar un minuto y decantar el sobrenadante. Añadir 1,5 mL en el mismo tubo *Eppendorf*, centrifugar y decantar, eliminando acuradamente el medio sobrenadante.
3. Resuspender el sedimento en 100 µL de tampón de lisis celular frío (50 mM glucosa; 10 mM EDTA a pH 8; 25 mM Tris HCl a pH 8). Poner en hielo.
4. Añadir 200 µL de solución de desnaturalización (NaOH 0,2 M; SDS 1%). Esta solución

debe estar recién preparada. Se mezcla el contenido invirtiéndolo rápidamente hasta que quede casi transparente. Se mantiene en hielo 5 minutos.

5. Añadir 150 μL de tampón de neutralización frío (11,5 mL ácido acético glaciado; 60 mL acetato potásico 5M; 28,5 mL H_2O destilada). Agitar en posición invertida hasta formar un precipitado y dejar en reposo durante 10 minutos en hielo.

6. Centrifugar durante 10 minutos y transferir el sobrenadante a otro tubo *Eppendorf*.

7. Desproteínizar con un volumen de fenol:cloroformo (1:1). Agitar vigorosamente hasta su homogeneización.

8. Centrifugar cinco minutos y transferir la fase acuosa a otro tubo *Eppendorf*.

9. Desproteínizar con un volumen de cloroformo-isoamilalcohol (24:1). Mezclar por inversión.

10. Centrifugar 5 minutos y transferir la fase acuosa a otro tubo *Eppendorf* con un volumen de isopropanol frío. Mezclar por inversión y dejar un mínimo de 15 minutos a -20°C .

11. Centrifugar 10 minutos a 4°C y eliminar el sobrenadante.

12. Añadir 500 μL de etanol al 70%. Centrifugar 5 minutos.

13. Eliminar completamente el sobrenadante.

14. Disolver en 20 μL de tampón TE a pH 7,5 y añadir 0,5 μL de solución de RNasa A (10mg/mL). Incubar 30 minutos a 37°C .

2.2.7. Extracción de DNA genómico

1. Llenar un homogeneizador con 0,15-0,2 g de moscas adultas anestesiadas.

2. Añadir 5 mL de tampón 1 (100mM NaCl; 10mM EDTA; 20mM TrisHCl a pH 7,5; 0,5% Tritón X-100) y homogeneizar con un pistón de máxima velocidad unas diez veces.

3. Pasar el producto homogeneizado por una malla de nylon a un tubo estéril de 10 mL.
4. Añadir 5 mL de tampón 1 para limpiar el pistón y la malla.
5. Centrifugar a 4.000 rpm a 4°C durante 10 minutos. Descartar el sobrenadante.
6. Resuspender el pellet de núcleos en 10 mL de tampón 2 (100mM NaCl; 10mM EDTA; 20mM TrisHCl a pH 7,5).
7. Repetir 3 veces los pasos 5 y 6.
8. Resuspender el pellet de núcleos limpios en 1,5 mL de tampón 2.
9. Añadir 0,75 mL de SDS 4% y 0,75 mL de perclorato sódico 4 M.
10. Agitar suavemente hasta formar un líquido viscoso durante 3 minutos.
11. Desproteínizar con un volumen de fenol:cloroformo. Agitar suavemente hasta que el líquido se vuelva homogéneo.
12. Centrifugar a 4.000 rpm a 4°C durante 5 minutos.
13. Transferir la fase acuosa a un tubo nuevo.
14. Repetir la desproteínización con cloroformo:alcohol isoamílico (24:1).
15. Añadir 2,5 volúmenes de etanol absoluto frío y agitar suavemente. Debe generarse un precipitado en forma de hilos (DNA).
16. Pasar el DNA a un tubo *Eppendorf*.
17. Centrifugar 10 minutos a 13.000 rpm y eliminar el sobrenadante.
18. Añadir 500 µL de etanol al 70%.
19. Centrifugar durante 5 minutos. Eliminar el sobrenadante.
20. Dejar secar el pellet y resuspender el DNA en 50-100 µL de H₂O destilada.

2.2.9. Secuenciación

Tanto los productos de PCR como los insertos de clones (en vectores *pGEM* y *pBSK*) que se secuenciaron se mandaron al servicio de secuenciación de Macrogen en Corea (http://www.macrogen.com/eng/sequencing/sequence_main.jsp). Los productos de PCR se han secuenciado con los mismos cebadores que se usaron para la amplificación. Los insertos en clones se han secuenciado desde los cebadores universales M13 incluidos en los vectores usados y que rodean los sitios de clonación múltiple.

Por otro lado, los extremos de los insertos de los clones BAC que se secuenciaron, se mandaron al servicio de GATC-biotech en Alemania (<http://www.gatc-biotech.com>). Estos extremos fueron secuenciados utilizando los cebadores universales T7 y SP6 incluidos en el vector *pTARBAC2.1* (vector usado en la genoteca).

2.3. Clones BAC

Los clones BAC usados en este trabajo, se han seleccionado de la genoteca CHORI-225 de DNA genómico de *D. buzzatii* (<http://bacpac.chori.org/>). Dicha base de datos dispone de los *fingerprints* de las digestiones de los clones BAC que han permitido estimar el tamaño de sus insertos, así como determinar el grado de solapamiento entre ellos. La hibridación *in situ* heteróloga de clones de *D. buzzatii* en preparaciones cromosómicas de *D. repleta* (González *et al.* 2005), permitió seleccionar aquellos BAC que contenían los puntos de rotura de las inversiones $2m$ y $2n$. Estos clones producen una señal única cuando se hibrida en los cromosomas invertidos de *D. buzzatii* y dos señales cuando se hibridan con cromosomas no invertidos de otras especies, como *D. repleta*.

Por otro lado, se ha usado el BAC 40C11. Dicho BAC fue previamente secuenciado y

anotado en los trabajos de Negre *et al.* (2005) y Delprat *et al.* (2009) y contiene la región del punto de rotura proximal de la inversión $2z^7$ de *D. buzzatii*.

2.4. Hibridación *in situ*

Las hibridaciones *in situ* de este trabajo se realizaron como en Montgomery *et al.* (1987) y Ranz *et al.* (1997). La técnica de hibridación *in situ* se basa en la hibridación específica de una sonda nucleotídica marcada. Las sondas fueron marcadas con biotina-16-dUTP (ROCHE) con el sistema de cebadores de hexanucleótidos al azar (*random primers*) y se hibridaron en preparaciones de cromosomas politénicos de las glándulas salivares de diversas especies de *Drosophila*. Las hibridaciones interespecíficas (heterólogas) se realizaron a 25°C, mientras que las intraespecífica (homólogas) a 37°C. Se han utilizado dos tipos de sonda, los clones BAC de la genoteca CHORI-225 de *D. buzzatii* (González *et al.* 2005) y los productos de PCR generados en el paseo cromosómico para acotar los puntos de rotura.

Además de la preparación del material biológico y el marcado de las sondas, los pasos básicos del proceso de hibridación *in situ*, descritos por Leitch (1994), consisten en: prehibridación, hibridación, posthibridación, revelado y visualización.

2.4.1. Preliminares de la técnica de hibridación *in situ*

a) Preparaciones cromosómicas

Los cromosomas politénicos de las preparaciones se han obtenido de las glándulas salivares de distintas especies de *Drosophila*. Antes de obtener preparaciones de cromosomas politénicos, es necesario tratar portaobjetos y cubres para aumentar la adherencia del material biológico a fin de evitar su pérdida durante el proceso de hibridación *in situ*.

Siliconado de cubreobjetos

1. Hervir ácido clorhídrico 1N en un vaso de precipitados.
2. Colocar 50-100 cubreobjetos y removerlos con una varilla mientras hierven durante 5 minutos.
3. Aclarar los cubreobjetos con H₂O destilada y dejar secar 20 minutos.
4. Sumergir los cubreobjetos uno a uno en etanol y dejar secar.
5. En una campana de gases, pasar los cubreobjetos uno a uno por una solución de Repel-Silane (20 g/L dimetilclorosilano en 1,1,1-triclorometano) y dejar secar dentro de la campana.
6. Lavar los cubreobjetos uno a uno con etanol y dejar secar. Guardarlos a temperatura ambiente en una caja protegidos del polvo.

Tratamiento de portaobjetos

1. Hervir ácido clorhídrico 1N en un vaso de precipitados.
2. Colocar los portaobjetos de 20 en 20 y dejar hervir durante 5 minutos.
3. Poner los portaobjetos en una cubeta con H₂O destilada. Sacarlos y aclararlos individualmente bajo un chorro de agua destilada. Opcionalmente se pueden dejar toda la noche en una cubeta con H₂O destilada.
4. Secar los portaobjetos durante una hora.
5. Poner los portaobjetos en una cubeta con una solución SSC Denhart (0,012 g polivinilpirrolidona; 0,012 g ficoll; 0,012 g albúmina de suero bovino; 90 mL 20xSSC (3 M NaCl; 0,3 M Na₃C₆H₅O₇ x 2H₂O); 600 mL H₂O destilada) en un baño a 65°C. Dejarlos en esta solución entre 2 y 3 horas.
6. Poner los portaobjetos en una cubeta con etanol:ácido acético (3:1) durante 20 minutos.
7. Sumergir los portaobjetos 2 segundos unas 10 veces en una cubeta con etanol al 95%.

8. Dejar secar y guardarlos a 4°C.

Disección de glándulas salivares

1. Colocar una gota de ácido acético al 45% (recién preparado) sobre un portaobjetos limpio.
2. Poner sobre el acético una larva de tercer estadio y diseccionarla bajo la lupa con dos pinzas: unas sujetando la larva por la parte distal y otras por la parte proximal. Estirar en línea recta de modo que las glándulas salivares sean visibles.
3. Con dos alfileres finos se debe eliminar la grasa que queda adherida a las glándulas.
4. Poner sobre un cubreobjetos siliconado, 8 μL de solución ácido láctico:H₂O destilada:ácido acético (1:2:3).
5. Transferir las glándulas a esta mezcla dónde se dejan 8-10 minutos.
6. Dejar caer con cuidado un portaobjetos tratado sobre el cubreobjetos.
7. Girar el portaobjetos y mirarlo a través del microscopio. Deben verse los núcleos y dentro de ellos los cromosomas separados por la acción del ácido láctico.
8. Picar sobre el cubreobjetos con una punta de pipeta para romper los núcleos y separar bien los cromosomas.
9. Poner un pliegue de papel absorbente sobre el cubreobjetos y presionar con el pulgar. Hay que evitar el desplazamiento del cubreobjetos para evitar que los cromosomas se rompan.
10. Colocar el portaobjetos en una carpeta plana que se dejará en posición horizontal toda la noche.
11. Sumergir un portaobjetos en nitrógeno líquido hasta que este deje de hacer burbujas.
12. Colocar el filo de una cuchilla entre una esquina del cubreobjetos y el portaobjetos para hacer saltar el cubreobjetos.

13. Colocar el portaobjetos en una cubeta con etanol al 95% un mínimo de 15 minutos.
14. Dejar secar los portaobjetos y guardarlos a 4°C.

b) Marcado de las sondas con el sistema de cebadores de hexanucleótidos al azar (*random primers*)

Las sondas a hibridar *in situ*, deben estar marcadas con biotina 16-dUTP con este sistema.

1. Desnaturalizar 500 ng de DNA de la sonda al “baño María” a 100°C durante 10 minutos.
2. Añadir la siguiente mezcla: 2 µL de hexanucleótidos; 2 µL de biotina 16-dUTP; 2 µL de dCTP, 2 µL de dGTP; 2 µL de dATP; 1 µL de enzima Kleenow y H₂O destilada hasta llegar a los 20 µL.
3. Dejar a 37°C durante toda la noche.
4. Añadir 2 µL de EDTA 0,2M pH 8.
5. Añadir 2,5 µL de CINi.
6. Añadir 70 µL de etanol absoluto frío.
7. Dejar en el refrigerador a -20°C como mínimo durante dos horas.
8. Centrifugar 15 minutos a 13.000rpm a 4°C.
9. Descartar el etanol y añadir 40 µL de etanol al 70% frío.
10. Centrifugar 5 minutos a 13.000rpm a 4°C.
11. Descartar el etanol y secar el pellet en bomba de vacío.
12. Añadir entre 25 y 50 µL de H₂O destilada.

2.4.2. Prehibridación

Durante la prehibridación se eliminan las proteínas básicas que podrían interactuar con la sonda y se deshidrata el material biológico sumergiéndolo en etanoles a diferentes concentraciones. Este último paso asegura que la sonda no se disuelva en ningún resto de solución de prehibridación.

1. Incubación de las preparaciones cromosómicas en 2xSSC durante 30 minutos en un baño a 65°C.
2. Poner las preparaciones en una cubeta con etanol al 70% durante 5 minutos.
3. Pasar las preparaciones a otra cubeta de etanol al 70% durante otros 5 minutos.
4. Pasar las preparaciones a otra cubeta con etanol al 95% durante 5 minutos.
5. Dejar secar las preparaciones cromosómicas un mínimo de 15 minutos.

2.4.3. Hibridación

Implica tanto la desnaturalización de la sonda como del material sobre el que se va a hibridar. Se han usado dos temperaturas de hibridación diferentes: 25°C cuando para las hibridaciones las preparaciones y la sonda no correspondían a la misma especie y 37°C cuando se trataba de la misma especie.

1. Poner las preparaciones en una cubeta con solución de desnaturalización (NaOH 0,07 N) durante 2 minutos.
2. Pasar las preparaciones a una cubeta con 2xSSC durante 5 minutos. Repetir dos veces.
3. Poner las preparaciones en una cubeta con etanol 70% durante 5 minutos.
4. Pasar las preparaciones a otra cubeta con etanol al 70% durante 5 minutos.
5. Pasar las preparaciones a otra cubeta con etanol al 95% durante 5 minutos.

6. Dejar secar las preparaciones durante al menos 15 minutos.
7. Desnaturalizar la sonda incubándola a 65°C durante 15 minutos.
8. Hacer un pulso para que baje el líquido evaporado y mantener la sonda en hielo.
9. Añadir 20 µL de sonda a la preparación cromosómica.
10. Poner un cubreobjetos sobre la zona donde se ha añadido la sonda. Si quedan burbujas, eliminarlas con la punta de la pipeta.
11. Colocar la preparación cromosómica en una cámara húmeda que se incubará un mínimo de 14 horas a 25°C o a 37°C.

2.4.4. Posthibridación

La temperatura, la concentración de sales de los lavados y la temperatura de hibridación en este paso determinan la astringencia con la que se está llevando a cabo la hibridación, es decir, el porcentaje aproximado de nucleótidos correctamente apareados entre la sonda y el DNA sobre el que se hibrida. La astringencia aumenta con el aumento de la temperatura o la disminución de la concentración de sales.

1. Poner las preparaciones cromosómicas en una cubeta con 2xSSC a 37°C durante 10 minutos.
2. Pasar las preparaciones cromosómicas a otra cubeta con 2xSSC a 37°C durante otros 10 minutos.
3. Pasar las preparaciones cromosómicas a una cubeta con 2xSSC a temperatura ambiente durante 10 minutos. Repetir la operación.
4. Poner las preparaciones en una cubeta con 1xPBS durante 5 minutos.

2.4.5. Revelado

El revelado se basa en la utilización de la proteína avidina con alta afinidad a la biotina. Esta proteína lleva acoplada la enzima peroxidasa. La avidina se unirá a las moléculas de biotina y al añadir el sustrato de la peroxidasa se produce la reacción, liberándose cationes que son captados por un cromógeno, la diamimobencidina (DAB), que precipita dando lugar a una señal colorimétrica.

1. Añadir a 0,21 μL de solución I (10 mL Tris 50 mM a pH 7,6; 4% de albúmina de suero bovino) 3,5 μL de reactivo A y 3,5 μL de reactivo B del kit Elite (Vector).
2. Añadir a cada preparación 20 μL de la solución anterior.
3. Poner un cubreobjetos sobre la zona donde se ha añadido la solución. Si quedan burbujas, eliminarlas con la punta de la pipeta.
4. Poner las preparaciones en una cámara húmeda e incubar a 37°C durante 45 minutos.
5. Poner las preparaciones en 1xPBS durante 10 minutos. Repetir dos veces.
6. Añadir a cada preparación 500 μL de solución III (5 μg de diaminobencidina; 1 mL de Tris HCl 50 mM a pH 7,6; 1 μL de H_2O_2).
7. Poner las preparaciones en una cámara húmeda e incubar a 37°C durante 45 minutos.
8. Poner las preparaciones en 1xPBS de 5 minutos a una hora.
9. Preparar la solución de tinción añadiendo 5 mL de Giemsa a 199 mL de tampón P (40% de $\text{NaH}_2\text{PO}_4\text{H}_2\text{O}$ 50 mM; 60% $\text{Na}_2\text{HPO}_4 \cdot 12 \text{H}_2\text{O}$ 50 mM).
10. Poner las preparaciones en la tinción durante 1 minuto.
11. Poner las preparaciones en una cubeta con H_2O destilada. Repetir en otra cubeta de H_2O destilada.
12. Dejar secar las preparaciones.
13. Añadir a cada preparación una gota de montador de preparaciones biológicas y poner

un cubreobjetos.

2.4.6. Visualización y localización de las señales de hibridación

Las preparaciones se han observado con un microscopio Nikon Optiphot-2 y se tomaron fotografías de las preparaciones cromosómicas hibridadas con la cámara Nikon Coolpix-4500.

Las señales de hibridación *in situ* se han localizado y numerado a partir del patrón de bandas determinado por el cromosoma 2 de *D. repleta* (Ruiz y Wasserman 1993, González *et al.* 2005, 2007), teniendo en cuenta las reorganizaciones, inversiones y cambios de orientación propios de cada linaje.

2.5. Electroforesis de campo pulsante (PFGE)

Se ha usado el campo pulsante *CHEF-DR III*® de BIORAD. El campo pulsante consiste en separar por peso molecular distintos fragmentos de DNA. A diferencia de los geles de electroforesis, el campo pulsante permite separar con gran resolución fragmentos de elevado peso molecular, mediante el cambio continuo de voltaje entre dos puntos distintos en el polo positivo. Las muestras digeridas (según el protocolo de digestión de DNA con enzimas de restricción) se cargan en un gel al 1% (1 g de agarosa *Certified*® de BIORAD en 100 mL de tampón TBE 0,5x).

La fuente electroforética debe rellenarse con 2,2 L de tampón TBE 0,5x. Los parámetros de la fuente de poder se han establecido optimizando la separación de fragmentos comprendidos entre 0,2 y 23 Kb, puesto que específicamente se ha usado para separar clones BAC digeridos.

- Temperatura del refrigerador: 14°C (encender ½ hora antes de usar para asegurar que el tampón ha alcanzado esta temperatura).
- Flujo de la bomba: 70 (equivalente a 0,8 L por minuto).
- Tiempo de intercambio de voltaje inicial: 0,1 segundos.
- Tiempo de intercambio de voltaje final: 6 segundos.
- Ángulo entre los puntos de intercambio: 120°C.
- Voltaje: 6 V/cm.
- Tiempo: 12 horas.
- Una vez finalizado, el gel se tiñe normalmente con bromuro de etidio (0,5 µg/mL).

2.6. Hibridación de Southern

2.6.1. Marcado y detección de sondas con DIG

Se han marcado las sondas para hibridar en la técnica de hibridación de Southern con DIG 16-dUTP, según el método de cebadores al azar (véase sección 2.6.1).

2.6.2. Transferencia

La transferencia del DNA del gel de agarosa a la membrana se ha realizado siguiendo el protocolo descrito por Sambrook (1989), mientras que para el revelado de las sondas se ha utilizado el kit comercial de ROCHE y su protocolo.

1. Después de realizar la electroforesis del DNA digerido con enzimas de restricción en un gel de agarosa del 0,7%, se debe teñir el gel con bromuro de etidio e irradiar con luz ultravioleta para realizar una fotografía con una regla fluorescente que servirá de guía para

analizar los resultados.

2. Se sumerge el gel en solución de desnaturalización (0,5M NaOH; 1,5M NaCl) durante una hora en agitación suave.
3. Eliminar la solución de desnaturalización y añadir el mismo volumen de la solución de neutralización (0,5M Tris ClH a pH 7,5; 1,5M NaCl) durante una hora, a temperatura ambiente en agitación suave.
4. Preparar un recipiente lleno de 10x SSC. Colocar sobre el recipiente una plataforma de acrílico. Sobre la plataforma colocar papel *Wattman* 3MM de manera que sus extremos estén dentro de la solución. Empapar el papel en 10x SSC y eliminar con una varilla de vidrio las burbujas que se hayan quedado entre el papel y la plataforma.
5. Colocar el gel invertido (boca abajo) sobre el papel húmedo y eliminar las burbujas.
6. Hacer un marco de papel de parafilm alrededor del gel.
7. Colocar sobre el gel un trozo de membrana de nylon previamente humedecida en 2xSSC y eliminar las burbujas. La membrana debe llevar una pequeña marca en su extremo izquierdo para identificar la cara sobre la que se ha transferido el DNA.
8. Colocar sobre la membrana dos trozos de papel *Wattman* 3MM del mismo tamaño que la membrana previamente humedecidos con 2xSSC y eliminar las burbujas.
9. Colocar encima de ellos colocar una pila de 10 centímetros de papel absorbente y sobre ella un peso de 500 g. Dejar transferir de 16-20 horas.
10. Desmontar el dispositivo de la transferencia y marcar en la membrana los pocillos del gel con una aguja enmangada.
11. Sumergir la membrana en 2xSSC durante 5 minutos a temperatura ambiente en agitación.
12. Dejar secar la membrana a temperatura ambiente. Irradiar con luz ultravioleta durante 3 minutos cada lado de la membrana para fijar el DNA.

2.6.3. Prehibridación e hibridación

1. Colocar la membrana en un tubo de hibridación (*roller*) con 20 mL de solución de prehibridación (5xSSC; 50% formamida; 0,1% N-laurilsarcosina; 0,02% SDS; 2,5% agente bloqueante) precalentada a 42°C.
2. Mantener el tubo de hibridación en rotación durante 1 hora a 42°C.
3. Desnaturalizar la sonda calentándola a 95 °C durante 10 minutos y colocándola en hielo durante 3 minutos.
4. Eliminar la solución de prehibridación y añadir 10 mL de solución de hibridación (solución de prehibridación con 5-25 ng de DNA de sonda/mL). Mantener en rotación durante 12-16 horas a la temperatura de hibridación.

2.6.4. Posthibridación y detección

1. Extraer la membrana del tubo de hibridación y sumergirla en 2xSSC/0,1% SDS durante 15-20 minutos a temperatura ambiente y en agitación. Repetir este paso una vez más.
2. Sumergir la membrana en 0,1x SSC/0.1% SDS durante 20 minutos a 68°C en agitación. Repetir este paso una vez más.
3. Lavar la membrana durante 1-5 minutos con tampón 1 (0,1 M ácido maleico; 0,15 M NaCl; pH 7,5).
4. Mantener durante 30 minutos en el tampón 2 (1% de agente bloqueante en tampón de ácido maleico).
5. Mantener durante 30 minutos en el tampón 2 con Anti-DIG-AP conjugado (1:5.000).
6. Lavar dos veces durante 15 minutos con tampón 1.

7. Equilibrar durante 3 minutos en tampón 3.
8. Eliminar el tampón 3 y añadir sobre la membrana la solución de coloración (200 μ L en 10 mL de tampón 3). Mantener a oscuras y temperatura ambiente entre 10 minutos y 20 horas hasta que se puedan diferenciar las bandas claramente.
9. Eliminar la solución de coloración con H₂O destilada.
10. Dejar secar la membrana para su posterior análisis.

2.7. Búsquedas de similitud y alineamiento

Los extremos de los clones BAC y las PCRs secuenciadas de *D. buzzatii* se han usado como secuencia problema en búsquedas de similitud para localizarlas en el genoma de *D. mojavensis* (*CAF1 Assembly*, Febrero 2006). Estas búsquedas de similitud se han realizado mediante el programa blastn implementado en DroSpeGe (Gilbert 2005) disponible en <http://insects.eugenes.org/>. DroSpeGe dispone de una versión actualizada del ensamblaje, el alineamiento y la anotación de los 12 genomas actualmente secuenciados del género *Drosophila*. La descarga de algunos fragmentos del genoma de *D. mojavensis* se ha realizado desde dicha base de datos. El alineamiento de secuencias de *D. buzzatii* con secuencias descargadas de *D. mojavensis* se ha realizado con el programa bl2seq (Tatusova y Madden 1999) implementado en la página del NCBI (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). Por norma general, sólo se han considerado como significativos los resultados con un tamaño mínimo de 40 pb y E-value $\leq 1e-03$. Por otro lado, los alineamientos entre múltiples secuencias se han realizado con el programa MUSCLE (Edgar 2004) disponible en la página web del EBI (<http://www.ebi.ac.uk/Tools/muscle/index.html>). Este programa también se ha utilizado para el reconocimiento de terceras bases de codones cuando se han alineado regiones codificantes. Los cambios sinónimos y no sinónimos en estas regiones codificantes se han

determinado mediante el análisis de estos alineamientos con el programa DnaSP (Librado y Rozas 2009) disponible en www.ub.edu/dnasp/.

2.8. Secuenciación y anotación de los clones BAC 1N19 y 20019

Dos clones BAC fueron totalmente secuenciados, el BAC 1N19 (punto de rotura AC) y el BAC 20019 (punto de rotura BE). Se secuenciaron por el proceso de secuenciación aleatoria, realizando una subgenoteca para cada clon. El servicio de secuenciación de Macrogen (<http://www.macrogen.com/>) realizó la genoteca, la mayor parte de la secuenciación y el ensamblaje de las lecturas en 5 *contigs* (2 para el BAC 1N9 y 3 para el BAC 20019). El cierre de los huecos entre estos *contigs* se realizó en nuestro laboratorio mediante amplificación y secuenciación de los productos de las reacciones PCR. Algunos detalles de la secuenciación de los dos clones BAC están compilados en la Tabla 2.5.

La anotación se realizó utilizando los programas de predicción de genes GENEID (Parra *et al.* 2000) y GENSCAN (Burge y Karlin 1997). Las predicciones se compararon con los modelos de GLEANR anotados en el genoma de *D. mojavensis* disponible en DroSpeGe (Gilbert 2005) y las anotaciones *D. melanogaster* disponible en FlyBase en <http://flybase.org> (Tweedie *et al.* 2009). Las predicciones iniciales fueron posteriormente modificadas a mano para ajustar los sitios de procesamiento, el marco de lectura y los codones de inicio y stop.

Para la anotación de los elementos transponibles (TE) y secuencias repetidas en los dos clones BAC secuenciados se llevaron a cabo búsquedas de similitud por fragmentos que se contrastaron con las bases de datos de secuencias repetidas y TE del NCBI (Tatusova y Madden 1999), DPDB (Casillas *et al.* 2005), Repbase (Jurka *et al.* 2005), y DroSpeGe (base de datos basada en ReAs, Gilbert 2005). También en este caso, sólo se han considerados significativos los resultados con un tamaño mínimo de 40 pb y E-value $\leq 1e-03$. Un análisis

más detallado se realizó en las regiones de los puntos de rotura. Mediante BLAST entre dos secuencias (bl2seq) las propias secuencias de los puntos de rotura fueron alineadas con la secuencia del elemento transponible con mayor identidad en la búsqueda anterior.

Tabla 2.5: Especificaciones del proceso de secuenciación de los clones BAC 1N19 y 20O19.

Clon BAC	1N19	20O19
Punto de rotura que contiene	AC	BE
Tamaño del clon BAC (pb)	138.724	143.293
Tamaño de los fragmentos seleccionados de las digestiones parciales para la construcción de la genoteca (Kb)	Entre 2 y 5	Entre 2 y 5
Número de lecturas en la secuenciación aleatoria	1536	1344
Número de lecturas excluidas	186	148
Tamaño medio de la secuenciación por lectura (pb)	869 ± 140	900 ± 150
Cobertura	8.47x	7.56x
Número de <i>contigs</i>	2 (6 y 7)	3 (3, 4 y 5)
Huecos (<i>Gaps</i>)	1	2
PCR diseñadas para cerrar el ensamblaje (véase Tabla 2.1 para los cebadores)	6-7	4-3 3-5

2.9. Extracción de mRNA

Se obtuvo RNA total de distintas fases del desarrollo de *D. buzzatii*: embrión (E), larvas (L), pupa (P) y adultos (A).

La recolección de embriones se realizó en distintos tiempos (1,30h, 2,15h, 2,30 h y 22h). En pequeñas cámaras de recolección se disponen placas de Petri con LB-agar y colorante

oscuro (para poder diferenciar los embriones claros). Una vez solidificado, se añade un poco de levadura en el centro. Se dejan los adultos unas 4 horas para que se acostumbren al nuevo medio. Se sustituye la placa de Petri (mismo contenido) y se recolectan los huevos en el tiempo deseado. Las larvas, las pupas y los adultos machos se pueden seleccionar de los frascos directamente. Los adultos hembras deben ser vírgenes, por lo que de un frasco sin adultos se irán recolectando todas las hembras recién salidas de las pupas antes que sean sexualmente maduras. Al no estar fecundadas, se evitan posibles contaminaciones por RNA de los embriones.

Las extracciones de RNA se han realizado según el protocolo del kit *RNeasy* de QIAGEN. Todos los pasos se realizan a temperatura ambiente y en lugares libres de RNAsa.

1. Colocar muestras del estadio elegido en el homogeneizador. Añadir 600 μ L de tampón RLT. Triturar.
2. Transferir el lisado a un tubo *Eppendorf* para centrifuga.
3. Centrifugar 3 minutos a máxima velocidad. Transferir el sobrenadante a un nuevo tubo *Eppendorf*.
4. Añadir 600 μ L de etanol al 70% y mezclar inmediatamente con la pipeta. Transferir a una columna *RNeasy* (con el colector).
5. Centrifugar 15 segundos a 10.000 rpm. Vaciar el colector.
6. Añadir 700 μ L de tampón RW1 a la columna.
7. Centrifugar 15 segundos a 10.000 rpm. Vaciar el colector.
8. Colocar la columna en un nuevo colector.
9. Añadir 500 μ L de tampón RPE a la columna.
10. Centrifugar 15 segundos a 10.000 rpm. Vaciar el colector.
11. Añadir 500 μ L de tampón RPE a la columna.
12. Centrifugar 2 minutos a 10.000 rpm. Vaciar el colector.

13. Centrifugar 1 minuto adicional a máxima velocidad.
14. Colocar la columna en un nuevo tubo *Eppendorf*. Pipetear 30 μL de H_2O (libres de RNAsa) directamente a la membrana de la columna. Esperar al menos 1 minuto.
15. Centrifugar 1 minuto a 10.000 rpm. Desechar la columna y conservar el producto eluido (RNA) a -20°C .

2.10. RT-PCR

Las RT-PCR son semicuantitativas y retrotranscriben el mRNA a cDNA mediante un sistema de cebadores de hexanucleótidos al azar (*random primers*) y la enzima retrotranscriptasa para poder usar como molde en posteriores reacciones PCR normales. Se realizan dos alícuotas, una RT+ (se añade enzima retrotranscriptasa) y otra RT- (a la que no se le añade enzima retrotranscriptasa). En caso de que las alícuotas RT- haya producto se debe a DNA genómico residual.

1. Añadir en un tubo *Eppendorf*: 8 μL RNA + 1 μL tampón 10x de DNAsa I + 1 μL DNAsa I (volumen total 10 μL) y dejar 10 minutos a temperatura ambiente.
2. Inactivar DNAsa I (añadir 1 μL EDTA 25 nM). Incubar 10 minutos a 65°C .
3. Dividir el volumen en dos tubos *Eppendorf* (RT+ y RT-) de PCR (5 μL en cada uno).
4. Añadir a cada alícuota: 1 μL *Oligo-PrimerT* (dT_s) + 1 μL dNTPs (10 mM) + 6 μL H_2O DEPC (volumen total 13 μL).
5. Colocar los tubos *Eppendorf* al termociclador 5 minutos a 65°C .
6. Añadir 2 μL 0.1x DTT + 5 μL 5x tampón de DTT.
7. Colocar los tubos *Eppendorf* al termociclador 2 minutos a 42°C .
8. Añadir 1 μL enzima retrotranscriptasa a las alícuotas RT+.
9. Colocar los tubos *Eppendorf* al termociclador 50 minutos a 42°C .

10. Colocar los tubos *Eppendorf* al termociclador 15 minutos a 70°C.

2.11. RACE-PCR

El experimento de RACE permite localizar el origen y el final de la transcripción de un gen. Las 5'RACE y 3'RACE amplifican específicamente las 5'UTR y 3'UTR hasta el origen y el final de un transcrito, respectivamente. Como en la RT-PCR, se obtiene cDNA de una muestra de mRNA. En este caso, para obtener cDNA no se usan cebadores de hexanucleótidos al azar como para las RT-PCR, sino adaptadores específicos que se unen al principio y al final de los transcritos. Estos adaptadores permiten retrotranscribir los extremos (5' y 3') de todos los mRNA a cDNA (véase la Figura 2.1 como esquema general). Dos PCR usando este cDNA como molde permiten obtener un producto de las regiones distales de los transcritos de interés. La primera PCR (*Outer*) usa un cebador que se une al adaptador retrotranscrito en el cDNA con un cebador específico del transcrito de interés (primer y último exón, para la 5'UTR y la 3'UTR, respectivamente). La segunda PCR (*Inner*) amplifica la primera PCR (*Outer*) como molde usando cebadores anidados. Las RACE-PCR se han realizado según el protocolo del kit *FirstChoice RLM-RACE*[®] de Ambion (AM1700).

2.11.1. 3' RACE

El protocolo de la 3'RACE está optimizado para una cantidad de 1 µg de RNA total.

a) Retrotranscripción

1. Añadir en un tubo *Eppendorf* libre de RNAsa en hielo: 2 µL de RNA (1 µg) + 4 µL dNTP (10 mM) + 2 µL de adaptador 3'RACE + 2 µL tampón 10x de enzima retrotranscriptasa + 1 µL Inhibidor de RNAsa + 1 µL M-MLV retrotranscriptasa + 8 µL de

H₂O (libre de RNAsa).

2. Mezclar suavemente y hacer un pulso con la centrífuga. Incubar a 42°C durante una hora. Se puede guardar el producto (cDNA) a -20°C o proceder a la PCR.

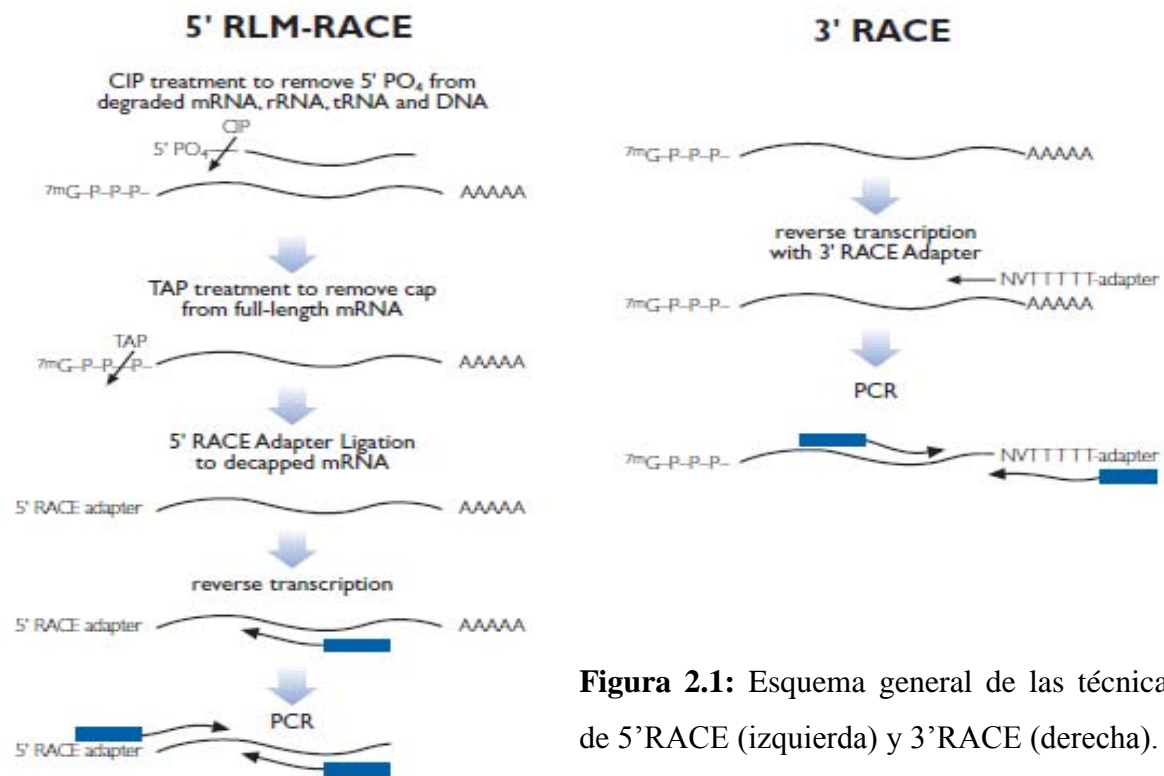


Figura 2.1: Esquema general de las técnicas de 5'RACE (izquierda) y 3'RACE (derecha).

b) Outer-PCR

1. Añadir en un tubo *Eppendorf* de PCR: 1 µL de reacción del paso anterior (cDNA) + 5 µL tampón 10x de la Taq polimerasa + 4 µL dNTP (10 mM) + 2 µL del cebador del último exón del gen específico (10 µM) + 2 µL del cebador 3 'Outer RACE + 1,25 U de Taq polimerasa (0,25 µL de 5U/µL) + completar con H₂O (libre de RNAsa) hasta obtener un volumen final de 50 µL. 2. PCR (véase sección 2.2.2).

b) Inner-PCR

1. Añadir en un tubo *Eppendorf* de PCR: 1 μ L de reacción del paso anterior (*Outer-PCR*) + 5 μ L tampón 10x de la Taq polimerasa + 4 μ L dNTP (10 mM) + 2 μ L del cebador anidado al usado en la *Outer-PCR* (10 μ M) + 2 μ L del cebador 3'*Inner* RACE + 1,25 U de Taq polimerasa (0,25 μ L de 5U/ μ L) + completar con H₂O (libre de RNAsa) hasta obtener un volumen final de 50 μ L.
2. PCR (véase sección 2.2.2).

2.11.2. 5' RACE

Este protocolo esta optimizado para una cantidad de 10 μ g de RNA total.

a) Tratamiento con CIP

1. Añadir a un tubo *Eppendorf* de centrífuga: 10 μ L de RNA Total + 2 μ L 10x tampón CIP + 2 μ L CIP + completar con H₂O (libre de RNAsa) hasta volumen final de 20 μ L.
2. Mezclar suavemente, hacer un pulso en la centrífuga. Incubar a 37 °C durante una hora. Se puede almacenar la muestra a -20 °C.
3. Añadir a la reacción: 15 μ L de solución de acetato de amonio + 115 μ L H₂O (libre de RNAsa) + 150 μ L de fenol:cloroformo.
4. Mezclar y centrifugar 5 minutos a máxima velocidad a temperatura ambiente. Transferir la fase acuosa a un nuevo tubo *Eppendorf*.
5. Añadir 150 μ L cloroformo. Mezclar y centrifugar 5 minutos a máxima velocidad a temperatura ambiente. Transferir la fase acuosa a un nuevo tubo *Eppendorf*.
6. Añadir 150 μ L de isopropanol, mezclar y dejar en hielo durante 10 minutos.
7. Centrifugar a máxima velocidad durante 20 minutos. Enjuagar el pellet con 0,5 mL de etanol al 70% frío. Mezclar y centrifugar 5 minutos a máxima velocidad.

8. Eliminar el etanol con cuidado y permitir que el precipitado se seque al aire libre.
9. Resuspender el pellet en 11 μL de H_2O (libre de RNAsa). Se puede almacenar la muestra a -20°C .

b) Tratamiento con TAP

1. Añadir al tubo *Eppendorf* de la reacción anterior: 1 μL 10x tampón TAP + 2 μL TAP.
2. Mezclar suavemente, hacer un pulso en la centrífuga. Incubar a 37°C durante una hora.
Se puede almacenar la muestra a -20°C .

c) Ligación del adaptador 5' RACE

1. Añadir a un tubo *Eppendorf*: 8 μL de la reacción anterior (RNA tratado con CIP/TAP) + 1 μL adaptador 5' RACE + 1 μL 10x tampón RNA ligasa* + 2 μL T4 RNA ligasa (2,5 U/ μL).

* (Antes de usar, calentar el tampón rodándolo entre los guantes para resuspender cualquier precipitado. Dado que este tampón contiene ATP, no se recomienda la exposición a un temperatura superior a los 37°C).

2. Mezclar suavemente, hacer un pulso en la centrífuga. Incubar a 37°C durante una hora.
Se puede almacenar la muestra a -20°C .

d) Retrotranscripción

1. Añadir en un tubo *Eppendorf* de centrífuga libre de RNAsa en hielo: 2 μL de RNA (10 ng) + 4 μL dNTP (10 mM) + 2 μL de adaptador 3'RACE + 2 μL tampón 10x de retrotranscriptasa + 1 μL de inhibidor de RNAsa + 1 μL M-MLV retrotranscriptasa + 8 μL H_2O (libre de RNAsa).
2. Mezclar suavemente y hacer un pulso en la centrífuga. Incubar a 42°C durante una hora.

Se puede guardar el producto (cDNA) a -20°C o proceder a la PCR.

b) *Outer*-PCR

1. Añadir en un tubo *Eppendorf* de PCR: 1 μL de reacción del paso anterior (cDNA) + 5 μL tampón 10x de la Taq polimerasa + 4 μL dNTP (10mM) + 2 μL del cebador del exón del gen específico (10 μM) + 2 μL del cebador 5'*Outer*-RACE + 1,25 U de Taq polimerasa (0,25 μL de 5U/ μL) + completar con H_2O (libre de RNAsa) hasta volumen final de 50 μL . 2. PCR (véase sección 2.2.2).

b) *Inner*-PCR

1. Añadir en un tubo *Eppendorf* de PCR: 1 μL de reacción del paso anterior (*Outer*-PCR) + 5 μL tampón 10x de la Taq polimerasa + 4 μL dNTP (10 mM) + 2 μL del cebador anidado al usado en la *Outer*-PCR (10 μM) + 2 μL del cebador 5'*Inner*-RACE + 1,25 U de Taq polimerasa (0,25 μL de 5U/ μL) + completar con H_2O (libre de RNAsa) hasta volumen final de 50 μL .

2. PCR (véase sección 2.2.2).

2.12. Detección de RNA de doble cadena (dsRNA)

Un dsRNA se compone de dos de fragmentos de RNA de cadena simple (ssRNA) complementarios. El protocolo de detección de RNA de doble cadena es similar al de la RT-PCR. El mRNA total se retrotranscribe a cDNA y, a continuación, se realiza una PCR normal usando este cDNA como molde. En la detección de dsRNA, el mRNA se digiere con una RNAsa previamente a la retrotranscripción. La RNAsa digerirá todo el RNA de la cadena simple (ssRNA) pero no el de doble cadena (dsRNA). Después de inactivar la RNAsa, y

desnaturalizar los dímeros de RNA se realiza la RT-PCR. Sólo se retrotranscribe el dsRNA protegido de la actividad RNAsa. De este modo, sólo se retrotranscribirá el dsRNA no digerido.

Se reparte la muestra en dos alícuotas en dos tubos distintos, una con RNAsa (C) y otra sin RNAsa (S). Cada tubo se reparte en distintas alícuotas para los distintos controles, añadiendo enzima retrotranscriptasa (RT+) o no (RT-), sin desnaturalizar entre la digestión y la RT-PCR (N) y usando cebadores de hexanucleótidos al azar (*random primers*) en vez de cebadores *Oligo-T* (OT).

Aparte del control de contaminación por DNA genómico (sin retrotranscriptasa, RT), se han añadido controles para todos los puntos del protocolo. De este modo se prepararon dos muestras, una con RNAsa (muestra S) y otra sin (muestra C). A su vez, cada muestra fue dividida en cuatro alícuotas distintas para realizar controles en cada paso. Las alícuotas + tienen retrotranscriptasa. Las alícuotas – no tienen retrotranscriptasa. Las alícuotas N no hacen un ciclo de desnaturalización a 94°C previo a la RT-PCR. Finalmente, se han distinguido dos tipos de cebadores para la RT-PCR. Las alícuotas T se retrotranscriben con cebadores OligoT en vez de cebadores al azar. La RNAsa tiene actividad endonucleasa 3'-5'. Los cebadores OligoT permiten la retrotranscripción desde la poli-A de los mRNA y los cebadores al azar (*random primers*) retrotranscriben distintos fragmentos al azar desde distintos puntos del mRNA.

La Tabla 3.20 recoge las especificaciones propias de cada alícuota. Las PCR-RT no incluyen medidas de desnaturalización antes de empezar y la enzima RT no podrá amplificar.

1. Preparar las dos alícuotas. En un tubo *Eppendorf* de centrífuga (alícuota S) se añade: 12 µL de RNA total (~10 µg) + 8 µL de H₂O (libre de RNAsa). En otro tubo *Eppendorf* de centrífuga (alícuota C) se añade: 2 µL de RNA total (~10 µg) + 2 µL 10x tampón RNAsa + 0,5 µL RNAsa ONE (1,5 U de Promega) + 5,5 µL de H₂O (libre de RNAsa).

2. Incubar las dos alícuotas a 37°C durante 5 minutos.
3. Añadir a las dos alícuotas: 1,5 µL de SDS (10%) + 0,5 µL de tRNA de levadura (de un *stock* 10 µg/ µL) + 6 µL de acetato de sodio (3M) + 150 µL de etanol al 100%. Dejar a -20°C un mínimo de 20 minutos.
4. Centrifugar las dos alícuotas 10 minutos a 12.000 rpm. Eliminar el sobrenadante.
5. Añadir a las dos alícuotas 1 mL de etanol al 70%. Mezclar y centrifugar 5 minutos a 12.000 rpm. Eliminar el etanol con cuidado y permitir que el precipitado se seque al aire.
6. Resuspender los pellets en 17 µL de H₂O (libre de RNAsa).
7. Añadir a las dos alícuotas: 1 µL tampón 10x de DNAsa I + 1 µL DNAsa I y dejar 15 minutos a temperatura ambiente.
8. Inactivar DNAsa I (Añadir 1 µL EDTA 25 nM). Incubar 10 minutos a 65°C.
9. Dividir el volumen total (21 µL) de cada tubo *Eppendorf* (S y C) en 4 alícuotas de 5 µL de volumen del volumen total. En total habrá 8 alícuotas: S+, S-, SN, ST y C+, C-, CN, CT.
10. Desnaturalizar las alícuotas: S+, S-, ST y C+, C-, CT durante 10 minutos a 97°C.
11. Añadir a las alícuotas ST y CT: 1 µL de *Oligo-PrimerT* (dT_s).
12. Añadir a las alícuotas S+, S-, SN y C+, C-, CN: 1 µL de cebadores de hexanucleótidos al azar.
13. Añadir a todas las alícuotas: 1 µL dNTPs (10 mM) + 6 µL H₂O DEPC.
14. Colocar los tubos *Eppendorf* en el termociclador 5 minutos a 65°C.
15. Añadir a todas las alícuotas: 2 µL 0,1x DTT + 5 µL 5x tampón de DTT.
16. Colocar los tubos *Eppendorf* en el termociclador 2 minutos a 42°C.
17. Añadir las alícuotas S+, SN, ST y C+, CN, CT 1 µL enzima retrotranscriptasa.
18. Colocar los tubos *Eppendorf* en el termociclador 50 minutos a 42°C.
19. Colocar tubos *Eppendorf* en el termociclador 15 minutos a 70°C.

2.13. Hibridación de Northern

La hibridación de Northern consiste en la hibridación de una sonda marcada en una membrana que contiene mRNA. El gel de agarosa debe realizarse en condiciones de desnaturalización, evitando las estructuras secundarias del mRNA. Las condiciones desnaturalizantes se consiguen químicamente añadiendo formaldehído y formamida en el gel y en las muestras de la carga. Cabe destacar que, a diferencia de la hibridación de Southern, las membranas de las hibridaciones de Northern son especulares a los carriles de los geles. En esta técnica, se debe trabajar con condiciones de máxima esterilidad, por lo que se evita cualquier manipulación innecesaria. Cuando se prepara la transferencia, el gel no se invierte y se deja cara arriba, por tanto, la membrana quedará transferida aparentando el orden inverso.

Se realizaron tres rondas de hibridación pero no se obtuvo señal hasta la última. En cada ronda, se añadía una nueva membrana. De este modo, la primera ronda se hizo con la membrana 1. La ronda dos con las membranas 1 y 2. La ronda 3, a la que también se añadió un Dot Blot, se realizó con las membranas 1, 2 y 3. Las sondas, así como sus concentraciones en cada hibridación, se muestran en la Tabla 2.6. Las especificaciones y diferencias entre cada membrana, se muestran en la Tabla 2.7.

No tenemos controles para discernir cuál fue el paso crítico que no nos permitió observar señal en las dos primeras rondas pero algún paso quedó optimizado en la tercera ronda. Este paso no tuvo que ver con la composición del gel o de la membrana porque la membrana 1, que no hibridó al principio, lo hizo en la tercera ronda. Las membranas se reutilizaron de una ronda a la siguiente eliminando la sonda hibridada (incluso cuando no se obtuvo señal) utilizando 0,1% SDS durante 20 minutos a temperatura de ebullición.

Tabla 2.6: Sondas y cantidad (entre paréntesis) usadas en cada membrana en cada ronda de hibridación de Northern. Se añade H₂O para obtener un volumen total de 9 µL. Las concentraciones de cada sonda son: PCR7-3': 5,7 ng/µL; PCR2-3: 7,7 ng/µL; PCR5-7: 15,3 ng/µL; *Gapdh*: 7,3 ng/µL.

Ronda	Membrana			
	1	2	3	Dot Blot
1	PCR 7-3' (4 µL)	-	-	-
2	<i>Gapdh</i> (6 µL)	PCR 5-7 (5 µL)	-	-
3	PCR 5-7 (5 µL)	PCR 2-3 (5 µL)	PCR 5-7 (5 µL)	PCR 2-3 (5 µL)

Tabla 2.7: Composición de los geles para transferir a cada membrana de la hibridación de Northern. Las especificaciones de las diluciones de las muestras cargadas en los pocillos y el tipo de tampón para la electroforesis, también se muestran. *(*Northern Formaldeyde Load Dye* de Ambion). ** (10x *Running buffer for Formaldeyde gels* de Ambion).

Gel	Composición del gel	Dilución de la muestra	Tampón**
1°	MOPS	3V de <i>Loading Dye</i> *	MOPS
	+Formaldehido	+Formaldehido	+Formaldehido
		+Formamida	
		+Bromuro de etidio	
2°	MOPS	3V de <i>Loading Dye</i> *	MOPS
	+Formaldehido	+Formaldehido	
		+Formamida	
		+Bromuro de etidio	
3°	MOPS	3V de <i>Loading Dye</i> *	MOPS
	+Formaldehido	+Formaldehido	
	+Bromuro de etidio	+Formamida	

Apenas hay diferencia entre las especificaciones de cada gel. Las mayores diferencias se encuentran en el marcado de la sonda y en los lavados post-hibridación. El marcaje siempre se realizó con P32, pero la alícuota de la tercera ronda es distinta a la de las dos anteriores. En cuanto a los lavados post-hibridación, en las dos primeras rondas se realizaron dos enjuagues con 2XSSC/0,1SDS y un lavado con 0,5XSSC/0,1SDS a 65°C durante 20 minutos. En la última ronda, sólo se realizaron dos lavados a 65°C durante 15 minutos con 2XSSC/0,1SDS.

2.13.1. Preliminares de la técnica de hibridación de Northern

En la técnica hibridación de Northern, la esterilidad y las condiciones libres de RNAsa deben ser extremas. Todos los utensilios, así como el espacio para trabajar y los aparatos deben limpiarse a fondo.

1. Lavar con agua y jabón.
2. Dejar en remojo en una solución de SDS.
3. Rociar con *RNAsaway* para eliminar RNAsa.
4. Enjuagar con H₂O-DEPC.
5. Dejar secar al aire en una zona libre de posibles contaminaciones.

2.13.2. Marcado de las sondas con P32 radioactivo

La sonda para hibridar en la hibridación de Northern ha sido marcada con P32 radioactivo con el kit de ROCHE mediante el sistema de cebadores de hexanucleótidos al azar (*random primers*). Las sondas utilizadas en cada membrana de la hibridación del Northern se muestran en la Tabla 2.6.

1. Añadir la sonda a un tubo *Eppendorf* libre de RNAsa para PCR. Debe haber 25 ng de

sonda como mínimo. Completar el volumen final de la sonda hasta 9 μL con H_2O libre de RNAsa.

2. Desnaturalizar a 95°C durante 10 minutos. Dejar en hielo.
3. Añadir la siguiente mezcla: 1 μL de dATP, 1 μL de dGTP, 1 μL de dTTP, 2 μL de *Reaction Mixture* (mezclar con el vórtex durante 2 minutos antes de usarlo), 1 μL de enzima Klenow y 5 μL de P32.
4. Mezclar y programar un ciclo de 37°C durante 1 hora.
5. Purificar el producto de PCR mediante columnas de filtración (Protocolo de QIAGEN. Cat.No./ID28106).

2.13.3. Electroforesis

a) Muestras

Añadir a cada muestra ($\sim 4 \mu\text{L}$ RNA) 2 μL de MOPS 10x* + 1,2 mL de formaldehído + 10 mL de formamida al 37%. Incubar a 65°C durante 15 minutos. *(Cat.No/ID8671 de Ambion. Incluye acetato sódico y EDTA).

Dejar en hielo y cargar 3 volúmenes de *Loading Dye (Northern Formaldehyde Load Dye* de Ambion) que incluye los colorantes cyanol-xileno y azul de bromofenol. En algunos geles de las hibridaciones de Northern hemos cargado también en este punto 0,5 μL de bromuro de etidio (véase Tabla 2.7). (*stock* al 0,5 $\mu\text{g}/\mu\text{L}$). Los demás se han teñido posteriormente (véase sección 2.2.4). Las concentraciones de las muestras cargadas en los geles de hibridación de Northern se muestran en la Tabla 2.8.

b) Gel

Mezclar 3 g de agarosa (1,2 g de agarosa por cada 100 mL de tampón) en 215 mL de

H₂O libre de RNAsa + MOPS (10x) 7,5 mL (0,6 M) formaldehído al 37%. El marcador de RNA cargado es de Promega (G3191) con un patrón de bandas desde 0,28 a 6,58 Kb.

Se aplica entre 35 y 45 V al gel durante toda la noche en tampón 1x MOPS. En algunos casos se ha añadido 1,1% de formaldehído (véase Tabla 2.7).

Antes de teñir los geles, se han lavado intensamente con agua libre de RNAsa para eliminar el formaldehído. La tinción se realiza sumergiendo los geles durante 1 hora en una solución de bromuro de etidio (0,5 µg/mL) en H₂O libre de RNAsa.

Tabla 2.8: Concentraciones [ng/µL] y cantidades cargadas en los pocillos de los geles usados para cada membrana de la hibridación de Northern. * µL usados en un volumen total de 500 µL (* 10 mM NaOH + 1 mM EDTA).

Muestras	Concentración (ng/µL)	Membrana (µL)			
		1	2	3	Dot Blot
m. marcador de RNA	stock	3	3	3	-
1. mRNA <i>D. mojavensis</i>	720	10	8	3	25*
2. Huevo (22h)	70	10	8	3	-
3. Larva	1.067	5	4,5	3	-
4. Pupa	1.267	5	4,5	3	-
5. Macho	1.139	5	4,5	3	-
6. Hembra	1.400	5	4,5	3	15*
7. DNA <i>D.buzzatii</i>	1.300	10	8	3	25*

2.13.4. Transferencia

1. Después de realizar la tinción del gel se toma una fotografía colocando una regla en el costado que servirá de guía para analizar los resultados (el marcador se ve este momento, pero no se verá en la membrana porque no hibridará con la sonda).

2. Lavar un par de veces con H₂O libre de RNAsa para eliminar el bromuro de etidio. Equilibrar con un baño de 10x SSC durante 30 minutos.
3. Preparar un recipiente lleno de 10x SSC. Colocar sobre el recipiente una plataforma de acrílico. Sobre la plataforma colocar papel *Wattman* 3MM de manera que los extremos queden sumergidos dentro de la solución. Empapar el papel en 10x SSC y eliminar con una varilla de vidrio las burbujas entre el papel y la plataforma.
4. Colocar el gel sobre el papel húmedo y eliminar las burbujas.
5. Hacer un marco de papel de parafilm alrededor del gel.
6. Colocar sobre el gel un trozo de membrana de nylon previamente humedecido (primero con H₂O libre de RNAsa y después con 10x SSC) y eliminar las burbujas. La membrana debe llevar una pequeña marca en su extremo izquierdo, para identificar la cara de la transferencia.
7. Colocar sobre la membrana una pila de papel *Wattman* 3MM previamente humedecido con 10x SSC y del mismo tamaño que la membrana; eliminar las burbujas.
8. Encima de la anterior pila, colocar una nueva pila de 10 centímetros de papel absorbente y poner encima un peso de 500 g. Dejar transferir de 16-20 horas.
9. Desmontar el dispositivo de la transferencia y marcar los pocillos del gel en la membrana con una aguja.
10. Equilibrar la membrana en 2x SSC durante 5 minutos a temperatura ambiente en agitación.
11. Dejar secar la membrana a temperatura ambiente. Irradiar la membrana con luz ultravioleta para fijar la muestra (0,45 segundos/120 mJoules).

2.13.5. Prehibridación e hibridación

1. Hervir 200 μ L de DNA de esperma de salmón (agente bloqueante) durante 10 minutos y dejar en hielo. Añadir 20 mL de tampón de hibridación (*Perfect Hyb Plus* de Sigma H7033).
2. Lavar la membrana con 2x SSC antes de introducirla en el tubo de hibridación. Introducir el tubo con la membrana y la anterior solución de prehibridación en el horno de hibridación a 65°C durante 1 hora.
3. Hervir 100 μ L de DNA de esperma de salmón y la sonda durante 10 minutos. Hacer un pulso en la centrífuga y poner en hielo. Añadir a ambos 10 mL de tampón de hibridación.
4. Eliminar la solución de prehibridación y añadir la solución de hibridación. Dejar hibridar toda la noche a 65°C.

2.13.6. Posthibridación y detección

a) Lavados

Los lavados se realizan de menos a más astringentes. Para decidir que lavados aplicar, se debe ir comprobando la radiación de la membrana. Si sigue siendo muy radioactiva se van astringiendo los lavados.

Extraer la membrana del tubo de hibridación y enjuagar dos veces por agitación con 2x SSC/0,1% SDS. Lavar la membrana con 0,5x SSC/0,1% SDS a 65°C durante 20 minutos. Lavar la membrana con 0,1x SSC/0,1% SDS a 65°C durante 20 minutos. Para evitar que la membrana se seque y desprenda radioactividad, sellarla con film.

b) Exposición

Se puede hacer una primera aproximación de la exposición (*Phosphoimager*). Colocar las membranas boca arriba en una *cassette* con una pantalla reactiva boca abajo y en contacto con la membrana. Dejar tres horas y comprobar. Si hay señal, se guardan las membranas en una nueva *cassette* con un film de captación de radiación. Dejar tres días y revelar.

c) Reciclaje de la membrana

Las membranas son reciclables hasta un máximo recomendable de tres hibridaciones. La sonda hibridada puede eliminarse lavando la membrana con 0,1% SDS durante 20 minutos a temperatura de ebullición.

3. RESULTADOS

3.1. Inversiones $2m$ y $2n$

3.1.1. Notación de los puntos de rotura

Las regiones de los puntos de rotura de las inversiones de $2m$, $2n$ y $2z^7$ fueron nombradas de acuerdo con la nomenclatura utilizada en trabajos anteriores (Wesley y Eanes 1994; Cáceres *et al.* 1999). Así, las tres regiones de los puntos de rotura de las inversiones $2m$, $2n$ fueron designadas provisionalmente como AB, CD y EF en los genomas no invertidos y como AC, BE y DF en los genomas invertidos (Figura 3.1). En ambos casos, A representa la región situada más distalmente en relación con el centrómero mientras que F corresponde a la región más proximal. La inversión $2z^7$ ocurrió en un cromosoma invertido $2mn$ e invierte el punto de rotura compartido BE (Figura 3.1).

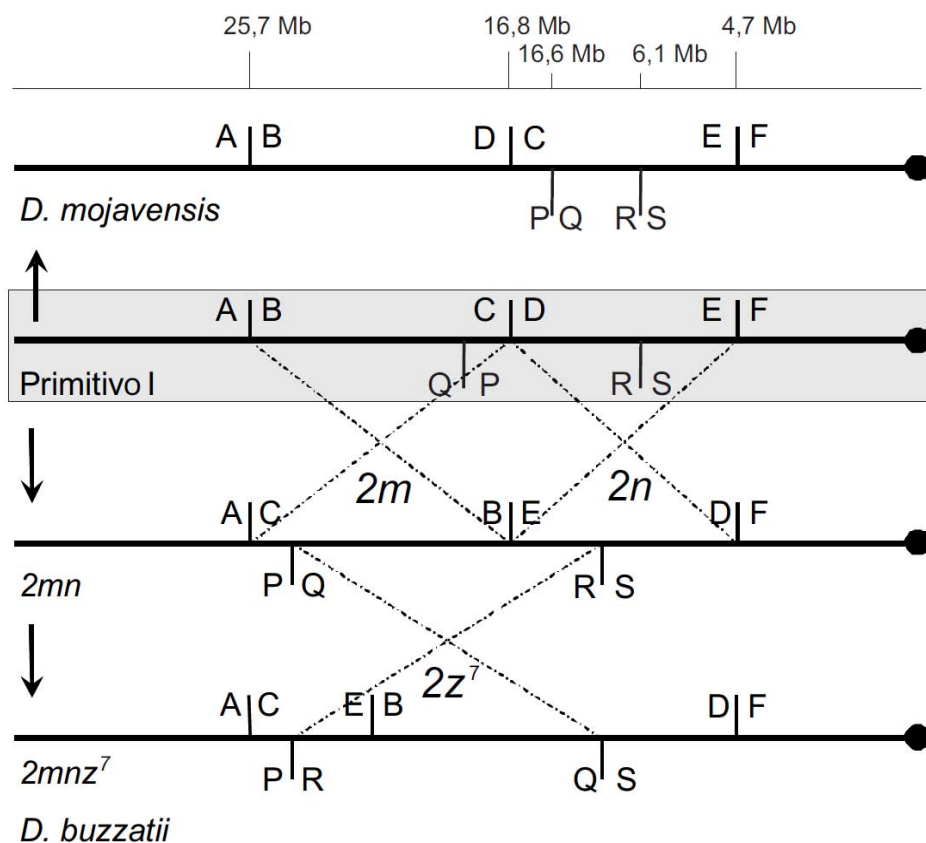


Figura 3.1: Vista esquemática de las inversiones del cromosoma 2 desde la ordenación del ancestro del grupo (Primitivo I) hasta *D. buzzatii*. También se muestra el orden de los puntos de rotura de las inversiones y su localización aproximada (Mb) en el *scaffold* 6450 de *D. mojavensis*.

3.1.2. Cartografía de los puntos de rotura en *D. mojavensis*

Existe una genoteca del genoma de *D. buzzatii* (especie con ambas inversiones). El mapa físico y la cartografía de los clones BAC de dicha genoteca fueron realizados por González *et al.* (2005). Tres clones BAC que contienen los puntos de rotura de las inversiones $2n$ y $2m$ fueron identificados por González *et al.* (2007). Los puntos de rotura fueron acotados en los clones BAC que los contienen para analizarlos. Para acotar los puntos de rotura, se cartografiaron y localizaron en la secuencia del genoma de *D. mojavensis* (Drosophila 12 Genomes Consortium 2007). El procedimiento fue similar al que ya se usó en Prazeres *et al.* (2009), que consta de cuatro etapas: (i) identificación de los clones BAC de *D. buzzatii* que contienen los puntos de rotura, (ii) secuenciación de los extremos de estos BAC y localización de las secuencias obtenidas en el genoma de *D. mojavensis* (este proceso permite acotar las regiones de los puntos de rotura entre los extremos de los clones), (iii) caracterización de los puntos de rotura para reducir la distancia inicial caminando a través del genoma de *D. mojavensis* (paseo cromosómico) y (iv) aislamiento de las regiones de punto de rotura en *D. buzzatii* por PCR.

(i) Los clones BAC 14B19, 22B03 y 16H04 abarcan las regiones de los punto de rotura AC, BE y DF, respectivamente, en *D. buzzatii* (González *et al.* 2007). Se aislaron de la genoteca otros 2 clones solapantes con cada uno de los 3 anteriores para trabajar con un total de tres clones BAC por punto de rotura. Como en González *et al.* (2007), se comprobó mediante hibridación *in situ* que también contuvieran los puntos de rotura. Un clon BAC del genoma de *D. buzzatii* produce una única señal cuando se hibrida en dicha especie, pero dos señales si contiene un punto de rotura y se hibrida en un genoma de *D. repleta* sin la inversión (véase Figura 1.5 en Introducción). De este modo, los clones 1N19, 14B19 y 15L19 contienen el punto de rotura AC, los clones 20O19, 22B03 y 2N19 contienen el punto de rotura BE y los

clones 16H04, 14E21 y 8C14 contienen el punto de rotura DF (Tabla 3.1).

ii) Se ha logrado secuenciar diez extremos de BAC (de los 18 posibles) y cartografiarlos con búsquedas de similitud en el genoma *D. mojavensis* (Tabla 3.2). Todas las secuencias han sido ancladas en el *scaffold* 6540 de *D. mojavensis* (~34 Mb de largo), correspondiente al cromosoma 2. La disposición de los extremos de los BAC en el *scaffold* nos permitió localizar las regiones de los puntos de rotura AB, CD y EF en *D. mojavensis* cerca de las coordenadas 25,7 Mb, 16,8 Mb y 4,7 Mb, respectivamente. Hay que tener en cuenta que en *D. mojavensis* el *scaffold* 6540 aumenta la numeración del centrómero a los telómeros (Schaeffer *et al.* 2008). Cada región del punto de rotura se redujo a un segmento específico acotado con la ayuda de al menos dos extremos anclados de dos BAC distintos. La región del punto de rotura AB se encontró dentro de los 152.933 pares de bases limitados por los extremos de BAC 22B03-T7 (25.671.752 pb) y 14B19-T7 (25.825.187 pb), la región del punto de rotura CD se encuentra dentro de los 91.453 pares de bases entre los extremos de los BAC 1N19-T7 (16.861.229 pb) y 14E21-SP6 (16.952.590 pb); la región del punto de rotura EF se encuentra dentro de los 111.770 pares de bases limitado por los extremos de los BAC 16H04-T7 (4.692.720 pb) y 20O19-SP6 (4.804.495 pb) (Figura 3.2).

Tabla 3.1: Lista de los clones BAC de *D. buzzatii* usados en este trabajo que contienen los puntos de rotura de las inversiones de $2m$ y $2n$. Estos clones hibridaron en una sola posición en el cromosoma 2 de *D. buzzatii* mientras que produjo dos señales de hibridación en el cromosoma 2 de *D. repleta*. Las coordenadas citológicas se refieren al mapa citológico del cromosoma 2 de *D. repleta* (Wharton 1942).

Punto de rotura	BAC^a	Tamaño (pb)^b	Coordenadas citológicas en <i>D. buzzatii</i>	Coordenadas citológicas en <i>D. repleta</i>
AC	1N19	138.197	D3e/F2a	D3e y F2a
AC	14B19	140.515	D3e/F2a	D3e y F2a
AC	15L19	133.077	D3e/F2a	D3e y F2a
BE	20O19	142.697	D3e/F6h	D3e y F6h
BE	22B03	128.945	D3e/F6h	D3e y F6h
BE	2N19	191.034	D3e/F6h	D3e y F6h
DF	16H04	124.185	F6h/F2a	F6h y F2a
DF	14E21	195.292	F6h/F2a	F6h y F2a
DF	8C14	156.485	F6h/F2a	F6h y F2a

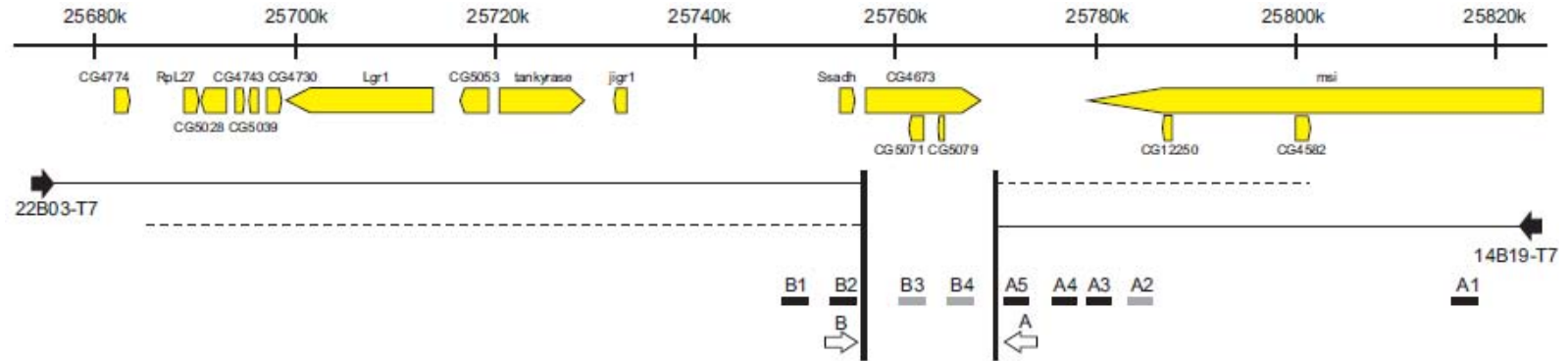
^a Los clones 14B19, 20O19 y 16H04 fueron hibridados por González *et al.* (2007).

^b Estimaciones de los tamaños de los insertos en los clones BAC tomadas del mapa físico de los clones BAC de *D. buzzatii* disponible en “Genome Sciences Centre” (<http://www.bcgsc.ca/ice>).

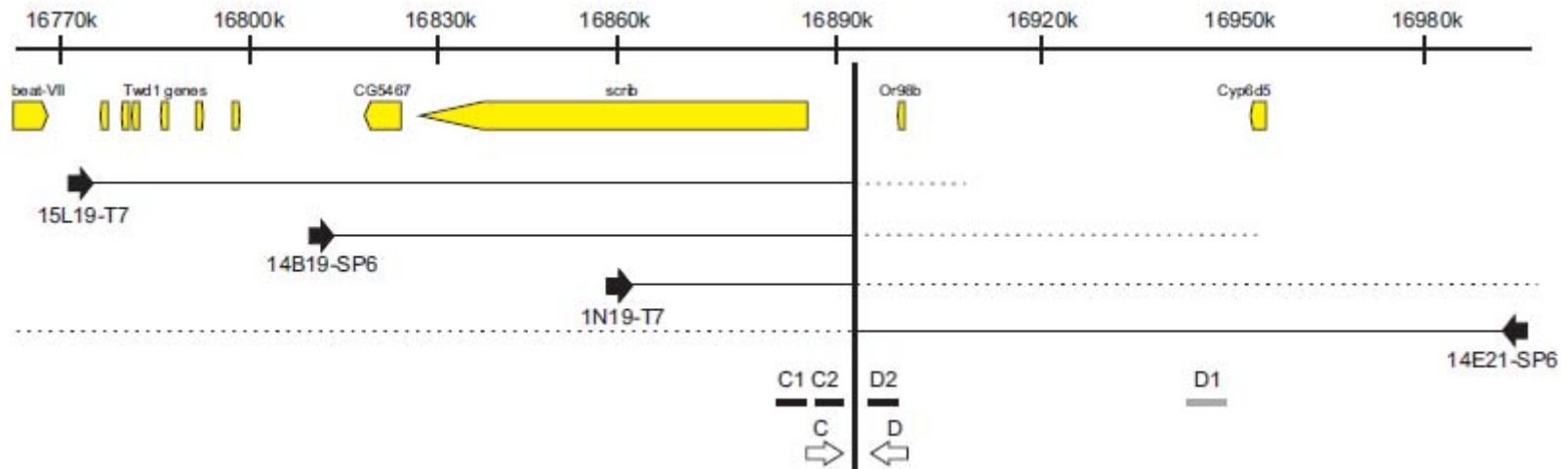
Tabla 3.2: Localización de los extremos de los clones BAC de *D. buzzatii* en el genoma de *D. mojavensis*. Todas las coordenadas del genoma de *D. mojavensis* pertenecen al *scaffold* 6540.

Punto de rotura		Coordenadas en <i>D. mojavensis</i>					
BAC	extremo	(pb)	Coordenadas	Anotación	E-value	Identidad (%)	
AC	14B19	T7	800	25.824.685-25.824.451	<i>msi</i>	2e-89	226/240 (94,1)
AC	1N19	T7	711	16.861.094-16.861.229	<i>scrib</i>	7e-33	116/135 (85,9)
AC	14B19	SP6	745	16.810.607-16.811.089	Intergénico (<i>Twd1-CG5467</i>)	2e-88	408/500 (81,6)
AC	15L19	T7	640	16.761.866-16.762.001	Intergénico (<i>beat-VII-Twd1</i>)	9e-87	203/219 (92,7)
BE	22B03	T7	569	25.671.558-25.671.752	<i>CG4774</i>	7e-47	169/200 (84,5)
BE	2N19	SP6	513	4.823.128-4.822.869	<i>slo</i>	7e-43	220/267 (82,4)
BE	20O19	SP6	578	4.804.621-4.804.495	<i>slo</i>	4e-56	123/127 (96,8)
DF	14E21	SP6	632	16.952.682-16.952.590	Intergénico (<i>Cyp6d5-beat-VI</i>)	1e-16	79/93 (84,9)
DF	16H04	T7	774	4.692.135-4.692.725	<i>CG9363</i>	0.0	535/592 (90,4)
DF	14E21	T7	841	4.644.326-4.644.563	Intergénico (<i>alpha-ManII-ps</i>)	3e-98	224/238 (94,1)

Región del punto de rotura AB



Región del punto de rotura CD



Región del punto de rotura EF

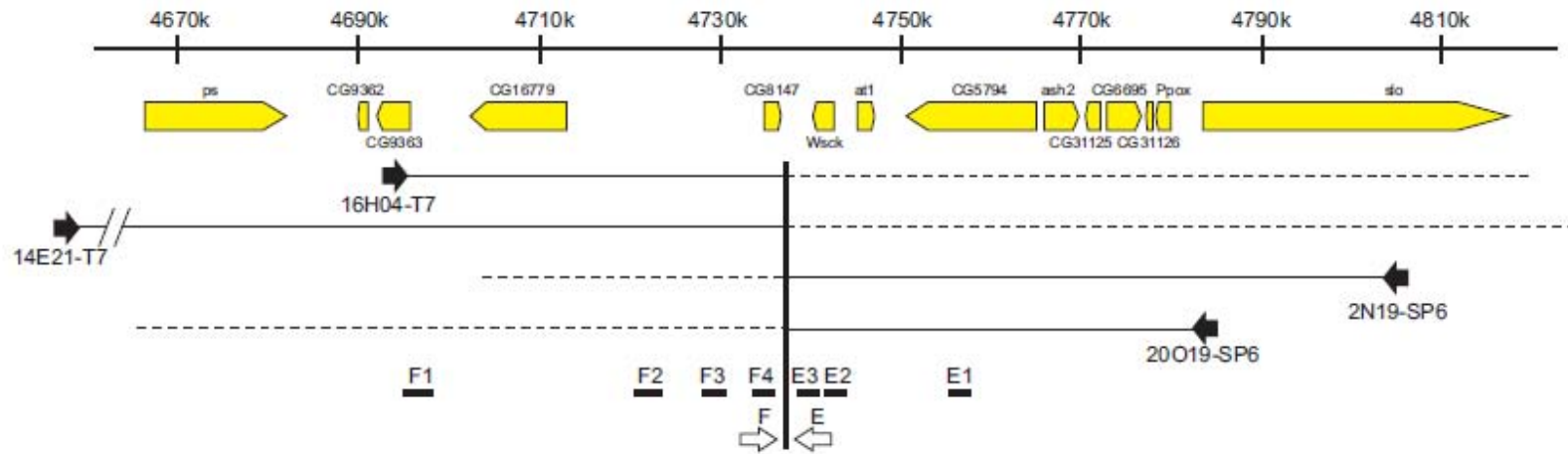


Figura 3.2: Vista esquemática del paseo cromosómico realizado para localizar los puntos de rotura de las inversiones $2m$ y $2n$. Las flechas negras muestran la localización de las secuencias de los extremos de los clones BAC. Las líneas discontinuas muestran la longitud de los clones BAC, aunque no se encuentran enteramente en estas regiones del genoma dado que están interrumpidos por los puntos de rotura. Sólo muestran la distancia máxima dónde los puntos de rotura pueden ser localizados, ya que dichos clones BAC los contienen. Los rectángulos negros corresponden a las PCR bien caracterizadas en *D. mojavensis*, mientras que los grises no produjeron señales de hibridación *in situ*. Las flechas blancas indican la situación de los diferentes cebadores diseñados para tratar de amplificar los puntos de rotura en *D. buzzatii*.

(iii) Con el fin de localizar con precisión el punto de rotura dentro de cada fragmento acotado en el *scaffold* 6540, se diseñaron distintas PCR a lo largo de los fragmentos de *D. mojavensis* acotados por los extremos anclados de los BAC. Estas PCR fueron amplificadas usando los BAC como molde y posteriormente utilizadas como sondas en hibridaciones *in situ* en cromosomas de *D. buzzatii* para determinar su ubicación. Las PCR, sus coordenadas en el genoma de *D. mojavensis* y su localización *in situ* se muestran en la Tabla 3.3. Las secuencias de los cebadores de estas PCR se muestran en la Tabla 2.1 de Materiales y Métodos. Este proceso nos permitió avanzar desde los extremos hasta la caracterización de los puntos de rotura en *D. mojavensis* (paseo cromosómico). La región del punto de rotura AB (punto de rotura distal de la inversión $2m$) se encuentra en la región entre las sondas B2 y A4 (16.190 pb). No se consideró inicialmente la sonda A5 porque a pesar de que la señal de la hibridación *in situ* parece estar localizada correctamente en el cromosoma se usó como sonda el producto amplificado en *D. mojavensis* (al no haber amplificado en *D. buzzatii*). La región del punto de rotura AB, se localizó entre los genes *msi* y *Ssadh* (17.800 pb). Entre ambos genes, está anotado el gen *CG4673*, que incluye dos genes anidados, *CG5071* y *CG5079* (véase Figura 3.2). Las sondas B3 y B4, diseñadas en el interior de este fragmento acotado no proporcionaron resultados concluyentes (Tabla 3.3) y no se tomaron en cuenta. No fue posible verificar si el punto de rotura está localizado aguas arriba, aguas abajo o interrumpiendo la unidad transcripcional del gen *CG4673*. La región del punto de rotura CD (punto de rotura compartido entre $2m$ y $2n$) se encuentra en el segmento de 3.297 pb entre las sondas de C1 y D2, en la región intergénica aguas arriba del gen *scrib* (C1) y aguas abajo del gen *Or98b* (D2). En este caso, tampoco se consideró inicialmente la sonda C2 por el mismo motivo que A5. Por último, el punto de rotura EF (punto de rotura proximal de la inversión $2n$) se encontró el segmento de 2.138 pb entre las sondas de F4 y E3, aguas abajo de los genes *CG8147* (F4) y *Wsck* (E3).

Tabla 3.3: PCR e hibridaciones *in situ* del paseo cromosómico desde los extremos de los clones BAC. Las reacciones de PCR se realizaron usando los propios clones BAC como molde. Véase Tabla 2.1 para las secuencias de los cebadores y la Figura 3.2 para la localización de las PCR. (-) indica que no se obtuvo producto de PCR. Las hibridaciones *in situ* se realizaron con el producto amplificado de las PCR como sondas.

* Las sondas para las hibridaciones *in situ* se obtuvieron de las PCR usando *D. mojavensis* como molde al no obtener producto con los BAC.

PCR	Coordenadas de los cebadores en <i>D. mojavensis</i>	Anotación	Tamaño de la PCR en <i>D. mojavensis</i> (pb)	Tamaño de la PCR usando el BAC 16H04 (pb)	Tamaño de la PCR usando el BAC 20O19 (pb)	Localización de la señal de hibridación en <i>D. buzzatii</i>
F1	4.693.132-4.695.292	<i>CG9363</i>	2.161	~2.100	-	F6h/F2a
F2	4.722.897-4.724.593	Intergénico (<i>CG16779-CG8147</i>)	1.697	~1.800	-	F6h/F2a
F3	4.730.053-4.730.865	Intergénico (<i>CG16779-CG8147</i>)	813	~820	-	F6h/F2a
F4	4.735.053-4.736.519	<i>CG8147</i>	1.467	~1.500	-	F6h/F2a
E3	4.738.657-4.739.571	Intergénico (<i>CG8147-Wsck</i>)	915	-	~1.200	D3e/F6h
E2	4.741.543-4.743.800	<i>Wsck</i>	2.258	-	-	D3e/F6h *
E1	4.755.646-4.757.683	<i>CG5794</i>	2.038	-	~2.000	D3e/F6h

PCR	Coordenadas de los cebadores en <i>D. mojavensis</i>	Anotación	Tamaño de la PCR en <i>D. mojavensis</i> (pb)	Tamaño de la PCR usando el BAC 1N19 (pb)	Tamaño de la PCR usando el BAC 16H04 (pb)	Localización de la señal de hibridación en <i>D. buzzatii</i>
D1	16.963.581-16.962.296	<i>Cyp6d5</i>	1.286	-	~1.500	Sin señal
D2	16.895.728-16.893.264	<i>Or98b</i>	2.465	-	~2.400	F6h/F2a
C2	16.891.713-16.889.967	Intergénico (<i>scrib-Or98b</i>)	1.747	-	-	D3e/F2a*
C1	16.890.058-16.887.461	<i>scrib</i>	2.598	-	~2.100	D3e/F2a
PCR	Coordenadas de los cebadores en <i>D. mojavensis</i>	Anotación	Tamaño de la PCR en <i>D. mojavensis</i> (pb)	Tamaño de la PCR usando el BAC 20O19 (pb)	Tamaño de la PCR usando el BAC 1N19 (pb)	Localización de la señal de hibridación en <i>D. buzzatii</i>
B1	25.752.005-25.753.543	Intergénico (<i>jigr1-Ssadh</i>)	1.539	~1.500	-	D3e/F6h
B2	25.757.303-25.758.924	<i>Ssadh</i>	1.622	~1.850	-	D3e/F6h
B3	25.761.704-25.764.947	<i>CG5071+CG5079</i>	3.244	-	-	Sin señal*
B4	25.768.418-25.769.198	<i>CG4673</i>	781	~750	~750	Sin señal*
A5	25.769.180-25.773.097	Intergénico (<i>CG4673-msi</i>)	3.918	-	-	D3e/F2a*
A4	25.775.114-25.776.743	Intergénico (<i>CG4673-msi</i>)	1.630	-	~3.500	D3e/F2a
A3	25.776.724-25.778.892	<i>msi</i>	2.169	-	~2.050	D3e/F2a
A2	25.783.641-25.786.366	<i>msi</i>	2.726	-	-	Sin señal *
A1	25.822.631-25.824.594	<i>msi</i>	1.964	-	~2.000	D3e/F2a

3.1.3. Aislamiento de los puntos de rotura en *D. buzzatii*

Una vez que las regiones de los puntos de rotura fueron acotadas en el genoma de *D. mojavensis* (AB, CD y EF) tratamos de aislar las tres regiones de los puntos de rotura de las inversiones $2m$ y $2n$ en *D. buzzatii* (AC, BE y DF). Distintos cebadores (véase Tabla 2.2 en Materiales y Métodos) fueron diseñados en el interior de las PCR caracterizadas y localizadas por hibridación *in situ* en *D. mojavensis* (flechas negras en la imagen detallada de la Figura 3.3). Los cebadores se diseñaron en las regiones conservadas entre *D. mojavensis* y *D. melanogaster*. Se espera que estas regiones también estén conservadas en *D. buzzatii*. Los cebadores diseñados en las regiones A y B de *D. mojavensis* se encuentran aguas abajo (cebador A) y aguas arriba (cebador B) del gen *CG4673*. De este modo, la amplificación de las regiones de los puntos de rotura en *D. buzzatii* incluirá dicho gen aún no caracterizado. Los cebadores fueron combinados según la orientación esperada en la Figura 3.1 para tratar de amplificar por PCR los puntos de rotura en *D. buzzatii*. Sólo el punto de rotura DF pudo ser amplificado con éxito dando lugar a un producto de PCR de ~4,4 Kb. Este producto corresponde a la región intergénica entre los genes *Or98b* (*CG1867*) y *CG8147*. Las regiones de los puntos de rotura AC y BE no pudieron ser amplificadas incluso después de varios intentos utilizando cebadores alternativos. Aparte de los cebadores usados inicialmente, se intentaron amplificar los puntos de rotura con los demás cebadores correctamente orientados (flechas blancas en la Figura 3.3) correspondientes a los cebadores que se usaron para fabricar las sondas (A4, A5, C1, C2, B2, E2 y E3). Se combinaron entre ellos y con los cebadores inicialmente diseñados (A, C, B y E). En ningún caso se consiguieron amplificar los puntos de rotura en *D. buzzatii*. Finalmente, se intentó localizar en *D. buzzatii* el gen *CG4673* de la región del punto de rotura AB de *D. mojavensis*. Existen dos posibilidades: (i) que el punto de rotura interrumpa la unidad transcripcional o (ii) que se encuentre en la región intergénica (ya

sea aguas arriba o aguas abajo del gen *CG4673*). Para estas PCR, se incluyeron los cebadores usados para amplificar las sondas B4 y B3 (Figura 3.3). Si la interrupción ocurre entre los genes *msi* y *CG4673*, las PCR B3 y B4 serán contiguas a B2 y E3. En el caso de que el punto de rotura AB se encuentre entre los genes *CG4673* y *Ssadh*, las PCR B3 y B4 serán contiguas a A4 y C1. Para comprobar qué PCR están contiguas a las PCR B3 y B4 se amplificaron PCR combinando los cebadores de este par de PCR con los cebadores de las posibles PCR contiguas. Ninguna de estas PCR para localizar B3 y B4 tuvo éxito. La falta de amplificación podría deberse a que las secuencias de los cebadores de las PCR B3 y B4 diseñados en *D. mojavensis* no estuvieran suficientemente conservados con la secuencia de *D. buzzatii*.

3.1.4. Electroforesis de campo pulsante e hibridación de Southern

Como método alternativo para aislar los puntos de rotura en *D. buzzatii* optamos por digerir los BAC y detectar los fragmentos que contuvieran los puntos de rotura restantes para ser clonados y secuenciados. Para llevarlo a cabo realizamos los siguientes pasos: (i) digerir los clones BAC con distintas enzimas de restricción, (ii) separar los fragmentos resultantes mediante electroforesis de campo pulsante (PFGE) y (iii) realizar una hibridación de Southern con las sondas bien caracterizadas y más cercanas a los puntos de rotura. El fragmento de la digestión que contenga el punto de rotura, será la banda del patrón de restricción del BAC que hibride con ambas sondas usadas aguas arriba y aguas abajo del punto de rotura. Por tanto, una banda que hibride con ambas sondas (a ambos lados del punto de rotura) no quedará interrumpida por la diana de restricción de dicha enzima. El campo pulsante funciona como los geles de agarosa pero pudiendo separar bandas de mayor peso molecular (véase Materiales y Métodos). Para facilitar la recuperación de dicha banda y evitar una secuenciación no relacionada, se priorizó la banda doblemente hibridada de menor tamaño.

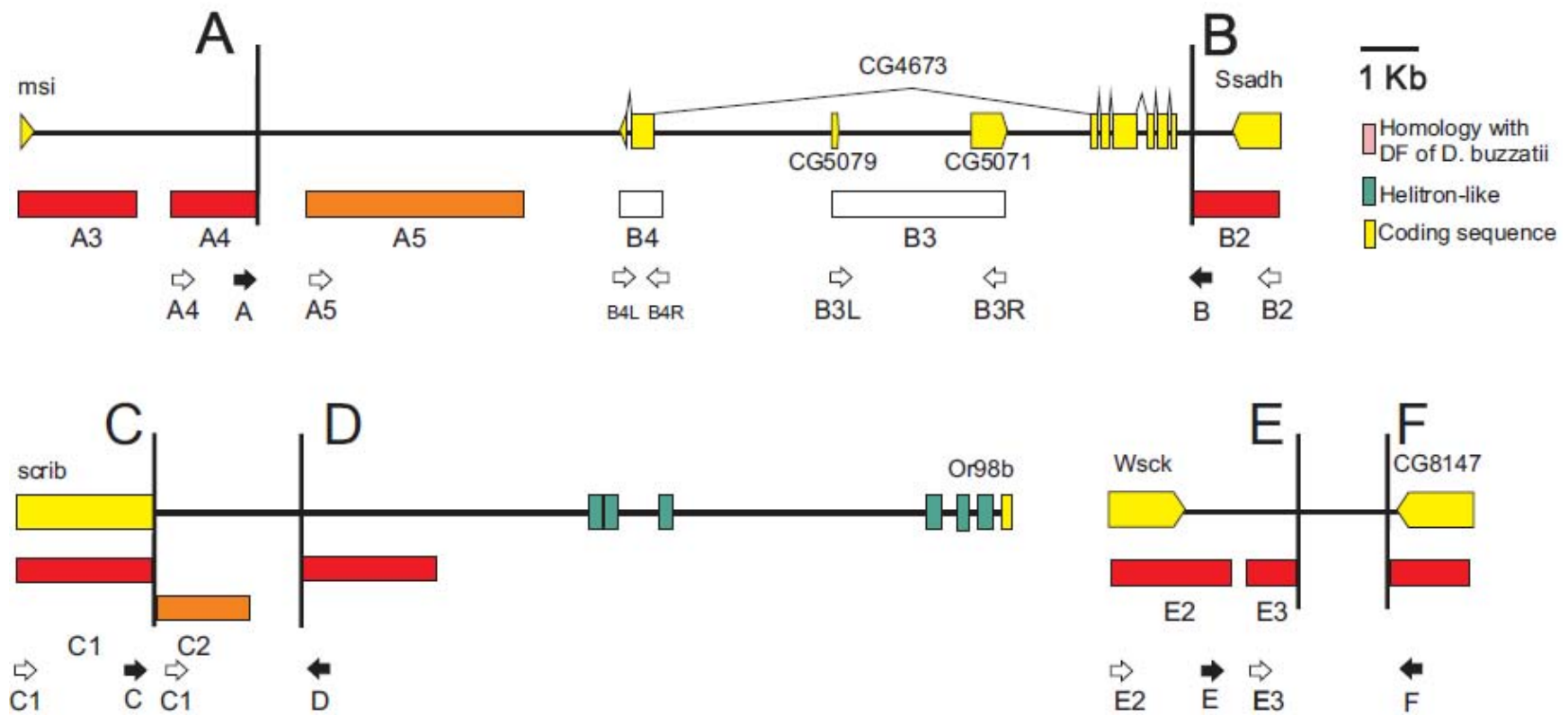


Figura 3.3: Vista detallada de las regiones de los puntos de rotura en *D. mojavensis*. Las PCR correctamente caracterizadas se muestran en rojo. Las PCR de color naranja no amplificaron en los clones BAC y las PCR de color blanco no produjeron señal de hibridación. Las flechas indican los cebadores de las PCR utilizadas en la cartografía en *D. mojavensis* (blanco) y para amplificar los puntos de rotura en *D. buzzatii* (negro). Las líneas verticales limitan las regiones de los puntos de rotura.

Sólo las sondas amplificadas e hibridadas con especificidad y bien caracterizadas en *D. buzzatii* tanto por PCR como por señal de hibridación *in situ* fueron usadas para la hibridación de Southern. El tamaño de la sonda no condicionó nuestra elección ya que la técnica de hibridación de Southern, a diferencia de la hibridación *in situ*, es suficientemente sensible.

Por un lado, se hibridaron las sondas A4 y C1 en las digestiones del BAC 1N19, y por otro, las sondas B2 y E3 en la digestiones del BAC 20O19. En CD, la sonda C2 está 1.655 pb más cerca del punto de rotura en *D. mojavensis*. Sin embargo, a pesar de que C2 se hibridara correctamente no amplificó en *D. buzzatii*. Por otro lado, entre las sondas B2 y A4 del punto de rotura AB de *D. mojavensis* seleccionadas para la hibridación en el Southern, se diseñaron otras tres sondas (A5, B3 y B4). Esta región corresponde al gen *CG4673* y las hibridaciones *in situ* no fueron concluyentes para localizar dicho gen en la región A o B. Aunque la distancia entre B2 y A4 es muy grande (16.190 pb), son las PCR más cercanas al punto de rotura y al gen *CG4673* (sin incluirlo). El resultado de la hibridación de Southern debería permitir determinar la localización de esta región correspondiente al gen *CG4673* y que aún está por localizar en el genoma de *D. buzzatii*.

En total, se realizaron 8 hibridaciones de *Southern*, 4 por cada BAC (Tabla 3.4). Para cada BAC, dos se hibridaron con la sonda aguas arriba y dos con la sonda localizada aguas abajo del punto de rotura. Así, dos hibridaciones de Southern corresponden a distintas digestiones del BAC 1N19 hibridado con la sonda A4 (Southern 1 y 5) y dos con la sonda C1 (2 y 6). Otras dos hibridaciones de Southern corresponden a distintas digestiones del BAC 20O19 hibridado con la sonda B2 (Southern 3 y 7) y dos con la sonda E3 (4 y 8). Para cada BAC hibridado con la misma sonda, se probaron 11 patrones de restricción distintos repartidos en los dos (12 en el caso de la sonda A4) más 4 controles negativos. El criterio de elección de las enzimas de restricción pretendía cubrir un amplio espectro de tamaños de los fragmentos. De este modo, se eligieron distintas enzimas de restricción con frecuencias de corte variadas.

Tabla 3.4: Hibridación de Southern. Para cada hibridación, se muestra la sonda y la enzima usadas para la digestión de los clones BAC en cada carril. Para cada hibridación se indica el tamaño (en Kb) de la banda hibridada con la sonda correspondiente. (M): Hibridación múltiple. (-) Sin hibridación. Los controles negativos se han realizado con la sonda correspondiente en la Tabla pero con el BAC que no la contiene.

Hibridación	1	2	5	6	3	4	7	8	
BAC	1N19				20O19				
Sonda	A4	C1	A4	C1	B2	E3	B2	E3	
Carriles	<i>1</i>	<i>BamHI</i>		<i>EagI+ApaI</i>		<i>BamHI</i>		<i>Clai</i>	
		17Kb	55Kb	14Kb	4Kb	14Kb	-	6Kb	4Kb
	<i>2</i>	<i>EagI</i>		<i>Clai</i>		<i>EagI</i>		<i>EcoRI</i>	
		14Kb	20Kb	M	M	12Kb	12Kb	M	5.5Kb
	<i>3</i>	<i>ApaI+ BamHI</i>		<i>EcoRI</i>		<i>ApaI</i>		<i>EcoRV</i>	
		17Kb	-	11Kb	M	40Kb	40Kb	22Kb	22Kb
<i>4</i>	<i>ApaI+SmaI</i>		<i>EcoRV</i>		<i>SmaI</i>		<i>PstI</i>		
	50Kb	28Kb	-	6Kb	50Kb	50Kb	7,5Kb	7,5Kb	
<i>5</i>	<i>SmaI+BamHI</i>		<i>PstI</i>		<i>ApaI+SmaI</i>		<i>SacI</i>		
	-		11Kb	17Kb	14Kb	14Kb	33Kb	33Kb	
<i>6</i>			<i>SacI</i>				<i>SalI</i>		
			M	21Kb			8Kb	8Kb	
Controles	<i>7</i>		<i>SalI</i>				<i>EagI</i> (1N19)		
			M	M			20Kb	-	
	<i>8</i>		<i>EagI</i> (20O19)						
			-	-					

El control negativo consistió en hibridar la sonda de un punto de rotura en un patrón de restricción del BAC que carecía de él. La elección de la enzima para el control negativo fue aleatoria puesto que no debe haber señal de hibridación. En este caso, se hibridó el BAC 20O19 (digerido con *EagI*) con las sondas A4 y C1 mientras que el BAC 1N19 (digerido con

EagI) se hibridó con las sondas B2 y E3. Los geles con el patrón de restricción y las membranas hibridadas de los Southern se muestran en la Figura 3.4.

No se han tomado en cuenta los patrones no hibridados ni con señales múltiples. Estos problemas suceden en algunas digestiones puntuales por lo que se descarto la inespecificidad o repetición a lo largo del BAC de alguna de las sondas.

AC: Los Southern 1 y 5 se hibridaron con la sonda A4 mientras que los Southern 2 y 6 se hibridaron con la sonda C1 en digestiones del BAC 1N19 (menos los controles, que se hibridaron con el 20O19). El 1 y el 2 por un lado y el 2 y el 6 por otro, comparten patrón de restricción. De sus 23 hibridaciones en el BAC1N19 se obtuvieron 14 señales únicas (7 con la sonda A4 y 7 con la sonda C1), 6 hibridaciones múltiples y en 3 casos no se obtuvo hibridación. Sólo se pudieron comparar 5 casos (donde se obtuvo hibridación única para ambas sondas en el mismo patrón de digestión). Ninguna banda fue hibridada por ambas sondas en un mismo patrón de restricción. En el octavo carril de los Southern 5 y 6 se hibridaron las mismas sondas pero en una digestión del BAC 20O19 (controles negativos). En ningún caso hubo señal de hibridación.

Al no obtener una única banda de alguno de los patrones de restricción que hibridara con ambas sondas (y que contuviera el punto de rotura), se seleccionaron aquellas bandas hibridadas de menor tamaño. De este modo, deberíamos recuperar ambos fragmentos, secuenciarlos y, posteriormente comprobar que fueran correlativos mediante una PCR que los uniera. De este modo, se decidió recuperar las bandas hibridadas de 4 Kb y 14Kb del patrón obtenido con la doble digestión de *EagI* y *ApaI* (carril 1 de los Southern 5 y 6).

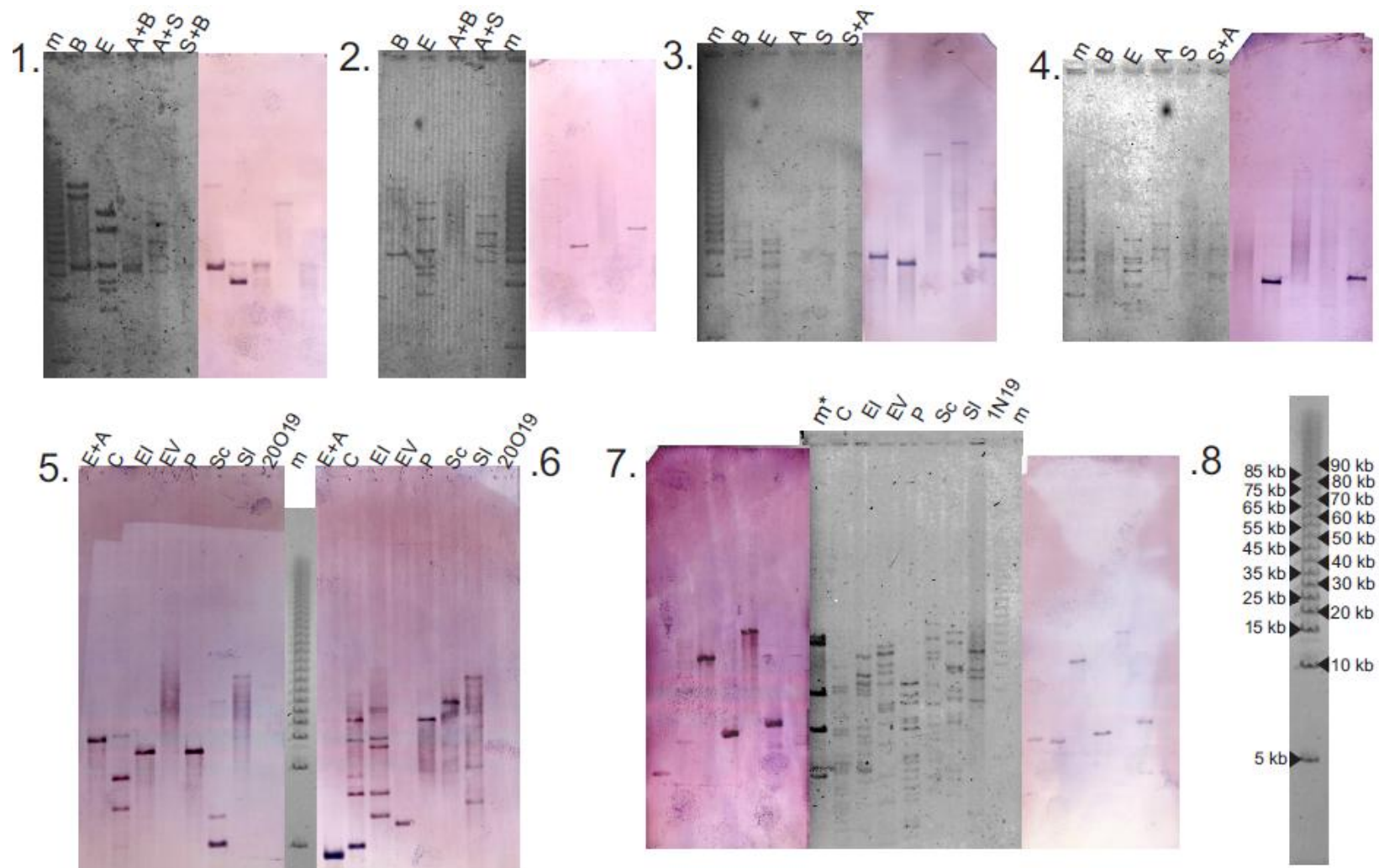


Figura 3.4: Hibridación de Southern llevadas a cabo para analizar los puntos de rotura AC y BE. Se muestran tanto los patrones de restricción obtenidos de la digestión de los clones BAC como las membranas resultantes de la hibridación (véase texto para más detalles). El BAC digerido, la sonda utilizada, la enzima de restricción y el tamaño de la banda hibridada se especifican en la Tabla 3.4. El tamaño (Kb) de las bandas del marcador también se muestra en esta Figura (derecha y abajo). m: marcador de Southern, m*: marcador lambda *HindIII*, B: *BamHI*, S: *SmaI*, E: *EagI*, A: *ApaI*, C: *ClaI*, EI: *EcoRI*, EV: *EcoRV*, P: *PstI*, Sc: *SacI*, SI: *SalI*.

BE: Los Southern 3 y 7 se hibridaron con la sonda B2 mientras que los Southern 4 y 8 se hibridaron con la sonda E3 en digestiones del BAC 20O19 (menos los controles, que se hibridaron con el 20O19). El 3 y el 4 por un lado y el 7 y el 8 por otro, comparten patrón de restricción. De sus 22 hibridaciones, se obtuvieron 20 señales únicas (11 con la sonda B2 y 9 con la sonda E3), pudiendo comparar 9 patrones de restricción (de los que se obtuvo señal única para ambas sondas). En este caso, en 8 de los 9 patrones hubo señal en la misma banda cuando se hibridó con las dos sondas. La banda de menor tamaño hibridada con las dos sondas, es de 7,5 Kb correspondiente a la digestión del BAC con la enzima *PstI* (carril 3 de los Southern 7 y 8). Este fragmento de digestión contiene la secuencia de las sondas E3 y B2 y por tanto el punto de rotura BE.

También se observó una señal de hibridación débil en uno de los controles negativos. La hibridación con la sonda B2 en el patrón de restricción del BAC 1N19 digerido con la enzima *EagI* produjo una señal débil de aproximadamente 20 Kb. Se trata de la misma banda que cuando se hibridó el mismo patrón de restricción pero con la sonda C1 (Southern 2). La sonda B2 corresponde a un fragmento de la región intergénica de los genes *CG4673* y *Ssadh* de *D. mojavensis*, originalmente caracterizada en el punto de rotura BE de *D. buzzatii*. Esta banda también se seleccionó.

Las bandas seleccionadas de las digestiones fueron recuperadas del gel de agarosa. Para comprobar que las bandas recuperadas fueran las hibridadas en el Southern, se realizó una batería de PCR. Las bandas recuperadas se usaron como molde de las PCR. Los cebadores se usaron los mismos que se usaron para obtener la sonda. Todas las PCR amplificaron correctamente y con el tamaño esperado. Por otro lado, también amplificó, aunque con mucho menos producto, el control negativo del Southern. Es decir, que la PCR B2, además de con la banda recuperada de 7,5 Kb de la digestión del BAC 20019 con *PstI*, también amplificó usando el fragmento de 20 Kb de la digestión del BAC 1N19 con *EagI*.

Nos propusimos clonar y secuenciar estas bandas recuperadas, aunque finalmente, la cantidad de producto recuperada del gel del campo pulsante fue suficiente para una PCR pero no para una ligación. Además, aunque se intentó con distintos vectores para fragmentos de gran tamaño, los fragmentos seleccionados resultaron demasiado grandes para el proceso de clonación.

Aunque el objetivo de clonar y secuenciar las bandas seleccionadas del gel de agarosa del campo pulsante no pudo materializarse, nos fue igualmente informativo.

Las sondas hibridadas en los patrones de restricción de los BAC digeridos señalan la banda que la contienen. En el punto de rotura AC, para todos los patrones de restricción, las dos sondas (A4 y C1) hibridan en bandas distintas. En cambio, en ocho de las nueve digestiones del BAC 20O19 que se pudieron comparar para el punto de rotura BE, ambas sondas (B2 y E3) hibridaron en la misma banda. A priori, se puede concluir que, en *D. buzzatii*, la distancia entre las sondas B2 y E3 es mucho menor de la que separa las sondas A4 y C1.

Concretamente, podemos determinar que, la banda hibridada con ambas sondas más pequeña nos indicará la distancia máxima que las puede separar. Para el punto de rotura BE, la banda hibridada con ambas sondas más pequeña se obtuvo de la digestión del BAC 20O19 con *Pst*I (~7,5 Kb). Si consideramos que el tamaño de las sondas B2 y E3 es 1.622 pb y 915 pb respectivamente y se lo deducimos a la banda de ~7,5 Kb, obtenemos que la distancia que las separa debe ser de ~5 Kb como máximo.

Por otro lado, nos es imposible determinar la distancia que separa las sondas A4 y C1 en *D. buzzatii*, aunque a la luz de los resultados del Southern, nos resignamos a que se tratará de un trozo bastante extenso.

Con esta información, volvemos a la ubicación del punto de rotura en el segmento AB en *D. mojavensis*. Hemos visto que la región del punto de rotura BE en *D. buzzatii* no es

mayor a ~5 Kb. Por lo tanto, la mayor parte de las 16 Kb que incluyen el gen *CG4673* y que separan las sondas A4 y B2, no pueden localizarse en este punto de rotura. En cambio, si este gen se localizase en la región A en *D. mojavensis*, lo haría en el punto de rotura AC en *D. buzzatii*. La distancia esperada entre las sondas A4 y C1 en *D. buzzatii* es consistente con la idea de que este gen se encuentre en dicha región. En estas 16 Kb, al menos existiría una diana para cada una de las enzimas de restricción utilizadas en las digestiones. Por esta razón, las sondas A4 y C1 siempre hibridarían en bandas distintas en el mismo patrón de restricción del BAC 1N19.

Finalmente, podemos considerar que, si el punto de rotura (BE) es, como máximo de ~5 Kb de longitud, la coincidencia observada citológicamente de las inversiones $2m$ y $2n$ también lo es molecularmente. Esta distancia, que aún puede ser menor, es imposible de detectar con técnicas citológicas convencionales, por lo que se justifica plenamente la elección de estas dos inversiones para abordar el problema de la reutilización de los puntos de rotura. La posibilidad que ambas inversiones sean independientes o solapantes es nula o en una fracción de secuencia muy pequeña.

Posible duplicación

Por otro lado, no se esperaba señal alguna en los controles negativos de las hibridaciones de Southern. Al hibridar la sonda B2 con la digestión del BAC 1N19 se observó una señal débil (Figura 3.4). La banda de 20 Kb hibridada fue la misma que hibridó cuando se usó la sonda C1 en el mismo patrón de digestión con el mismo BAC. Por lo tanto, la sonda B2 se encontraría cerca de la sonda C1, o por lo menos, a menor distancia de lo que se encuentra la sonda A4. Si fuera una duplicación, podría estar implicada en las inversiones. Se trataría de una fracción duplicada representada por la sonda B2, aunque, por el momento, desconozcamos

la extensión de dicha repetición. Además, otras dos sondas próximas (B3 y B4 diseñadas en el gen *CG4673*) no pudieron ser correctamente caracterizadas, y podrían estar también implicadas en la duplicación. Las duplicaciones son habituales en las inversiones. Algunos trabajos de análisis molecular de puntos de rotura han descrito duplicaciones en *Anopheles* (Sharakhov *et al.* 2006) y en *Drosophila* (Matzkin *et al.* 2005, Ranz *et al.* 2007). En estos casos, la duplicación parcial o total de estos genes próximos a los puntos de rotura está acompañada de la pseudogeneización de una de las copias. En nuestro caso, aun si el fragmento se ha duplicado, la debilidad en la señal indicaría menor similitud de la sonda B2 con el BAC 1N19 que en comparación con el BAC 20O19. Lógicamente, la similitud será menor si la duplicación ha originado un pseudogen, menos constreñido y, por tanto, más degenerado y con menor similitud al usar la sonda del gen funcional.

Contra esta posible explicación de los resultados, debemos puntualizar que éstos podrían deberse a causas distintas a una duplicación. Simplemente, el gen *CG4673*, podría formar parte de una familia de genes o tener estructura similar a otro gen que se encuentre en el BAC 1N19. Observaríamos menor señal al no tratarse del mismo gen a pesar de compartir regiones homólogas. La amplificación de la PCR B2 usando el BAC 1N19 se explicaría si los cebadores se hubieran diseñado en la región homóloga entre ambos genes.

3.1.5. Secuenciación de los clones BAC 1N19 y 20O19

El último recurso que nos permitía el análisis de los puntos de rotura de las inversiones $2m$ y $2n$ consistía en la secuenciación completa de los clones BAC, 1N19 y 20O19, que contienen los puntos de rotura AC y BE respectivamente (véase la Tabla 2.5 en Materiales y Métodos para los detalles de la secuenciación de los clones BAC). Para la secuenciación de cada clon, se construyó una subgenoteca con fragmentos de entre 2 y 5 Kb procedentes de

digestiones parciales con una cobertura aproximada del 8x. El ensamblaje de las lecturas de secuenciación permitió obtener finalmente dos *contigs* para el BAC 1N19 (*contig* 6 de 95.364 pb y *contig* 7 de 43.278 pb) y tres para el BAC 20O19 (*contig* 3 de 33.718 pb, *contig* 4 de 35.564 pb y *contig* 5 de 73.711 pb). Los 5 *contigs* fueron comparados y localizados en el genoma de *D. mojavensis* (Tabla 3.5). Los *contigs* del BAC 20O19 quedan ordenados como 4-3-5, siendo este último el que contiene el punto de rotura BE. En *D. mojavensis*, los huecos entre las regiones homólogas a los extremos de los *contigs* son de 156 pb (*contigs* 4 y 3) y 628 pb (*contigs* 3 y 5). Se diseñaron cebadores en los extremos de los *contigs* para unirlos (Tabla 2.1), los cebadores 4CL y 3CR para cerrar el espacio entre los *contigs* 4 y 3 y los cebadores 3CL y 5CR para cerrar el espacio entre los *contigs* 3 y 5. En el BAC 1N19 el *contig* 6 contiene el punto de rotura. En *D. mojavensis*, el hueco entre las regiones homólogas a los extremos de los *contigs* 6 y 7 es de 112 pb. También aquí se diseñaron cebadores para unir ambos *contig* (cebadores 6CL-7CR).

Las tres PCR para unir los *contigs* fueron secuenciadas. Dichas secuencias fueron alineadas con los extremos de los *contigs* para ensamblarlos y cerrar los huecos. De este modo, los insertos de los clones BAC de *D. buzzatii* son 138.724 pb (1N19) y de 143.293 pb (20O19). Su tamaño real se aproxima bastante a la estimación inicial (véase Tabla 3.1). En total supone haber secuenciado 282.818 pb del genoma de *D. buzzatii*. La región homóloga en *D. mojavensis* es unas 14 Kb mayor (véase Tabla 3.5). Al menos esta región del genoma de *D. buzzatii* sería un ~5% menor. El tamaño del genoma y el elemento E de Muller de *D. mojavensis* es de 193,8 Mb y 34,148 Mb, respectivamente. Con la proporción calculada para el contenido de los clones BAC, podemos estimar que el tamaño del cromosoma 2 de *D. buzzatii* será de ~32,44 Mb ($34,148 \times 95/100$) y su genoma de ~184,11 Mb ($193,8 \times 95/100$).

Tabla 3.5: Localización por similitud de los *contigs* de los BAC de *D. buzzatii* secuenciados en el genoma de *D. mojavensis*. Todas las coordenadas del genoma de *D. mojavensis* pertenecen al *scaffold* 6540. Los *contigs* 6 y 5 contienen los puntos de rotura AC y BE respectivamente, por lo que se especifica qué fragmento del *contig* se localiza en cada región del genoma de *D. mojavensis*. El Total incluye los tamaños de las PCR para cerrar los huecos entre *contigs*.

BAC	Contig	Tamaño (pb)	Localización por similitud con <i>D. mojavensis</i> (pb)	Tamaño equivalente en <i>D. mojavensis</i> (pb)
1N19	6	95.364	1-24.805:16.861.025-16.891.876	30.851
			25.270-95364: 25.757.536-25.825.832	68.297
	7	43.278	25.825.944-25.872.853	46.909
	Total	138.724	16.861.025-25.872.853	146.168
20O19	4	33.718	25.679.832-25.715.958	36.126
	3	35.564	25.716.114-25.750.958	34.844
	5	73.711	1-9.796: 25.769.312-25.756.520	17.726
			11.030-73.711: 4.738.649-4.805.028	66.379
	Total	143.293	25.679.832-4.805.028	155.859

3.1.6. Anotación y análisis de los clones BAC 1N19 y 20O19

Los dos clones BAC secuenciados fueron anotados usando los programas de predicción de genes GENEID y GENSCAN. Además, se usó como referencia el genoma de *D. mojavensis* (que es el genoma secuenciado y anotado más próximo a *D. buzzatii*). También se usó el genoma *D. melanogaster*, del cual se dispone de una anotación más detallada y actualizada así como de evidencia experimental procedente de datos de expresión.

Una comparación de la organización de genes de los dos clones BAC con las regiones homólogas en *D. mojavensis* revela que la mayoría de los genes poseen una estructura similar y misma orientación en las dos especies (con la excepción es *CG4673*, véase más adelante). Cada clon contiene un punto de rotura único que se encuentra en la región esperada inferida en la cartografía de alta resolución realizada con anterioridad. El contenido de los genes se muestra en la Figura 3.5. Para facilitar la comprensión de dicha Figura, el clon 1N19 se dispone en orientación CA.

1N19: Un total de 5 genes fueron anotados en este BAC, tres completos (*CG5071*, *CG12250* y *CG4582*), los primeros 5 exones del gen *scribbled* (*scrib*, *CG4562*) y los exones del 2 al 5 del gen *musashi* (*msi*, *CG5099*) (Figura 3.5). En *D. melanogaster*, el gen *msi* codifica dos transcritos, RA y RB. El BAC 1N19 de *D. buzzatii* contiene los tres exones (3-5) que codifican el transcrito RA y cuatro de los cinco exones (2-5) que codifican el transcrito RB. El exón 1 (RB) no queda incluido en este BAC. La organización compleja de los genes *msi*, *CG12250* y *CG4582* parece conservada entre *D. buzzatii*, *D. melanogaster* y *D. mojavensis* (aunque la anotación actual de este gen en *D. mojavensis* no lo muestra). El gen *CG4582* está totalmente anidado en el intrón 2 del gen *msi* mientras que algunos exones del gen *CG12250* quedan intercalados con los de *msi*. En la actual anotación del genoma *D. mojavensis*, el corto exón 1 (63 pb) del gen *CG12250* no está presente (el modelo génico carece del codón de inicio ATG), dando la falsa impresión de que este gen está completamente anidado dentro del intrón 2 del gen *msi*. Sin embargo, hemos detectado y anotado aguas abajo del codón stop de *msi*, el exón inicial putativo del gen *CG12250* que se conserva entre *D. mojavensis* y *D. buzzatii*. Esta reanotación de la estructura de estos genes, reubicaría el punto de rotura AB entre los genes *CG12250* (en substitución del gen *msi*) y *Ssadh*.

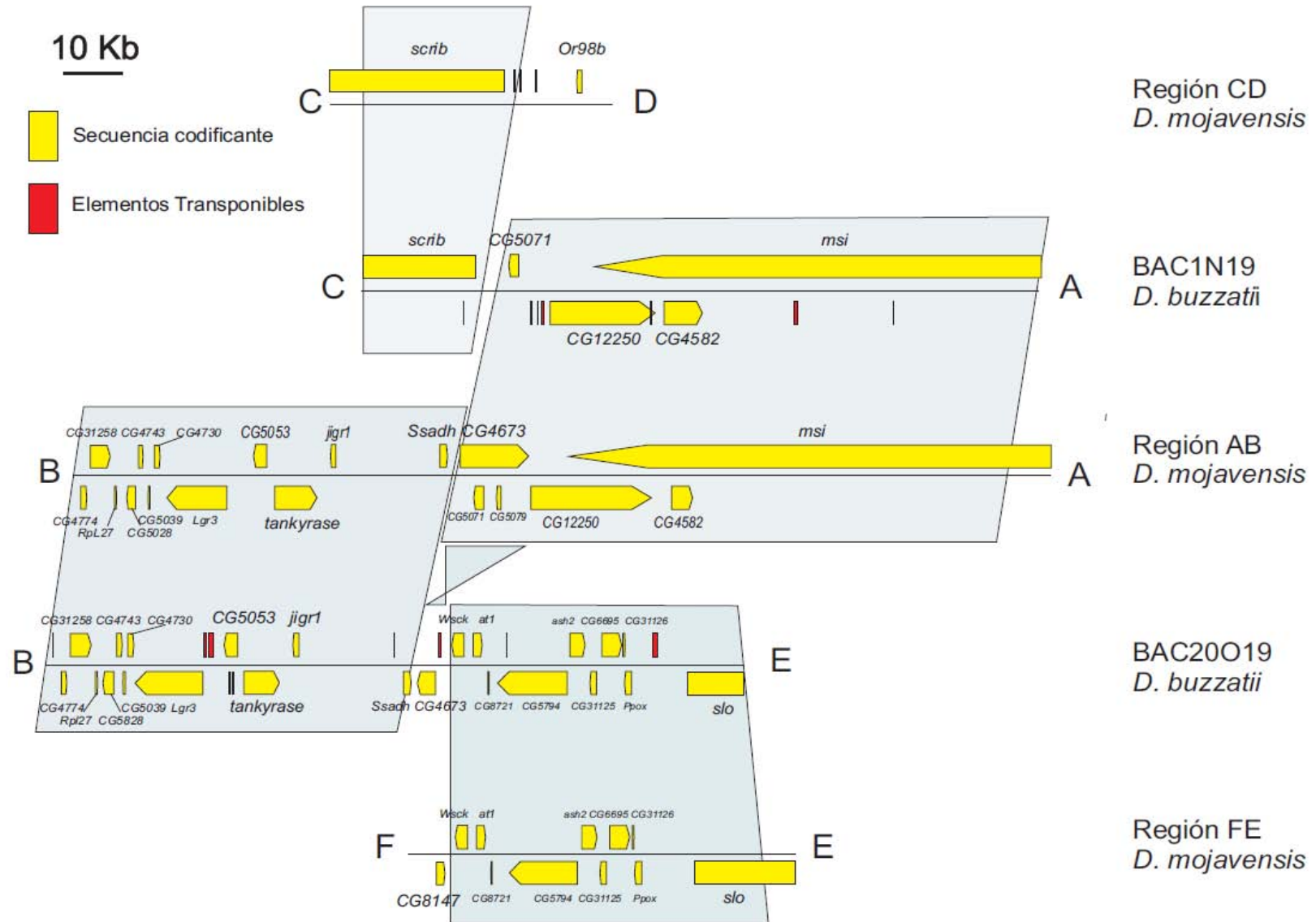


Figura 3.5: Comparación de los dos clones BAC secuenciados y anotados en *D. buzzatii* (1N19 y 20019) con las regiones homólogas del scaffold 6540 de *D. mojavensis*.

20019: Este BAC contiene las secuencias codificantes completas de 22 genes: *CG4774*, *CG31258*, *Rpl27* (*CG4743*), *CG5828*, *CG4743*, *CG5039*, *CG4730*, *Lgr3* (*CG31096*), *CG5053*, *tankyrase* (*CG4719*), *jigr1* (*CG17383*), *CG4585*, *CG4673*, *Wsck* (*CG31127*), *At1* (*CG6668*), *CG8721*, *CG5794*, *ash2* (*CG6677*), *CG31125*, *CG6695*, *CG31126* y *Ppox* (*CG5796*). Además, contiene los 7 primeros exones del gen *slowpoke* (*slo*, *CG10693*) un gen de 45.122 pb de longitud con 28 exones anotados en *D. mojavensis* (Figura 3.5).

La región localizada entre los genes *Ssadh* y *CG12250* en *D. mojavensis* (gen *CG4673* y sus genes anidados *CG5071* y *CG5079*) parece estar repartida entre ambos BAC. El gen *CG4673* está localizado en el BAC 20019 en *D. buzzatii*. En cambio, el gen anidado *CG5071* y un fragmento con similitud con el también gen anidado *CG5079* fueron anotados en el BAC 1N19 (ver más adelante).

También se realizó una búsqueda de elementos transponibles (TE) en ambos clones BAC. En esta sección, por coherencia con la Figura 3.5, el BAC 1N19 se enumera en orientación CA. En la Tabla 3.6 se muestra la anotación del clon BAC 1N19 y en la Tabla 3.7 se muestra la anotación del clon BAC 20019. En estas Tablas se muestra el tipo de elemento, su localización y la base de datos que nos permitió su anotación. Además, también se especifica la cantidad de pares de bases correspondiente a región putativamente transcrita así como el CDS. La información de estas Tablas se muestra en ventanas de 5 Kb. El análisis de la densidad de transcripción, CDS y TE, se muestra gráficamente en la Figura 3.6. El punto de rotura CA (región entre los genes *scrib* y *CG12250*) se localiza entre las ventanas 20-25 Kb y 35-40 Kb, coincidiendo con la menor densidad de transcripción. El pico de CDS dentro de esta sección de menor densidad de transcripción corresponde al gen *CG5071* (véase Figura 3.5). En cuanto al punto de rotura BE (región entre los genes *Ssadh* y *Wsck*) se localiza entre las ventanas 70-75 Kb y 80-85 Kb. En este caso, el pico de CDS dentro del punto de rotura corresponde al gen *CG4673*.

En la Tabla 3.8 se agrupa la información para las distintas regiones de cada clon BAC. Las densidades de TE se dividen en los que se encuentran aguas arriba del punto de rotura, aguas abajo del punto de rotura y dentro de éste (véase Figura 3.1). De este modo, los TE del clon BAC 1N19 se agrupan en los que están fuera de la inversión $2m$ (región A, aguas abajo del punto de rotura distal), dentro de ella (región C aguas arriba del punto de rotura proximal) y punto de rotura AC. Los TE del clon BAC 20O19 se dividen en los que están aguas abajo del punto de rotura compartido (región B), aguas arriba del punto de rotura compartido, y los del punto de rotura BE. Además de la densidad de pares de bases de cada región que ha sido anotada como TE, se especifica la proporción que ha sido anotada en regiones intergénicas y en regiones intrónicas. Sólo las regiones C y A (ambas del mismo punto de rotura) tienen mayor proporción de TE anotados en regiones intrónicas que en las intergénicas.

Las regiones C y B pertenecen a la inversión $2m$ y presentan una densidad de TE de 0,30% y 2,78%, respectivamente. Si calculamos la densidad conjuntamente, observamos que dentro de la inversión $2m$, la densidad de TE es del 2,19%. Esta densidad es mayor que la observada en el interior de la inversión $2n$, representada por la región E (1,65%). Ambas inversiones no tienen una densidad muy distinta de la observada fuera de las inversiones. La única región fuera de las inversiones es la región A (delimita la inversión $2m$ en su región distal) con una densidad del 1,67%. En las regiones de los puntos de rotura se observa la mayor densidad de TE (5,92% en AC y 6,33% en BE).

El mismo análisis de densidad se ha calculado para los pb correspondientes a CDS y al transcrito conceptual. Las regiones con menor proporción de pb transcritos corresponden a los fragmentos de los puntos de rotura (13,6% en AC y 44,90% en BE), aunque no la proporción de CDS. Este efecto se debe a que ambas regiones de los puntos de rotura, tienen un gen muy compactado. En AC, el gen *CG5071* está codificado en un único exón. En BE, el gen *CG4673* está codificado en 8 exones y se encuentra muy compactado debido a la pérdida del intrón 6.

Tabla 3.6: Anotación del clon BAC 1N19. La información se divide en ventanas de 5 Kb. Para cada ventana, se indica cuántos pares de bases corresponden a elementos transponibles (TE) anotados, cuántos corresponden a CDS y cuántos a transcrito. La anotación de los TE incluye su localización en el *contig* y la base de datos que los ha permitido anotar (RepeatMasker: RM).

Ventana	TE (pb)	Nombre	Base de datos	Localización		Tamaño (pb)	CDS (pb)	Transcripción (pb)
1-5kb	0						634	5.000
5-10Kb	0						583	5.000
10-15Kb	0						118	5.000
15-20Kb	0						0	5.000
20-25Kb	70	<i>TART</i>	RM	20.784	20.853	70	162	3.072
25-30Kb	155	<i>BuT5</i>	NCBI	25.138	25.198	61	175	175
		<i>BuT5</i>	NCBI	25.250	25.289	40		
		<i>BuT5</i>	NCBI	25.486	25.539	54		
30-35Kb	273	<i>Helitron</i>	RM	32.564	32.634	71	1.829	1.829
		<i>Helitron</i>	RM	32.683	32.790	108		
		<i>ISBu</i>	DPDB	33.557	33.650	94		
35-40Kb	480	<i>Jockey</i>	RM	36.035	36.103	69	63	1.585
		<i>Jockey</i>	RM	36.163	36.234	72		
		<i>ISBu</i>	DPDB	37.053	37.392	340		
40-45Kb	0						0	5.000
45-50Kb	80	<i>TART</i>	RM	47.862	47.941	80	1.239	5.000
50-55Kb	129	<i>TART</i>	RM	53.577	53.705	129	777	5.000

55-60Kb	235	<i>ISBu</i>	DPDB	55.200	55.264	65	414	5.000
		<i>Helitron</i>	RM	59.326	59.368	43		
		<i>Helitron</i>	RM	59.394	59.520	127		
60-65Kb	0						214	5.000
65-70Kb	0						1.502	5.000
70-75Kb	0						0	5.000
75-80Kb	0						0	5.000
80-85Kb	0						0	5.000
85-90Kb	594	<i>ISBu</i>	DPDB	89.160	89.298	139	0	5.000
		<i>ISBu</i>	DPDB	89.354	89.808	455	0	5.000
90-95Kb	53	<i>ISBu</i>	DPDB	90.569	90.621	53	45	5.000
95-100Kb	0						0	5.000
100-105Kb	0						0	5.000
105-110Kb	193	<i>TART</i>	RM	106.424	106.467	45	0	5.000
		<i>Helitron</i>	RM	108.643	108.790	148		
110-115Kb	0						0	5.000
115-120Kb	0						0	5.000
120-125Kb	0						0	5.000
125-130Kb	0						0	5.000
130-135Kb	0						0	5.000
135-140Kb	0						0	5.000

Tabla 3.7: Anotación del clon BAC 20O19. La información se divide en ventanas de 5 Kb. Para cada ventana, se indica cuántos pares de bases corresponden a elementos transponibles (TE) anotados, cuántos corresponden a CDS y cuántos son transcrito putativo. La anotación de los TE incluye su localización en el *contig* y la base de datos que los ha permitido anotar (RepeatMasker: RM).

Ventana	TE (pb)	Nombre	Base de datos	Localización		Tamaño (pb)	CDS (pb)	Transcripción (pb)
1-5kb	66	<i>Mini-Me</i>	RM	1.072	1.137	66	1.451	1.637
5-10Kb	0						3.908	4.207
10-15Kb	0						2.852	3.186
15-20Kb	0						2.708	4.189
20-25Kb	0						72	5.0.00
25-30Kb	0						838	5.000
30-35Kb	1.329	<i>ISBu</i>	RM-DPDB	31.998	32.138	141	88	1.684
		<i>ISBu</i>	RM-DPDB	32.183	32.350	168		
		<i>ISBu</i>	RM-DPDB	32.428	32.640	213		
		<i>ISBu</i>	DPDB	32.885	33.551	667		
		<i>ISBu</i>	DPDB	33.614	33.753	140		
35-40kb	518	<i>TART</i>	RM	37.149	37.288	140	2.185	3.063
		<i>TART</i>	RM	37.356	37.581	226		
		<i>TART</i>	RM	38.116	38.267	152		
40-45Kb	0						1.663	5.000
45-50Kb	100	<i>TART</i>	RM	47.911	48.010	100	3.063	3.312
50-55Kb	0						802	932

55-60Kb	0						0	0
60-65Kb	0						0	0
65-70Kb	0						0	0
70-75kb	58	<i>TART</i>	RM	70.923	70.980	58	1.470	1.535
75-80Kb	142	<i>BuT5</i>	DPDB	79.495	79.613	142	1.983	3.701
80-85Kb	438	<i>BuT5</i>	DPDB	79.644	80.023	380	1.938	2.003
		<i>BuT5</i>	DPDB	80.067	80.124	58		
85-90Kb	0						2.091	2.284
90-95Kb	249	<i>Helitron</i>	RM	91.677	91.724	48	2.252	2.931
		<i>TART</i>	RM	93.780	93.980	201		
95-100Kb	0						4.016	5.000
100-105Kb	0						4.537	5.000
105-110Kb	0						2.508	4.242
110-115Kb	0						3.615	3.841
115-120Kb	0						2.805	4.018
120-125Kb	0						0	0
125-130Kb	756	<i>ISBu</i>	RM	126.008	126.763	756	0	0
130-135Kb	0						384	3.213
135-140Kb	0						591	5.000
140-145Kb	0						92	5.000

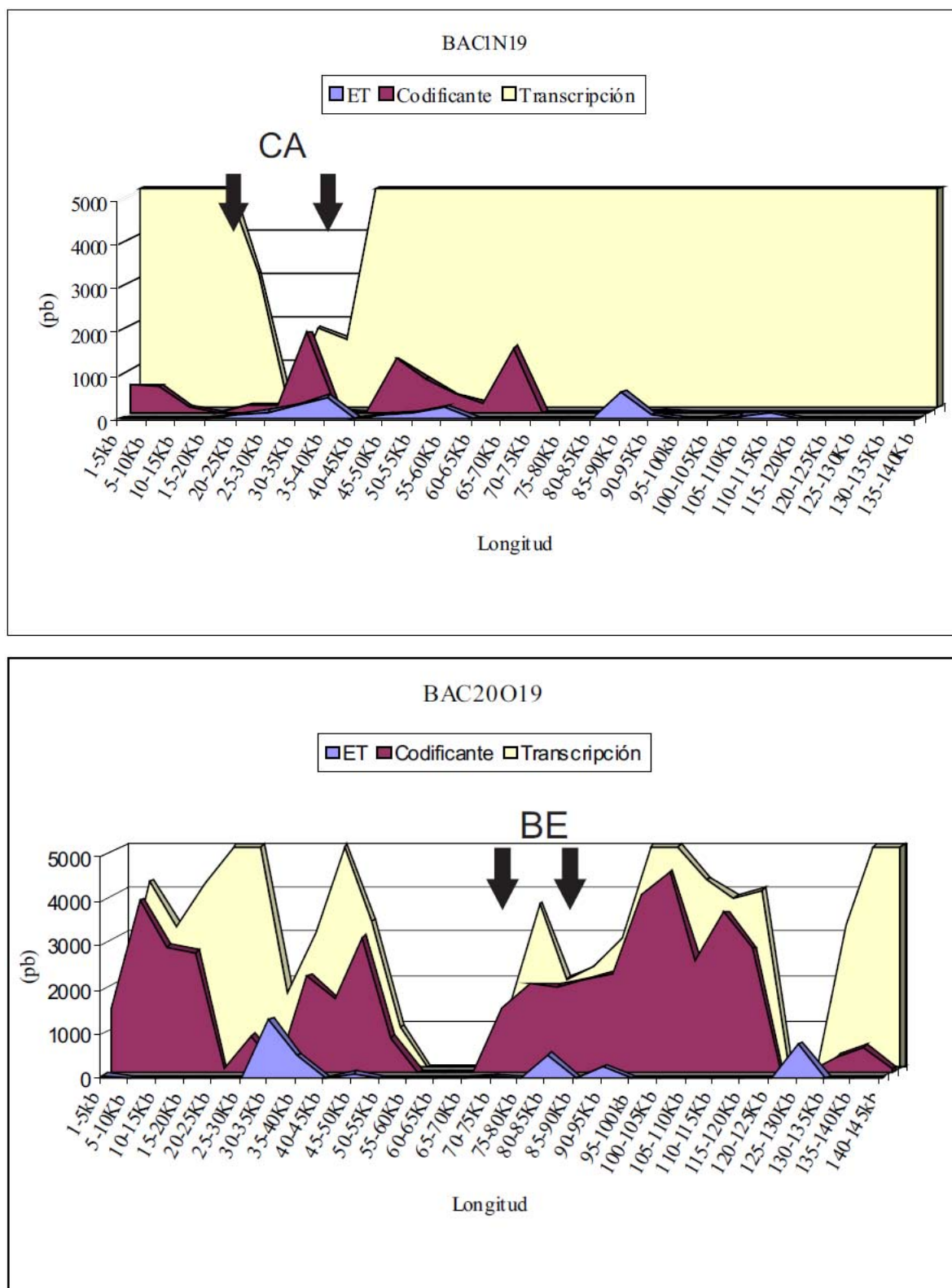


Figura 3.6. Análisis del contenido de los clones BAC 1N19 (arriba) y 20O19 (abajo). El contenido se muestra en ventanas de 5 Kb mostrando la cantidad de elementos transponibles (azul), CDS (granate) y región que se transcribe (amarillo).

Tabla 3.8: Resumen del contenido de cada BAC en elementos transponibles (TE), CDS y región transcrita (transcripción). Los TE se dividen en función de su localización (intergénica o intrónica). La información de cada BAC se subdivide en el contenido aguas abajo, aguas arriba y de la propia región del punto de rotura. En esta Tabla se muestra el número de pares de bases y la cantidad relativa en % que representan en la región. Hemos considerado como punto de rotura a toda la región entre los genes *scrib* y *CG12250* (AC) y entre *Ssadh* y *Wsck* (BE).

BAC	1N19 (138.724 pb)						20O19 (143.293 pb)					
Región	C		Punto de rotura AC		A		B		Punto de rotura BE		E	
	(1-23.072)		(23.073-38415)		(38.416-138.724)		(1-74.337)		(74.338-82.579)		(82.580-143.293)	
			(15.343)		(100.308)				(8.242)		(60.714)	
	pb	%	pb	%	pb	%	pb	%	pb	%	pb	%
Total TE	70	0,30	908	5,92	1.284	1,67	2.068	2,78	522	6,33	1.005	1,65
TE/Intergénica	0	0	908	5,92	0	0	1.453	1,95	522	6,33	804	1,32
TE/Intrónico	70	0,30	0	0	1.284	1,67	615	0,83	0	0	201	0,33
CDS	1.497	6,50	2.004	13,06	4.254	5,53	21.100	28,38	1.983	24,06	24.829	40,89
Transcripción	23.072	100,0	2.004	13,06	76.970	100,00	38.745	52,12	3.701	44,90	42.532	70,05

3.1.7. Análisis detallado de los puntos de rotura en *D. buzzatii*

Una imagen detallada de la organización de los genes de las regiones de los tres puntos de rotura tanto en *D. buzzatii* (cromosoma invertido) como en *D. mojavensis* (cromosoma no invertido) se muestra en la Figura 3.7.

Punto de rotura AC

En *D. mojavensis*, la región del punto de rotura AB comprende la región intergénica de 17.800 pb entre los genes *msi* y *Ssadh* incluyendo el gen *CG4673* y sus dos genes anidados en el intrón 6 (*CG5079* y *CG5071*) (véase Figura 3.3). A la luz de la nueva anotación del gen *CG12250* (región donde se encuentra la sonda A5), la región del punto de rotura entre *CG12250* y *Ssadh* se puede reducir a 12.650 pb.

En *D. buzzatii*, la región del punto de rotura AC comprende la región intergénica de 15.343 pb entre los genes *CG12250* y *scrib* (Figura 3.7). Si comparamos las regiones AC de *D. buzzatii* y la región AB de *D. mojavensis*, la similitud alcanza los 12.841 pb aguas arriba del inicio del gen *CG12250* (A). Cuando la misma región del punto de rotura AC de *D. buzzatii* se comparó con la región CD de *D. mojavensis*, la similitud alcanzó 1.808 pb aguas arriba del inicio del gen *scrib* (C). Estas observaciones localizan el punto de rotura de AC en un segmento de 693 pb, entre las coordenadas 113.151 y 113.843 del BAC 1N19 (orientación AC).

Sólo se detectó un ORF potencialmente funcional en esta región. Dicho ORF correspondería al gen *CG5071* situado 6.585 pb aguas arriba del codón inicial de *CG12250*. No se detectó ningún ORF putativo correspondiente al gen *CG5079*, aunque se detectó un pequeño fragmento con similitud con el gen *CG5079* de *D. mojavensis*. Por otro lado, ningún programa de predicción de genes detectó ningún ORF correspondiente al gen *CG4673* en este

punto de rotura. Sin embargo, algunos fragmentos similares al gen *CG4673* de *D. mojavensis* también pudieron ser anotados en esta región de *D. buzzatii*. Estos fragmentos aislados no constituyen una copia completa del gen y no se conserva ni la estructura ni el marco de lectura (ver más abajo).

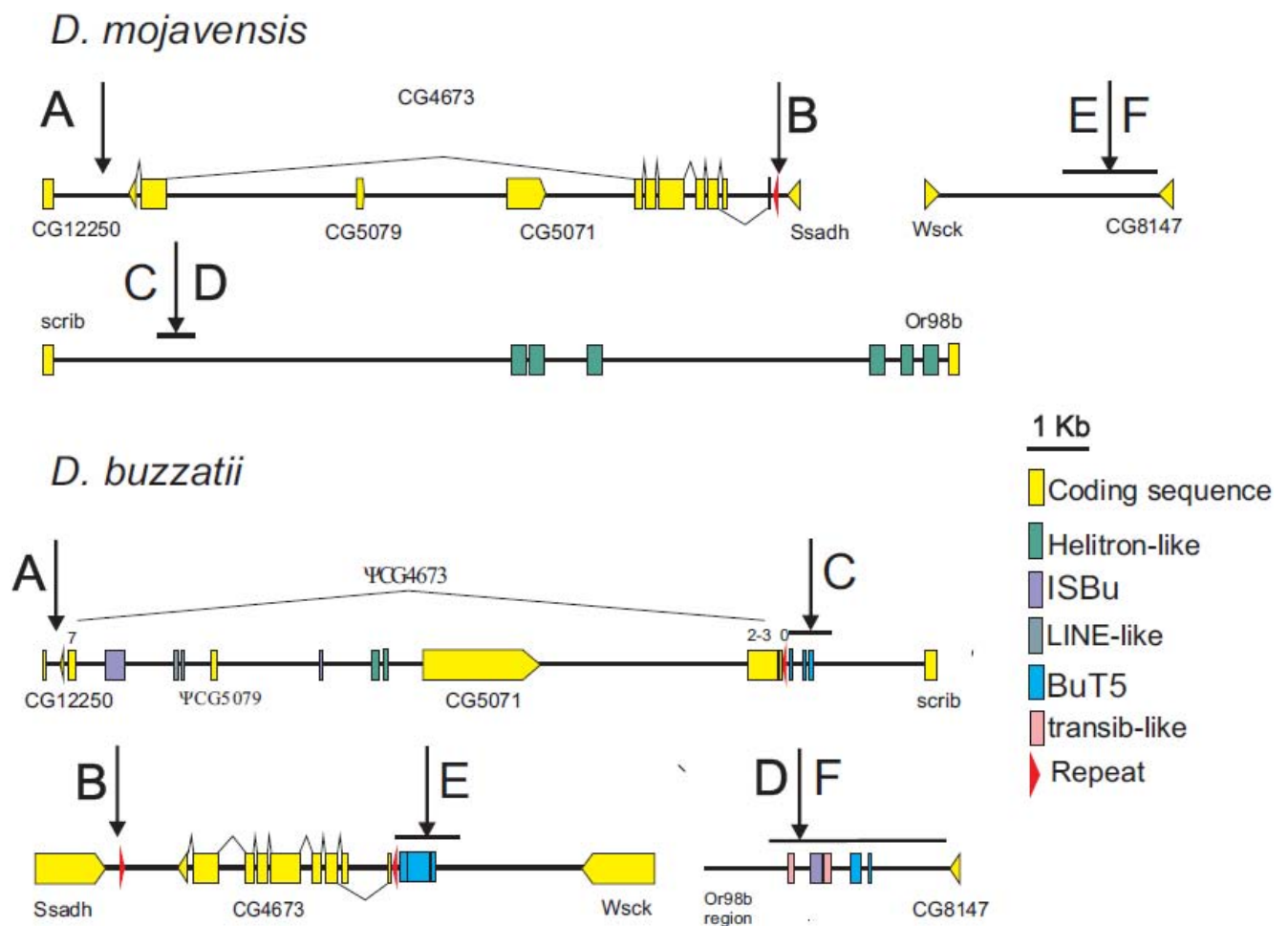


Figura 3.7.: Anotación de las regiones de los puntos de rotura de las inversiones $2m$ y $2n$ en *D. mojavensis* (cromosoma no invertido) y *D. buzzatii* (cromosoma invertido). Las líneas situadas debajo de las flechas negras muestran la junta del punto de rotura.

Punto de rotura BE

La región del punto de rotura BE se encuentra en la región intergénica de 8.242 pb entre los genes *Ssadh* (CG4685) y *Wsck* (CG31127) en el BAC 20O19 (Figura 3.7). La similitud con la región AB de *D. mojavensis* llega a 5.108 pb aguas abajo del codón stop de *Ssadh*. Por otra parte, la similitud con la región EF de *D. mojavensis* alcanza 2.019 pb aguas abajo del codón stop de *Wsck*. Así, el punto de rotura se encuentra en un segmento de 1.114 pb, entre las coordenadas 79.447 pb y 80.560 pb del BAC 20O19 (orientación BE).

Los programas de predicción de genes utilizados detectaron una óptima estructura codificante (ORF putativo) en esta región de *D. buzzatii* ortóloga al gen *CG4673* cuando se la compara con *D. mojavensis* y *D. melanogaster*. No hay rastro en esta región de *D. buzzatii* de los dos genes anidados en *CG4673* (*CG5071* y *CG5079*) en *D. mojavensis*. La comparación de la región del punto de rotura BE con la región AB de *D. mojavensis* nos permitió constatar que esta copia de *CG4673* está en una orientación opuesta a la encontrada en la región AB de *D. mojavensis*. En el genoma de *D. mojavensis*, el gen *CG4673* se encuentra aguas arriba de *CG12250* (A) y aguas abajo de *Ssadh* (B) con los tres genes en la misma cadena (la misma orientación). Sin embargo, esta copia de *CG4673* en la región BE de *D. buzzatii*, aunque aún está aguas abajo de *Ssadh*, está ahora en la cadena opuesta (orientación inversa). Los fragmentos de similitud con el gen *CG4673* en el punto de rotura AC de *D. buzzatii*, siguen aguas arriba del gen *CG12250* (A) y ambos están en la misma cadena.

Punto de rotura DF

El producto de la PCR de 4.400 pb que contiene el punto de rotura DF se comparó con la región de CD de *D. mojavensis*. Se detectaron dos bloques con similitud significativa (uno de 408 pb con el 87,7% de identidad y el E-value= 2e-147 y otro de 93 pb con el 70,9% de identidad y E-value= 9e-06). Aunque estas secuencias conservadas no corresponden a la

región codificante del gen *Or98b* o cualquier otro conocido, se encuentran aguas arriba de este en *D. mojavensis* y podrían incluir elementos potenciadores (*enhancers*) o regulador del gen. La similitud con la región de CD de *D. mojavensis* llega hasta la primeros 1.050 pb. Por otra parte, los últimos 165 pb de la secuencia del DF son homólogos al último exón y a la 3'UTR del gen *CG8147* de *D. mojavensis* (presente en la región EF). Así, el punto de rotura DF se encuentra en un segmento de 3.185 pb, entre las coordenadas 1.051 y 4.235 de la secuencia DF de *D. buzzatii* (Figura 3.7). Por lo tanto, el contenido del producto amplificado DF de *D. buzzatii* contiene únicamente región intergénica anotada en los puntos de rotura CD y EF de *D. mojavensis*. No hay ningún ORF adicional ni secuencia procedente del punto de rotura AB de *D. mojavensis*. El punto de rotura DF de *D. buzzatii* se localiza, como se esperaba, entre los genes *Or98b* y *CG8147*.

3.1.8. ORF en los puntos de rotura AC y BE de *D. buzzatii*

En *D. mojavensis* y *D. melanogaster* hay dos genes anidados anotados en la cadena complementaria del intrón 6 del gen *CG4673* (*CG5071* y *CG5079*).

El análisis de los puntos de rotura reveló identidad con el gen *CG4673* de *D. mojavensis* en ambos puntos de rotura de la inversión *2m* de *D. buzzatii* (Figura 3.7). En el punto de rotura BE en el BAC 20O19 parece existir una copia completa y potencialmente funcional del gen *CG4673* (véase sección 3.4). Sin embargo, esta copia en *D. buzzatii* carece de la práctica totalidad del intrón 6 y, por consiguiente, de los dos genes anidados (*CG5071* y *CG5079*).

Una repetición invertida de 121 pb (72,7% de identidad y E-value= 2e-16) se encuentra flanqueando al gen *CG4673* en el punto de rotura BE en *D. buzzatii*. Una repetición se encuentra 117 pb aguas abajo del codón stop de *Ssadh* y la otra se encuentra 3.137 pb aguas

abajo del codón stop de *Wsck*. Esta secuencia repetida se encuentra sólo una vez en el genoma de *D. mojavensis* y en la región del punto de rotura AC de *D. buzzatii* y está anotada como secuencia intergénica en 5' del gen *CG4673*.

Los programas de predicción de genes señalaron un ORF relacionado con el gen *CG5071* en el punto de rotura AC. En *D. mojavensis*, el gen *CG5071* está anotado como un gen de exón único que codifica una proteína de 225 aa. En *D. melanogaster*, el gen *CG5071* produce dos transcritos diferentes. El transcrito RA de *CG5071* contiene dos exones y codifica una proteína de 509 aa. El transcrito RB de *CG5071* no tiene señal de procesamiento y contiene un único exón que codifica una proteína de 680 aa. Ambos transcritos de *D. melanogaster* codifican proteínas de mayor tamaño que en *D. mojavensis*, la cual parece tener un codón prematuro de stop en la anotación de su transcrito. En *D. buzzatii*, el gen *CG5071* consta de un único exón de 2.004 pb de longitud que codifica una proteína de 668 aa con muy baja similitud con el gen de *D. mojavensis* (72% de identidad de los 225 aa alineados). De este modo, en *D. buzzatii* este gen es considerablemente mayor que en *D. mojavensis* pero similar al de *D. melanogaster* (transcrito RB). Por el contrario, ninguno de los programas utilizados detectaron ningún ORF relacionado con los genes *CG4673* y *CG5079* en este punto de rotura en el BAC 1N19. Sólo algunos fragmentos de similitud con los genes *CG4673* y *CG5079* pudieron ser anotados en este punto de rotura cuando se comparó esta región con el genoma de *D. mojavensis*. Se detectaron 16 pequeños bloques (de 45bp a 393 pb) con 65,2% al 88,9% de identidad, incluyendo fragmentos exónicos, del intrón 6 y de las UTR (Tabla 3.9). El orden de las coordenadas del *scaffold* 6540 de *D. mojavensis* es ascendente y la anotación del gen *CG4673* se dispone desde la 5'UTR hasta la 3'UTR. El fragmento del clon BAC 1N19 usado para la comparación, sin embargo, está en orientación opuesta. Dos fragmentos no siguen este patrón. Por un lado, cerca de los exones 2 y 3, hay dos fragmentos anotados como intrón 6 en *D. mojavensis* (25.759.443-25.759.569 pb y 25.759.465-25.759.577 pb) con identidad con dos

Tabla 3.9: Bloques de similitud entre el gen *CG4673* de *D. mojavensis* y la región del punto de rotura AC (BAC 1N19) de *D. buzzatii*. La búsqueda se realizó mediante BLAST (bl2seq) entre *D. buzzatii* (BAC 1N19:100.310..115.652) como secuencia problema y *D. mojavensis* (*scaffold_6540*:25.756.516..25.769.308) como secuencia alineada. Sólo las identidades con un tamaño mínimo de 40 pb y E-value $\leq 1e-03$ fueron incluidas en la lista.

Anotación de la región	Coordenadas en el scaffold 6540 de <i>D. mojavensis</i> (pb)	Coordenadas en <i>D. buzzatii</i> (pb)	Identidad (%)	E-value
5'UTR	25.756.516-25.756.613	12.802-12.704	72/99 (72,7)	2e-07
Exón 2-3	25.757.533-25.757.920	12.632-12.284	273/393 (69,5)	2e-37
Intrón 6	25.759.443-25.759.569	11.708-11.583	99/128 (77,3)	5e-21
Intrón 6	25.759.465-25.759.577	10.089-9.966	91/124 (73,4)	7e-13
Intrón 6	25.760.725-25.760.896	10.246-10.055	132/196 (67,3)	3e-10
Intrón 6	25.761.139-25.761.215	9.726-9.652	59/77 (76,6)	4e-09
Intrón 6	25.761.270-25.761.379	9.643-9.549	79/110 (71,8)	3e-10
CG5071	25.761.635-25.762.277	8.571-7.930	520/643 (80,9)	1e-173
CG5071	25.762.432-25.762.593	6.719-6.552	122/168 (72,6)	2e-20
Intrón 6	25.763.858-25.763.918	4.202-4.148	47/61 (77,0)	6e-07
CG5079	25.764.640-25.764.730	2.991-2.895	71/97 (73,2)	3e-10
Intrón 6	25.765.297-25.765.341	2.490-2.446	39/45 (86,7)	4e-09
Intrón 6	25.765.806-25.765.925	2.129-2.007	90/127 (70,9)	8e-12
Intrón 6	25.766.399-25.766.446	3.679-3.726	39/48 (81,2)	2e-06
Intrón 6	25.767.941-25.768.091	992-806	122/187 (65,2)	3e-10
Intrón 6	25.768.145-25.768.198	801-748	48/54 (88,9)	5e-14
Intrón 6	25.768.279-25.768.338	627-568	45/60 (75,0)	9e-05
Exón 7	25.768.450-25.768.573	518-396	98/125 (78,4)	4e-22
3' UTR	25.769.034-25.769.094	278-218	53/61 (86,9)	4e-15

fragmentos distintos en *D. buzzatii* (11.708-11.583 pb y 10.089-9.966 pb, respectivamente).

Los fragmentos de *D. mojavensis* son solapantes desde la posición 25.759.465 pb a la

25.759.569 pb, lo que podría indicar una pequeña duplicación en la región del intrón 6 de *D. buzzatii*. Por otro lado, un fragmento del clon BAC (de 3.679 a 3.726 pb) se localiza en orientación opuesta al patrón. Este fragmento corresponde a una región del intrón 6 aguas arriba del gen *CG5079* en *D. mojavensis*, pero en *D. buzzatii* se encuentra aguas abajo del gen. El algoritmo de búsqueda de promotores putativos en el genoma de *D. melanogaster* compilado en Flybase (Ohler *et al.* 2002) no detecta ningún posible promotor en este fragmento reorganizado. Ambas reorganizaciones encontradas en esta región del intrón 6 parecen ser eventos propios de la degradación que sufre la secuencia de los genes *CG5071* y *CG4673* en el punto de rotura AC. Ambos genes parecen pseudogeneizados, por lo que nos referiremos a partir de ahora como $\Psi CG4673$ y $\Psi CG5079$ para referirnos a las copias de este punto de rotura. En el punto de rotura AC, el gen *CG5071* y $\Psi CG5079$ han sido anotados entre los fragmentos más remotos de $\Psi CG4673$ en el punto de rotura AC, por lo que se mantienen anidados dentro de la copia degenerada.

3.1.9. Duplicación asociada a la inversión 2m

En el punto de rotura BE de *D. buzzatii* se ha anotado el gen *CG4673* mientras en AC se ha anotado el gen *CG5071* y fragmentos parciales de $\Psi CG5079$ y $\Psi CG4673$. Parece que existe un fragmento duplicado asociado a las inversiones *2m* y *2n* en *D. buzzatii*. Sin embargo, no parece existir ninguna duplicación funcional de los genes implicados.

Las anotaciones en ambos puntos de rotura, indican que el fragmento completo (incluyendo el gen *CG4673* y sus dos genes anidados) fue inicialmente duplicado. La pseudogeneización y la degradación se producirían más tarde. La duplicación no ha sido acompañada de un aumento en el número de copias funcionales de los genes en *D. buzzatii*. De hecho, el gen *CG5079* habría perdido completamente su funcionalidad. Sin embargo, dicha

duplicación ha sufrido dos reorganizaciones. Por un lado, el gen *CG4673* en *D. buzzatii* está ahora compactado debido a la pérdida de ~9 Kb de secuencia de intrónica. Por otra parte, el gen *CG5071*, anidado en *CG4673* en los genomas no invertidos, está ahora fuera de este en *D. buzzatii* (proceso que podríamos denominar desanidación).

La inesperada duplicación y su degradación parcial y distinta en cada punto de rotura, parece ser el motivo por el que las PCR diseñadas en estas regiones y utilizadas como sondas en la hibridación *in situ* no fueron concluyentes. También las PCR diseñadas para amplificar los puntos de rotura AC y BD en *D. buzzatii* fracasaron debido a las inesperadas longitud y orientación de los fragmentos.

3.1.10. Microinversión

En el genoma de *D. mojavensis*, el gen *CG4673* se encuentra aguas arriba de *CG12250* región (A) y aguas abajo de *Ssadh* (región B) con los tres genes en la misma cadena (la misma orientación). En *D. buzzatii*, Ψ *CG4673* en AC está en el mismo lugar que *CG4673* en el genoma no invertido de *D. mojavensis* (AB). Ambos están en la región A, aguas arriba y en la misma orientación de *CG12250*. Sin embargo, el gen *CG4673* de *D. buzzatii* está invertido respecto a la región AB punto de rotura de *D. mojavensis*. El gen *CG4673* en el punto de rotura BE, sigue aguas abajo de *Ssadh* (B) pero está ahora en la cadena opuesta (orientación inversa). Esta anotación parece indicar que el fragmento duplicado en el punto de rotura BE podría haber sufrido una microinversión.

3.1.11. Análisis detallado de los elementos transponibles en los puntos de rotura

Utilizando varias herramientas bioinformáticas (véase Materiales y Métodos) buscamos fragmentos de elementos transponibles (TE) y secuencias derivadas de su inserción en los puntos de rotura de las inversiones $2m$ y $2n$. Algunos TE han podido ser anotados en las regiones de los puntos de rotura, que incluyen toda la región intergénica entre los genes anotados aguas arriba y aguas abajo de éstos (Tabla 3.10 y Figura 3.7). Algunos de estos TE han podido ser anotados en las juntas de los puntos de rotura, es decir, que han sido anotados en la porción entre los bloques conservados. Si un TE ha participado de algún modo en el origen de la inversión, debe estar anotado en estas juntas.

En la región AC de *D. buzzatii* tres pequeños bloques con similitud con *BuT5* se anotaron en la junta del punto de rotura, aguas arriba de $\Psi CG4673$. Además otros TE se encontraron en la región de este punto de rotura. Aguas arriba del gen *CG5071*, dentro de la copia de $\Psi CG4673$, se anotaron dos fragmentos con similitud con *ISBu2*, otros dos similares a un elemento *LINE* (*BS2*) y, finalmente dos más similares a *Helitron1*. En la región BE, otros tres fragmentos de *BuT5* fueron anotados en la junta de este punto de rotura. Por último, dos fragmentos de *BuT5*, uno similar a *ISBu3* y dos fragmentos similares a *transib3* de *D. melanogaster* se encontraron en el punto de rotura DF. Todos ellos se han encontrado en la porción entre los bloques conservados D y F. La copia del elemento *ISBu* parece estar anidada dentro del fragmento anotado como *transib*.

Tabla 3.10: TE anotados en las regiones de los puntos de rotura AC, BE y DF de las inversiones $2m$ y $2n$. La búsqueda se realizó mediante BLAST (bl2seq), usando la secuencia específica del punto de rotura como secuencia problema y alineándola con la secuencia del TE de referencia. El elemento con mayor identidad fue el usado en la comparación (entre paréntesis en la Tabla). Sólo las identidades con un tamaño mínimo de 40 pb y E-value $\leq 1e-03$ fueron incluidas en la lista. Véase la Figura 3.7 para la localización de las copias en las regiones de los puntos de rotura.*: Estos E-value fueron obtenidos usando únicamente los fragmentos de los puntos de rotura como secuencia problema en vez de usar toda la región.

Región	Coordenadas en <i>D. buzzatii</i>	TE de referencia.	Identidad (%)	E-value
AC	1.024-1.371	<i>Dbuz</i> \ISBu2 (AF368867)	306/425 (72,0)	6e-74
	2.169-2.238	<i>Dmel</i> \BS2	55/79 (69,6)	1e-04
	2.295-2.349	<i>Dmel</i> \BS2	44/55 (80,0)	2e-07
	4.766-4.819	<i>Dbuz</i> \ISBu2(AF368867)	45/56 (80,3)	3e-08
	5.626-5.733	<i>Dmoj</i> \Helitron1	78/109 (71,5)	4e-07
	5.782-5.852	<i>Dmoj</i> \Helitron1	54/71 (76,0)	8e-09
	12.877-12.930	<i>Dbuz</i> \BuT5 (AF368894)	41/54 (75,9)	1e-06*
	13.127-13.166	<i>Dbuz</i> \BuT5 (AF368894)	34/41 (82,9)	1e-06*
	13.218-13.278	<i>Dbuz</i> \BuT5 (AF368894)	43/61 (70,5)	6e-04*
BE	5.158-5.276	<i>Dbuz</i> \BuT5 (AF368894)	88/123 (71,5)	7e-12*
	5.307-5.686	<i>Dbuz</i> \BuT5 (AF368894)	266/404 (65,8)	8e-24*
	5.730-5.787	<i>Dbuz</i> \BuT5 (AF368894)	49/65 (75,4)	2e-06*
DF	1.419-1.518	<i>Dmel</i> \Transib3	73/101 (72,3)	9e-10
	1.824-2.027	<i>Dbuz</i> \ISBu2 (AY900631)	182/204 (89,2)	5e-68
	2.042-2.172	<i>Dmel</i> \Transib3	94/133 (70,7)	2e-11
	2.522-2.700	<i>Dbuz</i> \BuT5 (AF368894)	122/179 (68,2)	9e-11
	2.787-2.836	<i>Dbuz</i> \BuT5 (AF368894)	40/55 (72,7)	0,0009

3.2. Inversión $2z^7$

3.2.1. Notación de los puntos de rotura

Con el fin de evitar la confusión con las regiones punto de rotura de las otras dos inversiones, las regiones de los puntos de rotura distal y proximal de la inversión $2z^7$ fueron nombradas PQ y RS en el cromosoma no invertido y PR y QS en el cromosoma invertido (véase Figura 3.1). En esta inversión, P corresponde a la región más distal mientras que S es la región más proximal. La inversión $2z^7$ solapa con $2m$ y $2n$ e invierte el punto de rotura BE compartido entre ellas.

3.2.2. Cartografía de los puntos de rotura en *D. mojavensis*

La región del punto de rotura proximal de la inversión $2z^7$ en *D. buzzatii* (QS) está contenida dentro del clon BAC 22N23 (González *et al.* 2007). Este clon es solapante con el clon BAC 40C11 que incluye el gen *proboscipedia* (*pb*) y ha sido totalmente secuenciado y anotado (Negre *et al.* 2005, Delprat *et al.* 2009). La comparación entre la organización de los genes anotados en el BAC 40C11 y los de la región ortóloga en el genoma de *D. mojavensis* (especie no invertida) nos permitió caracterizar dicho punto de rotura proximal. La primera mitad del BAC 40C11 corresponde a la región de la coordenada 16,6 Mb del *scaffold* 6540 de *D. mojavensis*, mientras que la otra mitad fue localizada cerca de la región en 6,1 Mb en el mismo *scaffold*. El punto de rotura QS en *D. buzzatii* fue localizado en un fragmento de 2.958 pb de tamaño (desde la posición 55.122 pb a 58.080 pb del BAC 40C11), correspondiente a la región intergénica aguas arriba del gen *CG14290* (Q) y aguas abajo del gen *Dlh* (S) (Figura 3.8). De este fragmento, 306 pb de la región intergénica aguas arriba del *CG14290* son

similares a *D. mojavensis*, dejando el punto de rotura QS en los restantes 2.652 pb. En *D. mojavensis*, adyacente y aguas arriba de *CG14290* (Q) está el 5' del gen *CG14645* (que correspondería a P). Adyacente y aguas abajo de *Dlh* (S) está el 5' de *CSN5* (que correspondería a R). Por tanto, en *D. mojavensis*, la región del punto de rotura PQ se encuentra en los 1.114 pb de secuencia intergénica entre *CG14645* y *CG14290* mientras que la región del punto de rotura RS se localiza en el segmento de secuencia intergénica de 1.738 pb entre *CSN5* y *Dlh* (Figura 3.8).

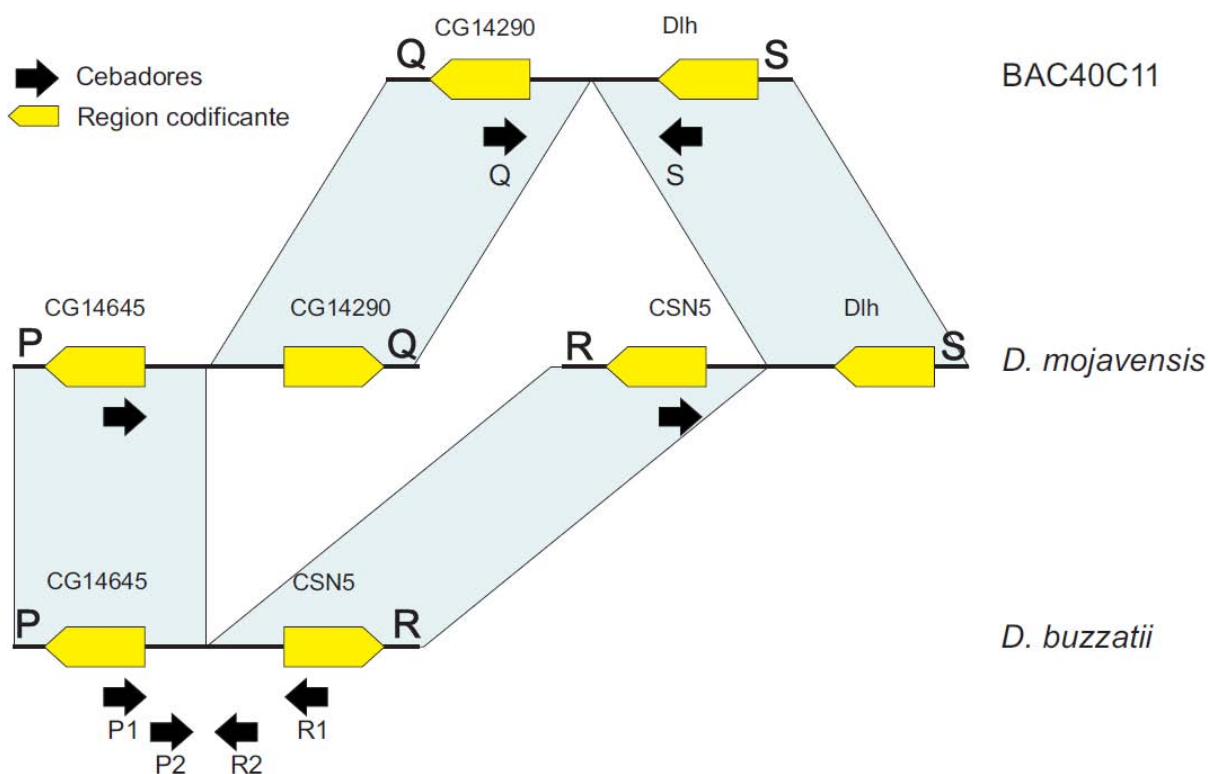


Figura 3.8: Estrategia utilizada para amplificar el punto de rotura distal PR de la inversión $2z^7$ en *D. buzzatii*. El punto de rotura proximal QS se localizó en el clon BAC 40C11 de *D. buzzatii* entre los genes *CG14290* (Q) y *Dlh* (S). Estos dos genes se localizaron en *D. mojavensis*: aguas arriba del gen *CG14290* está el gen *CG14645* (P) y aguas abajo del gen *Dlh* está el gen *CSN5* (R). Los cebadores (flechas negras) se diseñaron en los genes *CG14645* (P) y *CSN5* (R) para amplificar la región PR en *D. buzzatii*.

3.2.3. Aislamiento del punto de rotura distal en *D. buzzatii*

El punto de rotura proximal (QS) se encuentra en 2.652 pb entre los genes *CG14290* (Q) y *Dlh* anotados en el BAC 40C11. El punto de rotura distal (PR) en *D. buzzatii* esperamos que se encuentre en la región intergénica aguas arriba de *CG14645* (P) y la región aguas arriba de *CSN5* (R) (Figura 3.8). Se diseñó una PCR para corroborarlo. Los cebadores P y R (Tabla 2.2) fueron diseñados en estos genes y orientados como se espera según *D. mojavensis*. La secuencia resultante (1.619 pb) se comparó con *D. mojavensis*. Los primeros 246 pb fueron anotados como *CG14646* (P), mientras que los últimos 180 pb del fragmento fueron anotados como *CSN5* (R). Por tanto, el punto de rotura PR fue localizado en un fragmento 1.193 pb de región intergénica entre ambos genes. Sin embargo, la similitud de la región del punto de rotura PR de *D. buzzatii* y la región PQ de *D. mojavensis* llega hasta 484 pb aguas arriba del codón de inicio del gen *CG14645*. Cuando la misma región se compara con la región del punto de rotura RS de *D. mojavensis*, la similitud llega hasta 104 pb aguas arriba del gen de *CSN5*. Por tanto, el punto de rotura PR queda localizado en un fragmento de 605 pb en la región intergénica entre los genes *CG14645* (P) y *CSN5* (R) (Figura 3.8).

3.2.4. Análisis de los puntos de rotura

Se buscaron en los puntos de rotura de la inversión $2z^7$ en *D. buzzatii* utilizando varias herramientas bioinformáticas inserciones de elementos transponibles (TE), así como rastros derivados de estos (Tabla 3.11). Sólo dos pequeños fragmentos de diferentes TE (*Galileo* y *ISBu1*) habían sido previamente anotados en la región del punto de rotura proximal (BAC 40C11, Negre *et al.* 2005). Un análisis más detallado de esta región nos ha permitido ampliar y mejorar dichas anotaciones. Por otra parte, otros fragmentos no anotados anteriormente,

podieron ser detectados entre los genes *Dlh* y *CG14290* en *D. buzzatii*. Estos nuevos TE son similares a *BuT5*, *Dmoj_38* (tipo *CR1* de *D. melanogaster*) y *Dmoj_238* (tipo *Helena* de *D. melanogaster*). Hay cuatro bloques de similitud con *CR1* y los fragmentos de *ISBu1* y tipo *Helena* fueron anotados intercalados entre ellos (Figura 3.9). Aparentemente, estas dos inserciones son colonizadores secundarios anidados en la copia de *CR1*. Para la copia del elemento tipo *Helena* insertada, pudieron ser anotadas las duplicaciones del sitio diana de 9 pb (TSD) que produce su inserción. Sólo un fragmento de 329 pb de largo del extremo 3' de este elemento pudo ser anotado, aunque no se trata de una copia degenerada. La síntesis de este tipo de elementos consiste en una retrotranscripción desde el extremo 3'OH del mismo sitio de inserción desde donde se cebará la transcripción reversa (Ostertag y Kazazian 2001). El mRNA del elemento sirve de molde, retrotranscribiendo desde la poli-A (que queda insertada). Las inserciones en nuevas ubicaciones pueden ser completas o truncadas con el extremo 5' (*Death-On-Arrival*) si no se finaliza la retrotranscripción (Eickbush y Jamburuthugoda 2008). Por otro lado, la copia de *ISBu1* parece ser incompleta debido a una delección parcial.

Tabla 3.11: TE anotados en los puntos de rotura PR y QR de la inversión $2z^7$. La búsqueda se realizó mediante BLAST (bl2seq), usando la secuencia específica del punto de rotura como secuencia problema y alineándola con la secuencia del TE de referencia. El elemento con mayor identidad fue el usado en la comparación (entre paréntesis). Sólo las identidades con un tamaño mínimo de 40 pb y E-value $\leq 1e-03$ fueron incluidas en la lista. Véase Figura 3.9 para la localización de las copias en las regiones de los puntos de rotura.

Región	Especies	Coordenadas	TE	Identidad (%)	E-value
PR	<i>D. buzzatii</i>	482-616	<i>Dbuz\Newton</i> (AF368869)	100/146 (68,5)	7e-07
	<i>D. martensis</i>	480-541	<i>Dbuz\Newton</i> (AF368869)	50/63 (79,4)	5e-10
		686-766	<i>Dbuz\Newton</i> (AF368869)	64/82 (78,0)	9e-13

QS	<i>D. buzzatii</i>	394-480	<i>Dbuz\Newton</i> (AY756163)	66/88 (75,0)	2e-08
		502-704	<i>Dbuz\Galileo</i> (AY187769)	149/206 (72,3)	9e-26
		804-860	<i>Dbuz\BuT5</i> (AY187769)	49/57 (86,0)	2e-13
		893-934	<i>Dbuz\BuT5</i> (AY187769)	36/43 (83,7)	6e-07
		957-1.258	<i>Dmoj_38</i> (<i>Dme\CR1-like</i>)	261/308 (84,7)	2e-96
		1.258-1.584	<i>Dmoj_238</i> (<i>Dme\Helena-like</i>)	246/329 (74,8)	6e-63
		1.587-1.941	<i>Dmoj_38</i> (<i>Dme\CR1-like</i>)	307/397 (77,3)	7e-95
		1.944-2.138	<i>Dmoj_38</i> (<i>Dme\CR1-like</i>)	160/196 (81,6)	8e-50
		2.190-2.710	<i>Dbuz\ISBu1</i> (AF368869)	495/547 (90,5)	0
		2.711-2.899	<i>Dmoj_38</i> (<i>Dme\CR1-like</i>)	166/207 (80,2)	5e-51
	<i>D. martensis</i>	398-457	<i>Dbuz\Newton</i> (AY756163)	45/60 (75,0)	3e-07
		501-673	<i>Dbuz\Newton</i> (AY756163)	143/193 (74,1)	9e-33
		540-622	<i>Dbuz\BuT5</i> (AY187769)	64/84 (76,2)	5e-12
924-985		<i>Dbuz\Galileo</i> (AY756163)	54/62 (87,1)	5e-18	
<i>D. koepferae</i>	417-465	<i>Dbuz\Newton</i> (AY756163)	39/50 (78,0)	5e-04	
	500-671	<i>Dbuz\Newton</i> (AY756163)	143/184 (77,7)	3e-38	
	833-933	<i>Dbuz\BuT5</i> (AY187769)	86/116 (74,1)	6e-17	
RS	<i>D.mojavensis</i>	313-1.579	<i>Dmoj\Galileo</i> (BK006357)	1.215/1.267 (95,9)	0

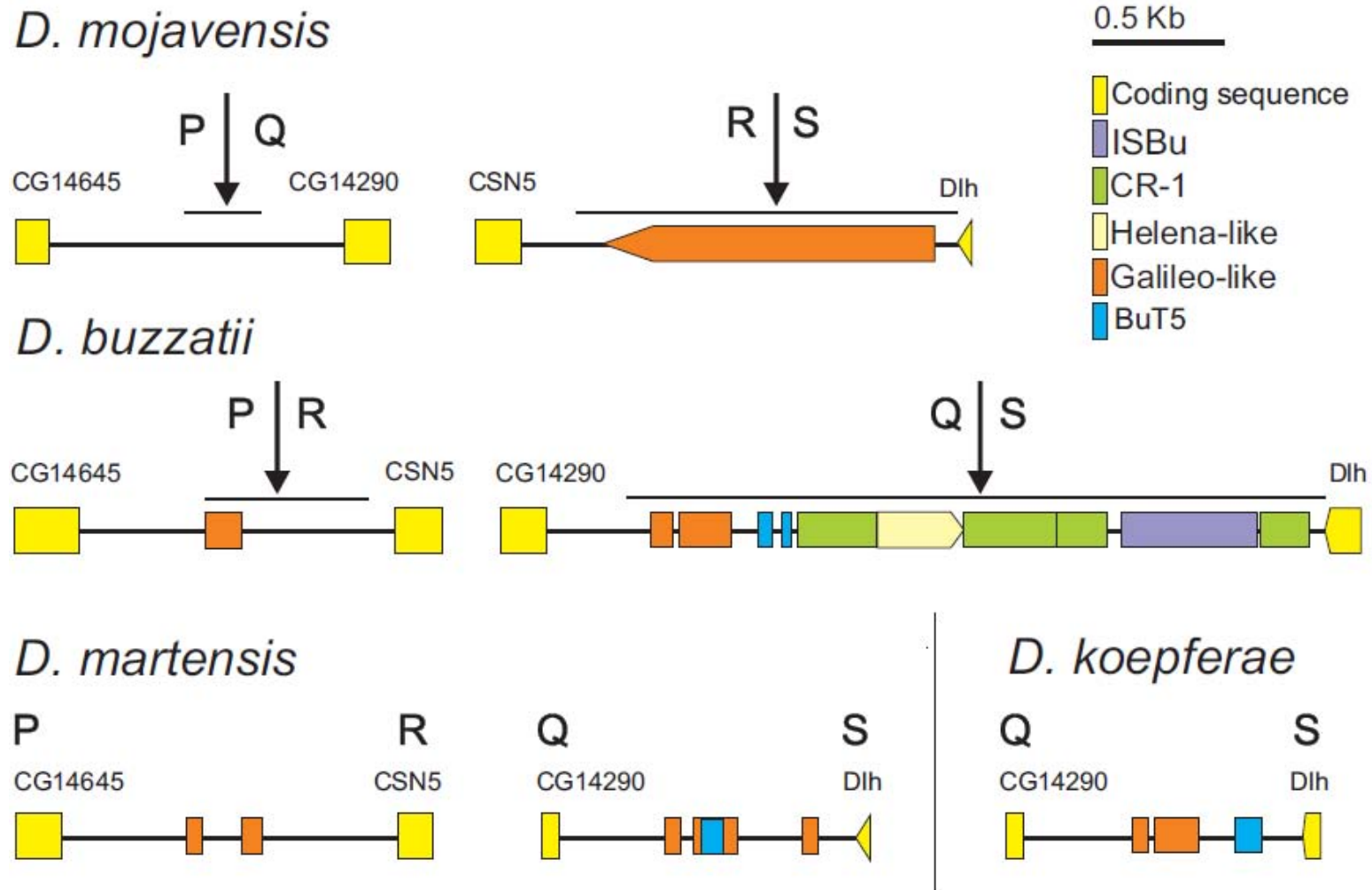


Figura 3.9: Anotación de las regiones de los puntos de rotura de la inversión $2z^7$. PQ y RS representan los puntos de rotura distal y proximal en *D. mojavensis* (cromosoma no invertido) mientras PR y QS son los puntos de rotura distal y proximal en *D. buzzatii*, *D. martensis* y *D. koepferae* (cromosoma invertido). Las líneas que se encuentran por debajo de las flechas negras muestran la junta del punto de rotura utilizando sólo la comparación entre los genomas de *D. mojavensis* y *D. buzzatii*.

El tiempo de inserción de los retrotransposones con LTR (*Long Terminal Repeats*) se puede estimar mediante el cálculo de la divergencia entre sus LTR, que son idénticas en el momento de la inserción (Brookfield 1986; Kapitonov y Jurka 1996; Costas y Naveira 2000; Bowen y McDonald 2001) suponiendo una tasa de sustitución sinónima (neutra). Pero en este caso, las copias encontradas en estos puntos de rotura son demasiado viejas y, por lo tanto, degeneradas. No se han podido identificar ni diferenciar claramente las LTR. Un método alternativo y aproximativo para estimar el tiempo de inserción de los elementos se basa en su presencia en otras especies filogenéticamente relacionadas. Si la inserción es reciente, pocas especies cercanas tendrán el elemento insertado. Por el contrario, cuanto más ancestral sean las inserciones, más especies tendrán rastro de estos. La edad aproximada de la inserción, la determinará el tiempo de divergencia con la especie más cercana sin rastro del elemento. Con el fin de comprobar la presencia de estos TE y su posible implicación en la inversión, también se aisló por PCR el punto de rotura QS en *D. martensis* y *D. koepferae*, dos especies del complejo *buzzatii* y también con la inversión $2z^7$ fijada (véase Figura 1.3 en Introducción). Se diseñaron nuevos cebadores en los genes *CG14290* y *Dlh* de *D. buzzatii* (véase Tabla 2.3 en Materiales y Métodos). Se amplificaron de 1.123 pb de *D. martensis* y 1.055 pb en *D. koepferae* correspondientes a dicho punto de rotura (Figura 3.9). La región del punto de rotura QS en estas especies es más pequeño que el fragmento equivalente en *D. buzzatii* al carecer de la mayoría de los fragmentos de elementos anotados de esta especie. No hay rastros de los fragmentos de los elementos *ISBu1*, tipo *CR1* ni tipo *Helena*. Sólo fueron detectados (en ambas especies) un fragmento insertado similar a *Galileo* y otro similar a *But5* (Figura 3.9).

Por otro lado, el punto de rotura PR también fue aislado en *D. martensis* (1.271 pb) usando para ello los mismos cebadores utilizados anteriormente para amplificar la misma región en *D. buzzatii*. También se anotó un pequeño fragmento similar a *Galileo* fue anotado en esta región.

3.3. Comparación de las regiones de los puntos de rotura en el género *Drosophila*

Las regiones de los puntos de rotura de las tres inversiones fueron comparadas en los genomas de las 12 especies secuenciadas de *Drosophila*. La ordenación de *D. mojavensis* (genoma sin las inversiones $2m$, $2n$ y $2z^7$) se tomó como referencia para determinar la orientación de los genes aguas arriba y aguas abajo de los puntos de rotura AB, CD, EF, PQ y RS en las especies secuenciadas del género (véase Figura 3.7). La conservación de los puntos de rotura y los linajes dónde se producen interrupciones se muestran en la Figura 3.10. La ordenación de los genes implicados en los puntos de rotura se muestra en la Tabla 3.12.

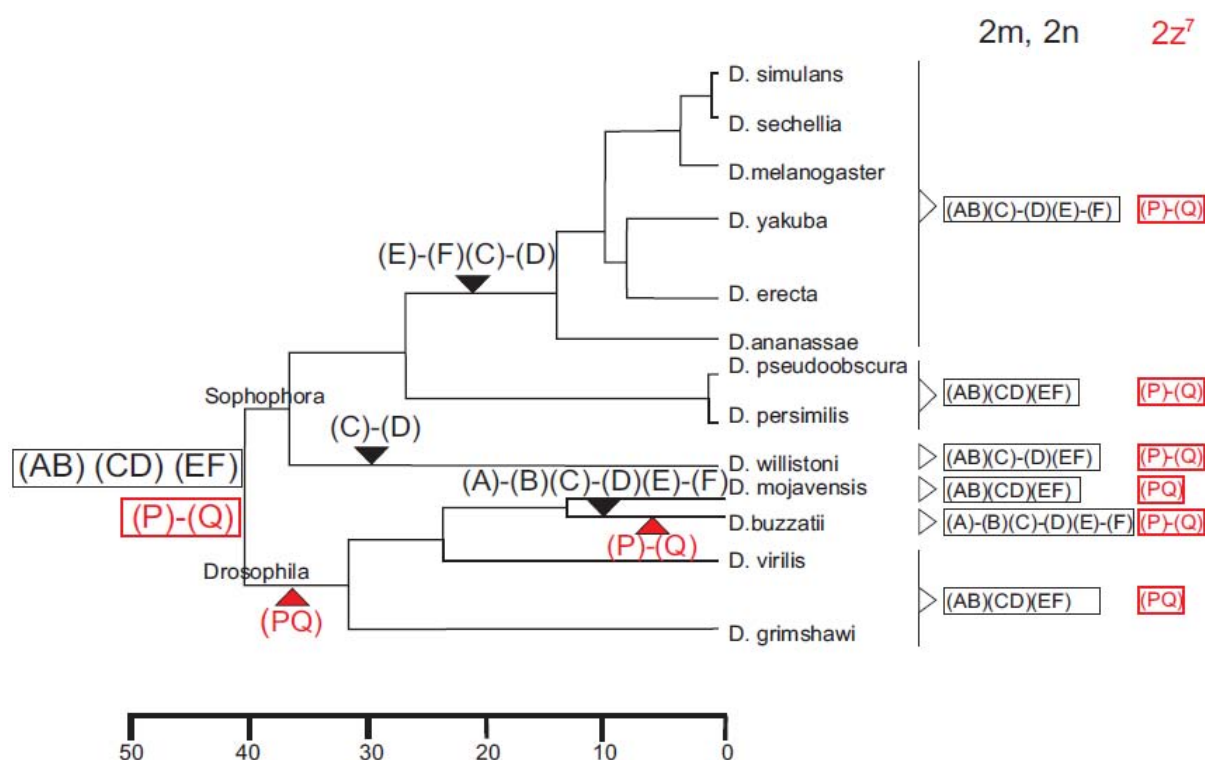


Figura 3.10: Conservación de las regiones de los puntos de rotura de las inversiones $2m$ y $2n$ (negro) y $2z^7$ (rojo) en el genoma de las 12 especies de *Drosophila* secuenciadas más *D. buzzatii*. Las cajas representan la organización en los genomas. También se muestra la hipotética ordenación ancestral en el género. Si las dos regiones están juntas, por ejemplo (AB), la región del punto de rotura se conserva con *D. mojavensis*. Si están separadas, por ejemplo (A)-(B), ambas regiones no son continuas. Los triángulos muestran en qué linaje se produjo la rotura que reorganiza estos puntos de rotura.

Tabla 3.12: Ordenación de los genes en las regiones de los puntos de rotura de las inversiones $2m$, $2n$ y $2z^7$ en las 12 especies secuenciadas de *Drosophila* más *D. buzzatii* y *Anopheles gambiae*. Se han tomado como referencia los genes de la ordenación del genoma de *D. mojavensis*. Para cada gen más cercano a los puntos de rotura se indica el gen colindante. *Para este análisis se utiliza el gen *CG14899* en vez del gen *Dlh* (véase texto para más detalles). (?) indica que no se ha podido localizar el gen colindante.

	A(CG12250)	B(Ssadh)	C(scrib)	D(Or98b)	E(Wsck)	F(CG8147)	P(CG14645)	Q(CG14290)	R(CSN5)	S(CG14899)*
<i>Dsim</i>	B(<i>Ssadh</i>)	A(CG12250)	CG6490	CG1866	<i>Syx18</i>	<i>RhoL</i>	CG31222	CG9776	CG14897	CG14898
<i>Dsec</i>	B(<i>Ssadh</i>)	A(CG12250)	CG6490	CG1866	<i>Syx18</i>	<i>RhoL</i>	CG31222	CG9776	CG14897	CG14898
<i>Dmel</i>	B(<i>Ssadh</i>)	A(CG12250)	CG6490	CG1866	<i>Syx18</i>	<i>RhoL</i>	CG31222	CG9776	CG14897	CG14898
<i>Dyak</i>	B(<i>Ssadh</i>)	A(CG12250)	CG6490	CG1866	<i>Syx18</i>	<i>RhoL</i>	CG31222	CG9776	CG14897	CG14898
<i>Dere</i>	B(<i>Ssadh</i>)	A(CG12250)	CG6490	CG1866	<i>Syx18</i>	<i>RhoL</i>	CG31222	CG9776	CG14897	CG14898
<i>Dana</i>	B(<i>Ssadh</i>)	A(CG12250)	CG6490	CG1866	<i>Syx18</i>	<i>RhoL</i>	CG31222	CG9776	CG14897	CG14898
<i>Dpse</i>	B(<i>Ssadh</i>)	A(CG12250)	D(<i>Or98b</i>)	C(<i>scrib</i>)	F(CG8147)	E(<i>Wsck</i>)	CG31222	CG9776	CG14897	CG14898
<i>Dper</i>	B(<i>Ssadh</i>)	A(CG12250)	D(<i>Or98b</i>)	C(<i>scrib</i>)	F(CG8147)	E(<i>Wsck</i>)	CG31222	CG9776	CG14897	CG14898
<i>Dwil</i>	B(<i>Ssadh</i>)	A(CG12250)	CG6089	CG6966	F(CG8147)	E(<i>Wsck</i>)	CG31222	CG9776	CG14897	CG14898
<i>Dmoj</i>	B(<i>Ssadh</i>)	A(CG12250)	D(<i>Or98b</i>)	C(<i>scrib</i>)	F(CG8147)	E(<i>Wsck</i>)	Q(CG14290)	P(CG14645)	S(CG14899)	R(CSN5)
<i>Dbuz</i>	C(<i>scrib</i>)	E(<i>Wsck</i>)	A(CG12250)	F(CG8147)	B(<i>Ssadh</i>)	D(<i>Or98b</i>)	R(CSN5)	S(CG14899)	P(CG14645)	Q(CG14290)
<i>Dvir</i>	B(<i>Ssadh</i>)	A(CG12250)	D(<i>Or98b</i>)	C(<i>scrib</i>)	F(CG8147)	E(<i>Wsck</i>)	Q(CG14290)	P(CG14645)	S(CG14899)	R(CSN5)
<i>Dgri</i>	B(<i>Ssadh</i>)	A(CG12250)	D(<i>Or98b</i>)	C(<i>scrib</i>)	F(CG8147)	E(<i>Wsck</i>)	Q(CG14290)	P(CG14645)	S(CG14899)	R(CSN5)
<i>Agam</i>	<i>Gyc88E</i>	CG34349	?	?	<i>Mipp1</i>	<i>Nup107</i>	CG31222	CG9776	?	?

3.3.1. Inversiones 2m y 2n

La región AB incluye los genes *CG12250* (A), *CG4673* y *Ssadh* (B). El gen *CG4673* se caracterizó en la cadena complementaria aguas arriba del gen *CG12250* y aguas abajo del gen *Ssadh* en *D. mojavensis* (véase Figura 3.7). El orden y la orientación de estos tres genes se conservan en las 12 especies secuenciadas del género *Drosophila* (todas no invertidas para 2m). La primera descripción de una rotura en esta región en todo el género es la que se refiere a las inversiones 2m y 2n (este trabajo). Por lo tanto, significa que no se conocen roturas en ninguna de las especies secuenciadas del género que interrumpan el contexto intergénico aguas arriba y aguas abajo del gen *CG4673*. Sólo en *D. buzzatii* (y las demás especies invertidas para 2m) encontramos que el gen *CG4673* no está en el contexto habitual.

Por el contrario, en las especies comparadas no se conservan ni la región CD ni la región EF. En el grupo *melanogaster*, aguas arriba del gen *scrib* (C) está anotado el gen *CG6490*, y aguas arriba del gen *Or98b* (D) se encuentra *CG1866*. En este mismo grupo, aguas abajo del gen *Wsck* (E) está el gen *Syx18* mientras que aguas abajo de *CG8147* (F) está anotado el gen *RhoL*. En *D. persimilis* y *D. pseudoobscura* (grupo *obscura*) la ordenación de los genes en los puntos de rotura es la misma que la anotada en *D. mojavensis*. En *D. willistoni*, mientras que la región EF se conserva como en *D. mojavensis*, la región CD está interrumpida aunque con distintos genes implicados a los anotados para el grupo *melanogaster*. Aguas arriba del gen *scrib* (C) está anotado el gen *CG6089* y aguas arriba de *Or98b* (D) está el gen *CG6966*. Aparte de las roturas en el linaje de *D. buzzatii* para las inversiones 2m y 2n, parece que la región CD se ha interrumpido al menos otras dos veces y la región EF al menos una más en las especies secuenciadas del género *Drosophila*.

3.3.2. Inversión 2z⁷

En la región RS del punto de rotura de la inversión 2z⁷, el gen *Dlh* (S) sólo está anotado en los genomas de *D. mojavensis* y *D. virilis* (Delprat *et al.* 2009). Para el análisis de esta región, se consideró el siguiente gen situado aguas arriba de *Dlh* (gen *CG14899*) como el más cercano al punto de rotura en el extremo S. Por lo tanto, para este análisis consideramos que la región del punto de rotura RS se encuentra aguas arriba de *CSN5* (R) y aguas abajo de *CG14899* (S). La región RS se conserva en las especies secuenciadas del subgénero *Drosophila*. En las especies del subgénero *Sophophora*, aguas arriba de *CSN5* está anotado el gen *CG14897* mientras que aguas abajo de *CG14899* está anotado el gen *CG14898*. Por lo tanto, al menos una reorganización anterior implicando esta región se produjo en el género *Drosophila*. No se ha podido utilizar ninguna especie externa al género para determinar cuál era la organización ancestral o el subgénero donde ocurrió dicha reorganización.

La región del punto de rotura PQ también se conserva en todas las especies secuenciadas del subgénero *Drosophila*. En las especies secuenciadas del subgénero *Sophophora*, el gen *CG31222* está anotado aguas arriba de *CG14645* (P), mientras que el gen *CG9776* está anotado aguas arriba de *CG14290* (Q). En este caso, *Anopheles gambiae*, genoma secuenciado y usado como grupo externo, tiene la misma ordenación que el subgénero *Sophophora*. Esta ordenación se correspondería a la organización ancestral. Por consiguiente, la región PQ habría participado al menos en una rotura en el subgénero *Drosophila* aparte de la descrita en la inversión 2z⁷. La interrupción descrita en la inversión 2z⁷ de *D. buzzatii* en esta región reutilizaría el punto de rotura anterior.

3.4. Gen CG4673

3.4.1. El gen CG4673 en *D. buzzatii*

En *D. melanogaster*, el gen *CG4673* tiene una longitud de 13.647 pb (3R: 21436735..21.450.382 [-]). Se han descrito cuatro transcritos alternativos y un total de 11 exones distintos. Tres transcritos son de 8 exones (RA, RB y RD) y uno de 7 (RC), aunque sólo RA y RB tienen una evidencia sólida. Los cuatro transcritos tienen en común los 7 últimos exones aunque el transcrito RD tiene una 3'UTR alternativa y su último exón se ha designado con otro número. Los transcritos RD y RB son iguales a excepción de dicha 3'UTR. En Flybase se han numerado los exones según su localización aunque, al estar el gen codificado en la cadena complementaria, el número de los exones es ascendente desde la 3'UTR. De este modo, el último exón se ha designado como 1 (en RD) o 2 (en RA, RB y RC). El primer y segundo exón son el 11 y el 8 (RB y RD) y 10 y 8 (RA), respectivamente. Así pues, los transcritos RB y RD sólo difieren con respecto al transcrito RA en el primer exón. El primer exón de RC, en cambio, es el 9 y continúa con el exón 7 (tercero en los demás transcritos). Los cuatro transcritos codifican 4 proteínas distintas de 652 aa (PA), 624 aa (PB), 553 aa (PC) y 624 aa (PD).

Actualmente en *D. mojavensis* únicamente están actualmente anotados los 8 exones que codifican el transcrito RA (11.726 pb, sin contar las UTRs) que codifica una proteína de 653 aa. Sin embargo, hemos detectado 678 pb aguas arriba del primer exón, un pequeño exón alternativo (15 pb, que en lo sucesivo denominaremos exón 0), también conservado en *D. virilis* y *D. grimshawii*, correspondiente al primer exón putativo que codifica el inicio del transcrito RB. Como en *D. melanogaster*, los transcritos RA y RB sólo difieren en el primer exón. El transcrito RB en *D. mojavensis* tiene un tamaño de 12.418 pb (sin UTR), codificando

una proteína de 624 aa.

En el punto de rotura BE (20O19) en *D. buzzatii*, se anotaron 9 exones del gen *CG4673*, 8 en el CDS de RA (3.015 pb) y 8 en el CDS de RB (3.701 pb). Ambos transcritos sólo difieren en el primer exón. La secuencia codificante del gen *CG4673* está altamente conservada con sus ortólogos en *D. melanogaster* y *D. mojavensis* pero el gen es considerablemente más corto debido a la reducción drástica del tamaño del intrón 6 (Tabla 3.13). A diferencia de Flybase, nosotros hemos numerado los exones desde la 5'UTR. En la Tabla 3.13 se han anotado las equivalencias de los nombres de cada exón con los de Flybase. El transcrito RA de *D. buzzatii* tiene una identidad nucleotídica del 90,6% con el de *D. mojavensis* y del 79,6% con el de *D. melanogaster* mientras que el transcrito RB tiene una identidad nucleotídica del 90,0% con el de *D. mojavensis* y del 78,9% con el de *D. melanogaster*. Ambos transcritos sólo difieren en el primer exón (exón 0 y exón 1 en los transcritos RB y RA, respectivamente). Las identidades en la comparación del exón 0 son menores que las del exón 1. De todos modos, el exón 0 tan sólo consta de 5 aa en *D. buzzatii* y *D. mojavensis* y de 6 aa en *D. melanogaster*, por lo que ambos transcritos están igualmente constreñidos con valores de ω (Ka/Ks) parecidos (0,0593 y 0,0557 en la comparación de los transcritos RB y RA entre *D. buzzatii* y *D. mojavensis* y 0,0489 y 0,0505 en la comparación de los transcritos RB y RA entre *D. buzzatii* y *D. melanogaster*). Las diferencias en el número de pares de bases anotados en cada exón son siempre múltiples de tres, por lo que en ningún caso se altera el marco de lectura ni los sitios de procesamiento. En los exones 7 y 8 también se ha anotado un triplete de diferencia con la anotación de *D. buzzatii* en *D. mojavensis* y *D. melanogaster*, respectivamente. El exón 3 es el más variable en número de pares de bases: 170 en *D. buzzatii*, 167 en *D. mojavensis* y 158 en *D. melanogaster*. Este exón es el menos constreñido de todos con los valores de ω más elevados (sin tener en cuenta el exón 0). En ambas comparaciones se observan los valores de ω más bajos para el exón 6, indicando que es

el más constreñido (0,0133 entre *D. buzzatii* y *D. mojavensis* y 0,0092 entre *D. buzzatii* y *D. melanogaster*).

En la Figura 3.11 se muestra el alineamiento (desde N-terminal a C-terminal) de las traducciones conceptuales del transcrito RA del gen *CG4673* de *D. melanogaster* (*Dmel*), *D. mojavensis* (*Dmoj*) y *D. buzzatii* (*Dbuz*). Se han señalado el fragmento traducido en cada exón (del 1 al 8). En la Figura 3.12 se muestra únicamente el alineamiento de la traducción conceptual de los exones 0 (específico del transcrito RB) y 2 (segundo exón compartido con RA).

Tabla 3.13: (Página siguiente). Estructura y comparación del gen *CG4673* de *D. buzzatii* en la región del punto de rotura en BE con las respectivas copias de *D. mojavensis* y *D. melanogaster*. Los exones del 1 al 8 codifican el transcrito RA. Los exones 0 y del 2 al 8 codifican el transcrito RB. *: Equivalencia con los nombres de los exones del Flybase. Los CDS no incluyen las UTR.

Región	Dbuz	Dmoj	<i>Dmel</i>	<i>Dbuz/Dmoj</i>			<i>Dbuz/Dmel</i>		
			(nombre)*	Identidad(%)	Ka/Ks	ω	Identidad(%)	Ka/Ks	ω
Exón 0	15	15	18 (11)	86,7	0,0834/0,6355	0,13123	66,7	0,2758/0,8240	0,3347
Intrón 0	846	853	1183						
Exón 1	102	102	102 (10)	95,1	0/0,2441	0	84,3	0,0532/0,8137	0,0654
Intrón 1	73	74	72						
Exón 2	234	234	234 (8)	93,2	0,0114/0,2965	0,0384	80,8	0,0713/1,0868	0,0656
Intrón 2	62	55	62						
Exón 3	170	167	158 (7)	85,0	0,0739/0,5360	0,1378	68,3	0,2572/1,9331	0,1330
Intrón 3	237	80	61						
Exón 4	529	529	529 (6)	89,6	0,0249/0,5084	0,0490	80,1	0,0475/2,1438	0,0221
Intrón 4	59	52	60						
Exón 5	199	199	199 (5)	93,0	0,0134/0,3081	0,0435	77,9	0,0779/1,9922	0,0391
Intrón 5	61	76	62						
Exón 6	161	161	161 (4)	89,4	0,0081/0,6078	0,0133	82,6	0,0163/1,7603	0,0092
Intrón 6	475	9.355	9.942						
Exón 7	416	413	416 (3)	91,5	0,0193/0,3722	0,0518	82,0	0,0652/0,9408	0,0693
Intrón 7	80	72	72						
Exón 8	157	157	160 (2/1)	87,9	0,0432/0,5301	0,081	77,1	0,0798/2,5959	0,0307
Transcrito RA	3.015	11.726	12.290						
CDS RA	1.968	1.962	1.959	90,6	0,0235/0,4218	0,0557	79,6	0,0723/1,4327	0,0505
Transcrito RB	3.701	12.418	13.317						
CDS RB	1.881	1.875	1.875	90,0	0,0259/0,4363	0,0593	78,9	0,0728/1,4874	0,0489

(pfam05020, E-value= $1e-65$). En esta proteína de *D. buzzatii* se encontró, aparte de este dominio ortólogo a la proteína NPL4 de humano, otro dominio en la región C-terminal aunque con baja significación (cl02656, E-value= 0.003). Dicho dominio corresponde al dedo de zinc de las proteínas RAN. El dominio NPL4 se codifica en los exones del segundo al séptimo. Por otro lado, el dedo de zinc de las proteínas RAN anotado en el extremo C-terminal se codifica en el exón 8. El dominio NPL4 está ampliamente conservado (incluyendo el dominio activo y el dedo de zinc N-terminal). El dedo de zinc C-terminal adicional se ha encontrado en metazoos y hongos pero parece ser menos constante en los demás organismos.



Nombre del dominio conservado homólogo	E-value
Dominio de la superfamilia NPL4	$5e-92$
Región de unión a dedo de zinc de la superfamilia NPL4	$1e-65$
zf-RanBP, dedo de zinc en proteínas de unión RAN	0.003
COG5100, multi-dominio NPL4 (<i>Nuclear Pore Localization Protein</i>)	$3e-103$

Figura 3.13: Dominios conservados (arriba) de la proteína PA del gen *CG4673* del punto de rotura BE en *D. buzzatii*. El nombre del dominio conservado y el E-value de la identidad con éste también se muestran (Tabla inferior). El multi-dominio NPL4 está compuesto por un dedo de zinc y un dominio activo de la superfamilia NPL4. Finalmente, un dedo de dedo de zinc homólogo a los de proteínas de unión a RAN, se encuentra en la región C-terminal.

3.4.3. NPL4 en el género Drosophila

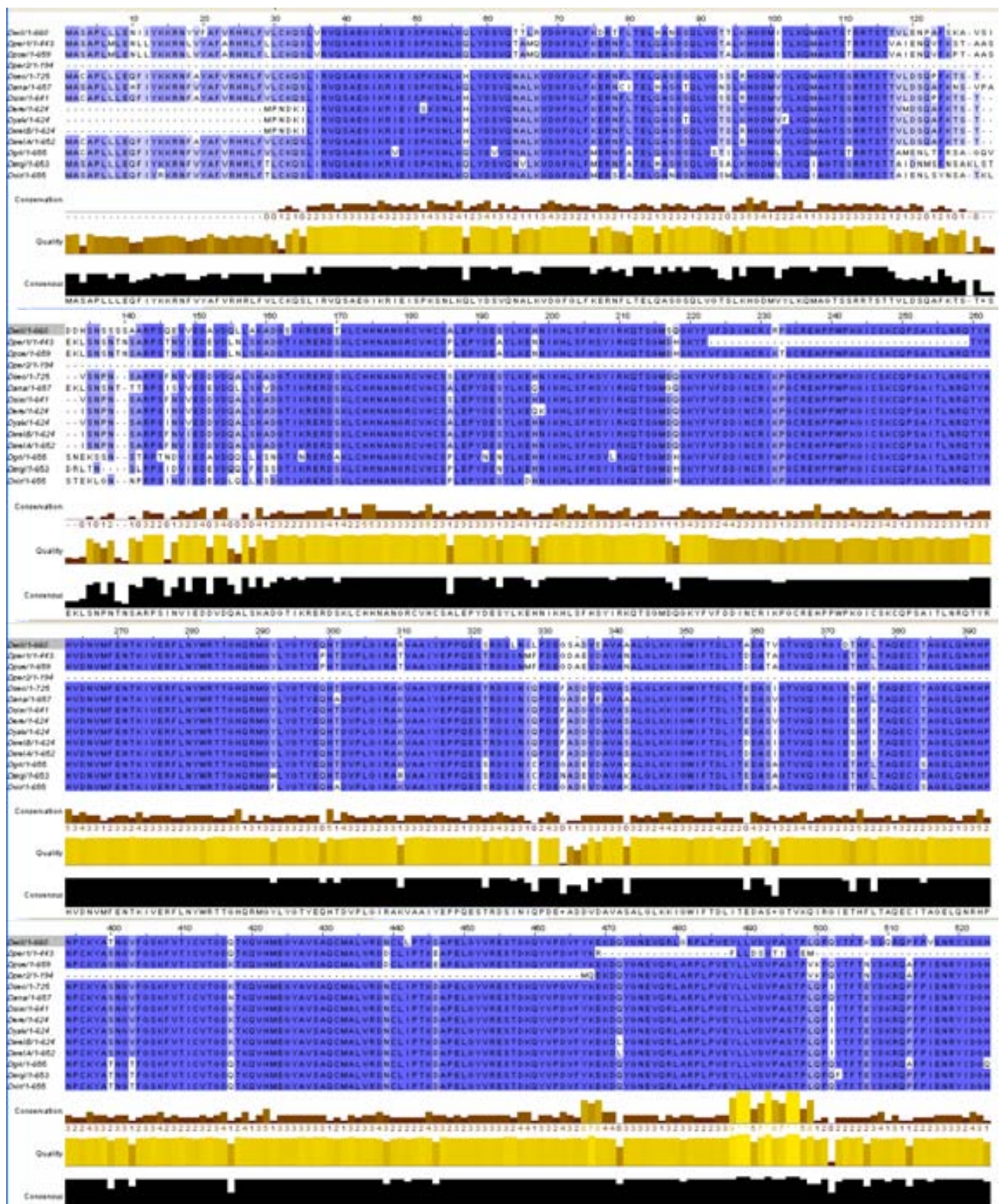
Los ortólogos del gen *NPL4* en los 12 genomas secuenciados de *Drosophila* están compilados en la Tabla 3.14. Para cada especie se muestra: el nombre del gen ortólogo, el número de exones anotados y el tamaño y transcrito conservado para dicho gen. Los ortólogos para los genes *CG5071* y *CG5079* anidados en el intrón 6 del gen ortólogo a *CG4673* también se muestran en dicha Tabla. En algunos casos, estos dos genes no están anotados como ortólogos de los genes *CG5071* y *CG5079* de *D. melanogaster* (*Berkeley Drosophila Genome Project*). Dicha anotación no los considera ortólogos debido a su baja identidad nucleotídica con los genes *CG5071* y *CG5079* por lo que no parecen genes muy constreñidos en la filogenia de *Drosophila*. Manualmente hemos revisado estos casos y han sido incluidos como ortólogos (marcados con un asterisco en la Tabla 3.14).

D. yakuba y *D. erecta* únicamente tienen anotado el transcrito RB mientras que el resto de especies, tan sólo tendrían anotado el transcrito RA. Sin embargo, es posible que se trate de una anotación incompleta como ha sucedido con *D. mojavensis* (véase sección 3.4.1). Las proteínas anotadas de las 12 especies secuenciadas (incluyendo ambos transcritos de *D. melanogaster*) y la de *D. buzzatii* han sido alineadas. Existe un alto grado de conservación (Figura 3.14) a excepción de las proteínas de *D. sechellia* (más larga) y de *D. persimilis* (truncada y renombrada como *Dper1* y *Dper2*).

Tabla 3.14. Nombre de los genes ortólogos a *CG4673*, *CG5079* y *CG5071* en las 12 especies secuenciadas del género *Drosophila*. Para los genes ortólogos a *CG4673* se muestran el número de exones anotados, la traducción conceptual conservada y su tamaño (*Berkeley Drosophila Genome Project*). *Genes considerados ortólogos por localización pero no anotados como tal según Flybase. El gen *CG4673* está anotado como truncado en *D. persimilis* (*Dper1* y *Dper2*).

Especie	Ortólogos de <i>CG4673</i>	Exones	Traducción (aa)	Ortólogos de <i>CG5079</i>	Ortólogos de <i>CG5071</i>
<i>D. simulans</i>	<i>Dsim\GD18187</i>	8	PA(641)	<i>GD21225</i>	<i>GD21226</i>
<i>D. sechellia</i>	<i>Dsec\GM10234</i>	12	PA (725)	<i>GM10252</i>	<i>GM10252*</i>
<i>D. melanogaster</i>	<i>Dmel\CG4673</i>	8	PA(652) PB (624)	<i>CG5079</i>	<i>CG5071</i>
<i>D. yakuba</i>	<i>Dyak\GE10682</i>	8	PB (624)	<i>GE23610</i>	<i>GE23611</i>
<i>D. erecta</i>	<i>Dere\GG12234</i>	8	PB (624)	<i>GG11414</i>	<i>GG11415</i>
<i>D. ananassae</i>	<i>Dana\GF17266</i>	8	PA (657)	<i>GF19920*</i>	<i>GF17698</i>
<i>D. pseudoobscura</i>	<i>Dpse\GA18345</i>	8	PA (659)	<i>GA18645*</i>	<i>GA18637</i>
<i>D. persimilis</i>	<i>Dper1\GL21676</i>	8	PA (443)	<i>GL22149*</i>	<i>GL22150</i>
	<i>Dper2\GL21677</i>	2	PA (194)		
<i>D. willistoni</i>	<i>Dwil\GK14396</i>	8	PA (660)	<i>GK18936*</i>	<i>GK12851</i>
<i>D. mojavensis</i>	<i>Dmoj\GI10357</i>	8	PA (653)	<i>GI22408*</i>	<i>GI22409</i>
<i>D. buzzatii</i>	<i>Dbuz\CG4673</i>	8	PA (655)	-	<i>CG5071</i>
<i>D. virilis</i>	<i>Dvir\GJ10209</i>	8	PA (655)	<i>GJ10968*</i>	<i>GJ10969</i>
<i>D. grimshawi</i>	<i>Dgri\GH18610</i>	8	PA (655)	<i>GH18983*</i>	-

Resultados



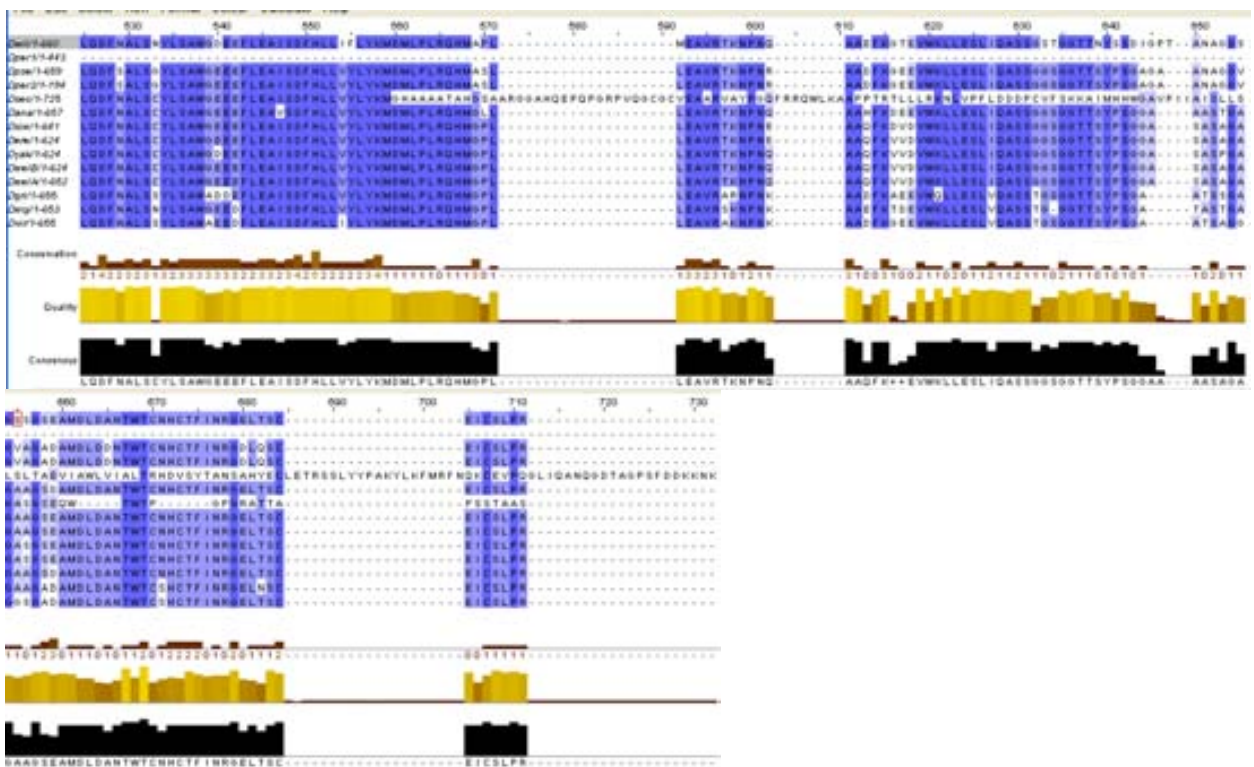
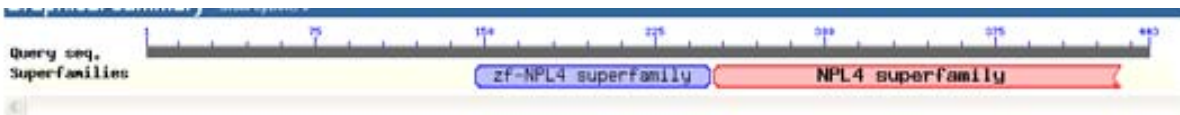


Figura 3.14: Comparación global de las proteínas codificadas por el gen *CG4673* en las 12 especies secuenciadas del género *Drosophila*. A la izquierda de cada panel se indica la especie y el tamaño de la proteína en aminoácidos (aa). Para *D. persimilis* se han tomado en cuenta ambas traducciones conceptuales (*Dper1*, segunda secuencia y *Dper2*, cuarta secuencia, véase texto para detalles). Para *D. melanogaster* se han incluido en el alineamiento las proteínas codificadas por los transcritos RB (décima secuencia) y RA (onceava secuencia). Las tres pistas bajo el alineamiento representan la conservación de la secuencia (marrón), la calidad, es decir, el número de aa para el análisis de cada sitio (amarillo) y finalmente la cantidad de aa para cada sitio de todas las secuencias que corresponden al consenso (negro). Debajo de la última pista se indica la secuencia consenso de la proteína.

Aparte de estas dos especies, los dominios conservados en las proteínas ortólogas son los mismos que los anteriormente encontrados en *D. buzzatii*. Por un lado, *D. sechellia* parece tener una terminación diferente debido a una mutación en el sitio de procesamiento del exón 6 que cambiaría el marco de lectura produciendo una proteína que carece del dominio del dedo de zinc (Figura 3.15). Por otro lado, *D. persimilis* parece tener el gen *CG4673* y en consecuencia el dominio NPL4 truncado, quedando anotado como dos transcritos independientes. Una mutación puntual incorpora un codón stop en la secuencia codificante. El alineamiento de la proteína (Figura 3.14) muestra que la proteína *Dper2* de *D. persimilis* comienza cuando termina *Dper1*. No parece que ambas mutaciones sean viables debido a la constricción que sufre esta proteína en todo el género. Al tratarse ambos casos de mutaciones puntuales, podría deberse a errores puntuales de la secuenciación.

Dper1\GL21676



Dper2\GL21677



Dsec\GM10234

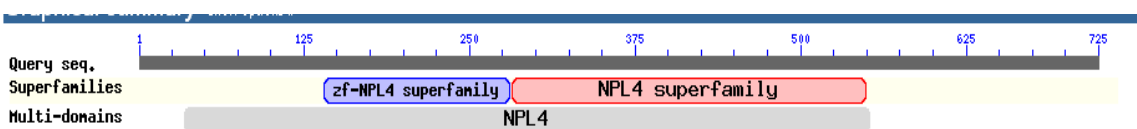


Figura 3.15: Dominios conservados de los transcritos putativos del gen *CG4673* de *D. persimilis* (*Dper1* y *Dper2*) y *D. sechellia*.

3.4.4. Región reguladora y promotora del gen *CG4673*

El gen *CG4673* se encuentra en un nuevo entorno genético en *D. buzzatii* y en las otras especies con la inversión *2m* (véase secciones 3.1.9 y 3.1.10). El transcrito y la proteína de este gen presentan un alto grado de conservación (véase sección 3.4.2) indicando un fuerte constreñimiento funcional. La región reguladora y promotora de este gen también podría estar fuertemente constreñida.

Sólo una pequeña fracción anotada como región intergénica se conserva aguas arriba y aguas abajo del gen *CG4673*. En el genoma invertido de *D. buzzatii* se encontraron dos copias del gen *CG4673*, quedando una en el punto de rotura AC y otra en el punto de rotura BE. La rotura que separa las regiones A y B en los genomas invertidos, se localiza justo antes del exón 0, y por consiguiente, la copia degenerada situada en el punto de rotura AC conserva la región aguas abajo (región intergénica entre *CG4673* y *CG12250*). Por el contrario, la copia funcional en BE en cambio, arrastró con la inversión la región aguas arriba (región intergénica entre *CG4673* y *Ssadh*). Por lo tanto, formulamos como hipótesis inicial que la región reguladora se encontraría en la región 5' (región intergénica entre los genes *CG4673* y *Ssadh*). La nueva región intergénica situada aguas arriba de la copia en AC (región intergénica entre los genes Ψ *CG4673* y *scrib*) no mantiene dicha región reguladora lo que podría explicar su degeneración al dejar de transcribirse.

Por otro lado, debido a la aparente microinversión de la copia funcional de *CG4673* en BE, la región aguas arriba (donde hipotéticamente debe encontrarse la región reguladora) se encuentra ahora también aguas abajo del gen. La microinversión no parece haber afectado la transcripción del gen *CG4673* debido al constreñimiento al que está sometido el gen en *D. buzzatii*. Dicho constreñimiento observado en el gen de *D. buzzatii* indicaría que éste sigue siendo funcional. El fragmento anotado en la región 5' del gen *CG4673* de *D. buzzatti* anotado

como región 5' del gen *CG4673* de *D. mojavensis*, comprende desde el codón de inicio hasta la repetición invertida. Por tanto, el promotor del gen debe encontrarse no más lejos en 5' que la región anotada como repetición invertida. Esta región delimita el fragmento anotado en 5' de los genes en especies no invertidas que la copia funcional en *D. buzzatii* sigue conservando aguas arriba. Esta pequeña región duplicada e invertida en la región aguas abajo del gen (BE), también pudo ser anotada en la región aguas arriba de Ψ *CG4673* en AC (véase Figura 3.7). Si la región reguladora se encuentra en este fragmento, no se interrumpiría la transcripción de la copia del gen en el punto de rotura AC y por consiguiente, tampoco existirían motivos que justificaran su degeneración.

La región 5' del gen *CG4673* (desde el exón 1 hasta el gen *Ssadh*) en *D. melanogaster* fue alineada con la región ortóloga de algunas especies del subgénero *Sophophora* y visualizada en el navegador VISTA (Figura 3.16). En las 5 especies comparadas, parece que existe una región altamente conservada dentro del intrón 0 (primer intrón del transcrito RB, entre el exón 0 y el exón 2) aparte del exón 1 y la 5'UTR del transcrito RA (también incluidos en el intrón 0). La anotación de la 5'UTR del transcrito RA termina en la posición 21.449.459 pb, mientras que dicha región conservada empieza en el siguiente nucleótido. El otro pico de conservación fuera de la transcripción putativa se observa sólo entre las especies del grupo *melanogaster* (especies con el transcrito RB anotado, véase Tabla 3.14) y se localiza justo antes de la 5' UTR del transcrito RB. Ni *D. ananassae* ni *D. persimilis* muestran conservación en esta región ni para el exón 0 del transcrito RB. Por otro lado, el algoritmo para la búsqueda de promotores putativos compilado en Flybase (Ohler *et al.* 2002) señala dos promotores putativos dentro del intrón 0 de *D. melanogaster* como posibles reguladores de este gen.

D. buzzatii no conserva esta secuencia, por lo que no fue alineada en la Figura 3.16 pero la región promotora putativa del gen *CG4673* se espera aguas arriba y próxima al inicio de la 5'UTR, de acuerdo con las alineaciones de otros genomas del género *Drosophila*.

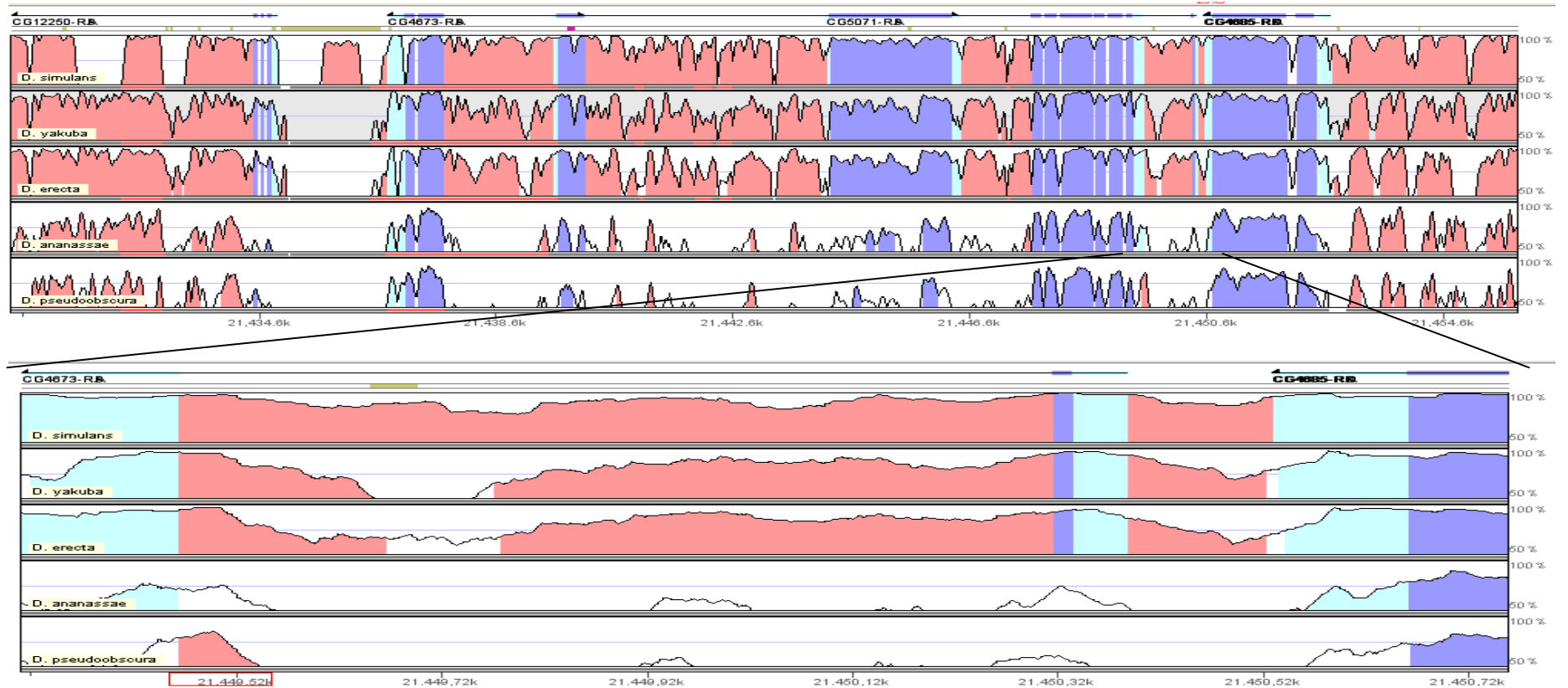


Figura 3.16: Alineamiento nucleotídico de *D. simulans*, *D. yakuba*, *D. erecta*, *D. ananassae* y *D. pseudoobscura* con *D. melanogaster* (genoma usado como referencia que no aparece en el alineamiento) de la región del gen *CG4673* (transcritos RA y RB) incluyendo los genes *CG12250* y *Ssadh* (*CG4685*) con el navegador VISTA. La altura de los picos indica el grado de conservación de cada especie alineada respecto a *D. melanogaster*. Los picos azules corresponden a regiones anotadas como exones, los rosados las regiones no codificantes y los azul celeste las UTR. En la parte inferior, se muestra un detalle de la región desde la 5'UTR del transcrito RA de *CG4673* hasta la 3'UTR y parte del último exón del gen *Ssadh*. La conservación, en color rosado, observada entre la 5'UTR de *CG4673*\RA y el exón 0 de *CG4673*\RB corresponde al intrón 0. La conservación, en color rosado, observada entre la 5'UTR de *CG4673*\RB y la 3'UTR de *Ssadh* corresponde a la región intergénica entre ambos genes. *D. simulans*, *D. yakuba* y *D. erecta* conservan el exón 0 y la 5'UTR del transcrito RB.

La posibilidad de que esta región pudiera ser la región de regulación y el inicio de transcripción es consistente con la anotación realizada en *D. buzzatii* ya que esperamos que la regulación del gen se encuentre entre el codón inicio y el punto de rotura de la microinversión.

El intrón 0 pudo ser anotado en el gen *CG4673* de *D. buzzatii*. Además, $\Psi CG4673$ en el punto de rotura AC también tiene similitud con este intrón. Si la región reguladora putativa del transcrito RA se encuentra en dicho intrón, tanto *CG4673* como $\Psi CG4673$ podrían seguir siendo regulados independientemente de las reorganizaciones que han tenido lugar. Por lo tanto, no existen evidencias, como mínimo a nivel de regulación, sobre las causas que promovieron la pseudogeneización del gen en el punto de rotura AC.

Para la detección de promotores putativos que pudieran regular el gen *CG4673* y el pseudogen $\Psi CG4673$ de *D. buzzatii*, las regiones completas de los dos puntos de rotura de AC (15.343 pb entre los genes *CG12250* y *scrib*) y BE (8.242 pb entre los genes *Ssadh* y *Wsck*) han sido tomadas en cuenta. Ambos fragmentos han sido analizados con el programa McPromoter (*The Markov Chain Promoter Predicción Server at Duke University, Institute for Genome Sciences and Policy*). Las secuencias se analizaron según la Figura 3.7, por lo que en ambos puntos de rotura, *CG4673* y $\Psi CG4673$ se localizan en la cadena en sentido 3'-5' (cadena complementaria). De este modo, el sentido de la Figura 3.7 (directo) es opuesto al sentido de la transcripción en ambos casos (reverso). Por otra parte se estableció un umbral de significación de 0,035 que corresponde a un nivel de alta sensibilidad para detectar promotores candidatos en estas regiones. La localización de los picos de los posibles promotores se muestra en la Figura 3.17.

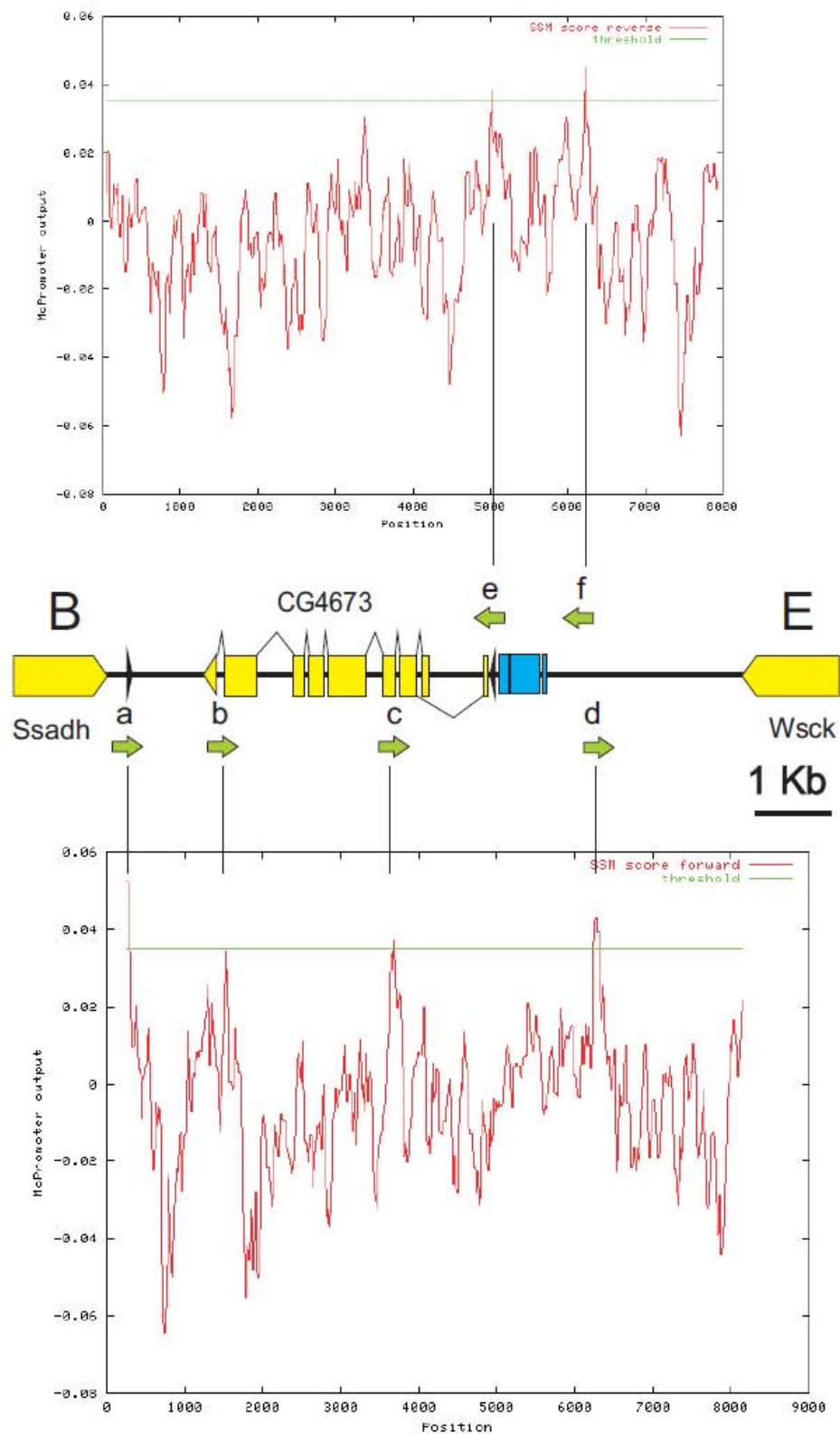


Figura 3.17: (ver Página siguiente).

Figura 3.17: (Página anterior): Predicción bioinformática de promotores. Análisis de los 8.355 pb del punto de rotura BE de *D. buzzatii* (región entre los genes *Ssadh* y *Wsck*, no incluidos) mediante el programa McPromoter. En la secuencia, están anotados los exones de los genes (amarillo), los fragmentos de *BuT5* (azul) y las repeticiones invertidas (triángulos invertidos negros). Se señalan los promotores putativos (flechas verdes) cuando se supera el umbral de significación de 0,035 (línea verde en los gráficos). En el gráfico superior, se observan dos picos de significación en el sentido reverso de la secuencia (*e*, *f*). En el gráfico inferior, se observan cuatro picos de significación en el sentido directo de la secuencia (*a*, *b*, *c*, *d*).

En la región del punto de rotura AC no se detectaron promotores putativos ni directos ni reversos en *D. buzzatii*. Si se reduce el umbral de aceptación de la significación a 0,03, el programa sugiere un promotor dentro de *ΨCG4673* y en el sentido directo aguas arriba del gen *CG5071*. La interpretación más plausible es que se trate del promotor de dicho gen anidado.

En el punto de rotura BE, distintos promotores putativos fueron detectados (Figura 3.17), cuatro en el sentido directo (*a*-0,05546, *b*-0,03508, *c*-0,03668 y *d*-0,04454) y dos en el sentido reverso (*e*-0,04066, *f*-0,05057). Los promotores *b* y *c* (los menos significativos) se localizan en las regiones exónicas del gen *CG4673* (exones 7 y 3 respectivamente), por lo que se consideran artefactos. El promotor *d* se localiza en la región intergénica entre *CG4673* y *Wsck* pero orientado hacia este último. Por último, el promotor más significativo (*a*) se localiza aguas abajo del gen *CG4673* y orientado hacia éste. Este promotor se localizaría en la secuencia anotada como repetición invertida en 3' del gen *CG4673*. En el sentido reverso, los promotores putativos *e* y *f* se encuentran aguas arriba del gen y orientados hacia éste, pudiendo estar involucrados con su regulación. El promotor *e* se localizaría en la secuencia anotada como repetición invertida en 5' del gen *CG4673*.

3.4.5. Repeticiones invertidas

Una repetición invertida de 121 pb flanquea el gen *CG4673* en el punto de rotura BE de *D. buzzatii* (véase Figura 3.7). Estas secuencias aguas arriba y abajo del gen son idénticas entre si en un 72,7% de los 121 pb alineados (E-value= 2e-16). Ambas repeticiones invertidas han sido sugeridas como las regiones promotoras (véase Figura 3.17). En ambos casos, están orientadas al gen, uno aguas arriba (promotor *e*) y el otro aguas abajo (promotor *a*).

En la región 5' de $\Psi CG4673$ en AC (véase Figura 3.7) se pudieron anotar 92 pb de esta secuencia. En *D. mojavensis* 93 pb de esta región fueron encontrados en la región intergénica en 5' del *CG4673*. En *D. mojavensis* se ha anotado como una secuencia única al no localizarse en la región 3' del gen *CG4673* como en *D. buzzatii* (véase Figura 3.7). Esta secuencia forma parte del fragmento duplicado entre los puntos de rotura AC y BE asociado a la inversión *2m*. Esta reorganización no explica aparentemente por qué esta secuencia también está duplicada en BE (repetición invertida) flanqueando el gen *CG4673*. La secuencia de la repetición invertida podría corresponderse originalmente a la región promotora del gen. De este modo, dos promotores rodearían el gen en el punto de rotura BE, y otro se encontraría en la región 5' del gen en el punto de rotura AC (aunque en este caso el programa McPromoter no detectó nada).

Las regiones intergénicas que flanquean el gen *CG4673* se secuenciaron en otro genoma invertido, *D. koepferae* (Figura 1.3). La región intergénica en 5' se amplificó con un cebador en el exón 1 (*firstL*) y otro en el gen adyacente aguas arriba (*WschR*). La región intergénica en 3' se amplificó con un cebador en el exón 8 (*lastR*) y otro en el gen adyacente aguas abajo (*SsadhL*). Los cebadores fueron diseñados en las regiones más conservadas de *D. buzzatii* con *D. mojavensis* para poder amplificar el genoma de *D. koepferae* (Figura 3.18). La secuencia de los cebadores se muestra en la Tabla 2.4 de Materiales y Métodos). Ambas

regiones intergénicas fueron correctamente amplificadas en *D. koepferae*: ~3.600 pb la región intergénica aguas arriba (4.074 pb en *D. buzzatii*) y ~1.100 pb la región intergénica aguas abajo (1.261 pb en *D. buzzatii*). Cuando estas regiones intergénicas se alinearon entre si mediante bl2seq, se detectó una repetición invertida como en *D. buzzatii*. Esta repetición invertida de 123 pb (121 pb en *D. buzzatii*) también flanquea el gen *CG4673*. Estas secuencias aguas arriba y abajo del gen son idénticas entre si en un 74,4% de los 123 pb alineados (E-value= $2e-20$). Si esta secuencia fuera intergénica y no estuviera implicada en la regulación del gen, no debería sufrir ningún tipo de constreñimiento y la cantidad de los cambios acumulados al comparar dos secuencias debería ser mayor cuanto más tiempo hubieran estado divergiendo. La mayor cantidad de cambios se observaría al comparar la secuencia en 5' del gen *CG4673* de *D. mojavensis* con las de las especies invertidas. Posteriormente tuvo lugar la inversión *2m* y la duplicación del gen *CG4673* (incluyendo esta región en 5'). Esto supone empezar a acumular diferencias entre las secuencias a 5' del gen en AC y a 5' del gen en BE dentro de las especies invertidas. La duplicación de esta secuencia a ambos lados del gen empieza a hacer divergir las secuencias a 5' y 3' del gen funcional. Finalmente, la divergencia entre las especies *D. buzzatii* y *D. koepferae* permitiría acumular diferencias entre las secuencias de estas especies.

Las secuencias anotadas como repetición invertida en la regiones 5' y 3' del gen *CG4673* de *D. koepferae* y *D. buzzatii* (*Dkoe*\5', *Dbuz*\5', *Dkoe*\3' y *Dbuz*\3') y en 5' de Ψ *CG4673* y *D. mojavensis* (*Dbuz*\5'AC y *Dmoj*) fueron comparadas a pares usando el programa de alineamiento MUSCLE (Figura 3.19). Las identidades entre las secuencias en cada comparación se muestran en la Tabla 3.15. Se observan pocos cambios entre las secuencias *Dbuz*\5' y *Dkoe*\5', divergidas hace menor cantidad de tiempo (identidad del 90,8%). En cambio, la cantidad de cambios acumulados en las demás comparaciones no es consistente con la cronología de los eventos en caso de que esta secuencia fuera neutra. Se

observa que las secuencias en 5' del gen funcional, *Dbuz*\5' y *Dkoe*\5' presentan pocos cambios respecto a la secuencia de *D. mojavensis* (identidades del 89,2 y 87,1% respectivamente). En cambio, estas mismas secuencias acumulan más cambios en la comparación con la secuencia en 5' del pseudogen (*Dbuz*\5'AC), (identidades del 68,5 y 63,4% respectivamente), y algunos menos con la repetición anotada en 3' del gen (identidades del 72,7 y 75,6% respectivamente).

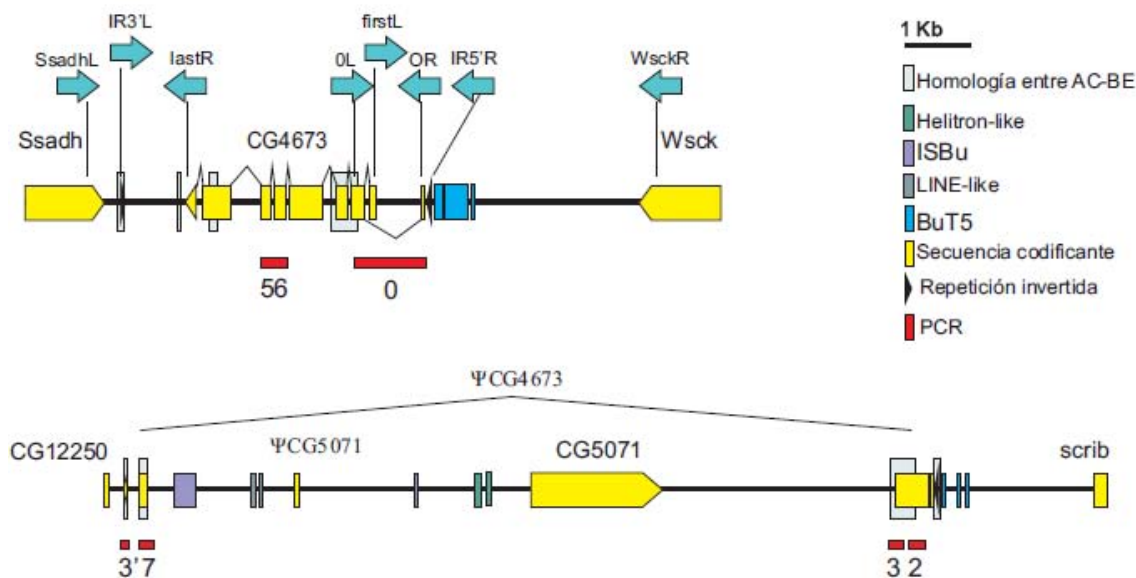


Figura 3.18: Regiones de los puntos de rotura BE (arriba) y AC (abajo) de *D. buzzatii*. Los fragmentos de AC que muestran similitud con BE (véase Tabla 3.17 más adelante) corresponden a los recuadros en azul claro. Las RT-PCR (líneas rojas) diseñadas se muestran en la región que amplifican. Las PCR 56 y 0 amplifican la copia funcional de *CG4673* (BE). Las PCR 3', 7, 3 y 2 amplifican Ψ *CG4673* (AC). También se muestran en la Figura la región y orientación de los cebadores diseñados en estas regiones.

A)

```

Dbuz\5'      TAACAAACCTCTTCAACAATTAGCATTGGTATTTTATCTCATAAACTTTTAGTTACATT 60
Dkoe\5'      -----ACAATTAGCAGCATTGGTATTTTATCTTATAAACTTTTAGTAACATT 48
Dmoj         -----AGCATTCGGTATTTTATCT-----TTTGTAGTACCATT 32
Dkoe\3'      -----ACAATTAATAGCATTCGATATTTTATCTTATTGACTTTTATACATATT 48
Dbuz\3'      TAACAGAGTTATTCAATTAATAGCATTTCGATATTTTATCTTATTGGCTTTTATTCGTATT 60
Dbuz\5' AC   -----GTATTTTATCCTATTGGCTTTCAGCTACATT 31
                *****          *** *      ***

Dbuz\5'      TGATTACTGTGTGCATGGTCACACTTACTCCCCACTGTCTACACAAGTGCTGATTTGCTT 120
Dkoe\5'      TTCTTACTGTATGCATGGTCACACTTACTCCCCACTGTCTACACAAGTGCTGATTTGCTT 108
Dmoj         AGAATAGTGTGCGCATGGTCACACTTAGTCCCTACTGTCTACACAAGTGCTGTTTGTCTT 92
Dkoe\3'      TGAATATTGAGCGAATGGTCACACTAAGTCCCTATTGCCTACACAAACGCTGCTTTGCTT 108
Dbuz\3'      TGAATATTATGCGAATGGTCACACTAAATCCCTATTGCCTTAACAAACGCTGCTTTGCTT 120
Dbuz\5' AC   TCGCTCTTGTGCGAACGGTCACACTTAAATTGTATAGCCAACAAAAGTTCTGCTTTCCTT 91
                * *      * * *****          * * *      * * *      * * *      * *

Dbuz\5'      C----- 121
Dkoe\5'      CCGAGTGCTATTTT-- 123
Dmoj         C----- 93
Dkoe\3'      TCAAATTCGTTATTTT 125
Dbuz\3'      C----- 121
Dbuz\5' AC   C----- 92
    
```

B)

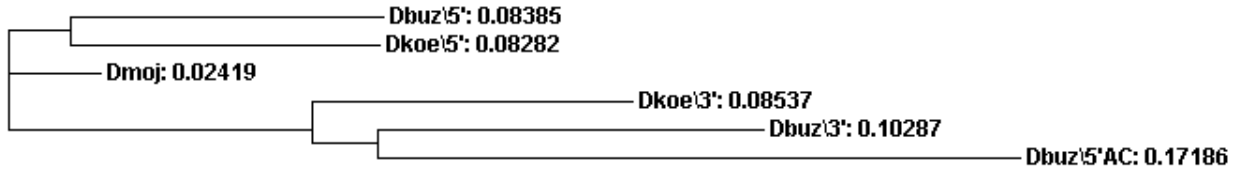


Figura 3.19: Comparación de las secuencias anotadas como repetición invertida en la regiones 5' y 3' del gen *CG4673* de *D. koepferae* y *D. buzzatii* (*Dkoe\5'*, *Dbuz\5'*, *Dkoe\3'* y *Dbuz\3'*), y ortólogas en 5' de Ψ *CG4673* (*Dbuz\5' AC*) y en *D. mojavensis* (*Dmoj*). A) Alineamiento de todas las secuencias. B) Filograma. En primer lugar se han calculado las distancias (% de divergencia) entre todas las secuencias a pares a partir de los datos del alineamiento en la Tabla 3.15. Posteriormente se ha aplicado el método de *Neighbour Joining* (Saitou y Nei 1987) a la matriz de distancias para calcular el diagrama de ramificación. La cantidad de cambios se muestra en cada rama.

Tabla 3.15: Identidad entre las secuencias anotadas como repetición invertida en la regiones 5' y 3' del gen *CG4673* de *D. koepferae* y *D. buzzatii* (*Dkoe*\5', *Dbuz*\5', *Dkoe*\3' y *Dbuz*\3'), y en 5' de Ψ *CG4673* (*Dbuz*\5' AC) y de *D. mojavensis* (*Dmoj*).

Secuencia 1	Tamaño (pb)	Secuencia 2	Tamaño (pb)	Identidad (%)
<i>Dbuz</i> \5'	121	<i>Dbuz</i> \3'	121	72,7
<i>Dbuz</i> \5'	121	<i>Dbuz</i> \5'AC	92	68,5
<i>Dbuz</i> \5'	121	<i>Dmoj</i>	93	89,2
<i>Dbuz</i> \5'	121	<i>Dkoe</i> \3'	125	76,6
<i>Dbuz</i> \5'	121	<i>Dkoe</i> \5'	123	90,8
<i>Dbuz</i> \3'	121	<i>Dbuz</i> \5'AC	92	69,2
<i>Dbuz</i> \3'	121	<i>Dmoj</i>	93	72,0
<i>Dbuz</i> \3'	121	<i>Dkoe</i> \3'	125	91,6
<i>Dbuz</i> \3'	121	<i>Dkoe</i> \5'	123	71,3
<i>Dbuz</i> \5'AC	92	<i>Dmoj</i>	93	63,0
<i>Dbuz</i> \5'AC	92	<i>Dkoe</i> \3'	125	71,1
<i>Dbuz</i> \5'AC	92	<i>Dkoe</i> \5'	123	63,4
<i>Dmoj</i>	93	<i>Dkoe</i> \3'	125	73,0
<i>Dmoj</i>	93	<i>Dkoe</i> \5'	123	87,1
<i>Dkoe</i> \3'	125	<i>Dkoe</i> \5'	123	75,6

En la Figura 3.19, también se observa el filograma entre estas secuencias. Un filograma es un diagrama de ramificación (árbol) que estima la filogenia, las longitudes de rama son proporcionales a la cantidad de cambio evolutivo, pero no está enraizado ni se muestra el ancestro. En el filograma de la Figura 3.19, observamos que las secuencias en 5' de los genes *CG4673* funcionales (*Dbuz*\5', *Dkoe*\5' y *Dmoj*) sufren muy pocos cambios (ramas cortas), lo que hace que queden agrupadas independientemente de la especie. La secuencia en 5' de Ψ *CG4673* (*Dbuz*\5'AC) es la secuencia que más cambios acumula y se agrupa con las secuencias localizadas en 3' del gen en *D. buzzatii* (*Dbuz*\3') y *D. koepferae* (*Dkoe*\3').

El número de sustituciones sinónimas por sitio sinónimo (Ks) entre los genes *CG4673* de *D. buzzatii* y *D. mojavensis* se estimó en la Tabla 3.13 (0,422). Si consideramos un tiempo promedio de divergencia de 12,5 millones de años (ma, Figura 1.3) podemos estimar la tasa de cambio nucleotídico neutro por sitio y por millón de años (r). El valor de r es igual a Ks dividido por 2 veces el tiempo de divergencia. Por tanto, la r estimada entre ambas especies será de $0,0168 \times 10^{-6}$. Hemos aplicado este valor de r para estimar distintos tiempos de divergencia, comparando las secuencias anotadas como repetición invertida en 5' del gen *CG4673* (Tabla 3.16). Se han ordenado las comparaciones según el orden en que se produjeron los eventos y deberían observarse tiempos decrecientes.

Los tiempos de divergencia entre *D. mojavensis* y las especies invertidas no se ajustan en absoluto a lo esperado en secuencias neutras. La comparación de las secuencias anotadas en la región 5' del gen *CG4673* en especies invertidas (*Dkoe*\5', *Dbuz*\5') con (*Dmoj*) debería ser la más divergente puesto que llevan más tiempo acumulando cambios. Por otro lado, las comparaciones de *D. mojavensis* con las secuencias que no se encuentran en la región 5' de un gen funcional (*Dbuz*\5'AC, *Dkoe*\3' y *Dbuz*\3') parecen haber acumulado más cambios. Los demás tiempos estimados, sí se ajustan a los tiempos de divergencia previamente descritos (Figura 1.3) y son correlativos a los eventos ocurridos. *Dbuz*\5'-*Dbuz*\5'AC ($12,14 \times 10^{-6}$) empezaron a divergir cuando se produjo la inversión *2m*. Las secuencias *Dbuz*\5'-*Dbuz*\3' y *Dkoe*\5'-*Dkoe*\3' ($10,10 \times 10^{-6}$ y $8,78 \times 10^{-6}$, respectivamente) divergen desde que ocurrió la microinversión de la duplicación. Finalmente, las comparaciones *Dbuz*\5'-*Dkoe*\5' y *Dbuz*\3'-*Dkoe*\3' ($2,91 \times 10^{-6}$ y $2,63 \times 10^{-6}$, respectivamente) describen el momento de divergencia entre las especies *D. buzzatii* y *D. koepferae*.

Tabla 3.16: Número de sustituciones nucleotídicas (K) y tiempos de divergencia (T) comparando dos a dos las secuencias anotadas como repetición invertida en la regiones 5' y 3' del gen *CG4673* de *D. koepferae* y *D. buzzatii* (*Dkoe*\5', *Dbuz*\5', *Dkoe*\3' y *Dbuz*\3'), y en 5' de Ψ *CG4673* (*Dbuz*\5' AC) y de *D. mojavensis* (*Dmoj*). K se ha calculado utilizando la fórmula de Jukes y Cantor. T se ha estimado utilizando la tasa de sustitución sinónima ($\mu = 0.0168 \times 10^{-6}$). No se han diferenciado sustituciones sinónimas de no sinónimas al no tratarse de CDS (sin marco de lectura).

Secuencia 1	Secuencia 2	Identidad (%)	K	T
Divergencia <i>D. mojavensis</i>/especies invertidas				
<i>Dbuz</i> \5'	<i>Dmoj</i>	83/93 (89,2)	0,1166	$3,47 \times 10^{-6}$
<i>Dkoe</i> \5'	<i>Dmoj</i>	81/93 (87,1)	0,1415	$4,21 \times 10^{-6}$
<i>Dmoj</i>	<i>Dbuz</i> \5' AC	58/92 (63,0)	0,5099	$15,17 \times 10^{-6}$
<i>Dmoj</i>	<i>Dbuz</i> \3'	67/93 (72,0)	0,3505	$10,43 \times 10^{-6}$
<i>Dmoj</i>	<i>Dkoe</i> \3'	68/93 (73,0)	0,3347	$9,96 \times 10^{-6}$
Tiempo de la inversión 2m				
<i>Dbuz</i> \5'	<i>Dbuz</i> \5' AC	63/92 (68,5)	0,4080	$12,14 \times 10^{-6}$
<i>Dbuz</i> \3'	<i>Dbuz</i> \5' AC	63/91 (69,2)	0,3965	$11,79 \times 10^{-6}$
Tiempo de la microinversión				
<i>Dbuz</i> \5'	<i>Dbuz</i> \3'	88/121 (72,7)	0,3394	$10,10 \times 10^{-6}$
<i>Dkoe</i> \5'	<i>Dkoe</i> \3'	93/123 (75,6)	0,2951	$8,78 \times 10^{-6}$
Divergencia <i>D. buzzatii</i>/<i>D. koepferae</i>				
<i>Dbuz</i> \5'	<i>Dkoe</i> \5'	99/109 (90,8)	0,0981	$2,91 \times 10^{-6}$
<i>Dbuz</i> \3'	<i>Dkoe</i> \3'	99/108 (91,6)	0,0890	$2,63 \times 10^{-6}$

3.4.6. Expresión del gen *CG4673* de *D. buzzatii*

La secuencia anotada como repetición invertida en la región 5' del gen *CG4673* de *D. buzzatii* no evoluciona de forma neutra y ha sido destacada como un posible promotor. Esta secuencia también se anotó en la región 3' del gen *CG4673* y en la región 5' Ψ *CG4673* del

punto de rotura AC. De este modo, dos promotores podrían estar flanqueando el gen *CG4673* en el punto de rotura BE, y otro se encontraría en la región 5' del gen en el punto de rotura AC pudiendo inducir hasta tres transcritos distintos: dos transcribiendo la misma región (gen en BE) pero en sentidos opuestos (uno en el sentido 5'-3' y otro antisentido 3'-5') y un tercero procedente del pseudogen en el punto de rotura AC. Sólo el transcrito 5'-3' del punto de rotura BE conserva la estructura con el gen *CG4673* y sería potencialmente funcional. Para conocer si existe actividad transcripcional desde los promotores putativos en la región 5' de *CG4673* y $\Psi CG4673$ en *D. buzzatii*, se diseñaron cebadores específicos para cada una con el fin de amplificar sus hipotéticos mRNA mediante RT-PCR.

3.4.6.1. Amplificación específica de las copias del gen *CG4673* en *D. buzzatii*

Hemos amplificado específicamente las regiones de similitud entre el pseudogen $\Psi CG4673$ y el gen *CG4673* (Tabla 3.17). Existen cuatro fragmentos con similitud entre ellos, uno en la 3'UTR, uno en el exón 7, uno que incluye parte de los exones 2 y 3 (y el intrón 2) y finalmente entre las secuencias anotadas como repetición invertida. Estos fragmentos constituyen las regiones menos degeneradas del pseudogen $\Psi CG4673$ de la región AC de *D. buzzatii*. Usando el valor de r estimado (véase sección 3.4.5.), hemos calculado el número de sustituciones nucleotídicas observadas (K) y el tiempo de divergencia entre los genes *CG4673* y $\Psi CG4673$ usando estos bloques de similitud entre los puntos de rotura AC y BE de *D. buzzatii* (Tabla 3.18). En esta Tabla se diferencia la comparación realizada en las regiones anotadas como codificantes y como no codificantes en *CG4673*, aunque no se han diferenciado sustituciones sinónimas de no sinónimas al no conocer el marco de lectura de $\Psi CG4673$. El tiempo de divergencia entre las regiones no codificantes de la duplicación se

ajusta a las estimas previas (véase Figura 1.3 en Introducción) y son parecidos a los tiempos estimados en la Tabla 3.16.

Tabla 3.17: Bloques de similitud entre los puntos de rotura AC y BE de *D. buzzatii*. Sólo las identidades con un tamaño mínimo de 45 pb y E-value $\leq 1e-04$ fueron tomados en cuenta. La identidad en la repetición invertida corresponde entre la anotada en 3' del gen *CG4673* en el punto de rotura BE con la anotada en 5' del gen *CG4673* en el punto de rotura AC.

Región	Coordenadas en el punto de rotura BE	Coordenadas en el punto de rotura AC	Identidad (%)	E-value
Repetición invertida	313-208	12.733-12.841	80/111 (72,1)	6e-08
Exones 2-3 (intrón2)	3.637-4.034	12.284-12.633	274/399 (68,7)	1e-41
Exón 7	1.714-1.838	395-518	98/125 (78,4)	4e-23
3' UTR	1.190-1.242	218-270	47/53 (88,8)	6e-14

Tabla 3.18: Número de sustituciones nucleotídicas (K) y tiempo de divergencia (T) entre los genes *CG4673* y Ψ *CG4673* de *D. buzzatii*. K se ha calculado utilizando la fórmula de Jukes y Cantor. T se ha estimado utilizando la tasa de sustitución sinónima ($r = 0.0168 \times 10^{-6}$). No se han diferenciado substituciones sinónimas de no sinónimas al no conocer el marco de lectura de Ψ *CG4673*.

<i>CG4673 D. buzzatii</i> (BE)- Ψ <i>CG4673 D. buzzatii</i> (AC)	K	T
Total	0,268	$7,97 \times 10^6$
Secuencia codificante (exones 3, 2, 7, 0)	0,244	$7,26 \times 10^6$
Secuencia no codificante	0,319	$9,49 \times 10^6$

Los cebadores fueron diseñados en estas regiones de mayor similitud, pero de forma que su posición 3' quedara en una de las zonas de menor identidad y que su amplificación fuera específica para el gen *CG4673* o el pseudogen Ψ *CG4673* (Figura 3.20). El nombre de la PCR corresponde a la región en la que fueron diseñados (PCR 2, 3, 7, 3'UTR). Todos estos cebadores amplifican específicamente en la copia del gen de AC para detectar si existe transcrito en alguna de estas regiones. Otros dos cebadores, 6L y 5R, fueron diseñados en el exón 6 y exón 5 del gen del punto de rotura BE. Esta PCR incluye el intrón 5 y amplifica específicamente el gen funcional (*CG4673*) (los exones 5 y 6 no están en AC). Adicionalmente, otros dos cebadores, 0R (en el exón 0) y 0L (en el exón 2) fueron diseñados para amplificar el inicio putativo del transcrito RB (PCR 0). Esta PCR también amplifica específicamente en la copia BE. Las PCR diseñadas y la región que amplifican se muestran en la Figura 3.18. Las secuencias de los cebadores se muestran en la Tabla 2.4.

3.4.6.2. RT-PCR del mRNA del gen *CG4673* en *D. buzzatii*

Para realizar la RT-PCR del gen *CG4673*, se obtuvo mRNA total de distintas fases del desarrollo de *D. buzzatii*: embrión (E), larvas (L), pupa (P) y adultos (A). Además, los embriones fueron recolectados de diferentes tiempos (1,30h; 2,15h; 2,30 h y 22h). Para el mRNA de los adultos, se distinguieron machos y hembras vírgenes. Por otro lado, también se obtuvo mRNA total de una mezcla de estadios de *D. buzzatii*.

Como control de la retrotranscripción, se amplificó el gen *Gadph* de *D. buzzatii* con los cebadores H1 y H2 (tomados de Puig *et al.* 2004, Tabla 2.4). La RT+ debe amplificar 438 pb, pero no el control negativo (RT-), al que no se le añade enzima retrotranscriptasa (no se amplifica cDNA). Si en el control negativo (RT-) hay producto, se debe a DNA genómico residual. También se ha añadido un control positivo de los cebadores, amplificando también el DNA genómico como molde para comprobar la eficacia de éstos. Hemos realizado dos tipos de RT-PCR. Por un lado, hemos amplificado las PCR diferenciales (2, 3, 56, 7 y 3'UTR) para cada uno de los mRNA obtenido a lo largo de las distintas etapas. Finalmente, se realizó la PCR 0 (amplificación desde el exón 2 al exón 0 del gen *CG4673*) para comprobar si el transcrito RB se produce en *D. buzzatii*.

Las PCR diferenciales (2, 3, 56, 7 y 3'UTR) amplificaron en todas las etapas. Las amplificaciones de estas PCR y el control de *Gapdh* para todas las alícuotas de los mRNA de los distintos estadios se muestran en la Figura 3.21.

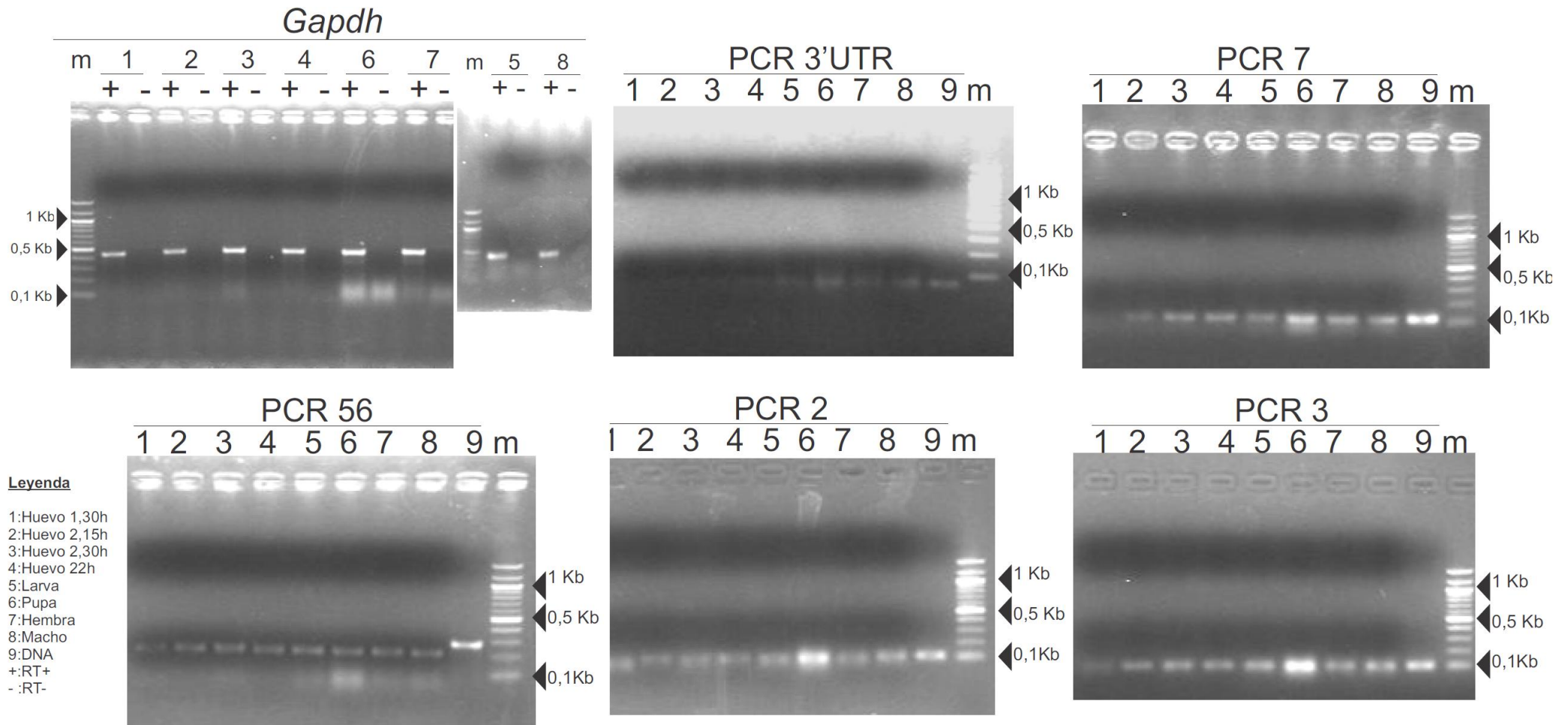


Figura 3.21: Experimentos de RT-PCR. Para cada extracción de mRNA (véase leyenda), se han realizado un control positivo con retrotranscriptasa (+) y uno negativo sin retrotranscriptasa (-) amplificado el gen *Gapdh*. El gen ha sido correctamente amplificado en todos los controles positivos (+) pero no en los negativos (-), evidenciando que no hay DNA residual (véase Materiales y Métodos). Para cada extracción y DNA genómico (control positivo de los cebadores) se han realizado las PCR 3'UTR, 2, 3, 7 y 56. Todas las PCR han amplificado en los distintos mRNA y DNA. El producto de amplificación de la PCR 56 en DNA genómico es ligeramente más grande que en las extracciones de mRNA al contener el intrón 5.

Todos los productos de las PCR fueron recuperados y secuenciados para verificar la copia que se estaba amplificando. Los tamaños de los productos de PCR así como la copia que amplifican (verificado por secuenciación) se indican en la Tabla 3.19. La PCR 56 amplificó un fragmento de *CG4673*, mientras que las demás PCR amplificaron el pseudogen. En todos los distintos tiempos seleccionados del desarrollo se transcriben las dos copias de *D. buzzatii* (Figura 3.21). Estas PCR son cualitativas, la intensidad de la banda en el gel no proporciona información sobre los niveles o diferencias de expresión de los transcritos entre los distintos estadios. Los tamaños de los productos de las RT-PCR corresponden al tamaño esperado y obtenido en el control positivo con DNA genómico. La PCR 56 amplificada en DNA genómico es ligeramente más grande (276 pb) respecto a las que amplifican cDNA (215 pb) al no contener el intrón 5 del gen *CG4673* del punto de rotura BE.

Estos resultados indican que habría al menos dos transcritos relacionados con el gen *CG4673* ya que ambas copias fueron amplificadas a partir del mRNA total (una procedente de la región AC y otra de BE).

Por otro lado, la PCR 0 se amplificó usando la mezcla de mRNA de *D. buzzatii* para verificar si existe el transcrito RB conceptual. El transcrito RB difiere del transcrito RA sólo en su primer exón. De este modo, para diseñar una PCR específica para este transcrito, hemos diseñado un cebador en el exón 0 (cebador 0R) y otro en el exón 2 (cebador 0L) (véase Figura 3.18). Las secuencias de dichos cebadores se muestran en la Tabla 2.4 en Materiales y Métodos). La RT-PCR 0 amplifica el exón 0 y el 2 (no amplifica en exón 1 ni el intrón 0) por lo que, al menos en *D. buzzatii* existen los transcritos RA y RB del gen *CG4673*.

Finalmente, se realizaron cuatro RT-PCR combinando los cebadores anteriores (Tabla 3.19). Al no haber observado diferencias en los distintos tiempos ni estadios en las RT-PCR, tan sólo se utilizaron un par de etapas para estas combinaciones. Estas PCR fueron realizadas usando DNA genómico (control positivo) y cDNA de hembras vírgenes y los embriones de 22

Tabla 3.19: Tamaños de los productos de RT-PCR y cebadores utilizados. La posición y orientación de los cebadores se muestran en la Figura 3.20. El nombre de cada cebador se corresponde al exón en el que fueron diseñados. Las RT-PCR se realizaron a partir de mRNA total de los estadios correspondientes (en la Tabla se muestra sólo un resultado al amplificar todos los estadios el mismo tamaño). Los pb esperados en la amplificación de cDNA, sólo se muestran para *CG4673* en BE al no conocer a priori los posibles lugares de procesamiento en AC. Las procedencias de las amplificaciones corresponden a la verificación por secuenciación. Se muestran los tamaños esperados hipotéticos de los productos amplificados en DNA y cDNA para cada PCR en cada región (si corresponde). X: PCR no amplificables al no contener la región del cebador.

RT-PCR							
PCR	Cebador1	Cebador2	pb esperado en DNA AC	pb esperado en DNA BE	pb amplificado en DNA (punto de rotura)	pb esperado en cDNA BE	pb amplificado en DNA (punto de rotura)
0	0R	0L	X	1.080	~1.000	235	~400 (BE)
2	2R	2L	106	109	~106 (AC)	109	~110 (AC)
3	3R	3L	100	103	~100 (AC)	103	~110 (AC)
7	7R	7L	116	117	~116 (AC)	117	~120 (AC)
3'UTR	3'R	3'L	68	68	~68 (AC)	68	~80 (AC)
56	5L	6R	X	276	~280 (BE)	215	~220 (BE)
<i>Gapdh</i>	H1	H2	X	X	~438	X	~450
7-3'	7R	3'L	304	571	~400 (AC)	571	~400 (AC)
2-3'	2R	3'L	12.421	2.852	No amplifica	1.870	No amplifica
2-3	2R	3L	277	327	~300 (AC)	265	~300 (AC)
5-7	2R	6L	X	1.577	~1.500 (BE)	1.158	~1.100 (BE)

horas como molde. Tres de ellas, las PCR 7-3', 2-3' y 2-3 se amplificaron a temperatura restrictiva para que fueran específicas de *ΨCG4673*. La PCR 2-3' no amplificó ni con genómico (tamaño esperado de 12.426 pb) ni con cDNA. La PCR 2-3 amplificó la región del intrón 2 usando el cDNA como molde por lo que al menos, esta señal de procesamiento estaría degenerada en el transcrito de *ΨCG4673*. Una cuarta RT-PCR (PCR 5-7) se amplificó a temperatura flexible (50°C) para permitir la unión del primer 7L a la secuencia del gen *CG4673* (la región del cebador 5R no se encuentra en la copia *ΨCG4673*). El producto amplificó con el molde de DNA genómico (~1.600 pb) y con el cDNA (~1.000 pb) correspondientes a la copia del gen *CG4673* en la región BE. El producto obtenido usando como molde el cDNA amplifica sin los intrones 5 y 6.

3.4.6.3. Detección de dsRNA del gen *CG4673* en *D. buzzatii*

El experimento de RT-PCR permitió amplificar dos transcritos de distinta procedencia en *D. buzzatii*. Por un lado, un transcrito proviene de la copia del gen *CG4673* de la región del punto de rotura BE y el otro, lo haría procedente del pseudogen *ΨCG4673* en la región del punto de rotura AC. Incluso pueden ser más, ya que la región promotora putativa de la copia del gen en el punto de rotura BE también se encuentra duplicada. Por tanto, un tercer transcrito independiente podría ser producido si *CG4673* en BE se transcribe en ambas direcciones. Estos transcritos no necesariamente han de ser traducidos para desempeñar una función o tener actividad bioquímica. De hecho, podrían estar relacionados entre ellos modulando su propia actividad o expresión. Para detectar si existe interacción entre ellos, se diseñó un experimento de detección de RNA de doble cadena (dsRNA). Un dsRNA se compone de dos fragmentos de RNA de cadena simple (ssRNA) y complementarios que se unen para crear una molécula de doble cadena. Se han amplificado las PCR 2, 3 y 7 diseñadas en *ΨCG4673* de *D. buzzatii*.

Además, se ha amplificado el gen *Gapdh* como control de la extracción del mRNA. Cada PCR se ha realizado con distintas condiciones (véase Materiales y Métodos). Sólo las alícuotas S se han realizado con RNAsa, por lo que *Gapdh* (mRNA de cadena simple) sólo debe amplificar en las alícuotas C. En la Tabla 3.20 se indican las PCR realizadas y las especificaciones para cada una de ellas.

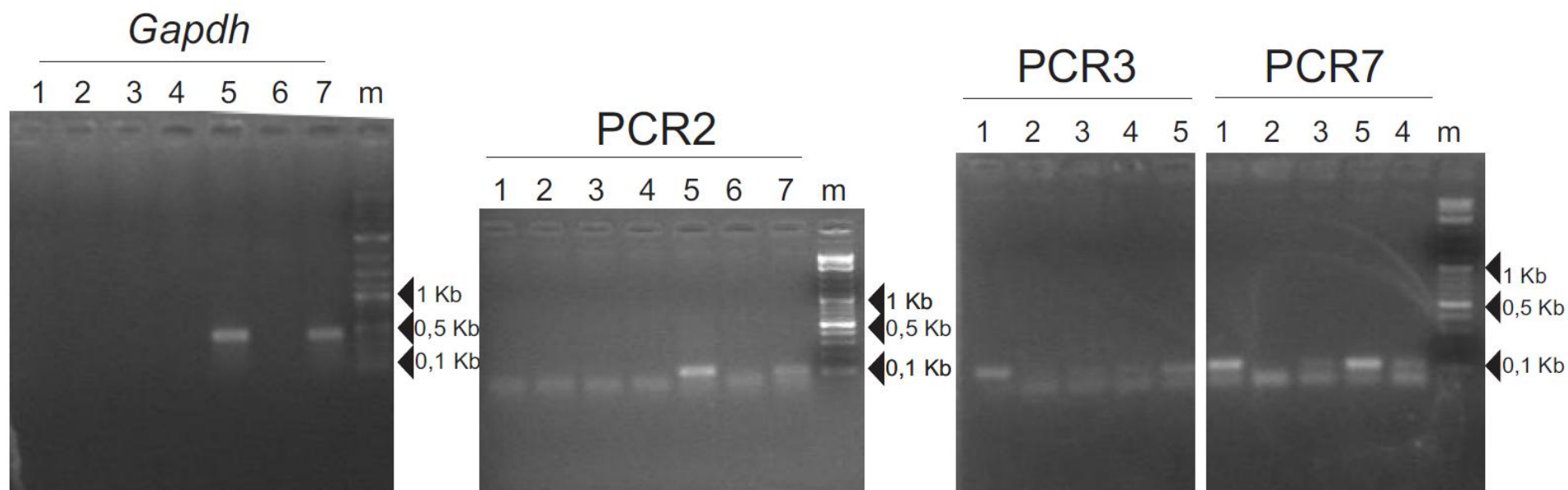
Los geles de la detección de dsRNA se muestran en la Figura 3.22. El gen *Gapdh* no amplifica en ninguna de las alícuotas digeridas con RNAsa (S). Tampoco ha amplificado en C- (no hay retrotranscriptasa) evidenciando que no hay DNA residual (véase Materiales y Métodos). En cambio, se obtuvo producto para las alícuotas C + y CT independientemente de los cebadores usados (cebadores OligoT o al azar), tal y como se espera en una RT-PCR de un mensajero de cadena sencilla (véase sección 3.4.6.2). La PCR 2, se comporta como el gen *Gapdh*, amplificando en las mismas alícuotas. Al menos en la porción de la PCR 2, no hay RNA de doble cadena relacionado con *ΨCG4673* ya que no está protegida de la digestión con RNAsa. La intensidad de la banda de la alícuota C+ es mucho mayor que la de CT, por lo que parece que en este caso, amplificar con cebadores al azar es mejor (se obtiene mayor cantidad de cDNA) que usando OligoT. Este efecto es debido a la localización de la PCR 2 (cercana a la región 5' del mensajero). Los cebadores OligoT, amplifican desde la cola poli-A. Si los productos de amplificación de cDNA no amplifican la práctica totalidad del mRNA, la PCR posterior tendrá menos molde para amplificar.

Las PCR 3 y 7 no se amplificaron en las alícuotas C (menos en la del control C+), puesto que el patrón de alícuotas amplificadas será el mismo que las PCR 2 y *Gapdh*. Claramente, se obtuvo producto en la alícuota S+ y C+. Además, se amplificaron débilmente las alícuotas SN y ST. La amplificación con desnaturalización previa facilita la obtención de producto aunque la temperatura a la que se realiza la RT-PCR es suficiente para mantener las dos hebras de RNA separadas durante el proceso y obtener un mínimo producto.

Tabla 3.20: Detección de dsRNA. Lado izquierdo: especificaciones de las distintas alícuotas. RP: cebadores al azar (*random primers*); OT: cebadores Oligo-T. Lado derecho: resultado de la PCR realizada con cada alícuota, señalando cuando se obtuvo producto (√) y cuando no (x). Los controles C- y CT sólo se han realizado para la PCR 2 y no para las PCR 3 y 7 (-).

Especificaciones de las Alícuotas					PCR				
Nombre	Enzima		Desnaturalización		Cebador en la RT	<i>Gapdh</i>	2	3	7
	RNAsa	RT	pre RT						
S+	√	√	√		RP	X	X	√	√
S-	√	X	√		RP	X	X	X	X
SN	√	√	X		RP	X	X	X	X
ST	√	√	√		OT	X	X	X	X
C+	X	√	√		RP	√	√	√	√
C-	X	X	√		RP	X	X	-	-
CT	X	√	√		OT	√	√	-	-

La amplificación de las alícuotas S+ indica que en ciertos lugares (al menos en las regiones de las PCR 3 y 7), existe mRNA de doble cadena que se protege de la digestión con RNAsa. Por tanto, al menos habría dos transcritos relacionados. Uno de ellos proviene de la copia *ΨCG4673* ya que las PCR que amplifican mRNA no degradado amplifican específicamente el pseudogen (las PCR 3 y 7, son específicas de esta copia). El otro transcrito detectado con este experimento no puede ser el transcrito de la copia *CG4673* puesto que no es complementario. Ambas copias están orientadas en la misma dirección y transcriben en la cadena complementaria. Si ambos transcritos son 5'-3', no pueden formar moléculas de RNA de doble cadena. Uno de los dos transcritos detectados en el experimento de dsRNA no se detectó en el experimento de RT-PCR. Con las RT-PCR, se detectó actividad transcripcional de ambas copias del gen *CG4673* de *D. buzzatii*.



Leyenda: 1:S+ 3:SN 5:C+ 7:CT
2:S- 4:ST 6:C- m: marcador

Figura 3.22: Experimento para la detección de dsRNA. Cada PCR se ha realizado en extracciones de mRNA total de *D. buzzatii*. Para cada extracción se ha amplificado el gen *Gapdh* y las PCR 2, 3, 7. Las alícuotas S tienen RNAsa pero las C no han sido previamente digeridas. Se ha realizado un control de DNA genómico residual, (+, con retrotranscriptasa y - sin retrotranscriptasa). Las alícuotas T usan cebadores OligoT en vez de cebadores al azar. *Gapdh* y la PCR 2, sólo amplifican en los carriles C+ y CT. Las PCR 3 y 7 amplifican las alícuotas S+, C+ y, con menor producto, las ST y SN.

Hace falta al menos un tercer transcrito en dirección 3'-5' que complemente con el transcrito codificado en la copia $\Psi CG4673$ de AC. Este tercer transcrito podría provenir de la transcripción antisentido de la copia del gen $CG4673$ de la región BE. Alguno de los promotores putativos (*a*, *b*, *c*, véase sección 3.4.4) encontrados en el sentido directo en el punto de rotura BE podría estar transcribiendo algún fragmento 3'-5' (antisentido de la codificación del gen). Para detectar si realmente podría haber 3 transcritos de las distintas copias del gen $CG4673$ de *D. buzzatii*, se realizó una hibridación de Northern.

3.4.6.4. Detección de los transcritos del gen $CG4673$ en *D. buzzatii* mediante hibridación de Northern

La hibridación de Northern debería permitirnos detectar cuántos transcritos relacionados con los distintos promotores putativos encontrados en $CG4673$ y $\Psi CG4673$ existen en *D. buzzatii*.

Se han realizado tres rondas de hibridación. Véase Materiales y Métodos para las especificaciones de cada membrana en cada ronda de hibridación. Sólo se obtuvo hibridación en la tercera ronda. En los carriles de los geles para transferir a las membranas de hibridación se cargaron todos los mRNA de los distintos estadios y se añadieron dos controles (mezcla de todos los estadios de mRNA de *D. mojavensis* y DNA genómico de *D. buzzatii*). En el Dot Blot se hibridaron tres muestras, mRNA de adulto hembra de *D. buzzatii* (b1 en la membrana), mRNA *D. mojavensis* (d1) y DNA genómico de *D. buzzatii* (f1) En las correspondientes casillas de la tercera fila se colocaron tres diluciones de las anteriores muestras (b3, d3 y f3 respectivamente). Las membranas 1 y 3 se hibridaron con la sonda 5-7 (véase Tabla 3.19). El Dot Blot y la membrana 2 se hibridaron con la sonda 2-3. La sonda 2-3 ha quedado sin

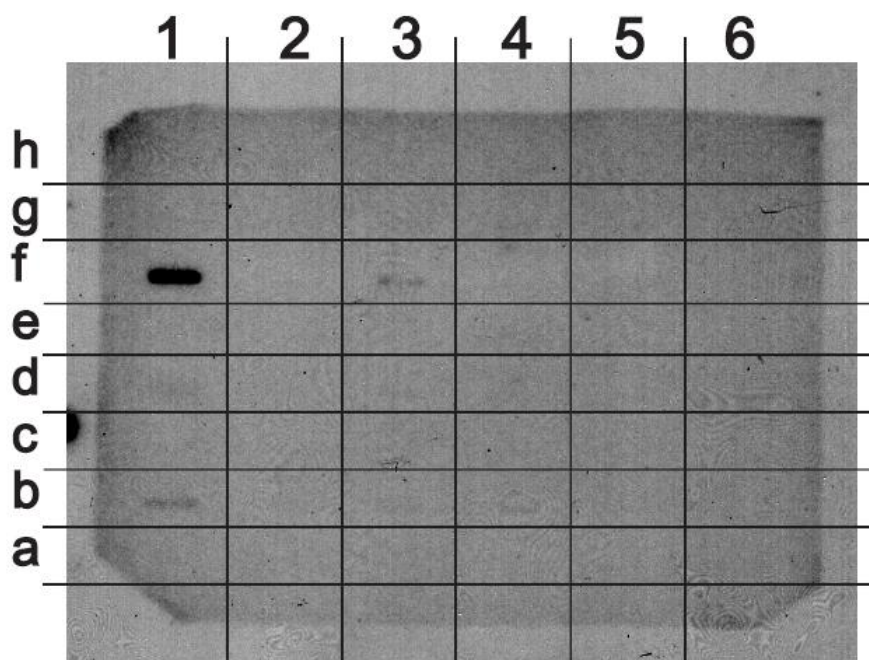
hibridar en la membrana 2. No obstante, lo hizo en el Dot Blot, por lo que la falta de hibridación no se debe a problemas de tamaño de la sonda o inespecificidad de ésta.

La hibridación en el Dot Blot no es completa (Figura 3.23). Las señales en las diluciones (carril 3) apenas se aprecian. La mayor señal se obtuvo en el DNA genómico de *D. buzzatii* (f1). La señal en *D. mojavensis* (d1) es prácticamente inexistente, probablemente debido a la excesiva divergencia de la sonda. En b1 (mRNA de adulto hembra de *D. buzzatii*) se obtuvo una señal muy débil pero corrobora que la sonda funcionaba en mRNA, es decir que no habría motivo para pensar que dicha sonda no debiera funcionar en la hibridación de Northern (membrana 2).

Las membranas 1 y 3 se hibridaron con la sonda PCR 5-7 (Figura 3.23). La principal diferencia entre ellas es la cantidad de mRNA cargado en el gel (véase Tabla 2.8 en Materiales y Métodos), siendo mucho menor la cantidad de muestra en el gel para la membrana 3. La diferencia se hace evidente en el revelado de la hibridación, dado que tanto la intensidad como el número de hibridaciones es menor en la membrana 3.

En la membrana 1, se obtuvo señal en los dos controles. La señal en el carril con DNA de *D. buzzatii* es débil debido a que el DNA no penetra bien en las condiciones desnaturizantes de los gels para RNA, por lo que todo el material se concentra al principio del pocillo. En el otro control cargado con RNA de *D. mojavensis* se obtiene una señal única y clara correspondiente al único transcrito del gen *CG4673* en esta especie (~1.100 pb). Las demás señales en la membrana 1 (hibridaciones con los mRNA de los distintos estadios de *D. buzzatii*) sufren una gradación de intensidad que atribuimos no a la cantidad de expresión, sino a la concentración de la muestra (véase Tabla 2.8 en Materiales y Métodos) al tratarse de una técnica cualitativa. En todos los carriles con mRNA de *D. buzzatii* se observa una señal de hibridación principal (~1.100 pb) parecida a la señal obtenida en la hibridación en el transcriptoma de *D. mojavensis*.

A: Dot Blot



	1	2	3	4	5	6
h	-	-	-	-	-	-
g	-	-	-	-	-	-
f	25 µL DNA <i>D. buzzatii</i> (720 ng/µL)/500 µL	-	50 µL c1/500 µL	-	-	-
e	-	-	-	-	-	-
d	25 µL mRNA <i>D. mojavensis</i> (720 ng/µL)/500 µL	-	50 µL c1/500 µL	-	-	-
c		-		-	-	-
b	15 µL mRNA hembra (1.400 ng/µL)/500 µL		50 µL b1/500 µL	-	-	-
a	-	-	-	-	-	-

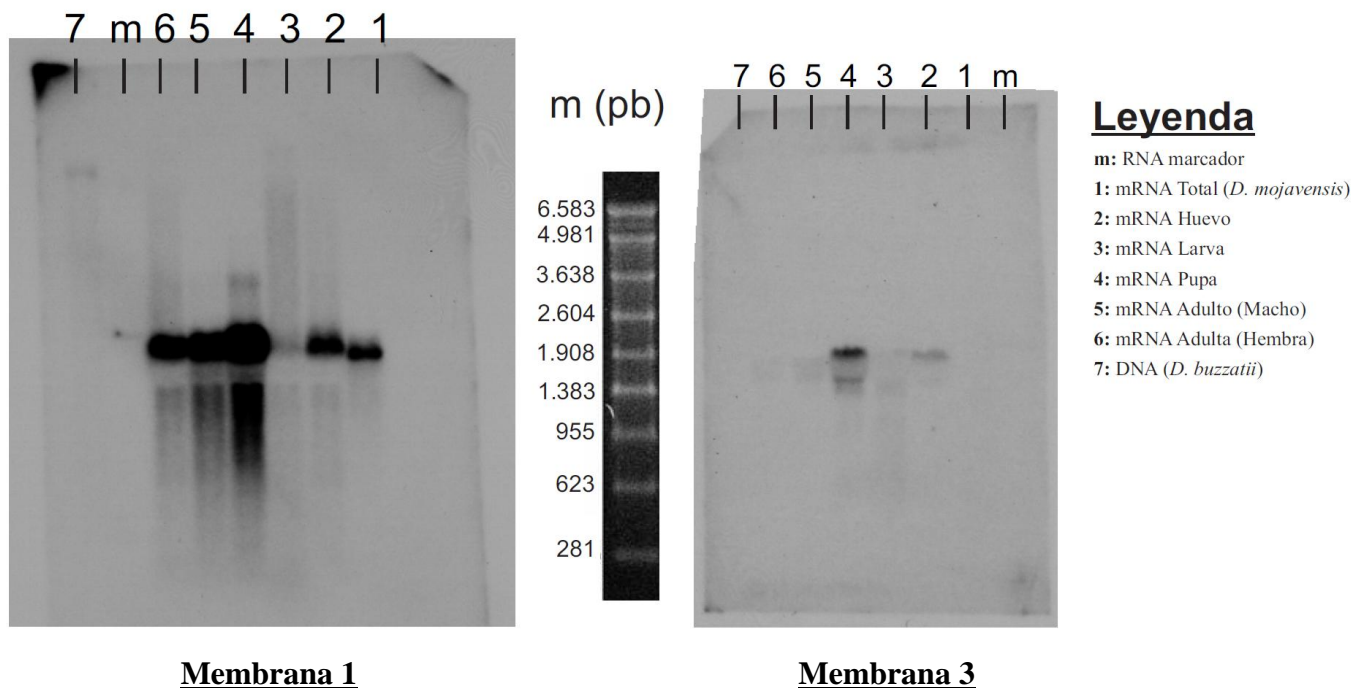
B. Hibridación de Northern

Figura 3.23: A) Dot Blot. El contenido y las concentraciones del material para cada cuadrícula de la membrana se especifican en la Tabla inferior de la hibridación del Dot Blot. B) Hibridación de Northern. Se muestran las membranas 1 y 3. Las equivalencias de peso molecular de cada banda hibridada se muestran en el marcador de peso molecular entre las membranas. El mRNA para cada carril se muestra en la leyenda.

Además, en los carriles mejor hibridados (huevo, pupa y ambos adultos) aparece una segunda banda de hibridación de menor tamaño (~950 pb), aunque su interpretación es dudosa al camuflarse entre una mancha de degradación de mRNA. Finalmente, en el mRNA de pupa (donde la intensidad de la hibridación es mayor) también se observa una señal de hibridación de mayor tamaño que la principal (~2.000 pb).

Si comparamos esta membrana con la membrana 3 (misma sonda hibridada) vemos que la intensidad es menor. En los carriles de los dos controles, que ya en la membrana 1 eran los hibridados con menos intensidad, no se observa señal. En los carriles con RNA de pupa y

huevo (4 y 2, respectivamente) se observa la señal de hibridación principal (~1.100 pb). En el carril 4, se observa claramente la segunda señal de hibridación de menor tamaño que estaba oculta en la membrana 1. Por el contrario, la tercera señal (la más débil y de mayor tamaño en el carril 4 de la membrana 1) ha desaparecido.

Si compilamos el resultado de ambas membranas, observamos, al menos en pupa (carril 4) tres señales distintas relacionadas con la sonda 5-7 del gen *CG4673* de *D. buzzatii*. La señal principal también se observa en el mRNA de *D. mojavensis* y podría corresponderse con el mRNA de *CG4673*. Adicionalmente, una señal de mayor tamaño se observa en la membrana 1 y una de menor tamaño en la membrana 3.

3.4.6.5. RACE-PCR del *CG4673* en *D. buzzatii*

El experimento de RACE permite localizar el origen y el final de la transcripción de un gen. Las 5'RACE y 3'RACE amplifican específicamente las 5'UTR y 3'UTR hasta el origen y el final de un transcrito, respectivamente. Como en la RT-PCR, se obtiene cDNA de una muestra de mRNA. En este caso, para obtener cDNA no se usan cebadores al azar (*random primers*) como para las RT-PCR, sino adaptadores específicos que se unen al principio y al final de los transcritos. Estos adaptadores permiten retrotranscribir los extremos (5' y 3') de todos los mRNA a cDNA. Dos PCR usando este cDNA como molde permiten obtener producto de las regiones distales de los transcritos de interés. La primera PCR (*Outer*) usa un cebador que se une al adaptador retrotranscrito en el cDNA con un cebador específico del transcrito de interés (primer y último exón, para la 5'UTR y la 3'UTR, respectivamente). La segunda PCR (*Inner*) amplifica la primera PCR (*Outer*) como molde usando cebadores anidados. Las secuencias de los cebadores para la RACE-PCR se muestran en la 2.4. Para amplificar la *Outer* PCR de las UTR 5' y 3' del gen *CG4673* se usaron los cebadores 1L

(primer exón), 5' *Outer* (adaptador en 5') y 8R (ultimo exón), 3' *Outer* (adaptador en 3') respectivamente. Las 3' y 5' UTR amplificaron correctamente en *D. buzzatii*, aunque sólo la 3'RACE pudo ser completamente secuenciada.

3' UTR: El fragmento secuenciado (PCR 8Rint-3' *Inner*) fue secuenciado y alineado con las regiones 3' del gen *CG4673* de distintas especies secuenciadas (*D. grimshawii*, *D. virilis* y *D. mojavensis*), y de *D. buzzatii* (BE). La región de la señal de poliadenilación (poli-A) se localiza 196 pb aguas abajo del codón stop (Figura 3.24) y se conserva en todas las especies alineadas. La 3'RACE amplifica hasta la señal de poli-A (donde se localiza el Adaptador 3' de la 3'RACE), aunque el final de la transcripción suele prolongarse hasta 30 pb después de esta señal. La conservación entre las especies alineadas se mantiene 24 pb adicionales que podrían corresponderse al final putativo de la transcripción.

```

Dgri      -----TCAGCTTAGCGCGAAGAAAAATCTGGTGTTAAAACGAAACTAAAGATT 49
Dvir      -----TCAGCTTAGCGCGAAGAAAAATCTGGTGTTAAAACGAAACTAAAGATT 49
Dmoj      -----TCAGCTTAGCGCGAAGAAAAATCTGGTGTTAAAACGAAACTAAAGATT 49
Dbuz      AGCAGCGTACTTCAGCTTAGCGCGAAGAAAAATCTGGTGTTAAAACGAAACTAAAGATT 60
Dbuz\3' RACE -----GAAAAATCTGGTGTTAAAACGAAACTAAAGATT 34
                        *****

Dgri      CAATCATACTACT----- 63
Dvir      CAATCATACTACT----- 63
Dmoj      CAATCATACTACTCTACACATATATATAGCAGAGGGGCACATGGATAAAGTTAAAGCCA 109
Dbuz      CAATCATACTACTCTACACATATATATAGCAGAGGGGCACATGGATAAAGTTAAAGCCA 120
Dbuz\3' RACE CAATCATACTACTCTACACATATATATAGCAGAGGGGCACATGGATAAAGTTAAAGCCA 94
                        *****

Dgri      -----TTCTCTATAAATGTAATCTTGTATAAGTTTATAACAA 101
Dvir      -----CGCATACCGTACGCCTTCTCTATAAATGTAATCTTGTATAAGTTTATAACAA 116
Dmoj      ATC--ATCGCATACCGTACGCCTTCTCTATAAATGTAATCTTGTATAAGTTTATAACAA 167
Dbuz      ATCCAATCGCATACCGTACGCCTTCTCTATAAATGTAATCTTGTATAAGTTTATAACAA 180
Dbuz\3' RACE ATCCAATCGCATACCGTACGCCTTCTCTATAAATGTAATCTTGTATAAGTTTATAACAA 154
                        *****

Dgri      AATTTAATTTATTTCTAAATAAATTTAAATGTTGAGT 139
Dvir      AATTTAATTTATTTCTAAATAAATTTAAATGTTGAGT 154
Dmoj      AATTTAATTTATTTCTAAATAAATTTAAATGTTGAGT 218
Dbuz      AATTTAATTTATTTCTAAATAAATTTAAATGTTGAGT 235
Dbuz\3' RACE AATTTAATTTATTT
                        *****

```

Figura 3.24: Alineamiento de la región 3' conservada a partir del codón stop del gen *CG4673* de las especies *D. grimshawii* (*Dgri*), *D. virilis* (*Dvir*), *D. mojavensis* (*Dmoj*), *D. buzzatii* (*Dbuz*) y la 3'RACE de *D. buzzatii* (*Dbuz\3'RACE*). Se muestra la región de poliadenilación (gris) conservada en todas las especies alineadas.

5'UTR: El fragmento amplificado (PCR 1Lint-5'Inner) no pudo ser completamente secuenciado. Después de múltiples intentos, los cromatogramas de secuenciación resultan de calidad insuficiente y sólo se han podido secuenciar ~200 pb desde el codón inicio del transcrito RB. Se ha intentado localizar el inicio putativo de la transcripción mediante la conservación nucleotídica entre distintas especies. Se ha completado el alineamiento con las regiones 5' de las especies *D. koepferae* y *D. martensis* (ambas invertidas para *2m* y con la duplicación del gen *CG4673*). La región 5' del gen *CG4673* en *D. koepferae* ya fue secuenciada para la comparación de las repeticiones invertidas (véase sección 3.4.4). La región 5' del gen *CG4673* de *D. martensis* se ha amplificado con un cebador en el primer exón orientado hacia la región aguas arriba (firstL) y otro en la propia repetición invertida anotada en 5' del gen *CG4673* funcional (IR5'R) (véase Figura 3.18). Las secuencias de los cebadores se muestran en la Tabla 2.4 en Materiales y Métodos. El alineamiento se completa con las regiones 5' del gen *CG4673* de *D. virilis* y *D. mojavensis* (Figura 3.25). La secuencia de *D. virilis* no alinea correctamente (especie más alejada filogenéticamente), por lo que sólo se han usado para el alineamiento las regiones con identidad localizadas mediante BLAST con la secuencia de *D. buzzatii*.

La región alineada en la Figura 3.25 comprende desde el exón 1 hasta la región anotada como repetición invertida en 5' del gen *CG4673* (incluyendo el codón inicio y el exón 0 del transcrito RB). Este fragmento de *D. buzzatii* se ha usado para localizar inicios de transcripción (TSS) putativos con el programa Neural Network Promoter Prediction (Reese 2001) implementado en el *Berkeley Drosophila Genome Project*. Se han considerado todos aquellos inicios de transcripción propuestos para este fragmento con una puntuación superior a 0,95.

En la Figura 3.25 se muestran anotados en gris (TSS1 y TSS4) con las cajas TATA putativas a 30 pb aguas arriba. Ni estos TSS ni sus cajas TATA se conservan en todas las

especies comparadas. Paralelamente, hemos localizado otros dos inicios de transcripción putativos (TSS2 y TSS3) con las cajas TATA también conservadas entre las especies comparadas. En el TSS3 también ha sido anotado el supuesto elemento BR (sitio de unión de algunos factores de transcripción y que se localiza justo aguas arriba de la caja TATA). Hay otra región altamente conservada en las 5 especies comparadas (regiones en color rosa en la Figura 3.25). Esta secuencia podría tratarse de un promotor proximal de los TSS 1 o 2, un promotor aguas abajo (DPE) de los TSS 3 o 4 o un potenciador (*enhancer*).

Se han combinado los cebadores de la RACE con los cebadores diseñados en el inicio y final del transcrito *CG4673* para intentar amplificar los extremos (inicio y final) de los otros dos transcritos detectados en el Northern. Los extremos del transcrito antisentido del gen *CG4673* amplificarían con las combinaciones PCR 1L-5'Inner y PCR 8R-3'Inner. Sin embargo no se obtuvo producto de amplificación para estas PCR. Por otro lado, al haber amplificado transcrito procedente de la copia AC en el anterior experimento de RT-PCR se intentó combinar los cebadores de la RACE con los cebadores específicos diseñados en la copia de AC. En este caso, los cebadores *Outer* fueron reutilizados de las RT-PCR (2L, para la 5'UTR y 7R, para la 3'UTR) y los anidados fueron nuevamente diseñados (2Lint y 7Rint). Como para la amplificación de los extremos del gen *CG4673* en BE, se ensayaron ambos sentidos ya que, aunque se detectó transcripción, no se conoce el sentido de ésta. Tampoco en este caso se obtuvo producto para estas RACE PCR.

```

Dmar -----
Dkoe -----
Dbuz CGATTGTTTACATAACGTGAAGAGACGATGTCGCACAAAGGCATAAACGAAATTCCGTTT 60
Dmoj CGATTGTTTACATAACGTAAAGAGACGATGCCGCACAAAGGCATAAACGAAATTTCGTTT 60
Dvir CGATTGTTTACATAACGTAAAGAGACGATGGCGCACAAAGGCATAAACAAAAATTGCGCTT 60

Dmar TATCGGCATGTACTGTACGAGACAGTGAGG-GCNNTGGA CAT GATGTTGATGATG---TT 56
Dkoe -----
Dbuz CTTGTAGATGAACTGCTCGAGCAGCAAAGGGGCGCTGGC CAT GATGTTGATGATGTTCTT 120
Dmoj CTTGTAGATGAATTGCTCGAGCAGCAAAGGAGCGCTGGC CAT GGTGTTCTTGCTG---TT 117
Dvir CCTATAGATGAATTGCTCGAGCAGCAGCGGTGCACTGGC CAT GTTGTCTTCTTGGTGT 120

START RA

Dmar TNNNNGCC-----TGCTGCTGTTGGTGTGTTAAACG-AGCTTGAGNNNNNTGGA-A 107
Dkoe -----
Dbuz GTATTGGCCT---GCTGCTGTTGGTGTGTTAAACGTAGTTTGGCTACTTTGG-A 175
Dmoj GTCTTGGCCTGCT-GTTGCTGTGGTGTGTTATTGTTAGCGTAGCTTGAGCTACTTTGGA-T 175
Dvir GTTTTGGCCAAATGGCTGCGCTGTTGTTGTTGTTAAAGCGTAGCTTACGCTACTTTGGTT 180

Dmar -AGCCTAAAAGGCCTTAACTTAGTGTATTGAACTGTACGACATGAAACCGCGCATTTTAA 166
Dkoe ---CCTAAGAGGCTT---AGTTAG-GTATTGAAGTGTACGACATGAAACCGCGCATTTTAA 53
Dbuz TGGCCTAAAAGGCCTTATACTTAGTGTATTGAACTGTACGACATGAAACCGCGCATTTTAA 235
Dmoj TAGCCTAAAAGGCCTTAACTTGGTGTATTGAACTGTACGACATGAAACCGCGCATTTTAA 235
Dvir TAGCCTAAAAGGCCGGAT-CCTAGCGTATTGCACTGTACGACATGCATGCGCGCATTTTAA 239
      ***** ** * ***** * * ***** * ***** *

TSS 1

Dmar CTGTTTTCCGAAAACAATT--CGAACCAATATCG--TTT TTTGGCAT TACATACAAT 222
Dkoe CTG-CTTCTGGAAAACAATT--CGTACCAATATCG--TTT TTTGGCAT TACATACAAT 108
Dbuz CTG-CAACAGGAAAACAATT--CGTACCAATATCG--TTT TTTGGCAT TACTTACAAT 290
Dmoj CTG-CAACAGGAAAACAATT--CGTACCAATATCGTATT TTTGGCAT TGCTCACAAA 292
Dvir CTG-CAGCAGCAAAACGAATTTCCGTACCAAA-----TAT TTTGGCAT TATTTGCAAA 291
      *** * * ***** * ** * ***** * ***** * ***** *

TSS 2

Dmar ATGTAATAAATGTGTTAAGACTGTGCCAAAAAGACAAAAGGTTCCCACTGCCATGGTCA 282
Dkoe ATGTAATAAATGTGTTAGGACTGTGCCAAAAAGAGAAAAGGTTCCCACTGCCATGGTCA 168
Dbuz ATGTAATAAATGTGTTAAGACTGTGCCAAAAAGAGAAAAGGTTCTCCCTGCCATGGTCA 350
Dmoj ATGTAATAATATGCGATAGGACTGTGCCAAAAAGAGAAAAGGTTCTCACTGTGATGGCCA 352
Dvir ATGTAATAAATGTGTTA--ATTGTGCCAAAAAGAGAACAGTT-GCACTGCCATGGTCA 348
      ***** * * * * * * ***** * * * * * * * * * * * *

Dmar CAAATGTACACGATTGCAGCTATTTACCGCCTCGCCTCTTCTTTTCATCCCATCTATCGG 342
Dkoe CAAATGTACACGATTGCAGCTATTTACCGTCTCGCCTCGTCTTTTCATCCCATCTATGGG 228
Dbuz CAAATGTACACGATTGCAGCTATTTACCGCCTCGCCTCTTCTTTTCATCCCATCTATGGG 410
Dmoj CAAAAATATACGATTGTATACATACACCGCGGCGCCTCC-----TCTATCGG 399
Dvir CAA----- 351
      ***

Dmar TTCGTGGCTTTGGTTACAAAACTACAAGTCATTGCACCTTTGTTTGTGTGTGTGTCCTG 402
Dkoe TTCGTGGCTTTGGTTACAAAACTACAAGTCATTGTACCTTTGTTTGTGTGTGTGTCCTG 288
Dbuz TTCGTGGCTTTGGTTACAAAACTACAAGTCATTGTACCTTTGTTTGTGTGTGTGTCCTG 470
Dmoj TCCGTGCTTTGGTTGCAAAATCTACAAGTCATTGTACCTTTGTTTGTGTGTGTCCTG 459
Dvir -----GCTTTGGCTGCAAAATCTACAAGTCATTGTACCTTTGTTTGTGTGT----- 397
      ***** * ***** ***** ***** ***** ***** *****

Dmar CATCGCGTATGTTGTCAAGCTTCGCGGTGGCAACGAATTGGTACTGACCG-ACATATTGA 461
Dkoe CATCGCGTATTTGTCAAGCTTCGCGGTGCAACGAATTGGTACTGACCGTACATATTGA 348
Dbuz CATCGCGTATGTTGTCAAGCTTCGAGAGGAACGAATTACTACTGATCGTACATATTGA 530
Dmoj CGTCGTGCTGTTGTCCGCGTTCGCGGTGGCAACGAATTGCACTGACCGTACTCATCGA 519
Dvir -----

```

<i>Dmar</i>	CTGACTGTC-----TGCTGCAAAACCGTCACACAATCAATACAATAAT	505
<i>Dkoe</i>	CTGACTGTC-----TGCTGCAAAACCGTCGCACAATCAATACAATA--	390
<i>Dbuz</i>	CTGAC-----TGCCGCAAAACCGTCGCACAATCAATACAATAAT	570
<i>Dmoj</i>	CGGATTCTCCATTATGCACTACACTTCCCGCATAACCGGTGCAGAATTAATACAATAAT	579
<i>Dvir</i>	-----	
TSS 3		
<i>Dmar</i>	GTGTTAAAAACAAAACAAAGCAAATTTGTTTTTGGGCTG---CAACAGAGCGTTTGCG	562
<i>Dkoe</i>	GTGTTAAAAACAAAACAAAGCC-ATTTTCTTTTCGGCTG---CAACAGAGCGTTTGCG	446
<i>Dbuz</i>	GTGTTAAAAACAAAACAAAG---TATTTTTCGGCTG---CAACAGAGCGTTTGCG	624
<i>Dmoj</i>	TCAACAAAATAAGAA-----ATTTTCTTGTGCAACTGCAACAACAGAGCGTTTGCG	631
<i>Dvir</i>	-----	
TSS 3		
<i>Dmar</i>	AAAATGTTATATGCCGGCGTTTGTAAACAATCAGACAACGCAGAACTGCGTCAACCATAC	622
<i>Dkoe</i>	AAAATGTTATATGCCGGCGTTTGTAAACAATCAGACAACGCAGAACGGCGTCAACTATAT	506
<i>Dbuz</i>	AAAATGTTATATGCCGGCGTTTCGAAACAATCTGACAACGCAGAACGCCGTCAACTATAT	684
<i>Dmoj</i>	AAAATGTTATATGCCGGCGTTTGTAAAGCGAACAGAAAGCACAGCATGGCGTCAGTATAT	691
<i>Dvir</i>	-----	
TSS 4		
<i>Dmar</i>	ATCCTTCATGTATGCCGTGCTACAAAACCTCACAATTATTTACCGAATAATTACTG-ACT	681
<i>Dkoe</i>	ATCCTTCATGTGTGCAGTGTATGAAGCTGCGTATATTTTACCGCATAATTATTG-ATT	565
<i>Dbuz</i>	ATCCTTCATGTATGCAGTGTAGAAAGCTCAGTATATTTTACCACATATTTATTGAAAT	744
<i>Dmoj</i>	ACTCTTCGAGTATACAATTGTAGAGTGTGCTCGATTCTGCCTTGTAACCTAATTATTG-CTT	750
<i>Dvir</i>	-----	
<i>Dmar</i>	TCTATCAATCTATTATCAGTTTACTTACAATTTATCAGACATGCTT-----AAAAAA	736
<i>Dkoe</i>	TCTATCAATTCATTTATCAGTTTACTTACAATTTATCAGACATGCTT-----AAAAAA	620
<i>Dbuz</i>	TCTATCAATCCATTTATCAGTTTACTTACAATTTGATCAGACATGCTT-----AAAAAA	799
<i>Dmoj</i>	TCTATCACTCCAATTAGCATTACTTACAATTTATCGGACATGCTTCA---AAAAAA	807
<i>Dvir</i>	-----TTTACTTACAATTTATCAGACATGCTTCAA--AATAAA	435
	***** **	
START RB		
<i>Dmar</i>	TAGCACTTGAAGCAAATCATCACTTGTGTAGACAGTTGGGAAGTA-AGTGTGACCCTGC	795
<i>Dkoe</i>	TAGCACTCGGAAGCAAATCAGCACTTGTGTAGACAGT-GGGGAGTA-AGTGTGACCATGC	678
<i>Dbuz</i>	TAGCACTCGGAAGCAAATCAGCACTTGTGTAGACAGT-GGGGAGTA-AGTGTGACCATGC	857
<i>Dmoj</i>	TTACACTTGAAGCAAACAGCACTTGTGTAGACAGT-AGGGACTA-AGTGTGACCATGC	865
<i>Dvir</i>	TTACACTTGAAGCAAAGCAGCACTTGTATAGACAGT-ATAGATTA-AGTGTGACCATGC	493
	* **** * ***** ** ***** ***** * ** ***** **	
<i>Dmar</i>	AC---TACCAAAATGTATTTCAAAGT-----	818
<i>Dkoe</i>	ATACAGTAAGAAAATGTTACTAAAAGTTTATAAGATAAAAATACCAAAATGCTAATTGT	738
<i>Dbuz</i>	ACACAGTAATCAATGTAAGTAAAAGTTTATGAGATAAAAATACCAAA---TGCTAATTGT	914
<i>Dmoj</i>	GCACA-----	870
<i>Dvir</i>	TCACA-----	498

Figura 3.25: Alineamiento de la región 5' del gen *CG4673* de las especies *D. martensii* (*Dmar*), *D. koepferae* (*Dkoe*), *D. buzzatii* (*Dbuz*), *D. grimshawii* (*Dgri*), *D. mojavensis* (*Dmoj*) y *D. virilis* (*Dvir*). Se muestran los codones de inicio (START) para los transcritos RB y RA (rojo) y los exones 1 y 0 (amarillo). Se han señalado todos los sitios de inicio de transcripción (TSS) putativos y las cajas TATA ~30 pb aguas arriba (color más oscuro). Los TSS 1 y 4 (gris) son predicciones con puntuación superior al 0,95% (*Berkeley Drosophila Genome Project*). Los TSS 2 y 4 (verde) no han sido predichos pero son regiones conservadas y candidatas a TSS. El elemento BR también se ha señalado para el TSS 3. Una región altamente conservada se ha señalado en rosa. La repetición invertida y límite de la región promotora en 5' se señala en azul.

4. DISCUSIÓN

4.1. Inversiones $2m$ y $2n$

4.1.1. Cartografía de los puntos de rotura

La región de los puntos de rotura en el genoma no invertido de *D. mojavensis* se localizó inicialmente mediante la búsqueda bioinformática con los extremos de los clones BAC secuenciados. Diez de los 18 extremos de los 9 BAC se secuenciaron y localizaron correctamente en el genoma de *D. mojavensis*. Las regiones de los puntos de rotura se localizaron en entre los extremos 14B19-T7 (A) y 20O19-T7 (B), 20O19-SP6 (E) y 16H04-T7 (F), 1N19-T7/14B19-SP6 (C) y 16H04-SP6 (D). Las PCR diseñadas en los fragmentos de la secuencia del *scaffold* 6450 de *D. mojavensis* limitados por los extremos de los clones BAC, permitió acotar las regiones de la cartografía (véase Resultados para más detalles). En *D. mojavensis* la región del punto de rotura EF se localizó en la secuencia intergénica de ~2 Kb entre los genes *Wsck* (E) y *CG8147* (F). El punto de rotura CD quedó cartografiado en un fragmento de ~3 Kb de la región intergénica entre los genes *scrib* (C) y *CG8147* (*Orb98b*). Finalmente, el punto de rotura AB se encontró en un fragmento de ~18 Kb entre los genes *msi* (A) y *Ssadh* (B). Posteriormente, la reanotación del genoma de *D. mojavensis*, permitió acotar dicha región a ~15 Kb entre los genes *CG12250* (A) y *Ssadh* (B). Los genes *CG12250* y *msi* son solapantes con orientación opuesta. La PCR A4, que había sido diseñada en una fracción conservada de la región intergénica entre los genes *msi* y *CG4673*, corresponde con el inicio del gen *CG12250* (véase Figura 3.3 en Resultados). Inicialmente, el gen *CG4673* y sus dos genes anidados (*CG5079* y *CG5071*) no pudieron ser cartografiados en ningún bloque sinténico, por lo que el punto de rotura se ubicó en una región que incluyó dicho conjunto de genes.

4.1.2. Aislamiento y secuenciación de los puntos de rotura

La distancia entre las sondas A4 (*CG12250*) y B2 (*Ssdah*) de la cartografía del punto de rotura AB no pudo ser reducida. Una batería de cebadores fueron diseñados para intentar posicionar el fragmento que incluye las PCR B3 y PCR B4 (véase Figura 3.3 en Resultados) pero no se consiguió ubicarlo contiguo al bloque A o al B. Alternativamente, se pensó que dicha región podría albergar el punto de rotura AB, interrumpiendo la unidad de transcripción o su regulación. Este efecto de posición suele ser detrimental aunque la alteración del patrón de expresión ha sido observada en las inversiones *In(3L)Payne* (Wesley y Eanes 1994) y *2j* (Puig *et al.* 2004) en *Drosophila*.

Esta distancia fue insalvable para amplificar por PCR los puntos de rotura AC y BE en *D. buzzatii* y, sólo el punto de DF rotura pudo ser secuenciado inicialmente (ver más abajo).

4.1.2.1. Hibridación de Southern

Abordamos el problema de la caracterización de la región del punto de rotura AB con una hibridación de Southern (véase Tabla 3.4 en Resultados) con las sondas más cercanas a la cartografía de los puntos de rotura (A4, B2, C1 y E3). En las hibridaciones con las sondas A4 y C1 en las digestiones del clon BAC 1N19 (punto de rotura AC) no se obtuvo ningún fragmento que las contuviera a ambas. Por el contrario, cuando se hibridaron las sondas B2 y E3 en las digestiones del clon BAC 20019 (punto de rotura BE), se obtuvieron 8 patrones de restricción donde el fragmento hibridado era el mismo para ambas sondas.

Al comparar la cantidad de bandas únicas hibridadas para ambas sondas se concluyó que la distancia entre las sondas A4 y C1 debía ser mucho mayor a la que separaba las sondas B2 y E3 en *D. buzzatii*. Concretamente, el fragmento de la digestión del BAC 20019 más

pequeño hibridado con las sondas B2 y E3 indicaría la distancia máxima que las puede separar. El fragmento más pequeño hibridado con ambas sondas fue el obtenido de la digestión con *Pst*I y tenía un tamaño de ~7,5 Kb, tamaño muy inferior al necesario para albergar la mayor parte de las ~15 Kb que separan las sondas A4 y B2 en *D. mojavenis*. De este modo, el punto de rotura AB en *D. mojavenis* debería encontrarse más cerca de la sonda B2. Por tanto, en *D. buzzatii*, la mayor parte de las 15 Kb que incluyen el gen *CG4673* se localizarían contiguas al límite del bloque sinténico A (entre las sondas A4 y C1).

Alternativamente, si la inversión $2m$ se hubiera generado primero, el límite del segmento B podría contener las 15 Kb que separan las sondas A4 y B2. En este caso, el punto de rotura AB en *D. mojavenis* se encontraría cercano a la sonda A4. La inversión $2n$ se produciría solapando una ordenación intermediaria AC-BD-EF. Si el punto de rotura distal de la inversión $2n$ fuera imbricado al punto de rotura proximal de la inversión $2m$, la ordenación final podría ser AC-B'E-D(B'')F donde B'' correspondería a las 15 Kb representadas por las sondas B3 y B4. Sin embargo, esta hipótesis quedó descartada porque en el análisis preliminar del punto de rotura DF en *D. buzzatii*, no se encontraron trazas del punto de rotura AB de *D. mojavenis*.

La sonda B2 fue diseñada en el límite del bloque sinténico B y abarcaba parte del gen *Ssadh* y la región intrónica de éste con el gen *CG4673*. Esta sonda hibridó normalmente con las digestiones del BAC 20O19, pero también hibridó débilmente en el control negativo de la hibridación de Southern con el BAC 1N19. Esta señal sería consistente con una duplicación que, al menos contendría dicha sonda. La debilidad de la señal puede deberse a dos motivos. Por un lado, la duplicación podría incluir sólo parcialmente la sonda. Por otro lado, la duplicación podría tener menor similitud con la sonda debido a la degradación y por tanto, menor afinidad con esta. La sonda B2 hibridó con el mismo fragmento de la digestión del BAC 1N19 con el que hibrida la sonda C1, pero no con el que hibrida la sonda A4. Esta

coincidencia colocaría la posible duplicación (representada por la sonda B2) más cerca del gen *scrib* (sonda C1) que al gen *CG12250* (sonda A4). Finalmente, la única solución a la amplificación y secuenciación de los puntos de rotura AC y BE de *D. buzzatii* fue secuenciar los dos BAC que contenían dichos puntos de rotura.

4.1.2.2. Anotación de los clones BAC 1N19 y 20019

Las secuencias de los clones BAC fueron anotadas y la organización génica comparada con la de las regiones homólogas en el genoma de *D. mojavensis* (véase Figura 3.5 en Resultados).

Genes

La mayor parte del clon BAC 1N19 es ortóloga a la región A de *D. mojavensis* mientras que la mitad del clon BAC 20019 es ortóloga a la región B y la otra mitad a la región E de *D. mojavensis* (véase Figura 3.5 en Resultados). El contenido génico de los clones BAC está altamente conservado con el genoma de *D. mojavensis*. El número, orden, marco de lectura, posición y orientación de los genes anotados en los clones BAC se conservan en *D. mojavensis*. Los clones BAC no contienen por tanto, ningún otro punto de rotura de ninguna otra reorganización ni límite de bloque sinténico aparte del límite entre A y C (BAC 1N19) y entre B y E (20019).

Los puntos de rotura se localizaron en los lugares que la cartografía predijo. El punto de rotura AC fue localizado entre los genes *CG12250* y *scrib* y el punto de rotura BE entre los genes *Ssadh* y *Wsck*. El gen *CG4673* y los dos genes anidados *CG5071* y *CG5079*, que no habían podido ser localizados anteriormente en *D. buzzatii*, fueron localizados en las regiones de los puntos de rotura. El gen *CG5071* fue anotado en el punto de rotura AC. Sólo un

fragmento con homología con el gen *CG5079* de *D. mojavensis* pudo ser anotado también en el punto de rotura AC. Por el contrario, el gen *CG4673*, compactado por la ausencia del intrón 6, se anotó en el punto de rotura BE. En el punto de rotura AC también se encontraron múltiples fragmentos con similitud a *CG4673*. Todas estas observaciones pueden explicarse mediante una duplicación del segmento que contiene *CG4673* y los dos genes anidados *CG5071* y *CG5079*. Posteriormente, la pérdida del intrón 6 del gen *CG4673* en BE, provocaría la pérdida de sus genes anidados. En el punto de rotura AC, los genes *CG4673* y *CG5079* habrían pseudogenizado y sólo el gen *CG5071* permanecería funcional. Por otro lado, el gen *CG4673* anotado en BE tiene una orientación opuesta respecto a la región B (gen *Ssadh*) de *D. mojavensis*. En el genoma no invertido, el gen *Ssadh* y *CG4673* están en la misma cadena (véase Figura 3.5 en Resultados). En *D. buzzatii*, *Ssadh* sigue en la misma cadena y orientación, pero *CG4673* está anotado en la cadena opuesta. Esta anotación indica que la duplicación del fragmento podría estar además microinvertida (ver más abajo).

En la Tabla 3.8, se muestra la densidad de CDS y pb putativamente transcritos de cada región de ambos clones BAC. Las regiones C, B y E se localizan en el interior de las inversiones $2m$ y $2n$ mientras que la región A se localiza en la porción distal fuera de la inversión $2m$ (véase Figura 3.1 en Resultados). Las duplicaciones se han englobado dentro de las regiones de los puntos de rotura. De este modo, la proporción de CDS en los puntos de rotura se refiere al gen *CG5071* (punto de rotura AC) y a *CG4673* (punto de rotura BE). Estas regiones tienen la menor proporción de transcripción (13,6% en AC y 44,90% en BE), aunque no la de CDS. Este efecto se debe a que ambas regiones de los puntos de rotura, tienen un gen abundante en pb exónicos pero no en intrónicos. En AC, el gen *CG5071* está codificado por un sólo exón. En BE, el gen *CG4673* está codificado por 8 exones, pero se encuentra muy compactado debido a la pérdida del intrón 6. En cualquier caso, los puntos de rotura se observan en las regiones de menor densidad génica. Ninguna de las inversiones descritas en *Drosophila*

(véase Tabla 1.1 en Introducción) interrumpe una unidad transcripcional. La selección actúa contra las inversiones con efectos de posición desfavorables por lo que se esperan mayor cantidad de puntos de rotura en regiones con densidad génica baja (Sperlich 1963).

Elementos transponibles

Los TE anotados en los dos clones BAC secuenciados se muestran en las Tablas 3.6 y 3.7 de Resultados. En la Tabla 3.8, se muestra la densidad de los elementos en cada región para ambos clones BAC. En la Tabla 4.1 hemos agrupado esta información para los TE dentro de las inversiones $2m$ y $2n$ (regiones C, B y E), fuera de ellas (región A), en las regiones de los puntos de rotura (BE y AC) y en el total de los dos clones BAC secuenciados.

Tabla 4.1: Resumen del contenido en elementos transponibles de los dos BAC secuenciados organizado y agrupado en (i) dentro de las inversiones (regiones B, C y E), (ii) fuera de las inversiones (región A) y (iii) en puntos de rotura (AC y BE).

	dentro de las inversiones (regiones B, C y E)	fuera de las inversiones (región A)	puntos de rotura (AC y BE)	TOTAL
pb	158.123	76.970	23.585	258.678
pb TE	3.143	1.284	1.430	5.857
% TE	1,98	1,67	6,06	2,26
n° TE	9	7	6	22
n°	56,92	90,94	254,40	85,05
inserciones/Mb				
Tamaño medio de cada TE (pb)	349,22	183,42	357,5	266,22

Hemos analizado las diferencias observadas en las densidades de TE en las regiones de los clones BAC. En las regiones de los puntos de rotura, se observa una densidad de 6,06% de TE, lo que significa tres veces mayor densidad que fuera de éstos. Hemos realizado un ANOVA para comparar las densidades observadas en las regiones fuera de las inversiones, dentro de las inversiones y en los puntos de rotura. Las densidades observadas son significativamente distintas ($P=0,0335$) (Tabla 4.2). Sin embargo, la práctica totalidad de la región C (inversión 2*m*) del BAC 1N19, está anotada como gen *scrib* (véase Figura 3.5 en Resultados). Podríamos estar observando menor densidad dentro de las inversiones debido a que la selección no haya permitido más inserciones dentro del gen. Sólo 436 de los 1.572 elementos *D. melanogaster* han sido anotados dentro de regiones que se transcriben (Kaminker *et al.* 2002).

Para analizar las proporciones de TE en cada región, hemos comparado las ventanas de 5 Kb descritas en las Tablas 3.6 y 3.7. Hemos organizado las ventanas en grupos según la región a la que pertenecen. De este modo, hemos considerado las proporciones de 30 ventanas dentro de inversiones, 20 fuera de las inversiones y 5 en los puntos de rotura. Sólo la comparación de las proporciones entre los TE fuera de las inversiones y en los puntos de rotura es significativa ($P=0,0039$) (Tabla 4.2).

Sin embargo hay que tener en cuenta que este análisis no contabiliza el número de TE insertados, sino la proporción de pares de bases anotadas como tal. El sesgo observado podría deberse simplemente a que las inserciones en los puntos de rotura son más recientes y, por tanto, menos degeneradas. Es decir, con fragmentos de TE anotados más largos.

Tabla 4.2: Análisis de la varianza (ANOVA) de un factor entre la cantidad de TE anotados en cada región de los dos BAC secuenciados. 1. Comparación entre grupos: Se han comparado las densidades de los tres grupos usando la proporción de TE (%) de cada región de la Tabla 3.8: (i) dentro de inversiones (densidades de C, B y E); (ii) fuera de inversiones (densidad de A); (iii) puntos de rotura (densidades de AC y BE). Las comparaciones 2, 3 y 4 se han realizado usando la proporción de TE en cada ventana de 5 Kb de las Tablas 3.6 y 3.7 correspondientes a las regiones dentro de las inversiones, fuera de las inversiones y en puntos de rotura. 2. Comparación dentro/fuera; 3. Comparación fuera/punto de rotura; 4. Comparación dentro/punto de rotura. *P<0,05.

1. Análisis de varianza de un factor entre dentro, fuera y punto de rotura

GRUPOS	n°porcentajes	Suma	media	Varianza
Dentro	3	4,73	1,576667	1,541633
Punto rotura	2	12,25	6,125	0,08405
Fuera	1	1,67	1,67	-

	SS	df	MS	F	P-value	F crit
Entre grupos	27,30737	2	13,65368	12,93241	0,033506	9,552094
Dentro grupos	3,167317	3	1,055772			

2. Análisis de varianza de un factor entre dentro y fuera

GRUPOS	n°ventanas	TE(pb)	media de TE (pb)	Varianza
Dentro	30	3146	104,8667	81407,29
Fuera	20	1284	64,2	20459,33

	SS	df	MS	F	P-value	F crit
Entre grupos	19845,33	1	19845,33	0,346449	0,55889	4,042652
Dentro grupos	2749539	48	57282,06			

3. Análisis de varianza de un factor entre fuera y punto de rotura

GRUPOS	n°ventanas	TE(pb)	media de TE (pb)	Varianza
Fuera	20	1284	64,2	20459,33
Punto rotura	5	1488	297,6	24533,3

	SS	df	MS	F	P-value	F crit
Entre grupos	217902,2	1	217902,2	10,29402	0,003899	4,279344
Dentro grupos	486860,4	23	21167,84			

4. Análisis de varianza de un factor entre dentro y punto de rotura

GRUPOS	n°ventanas	TE(pb)	media de TE (pb)	Varianza
Punto rotura	5	1488	297,6	24533,3
Dentro	30	3146	104,8667	81407,29

	SS	df	MS	F	P-value	F crit
Entre grupos	159197,7	1	159197,7	2,136496	0,153288	4,139252
Dentro grupos	2458945	33	74513,47			

Las diferencias observadas podrían deberse al tamaño de los TE, y no a la densidad de éstos. En la Tabla 4.1 también se muestra la estima del número de inserciones de TE en cada región. Cuando dos fragmentos del mismo tipo de TE han sido anotados a menos de 0,5 Kb el uno del otro, hemos considerado que se trata de la misma copia insertada. De este modo, en el BAC 1N19, observamos que sólo hay un elemento en la región C (*TART*), cinco en el punto de rotura AC (*BuT5*, *Helitron*, 2 *ISBu* y *BS2*) y siete en la región A (2 *ISBu*, 3 *TART* y 2 *Helitron*) (véase Tabla 3.6 en Resultados). Del mismo modo, para el BAC 20O19, observamos que hay 5 elementos en la región B (*Mini-me*, *ISBu*, 3 *TART*), uno en el punto de rotura BE (*BuT5*) y 3

en la región E (*Helitron*, *TART* y *ISBu*) (véase Tabla 3.7 en Resultados). En total, observamos 9 TE dentro de las inversiones, 7 fuera de las inversiones y 6 en los puntos de rotura. Si normalizamos el número de TE por el tamaño de cada región, observamos que la tasa de inserción por Mb en cada región será de 56,92 TE/Mb dentro de las inversiones; 90,94 TE/Mb fuera de las inversiones y 254,40 TE/Mb en los puntos de rotura (véase Tabla 4.1). Si dividimos el número de pb anotados como TE en cada región entre el número de copias insertadas, obtenemos el tamaño medio de cada elemento. La media de tamaños es similar dentro de inversiones (349,22 pb) y en las regiones de los puntos de rotura (357,5 pb) pero inferior fuera de las inversiones (183,42 pb).

Por tanto, la mayor densidad observada en los puntos de rotura podría deberse al mayor tamaño de los TE. De todos modos, la tasa de inserción por Mb en las regiones de los puntos de rotura es 4 veces mayor que dentro de las inversiones y 2,5 veces mayor que fuera de ellas.

Estos resultados indican que los TE anotados en los puntos de rotura y dentro de inversiones están menos degenerados. Este efecto observado tiene dos explicaciones. Por un lado, estas inserciones podrían haberse producido más recientemente en el tiempo. Por otro lado, la degeneración de estos elementos, se habría visto frenada en regiones de baja recombinación, quedando protegidos de su eliminación. De todos modos, no esperamos diferencia en la densidad entre las regiones comparadas ya que la baja recombinación observada en los heterocariotipos tiende a acumularse más allá de los puntos de rotura y no sólo dentro de las inversiones (Andolfatto *et al.* 2001). Además, la tasa de eliminación de los TE se reduce en los heterocariotipos debido a la reducción de la recombinación sólo cuando las inversiones son polimórficas (Montgomery *et al.* 1987; Langley *et al.* 1988; Lyttle y Haymer 1992; Cáceres *et al.* 2001; Bartolomé *et al.* 2002; Casals *et al.* 2006). Cuando la inversión aumenta en frecuencia y el número de éstos se reduce, la eliminación de TE por recombinación ectópica sucede normalmente (Petrov *et al.* 2003).

La significación observada en la densidad de los TE en los puntos de rotura podría deberse sólo a la densidad de unidades de transcripción. Las regiones de los puntos de rotura AC y BE son las regiones con menor densidad génica, por lo que podrían acumular más cantidad de TE al ser sus inserciones aquí menos detrimentales (Kaminker *et al.* 2002).

Por otro lado, hemos analizado conjuntamente los resultados observados en el total de los dos clones BAC (considerando todas las regiones). Kaminker *et al.* (2002), estudiaron la fracción de TE en la porción eucromática del genoma de *D. melanogaster*. La densidad de TE en la eucromatina es del 3,86% y la media de tamaño de los 1.572 TE anotados es de 2.869 pb. La estima del número de inserciones por Mb resultó de 13,46 elementos. La densidad de TE en los dos clones BAC de *D. buzzatii* es de 2,26%, algo menor que la observada en la región eucromática del genoma de *D. melanogaster*. Sin embargo, la media de tamaño estimado para los TE anotados en Kaminker *et al.* (2002) muy superior a los 266,22 pb observados en los clones BAC de *D. buzzatii* (véase Tabla 4.1). De este modo, la diferencia observada en la densidad, podría deberse al tamaño y no al número de inserciones. En los clones BAC de *D. buzzatii* observamos 22 inserciones. El número promedio de inserciones es de 85,05 TE por Mb. Esta estima es muy superior a las 13,46 inserciones por Mb calculada para *D. melanogaster*. Esta observación podría deberse a que anotaciones de fragmentos muy separados se hayan considerado parte del mismo TE. De este modo, se explicaría porque observan mayor densidad y un orden de magnitud mayor el tamaño medio de los elementos anotados. Si sólo consideramos la fracción eucromática del elemento E de Muller de *D. melanogaster* (Kaminker *et al.* 2002), las observaciones son parecidas a las observadas para todo el genoma. Se han contabilizado 288 TE con una media de tamaño de 3.010 pb y 10,33 inserciones por Mb. Las diferencias entre el trabajo de Kaminker *et al.* (2002) y nuestros resultados no se deben por tanto a diferencias entre cromosomas. Sin embargo estas estimas consideran todo el genoma de *D. melanogaster* mientras que la que obtenemos en este trabajo

sólo considera una fracción muy pequeña del genoma de *D. buzzatii* que podría sesgar la observación.

4.1.3. Generación de las inversiones $2m$ y $2n$

La caracterización y el análisis molecular de los puntos de rotura de una inversión no siempre permiten resolver el origen de ésta. En numerosas inversiones, no se pudo resolver el mecanismo que las ha generado. La presencia de TE en los puntos de rotura de una inversión no siempre se ha asociado a recombinación ectópica o inserción híbrida. Un elemento en la región del punto de rotura puede no haber intervenido en el origen de la inversión y haberlo colonizado posteriormente (Cáceres *et al.* 2001; Casals *et al.* 2006). Sólo cuando éste se anota justo en la junta entre bloques conservados puede haber participado en el origen de la inversión. Del mismo modo, no siempre se puede identificar la junta. La presencia de los elementos tipo *Helitron* y *TART* en los puntos de rotura de la inversión *5g* (Prazeres *et al.* 2009) y *BS* y *CR1a* en la inversión *3R(8)*, no fueron anotados en la junta, por lo que no fueron relacionados con el origen de la inversión. El complejo de secuencias repetitivas en los puntos de rotura de *2La* (Sharakhov *et al.* 2006), el elemento *G* en el punto de rotura proximal de *In(3R)Payne* (Matzkin *et al.* 2005), el elemento tipo *LINE* en el punto de rotura distal de *In(2L)* (Andolfatto *et al.* 1999) y el elemento *Odysseus* en el punto de rotura distal de *2Rd'* (Mathiopoulos *et al.* 1998) aún estando anotados en las juntas entre bloques conservados no pudieron ser relacionados con el mecanismo que ha generado las inversiones (véase Tabla 1.1 en Introducción).

4.1.3.1. Duplicación asociada a la inversión 2m

La anotación de los clones BAC 1N19 y 20O19 permitió detectar una duplicación asociada a la inversión 2m. En el punto de rotura proximal (BE), sólo el gen *CG4673* fue anotado y es putativamente funcional (véase sección 4.1.2.2). En el punto de rotura distal de la inversión 2m (AC), sólo el gen *CG5071* sigue siendo putativamente funcional. En este caso, algunos fragmentos de identidad con los genes *CG4673* y *CG5079* de *D. mojavensis* pudieron ser anotados. Las anotaciones más remotas de $\Psi CG4673$ contienen el gen *CG5071* y el pseudogen $\Psi CG5079$.

Aunque algunos genes anotados estén degradados, parece razonable suponer que el fragmento entero (incluyendo el gen *CG4673* y los anidados *CG5071* y *CG5079*) fue inicialmente duplicado. La pseudogenización de $\Psi CG4673$ y $\Psi CG5079$ en el punto de rotura AC, así como la pérdida del intrón 6 de *CG4673* en el punto de rotura BE, ocurrieron más tarde. De este modo, la duplicación no ha provocado el aumento en el número de copias de estos genes en *D. buzzatii*. De hecho, el gen *CG5079* se ha perdido en ambos puntos de rotura. Un fragmento del punto de rotura AC con identidad con el intrón 6 del gen *CG4673* de *D. mojavensis* fue anotado microinvertido (véase Tabla 3.9 en Resultados). Este fragmento se localiza en la región 5' de $\Psi CG5079$ en AC en *D. buzzatii* y podría ser el motivo de la degradación de este gen. Sin embargo, en esta región no hay anotado ningún promotor en los genomas de *D. mojavensis* y *D. melanogaster*. La duplicación supone otros dos eventos en el genoma de *D. buzzatii*. En primer lugar, el gen *CG4673* se ha compactado debido a que la copia putativa funcional localizada en BE ha perdido las ~9 Kb de la secuencia del intrón 6 (véase Tabla 3.13 en Resultados). Esta compactación podría suponer una transcripción más eficiente y rápida en los genomas invertidos. En segundo lugar, el gen *CG5071*, anidado en *CG4673* en *D. mojavensis*, no se encuentra anidado en *D. buzzatii*.

Estos resultados observados en la secuenciación de los dos clones BAC, son coherentes con las conclusiones obtenidas en la hibridación de Southern. Por un lado, el fragmento de ~15 Kb correspondiente al gen *CG4673* (aunque éste esté degenerado) se ha anotado entre las sondas A4 y C1. Por otro lado, una copia compactada del gen *CG4673* también se ha anotado en el punto de rotura BE (entre las sondas B2 y E3), produciendo la señal débil obtenida en el control negativo. La sonda B2 delimitaba la cartografía del bloque sinténico B porque amplifica el último exón del gen *Ssadh* y la región intergénica con el gen *CG4673*. La amplificación y la señal fuerte en la hibridación de Southern en el BAC20019 de esta PCR indicaron que se había diseñado en el bloque sinténico B. Sin embargo, la región intergénica que amplifica la PCR B2 corresponde a la región aguas arriba del gen *CG4673*, también presente en el punto de rotura AC y, por tanto, se obtuvo una señal débil en las hibridaciones de esta sonda en el BAC 1N19. Tal y como se discutió en las conclusiones de la hibridación de Southern, la homología de la sonda B2 en el BAC 1N19 se localizó más cerca de la sonda C1 (gen *scrib*) ya que, entre B2 y A4 (*CG12250*) está Ψ *CG4673*.

La presencia de duplicaciones en los puntos de rotura como consecuencia del mecanismo de inversión ha sido descrita en trabajos anteriores. En el grupo *melanogaster*, 18 de las 29 inversiones fijadas entre *D. melanogaster* y *D. yakuba* se han asociado a duplicaciones de genes (Ranz *et al.* 2007). La mayoría de estas duplicaciones se encuentran invertidas (17 de 18). Por otro lado, sólo 6 procederían de ambos puntos de rotura, mientras que en 12 casos se detectó un fragmento duplicado e invertido de un sólo punto de rotura. También en la inversión *In(3R)Payne* (Matzkin *et al.* 2005) se observaron duplicaciones invertidas de ambos puntos de rotura. Cuando ambos puntos de rotura se duplican, la explicación más plausible consiste en la rotura escalonada de ambos puntos de rotura y reparación errónea del fragmento (Krimbas y Powell 1992; Ranz *et al.* 2007). En la inversión *2La* en el complejo *Anopheles gambiae* (Sharakhov *et al.* 2006) se detectó un fragmento

duplicado de sólo un punto de rotura y la duplicación anotada en *Xe* en *D. mojavensis* (Cirulli y Noor 2007) es anterior al origen de la inversión (Runcie y Noor 2009). Para ambas inversiones, la rotura y reparación es consistente pero pueden explicarse por otros mecanismos. La inversión *3R(6)* de *D. melanogaster* (MacDonald y Long 2006) es la única inversión con duplicaciones directas. Ranz *et al.* (2007) explican que una microinversión habría ocurrido para explicar dicha observación. No se ha descrito ningún mecanismo que produzca duplicaciones directas en los puntos de rotura. La duplicación directa encontrada en la inversión *2m* también parece ser debida a la microinversión de la duplicación (véase sección 4.1.2.2). Cuando sólo uno de los puntos de rotura está duplicado, aparte del mecanismo de NHEJ (un punto de rotura escalonado y otro de doble cadena), el mecanismo de inserción híbrida entre distintas copias de un TE (Chalmers y Kleckner 1996; Gray *et al.* 1996; Gray 2000) también puede explicar dichas duplicaciones invertidas.

Mecanismos de duplicación

Inserción híbrida

En el sistema clásico de transposición de cortar-pegar, la transposasa escinde el TE del genoma y éste se inserta en una nueva diana. Alternativamente, la escisión errónea (transposición alternativa) implica el extremo 5' de un elemento y el extremo 3' de otro cercano (Gray *et al.* 1996). La inserción híbrida de este elemento en una nueva diana provoca distintos tipos de reorganizaciones (Zhang y Paterson 2005). Este mecanismo también puede generar duplicaciones. En el caso de la inversión *2m*, la duplicación observada se produciría con la transposición alternativa de un elemento insertado aguas arriba del gen *CG4673* y otro elemento en la misma orientación aguas abajo del gen *CG4673* en un cromosoma homólogo (Figura 4.1).

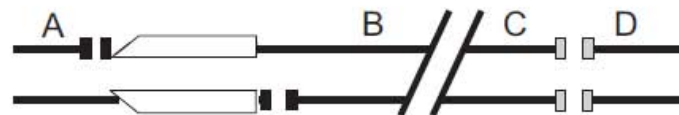
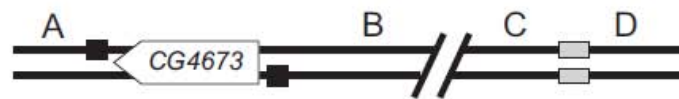
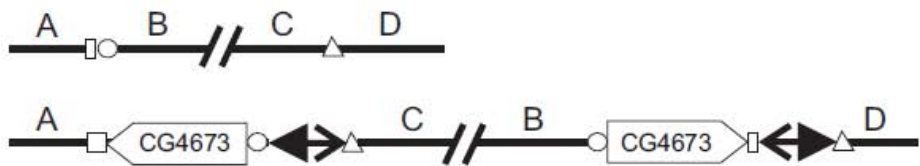
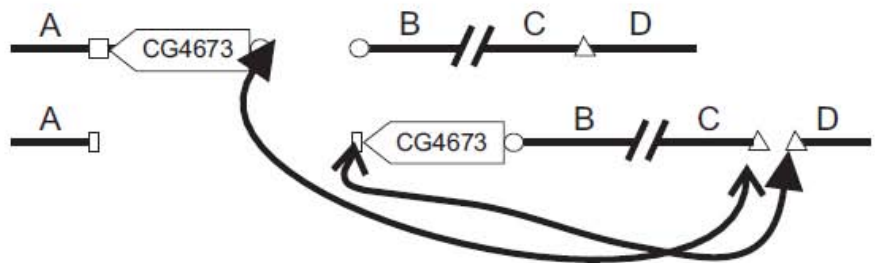
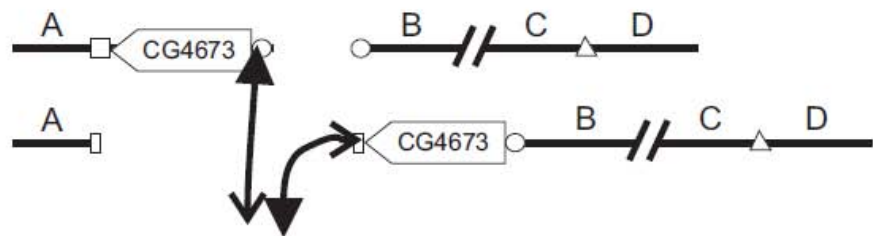
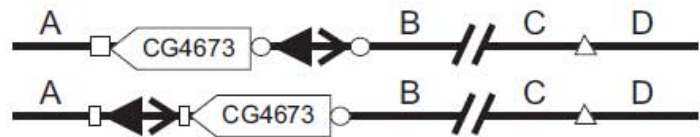


Figura 4.1: (Página anterior). Posibles mecanismos de inversión del fragmento B-C (inversión $2m$) con duplicación del gen *CG4673*. En ambos casos, la duplicación queda en orientación opuesta. Izquierda, transposición alternativa. Derecha, rotura más reparación por NHEJ. En el primer mecanismo, las flechas indican copias de TE en la misma orientación (círculo y cuadrado señalan los sitios de inserción de éstos). Primero se produce la transposición alternativa con escisión de extremos complementarios de cada una de las copias. La inserción híbrida en la nueva diana (triángulo) produce un cromosoma homólogo o cromátida con una inversión y una duplicación asociada. Los otros dos extremos podrían unirse observando una delección. Los TE quedan en orientación opuesta. En el segundo mecanismo, se muestran las dos hebras de la cadena. Los cuadrados negros indican dónde se produce una rotura de una única hebra (SSB) y los cuadrados grises dónde se produce una rotura de doble hebra (DSB). Después de la inversión, se reparan los extremos (flechas negras) y las cadenas de hebra sencilla (flechas blancas).

La inversión con la duplicación del gen *CG4673* se produciría si la inserción del elemento híbrido se produjera en el punto de rotura CD. En la anotación de los puntos de rotura, distintas copias del elemento transponible *BuT5* fueron anotadas aguas abajo de las dos regiones duplicadas como se espera en el modelo de inserción de un elemento híbrido (véase Figura 3.7 en Resultados). En la inversión *X(1)* de *D. melanogaster* (Ranz *et al.* 2007), como en la inversión $2m$, se ha descrito una duplicación invertida además de repeticiones directas de TE en los puntos de rotura. El mecanismo propuesto para esta inversión fue la rotura escalonada. Los TE suelen colonizar los puntos de rotura debido a la baja recombinación y su incapacidad para eliminarlos (Langley *et al.* 1988; Ashburner 1989; Montgomery *et al.* 1991; Charlesworth 1996; Goldman y Lichten 1996). Aun cuando el mecanismo de inversión más plausible es la reparación errónea de un fragmento se han encontrado elementos transponibles. En las inversiones con duplicaciones invertidas originadas por NHEJ *In(3R)Payne* (Matzkin *et al.* 2005) y *3R(8)* (Ranz *et al.* 2007) se han descrito TE en los puntos de rotura. También en la

inversión *5g* en *D. buzzatii* (Prazeres *et al.* 2009), aunque el mecanismo más plausible fuera el resultado limpio de cortar-pegar, los puntos de rotura también han sido colonizados por TE a posteriori (véase Tabla 1.1 en Introducción). En las inversiones *2Rd'* (Mathiopoulos *et al.* 1998), *In(2L)* (Andolfatto *et al.* 1999) y *2La* (Sharakhov *et al.* 2006) también han sido anotados distintos TE aunque no se ha descartado su implicación en el mecanismo que las ha generado (véase Tabla 1.1 en Introducción).

Además, el propio *BuT5* se espera en inversiones y puntos de rotura como un clásico colonizador secundario (Cáceres *et al.* 2001; Casals *et al.* 2006; Delprat *et al.* 2009).

Rotura escalonada

Las inversiones producidas por NHEJ también pueden producir duplicaciones al reparar las roturas escalonadas. La rotura escalonada está compuesta por dos roturas SSB cercanas (una en cada hebra) produciendo extremos cohesivos. La reparación errónea invierte el fragmento situando cada una de las dos hebras del extremo cohesivo flanqueando dicho fragmento invertido. La reparación de estas hebras de cadena simple genera una duplicación en orientación opuesta que rodea la inversión (véase Figura 1.1 en Introducción). En la inversión *2m*, el extremo distal sería la rotura cohesiva y la proximal de cadena doble (Figura 4.1). Las roturas de una hebra de la rotura escalonada se producirían aguas abajo y aguas arriba del gen *CG4673*. Por tanto, el extremo cohesivo incluye todo el gen *CG4673*. La distancia entre ambas roturas de cadena simple y por tanto el tamaño de la duplicación es de ~13 Kb. Las duplicaciones descritas en inversiones originadas por este mecanismo son de menor tamaño. En la comparación de los genomas entre *D. melanogaster* y *D. yakuba*, el rango de estas duplicaciones varía entre 46 pb y 1,9 Kb, con 321 pb de promedio (Ranz *et al.* 2007). La duplicación descrita en la inversión *2La* en *A. gambiae* (Sharakhov *et al.* 2006) llega hasta los 750 pb mientras que en la inversión *In(3R)Payne* (Matzkin *et al.* 2005), la

duplicación en el punto de rotura distal es de ~2 Kb y la del punto de rotura proximal se ha anotado en ~7 Kb.

Microinversión

Los dos modelos que hemos discutido generan duplicaciones con orientación invertida (Figura 4.1). Sin embargo, la duplicación asociada con la inversión $2m$ que incluye el gen *CG4673* está en la misma orientación en los dos puntos de rotura AC y BE. Por tanto, la duplicación ocurrió primero con la inversión $2m$ y, posteriormente, este fragmento duplicado se microinvirtió. La comparación con el genoma de *D. mojavensis* indica que la microinversión ocurrió en la duplicación proximal (véase sección 4.1.2.2). Por lo tanto, la microinversión cambió la orientación del gen *CG4673* en el punto de rotura BE. Casos similares de microinversión han sido descritos previamente para explicar la orientación de los genes duplicados que flanquean una inversión (Macdonald y Long 2006; Ranz *et al.* 2007).

Una repetición invertida de ~120 pb fue anotada flanqueando el gen *CG4673* en el punto de rotura BE. Esta secuencia repetida está anotada como copia única en la región 5' del gen *CG4673* en *D. mojavensis* y en la región 5' de $\Psi CG4673$ en el punto de rotura AC. No podemos por lo tanto invocar la recombinación ectópica entre las repeticiones invertidas como mecanismo que generó la microinversión. A diferencia de la inversión $2m$, sólo se anotaron trazas de TE en la región aguas abajo de la duplicación proximal de la microinversión (véase Figura 3.7 en Resultados) y no en ambas regiones y en orientación opuesta como se describió en el modelo de inserción híbrida (Figura 4.1). Sin embargo, la rotura escalonada y NHEJ produciría la microinversión de esta duplicación y explicaría la presencia de repeticiones invertidas flanqueando el gen *CG4673* en *D. buzzatii*.

El gen *CG4673* en el intermediario no microinvertido se localizaría en orientación directa 5'-3'. En *D. mojavensis*, la secuencia de la repetición ha sido anotada en la región 5'.

De este modo, en esta microinversión, la rotura escalonada procedería del punto de rotura distal (aguas arriba de *CG4673*). Al no haber encontrado duplicaciones procedentes de la rotura proximal (aguas abajo de *CG4673* en BE) se entiende que, como en la inversión *2m*, la rotura aquí fue de doble cadena. La duplicación de la inversión *2m* también procede del punto de rotura distal.

La duplicación inesperada, la degradación del gen *CG4673* en AC y la microinversión explican porque las PCR para intentar amplificar los puntos de rotura en *D. buzzatii* no funcionaron. Los intentos de localizar el gen *CG4673* en AC habrían fracasado debido a la degradación del gen. Los cebadores diseñados en *CG4673* no tienen homología con la copia pseudogenizada anotada en AC. Cuando intentamos amplificar el gen en BE, los resultados fueron negativos debido a la orientación inesperada causada por la microinversión.

4.1.3.2. Inversión 2n

No hay duplicaciones asociadas a la inversión *2n*. Algunos fragmentos de TE fueron anotados en ambos puntos de rotura de la inversión. La copia de *ISBu* anotado en el punto de rotura proximal (DF) está anidado en un elemento tipo *transib*. *BuT5* ha sido anotado tanto en el punto de rotura distal (BE) como en el proximal (DF) de esta inversión. El origen de esta inversión puede explicarse, por tanto, con distintos mecanismos en función de la implicación de este elemento. Por un lado, la recombinación ectópica entre las dos copias del elemento *BuT5* y, por otro, de nuevo NHEJ aunque con roturas de doble cadena.

La recombinación ectópica entre copias de *BuT5* no ha sido descrita en ninguna inversión anterior. Sin embargo, no hay motivos que impidan pensar que pueda ocurrir recombinación entre dos copias de este elemento. Debido al alto grado de degradación de estas copias, no se pudo diferenciar si se trata de copias quiméricas, por lo que no podemos estar

seguros de si la recombinación fue la causa de la inversión $2n$. En los dos clones BAC secuenciados, sólo se ha anotado *BuT5* en la región de los puntos de rotura. Si la inserción de este elemento fuera secundaria, esperaríamos encontrar alguna copia de este elemento también en el interior de las inversiones. Casals *et al.* (2006), analizaron la distribución de este elemento mediante hibridación *in situ* en distintas cepas de *D. buzzatii*. En el cromosoma 2 de la cepa st-1 de *D. buzzatii* sólo se encontró este elemento en la región heterocromática del centrómero y otras dos copias que podrían tratarse de las encontradas en este trabajo en los puntos de rotura AC y DF de las inversiones $2m$ y $2n$. En cualquier caso, la copia anotada en el punto de rotura BE en este trabajo no fue detectada en Casals *et al.* (2006). Parece poco probable encontrar este elemento sólo en los puntos de rotura por azar. De todos modos, *BuT5* si ha sido descrito como un clásico colonizador secundario. En el mismo trabajo (Casals *et al.* 2006), encontraron algunas copias mas de *BuT5* en los puntos de rotura y dentro de las inversiones en las cepas j-2, jq⁷-4, jz³-6 y y³-1. En la inversión $2q^7$ (Casals *et al.* 2003) se encontraron copias de *BuT5* aparentemente no relacionadas con el origen de la inversión en ambas regiones de los puntos de rotura. En la inversión $2z^3$ (Delprat *et al.* 2009) también se encontraron copias de *BuT5* en ambos puntos de rotura aunque la copia anotada en el punto de rotura proximal no está en la junta y también se encontró en el cromosoma parental no invertido. En la inversión $2j$ (Cáceres *et al.* 1999) se encontró un elemento *BuT5* en el punto de rotura proximal de una de las líneas secuenciadas. En la inversión $2z^7$ (en este trabajo) también se ha encontrado una copia de *BuT5* fijada en el punto de rotura proximal en *D. buzzatii*, *D. koepferae* y *D. martensis*.

Por otro lado, en los puntos de rotura de las inversiones $3R(8)$ (Ranz *et al.* 2007) y *In(3R)Payne* (Matzkin *et al.* 2005) originadas por NHEJ también se han encontrado TE no relacionados con la inversión. De este modo, la rotura y reparación errónea no puede ser descartada como el mecanismo que hubiera originado la inversión $2n$. Debido a la ausencia de

duplicaciones asociadas a esta inversión, en este caso ambos puntos de rotura serían de doble cadena.

4.1.4. Relación entre las inversiones $2m$ y $2n$

Los posibles mecanismos descritos para las inversiones $2m$ y $2n$ pueden ser analizados de forma conjunta. A priori, no tenemos ninguna evidencia de la cronología de ambas inversiones. El paso intermedio e hipotético para llegar desde el Primitivo I del grupo a *D. buzzatii*, bien podría ser AB-CE-DF (si $2n$ hubiera ocurrido primero) o AC-BD-EF (si $2m$ hubiera ocurrido primero). No se han observado trazas de imbricación o independencia entre ellas. Por tanto, el orden de las dos inversiones no afecta a la conclusión que la reutilización del punto de rotura compartido entre ambas inversiones es completa. Si la reutilización no fuera completa, en alguno de los puntos de rotura de *D. buzzatii* se encontraría porciones procedentes de los tres puntos de rotura caracterizados en *D. mojavensis*. Estas porciones pueden ser lo suficientemente pequeñas para no haber sido localizadas mediante la hibridación *in situ* pero tal vez se habrían observado a nivel molecular ya que se han podido anotar las demás regiones intergénicas. En el caso que las dos inversiones se originaran con puntos de rotura independientes, el BAC 20O19, aparte de las regiones anotadas como B y E, debería contener entre ellas, una fracción parcial de la región del punto de rotura CD (Figura 4.2).

Por otro lado, también podría haber ocurrido imbricación (solapamiento) entre la porción proximal de la inversión $2m$ y la porción distal de la inversión $2n$. En este caso, una fracción anotada en *D. mojavensis* como región E se hubiera localizado en el punto de rotura AC de *D. buzzatii* ($2m$ distal) si la inversión $2n$ se hubiera producido primero. Si la inversión $2m$ se hubiera generado primero, una fracción anotada en *D. mojavensis* como región E habría sido anotada en el punto de rotura DF ($2n$ proximal) (Figura 4.2).

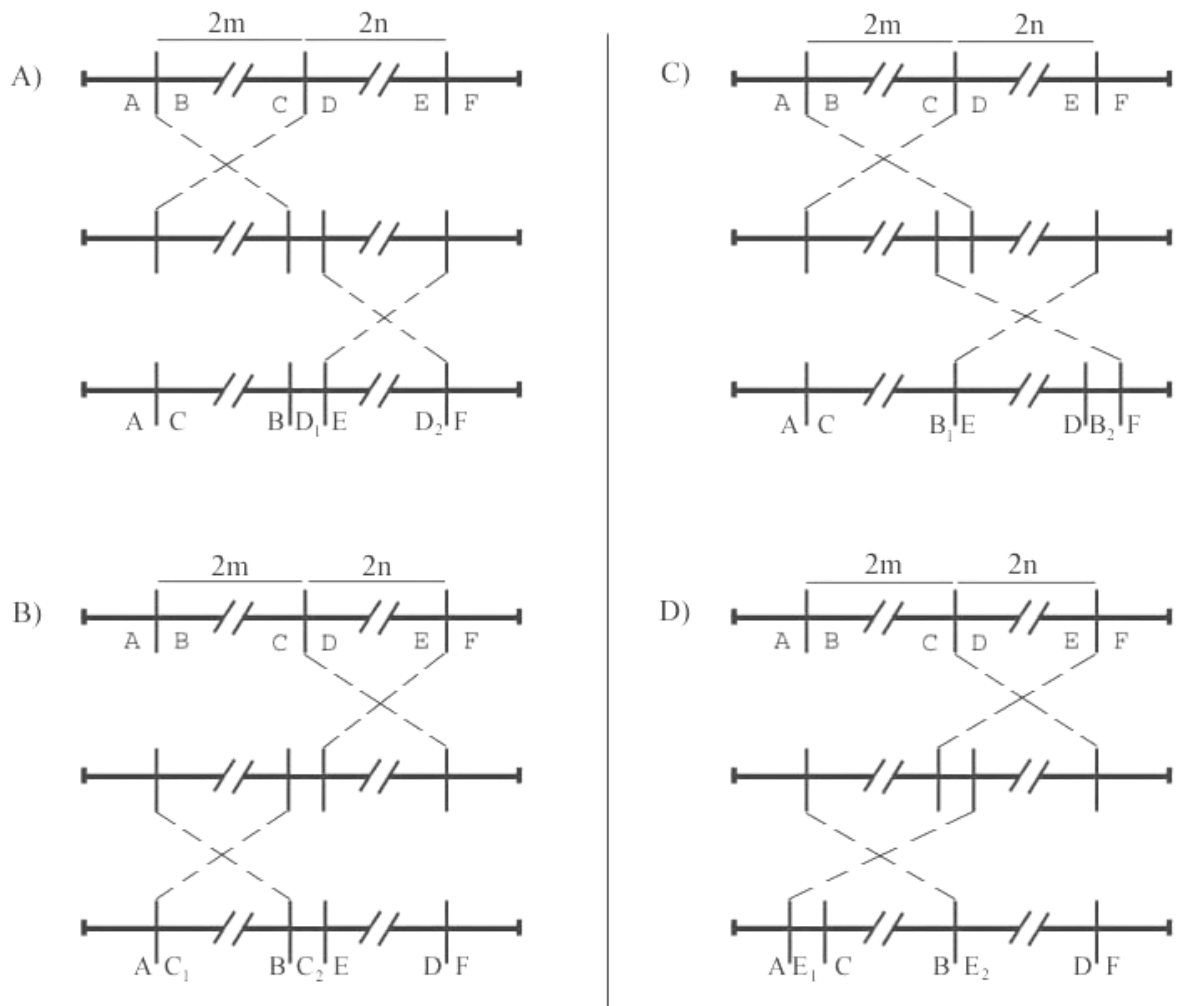


Figura 4.2: Posibles relaciones entre las inversiones $2m$ y $2n$ y el contenido molecular esperado en los puntos de rotura de *D. buzzatii* para los casos de imbricación e independencia. A) Independencia siendo $2m$ anterior a $2n$. B) Independencia siendo $2n$ anterior a $2m$. C) Imbricación con $2m$ anterior a $2n$. D) Imbricación con $2n$ anterior a $2m$.

4.1.4.1. Reutilización del punto de rotura

El modelo Nadeau y Taylor (1984) propone que los puntos de rotura de las reorganizaciones se distribuyen al azar. Por el contrario, regiones particulares del genoma revelan un exceso de pequeñas reorganizaciones y una extraordinaria densidad de puntos de

rotura en la comparación de los genomas de humano y ratón (Bailey *et al.* 2004) y en *Drosophila* (Wasserman 1992; Bhuktar *et al.* 2008). El modelo de distribución no al azar de los puntos de rotura defiende que hay regiones que acumulan más roturas y éstas tienden a reutilizarse (Pevzner y Tesler. 2003; Bourque *et al.* 2004). Este modelo postula que las microinversiones y la reutilización son la razón por el que el primer modelo de distribución al azar descrito por Ohno (1973) no es válido. La acumulación de puntos de rotura en la región BE de *D. buzzatii* y la pequeña microinversión de la duplicación, son consistentes con el modelo de distribución no aleatoria de los puntos de rotura. La compleja reorganización compuesta por las inversiones $2m$, $2n$ y la microinversión implica 4 puntos de rotura en vez de los 6 esperados: $2m$ distal, microinversión distal, $2n$ proximal y uno compartido ($2m$ proximal, microinversión proximal y $2n$ distal). Por tanto, cada punto de rotura se está utilizando 1,5 veces, tal y como Bhuktar *et al.* (2008) promediaron para el género *Drosophila*.

Existen diversas explicaciones para la reutilización observada en los puntos de rotura. En humanos, se ha descrito que repeticiones de CCG y regiones ricas en AT (Schwartz *et al.* 2006), podrían ser la causa de la reutilización. Estas regiones frágiles (Cohen *et al.* 1996) no se han descrito en *Drosophila*, y no parecen ser la causa de la reutilización del punto de rotura de las inversiones $2m$ y $2n$. La hipótesis selectiva de Becker y Lenhard (2007), propone que los puntos de rotura se acumulan en regiones donde no se acumulan unidades de transcripción o de regulación. Parece difícil pensar que en el cromosoma 2 de *D. buzzatii* pudiera haber tan pocas regiones que pudieran ser interrumpidas. Finalmente, la reutilización de puntos de rotura se ha descrito por mecanismos mediados entre repeticiones (Stankiewicz y Lupski 2002). Los puntos de rotura se reutilizan por reorganizaciones mediadas por mecanismos de recombinación homóloga no alélica (NARH) (Gu *et al.*, 2008). Se ha propuesto que los TE tienden a acumularse en las regiones de los puntos de rotura (Cáceres *et al.* 2001; Peng *et al.* 2006) y estos actuarían como nuevo sustrato en el mecanismo de recombinación ectópica

generando nuevas reutilizaciones (González *et al.* 2007). Después del primer evento de reorganización, el segundo reutiliza secuencialmente un punto de rotura del anterior. El mecanismo de rotura más reparación errónea (NHEJ) no se modula por secuencias repetitivas y dos DSB no se esperan en la misma región por azar. Aunque aparentemente no existen dianas específicas para las roturas, cierta arquitectura del genoma (palíndromos, estructuras cruciformes...) pueden catalizar este mecanismo (Stankiewicz *et al.* 2003). Sin embargo, después de la rotura en estas regiones, esta arquitectura catalizadora se divide entre los puntos de rotura, por lo que no se ve fácilmente cómo será capaz de motivar un nuevo evento de rotura. Por otro lado, estos agentes si pueden estar conservados entre especies, por lo que los casos de recurrencia de los puntos de rotura, si se espera que se produzcan con este mecanismo (Ruiz-Herrera y Robinson 2007). En el trabajo de Ranz *et al.* (2007) se analizó molecularmente las recurrencias en distintas especies del grupo melanogaster. El mecanismo de las inversiones recurrentes se debía a rotura más NHEJ.

Claramente, las inversiones $2m$ y $2n$ reutilizan un punto de rotura. Podemos analizar de forma conjunta los mecanismos plausibles descritos para las inversiones $2m$ y $2n$ a la luz de las explicaciones para la reutilización de puntos de rotura.

4.1.4.2. Relación entre los mecanismos propuestos

En la Figura 4.3 se muestra un modelo de generación de las inversiones $2m$ y $2n$. Ambas inversiones se muestran simultáneas al no tener datos de la cronología. Sólo la microinversión sabemos es posterior a la inversión $2m$ al invertir la duplicación consecuencia de ésta. Del mismo modo, tampoco habría datos de la cronología entre la inversión $2n$ y la microinversión. No todos los mecanismos descritos para las inversiones $2m$ y $2n$ son consistentes entre si, por lo que debemos diferenciar cuando $2m$ sea anterior a $2n$ y viceversa.

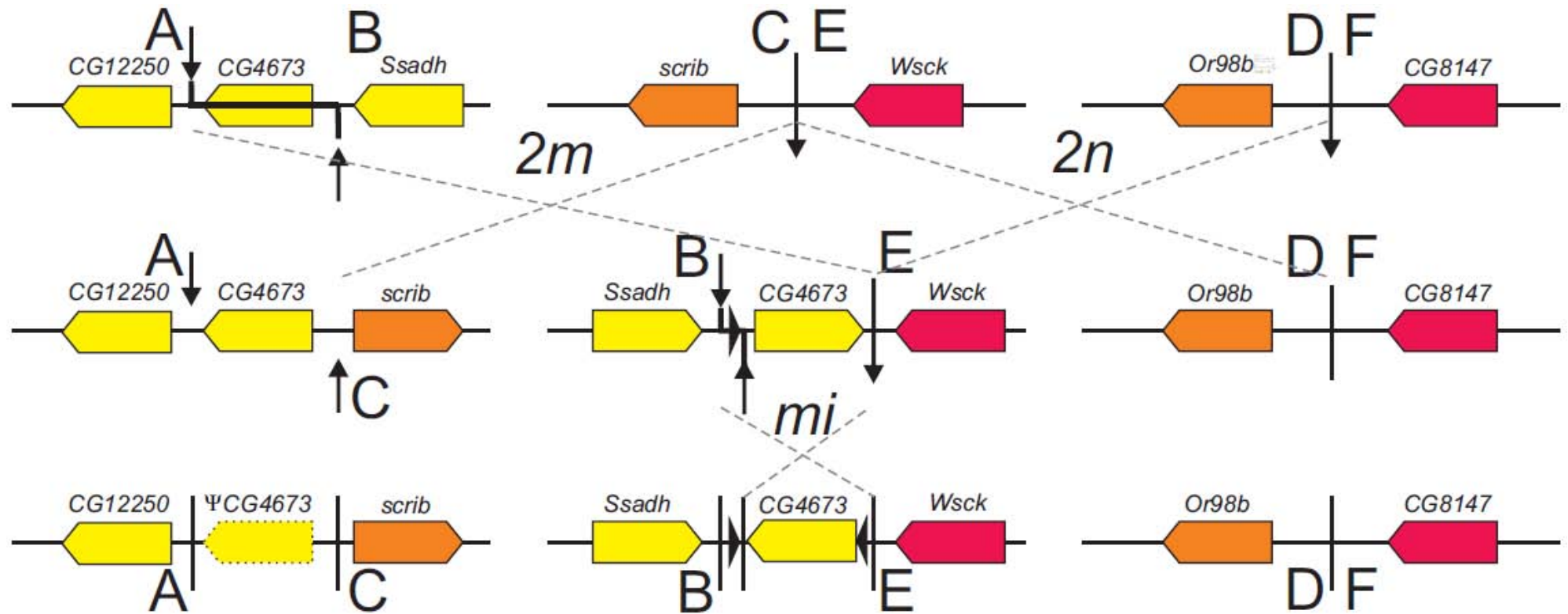


Figura 4.3: Mecanismo propuesto de generación del cromosoma $2mn$ de *D. buzzatii*. Se muestran las inversiones $2m$ y $2n$ y la microinversión para llegar desde AB-CD-EF (Primitivo I del grupo) a AC-BE-DF (*D. buzzatii*). Las inversiones $2m$ y $2n$ se muestran simultáneas al no conocer su cronología. La inversión $2m$ genera una duplicación (CG4673) debido a la rotura escalonada en la región del punto de rotura AB. La microinversión (*mi*) de la duplicación en BE y la pseudogenización del gen CG4673 en AC ocurren posteriormente. Los triángulos negros indican la duplicación asociada a la microinversión (*mi*) debido a la rotura escalonada del punto de rotura distal.

Inversión $2m$ anterior a $2n$

El origen de la inversión $2m$ es consistente con la inserción híbrida y NHEJ con rotura escalonada (véase sección 4.1.3). En segundo lugar, se produciría la inversión $2n$, por recombinación ectópica o por rotura escalonada. Ambas inversiones no pueden haberse producido por rotura más reparación, ya que dos roturas de doble cadena no se esperan en el mismo lugar por azar. Con esta sucesión de eventos, no podríamos explicar la reutilización completa observada. Del mismo modo, aunque la inversión $2m$ se hubiera producido por inserción híbrida, no se explicaría la coincidencia de un punto de rotura de doble cadena en el mismo lugar. Por tanto, sólo nos quedan como posibles mecanismos, que independientemente al mecanismo por el que se haya originado la inversión $2m$ (NHEJ o inserción híbrida), si la inversión $2m$ fuera anterior, sólo el mecanismo de recombinación ectópica podría explicar la inversión $2n$. La copia de *BuT5* en BD (punto de rotura intermedio hipotético) podría ser la que habría intervenido en la inserción híbrida o tratarse de una nueva copia secundaria que habría colonizado el punto de rotura proximal de la inversión $2m$ durante la fase de baja recombinación de los heterocariotipos (Sturtevant 1917; Langley *et al.* 1988; Ashburner 1989; Montgomery *et al.* 1991; Charlesworth 1996; Goldman y Lichtenstein 1996; Armengol *et al.* 2003). Esta copia recombinaría con otra copia de *BuT5* situada en la región del punto de rotura proximal de la inversión $2n$ (EF).

Inversión $2n$ anterior a $2m$

El origen de la inversión $2n$ es consistente con la recombinación ectópica o la reparación NHEJ con roturas de doble cadena (véase sección 4.1.3). Posteriormente, se produciría la inversión $2m$, por inserción híbrida o NHEJ con rotura escalonada. Si la inversión $2n$ se produjera mediante recombinación ectópica entre copias de una familia de TE, de nuevo, no habría explicación para la coincidencia con la inversión $2m$ si ésta se generara

por rotura mas reparación. Tampoco podría explicarse la reutilización si ambas inversiones se hubieran producido por rotura más reparación, ya que dos roturas de doble cadena no se esperan en el mismo lugar por azar. Por otro lado, una de las copias del TE que habría originado la inversión $2n$ sí podría estar implicada en el mecanismo de inserción híbrida de la inversión $2m$. Finalmente, también podría explicarse la reutilización observada si la inversión $2n$ se hubiera generado por rotura más reparación. Una copia de *BuT5* acumulada en el punto de rotura distal de la inversión $2n$ en la fase de baja recombinación, produciría por inserción híbrida la inversión $2m$.

Simultaneidad de las inversiones $2m$ y $2n$

Alternativamente, la reutilización completa observada, sería consistente con que ambas inversiones se produjeran simultáneamente, de forma sincronizada, mediante un único mecanismo que resulte en una única reorganización cromosómica formada de dos inversiones. En este caso, las tres roturas serían simultáneas generando dos fragmentos. La reparación errónea de ambos fragmentos produciría una inversión doble. La filogenia es consistente con esta opción, puesto que no se conocen especies que tengan sólo una de las inversiones.

La reparación de tres roturas simultáneas puede resolverse de cuatro formas distintas. La correcta reparación de ambos fragmentos, los dejaría en su posición original (no se observaría ninguna reorganización). En otros dos casos, sólo un fragmento se invertiría (uno se repararía correctamente y otro no y viceversa), observando una inversión simple. Por último, ambos fragmentos podrían ser mal reparados (caso que estaríamos observando en el estudio de las inversiones $2m$ y $2n$). Por tanto, sólo el 25% de los casos cuando se producen tres roturas simultáneas, producirán una falsa reutilización observable. Además, las reorganizaciones simultáneas se observarían en baja frecuencia porque es más probable que ocurran dos roturas al mismo tiempo, que tres. También hay que considerar que cuantas más roturas se produzcan,

más aumenta la posibilidad de efectos perjudiciales por interrupción de unidades de transcripción o de regulación. Todas las especies invertidas para una inversión, lo son para las tres ($2m$, $2n$ y la microinversión) estando en todos los casos fijadas. Si no estuvieran relacionadas, probablemente se podría esperar mayor heterogeneidad.

4.2. Inversión $2z^7$

4.2.1. Elementos transponibles anotados en los puntos de rotura

Distintos fragmentos de elementos transponibles (TE) han sido anotados en ambos puntos de rotura de la inversión $2z^7$ de *D. buzzatii* (véase Figura 3.9 en Resultados). En el punto de rotura PR, sólo se anotó un fragmento de un elemento tipo *Galileo*. Por el contrario, distintos elementos se encontraron en el punto de rotura QS. También en este punto de rotura se encontró un fragmento de elemento tipo *Galileo*, siendo éste el único elemento encontrado en ambos puntos de rotura de la inversión $2z^7$ de *D. buzzatii*. Los demás fragmentos anotados en el punto de rotura son de mayor tamaño. Además, estos elementos fueron anotados comparando la secuencia de *D. buzzatii* con los elementos caracterizados en *D. melanogaster* mientras que los fragmentos anotados tipo *Galileo* fueron comparados con copias caracterizadas en el mismo genoma de *D. buzzatii* (véase Tabla 3.11 en Resultados). Aun así, los fragmentos de elementos distintos de *Galileo* tienen mayor similitud con elementos de otros genomas que los fragmentos tipo *Galileo* con copias caracterizadas en el propio genoma de *D. buzzatii*. Por tanto, los TE tipo *Galileo* parece que están más degenerados y su inserción en este punto sea menos reciente que los demás.

La presencia de un TE en un punto de rotura, no lo implica necesariamente en el mecanismo que generó la reorganización. La colonización secundaria de elementos

transponibles en puntos de rotura de inversiones es habitual (Montgomery *et al.* 1991; Cáceres *et al.* 2001; Richards *et al.* 2005; Peng *et al.* 2006; Casals *et al.* 2006).

Las inserciones de los elementos transponibles con LTR se pueden datar mediante la comparación de sus LTR, idénticas en el momento de la inserción (Brookfield 1986; Kapitonov y Jurka 1996; Costas y Naveira 2000; Bowen y McDonald 2001). Ninguno de los TE anotados tiene LTR, por lo que no podemos datar las inserciones mediante la comparación de éstas. Aun así, la anidación de estos elementos, proporciona información acerca de su cronología de inserción (Cáceres *et al.* 2001; Casals 2006). Los elementos tipo *Helena* e *ISBu1* son colonizadores secundarios del punto de rotura proximal puesto que claramente están insertados anidados en el elemento tipo *CR1* (véase Figura 3.9 en Resultados). La inserción de elementos dentro de otros es habitual en *Drosophila* (Kaminker *et al.* 2002). Además, la duplicación de la diana de inserción (TSD) que produce el elemento tipo *Helena* es parte del elemento *CR1*. De todos modos, al no poder datar las inserciones de los demás TE anotados, no podemos saber si se produjeron antes o después de la inversión $2z^7$. Sin embargo, el estudio de la presencia de estos elementos en otras especies invertidas nos permitió saber cuales se insertaron antes del evento de reorganización. Los elementos anotados en el punto de rotura QS de *D. buzzatii* (*ISBu1*, tipo *CR1* y tipo *Helena*) no se encontraron en ninguna de las otras dos especies estudiadas (*D. koepferae* y *D. martensis*, véase Figura 3.9 y Tabla 3.11 en Resultados). La divergencia de *D. martensis* ocurrió antes de la separación entre los linajes de *D. buzzatii* y *D. koepferae* (véase Figura 1.3 en Introducción). Si estos elementos se hubieran insertado antes de la inversión, dos sucesos independientes deberían haber ocurrido para eliminarlos completamente. Una delección habría sucedido en el linaje de *D. koepferae* y el otra en el linaje de *D. martensis* (ambas especies invertidas y sin estos TE anotados). Un evento único de la colonización secundaria de estos elementos en el linaje *D. buzzatii* parece más plausible. En consecuencia, estos elementos no estarían implicados ni en el mecanismo ni en

el origen de la inversión. Por el contrario, como en *D. buzzatii*, en todos los puntos de rotura secuenciados de especies con la inversión (punto de rotura RS y QS de *D. martensis* y QS de *D. koepferae*) se encontraron fragmentos de elementos tipo *Galileo*. Por otro lado en el punto de rotura QS de las tres especies con la inversión (*D. buzzatii*, *D. koepferae* y *D. martensis*), ha sido anotado un fragmento de *BuT5*. En *D. martensis*, este elemento ha sido anotado anidado en un fragmento de elemento tipo *Galileo* por lo que, al menos en este genoma, también se trata de una colonización secundaria en el elemento tipo *Galileo*.

4.2.2. Mecanismo de generación de la inversión $2z^7$

El elemento tipo *Galileo* encontrado en los puntos de rotura de las especies con la inversión $2z^7$ (véase sección anterior), podría estar relacionado con el origen de ésta. Este elemento se ha anotado dentro de las juntas entre los bloques conservados con *D. mojavensis*. Los elementos transponibles han sido relacionados con el origen de las inversiones (Langley *et al.* 1988; Ashburner 1989; Charlesworth 1996; Goldman y Lichten 1996). Los elementos *hobo* (Lyttle y Haymer 1992), el elemento *P* (Regner *et al.* 1996) o *Penelope* y *Ulysses* en *D. virilis* (Evgen'ev *et al.* 2000) fueron asociados a los puntos de rotura de distintas inversiones naturales. En inversiones experimentales producidas en el laboratorio, los elementos *copia* y *roo* también fueron asociados a los puntos de rotura (Montgomery *et al.* 1991). En el estudio de las inversiones *In(4)a* en *D. americana* (Evans *et al.* 2007), *Arrowhead* en *D. pseudoobscura* (Richards *et al.* 2005) y $2j$, $2q^7$ y $2z^3$ en *D. buzzatii* (Cáceres *et al.* 1999; Casals *et al.* 2003 y Delprat *et al.* 2009, respectivamente), fueron descritas repeticiones homólogas en las juntas de ambos puntos de rotura. Concretamente, en estos trabajos se concluye que la recombinación ectópica entre dos copias del elemento *Galileo* originó dichas inversiones (véase Tabla 1.1 en Introducción). La implicación de *Galileo* como sustrato de la recombinación ectópica para

explicar el origen de la inversión $2z^7$ es plenamente consistente con las inversiones anteriormente descritas en *D. buzzatii*. La inserción del elemento *BuT5* parece ser secundaria a *Galileo*. La copia de este elemento en el punto de rotura QS en *D. martensis* se ha localizado dentro de una anotación de un elemento tipo *Galileo* (véase Figura 3.9 en Resultados). Si la inserción del elemento *BuT5* se hubiera producido antes de la inversión, disminuiría la región de similitud de *Galileo* y, por tanto, reduciría la probabilidad que éste fuera sustrato de la recombinación ectópica. Por otro lado, tres eventos de inserción de *BuT5* en el mismo sitio en tres linajes son poco probables por lo que, probablemente, habrá colonizado el punto de rotura poco después de la aparición de la inversión y antes de la divergencia del linaje de *D. martensis*.

En Delprat *et al.* (2009), las copias de *Galileo* encontradas en ambos puntos de rotura eran quiméricas y con las TSD intercambiadas tal y como se espera en el mecanismo de recombinación ectópica. La detección de copias quiméricas depende de las diferencias nucleotídicas entre las copias parentales. Si las copias parentales son muy similares o no disponemos de copias intactas no-recombinantes para comparar, no se podrá detectar el carácter quimérico de las copias resultantes. En el caso de la inversión $2z^7$, no se pudieron distinguir copias enteras y mucho menos quiméricas probablemente debido a la edad de la reorganización.

Finalmente, en la región del punto de rotura RS en *D. mojavensis* (véase Figura 3.9 en Resultados) también se anotó una copia de un elemento tipo *Galileo*. Este fragmento se anotó usando una copia del elemento caracterizada en *D. mojavensis*. El fragmento anotado en *D. mojavensis* parece estar menos degradado. Por un lado, el fragmento anotado es mucho mayor, por otro, la identidad de esta copia con la usada para anotarlo es mucho mayor (95,9%) que la identidad de los fragmentos anotados en *D. buzzatii* con las copias caracterizadas en *D. buzzatii* (cerca al 70%). La presencia de un elemento de la misma familia y en el mismo

lugar que los encontrados en los genomas invertidos podría explicarse de distintas formas. Por un lado, esta copia anotada en el punto de rotura RS de *D. mojavensis* podría ser una de las copias parentales de la recombinación ectópica que produjo la inversión $2z^7$. Por otro lado, el tamaño del fragmento anotado en *D. mojavensis* y su identidad nucleotídica con otras copias indicaría que su inserción fue posterior a la de la copia en los genomas invertidos. Por tanto, esta copia podría haberse insertado en el punto de rotura RS del linaje no invertido posteriormente a la inversión $2z^7$. En este caso, al tratarse del mismo tipo de elemento hablaríamos de recurrencia en el sitio de inserción para los elementos de la familia *Galileo*. Una copia se insertó en la región del punto de rotura RS en el linaje de *D. mojavensis* y la otra, en el linaje de la especie con la inversión $2z^7$. Estos sitios y motivos preferenciales de inserción han sido descritos para algunos elementos (Zhang y Spradling 1994; Vigdal *et al.* 2002; Linheiro y Bergman 2008) y retrovirus (Wu y Burgess 2004).

4.2.3. Posición y consecuencias de los puntos de rotura

En *D. mojavensis*, los puntos de rotura distal (PQ) y proximal (RS) de la inversión $2z^7$ se encuentran en las regiones intergénicas entre los genes *CG14645* y *CG14290* y entre los genes *CSN5* y *Dlh*, respectivamente (véase Figura 3.9 en Resultados). En *D. buzzatii* el punto de rotura proximal (QS) se encuentra en la región intergénica entre 306 pb aguas arriba del codón inicio del gen *CG14290* (Q) y *Dlh* (S). Ambos genes están anotados en el BAC 40C11 (Negre *et al.* 2005; Delprat *et al.* 2009). El gen *Dlh* es un gen de cuatro exones y esta conservado con *D. mojavensis* (Delprat *et al.* 2009). La proteína putativa del gen *CG14290* es idéntica a la de *D. mojavensis* en un 93.5%. Esta conservación indica que no se está interrumpiendo ninguna unidad de transcripción y ambos genes siguen siendo funcionales.

Por otro lado el punto de rotura distal (PR) en *D. buzzatii* se encontro en un fragmento

a 484 pb aguas arriba del codón de inicio del gen *CG14645* (P) y a 104 pb aguas arriba del gen *CSN5* (R) anotado como region intergénica en *D. mojavensis*. En este caso, no se tiene la secuencia de entera de estos genes, sólo el primer exón de cada uno. La conservación de los codones de inicio y la alta conservación nucleotídica del primer gen con *D. mojavensis*, sugieren que tampoco estos genes han dejado de ser funcionales.

Los cuatro genes implicados han conservado el marco de lectura. Si las regiones promotoras de estos genes se interrumpieran en los genomas invertidos, se esperaría encontrar cierta pseudogenización al haber dejado de ser regulados y, por consiguiente, funcionales. Por tanto, ambos puntos de rotura de la inversión $2z^7$ parecen haberse producido en regiones intergénicas que no han alterado la transcripción de los genes colindantes. Sin embargo, no se puede descartar que los genes afectados se estén regulando de forma distinta que en los genomas no invertidos al no tener datos de su expresión.

4.3. Comparación de los puntos de rotura de las inversiones $2m$, $2n$ y $2z^7$ en el género *Drosophila*

Las inversiones $2m$, $2n$ y $2z^7$ sólo se han descrito en el complejo *buzzatii*. No se conocen otras especies fuera del complejo con las mismas inversiones. Sin embargo, la comparación del orden de los genes en las regiones de los puntos de rotura en los doce genomas de *Drosophila* secuenciados ha permitido observar otros puntos de rotura de otras reorganizaciones en las mismas regiones.

El orden de las regiones de los puntos de rotura PQ y RS corresponde al genoma no invertido de *D. mojavensis* para la inversión $2z^7$ (véase Figura 3.10 y Tabla 3.12 en Resultados). Las especies secuenciadas del subgénero *Drosophila* (*D. mojavensis*, *D. virilis* y *D. grimshawi*) conservan esta ordenación mientras que en las especies secuenciadas del

subgénero *Sophophora* las regiones PQ y RS están interrumpidas ((P)-(Q) y (R)-(S), respectivamente). Los genes en la ordenación de (P)-(Q) del subgénero *Sophophora* se conserva en el grupo externo secuenciado más cercano (*A. gambiae*). Los genes colindantes a las regiones P y Q en el grupo externo son los mismos que en el subgénero *Sophophora*. Por tanto, la ordenación hipotética del ancestro de *Drosophila* en este punto de rotura es la descrita en el subgénero *Sophophora*. En el caso de RS, no se ha podido determinar cuál era la ordenación de ningún grupo externo y por tanto, la del ancestro. La región PQ por tanto, no estaba conservada en el ancestro, por lo que esta región estuvo implicada en una rotura anterior. Por tanto, la rotura distal (PQ) de la inversión $2z^7$, estaría reutilizando un punto de rotura anterior ocurrido en el linaje al subgénero *Drosophila*. La reorganización que anteriormente implicó una rotura en la región RS podría haberse producido en las ramas de los subgéneros *Sophophora* o *Drosophila*. Si esta reorganización se hubiera producido en el subgénero *Drosophila* (como en el caso de PQ), el punto de rotura RS de la inversión $2z^7$ también estaría reutilizando un punto de rotura anterior.

En cuanto a las inversiones $2m$ y $2n$, observamos que la ordenación de los genes en el ancestro del género también sería AB, CD y EF (véase Figura 3.10 y Tabla 3.12 en Resultados). La región AB sólo se ha interrumpido debido a la inversión $2m$. Sin embargo, las regiones de los puntos de rotura CD y EF se han interrumpido a lo largo de la evolución del género. La región CD estaría implicada aparte de las inversiones $2m$ y $2n$, en dos roturas independientes el subgénero *Sophophora*, una en el linaje del grupo *melanogaster* y otra en el linaje del grupo *willistoni*. Estas dos regiones parecen ser recurrentes en la rotura al estar implicadas en reorganizaciones independientes en distintas ramas. La región EF también es recurrente al estar involucrada en una reorganización en el linaje del grupo *melanogaster*.

Por tanto, en las regiones de los puntos de rotura de las inversiones $2m$, $2n$ y $2z^7$ han sucedido otros 5 puntos de rotura en las especies secuenciadas del género *Drosophila*. Para

estimar si estas roturas se esperan por azar o hay un exceso de roturas en estas regiones, hemos calculado la probabilidad de encontrar el número de roturas observadas. Bhutkar *et al.* (2008), en la comparación del los genomas de *D. mojavensis* y *D. melanogaster* describieron 324 roturas en el elemento E de Muller. Según Tamura *et al.* (2004), la divergencia de ambas especies ocurrió hace 62,9 ma, lo que significa que el tiempo evolutivo acumulado entre *D. mojavensis* y *D. melanogaster* es de 125,8 ma. Podemos estimar que, considerando que el tiempo evolutivo acumulado en las 12 especies secuenciadas de *Drosophila* es de 402 ma (Tamura *et al.* 2004; Negre *et al.* 2007), ha habido ~1035,36 roturas ($324 \times 402 / 125,8$) en el elemento E de Muller en estas especies suponiendo las mismas tasas de rotura en todas las ramas.

Con estas roturas esperadas en el elemento E de Muller, podemos calcular el número de roturas esperadas en las regiones de los puntos de rotura de las inversiones $2m/2n$ y $2z^7$. El número de roturas esperadas en una región cromosómica es $N_R = N \cdot l$, donde N es el número total de roturas en el cromosoma y l es la longitud relativa de la región (González *et al.* 2000). Para calcular la longitud relativa se ha considerado el tamaño total de las regiones de puntos de rotura de las tres inversiones del genoma de *D. mojavensis* relativo a las 34.148 Kb de longitud de su cromosoma 2 (Bhutkar *et al.* 2008). Para la región del punto de rotura AB, se ha considerado las regiones intergénicas aguas arriba (1.339 pb) y aguas abajo (337 pb) del gen *CG4673*. El tamaño total de las regiones de los puntos de rotura en *D. mojavensis* es de 23,396 Kb por lo que la longitud relativa es $l = 23,396 / 34.148 = 6,85 \times 10^{-4}$. Por otro lado, también hemos calculado la longitud relativa considerando que los puntos de rotura de las inversiones no detrimentales sólo ocurran en las regiones intergénicas. Si consideramos que la estima de porción intergénica es del 38,2% (Crosby *et al.* 2007), la longitud relativa intergénica es de $l(i) = 23,396 / 13.044 = 1,79 \times 10^{-3}$. En consecuencia el número de roturas esperadas en el total de las regiones de los puntos de rotura de estas inversiones es $N_R = 1035,36 \times 6,85 \times 10^{-4} = 0,71$ y

$N_R(i) = 1035,36 \times 1,79 \times 10^{-3} = 1,8$. En la Tabla 4.3 se muestra el número de roturas observado en las especies secuenciadas del género *Drosophila* y *D. buzzatii*. En *D. buzzatii* hemos considerado las roturas observadas diferenciando dos clases de roturas observadas. Por un lado, hemos considerado roturas independientes (2 roturas por cada inversión) y, por otro, considerando reutilización. En este caso, hemos considerado los eventos de reutilización como una única rotura. La rotura escalonada de la región AB se ha considerado como una sola rotura ya que la rotura de una sola hebra, no produciría una rotura observable.

Tabla 4.3: Roturas esperadas y observadas para las regiones de los puntos de rotura de las inversiones $2m$, $2n$ (AB, CD y EF) y $2z^7$ (PQ y RS). Se muestran las longitudes relativas de las regiones de las inversiones $2m$ y $2n$ por un lado y la inversión $2z^7$ por otro. El tamaño de las regiones (L_R) y tamaños relativos (l) se han tomado considerando *D. mojavensis*. También se muestra la longitud relativa y roturas esperadas considerando sólo la porción intergénica del cromosoma 2 de *D. mojavensis*. Las roturas esperadas se han calculado usando las tasas de rotura por ma por Mb tomadas de Bhutkar *et al.* (2008) teniendo en cuenta el tamaño del genoma de *D. mojavensis* y *D. melanogaster*. Se ha considerado 402 ma de evolución en el género *Drosophila* y 10 ma de evolución de *D. buzzatii* desde el Primitivo I del grupo (Tamura *et al.* 2004; Negre *et al.* 2007). En esta Tabla también se muestran las roturas observadas (Obs.) en las especies secuenciadas del género y en *D. buzzatii*. Las roturas observadas en *D. buzzatii* se muestran considerando roturas independientes (Ind.) con dos puntos de rotura por cada inversión y teniendo en cuenta reutilización (Reu.) considerando una única rotura si se comparte. De este modo, el punto de rotura CD alberga 4 roturas considerando roturas independientes ($2m$ proximal + $2n$ distal + ambos de la micro inversión) y sólo 2 considerando reutilización ($2m$ proximal/ $2n$ distal/microinversión proximal + microinversión distal).

Región	L_R (pb)	l	N_R	$l(i)$	$N_R(i)$	Obs.	Obs. (<i>D. buzzatii</i>)	
						(género)	Ind.	Reu.
$2m+2n$	20.754	$6,1 \times 10^{-4}$		$1,59 \times 10^{-3}$		3	7	5
$2z^7$	2.642	$7,73 \times 10^{-5}$		$2,04 \times 10^{-4}$		2	2	2
TOTAL	23.396	$6,85 \times 10^{-4}$	0,71	$1,79 \times 10^{-3}$	1,8	5	9	7

Por otro lado podemos calcular la probabilidad de que ocurra n roturas en el conjunto de las regiones de los puntos de rotura mediante el desarrollo de la expresión binomial $[(1-l) + l]^N$ donde l corresponde a la longitud relativa y N al número total de roturas en el cromosoma (González *et al.* 2000). En la Tabla 4.4, se muestra las probabilidades de que sucedan n roturas en el total de las regiones de los puntos de rotura considerando la longitud relativa respecto el elemento E de Muller y con la longitud relativa sólo considerando las regiones intergénicas.

Tabla 4.4: Desarrollo de la expresión binomial $[(1-l) + l]^N$ para determinar la probabilidad de encontrar ninguna, una, dos, tres o cuatro roturas en las regiones de los puntos de rotura. Se ha considerado la longitud relativa (l) y longitud relativa intergénica ($l(i)$) de las regiones de los puntos de rotura de las inversiones $2m$, $2n$ (AB, CD y EF) y $2z^7$ (PQ y RS) en *D. mojavensis* (véase Tabla 4.3). La probabilidad de encontrar 5 roturas o más en estas regiones corresponde a $1 -$ la probabilidad de encontrar ninguna, una, dos, tres o cuatro roturas.

Roturas	Expresión binomial $[(1-l) + l]^N$	$l=6,85 \times 10^{-4}$	$l(i)=1,80 \times 10^{-3}$
n=0	$1 \times (1-l)^{1035} \times l$	0,4920	0,1549
n=1	$1035 \times (1-l)^{1034} \times l^1$	0,3490	0,2891
n=2	$535095 \times (1-l)^{1033} \times l^2$	0,1237	0,2696
n=3	$184251045 \times (1-l)^{1032} \times l^3$	0,0292	0,1674
n=4	$47536769610 \times (1-l)^{1031} \times l^4$	0,0052	0,0779
n≥5	$1 - [P(n=0) + P(n=1) + P(n=2) + P(n=3) + P(n=4)]$	0,0009	0,0411

Cuando consideramos la longitud relativa del conjunto de las regiones de los puntos de rotura de las inversiones $2m$, $2n$ y $2z^7$ respecto el cromosoma 2, observamos que la probabilidad de encontrar 5 roturas o más es significativamente reducida ($P=0,0009$). Sin embargo, las regiones seleccionadas corresponden a fracción intergénica. Cuando consideramos la longitud relativa respecto a la porción intergénica del cromosoma 2, aunque

significativa, la probabilidad de encontrar 5 roturas o más es mayor que en caso anterior ($P=0,0411$). Esta significación indica que por azar no esperamos tantas roturas en el conjunto de las regiones analizadas. Por otro lado, aunque la elección de estas regiones no es al azar porque contienen los puntos de rotura de las inversiones que estamos estudiando, observamos que también están interrumpidas en *D. buzzatii*. Las regiones de los puntos de rotura CD y EF se observan interrumpidas en *D. buzzatii* y recurrentemente en otros puntos del género. Al menos el punto de rotura PQ de la inversión $2z^7$ está reutilizando una region anteriormente interrumpida en el género *Drosophila*.

El punto de rotura PQ está involucrado en la inversión $2z^7$ originada por el mecanismo de recombinación ectópica. Sin embargo los puntos de rotura CD y DF de las inversiones $2m$, $2n$ están involucrados en inversiones mediadas por un mecanismo distinto de la recombinación ectópica. La asociación del mecanismo con la reutilización de los puntos de rotura, ya ha sido descrita en trabajos anteriores (Stankiewicz y Lupski 2002; Gu *et al.* 2008). Cuando una inversión se produce por primera vez, independientemente del mecanismo, estos puntos de rotura pueden convertirse en un *hotspot* (Cáceres *et al.* 2001; Peng *et al.* 2006) debido a la baja recombinación en los heterocariotipos (Montgomery *et al.* 1987; Eanes *et al.* 1992). Estas nuevas inserciones podrían generar nuevas reorganizaciones reutilizando el punto de rotura. Por tanto, el mecanismo de la segunda reorganización implica TE (ya sea por recombinación ectópica o inserción híbrida). El mecanismo de rotura mas NHEJ no explica la reutilización de un punto de rotura. En cambio, la susceptibilidad o la arquitectura específica del genoma que cataliza la rotura pueden ser conservadas entre linajes. En estas regiones conservadas, los puntos de rotura evolutivos tenderían a acumularse (Murphy *et al.* 2005). Ranz *et al.* (2007), concluyeron que el mecanismo de rotura mas reparación errónea genera las reorganizaciones con puntos de rotura recurrentes en el grupo *melanogaster*.

De este modo, los puntos de rotura recurrentes estarán implicados en reorganizaciones

NHEJ. Por el contrario, los reutilizados estarán implicados en mecanismos de NAHR. Los resultados observados en las inversiones de este trabajo, son consistentes con esta conclusión. La inversión $2z^7$ reutiliza al menos un punto de rotura (PQ) y el mecanismo más plausible que la ha originado es la recombinación ectópica.

Por otro lado, hasta cinco combinaciones de mecanismos fueron descritas como consistentes con el origen de las inversiones $2m$ y $2n$ (véase sección 4.1.4.2). Si la inversión $2m$ se hubiera originado primero, podía hacerlo por NHEJ (con rotura escalonada) o por inserción híbrida. En cualquier caso, la inversión $2n$ posterior, sólo podría originarse por recombinación ectópica. Si la inversión $2n$ se hubiera generado primero, esta podría hacerlo por recombinación ectópica o mediante NHEJ (con rotura de doble cadena). En cualquier caso, la inversión $2m$ posterior sólo lo haría mediante inserción híbrida. Finalmente, podría ocurrir la simultaneidad de ambas inversiones.

Los puntos de rotura CD y EF son recurrentes a lo largo del género *Drosophila* (véase 3.10 en Resultados). Por azar, no esperamos que la inversión $2m$ se genere primero por inserción híbrida si el punto de rotura CD es recurrente en el género. Del mismo modo, tampoco no esperamos que la inversión $2n$ se genere primero por recombinación ectópica. De este modo, sólo tres combinaciones son aún consistentes: i) $2m$ mediante NHEJ más inversión $2n$ por recombinación ectópica, ii) inversión $2n$ por NHEJ más inversión $2m$ por inserción híbrida y iii) la simultaneidad de ambas inversiones.

4.4. Comparación de las inversiones caracterizadas en *D. buzzatii*

La selección sobre los efectos de las inversiones en los cromosomas determina la frecuencia de éstas (Mettler *et al.* 1977; Stalker 1980). Por tanto, la frecuencia de una inversión dependerá de la adaptación de los efectos directos e indirectos de ésta y sus puntos

de rotura. A priori, la frecuencia de las inversiones no está relacionada con el mecanismo que la origina. La retención de combinaciones alélicas favorables (epistasia) en condiciones de supresión de la recombinación (Dobzhansky 1970; Charlesworth 1974) es independiente al mecanismo que origina una inversión. Los efectos directos en los cromosomas que producen cambios en la expresión génica como el efecto mutacional de los puntos de rotura (Sperlich 1963) o el efecto de posición por variegación (Spofford 1976; Eissenberg 1989; Henikoff 1990; Weiler y Wakimoto 1995) tampoco dependen del mecanismo que origina una inversión.

Sin embargo, pueden existir efectos de posición distintos de la localización del punto de rotura. Algunos efectos directos en los cromosomas no dependen del mecanismo *per se*, aunque sí de las causas y las consecuencias del mecanismo. Las duplicaciones asociadas a los mecanismos de inserción híbrida y rotura escalonada pueden incluir genes y dotarlos de nuevas funciones o subfuncionalizar las antiguas. También los TE que han intervenido en un mecanismo de recombinación ectópica pueden interrumpir o generar nuevos inicios de transcripción así como afectar sistemas de regulación.

4.4.1. Relación entre el mecanismo de inversión y el éxito evolutivo

Esperamos que, la frecuencia de la inversión dependa del efecto, no del mecanismo que la genera. No obstante, en *D. buzzatii* se observa cierto sesgo en la frecuencia de las inversiones fijadas en función del mecanismo que las ha generado. Las tres inversiones polimórficas $2z^3$ (Delprat *et al.* 2009), $2q^7$ (Casals *et al.* 2003) y $2j$ (Cáceres *et al.* 1999) y la fijada $2z^7$ están relacionadas con el mecanismo de recombinación ectópica. La otra inversión fijada no está relacionada con TE: $5g$ (Prazeres *et al.* 2009). La inversión $2m$, originada mediante rotura más NHEJ o inserción híbrida, tampoco comparte mecanismo con las inversiones polimórficas. Finalmente, sólo si la inversión $2n$ se originó por recombinación

ectópica, equilibraría la relación entre mecanismo y frecuencia. Aunque de todos modos, todas las inversiones polimórficas se habrían originado por recombinación. Si la inversión $2n$ se originara por NHEJ, acentuaría el desequilibrio observado, ya que claramente habría un mecanismo para las polimórficas que no sería frecuente en las fijadas. Sin embargo, una inversión polimórfica puede tener efecto positivo local. Esta inversión no se fijaría y se establecería en clinas (Dobzhansky 1971; Prevosti *et al.* 1988). La inversión polimórfica $2j$ muestra adaptación local en latitudes altas (Puig *et al.* 2004). De todos modos sorprende no encontrar en *D. buzzatii* inversiones con adaptación local generadas por un mecanismo distinto a la recombinación ectópica. En *D. melanogaster*, la inversión polimórfica *In(2L)t* se encuentra en mayor frecuencia en hábitats cálidos sin haberse originado por recombinación ectópica (Lemeunier y Aulard 1992; Andolfatto *et al.* 1999).

Para la observación sesgada en *D. buzzatii*, habría dos posibles explicaciones. La primera implica que exista una relación entre el mecanismo de inversión con su frecuencia. La segunda explicaría el sesgo observado debido a la frecuencia de los agentes que generan las inversiones.

En la primera explicación, la frecuencia de la inversión se asociaría a los lugares de las roturas de los distintos mecanismos. En general, las inversiones originadas por recombinación ectópica podrían tener efecto neutro con más frecuencia que las producidas por NHEJ. Si el punto de rotura se produce en una región donde la selección ya habría permitido la inserción de un TE, la posibilidad de que el efecto de la reorganización implicando dicho elemento siga siendo neutro, aumenta. Por contra, la rotura de doble cadena puede producirse en cualquier lugar del genoma generando efectos más heterogéneos distintos de la neutralidad. Aquellas reorganizaciones con efecto negativo desaparecerían, pero si las nuevas combinaciones fueran positivas, se fijarían con mayor facilidad.

En segundo lugar, el sesgo observado podría explicarse si la cantidad de reorganizaciones generadas con mecanismos de recombinación ectópica fuera proporcional a la presencia de TE. El mecanismo de rotura más NHEJ representaría la tasa basal esperada de reorganizaciones en un genoma. Esta tasa de reorganizaciones se modularía con la frecuencia y localización de los TE capaces de servir de agentes de la recombinación. Las inversiones $2m$, $2n$ y $5g$ podrían pertenecer a la inestabilidad basal, mientras que las tres inversiones polimórficos y $2z^7$ se originaron debido a la inestabilidad producida después de la invasión del elemento *Galileo* y su replicación en el genoma. Las observaciones en el genoma de *D. buzzatii* son consistentes con que la actividad de los TE sería la razón de las altas tasas de inversión en los genomas (Yamaguchi y Mukai 1974). Este argumento también podría explicar la diferencia observada en el número de inversiones entre especies y entre y dentro de cromosomas. Por otro lado, esta explicación sólo serviría para explicar el sesgo en las inversiones originadas por mecanismos mediados por TE. Según esta interpretación, esperaríamos una distribución aleatoria y homogénea de las inversiones originadas por la tasa basal. Sin embargo, el cromosoma 2 contiene el 70% de las inversiones descritas en el grupo (Wasserman 1992) y no presenta un exceso en el número de TE (Casals *et al.* 2005 y 2006). Aun así, esta explicación podría ser el motivo de las diferencias observada en el número de inversiones entre especies. Para rechazar o aceptar esta conjetura, se debería analizar el mecanismo de todas las inversiones de otras especies como nuestro grupo de trabajo ha realizado en *D. buzzatii*.

4.4.2. TE en los puntos de rotura de las inversiones de *D. buzzatii*

Independientemente al mecanismo que las ha generado, en todas las inversiones analizadas en *D. buzzatii* se han encontrado TE en ambos puntos de rotura.

4.4.2.1. Acumulación de TE en el punto de rotura proximal

Las inversiones originadas por recombinación ectópica $2z^3$ (Delprat *et al.* 2009), $2q^7$ (Casals *et al.* 2003) y $2j$ (Cáceres *et al.* 1999) parecen acumular mayor cantidad de TE en las juntas de los bloques conservados que las demás inversiones analizadas en *D. buzzatii*, la inversión $5g$ (Prazeres *et al.* 2009) y $2z^7$, $2m$ y $2n$ (en este trabajo). En la inversión $5g$ (Prazeres *et al.* 2009), ningún TE fue anotado en las juntas de los bloques conservados y sólo una copia de los elementos *Helitron* y *TART* fueron descritas en las regiones de los puntos de rotura distal y proximal, respectivamente. *BuT5* es el único TE presente en ambas juntas de la inversión $2m$, mientras que en la inversión $2n$, aparte de la copia de *BuT5*, se anotó un elemento *ISBu* y otro *transib* también en la junta proximal (DF). La diferencia en la cantidad de TE entre las distintas inversiones descritas en *D. buzzatii* podría deberse a la antigüedad de las inversiones. Dos factores contribuyen a la rápida desaparición de los TE. Singh y Petrov (2004) calcularon una tasa de desintegración cercana a la mitad de la fracción no codificante en 12 ma. De este modo, después de algún tiempo, no esperamos encontrar trazas de los TE en los puntos de rotura de las inversiones más antiguas. En segundo lugar, la selección actúa en contra de las anomalías cromosómicas causadas por recombinación ectópica que limita la propagación de los TE. Por tanto, la recombinación ectópica es un mecanismo selectivo contra las inserciones de TE (Petrov *et al.* 2003).

Por otro lado, los puntos de rotura proximales de las inversiones polimórficas ($2z^3$, $2q^7$ y $2j$) y la inversión $2n$ en *D. buzzatii*, parecen acumular mayor cantidad de TE. La región del punto de rotura proximal de la inversión $2z^7$ en *D. buzzatii* también parece acumular mayor cantidad de TE aunque no el de las otras especies invertidas analizadas *D. martensis* y *D. koepferae* (véase Figura 3.9 en Resultados). Dos hipótesis podrían explicar este efecto. Por un lado, que los puntos de rotura proximales sean más cercanos al centrómero y, por otro, a la

protección de los puntos de rotura por el solapamiento con otras inversiones. La primera explicación consiste en que la recombinación se reduce en regiones centroméricas y los TE tienden a acumularse en ellas (Tonzetich *et al.* 1988; Krimbas y Powell 1992; Casals *et al.* 2006). La reducción de la recombinación podría ser relativamente gradual desde el centrómero a lo largo de los cromosomas hasta su porción más distal. La degradación de los TE disminuiría en porciones proximales cercanas a los centrómeros y aumentaría en las porciones distales con mayor tasa de recombinación. De todos modos, esta observación debería ser generalizada, en cambio los elementos *BS* y *CR1* en la inversión *3R(8)* (Ranz *et al.* 2007) y *Odysseus* en la inversión *2Rd'* (Mathiopoulos *et al.* 1998) fueron anotados en el punto de rotura distal de la inversiones y ninguno en el proximal.

Por otro lado, este efecto observado, podría deberse a que una segunda inversión podría incluir el punto de rotura de una inversión anterior. Un punto de rotura anidado en una segunda inversión lo incluiría en una zona de baja recombinación (Navarro *et al.* 1997; Bartolomé *et al.* 2002; Armengol *et al.* 2003). Los TE acumulados (Langley *et al.* 1988; Ashburner 1989; Montgomery *et al.* 1991; Charlesworth 1996; Goldman y Lichten 1996; Cáceres *et al.* 2001; Casals *et al.* 2006) en este punto de rotura anidado no degradarían normalmente. Si sólo el punto de rotura proximal queda anidado en la segunda inversión, éste tendría mayor cantidad de TE que el punto de rotura distal de la misma inversión. Esta explicación es consistente con la cantidad similar de los puntos de rotura de la inversión *5g*, puesto que ninguna inversión posterior se ha producido en el cromosoma 5 en *D. buzzatii*. Por otro lado, la consecución de las inversiones en el cromosoma 2 de *D. buzzatii* no explicaría la mayor cantidad de TE en los puntos de rotura proximales. La cepa estándar (*st-1*) de *D. buzzatii* incluye las inversiones *2m*, *2n* y $2z^7$ (Figura 4.4). Sólo el punto de rotura BE (inversiones *2m* y *2n*) queda incluido en la inversión $2z^7$, por lo que no se explicaría la mayor cantidad de TE observada en los puntos de rotura DF (inversión *2n*) ni RS (inversión $2z^7$). La

inversión $2j$ se ha descrito sobre la ordenación $st-1$ y las inversiones $2z^3$ y $2q^7$ se han descrito sobre la ordenación $2j$, por lo que ninguna inversión solapa las inversiones polimórficas en las cepas $2jz^3$ y $2jq^7$ y tampoco quedaría justificada con esta explicación la mayor cantidad de TE encontrados en el punto de rotura proximal de éstas. Sólo en las cepas $2jq^7$ y $2jz^3$, los puntos de rotura de las inversiones $2n$, $2z^7$ y $2j$ quedarían solapados con otras inversiones. El presente trabajo se ha realizado con la cepa estándar por lo que el solapamiento de las inversiones no justifica la mayor cantidad de TE observados en los puntos de rotura proximales en las inversiones $2z^7$ y $2n$.

No parece que ninguna de los dos hipótesis propuestas puedan explicar satisfactoriamente la mayor observación de TE en los puntos de rotura proximales de las inversiones del cromosoma 2 de *D. buzzatii*.

4.4.2.2. Elemento BuT5

En la sección 4.1.3.2 se destacó la relación observada entre los elementos tipo *Galileo* y los puntos de rotura y *BuT5* (Casals *et al.* 2006). El elemento *BuT5* se ha encontrado en todas las inversiones descritas en el cromosoma 2 de *D. buzzatii*. En las inversiones $2j$ y $2z^7$, *BuT5* sólo se ha encontrado en el punto de rotura proximal mientras que en $2z^3$ sólo se ha encontrado en el distal. En la inversión $2q^7$ se encontraron copias de *BuT5* en ambos puntos de rotura. En las inversiones fijadas $2m$ y $2n$ se ha anotado *BuT5* en todos sus puntos de rotura. En las inversiones polimórficas y $2z^7$ (todas generadas por mecanismo de recombinación ectópica), *BuT5* se ha encontrado anotado junto a la presencia de elementos tipo *Galileo* (agente de la recombinación ectópica en todas ellas). Sólo este elemento es más frecuente en estas inversiones que el propio *BuT5* (Delprat *et al.* 2009). En cambio, en las inversiones $2m$ y $2n$, prácticamente sólo se ha encontrado este elemento y, además, no hay rastro de ningún

elemento tipo *Galileo*. En la inversión fijada *5g* en el cromosoma 5 originada por un mecanismo distinto a la recombinación ectópica, no se encontraron trazas ni de *Galileo* ni de *BuT5*.

No podemos descartar la hipótesis que *BuT5* no se observe por azar en las inversiones mediadas por mecanismos de recombinación ectópica entre copias de *Galileo*. La acumulación de *BuT5* en estos puntos de rotura podría no sólo deberse a la baja tasa de recombinación de los heterocariotipos (Navarro *et al.* 1997), sino también a la preferencia de inserción de éstos.

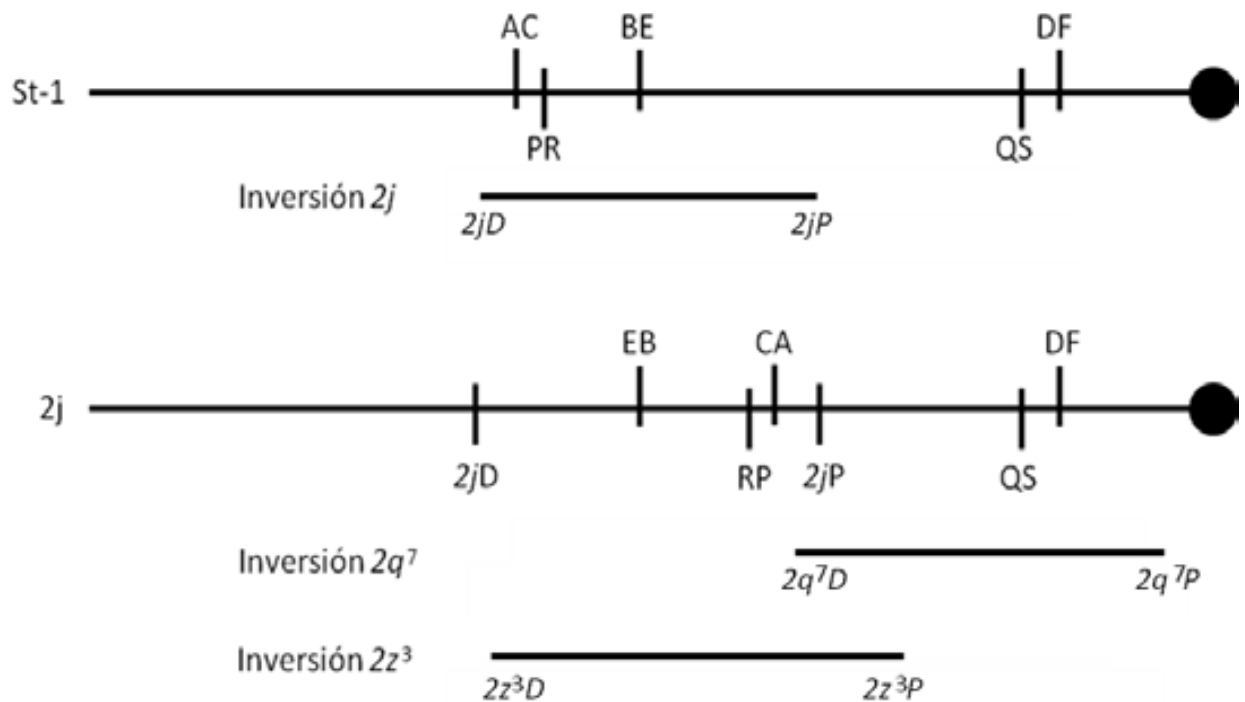


Figura 4.4. Esquema de la localización de los puntos de rotura de las inversiones descritas en el cromosoma 2 de *D. buzzatii*. Se muestran la ordenación estándar (st-1) y la ordenación 2j (con la inversión 2j) arriba y abajo, respectivamente. Los puntos de rotura de las inversiones polimórficas se han denominado D (distal) o P (proximal). En la ordenación estándar (st-1), se muestran los puntos de rotura de las inversiones $2m$, $2n$ y $2z^7$ (AC, BE, DF, PR y QS) y la localización de la inversión $2j$. En la ordenación 2j, se muestran los puntos de rotura de las inversiones $2m$, $2n$, $2z^7$ y $2j$ y la localización de las inversiones $2q^7$ y $2z^3$.

En las inversiones mediadas por *Galileo*, encontramos *BuT5* como colonizador secundario no del punto de rotura, sino de *Galileo*. Los elementos tipo *Galileo* podrían proporcionar dianas preferenciales de inserción a *BuT5*. Se ha descrito la inserción de algunos elementos en *Drosophila* en dianas preferenciales (O'Hare y Rubin 1983). Casals *et al.* (2006) encontraron cierta tendencia de estos TE a insertarse en las mismas bandas que otros miembros de la misma familia y los elementos tipo *Galileo*. En *D. virilis* también se encontró asociación entre los elementos *Penelope* y *Ulysses* (Zelentsova *et al.* 1999; Evgen'ev *et al.* 2000). En *Arabidopsis*, algunos TE con LTR también muestran inserción preferencial en otras inserciones (Witte *et al.* 2001).

De ser así, no habría evidencias de colonización secundaria de *BuT5* sin la presencia de elementos tipo *Galileo*. En las inversiones $2m$ y $2n$, por tanto, *BuT5* podría no ser un colonizador secundario. De este modo, sugerimos que *BuT5* podría no ser un colonizador secundario sino un colonizador de *Galileo*. La ausencia de *Galileo* en los puntos de rotura de estas inversiones, implicaría que no hay motivo para encontrar copias de *BuT5* si no es por azar a no ser que hubieran intervenido en el origen de las inversiones. Según esta hipótesis teórica, sólo cuando *BuT5* hubiera intervenido en ambas inversiones, se explicaría su presencia en los tres puntos de rotura de las inversiones $2m$ y $2n$. Cuando se relacionaban los mecanismos consistentes con ambos puntos de rotura (véase sección 4.1.4.2) se describían como consistentes las combinaciones i) $2m$ anterior mediante NHEJ más $2n$ por recombinación ectópica, ii) $2m$ anterior mediante inserción híbrida más $2n$ por recombinación ectópica iii) $2n$ anterior por NHEJ más inversión $2m$ por inserción híbrida y iv) la simultaneidad de ambas inversiones. La única posibilidad consistente con esta hipótesis es que la inversión $2m$ se haya generado primero mediante inserción híbrida y posteriormente la copia en el punto de rotura proximal recombinara con otra copia de *BuT5* para generar la inversión $2n$. Esta es la única opción que no sería consistente con el modelo de recurrencia descrito en la

sección 4.1.4.

4.5 NPL4

Cuando una reorganización tiene lugar en el genoma, puede no tener ningún efecto y su frecuencia dependerá de la deriva genética. Sin embargo, la selección también puede mantener polimorfismos estables así como promover su fijación o eliminación dependiendo de los efectos directos de la propia inversión y los puntos de rotura que la generan (Kirkpatrick y Barton 2006). Una inversión puede crear nuevas combinaciones de expresión, regulación o genes *de novo* a partir de fisiones de éstos así como el silenciamiento de una región por su yuxtaposición con la heterocromatina (variegación). La selección también modificará la frecuencia de una inversión si ésta incluye combinaciones de alelos favorables debido a la supresión de la recombinación (coadaptación).

En algunas inversiones se han descrito cambios en el patrón de expresión y regulación de genes (Spofford 1976; Eissenberg 1989; Henikoff 1990; Wesley y Eanes 1994; Weiler y Wakimoto 1995; Puig *et al.* 2004). En *Drosophila*, en la inversión *In(3R)Payne* (Kennington *et al.* 2006) se observó una fuerte selección epistática y efecto de posición por variegación en las inversiones *Bm2* (Belgovsky y Muller 1938) y *w[m4]* (Sitko 1938).

Se deduce que las reorganizaciones fijadas son beneficiosas o neutras. Las inversiones polimórficas pueden tener gradientes de frecuencia debido a sus distintos efectos en las distintas poblaciones de una misma especie en distintos ambientes, siendo beneficiosas sólo en ciertas condiciones (Dobzhansky 1971; Prevosti *et al.* 1988). Las inversiones *2m* y *2n* están fijadas en todo el complejo *buzzatii*, aunque desconocemos si lo hicieron por deriva o por haber conferido ventaja adaptativa. Hay una compleja cadena de eventos genómicos asociados a la inversión *2m*. Sin embargo, la duplicación no provoca el aumento en el número de copias

de ningún gen en *D. buzzatii*. La única copia del gen *CG4673* se encuentra en la región del punto de rotura BE mientras que el *CG5071* se anotó en la región del punto de rotura AC. También en AC, se localizaron fragmentos con identidad con los genes *CG5079* y *CG4673*, aunque ambos parecen degenerados. No se ha encontrado una copia funcional del gen *CG5079* en estas regiones del genoma de *D. buzzatii*. La pérdida de un gen es difícil de estudiar, más si en realidad su función no se ha perdido sino traspasado o el gen se ha retrotranspuesto y sigue siendo funcional. De todos modos, no encontramos información sobre la regulación, expresión ni función de este gen en ninguna base de datos. Esta falta de literatura al respecto podría indicar que su función no sería muy específica o insalvable.

El gen *CG5071*, codifica una proteína con actividad peptidil-prolil cis-trans isomerasa. En *D. buzzatii*, este gen no se encuentra anidado en *CG4673* como en los demás genomas de *Drosophila*. Aun así, conserva su estructura, orientación y localización. Por el contrario, el gen más interesante a estudiar que ha resultado afectado por estas reorganizaciones es sin duda el gen *CG4673*. En primer lugar, el gen *CG4673* se ha compactado debido a que la copia putativa funcional localizada en BE ha perdido las ~9 Kb de la secuencia del intrón 6. Esta compactación podría suponer una transcripción más eficiente y rápida en los genomas invertidos. Esta copia se encontró invertida respecto a los genomas sin la duplicación debido a la microinversión de éste (véase sección 4.1.3.1). Además, es el único gen incluido en la duplicación que se pudo anotar en ambos puntos de rotura ya que trazas de identidad con dicho gen pudieron ser anotadas en el punto de rotura AC.

4.5.1. Dominios y función

Los transcritos RA y RB del gen *CG4673* fueron correctamente anotados en el punto de rotura BE de *D. buzzatii*. Ambos transcritos están altamente conservados con los transcritos

ortólogos de *D. mojavensis* y *D. melanogaster* (véase sección 3.2 en Resultados) aunque el tamaño del gen es considerablemente más corto debido a la reducción del tamaño del intrón 6 (véase Tabla 3.12 en Resultados). El transcrito *CG4673-RA* de *D. buzzatii* tiene una identidad nucleotídica del 90,6% con el de *D. mojavensis* y del 79,6% con el de *D. melanogaster*. El transcrito *CG4673-RB* tiene una identidad nucleotídica 90,0% con el de *D. mojavensis* y 78,9% con el de *D. melanogaster*. Ambos transcritos sólo difieren en el primer exón (1-8 en el transcrito RA y 0 + 2-8 en el transcrito RB). Este primer exón podría tratarse de una señal específica de destino, pero el gen *CG4673* fue propuesto como un gen regulado por el sistema NMD (Hansen *et al.* 2009). El sistema NMD es un sistema de regulación de la traducción de mRNA (Culbertson y Leeds 2003). La regulación de la expresión NMD consiste en el procesamiento alternativo en la maduración de los mRNA. Estos mRNA contienen codones stop prematuros que los convierte en objetivo del sistema de degradación NMD. En *Drosophila*, se estableció que el 14% de los genes aumentaban su expresión 1,5 veces cuando el sistema NMD era eliminado en cepas mutantes (Rehwinkel *et al.* 2005). El estudio de estos genes aumentó los tipos de procesamientos alternativos que serían objetivos del sistema NMD como la inclusión o exclusión del primer exón, retención de intrones, procesamientos alternativos en las 5' y 3'UTR o poliadenilaciones específicas. La expresión de la transcripción RA del gen *CG4673* no se afectó en las cepas deficientes del sistema NMD, no así la transcripción RB (Hansen *et al.* 2009) significativamente aumentada. La carencia del exón 1 en la transcripción RB podría hacerla objeto del sistema NMD.

La traducción tentativa del transcrito *CG4673-RA* (véase Figura 3.11 en Resultados) fue utilizada para describir los dominios conservados (véase Figura 3.12 en Resultados). Ambas transcripciones (RA y RB) del gen *CG4673* anotadas en el punto de rotura BE de *D. buzzatii* codifican una proteína con un dominio NPL4, *Nuclear Pore Localization* (Botta *et al.* 2001; Bays *et al.* 2001). Ni el exón 0 (del transcrito RB) ni el exón 1 (del transcrito RA)

codifican parte de dicho dominio (codificado en los exones 2 al 7). El exón 8 (encontrado en ambas transcripciones) codifica para un dedo de zinc de las proteínas RAN adicional que sólo se conserva en la proteína NPL4 de metazoos, hongos y celomados.

La proteína NPL4 consta de un dominio activo NPL4 y un dedo de zinc N-terminal y fue inicialmente descrita como integrante del poro nuclear ya que se la involucró en el transporte nuclear. Bays *et al.* (2001) mediante ensayos de ubiquitinación, reveló que en realidad, la proteína NPL4 era requerida en el proceso de degradación asociada al retículo endoplasmático (sistema ERAD). Las cadenas de ubiquitina (substratos ERAD) han sido descritas como una señal específica de degradación. El complejo segregasa, está involucrado en el proceso de reconocimiento de distintas proteínas etiquetadas con ubiquitina. La proteína NPL4 forma parte de este complejo junto con las proteínas Cdc48p (VCP ATPasa) y Ufd1 (*Ubiquitin Fold Domain*) (Lundgren *et al.* 2005). Inicialmente se forma un dímero mediante la región central de la proteína NPL4 con Ufd1. Posteriormente, el conjunto interactúa con la VCP ATPasa por el extremo N-terminal de la proteína NPL4. Este complejo reconoce proteínas ubiquitinadas para desechar del retículo endoplasmático y las presenta al proteasoma 26S para su degradación. Bays y Hampton (2002) observaron que los elementos ubiquitinados y los substratos GFP se estabilizaban cuando se usaban RNA de interferencia para VCP y Ufd-1. Recientemente, le fueron atribuidas al complejo más funciones con la misma actividad catalítica, pero en distintos escenarios. La VCP ATPasa puede interactuar como una chaperona con especificidad para proteínas ubiquitinadas sintetizadas en el retículo para su retrotranslocación al núcleo (Lundgren *et al.* 2005). Finalmente, Cao *et al.* (2004) también pudo relacionar el complejo con el desensamblaje del huso mitótico y reensamblaje de la cubierta nuclear en interfase.

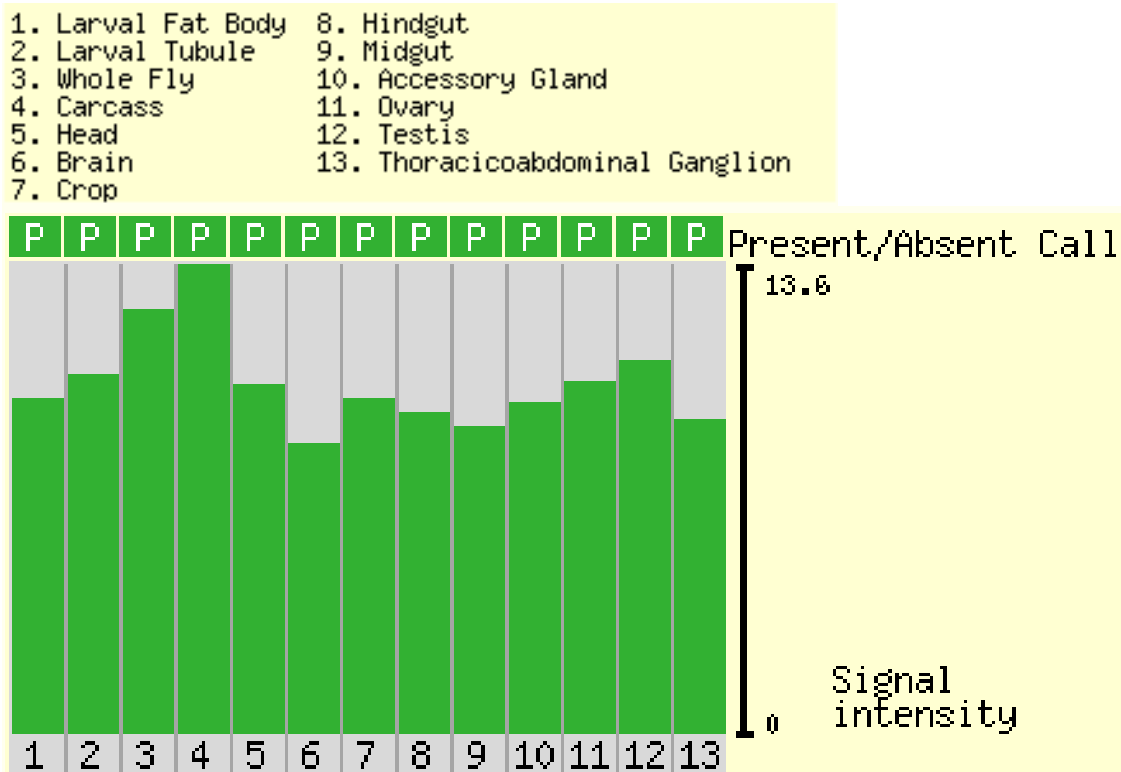
Existe un alto grado de conservación (véase Figura 3.14 en Resultados) entre las proteínas ortólogas a NPL4 de *D. buzzatii* de las 12 especies secuenciadas (incluyendo ambas

transcripciones de *D. melanogaster*). La proteína de *D. sechellia* es más larga y la de *D. persimilis* parece estar truncada (*Dper1* y *Dper2*). Por un lado, *D. sechellia* parece tener un final diferente debido a una mutación en el sitio de procesamiento en el exón 6 que cambiaría el marco de lectura produciendo una proteína que carece del dominio del dedo de zinc (véase Figura 3.15 en Resultados). En el caso de *D. persimilis*, hay anotados dos transcritos independientes debido a una mutación puntual que supone generar un codón stop en la secuencia codificante. El alineamiento de la proteína (véase Figura 3.13 en Resultados) muestra que la proteína *Dper2* de *D. persimilis* comienza cuando termina *Dper1*. No parece que ambas mutaciones sean viables debido a la constricción que sufre esta proteína en todo el género. Al tratarse ambos casos de mutaciones puntuales, creemos que se trata de errores puntuales de la secuenciación.

4.5.2. Datos de expresión

No existen diferencias significativas entre la expresión de los dos transcritos en *D. melanogaster* (*Life-Cycle Expression Data*, Arbeitman *et al.* 2002). Los datos de la expresión del gen *CG4673* de *D. melanogaster* para cada tejido y etapa del desarrollo se muestran en la Figura 4.5. No existen diferencias significativas en la expresión en el tiempo y en todas las etapas. La carcasa (tejido correspondiente a lo que queda del tórax después de haber extraído el intestino, los túbulos, las gónadas y el ganglio torácico-abdominal) tiene mayor expresión mientras que el cerebro y el intestino medio son los tejidos con menor cantidad de proteína.

A)



B)

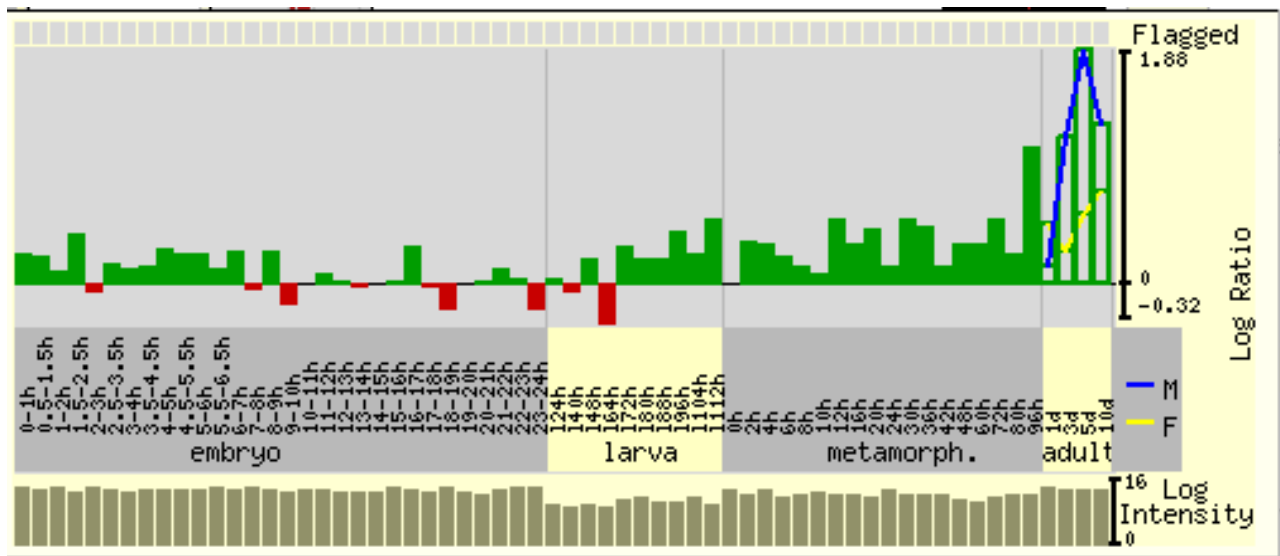


Figura 4.5: Datos de expresión de la transcripción RA del gen *CG4673* en *D. melanogaster*.

A) Expresión por tejidos. B) Expresión por estadios.

La RT-PCR 56 (PCR diseñada entre los exones 5 y 6 del gen *CG4673* de *D. buzzatii*) amplificó en todas las etapas del desarrollo ensayadas en *D. buzzatii* (véase Figura 3.21 en Resultados). Estos resultados son coherentes con la expresión en *D. melanogaster* y para un gen que codifica una proteína con implicación en los procesos metabólicos explicados anteriormente para NPL4. La RT-PCR 0 (PCR diseñada entre los exones 0 y 2 del gen *CG4673* de *D. buzzatii*) nos permitió comprobar que también existe transcrito RB del gen *CG4673* en *D. buzzatii* (véase Figura 3.21 en Resultados).

4.5.3. Región reguladora

El orden y la orientación de los genes *CG12250* (A) y *Ssadh* (B) implicados en el punto de rotura AB (incluyendo el gen *CG4673*) se conservan en las 12 especies secuenciadas del género *Drosophila* (todas no invertidas para *2m*). Por tanto, significa que no se conocen roturas en el género que interrumpan el contexto intergénico aguas arriba y aguas abajo del gen *CG4673*. Sólo en *D. buzzatii* (y especies invertidas para *2m*) encontramos que el gen *CG4673* no se encontraría en el contexto habitual.

El gen *CG4673* en *D. buzzatii* ha sufrido dos reorganizaciones. Por un lado, esta copia procede de una duplicación asociada a la inversión *2m*. Posteriormente, una microinversión explicaría porque la orientación de la duplicación es opuesta a la esperada. En la inversión *2m*, la duplicación interrumpe la región aguas abajo del gen *CG4673*. En la microinversión, se interrumpe también la región aguas arriba. Las repeticiones invertidas que flanquean la copia duplicada indican la región de los puntos de rotura de la microinversión. Como esta copia aún sigue siendo funcional, la región reguladora de este gen debe estar entre estas repeticiones invertidas. Los análisis de la conservación de esta región aguas arriba del gen en *D. buzzatii* y las especies del grupo *melanogaster* son consistentes en que dicha región reguladora se

encuentre en la inmediata región 5' del gen *CG4673* (véase sección 3.3.1 en Resultados). Las regiones aguas arriba y aguas abajo del gen *CG4673* fueron analizadas con el programa de predicción de secuencias reguladoras McPromoter señalando significativamente la propia repetición invertida anotada en 5' del gen *CG4673* como promotor putativo. La secuencia anotada como repetición invertida en 5' fue comparada con la región duplicada en el punto de rotura AC también en 5' de Ψ *CG4673* y las regiones ortóloga en *D. mojavensis*, *D. koepferae* y las propias repeticiones en la región 3' (véase Tablas 3.14 y 3.15 en Resultados). Aunque esta secuencia está anotada como región intergénica en *D. mojavensis*, se observó que los cambios acumulados no se corresponderían a una secuencia neutra. Esta región encuentra conservada cuando se comparan las secuencias anotadas en 5' de los genes funcionales de *D. buzzatii*, *D. koepferae* y *D. mojavensis*. La comparación de estas secuencias suponiendo evolución neutra para calcular el tiempo de divergencia de las especies invertidas respecto el linaje de *D. mojavensis* observamos tiempos cercanos a los 4 ma (véase Tabla 3.15 en Resultados). Esta estima está lejos del tiempo promedio de divergencia de 12,5 millones de años esperado (véase Figura 1.3 en Introducción). Asumimos por tanto, que la región reguladora del gen *CG43673* coincide con la repetición invertida anotada en *D. buzzatii*.

4.5.4. Transcritos funcionales de *CG4673* en *D. buzzatii*

En *D. buzzatii*, la secuencia reguladora putativa también se anotó en la región 5' de Ψ *CG4673* en AC (como parte del fragmento duplicado debido a la inversión *2m*) y en la región 3' de *CG4673* en BE (como parte del fragmento duplicado debido a la microinversión). Por tanto estas secuencias reguladoras podrían estar generando otros dos transcritos adicionales al mRNA funcional del gen *CG4673*. Por un lado, Ψ *CG4673* en AC podría tener capacidad plena o parcial para transcribirse y el gen *CG4673* podría producir un transcrito

antisentido. En el primer caso, el gen $\Psi CG4673$ podría haber sido funcional y los motivos que provocaron su degeneración deben ser distintos de la interrupción de la región reguladora. Sin embargo, en este caso, el programa McPromoter no señaló ningún promotor putativo en esta región. Por otro lado, el programa McPromoter si detectó un promotor putativo orientado hacia el gen $CG4673$ en la región reguladora putativa en la región 3' del gen $CG4673$ en BE. De este modo, dos promotores rodearían al gen en el punto de rotura BE.

Estos dos promotores putativos adicionales (en AC y en 3' del gen $CG4673$) podrían estar regulando la expresión de un transcrito, aunque bien podrían ser inactivos y tratarse de falsos positivos residuales. En este caso, se explicaría por qué el programa McPromoter aún es capaz de detectar significación en la región anotada en 3' y no en la copia AC. La región AC (inversión $2m$) proviene de un evento más antiguo que la región anotada en 3' (microinversión) y llevaría más tiempo acumulando cambios. Además, cuando estas secuencias se comparan dentro de las especies invertidas, se encuentran menos constreñidas que cuando se compararon las secuencias en 5' del gen $CG4673$ entre especies. Para conocer la funcionalidad de estas secuencias promotoras, hicimos distintos estudios de expresión referentes a los posibles transcritos alternativos derivados de las duplicaciones de la región reguladora.

4.5.4.1. Transcrito del pseudogen $\Psi CG4673$ en *D. buzzatii*

Las PCR 2, 3, 7 y 3'UTR (diseñadas en el pseudogen en las regiones con identidad con el gen $CG4673$) amplificaron correctamente en todos los estados del desarrollo (véase Figura 3.21 en Resultados). Por tanto, confirmamos que existe transcrito procedente del pseudogen y es consistente con que la secuencia anotada como repetición invertida sea la región promotora. La proteína NPL4 ha sido ampliamente estudiada y nunca se había descrito su duplicación en

ninguna especie. Por tanto, no parece que dos copias de este gen supusieran una ventaja evolutiva, fueran necesarias o incluso toleradas. Independientemente que se trate de un pseudogen (ha perdido la práctica totalidad del marco de lectura y los sitios de procesamiento descritos en la copia funcional) puede existir transcrito residual. Las duplicaciones de genes normalmente se resuelven con la pseudogenización de una de las copias. Aunque la degradación es evidente, no conocemos su grado de pseudogenización (estados “ghost” o “death” de las duplicaciones, Zheng y Gerstein 2007), y menos si su función reguladora no fue interrumpida. Un gen duplicado puede adquirir funciones parciales o subfuncionalizar las antiguas. Por otra parte, cuando la selección purificadora se relaja contra los constreñimientos funcionales en una copia duplicada, éstos pueden adquirir nuevas funciones. Sin embargo, algunos pseudogenes “dying” incluso conservan alguna actividad de transcripción aun sin tener función bioquímica.

4.5.4.2. Transcrito antisentido del gen *CG4673* en *D. buzzatii*

Aparte de los dos transcritos detectados (uno proveniente del gen *CG4673* y el otro de $\Psi CG4673$) podríamos esperar un tercer transcrito regulado por la repetición invertida anotada en la región 3' del gen *CG4673*. La región promotora en 5' de $\Psi CG4673$ no fue propuesta por el programa McPromoter y, aun así, se detectó transcrito. En cambio, la región reguladora putativa en 3' de *CG4673* si fue señalada significativamente por el programa McPromoter, por lo que transcrito procedente de esta región reguladora se espera con más motivo.

La secuencia de este tercer transcrito sería complementaria al mRNA funcional. No se diseñaron PCR específicas de sentido para detectar este transcrito al no conocer su longitud, origen y final o sitios de procesamiento. El experimento de hibridación de Northern, reveló la existencia de un tercer transcrito relacionado con la secuencia del gen *CG4673*. La hibridación

de Northern es cualitativa, así que las intensidades de las bandas no son proporcionales ni comparables.

La membrana 2 se hibridó con la sonda 2-3 aunque no se obtuvo señal de hibridación. La misma sonda hibridó correctamente en el Dot Blot, por lo que la falta de hibridación no se debe a problemas de tamaño o de inespecificidad de la sonda. En la hibridación de Dot Blot, la mayor señal se obtuvo en el DNA genómico de *D. buzzatii*. No hubo señal en *D. mojavensis*, probablemente debido a la excesiva divergencia de la sonda o a la concentración de ésta.

Las membranas 1 y 3 hibridaron correctamente con la sonda 5-7 (véase Figura 3.23 en Resultados). La principal diferencia entre ellas es la cantidad de mRNA cargado en el gel (véase Tabla 2.8 en Materiales y Métodos), siendo mucho menor la cantidad de muestra en el gel para la membrana 3. La diferencia entre la intensidad de la señal de estas hibridaciones en las membrana 1 y 3 con la hibridación de Dot Blot se debe a que la sonda PCR 2-3 corresponde a la copia de *ΨCG4673*, que incluso incluye el intrón 2. La sonda de la membrana 1 proviene de la copia de *CG4673* de *D. buzzatii*, con mayor identidad con la copia del gen en *D. mojavensis*.

Se realizaron dos controles. Por una lado, se hibridó la sonda en DNA genómico de *D. buzzatii*. La señal de hibridación con el DNA se observa al principio del carril ya que las condiciones usadas para el RNA no permiten la correcta penetración del DNA genómico en el gel. En el otro control se hibridó la misma sonda con el transcriptoma de una mezcla de todos los estadios de mRNA de *D. mojavensis*. El mRNA usado en los geles de hibridación de Northern procede del citoplasma y ha sido procesado, por lo que han eliminado los intrones y se les ha añadido la caperuza en la 5'UTR y la cola de poli-A después de la 3'UTR. Por tanto, aunque el gen *CG4673* en *D. buzzatii* no tenga el intrón 6 (~9Kb), los mRNA del gen en ambas especies tendrán un tamaño parecido. Sólo se obtuvo una señal de hibridación en el carril de *D. mojavensis*, por lo que la sonda 5-7 procedente de *CG4673* no hibrida

inespecíficamente con ningún otro gen con secuencia parecida o alguna transcripción de tamaño muy distinto procedente de algún procesamiento alternativo.

En la membrana 3 hay una buena señal de hibridación en los estadios de embrión, pupa y ambos adultos. La menor hibridación se obtiene en larva, aunque la intensidad de la señal no es proporcional a la expresión del gen. La señal de hibridación principal se localiza en la misma posición en los distintos estadios de *D. buzzatii* (~1.100 pb), y es parecida a la señal obtenida en la hibridación en el transcriptoma de *D. mojavensis*. Las señales de la hibridación principal de *D. buzzatii* quedan localizadas ligeramente por encima de la de *D. mojavensis*, indicando que el mensajero en *D. buzzatii* es ligeramente más largo. Esta señal corresponde al mRNA del gen *CG4673* en *D. buzzatii*. El tamaño de la señal es mayor al tamaño del CDS porque el mRNA que hibrida con la sonda está compuesto por el propio CDS más ambas UTR y la caperuza 5' y la cola poli-A en 3'. En los carriles de mRNA de pupa y adultos, una señal de hibridación aparece por debajo de la hibridación principal (~950 pb), aunque su interpretación es dudosa al camuflarse entre una mancha de degradación de mRNA. En la hibridación con mRNA de pupa, también se pudo observar una señal de hibridación de mayor tamaño que la principal (~2.000 pb). En la membrana 3 se realizó la misma hibridación con menores concentraciones de mRNA cargadas en el gel. La diferencia se hace evidente en el revelado de la hibridación, puesto que tanto la intensidad como el número de hibridaciones, es menor en esta membrana. En la membrana 3, sólo se observa señal en las hibridaciones con mRNA de pupa y el control con *D. mojavensis*. El tamaño de la señal de hibridación en el control indica que, de nuevo se está hibridando el mRNA del gen *CG4673*. En la hibridación con pupa, la señal más intensa también corresponde a este mRNA.

En esta membrana, la señal de mayor tamaño ha desaparecido mientras que claramente se observa la de menor tamaño por debajo de la hibridación principal sin la contaminación de la primera membrana.

Por tanto, al menos en el mRNA de pupa, se han observado tres señales. La señal principal corresponde con el gen funcional y se observa en ambas membranas y en el transcriptoma de *D. mojavensis*. Una señal de mayor tamaño se observa en la membrana 1 y una de menor tamaño en la membrana 3. Estas tres hibridaciones corroboran que hay tres transcritos con homología a la sonda utilizada del gen *CG4673*. Hay dos transcritos adicionales que no se observan en los genomas no invertidos (*D. mojavensis*), por lo que es consistente que se deban a las reorganizaciones ocurridas en *D. buzzatii*.

La hibridación de Northern ha detectado el transcrito del gen y el del pseudogen (previamente detectados mediante RT-PCR). El tercer transcrito detectado podría proceder de la tercera región promotora putativa aguas abajo del gen *CG4673* en *D. buzzatii*. Un transcrito procedería de la expresión del pseudogen en el punto de rotura AC y otros dos del punto de rotura BE, pero en sentidos opuestos (uno en el sentido 5'-3' y otro antisentido 3'-5'). Sólo el sentido 5'-3' del gen *CG4673* está completo y potencialmente funcional.

4.5.5. Relación entre los diferentes transcritos del gen *CG4673*

Un transcrito no necesariamente ha de ser traducido para tener rol o actividad bioquímica. Incluso cuando las transcripciones no llegan a traducirse, pueden estar asociadas con la regulación de la copia funcional. En el caso del transcrito del pseudogen detectado por RT-PCR e hibridación de Northern podría ser eliminado directamente. La falta del exón 1 de la transcripción RB del gen *CG4673* producida por el procesamiento alternativo podría ser una señal de degradación por el sistema de control de expresión NMD (véase sección 4.5.1). En $\Psi CG4673$, se pudo anotar el exón 0 pero no el 1. Por tanto, los transcritos procedentes de $\Psi CG4673$ podrían ser eliminados directamente por el mismo sistema que regula la expresión en el gen funcional. Sin embargo, no sabemos las señales de procesamiento del transcrito

producidos desde la región 3' del gen *CG4673* y no sabemos si incluye la secuencia del exón 1 o 0. De todos modos, al ser antisentido, la posible señal codificada en el exón 1 o 0, sería complementaria en este transcrito, por lo que no sería reconocido por este sistema de regulación. La complementariedad de ambos transcritos podría unirlos y formar moléculas de doble cadena, por lo que se diseñó el experimento de detección de dsRNA. La causa más común de dsRNA es la regulación por interferencia. Un mismo fragmento se transcribe en ambos sentidos regulado por dos promotores distintos. La transcripción antisentido regula la cantidad de mRNA libre que se traduce de la transcripción de un gen. Cuando un transcrito forma una molécula de doble cadena con un transcrito antisentido (dsRNA) no puede ser traducido normalmente.

El experimento de detección de dsRNA reveló que los transcritos relacionados con el gen *CG4673* forman moléculas de doble cadena. Alguno de los transcritos detectados podría estar modulando la actividad o expresión del transcrito procedente del gen funcional. El dsRNA se compone de dos de fragmentos de RNA de cadena simple (ssRNA) complementarios que se unen para crear una molécula de doble cadena. Se amplificó producto de PCR en las regiones correspondientes a los exones 3 y 7 (véase Figura 3.22 en Resultados) por lo que, al menos en estas regiones, existe doble cadena que se protege de la digestión con RNAasa. En cambio, la PCR 2 amplificó como en el control (*Gapdh*), deduciéndose que en esta región no habría molécula de doble cadena. Para interrumpir la traducción de un transcrito, basta una pequeña región de doble cadena. La región del exón 2 puede encontrarse como cadena simple y ser digerido hasta la región del exón 3, donde se encontraría protegido por la doble cadena.

La región promotora del gen *CG4673* se ha duplicado dos veces. La primera en la duplicación asociada a la inversión *2m* y la segunda con la duplicación asociada a la microinversión. Las tres regiones promotoras regulan tres transcritos relacionados con el gen

CG4673. Estos tres transcritos no tienen la misma orientación. La región promotora del transcrito procedente de *ΨCG4673* en el punto de rotura AC y la del transcrito funcional del punto de rotura BE están localizadas en 5', por lo que la orientación es directa (5'-3'). *ΨCG4673* y *CG4673* están orientadas en la misma dirección y al ser ambos transcritos 5'-3', no son capaces de formar moléculas de RNA de doble cadena. El transcrito complementario detectado debe ser antisentido (3'-5'). La región reguladora anotada en 3' del gen *CG4673* produciría un transcrito antisentido. Cualquiera de los dos transcritos con la región reguladora anotada en 5' podría crear moléculas de doble cadena con el transcrito antisentido. De todos modos, las PCR 3 y 7 fueron diseñadas específicamente para amplificar *ΨCG4673*. De este modo, el mRNA de doble cadena detectado proviene de la unión de los transcritos antisentido y del pseudogen. Por tanto, la molécula de doble cadena se estaría formando entre el transcrito 5'-3' de *ΨCG4673* y el transcrito 3'-5' del gen *CG4673*.

No podemos concretar la funcionalidad de los tres transcritos encontrados. Tampoco hemos podido averiguar cuántos de ellos llegan a traducirse. De todos modos, pensamos que sólo uno de ellos codifica realmente una proteína. Los otros dos transcritos podrían estar regulando el transcrito del gen *CG4673*. Dicho transcrito está conservado con todas las especies de *Drosophila* y su estreñimiento y la función descrita de su dominio conservado, hacen pensar que es plenamente funcional y necesario. Un transcrito antisentido (como el que podría producirse desde la región reguladora situada a 3' de *CG4673*) sería complementario y podría estar inactivando la traducción y síntesis de la proteína. Para evitar que este transcrito antisentido inactive un gen imprescindible, se habría conservado la transcripción procedente de *ΨCG4673*. Este transcrito, aunque no codifique para ninguna proteína funcional, tiene sentido 5'-3' como el del gen *CG4673*. Este transcrito, interactuaría con el transcrito antisentido de *CG4673*, formando moléculas de doble cadena de mRNA. Si este transcrito inactiva el transcrito antisentido, el transcrito del gen *CG4673* queda libre y se traduciría

normalmente. Aunque los resultados experimentales son consistentes con esta teoría de regulación, no conocemos las causas ni los casos en los que se produce. No hemos encontrado argumentos ni justificación para este sistema complejo de regulación. No parece que haya podido ser objeto de selección positiva. No hay evidencias que apunten a que este sistema de regulación confiera alguna ventaja o que haya sido el motivo de la fijación de las reorganizaciones $2m$ y $2n$. Más bien, pensamos que podría tratarse de un efecto colateral de algún otro evento que si fue seleccionado positivamente. Este sistema de regulación se habría formado de forma accidental, de forma paralela a otro evento que lo ha arrastrado (efecto *hitch-hiking*).

5. CONCLUSIONES

Las conclusiones de este trabajo son:

- 1) Se han localizado y cartografiado las regiones de los puntos de rotura de las inversiones $2m$, $2n$ (puntos de rotura AB, CD y EF) y $2z^7$ (PQ y RS) en el genoma de *D. mojavensis* (genoma no invertido) que representa la ordenación ancestral del grupo *repleta*. Los puntos de rotura no interrumpen ningún gen.
- 2) Se han aislado y secuenciado los puntos de rotura de las inversiones $2m$ y $2n$ en el genoma de *D. buzzatii* (AC, BE y DF). Se han aislado y secuenciado los puntos de rotura de la inversión $2z^7$ en los genomas de *D. buzzatii*, *D. martensis* y *D. koepferae* (PR y QS).
- 3) Para aislar los puntos de rotura AC y BE, se han secuenciado los clones BAC 20O19 y 1N19 de la genoteca de *D. buzzatii*. Los puntos de rotura en los dos clones BAC secuenciados acumulan mayor número de inserciones de elementos transponibles que fuera de éstos.
- 4) El análisis molecular confirma la reutilización citológica previamente observada entre las inversiones $2m$ y $2n$. Este trabajo es consistente con la distribución no al azar de los puntos de rotura.
- 5) La inversión $2m$ está asociada a una duplicación génica que incluye el gen *CG4673* y dos genes anidados en éste (*CG5071* y *CG5079*). En la posición original (punto de rotura AC), los genes *CG4673* y *CG5079* han pseudogeneizado. En la duplicación (punto de rotura BE), el gen *CG4673* es funcional pero no hay rastro de los genes anidados debido a la pérdida del intrón 6 (que incluye ambos genes anidados).
- 6) La duplicación asociada a la inversión está microinvertida. Una repetición invertida de 121 pb flanquea dicha duplicación. El origen de la microinversión de la

duplicación es consistente con la rotura escalonada más reparación por NHEJ generando estas repeticiones invertidas flanqueantes.

- 7) Por tanto, hasta tres puntos de rotura se localizan en la misma junta de la región del punto de rotura CD ($2m$ proximal, $2n$ distal y proximal de la microinversión).
- 8) Los puntos de rotura CD (punto de rotura compartido entre las inversiones $2m$ y $2n$) y EF son recurrentes a lo largo del género *Drosophila*. La región del punto de rotura AB no se ha observado interrumpida en ninguna de las especies secuenciadas del género *Drosophila*. La recurrencia se ha asociado a mecanismos NHEJ. El origen de las inversiones $2m$ y $2n$ es consistentes con tres mecanismos distintos.

(i) primero inversión $2m$ mediante NHEJ + inversión $2n$ por recombinación ectópica entre copias del elemento *BuT5*.

(ii) primero inversión $2n$ mediante NHEJ más inversión $2m$ por inserción híbrida.

(iii) que sean simultáneas. (1 rotura escalonada + 2 roturas DSB más reparación NHEJ de los dos fragmentos generados).

- 9) La recombinación ectópica (NAHR) entre copias del elemento *Galileo* es consistente para explicar el origen de la inversión $2z^7$ como el de las 3 inversiones polimórficas descritas anteriormente en *D. buzzatii*. El punto de rotura PQ está reutilizando un punto de rotura producido en el subgénero *Sophophora*.
- 10) El elemento *BuT5* se ha anotado en todos los puntos de rotura. En las inversiones $2m$ y $2n$ este elemento se ha anotado solo, mientras que en la inversión $2z^7$ se ha anotado acompañando a *Galileo*.

- 11) Podría existir una relación entre la presencia del elemento transponible *BuT5* en los puntos de rotura y su implicación en el origen de la inversión. Dicho elemento parece ser un clásico colonizador secundario del elemento *Galileo*, implicado en el origen de las inversiones $2z^7$, $2j$, $2q^7$ y $2z^3$. La presencia de *BuT5* sin *Galileo*, podría implicarlo en el origen de la inversión (inversiones $2m$ y $2n$).
- 12) Parece existir una relación entre la frecuencia de las inversiones descritas en *D. buzzatii* y el mecanismo que las ha originado. Las inversiones polimórficas descritas en *D. buzzatii* se han originado por recombinación ectópica. Sólo en las inversiones fijadas se ha observado un mecanismo distinto. La localización de los elementos transponibles podría generar con mayor frecuencia inversiones neutras.
- 13) Al menos la inversión $2m$ se han generado por un mecanismo distinto a la recombinación ectópica por lo que no es el mecanismo exclusivo del cromosoma 2 en *D. buzzatii* (como se había observado en las inversiones anteriormente descritas). Sin embargo si se observa cierta preferencia ya que la mayoría de inversiones descritas en este cromosoma se han originado con este mecanismo. La frecuencia de TE observada en los clones BAC secuenciados de *D. buzzatii* no es mayor que la frecuencia de TE descrita en los genomas secuenciados del género *Drosophila*. De este modo, esta preferencia en el mecanismo de recombinación ectópica no se debería a la densidad de TE en el cromosoma 2.
- 14) La inversión $2z^7$ parece haberse originado por recombinación ectópica entre copias de un elemento tipo *Galileo*, por lo que éste no sería un mecanismo reciente ni exclusivo de inversiones polimórficas.
- 15) La repetición invertida que flanquea la microinversión de la duplicación asociada a la inversión $2m$ coincide con el promotor putativo del gen *CG4673*. Se han localizado hasta tres posibles promotores: (i) en la región 5' del gen $\Psi CG4673$ en

el punto de rotura AC y (ii) en la región 5' y (iii) en la región 3' del gen *CG4673* del gen *CG4673* en el punto de rotura BE.

16) Hasta tres transcritos relacionados con el gen *CG4673* han sido detectados en *D. buzzatii* por uno sólo en el genoma no invertido de *D. mojavensis*. Uno procedería de la copia degenerada del gen (punto de rotura AC) y otros dos de la copia del gen funcional (punto de rotura BE).

17) Al menos dos de los tres transcritos detectados se relacionan entre sí formando moléculas de mRNA de doble cadena. Los transcritos antisentido (promotor en la región 3' del gen *CG4673*) y procedente del gen $\Psi CG4673$ podrían asociarse y regular el transcrito del gen funcional (promotor en la región 5' del gen *CG4673*).

6. REFERENCIAS

- Adams MD., SE. Celniker, RA. Holt, CA. Evans, JD. Gocayne, PG. Amanatides, SE. Scherer SE, PW. Li, RA. Hoskins, RF. Galle *et al.* (2000). "The genome sequence of *Drosophila melanogaster*". Science Mar 24;287(5461):2185-95.
- Andolfatto, P., J. D. Wall y M. Kreitman (1999). "Unusual haplotype structure at the proximal breakpoint of *In(2L)t* in a natural population of *Drosophila melanogaster*." Genetics 153(3): 1297-311.
- Andolfatto, P., F. Depaulis y A. Navarro (2001). "Inversion polymorphisms and nucleotide variability in *Drosophila*". FGenet. Res.77: 1-8.
- Arbeitman, MN., EE. Furlong, F. Imam, E. Johnson, BH. Null, SS. Baker, MA. Krasnow, MP. Scott, RW. Davis y KP. White (2002). "Gene expression during the life cycle of *Drosophila melanogaster*". Science Sep 27;297(5590):2270-5.
- Arcot, SS., Z. Wang, JL. Weber, PL. Deininger y MA. Batzer (1995). "*Alu* repeats: a source for the genesis of primate microsatellites". Genomics 1;29(1):136-44.
- Armengol, L., M. A. Pujana, J. Cheung, S. W. Scherer y X. Estivill (2003). "Enrichment of segmental duplications in regions of breaks of synteny between the human and mouse genomes suggest their involvement in evolutionary rearrangements." Hum Mol Genet 12(17): 2201-8.
- Ashburner, M. (1989). "*Drosophila*". A laboratory handbook. Cold Spring Harbor Laboratory Press. Cold Spring Harbor, NY
- Bailey, J. A., R. Baertsch, W. J. Kent, D. Haussler y E. E. Eichler (2004). "Hotspots of mammalian chromosomal evolution." Genome Biol 5(4): R23.
- Balanyà, J., L. Serra, G. W. Gilchrist, R. B. Huey, P. Pascual *et al.* (2003). "Evolutionary pace of chromosomal polymorphism in colonizing populations of *Drosophila subobscura*: an evolutionary time series". Evolution 57: 1837-1845.

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

Bays, NW., RG. Gardner, LP. Seelig, CA. Joazeiro y RY. Hampton (2001). "*Hrd1p/Der3p* is a membrane-anchored ubiquitin ligase required for ER-associated degradation". *Nat. Cell. Biol.* Jan;3(1):24-9.

Bays, NW. y RY. Hampton (2002). "Cdc48-Ufd1-Npl4: stuck in the middle with Ub". *Curr. Biol.* May 14;12(10):R366-71.

Becker, TS. y B. Lenhard B. (2007). "The random versus fragile breakage models of chromosome evolution: a matter of resolution". *Mol. Genet. Genomics* Nov;278(5): 487-91.

Belgovsky, M. L. y H. J. Muller (1938). "Further evidence of the prevalence of minute rearrangement and absence of simple breakage in and near chromocentral regions and its bearing on the mechanism of mosaicism and rearrangement". *Genetics* 23: 139-140.

Bhutkar, A., SW. Schaeffer, SM. Russo, M. Xu, TF. Smith y WM. Gelbart (2008). "Chromosomal rearrangement inferred from comparisons of 12 *Drosophila* genomes". *Genetics* Jul;179(3):1657-80.

Blanchette, M., T. Kunisawa y D. Sankoff (1996). "Parametric genome rearrangement." *Gene* 172(1): GC11-7.

Botta, A., C. Tandoi, G. Fini, G. Calabrese, B. Dallapiccola y G. Novelli (2001). "Cloning and characterization of the gene encoding human NPL4, a protein interacting with the ubiquitin fusion-degradation protein (UFD1L)". *Gene* Sep 5;275(1):39-46.

Bourque, G., P. A. Pevzner y G. Tesler (2004). "Reconstructing the genomic architecture of ancestral mammals: lessons from human, mouse, and rat genomes." *Genome Res* 14(4): 507-16.

- Bowen, NJ. y JF. McDonald (2001). "Drosophila euchromatic LTR retrotransposons are much younger than the host species in which they reside". *Genome Res.* Sep;11(9):1527-40.
- Brookfield, JF (1986). "The population biology of transposable elements". *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* Jan 29;312(1154):217-26.
- Brookfield, JF. (2004). "Evolutionary genetics: mobile DNAs as sources of adaptive change?" *Curr. Biol.* 14:R344-45.
- Burge, C. y S. Karlin (1997). "Prediction of complete gene structures in human genomic DNA". *J. Mol. Biol.* 268, 78-94.
- Cáceres, M., A. Barbadilla, M. y A. Ruiz (1997). "Generation of a widespread *Drosophila* inversion by a transposable element." *Science* 285(5426): 415-8.
- Cáceres, M., J. M. Ranz, A. Barbadilla, M. Long y A. Ruiz (1999). "Inversion length and breakpoint distribution in the *Drosophila buzzatii* species complex: Is inversion length a selected trait?" *Evolution* 51 1149-1155.
- Cáceres, M., M. Puig y A. Ruiz (2001). "Molecular characterization of two natural hotspots in the *Drosophila buzzatii* genome induced by transposon insertions". *Genome Res.*11(8):1353-64.
- Cao, K., R. Nakajima, HH. Meyer y Y. Zheng (2003). "The AAA-ATPase *Cdc48/p97* regulates spindle disassembly at the end of mitosis". *Cell* Oct 31;115(3):355-67.
- Capy, P. (1998). "Evolutionary biology. A plastic genome." *Nature* 396(6711): 522-3.
- Carson, H. L. (1992). "Inversions in Hawaiian *Drosophila*". *Drosophila inversions polymorphism*. Eds. C. B. Krimbas y J. R. Powell. CRC Press, Boca Ratón, FL. pp. 407-439.
- Casals, F., M. Cáceres y A. Ruiz (2003). "The *foldback*-like transposon *Galileo* is involved in the generation of two different natural chromosomal inversions of *Drosophila buzzatii*." *Mol Biol Evol* 20(5): 674-85.

Casals, F., M. Cáceres, MH. Manfrin, J. González y A. Ruiz (2005) "Molecular characterization and chromosomal distribution of *Galileo*, *Kepler* and *Newton*, three *foldback* transposable elements of the *Drosophila buzzatii* species complex". *Genetics* 169(4):2047-59.

Casals, F., J. González y A. Ruiz (2006). "Abundance and chromosomal distribution of six *Drosophila buzzatii* transposons: *BuT1*, *BuT2*, *BuT3*, *BuT4*, *BuT5*, and *BuT6*". *Chromosoma* Oct;115(5):403-12.

Casillas S., N. Petit y A. Barbadilla (2005). "DPDB: a database for the storage, representation and analysis of polymorphism in the *Drosophila* genus". *Bioinformatics*. Sep 1;21 Suppl 2:ii26-30.

Chalmers, R. M. y N. Kleckner (1996). "*IS10/Tn10* transposition efficiently accommodates diverse transposon end configurations." *Embo J* 15(18): 5112-22.

Charlesworth, B. (1974). "Inversion polymorphism in a two-locus genetic system". *Genet. Res.* 23: 259–280.

Charlesworth, B. (1996). "Background selection and patterns of genetic diversity in *Drosophila melanogaster*". *Genet. Res.* Oct;68(2):131-49.

Cirera, S., J. M. Martín-Campos, C. Segarra y M. Aguadé (1995). "Molecular characterization of the breakpoints of an inversion fixed between *Drosophila melanogaster* and *D. subobscura*." *Genetics* 139(1): 321-6.

Cirulli, ET. y MA. Noor (2007). "Localization and characterization of X chromosome inversion breakpoints separating *Drosophila mojavensis* and *Drosophila arizonae*. *J. Hered.* Mar-Apr;98(2):111-4.

Clayton, F. E. y Guest, W. C. (1986). "Overview of chromosomal evolution in the family of *Drosophilidae*". *The Genetics and Biology of Drosophila*, Vol. 3e. Eds. M. Ashburner, H. L. Carson y J. N. Thompson, Jr. Academic Press, London. pp. 1-38

- Coghlan, A. y K. H. Wolfe (2002). "Fourfold faster rate of genome rearrangement in nematodes than in *Drosophila*." *Genome Res* 12(6): 857-67.
- Cohen, O., C. Cans, M. Cuillel, J. L. Gilardi, H. Roth, M. A. Mermet, P. Jalbert y J. Demongeot (1996). "Cartographic study: breakpoints in 1574 families carrying human reciprocal translocations." *Hum Genet* 97(5): 659-67.
- Costas, J. y H. Naveira (2000). "Evolutionary history of the human endogenous retrovirus family *ERV9*." *Mol. Biol. Evol.* Feb;17(2):320-30.
- Coyne, J. A., B. C. Moore, J. A. Moore, J.R. Powell y C. E. Taylor (1992). "Temporal stability of third-chromosome inversion frequencies in *Drosophila persimilis* and *D. pseudoobscura*". *Evolution* 46: 1558–1563.
- Culbertson, MR. y PF. Leeds (2003). "Looking at mRNA decay pathways through the window of molecular evolution". *Curr. Opin. Genet. Dev.* Apr;13(2):207-14. Review.
- Craig NL. (1997). "Target site selection in transposition". *Annu. Rev. Biochem.* 66: 437-74.
- Craig, NL., R. Craigie, M. Gellert y AM. Lambowitz (2002). "Mobile DNA II". Washington, DC: Am. Soc. Microbiol. Press.
- Crosby MA., JL. Goodman, VB. Strelets, P. Zhang, WM Gelbart y Consorcio FlyBase (2007) "FlyBase: genomes by the dozen". *Nucleic. Acids. Res.* Jan;35(Database issue):D486-91.
- Dehal, P., P. Predki, A. S. Olsen, A. Kobayashi, P. Folta, S. Lucas, M. Land, A. Terry, C. L. Ecale Zhou, S. Rash, Q. Zhang, L. Gordon, J. Kim, C. Elkin, M. J. Pollard, P. Richardson, D. Rokhsar, E. Uberbacher, T. Hawkins, E. Branscomb y L. Stubbs (2001). "Human chromosome 19 and related regions in mouse: conservative and lineage-specific evolution." *Science* 293(5527): 104-11.
- Delprat, A., B. Negre, M. Puig y A. Ruiz (2009). "The transposon *Galileo* generates natural chromosomal inversions in *Drosophila* by ectopic recombination". *PLoS One* Nov 18;4(11).

Dernburg, A. F. (2001). "Here, there, and everywhere: kinetochore function on holocentric chromosomes." *J Cell Biol* 153(6): F33-8.

Dobzhansky, T. (1970). "Genetics of the Evolutionary Process". Columbia University Press, New York.

Dobzhansky, T. (1971). "Evolutionary oscillation in *Drosophila pseudoobscura*" pp. 109–133 in *Ecological Genetics and Evolution*, edited by R. Creed. Blackwell Scientific, Oxford.

Dobzhansky, Th. Y Powell, J. R. (1975). "The *willistoni* group sibling species of *Drosophila*" Vol. 3. Ed. R. C. King. Plenum Press, New York. *Handbook of genetics*. pp 589-622.

Drosophila 12 Genomes Consortium, AG. Clark, MB. Eisen, DR. Smith, CM. Bergman, B. Oliver, TA. Markow, TC. Kaufman, M. Kellis, W. Gelbart, VN. Iyer *et al.* "Evolution of genes and genomes on the *Drosophila* phylogeny". *Nature* Nov 8;450(7167):203-18.

Dunham, M. J., H. Badrane, T. Ferea, J. Adams, P. O. Brown, F. Rosenzweig y D. Botstein (2002). "Characteristic genome rearrangements in experimental evolution of *Saccharomyces cerevisiae*." *Proc. Natl. Acad. Sci. USA* 99(25): 16144-9.

Eanes, WF., C. Wesley y B. Charlesworth (1992). "Accumulation of *P* elements in minority inversions in natural populations of *Drosophila melanogaster*". *Genet. Res.* Feb;59(1):1-9.

Edgar, RC. (2004). "MUSCLE: multiple sequence alignment with high accuracy and high throughput". *Nucleic Acids Res.* Mar 19;32(5):1792-7.

Eggleston, WB., NR. Rim y JK. Lim (1996). "Molecular characterization of *hobo*-mediated inversions in *Drosophila melanogaster*." *Genetics* 144(2): 647-56.

Ehrlich, J., D. Sankoff y JH. Nadeau (1997). "Synteny conservation and chromosome rearrangements during mammalian evolution". *Genetics* Sep;147(1):289-96.

Eickbush, TH. y VK. Jamburuthugoda (2008). "The diversity of retrotransposons and the properties of their reverse transcriptases. *Virus Res.* Jun;134(1-2):221-34.

- Eissenberg, JC. (1989). "Position effect variegation in *Drosophila*: towards a genetics of chromatin assembly". *Bioessays* Jul;11(1):14-7.
- Evans AL., PA. Mena y BF. McAllister (2007). "Positive selection near an inversion breakpoint on the neo-X chromosome of *Drosophila americana*". *Genetics* 177: 1303–1319.
- Evgen'ev, M. B., H. Zelentsova, H. Poluectova, G. T. Lyozin, V. Veleikodvorskaja, K. I. Pyatkov, L. A. Zhivotovsky y M. G. Kidwell (2000). "Mobile elements and chromosomal evolution in the *virilis* group of *Drosophila*." *Proc. Natl. Acad. Sci. USA* 97(21): 11337-42.
- Eyre-Walker, A., P. D. Keightley, N. G. Smith y D. Gaffney (2002). "Quantifying the slightly deleterious mutation model of molecular evolution." *Mol Biol Evol* 19(12): 2142-9.
- Feschotte, C. y SR. Wessler (2001). "Treasures in the attic: rolling circle transposons discovered in eukaryotic genomes". *Proc. Natl. Acad. Sci. USA* 98:8923–24.
- Feschotte, C. y EJ. Pritham (2007). "DNA transposons and the evolution of eukaryotic Genomes". *Annu. Rev. Genet.* 41: 331–368.
- Finnegan, DJ. (1989). "Eukaryotic transposable elements and genome evolution". *Trends. Genet.* Apr;5(4):103-7.
- Fischer, G., S. A. James, I. N. Roberts, S. G. Oliver y E. J. Louis (2000). "Chromosomal evolution in *Saccharomyces*." *Nature* 405(6785): 451-4.
- Fontdevila A., Ruiz A., Alonso G., and Ocaña J. (1981) The evolutionary history of *Drosophila buzzatii*. I. Natural Chromosomal polymorphism in colonized populations of the Old World. *Evolution* 35: 148-157.
- Gilbert, D.G. (2005). "*DroSpeGe*, base de datos pública de los genomas secuenciados del genero *Drosophila*".
- Girard, L. y M. Freeling (1999). "Regulatory changes as a consequence of transposon insertion". *Dev. Genet.* 25:291–96.

Gladyshev, EA. y IR. Arkhipova (2007). "Telomere-associated endonuclease-deficient *Penelope*-like retroelements in diverse eukaryotes". Proc. Natl. Acad. Sci. USA May 29;104(22):9352-7.

Goffeau, A., B. G. Barrell, H. Bussey, R. W. Davis, B. Dujon, H. Feldmann, F. Galibert, J. D. Hoheisel, C. Jacq, M. Johnston, E. J. Louis, H. W. Mewes, Y. Murakami, P. Philippsen, H. Tettelin y S. G. Oliver (1996). "Life with 6000 genes." Science 274(5287): 546, 563-7.

Goldman, AS. y M. Lichten (1996). "The efficiency of meiotic recombination between dispersed sequences in *Saccharomyces cerevisiae* depends upon their chromosomal location". Genetics Sep;144(1):43-55.

González J., Nefedov M., Bosdet I., Casals F., Calvete O., Delprat A., Shin H., Chiu R., Mathewson C., Wye N., Hoskins R. A., Schein J. E., de Jong P., Ruiz A. (2005). "A BAC-based physical map of the *Drosophila buzzatii* genome". Genome Research 15: 885-892.

González, J., F. Casals y A. Ruiz (2007). "Testing chromosomal phylogenies and inversion breakpoint reuse in *Drosophila*". Genetics 175(1):167-77.

Goodier, JL y HH. Jr. Kazazian HH (2008). "Retrotransposons revisited: the restraint and rehabilitation of parasites". Cell Oct 3;135(1):23-35.

Gray, YH., MM. Tanaka y JA. Sved (1996). "*P*-element-induced recombination in *Drosophila melanogaster*: hybrid element insertion." Genetics 144(4): 1601-10.

Gray, YH. (2000). "It takes two transposons to tango: transposable-element-mediated chromosomal rearrangements". Trends Genet. Oct;16(10):461-8. Review.

Grewal, SI. y S. Jia (2007). "Heterochromatin revisited". Nat. Rev. Genet. 8:35-46.

Gu, W., F. Zhang y JR. Lupski (2008). "Mechanisms for human genomic rearrangements". Pathogenetics Nov 3;1(1):4.

- Hansen, KD., LF. Lareau, M. Blanchette, RE. Green RE, Q. Meng, J. Rehwinkel, FL. Gallusser, E. Izaurralde, DC. Rio, S. Dudoit y SE. Brenner (2009). "Genome-wide identification of alternative splice forms down-regulated by nonsense-mediated mRNA decay in *Drosophila*". PLoS Genet. Jun;5(6):e1000525.
- Hartl, DL. (1970). "Analysis of a general population genetic model of meiotic drive". Evolution 24: 538–545.
- Hartl, DL. (1997). "Mechanism of a case of genetic coadaptation in populations of *Drosophila melanogaster*". Proc. Natl. Acad. Sci. U S A Jan;74(1):324-8.
- Hasson, E. y WF. Eanes (1996). "Contrasting histories of three gene regions associated with *In(3L)Payne* of *Drosophila melanogaster*". Genetics Dec;144(4):1565-75.
- Hedges, DJ. y PL. Deininger (2007). "Inviting instability: Transposable elements, double-strand breaks, and the maintenance of genome integrity". Mutat. Res. Mar 1;616(1-2): 46-59.
- Henikoff, S. (1990). "Position-effect variegation after 60 years". Trends Genet. Dec;6(12):422-6.
- Hoffmann, AA. y LH. Rieseberg (2008). "Revisiting the Impact of Inversions in Evolution: From Population Genetic Markers to Drivers of Adaptive Shifts and Speciation?". Annu. Rev. Ecol. Evol. Syst. Dec 1;39:21-42.
- Jiang, N., Z. Bao, X. Zhang, SR. Eddy y SR. Wessle (2004). "Pack-*MULE* transposable elements mediate gene evolution in plants". Nature 431:569–73.
- Juretic, N., DR. Hoen, ML. Huynh, PM. Harrison y TE. Bureau (2005). "The evolutionary fate of *MULE*-mediated duplications of host gene fragments in rice". Genome Res. 15:1292–97.
- Jurka J., VV. Kapitonov, A. Pavlicek, P. Klonowski, O. Kohany O y J. Walichiewicz (2005). "Repbase Update, a database of eukaryotic repetitive elements". Cytogenet. Genome Res. 110(1-4):462-7. Review.

Kaminker, JS., CM. Bergman, B. Kronmiller, J. Carlson, R. Svirskas, S. Patel, E. Frise, DA. Wheeler, SE. Lewis, GM. Rubin, M. Ashburner y SE. Celniker (2002). "The transposable elements of the *Drosophila melanogaster* euchromatin: a genomics perspective". *Genome Biol.* 3(12).

Kapitonov, V. y J. Jurka (1996). "The age of *Alu* subfamilies". *J. Mol. Evol.* Jan;42(1):59-65.

Kapitonov, V. y J. Jurka (2001). "Rolling-circle transposons in eukaryotes". *Proc. Natl. Acad. Sci. USA* 98:8714–19.

Kapitonov, VV. y J. Jurka (2006). "Self-synthesizing DNA transposons in eukaryotes." *Proc. Natl. Acad. Sci. USA.* Mar 21;103(12):4540-5.

Kazazian, H. H., Jr. (2004). "Mobile elements: drivers of genome evolution." *Science* 303(5664): 1626-32.

Kennington, WJ., L. Partridge y AA. Hoffmann (2006). "Patterns of diversity and linkage disequilibrium within the cosmopolitan inversion *In(3R)Payne* in *Drosophila melanogaster* are indicative of coadaptation". *Genetics* Mar;172(3):1655-63.

Kidwell, MG. (2002). "Transposable elements and the evolution of genome size in eukaryotes". *Genetica*; 115:49-63.

Kidwell, MG. y DR. Lisch (2001). "Perspective: transposable elements, parasitic DNA, and genome evolution". *Evol. Int. J. Org. Evol.* 55:1–24.

Kim JM., S. Vanguri, JD. Boeke, A. Gabriel y DF. Voytas (1998). "Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete *Saccharomyces cerevisiae* genome sequence". *Genome Res.* May;8(5):464-78.

Kirkpatrick, M. y N. Barton (2006). "Chromosome inversions, local adaptation and speciation *Genetics* 173: 419–434.

- Krimbas C.B. (1992). "The inversion polymorphism of *Drosophila subobscura*". pp. 127-220 en *Drosophila inversion polymorphism*. Krimbas C.B. y Powell J.R eds. CRC Press, Boca Raton, FL.
- Krimbas C.B. y Powell J.R. (1992). "Drosophila inversion polymorphism". CRC Press, Boca Raton, FL.
- Lacadena, J. R. (1996) "Citogenética". Editorial Complutense, Madrid.
- Lander ES., LM. Linton, B. Birren, C. Nusbaum, MC. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, International Human Genome Sequencing Consortium *et al.* (2001). "Initial sequencing and analysis of the human genome". *Nature* Aug 2;412(6846):565.
- Langley, CH., E. Montgomery, R. Hudson, N. Kaplan y B. Charlesworth (1988). "On the role of unequal exchange in the containment of transposable element copy number". *Genet. Res.* Dec;52(3):223-35.
- Lee BS., NW. Kim y NR. Rim (2002). "Chromosome inversion—environment relationships in Korean populations of *Drosophila melanogaster*". *Korean J. Genet.* 24:9–20.
- Leitch, A. R., T. Schwarzacher, D. Jackson y I. J. Leitch (1994). "*In situ* hybridization: a practical guide". Bios Scientific Publishers, Oxford.
- Lemeunier, F. y Aulard, S. (1992) "Inversion polymorphism in *Drosophila melanogaster*". *Drosophila Inversion Polymorphism*. Eds. C. B. Krimbas y J. R. Powell. CRC Press, Boca Ratón, FL. pp. 339-405.
- Liberman, U. (1976). "Modifier theory of meiotic drive: Is Mendelian segregation stable?". *Theor. Popul. Biol.* 10: 127–132.
- Librado, P. y J. Rozas (2009). "DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* Jun 1;25(11):1451-2.

Lim, JK. y MJ. Simmons (1994). "Gross chromosome rearrangements mediated by transposable elements in *Drosophila melanogaster*". *BioEssays* 16:269–75.

Linheiro, RS. y CM. Bergman (2008). "Testing the palindromic target site model for DNA transposon insertion using the *Drosophila melanogaster* P-element". *Nucleic Acids Res.* Nov;36(19):6199-208.

Lippman, Z., AV. Gendrel, M. Black, MW. Vaughn, N. Dedhia *et al.* (2004). "Role of transposable elements in heterochromatin and epigenetic control. *Nature* 430:471–76.

Lisch, D. (2002). "Mutator transposons". *Trends Plant Sci.* 7:498-504.

Lobachev, KS., BM. Shor, HT. Tran, W. Taylor, JD. Keen, MA. Resnick y DA. Gordenin (1998). "Factors affecting inverted repeat stimulation of recombination and deletion in *Saccharomyces cerevisiae*". *Genetics* Apr;148(4):1507-24.

Lundgren, J., P. Masson, Z. Mirzaei y P. Young (2005). "Identification and characterization of a *Drosophila* proteasome regulatory network". *Mol Cell Biol.* Jun;25(11):4662-75.

Lyttle, T. W. y D. S. Haymer (1992). "The role of the transposable element hobo in the origin of endemic inversions in wild populations of *Drosophila melanogaster*." *Genetica* 86(1-3): 113-26.

Macdonald, SJ y AD. Long (2006). "Fine scale structural variants distinguish the genomes of *Drosophila melanogaster* and *D. pseudoobscura*". *Genome Biol.* 7(7):R67.

Marchler-Bauer A., JB. Anderson, F. Chitsaz, MK. Derbyshire, C. DeWeese-Scott, JH. Fong, LY. Geer, RC. Geer, NR. Gonzales, M. Gwadz, S. He, DI. Hurwitz, JD. Jackson, Z. Ke, CJ. Lanczycki, CA. Liebert, C. Liu, F. Lu, S. Lu, GH. Marchler, M. Mullokandov, JS. Song, A. Tasneem, N. Thanki, RA. Yamashita, D. Zhang, N. Zhang y SH. Bryant (2009). "CDD: specific functional annotation with the Conserved Domain Database". *Nucleic Acids Res.* Jan;37:D205-10.

- Mathiopoulos, K. D., A. della Torre, V. Predazzi, V. Petrarca y M. Coluzzi (1998). "Cloning of inversion breakpoints in the *Anopheles gambiae* complex traces a transposable element at the inversion junction." *Proc. Natl. Acad. Sci. USA* 95(21): 12444-9.
- Matzkin, L. M., T. J. Merritt, C. T. Zhu y W. F. Eanes (2005). "The structure and population genetics of the breakpoints associated with the cosmopolitan chromosomal inversion *In(3R)Payne* in *Drosophila melanogaster*." *Genetics* 170(3): 1143-52.
- McClintock, B. (1984). "The significance of responses of the genome to challenge." *Science* 226(4676): 792-801.
- Mettler, LE., RA. Voelker y T. Mukai (1977). "Inversion clines in populations of *Drosophila melanogaster*". *Genetics* Sep;87(1):169-176.
- Montgomery, E., B. Charlesworth y C. H. Langley (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: 39-41.
- Montgomery, E., SM. Huang, CH. Langley y BH. Judd (1991). "Chromosome rearrangement by ectopic recombination in *Drosophila melanogaster*: genome structure and evolution". *Genetics* Dec;129(4):1085-98.
- Muller, J. H. (1940). "Bearings of the *Drosophila* work on systematics". *The New Systematics*. Ed. J. Huxley. Clarendon Press, Oxford. pp. 185-268.
- Munté, A., J. Rozas, M. Aguadé y C. Segarra (2005). "Chromosomal inversion polymorphism leads to extensive genetic structure: a multilocus survey in *Drosophila subobscura*". *Genetics* Mar;169(3):1573-81.
- Mural, R. J., M. D. Adams, E. W. Myers, H. O. Smith, G. L. Miklos, R. Wides, A. Halpern, P. W. Li, G. G. Sutton, J. Nadeau, S. L. Salzberg, R. *et al.* (2002). "A comparison of whole-genome shotgun-derived mouse chromosome 16 and the human genome." *Science* 296(5573): 1661-71.

Murphy, WJ., DM. Larkin , A. Everts-van der Wind, G. Bourque, G. Tesler, L. Auvil, JE. Beever, BP. Chowdhary, F. Galibert, L. Gatzke *et al.* (2005). "Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps". *Science* Jul 22;309(5734):613-7.

Nadeau, J. H. y B. A. Taylor (1984). "Lengths of chromosomal segments conserved since divergence of man and mouse." *Proc. Natl. Acad. Sci. USA* 81(3): 814-8.

Navarro, A., E. Betrán, A. Barbadilla y A. Ruiz (1997). "Recombination and gene flux caused by gene conversion and crossing over in inversion heterokaryotypes". *Genetics* Jun;146(2):695-709.

Negre, B., S. Casillas, M. Suzanne, E. Sánchez-Herrero, M. Akam, M. Nefedov, A. Barbadilla, P. de Jong y A. Ruiz (2005). "Conservation of regulatory sequences and gene expression patterns in the disintegrating *Drosophila Hox* gene complex". *Genome Res.* May;15(5):692-700.

Negre, B. y A. Alfredo Ruiz (2007)" HOM-C evolution in *Drosophila*: is there a need for Hox gene clustering?" *Trends. Genet.* Feb;23(2):55-9. Epub 2006 Dec 26.

Nickerson, E. y D. L. Nelson (1998). "Molecular definition of pericentric inversion breakpoints occurring during the evolution of humans and chimpanzees." *Genomics* 50(3): 368-72.

O'Hare, K. y G. M. Rubin (1983). "Structure of *P* transposable elements and their sites of insertion and excision in the *Drosophila melanogaster* genome". *Cell* 34: 25-35.

Ohler, U., GC. Liao, H. Niemann y GM. Rubin (2002). "Computational analysis of core promoters in the *Drosophila* genome". *Genome Biol.* 3(12):research0087.

Ohno, S. (1973). "Ancient linkage groups and frozen accidents". *Nature.* 1973;244:259–262

Olvera, O. J., Powell, J. R., de la Rosa, M. E., Salceda, V. M., Gaso, M. I., Guzman, J., Anderson, W. W. y Levine, L. (1979). "Population genetics of Mexican *Drosophila*. VI.

- Cytogenetic aspects of the inversion polymorphism in *Drosophila pseudoobscura*". Evolution 33: 381-395.
- Orgel, L. E. y F. H. Crick (1980). "Selfish DNA: the ultimate parasite." Nature 284(5757): 604-7.
- Ostertag, E.M. y H.H. Jr. Kazazian (2001). "Biology of mammalian *LI* retrotransposons". Annu. Rev. Genet. 35:501–538.
- Painter, TS. (1934). "A new method for the study of chromosome aberrations and the plotting of chromosome maps in *Drosophila melanogaster*". Genetics May;19(3):175-88.
- Parra, G., E. Blanco y R. Guigó (2000). "GenenID in *Drosophila*". Genome Res. Apr;10(4):511- 5.
- Patterson, J.T y M. R. Wheeler (1942) "Description of new species of the subgenera *Hirtodrosophila* and *Drosophila*". Univ. Texas Publ. 4213: 67-109.
- Peng, Q., P. A. Pevzner y G. Tesler (2006). "The fragile breakage versus random breakage models of chromosome evolution." PLoS Comput Biol 2(2): e14.
- Petes, T. D. y C. W. Hill (1988). "Recombination between repeated genes in microorganisms." Annu Rev Genet 22: 147-68.
- Petrov DA., YT. Aminetzach, JC. Davis, D. Bensasson y AE. Hirsh (2003). "Size matters: non-LTR retrotransposable elements and ectopic recombination in *Drosophila*". Mol. Biol. Evol. Jun;20(6):880-92.
- Pevzner, P. y G. Tesler (2003a). "Genome rearrangements in mammalian evolution: lessons from human and mouse genomes." Genome Res 13(1): 37-45.
- Pevzner, P. y G. Tesler (2003b). "Human and mouse genomic sequences reveal extensive breakpoint reuse in mammalian evolution." Proc. Natl. Acad. Sci. USA 100(13): 7672-7.

Pfeiffer, P., W. Goedecke y G. Obe (2000). "Mechanisms of DNA double-strand break repair and their potential to induce chromosomal aberrations". *Mutagenesis* Jul;15(4):289-302.

Piegu, B., R. Guyo, N. Picault, A. Roulin, A. Saniyal *et al.* (2006). "Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice". *Genome Res.* 16:1262–1269.

Plasterk, RH. (1995). "Reverse genetics: from gene sequence to mutant worm". *Methods Cell Biol.* 48:59-80.

Powell, J. R. (1992). "Inversion polymorphism in *Drosophila pseudoobscura* and *Drosophila persimilis*". *Drosophila inversion polymorphism*. Eds. C. B. Krimbas y J. R. Powell. CRC Press, Boca Ratón, FL. pp. 73-126.

Powell, J. R. (1997). "Progress and Prospects in Evolutionary Biology: The *Drosophila* Model". Oxford University Press, New York.

Prakash. S. y RC. Lewontin (1968). "A molecular approach to the study of genic heterozygosity in natural populations. 3. Direct evidence of coadaptation in gene arrangements of *Drosophila*". *Proc. Natl. Acad. Sci. USA* Feb;59(2):398-405.

Prazeres da Costa, O., J. González y A. Ruiz (2009). "Cloning and sequencing of the breakpoint regions of inversion 5g fixed in *Drosophila buzzatii*". *Chromosoma* Jun;118(3):349-60.

Prevosti, A., G. Ribo, L. Serra, M. Aguadé, J. Balanyà, M. Monclus y F. Mestres (1988). "Colonization of America by *Drosophila subobscura*: Experiment in natural populations that supports the adaptive role of chromosomal-inversion polymorphism". *Proc. Natl. Acad. Sci. USA*. Aug;85(15):5597-5600.

Pritham, EJ., T. Putliwala y C. Feschotte (2007). "Mavericks, a novel class of giant transposable elements widespread in eukaryotes and related to DNA viruses". *Gene* 390:3–17.

- Pritham, EJ. y C. Feschotte (2007). "Massive amplification of rolling-circle transposons in the lineage of the bat *Myotis lucifugus*". Proc. Natl. Acad. Sci. USA 104:1895–1900.
- Puig, M., M. Cáceres y A. Ruiz (2004). "Silencing of a gene adjacent to the breakpoint of a widespread *Drosophila* inversion by a transposon-induced antisense RNA". Proc. Natl. Acad. Sci. USA Jun 15;101(24):9013-8.
- Rafael V. y G. Arcos (1989). "Subgrupo *inca*, un nuevo subgrupo del grupo *repleta*, con descripción de *Drosophila huancavilcae* N.SP. (Diptera, *Drosophilidae*)". Evol. Biol. 3:233-243.
- Ranz, J. M., C. Segarra y A. Ruiz (1997). "Chromosomal homology and molecular organization of Muller's elements D and E in the *Drosophila repleta* species group." Genetics 145(2): 281-95.
- Ranz, J. M., F. Casals y A. Ruiz (2001). "How malleable is the eukaryotic genome? Extreme rate of chromosomal rearrangement in the genus *Drosophila*." Genome Res 11(2): 230-9.
- Ranz, JM., D. Maurin, YS. Chan, M. von Grotthuss, LW. Hillier, J. Roote, M. Ashburner y CM. Bergman (2007). "Principles of genome evolution in the *Drosophila melanogaster* species group". PLoS Biol. Jun;5(6):e152.
- Reese, MG. (2001). "Application of a time-delay neural network to promoter annotation in the *Drosophila melanogaster* genome". Comput. Chem. 26(1),51-6.
- Regner, L. P., M. S. Pereira, C. E. Alonso, E. Abdelhay y V. L. Valente (1996). "Genomic distribution of *P* elements in *Drosophila willistoni* and a search for their relationship with chromosomal inversions." J Hered 87(3): 191-8.
- Rehwinkel, J., I. Letunic, J. Raes, P. Bork y E. Izaurralde (2005). "Nonsense-mediated mRNA decay factors act in concert to regulate common mRNA targets". RNA Oct;11(10):1530-44.
- Richards, S. Y. Liu, BR. Bettencourt, P. Hradecky, S. Letovsky, R. Nielsen, K. Thornton, MJ. Hubisz, R. Chen, RP. Meisel RP *et al.* (2005). "Comparative genome sequencing of

Drosophila pseudoobscura: chromosomal, gene, and cis-element evolution". Genome Res. 15(1):1-18.

Rieseberg, LH. (2001). "Chromosomal rearrangements and speciation". Trends. Ecol. Evol. 16: 351–358.

Rouse, J. y SP. Jackson (2002). "*Lcd1p* recruits *Mec1p* to DNA lesions *in vitro* and *in vivo*". Mol. Cell. Apr;9(4):857-69.

Ruiz, A., M. Santos, A. Barbadilla, JE. Quezada-Diaz, E. Hasson y A. Fontdevila (1997). "Genetic variance for body size in a natural population of *Drosophila buzzatii*". Genetics 128: 739-750.

Ruiz, A. y A. Fontdevila (1981). "Ecología y evolución del subgenero *mulleri* de *Drosophila* en Venezuela y Colombia". Acta Científica Venezolana 32: 338-345.

Ruiz, A. y M. Wasserman (1993). "Evolutionary cytogenetics of the *Drosophila buzzatii* species complex." Heredity 70 (Pt 6): 582-96.

Ruiz-Herrera, A., J. Castresana y TJ. Robinson (2006). "Is mammalian chromosomal evolution driven by regions of genome fragility?". Genome Biol. 7(12):R115.

Ruiz-Herrera A. y TJ. Robinson (2007). "Chromosomal instability in Afrotheria: fragile sites, evolutionary breakpoints and phylogenetic inference from genome sequence assemblies". BMC Evol. Biol. Oct 24;7:199.

Runcie DE. y MA. Noor (2009). "Sequence signatures of a recent chromosomal rearrangement in *Drosophila mojavensis*". Genetica May;136(1):5-11.

Russo CA., Takezaki N. y Nei M. (1995). "Molecular phylogeny and divergence times of *drosophilid* species". Mol Biol Evol.;12(3):391-404.

Saitou N. y M. Nei (1987). "The neighbor-joining method: a new method for reconstructing phylogenetic trees".Mol. Biol. Evol. Jul;4(4):406-25.

- Sambrook, J., E. F. Fritsch y T. Maniatis (1989). "Molecular cloning, a laboratory manual". Cold Spring Harbor Laboratory Press, New York.
- Sankoff, D. (1999). "Comparative mapping and genome rearrangement". En, From Jay L. Lush to Genomics: Visions for Animal Breeding and Genetics. (Dekkers, J.C.M. *et al.* eds), pp 124-134, CAB International.
- Sankoff, D. (2009). "The where and wherefore of evolutionary breakpoints". *J. Biol.*8:66
- Schaeffer, SW., MP. Goetting-Minesky, M. Kovacevic, JR. Peoples, JL. Graybill, JM. Miller, K. Kim, JG. Nelson y WW. Anderson (2003). "Evolutionary genomics of inversions in *Drosophila pseudoobscura*: evidence for epistasis". *Proc. Natl. Acad. Sci. USA* Jul 8;100(14):8319-24.
- Schaeffer, SW., A. Bhutkar, BF. McAllister, M. Matsuda, LM. Matzkin, PM. O'Grady, C. Rohde, VL. Valente, M. Aguadé, WW. Anderson *et al.* (2008). "Polytene chromosomal maps of 11 *Drosophila* species: the order of genomic scaffolds inferred from genetic and physical maps". *Genetics* Jul;179(3):1601-55.
- Schneuwly, S., R. Klemenz y W. J. Gehring (1987). "Redesigning the body plan of *Drosophila* by ectopic expression of the homoeotic gene *Antennapedia*." *Nature* 325(6107): 816-8.
- Schwartz, A., DC. Chan, LG. Brown, R. Alagappan, D. Pettay, C. Disteché, B. McGillivray, A. de la Chapelle y DC. Page (1998). "Reconstructing hominid Y evolution: X-homologous block, created by X-Y transposition, was disrupted by Yp inversion through *LINE-LINE* recombination". *Hum. Mol. Genet.* Jan;7(1):1-11.
- Schwartz, M., E. Zlotorynski y B. Kerem (2006). "The molecular basis of common and rare fragile sites". *Cancer Lett.* Jan 28;232(1):13-26.
- Sharakhov, I. V., B. J. White, M. V. Sharakhova, J. Kayondo, N. F. Lobo, F. Santolamazza, A. Della Torre, F. Simard, F. H. Collins y N. J. Besansky (2006). "Breakpoint structure reveals the unique origin of an interspecific chromosomal inversion (*2La*) in the *Anopheles gambiae* complex." *Proc. Natl. Acad. Sci. USA* 103(16): 6258-62.

Singh, ND. y D. Petrov (2004). "Rapid sequence turnover at an intergenic locus in *Drosophila*". *Mol. Biol. Evol.* Apr;21(4):670-80.

Sitko, P. O. (1938). "Mutation and aberration in inverted X-chromosomes". *Memoirs of Genetics, Acad. Sci. Ukr. SSR* 2:3-31.

Slotkin, RK. y R. Martienssen (2007). "Transposable elements and the epigenetic regulation of the genome". *Nat. Rev. Genet.* 8:272–285.

Sorsa, V. (1988). "Polytene Chromosomes in Genetic Research". John Wiley & Sons, New York.

Sperlich, D. (1963). "Chromosomal polymorphism and position effect in *Drosophila subobscura*". *Proc. 11th Int. Congr. Genet.* 1:5.

Sperlich D. y Pfriem P. (1986). "Chromosomal polymorphism in natural and experimental populations". *The genetics and Biology of Drosophila* Vol. 3 ed Academic Press, NY, pp 257-309. Eds. Ashburner M., Carson H.L. and Thompson Jr., J.N.

Spofford, JB. (1976). "Position effect variegation in *Drosophila*". In: Ashburner, M. y E. Novitski (eds). *The genetics and biology of Drosophila*, vol 1C. Academic Press, London New York San Francisco, pp 955±1018.

Spradling AC. y GM. Rubin (1981). "Drosophila genome organization: conserved and dynamic aspects". *Annu. Rev. Genet.*15:219-64.

Stalker, HD. (1976). "Chromosome studies in wild populations of *D. melanogaster*". *Genetics* Feb;82(2):323-47.

Stalker, HD. (1980). "Chromosome studies in wild populations of *Drosophila melanogaster* II. Relationship of inversion frequencies to latitude, season, wing-loading and flight activity". *Genetics* May;95(1):211-223.

- Stankiewicz, P. y JR. Lupski (2002). "Genome architecture, rearrangements and genomic disorders". *Trends Genet* Feb;18(2):74-82.
- Stankiewicz, P., CJ. Shaw, JD. Dapper, K. Wakui, LG. Shaffer, M. Withers, L. Elizondo, SS. Park y JR. Lupski (2003). "Genome architecture catalyzes nonrecurrent chromosomal rearrangements". *Am. J. Hum. Genet.* May;72(5):1101-16.
- Stone, W. S., W. C. Guest y F. D. Wilson (1960). "The Evolutionary Implications of the Cytological Polymorphism and Phylogeny of the *Virilis* Group of *Drosophila*." *Proc. Natl. Acad. Sci. USA* 46(3): 350-61.
- Sturtevant, A. H. (1917). "Genetic Factors Affecting the Strength of Linkage in *Drosophila*." *Proc. Natl. Acad. Sci. USA* 3(9): 555-8.
- Sturtevant, A. H. (1926). "A crossover reducer in *Drosophila melanogaster* due to inversion of a section of the third chromosome". *Biol. Zentralbl.* 46: 697-702.
- Sturtevant, A. H. y Plunkett, C. R. (1926). "Sequence corresponding third chromosome genes in *Drosophila melanogaster* and *Drosophila simulans*". *Biol. Bull.* 50: 56-60.
- Sturtevant, A. H. y T. Dobzhansky (1936). "Inversions in the Third Chromosome of Wild Races of *Drosophila pseudoobscura*, and Their Use in the Study of the History of the Species." *Proc. Natl. Acad. Sci. USA* 22(7): 448-50.
- Tamura, K., S. Subramanian y S. Kumar (2004). "Temporal patterns of fruit fly (*Drosophila*) evolution revealed by mutation clocks". *Mol. Biol. Evol.* Jan;21(1):36-44.
- Tatusova, TA. y TL. Madden (1999). "BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences". *FEMS Microbiol. Lett.* May 15;174(2):247-50.
- Tidon-Sklorz, R. y F. M. Sene (2001). "Two new species of the *Drosophila serido* sibling set (*Diptera, Drosophilidae*)". *Iheringia, Sér. Zool.* May no.90. Porto Alegre.

Tolchkov, EV., MD. Balakireva y VE. Alatortsev (1984). "Inactivation of the X chromosome region with a known fine genetic structure as a result of the variegated position effect in *Drosophila melanogaster*". *Genetika* 20:1846±1856.

Tolchkov, EV., IA. Kramerova, SA. Lavrov, VI. Rasheva, S. Bonaccorsi, VE. Alatortsev y VA. Gvozdev (1997). "Position-effect variegation in *Drosophila melanogaster* X chromosome inversion with a breakpoint in a satellite block and its suppression in a secondary rearrangement". *Chromosoma* Dec;106(8):520-5.

Tonzetich, J., T. W. Lyttle y H. L. Carson (1988). "Induced and natural break sites in the chromosomes of Hawaiian *Drosophila*." *Proc. Natl. Acad. Sci. USA* 85(5): 1717-21.

Tweedie, S., M. Ashburner, K. Falls, P. Leyland, P. McQuilton, S. Marygold, G. Millburn, D. Osumi-Sutherland, A. Schroeder, R. Seal, H. Zhang y FlyBase Consortium (2009). "FlyBase: enhancing *Drosophila* gene ontology annotations". *Nucleic. Acids Res.* Jan;37(Database issue):D555-9.

Vigdal, TJ., CD. Kaufman, Z. Izsvák, DF. Voytas y Z. Ivics (2002). "Common physical properties of DNA affecting target site selection of *sleeping beauty* and other *Tc1/mariner* transposable elements. *J. Mol. Biol.* Oct 25;323(3):441-52.

Volff, JN (2006). "Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes". *BioEssays* 28:913–922.

Wasserman, M. (1982). "Evolution of the *repleta* group". The genetics and biology of *Drosophila*. Editado por Ashburner, M., H.L. Carson y J. N. Thompson. Academic Press, New York, vol 3b, pp.61-139.

Wasserman M. (1992). "Cytological evolution of the *Drosophila repleta* species group". *Drosophila* inversion polymorphism. Edited by C.B. Krimbas and J.R. Powell. RC Press, Boca Raton, FL, pp. 455-52.

- Weeks, AR., S. W. McKechnie y AA. Hoffmann (2002). "Dissecting adaptive clinal variation: markers, inversions and size/stress associations in *Drosophila melanogaster* from a central field population". *Ecol. Lett.* 5: 756–763.
- Weiler, KS. y BT. Wakimoto (1995). "Heterochromatin and gene expression in *Drosophila*". *Annu. Rev. Genet.* 29:577-605.
- Wesley, CS. y WF. Eanes (1994). "Isolation and analysis of the breakpoint sequences of chromosome inversion *In(3L)Payne* in *Drosophila melanogaster*." *Proc. Natl. Acad. Sci. USA* 91(8): 3132-6.
- Witte CP., QH. Le, T. Bureau y A. Kumar (2001). "Terminal-repeat retrotransposons in miniature (TRIM) are involved in restructuring plant genomes". *Proc. Natl. Acad. Sci. USA* Nov 20;98(24):13778-83.
- White, M. J. D. (1978). "Modes of Speciation". W. H. Freeman and Company, San Francisco.
- Wu, X., y S. M. Burgess (2004). "Integration target site selection for retroviruses and transposable elements". *CMLS, Cell. Mol. Life Sci.* 61:2588–2596.
- Xing, J., H. Wang, VP. Belancio, R. Cordaux, P. L. Deininger y MA. Batzer (2006). "Emergence of primate genes by retrotransposon-mediated sequence transduction". *Proc. Natl. Acad. Sci. USA* 103:17608–17613.
- Yamaguchi, O. y T. Mukai (1974). "Variation of spontaneous occurrence rates of chromosomal aberrations in the second chromosomes of *Drosophila melanogaster*". *Genetics* Dec;78(4):1209-21.
- Yoon, JS. y RH. Richardson (1978). "A mechanism of chromosomal rearrangements: The role of heterochromatin and ectopic joining". *Genetics* Feb;88(2):305-316.
- Yoshida, J., K. Umezumi y H. Maki (2003). "Positive and negative roles of homologous recombination in the maintenance of genome stability in *Saccharomyces cerevisiae*." *Genetics* 164(1): 31-46.

Zelentsova, H., H. Poluectova, L. Mnjoian, G. Lyozin, V. Veleikodvorskaja, L. Zhivotovsky, MG. Kidwell y MB. Evgen'ev (1999). "Distribution and evolution of mobile elements in the *virilis* species group of *Drosophila*". *Chromosoma*. Dec;108(7):443-56.

Zhang, P. y AC. Spradling (1994). "Insertional mutagenesis of *Drosophila* heterochromatin with single *P* elements". *Proc. Natl. Acad. Sci. USA* Apr 26;91(9):3539-43.

Zhang, J. y T. Peterson (2005). "A segmental deletion series generated by sister-chromatid transposition of *Ac* transposable elements in maize". *Genetics* Sep Vol. 171, 333-344.

Zhang, J., C. Yu, V. Pulletikurti, J. Lamb, T. Danilova, DF. Weber J. Birchler y T. Peterson (2009). "Alternative *Ac/Ds* transposition induces major chromosomal rearrangements in maize". *Genes Dev*. Mar 15;23(6):755-65.

Zheng, D. Y M.B. Gerstein (2007). "The ambiguous boundary between genes and pseudogenes: the dead rise up, or do they?". *Trends in Genetics* Volume 23, Issue 5, 219-224.

Zhou, ZH., E. Akgün y M. Jasin (2001). "Repeat expansion by homologous recombination in the mouse germ line at palindromic sequences". *Proc. Natl. Acad. Sci. USA* Jul 17;98(15):8326-33.

Zhou, L., R. Mitra, P.W. Atkinson, A. B. Hickman, F. Dyda y N. L. Craig (2004). "Transposition of *hAT* elements links transposable elements and V(D)J recombination". *Nature* 432: 995–1001.

Páginas web/programas on line:

DNAsp: <http://www.ub.edu/dnasp/>

BLAST: <http://blast.ncbi.nlm.nih.gov/Blast.cgi>

MUSCLE: <http://www.ebi.ac.uk/Tools/muscle/index.html>

McPromoter : <http://tools.igsp.duke.edu/generegulation/McPromoter/>

GENEID: <http://genome.imim.es/geneid.html>

GENSCAN: <http://genes.mit.edu/GENSCAN.html>

RepeatMasker: <http://repeatmasker.org>